Financial and other support from Environment Canada is gratefully acknowledged.
Impressum

Copyright: the copyright of all papers is an exclusive right of the authors. No work can be reproduced without written permission of the authors.

Each session paper and the papers in workshop W12 in these proceedings was refereed by a member of the Editorial Board and two anonymous referees. Each workshop paper was refereed by a member of the Editorial Board.

ISBN: 978-88-9035-741-1

Published by the International Environmental Modelling and Software Society (iEMSs)
President: Alexey Voinov. Address: c/o IDSIA, Galleria 2, 6928 Manno, Switzerland.
E-mail: president@iemss.org. Website: http://www.iemss.org
Printed in Guelph, Canada

Date of Publication: December 28, 2010
Acknowledgements

The organizers are grateful for the support of The University of Guelph, the iEMSs executive, session and workshop organizers, and the more than 200 referees, who made this undertaking possible.

The Conference Organizers recognize the major financial support of the United States Geological Survey—Climate and Land Use Change Branch. Their support contributed to the success of the workshop W1. Effects of climate change: landscape modelling, providing decision support, and understanding uncertainties, organized by: Richard S. Sojda, Thomas C. Edwards, Karina Gibert, Mark Borsuk, Carmel Pollino and Tony Jakeman, and indirectly led to the development of several other climate change workshops and sessions.

The unexpectedly large response to the conference (50% more attendees than originally planned for) is a testament to the enthusiasm of the iEMSs community. The professionalism shown by the Westin staff and the staff at the various venues was praiseworthy.

The student volunteers from Endicott College (Daralyn Klemansky and Tony Gatto) showed panache in carrying out their respective tasks.

Thanks to everyone for their keen participation.
iEMSs 2010 ORGANIZATION

HONORARY CHAIR

John Lawrence
Director, Aquatic Ecosystem Management Research
Environment Canada

CONVENER AND GENERAL CHAIR

David A. Swayne
School of Computer Science
University of Guelph, Canada

VICE-CHAIRS – CONTRIBUTED TECHNICAL SESSIONS

Andrea Rizzoli
Istituto Dalle Molle di Studi sull’Intelligenza Artificiale
University of Lugano, Switzerland

Yongbo Liu
Department of Geography
University of Guelph, Canada

VICE-CHAIRS – WORKSHOPS

Alexey Voinov
International Institute for Geo-information Science and Earth Observation
Enschede, The Netherlands

Wanhong Yang
Department of Geography
University of Guelph, Canada

VICE-CHAIR – PEER REVIEWS

Tatiana Filatova
Water Engineering & Management
University of Twente,
The Netherlands

VICE-CHAIRS LIAISON

ANZ/ASIA

Susan Cuddy,
CSIRO Australia

UNIVERSITIES/CANADA

Aziz Guergachi
Ryerson University, Canada

VICE-CHAIR FINANCES

Jaroslav Mysiak
Fondazione Eni Enrico Mattei, Italy

TECHNICAL EDITOR

Simon Hood
School of Computer Science
University of Guelph, Canada
VICE-CHAIR LOCAL ARRANGEMENTS OTTAWA

Linda Robson
Endicott College, USA

INTERNATIONAL PROGRAM COMMITTEE

David Anderson, Canada                  Giovanni Molina, El Salvador
Robert Argent, Australia                Lachlan Newman, Australia
William Booty, Canada                   Bernd Page, Germany
Ralf Denzer, Germany                    Claudia Pahl Wostl, Germany
Steven Fry sing er, United States       Jaroslav Pokorny, Czech Republic
Giorgio Guariso, Italy                  Nigel Quinn, United States
Jiri Hrbecek, Czech Republic            Jaroslav Racek, Czech Republic
Harvey Johnstone, Canada                Andrea Rizzoli, Switzerland
David Lam, Canada                       Gerald Schmak, Austria
Luis Leon, Canada                       Roberto Soncini-Sessa, Italy
Lindsay Marshall, Canada                Julius Stuller, Czech Republic
Ann McMillan, Canada                    Alexey Voinov, The Netherlands

ORGANIZING COMMITTEE (SESSIONS AND WORKSHOPS)

Georgii Alexandrov, Japan                D. T. Mihailovic, Serbia
Xavier Flores Alsina, Canada             N. Mole, UK
Dan Ames, USA                           Dirk Muschalla, Canada
David Arctur, USA                       Jaroslav Mysiak, Italy
Andrea Atenzi, Italy                    Nick H. Ogden, Canada
Ioannis Athanasiadis, Switzerland        Tim Oxley, UK
Craig Aumann, USA                       Dawn Parker, USA
Eduardo Ayesa, Spain                    Abani Patra, USA
Mark Borsuk, USA                        George Percivall, USA
François Bousquet, France               Ajith H. Perera, Canada
Virginia Brilhante, Brazil              Jean-Michel Perraud, Australia
Lisa J. Buse, Canada                     Werner Pillmann, Germany
David Butler, UK                        Didrik Pinte, Belgium
Andrea Castelletti, Italy               Manel Poch, Spain
B. Chen, China                          Carmen Pollino, Australia
G.Q. Chen, China                        David Post, USA
Hubert Colas, Canada                     Joerg Preiss, Germany
Joachim Comas, Spain                    Nigel Quinn, USA
Ulises Cortés, Spain                    Marco Ratto, Italy
Barry Croke, Australia                  Wolfgang Rauch, Austria
Susan Cuddy, Australia                   Patrick Reed, USA
Wenbin Cui, Canada                      Stefan Reis, UK
Wilfred Cuff, Canada                    Andrea Rizzoli, Switzerland
Phillip Dibner, USA                     Barbara Robson, Australia
C. Ashton Drew, USA                     Ignasi Rodriguez-Roda, Spain
S0. Open Session
Organizers: Simon Hood, David Swayne

If there was no session or workshop which matched your research interests or proposed paper, your submission was entered in this session category.

S00. Tools and techniques for environmental modelling and software
Organizers: David Swayne, Simon Hood

This session, tools and techniques, was intended for papers of a background nature, whether novel uses of algorithms from spatial statistics or applications of supercomputers.

S1. Forest fire modeling and software
Organizer: Wenbin Cui

This session will concentrate on forest fire models and the development of their corresponding software. The models mainly include models of forest fire occurrence, forest fire growth, forest fire behaviour, forest fire effect, and more importantly forest fire regime simulation. There are increasing interests in forest fire modeling recently. There exist many forest fire models in all the areas listed above but software are relatively few because the models are mostly research models and were not (well) designed for other users. Thus many users (researchers) have to develop their own models because the existing models are not flexible enough to meet their particular needs even though they might meet some of their requirements. The goal of the session will be not only exploring forest fire modeling but also combining forest fire modeling and software development.

S2. Environmental fluid mechanics - theoretical, numerical and experimental approaches

Environmental Fluid Mechanics (EFM) is the scientific study of transport, dispersion and transformation processes in natural fluid flows on our planet Earth, from the microscale to the planetary scale. Stratification and turbulence are two essential ingredients of EFM. Stratification occurs when the density of the fluid varies spatially, as in a sea breeze where masses of warm and cold air lie next to each other or in an estuary where fresh river water flows over saline seawater. Turbulence is the term used to characterize the complex, seemingly random motions that continually result from instabilities in fluid flows. Turbulence is ubiquitous in natural fluid flows because of the large scales that these flows typically occupy. The processes studied by EFM greatly affect the quality of natural ecosystems. For this session papers reporting observational, experimental, numerical and theoretical investigations would be welcome. So the Session will be organized in two parts: Theoretical and Numerical aspects (Part 1) and Applicative, Software and Experimental issues (Part 2).

This session could tentatively cover the following topics:

- Diffusion, turbulent dispersion and mixing of environmental contaminants in natural and engineered water systems and in the atmosphere
- Processes at the environmental interfaces in soil, atmosphere and natural waters
- Turbulent flows
- Nonlinear processes in environmental fluid mechanics
- Two-phase and multiphase flows
- Stratified flows
- Transport of water and chemicals in the soil
S3. Modelling of dangerous phenomena and innovative techniques for hazard evaluation, mapping, and mitigation
Organizers: Giulio Iovine, John B. Rundle, David Yuen, Abani Patra

Several types of dangerous phenomena (either natural or man-made) pose serious risk in many parts of the world. Fundamental tasks in hazard evaluation include the prediction of:

- the area influenced by the phenomenon,
- its evolution in space and time, and
- the understanding of the triggering mechanisms.

This section mainly focuses on theoretical and numerical research, especially those supported by computer-assisted techniques of computation. Among the different approaches which might be taken, some mainly focus on the problem of time-sequential movements, by using either physical-based or empirical methods of analysis, while other methods attempt to predict the evolution of a given natural phenomenon. Studies concerning innovative methods of modelling and simulation for hazard mapping and prevention purposes are welcome. Contributions on new techniques of simulation and mapping, case studies, and novel methods of model calibration and validation, as well as on sensitivity analyses, are solicited. Comparative discussions on the potential, and the limits of different modelling approaches, are also within the scope of this session.

S4. The European Environmental Information Space: news and trends
Organizer: Kristina Voigt

The continuous work in the EnviroInfo community which is mainly focussing European science and technologies is closely related to the topics treated in other societies like e.g. the iEMSs (International Environmental Modelling and Software Society), the TIES (The International Environmetrics Society) etc. In the past two years a major effort has been made to intensify the collaboration in Europe by establishing a European Environmental Information Space.

The EU granted project ICT-ENSURE (Information and Communication Technologies - Environmental Sustainability Research) is the leading support action in the area "ICT for Environmental Sustainability Research". The main objectives are to extend the network of environmental sustainability research and to explore the structure and content of European research programmes relevant for sustainable development. Two main topics will be presented: an information system on European research programmes and research projects as well as full text databases on the proceedings' volumes in the field of Environmental Informatics.

Furthermore other European environmental modelling and informatics activities will be given in this session. Concerning the chemicals policy the REACH law and its environmetrical implications will be introduced. REACH is a new European Community Regulation on chemicals and their safe use (EC 1907/2006). It deals with the Registration, Evaluation, Authorisation and Restriction of Chemical substances. In addition, a software assistant to help operators in fulfilling theirs tasks concerning the European Emission Trading System (ETS) is presented. In a different paper the software Umberto I used to model a trigeneration system providing electricity, heating and cooling to a building, revealing important relevant structures and flows. It will be demonstrated in the session that theoretical environmental disciplines like models, information systems, and software tools are of upmost importance in the pursuit of sustainability. A paper on the topic of sustainable chemistry rounds up this session.
S5. Modeling and deciding with stakeholders
Organizers: Alexey Voinov, Raffaele Giordano, Jaroslav Mysiak, Francois Bousquet

Stakeholder engagement, collaboration, or participation, shared learning or fact-finding have become buzz words and hardly any environmental assessment or modeling effort today can be presented without some kind of reference to stakeholders and their involvement in the process. Stakeholder involvement became almost a "must". This is clearly a positive sign, however in far too many cases stakeholders are given only lip service and their engagement turns out to be quite formal. The linked session and workshop will explore the expanding field of participatory modeling and participatory decision analysis.

We intend to contribute to the emerging theory of stakeholder involvement, come up with some classifications and categorization of the multiple efforts in this area, and then focus on various applications and case studies. In particular we will consider the various modeling tools and frameworks that are available and will decide what are the benefits that they present for participatory efforts.

S6. Ecosystem services concept for environmental management
Organizers: Ralf Seppelt, Martin Volk, Ann van Griensven

How can environmental modelling support the implementation of the ecosystem services concept for environmental management? Towards appropriate methods for investigation, assessment, and implementation of ecosystem services in environmental management.

Topics

- Model based quantifying ecosystem services
- Use of ecosystem service modelling in regional case studies for environmental management
- Who is using it where? Examples for the implementation of the ecosystem services concept in environmental management (under stakeholder involvement)

The ecosystem services concept enables development of policies that integrate social, economic, and ecological perspectives. Integrated environmental models are prone to support this concept as those can integrate several ecological functions, support the analysis of trade-offs and have been developed to a state where stakeholders can be involved and management solution can be derived.

Thus in this session we seek for regional studies and model development, that support a biophysical founded assessment of ecosystem services. Thus this session will be interdisciplinary firstly by discussion studies focusing on abiotic (water provisioning, regulating, soil protection) as well biotic processes (provisioning services, crop production, pollination, biocontrol). We welcome project that performance those studies with modelling approach on different levels of complexity. Second, we seek for investigations that specifically analyze trade-off and off site effects. Thirdly, we appreciate to see how these results fed into a stakeholder process (for instance using multi criteria analysis and weighting of ecosystem services for assessment). We expect to get excellent examples of these issues to stimulate consistency and creativity of future ESS studies.

S7. Spatial agent-based models for socio-ecological systems
Organizers: Dawn C. Parker, Tatiana Filatova
Co-sponsor: Aberdeen Global Land Project Nodal Office on Integration and Modelling

Coupled socio-ecological systems are complex and operate on a variety of scales. Agent-based modeling is widely used in exploring how aggregated phenomena emerge from
interactions of different actors and processes in micro level. This is especially important for modeling land use and various issues related to interactions in socio-ecological systems where human behavior changes the environment. Agent-based techniques allow modeling of these interactions and feedbacks between aggregated outcomes of heterogeneous human behaviors and their affects on environment in a spatially explicit way. This session invites papers demonstrating application of agent-based modeling to land use problems at different scales, socio-ecological problems, and modeling heterogeneous human behavior and its impacts on environment and ecosystem services. The range of questions this session focuses on include: coupling of socio-economic and biophysical models, adding a behavioral component to land use models, exploring different policy incentives for land managers, environmental impacts of different land management behaviors, finding appropriate scale of modeling human behavior and environmental processes in agent-based models, and building empirical agent-based models.

S8. Integrating surface water quality models at the basin scale
Organizers: David Swayne, Yongbo Liu, William Booty, David Lam, Isaac Wong

Innovative and complex models have been developed for watershed models at field scale and for watersheds up to several hundred km². While some improvements have still to be made, the process models are increasingly accepted as reliable and realistic, if somewhat cumbersome in implementation and calibration.

All of these problems grow with the size of the area under observation. Currently, there is a "space barrier" beyond which the only practical estimators are loading estimates and crude spreadsheet models. Transport ability of models can be a problem within a large basin that straddles more than a few ecozones (e.g. the Lake Winnipeg Basin in North America).

This session seeks input for the best practices for watershed modelling, when applying small scale models to large scale basin transport and fate calculations.

S9. Modelling for Northern Environment's Sake: the models in need and the models in deed
Organizers: Georgii Alexandrov, Kaz Higuchi

The feature of Northern Environment is the lack of human domination. Indigenous population is relatively small and lives in balance with nature. This is a fragile balance, vulnerable to any technological intervention and climate change. Intensified development of the North may destroy this balance and make the indigenous population extinct. The purpose of this session is to identify the research lines in environmental modelling that may be essential for developing the knowledge that local communities will need to survive in a changing North.

Model use in the studies of Northern Environment (including case studies, decision models, and guidelines for involving stakeholders in model development)

- Environmental monitoring of the North (with respect to environmental impact assessments)
- Environmental management in the North (principles and conceptual frameworks)
- Ecological economics of indigenous lifestyle in the North
- Scenarios of the North development
- Scenarios of Northern climate change
S10. Integrated Modeling Technologies
Organizers: Andrea Rizzoli, Gerry Laniak, Gene Whelan, Dan Ames, Noha Gaber, Alexey Voinov

This session will use a traditional technical presentation format to provide participants an opportunity to learn about the state of the art and advances in the development of new and existing frameworks and tools for the development of integrated modeling applications and tools as well as applications and case studies. The session will cover the following:

- Frameworks: theory and practice. In this category I expect to see reports on the development of new and existing frameworks for the development of integrated modelling applications and tools. Both theoretical and practical aspects should be investigated.
- Tools: these are specific software systems (e.g. DSS, or sustainability assessment tools) made as instances of frameworks. These are not yet applications, since they display some generic features that make them re-usable across studies and applications.
- Applications and case studies: this is the place to report on concrete real-world applications of integrated modelling. Some of these applications might be based on the frameworks and tools previously introduced, but not necessarily. Particular attention should be devoted to the implications of uncertainty in the integration of large and diverse (in the sense of domain of application) models.
- Standards, ontologies and conventions. How to describe models, space, time, complexity?

S11. Sustainability appraisal – concepts, tools and outcomes
Organizers: Brian S. McIntosh, Keith Matthews, Stefan Sieber, Dagmar Haase

Sustainability and sustainable development have stood notionally at the heart of environmental policy and management since the early 1990s. However they remain deeply contested and often difficult to operationalise concepts when it comes to using them as a basis for policy or management action. The basic aim of this session will be to provide a forum for the presentation and critical discussion of how sustainability appraisal is being incorporated (or failing to be incorporated) into environmental policy and management action through the use of models and software. To do so the Organizers wish to invite contributions covering the full range of issues involved from concepts through models and tools to policy or management outputs and outcomes. The Organizers are keen to see contributions covering a wide range of sustainability perspectives from impact minimisation through learning and change to flexibility and resilience.

The session will be run as a traditional presentation session with accompanying workshop, from which the ambition is to produce a positioning paper for publication in Environmental Modelling and Software.

S12. Environmental monitoring: methods, models, designs and criteria of efficiency
Organizers: Marina G. Erechotchoukova, Peter A. Khaiter

Environmental monitoring is an important starting point in environmental modeling process. Monitoring provides observation data which are used for model calibration, validation and simulation experiments in order to draw conclusions on current and future states of the investigated environmental resource. Since monitoring systems operate in a changing environment and are subject to budgetary constraints, the evaluation of the efficiency and improvements in data collection are urgent. The interdisciplinary nature of a monitoring system dictates the necessity to consider different and sometimes
contradicting aspects of the system during its optimization. Formal optimization techniques require quantification of the effectiveness of a monitoring design in order to weigh it versus the design cost.

The session is aimed at bringing together researchers and practitioners who are interested in optimization and/or improvement of environmental monitoring and, particularly, in the development of efficiency criteria of monitoring programs, optimization methods and approaches to monitoring design, novice methods of observation data processing and application of simulation models to various aspects of environmental monitoring.

S13. Bringing together spatial models of energy, material and water flows in (semi-)natural and technical systems
Organizers: Ruediger Schaldach, Jennifer Koch

The main focus of current spatially explicit land-use system models is the simulation of geographical patterns and their environmental impacts and/or the exploration of interaction between human decision making and the environment. In both cases, the process-based simulation of flows of energy, water and chemical elements within (semi-)natural systems is of central importance. On the other hand, there are numerous simulation techniques available for energy and material flows within technical systems. These include event based models of logistic transport processes as well as methods from the field of Life Cycle Analysis, both representing (semi-)natural systems on a relatively aggregated level. Due to their strong linkage, there is the need for approaches that couple detailed models of land-use systems and technical systems. This session will concentrate on case studies and software solutions which aim at the integrated spatially explicit modeling of energy, material and water flows of both semi-natural and technical systems within geographical regions. Examples may include the coupling of land-use to logistic processes or the link between spatial traffic activities, air pollution and their environmental impacts.

S14. Modelling the coupled social-environmental and physical systems of urban water
Organizers: Vikas Kumar, D. N. Lerner, B. Harris, B. Surridge

Urban water management (UWM) has changed greatly in recent decades, and is no longer just about water supply, sewage and flood defence. It now includes economic and social regeneration, river restoration, ecological habitat creation, riverside urban development, and the provision of amenity and recreational facilities. Despite fine technical work and research on urban water supply engineering and economics, it often seems that such work has not provided a clear unified approach for combining the different approaches to modelling and decision-making for social-environmental and physical systems of urban water. The fundamental challenge that UWM presents to decision makers is how to balance competing and often conflicting environmental, social and economic demands that are placed on catchments. A key challenge for modelling which seeks to support the decision makers is to represent these complex systems in simple yet robust ways. The integrated model should help to negotiate the complex and often conflicting demands associated with UWM. In recent years, many modelling techniques with the potential to meet these challenges have emerged, ranging from graphical probabilistic techniques to agent based AI techniques. Several projects have also taken these techniques forward to the implementation stage and engaged with decision makers.

The aim of this session will be to present recent theoretical and applied work in the area of integrated modelling to support UWM. This session will seek to stimulate analysis and discussion around a number of important issues, including:
• Interfacing social and engineering models.
• Comparing different techniques.
• What are the issues for effective implementation?
• Acceptance - the challenges of the transition from prototype model to real life application.

**S15. Integrated Assessment Modelling of Air Quality and Climate Change across different spatial and temporal scales**
Organizers: Stefan Reis, Tim Oxley

This session is designed to attract model developers as well as stakeholders applying/using integrated assessment models for the design of integrated strategies to improve air quality and to combat climate change. Currently, a variety of integrated assessment models is being developed and applied for a wide range of science and policy questions. These operate on different spatial scales (global, regional, national and local) and - addressing key environmental problems with quite different temporal horizons - are applied to develop policy strategies for both immediate and long-term implementation. Environmental topics such as air pollution and climate change are in the focus of this session, but issues of scale are of particular relevance as well for cross-media effects of pollution, which are often key aspects of model integration and integrated policy design.

Papers submitted to this session should address aspects of integration in crossing spatial scales, dealing with issues of different time-scales and, first and foremost, demonstrate advanced levels of integration both from a model development and an application point of view. Pending the number and coverage of papers submitted, a dedicated special issue publication in an appropriate journal (e.g. Environmental Pollution, Environmental Modelling & Software) is intended.

**S16. Feedbacks in socio-environmental systems**
Organizers: Joerg A. Priess, Nina Schwarz, Sven Lautenbach

The dynamics of socio-environmental systems are driven by exogenous forces and by the interaction of endogenous system components, both within the social and environmental realms, as well as between them. In recent years, the number of models and modelling frameworks explicitly representing feedbacks has increased. Thus, a synthesis of chances and pitfalls in including feedbacks in such models is needed. Land use systems are an important example of socio-environmental systems and will be used as the main focus of the analysis.

Land use changes are on the one hand caused by a complex interaction of human and/or institutional land use demands and the environment which supports or limits human use in several aspects. On the other hand, land use changes and their effects at least partly influence the respective driving forces and future land-use decisions, e.g. by affecting the productivity of agricultural land, in- or decreasing the quality of life in (residential) urban areas, increasing accessibility and thereby facilitating the economic development of areas and so forth. Accordingly, feedback loops of (i) land use changes or (ii) changes of the state of the environment on the socio-economic drivers and land-use decisions are crucial for capturing at least some aspects of the complex socio-environmental system.

For this session, we invite papers on modelling case studies that explicitly deal with land-use related feedbacks in socio-environmental systems. Papers might cover, but are not restricted to, the following topics:

• land-use change
• farming
• forest management
Abstracts should briefly describe the research question, chosen modelling approach (including calibration and model coupling if applicable), implemented feedbacks with spatial and temporal scale, and results. This session is followed by a corresponding workshop which aims at synthesising results.

S17. Scientific workflow tools for the environmental domain - technologies and applications
Organizers: Susan Cuddy, Jean-Michel Perraud

The execution of large-scale environmental modelling exercises (multiple watersheds, multiple models, many data sources, multi-issue, multiple reporting requirements) requires the development and execution of complex scientific and reporting workflows. These may be embedded in an integrated modelling environment - or, more usually, bring together suites of different tools, datasets and processes. The development of scientific workflow tools appears to provide a practical solution for binding disparate data and toolsets together within a controllable framework.

This session invites papers that describe the development and use of scientific workflow tools for environmental modelling. Papers on all and any aspects - technical, governance, adoption, challenges, lessons learnt - are welcome.

S18. Environmental Impacts of Natural Offshore Obstacles
Organizers: Andrea Atzeni, Andrea Sulis

A problem that must be addressed in the design of coastal structure is prediction of the effects of the structure's presence on the shoreline. Breakwaters simulate "nature's way" of using natural obstacles to protect the shoreline, and a thorough understanding of shoreline responses to natural obstacles outside the laboratory can give insights and data to prevent over-design (e.g. tombolo formation), eroding downdrift beach and other negative environmental impacts from single or multiple breakwaters. Environmental and structural effects that control the mode (erosion and accretion) and the magnitude of the shoreline responses in the lee of natural obstacles are not well understood at present, perhaps these conditions having been affected by changing marine climate over recent decades.

This session is concerned with reporting on present knowledge of field observations, laboratory experiments, and numerical modeling studies of shoreline response to natural reefs and islands. Papers are encouraged that address topics in the following priority areas:

- Field observations and scale physical modeling studies to investigate nearshore circulation patterns and resulting shoreline responses;
- Development and application of theoretical models and software tools to simulate beach profiles evolution at various locations in the lee of obstacles.

S19. Modeling climate change impacts: water resources and agriculture
Organizers: Andrei Kirilenko, Xiaodong Zhang

Changing climate is likely to modify hydrologic resources of agricultural areas. The direct effects of these modifications include loss of arable land, loss of infrastructure, changing soil moisture regime. These changes add to the direct effects of shifting temperature and precipitation pattern on agricultural production. The balance between the positive effect
from expanding growing season and negative effect from decreasing soil water availability and higher frequency of meteorological droughts will determine the local impacts of climate change on yields; however the societal impact can be moderated by the adaptations. Amplified by the land use change, the effects of hydrological regime modifications are especially strong in the basins of terminal lakes, which form closed hydrologic systems.

We especially welcome presentations discussing the endorheic basins; however the session invites all abstracts focusing on modeling of climate change impacts on water resources of agricultural lands, including the following topics:

- modification in hydrological regime,
- impacts on the endorheic basins,
- water quality,
- impact on yield,
- societal impacts and adaptations,
- accounting for data, model and scenario uncertainty, and application of model ensembles.

**S20. Model development: the role of uncertainty and model diagnostics**
Organizers: Barry Croke, Hoshin Gupta, Thorsten Wagener, David Post, Ian Littlewood

Development of models requires understanding of the processes involved and the uncertainties in the data used, as well as techniques for evaluating model performance. Uncertainty in the data masks the signature of the system being modelled, limiting the degree of model development when using the top-down approach (building a model based on the signal contained in the data). For bottom-up models, data uncertainty leads to an inability to distinguish between equally plausible model structures. Development of suitable models therefore requires a balance between complexity/detail in representation of processes involved and the impact of data uncertainty. In regard to model performance, the evaluations should emphasise where a model has difficulty (and why), rather than where a model does well, so that improvements in model structure can be pursued. This session will explore these issues, and contributions are invited relating to model development and diagnostics under/for 3 conditions:

- gauged basins,
- ungauged basins, and
- basins under change (land cover and climate).

**S21. Intelligent and collaborative engineering of environmental knowledge: Software platforms, agents and semantics – a special session organized by i-Seek, International Workshop on "Intelligent Systems for Environmental (Knowledge) Engineering and EcoInformatics"**
Organizers: Ioannis N. Athanasiadis, Konstantinos Kotis, Andrea-Emilio Rizzoli, Ferdinando Villa

Models, and to a lesser extent datasets, embody sophisticated statements of environmental knowledge. Yet, both models and datasets rarely encode this knowledge in forms that are self-contained enough to be understood and used - by humans or machines - without the modeller’s mediation. Intelligent and collaborative systems that exploit semantic technologies, agent-based computing or modern software engineering principles can provide a remedy to the above situation. This session aims to bring together scientists reporting on recent advances on the field, especially on the following topics:
Environmental informatics (or enviromatics) is a maturing subject with interdisciplinary roots. The application of information and communication technology (ICT) to the environment is emerging as one of great importance as the health of our planet gains priority on research agendas. Ultimately, environmental information must be put into people's hands so that they can make decisions. How best to involve stakeholders, so that they can access the information they need and put it to use in a satisfying manner, remains a topic of inquiry. Underlying the larger benefits of enviromatics as a tool for policy decisions is the architecture that enables those decision making processes. To maximize the value of the infrastructure, interaction design must be an integral part of the architectural plan. How do we best employ metaphor in educating users and influencing their mental models? What are the ethical concerns involved and how can they be addressed? This design helps the user to improve the quality of the information that is produced, presented, and used. Contributions are sought for a special session on human factors in enviromatics. We will seek to put work on interaction design and human computer interaction into the specific context of environmental modelling and software, with the goal of understanding how to draw on and apply existing knowledge to environmental informatics so that efforts are focused on refinement and adaptation instead of reinvention. Topics include, but are not limited to:

- Usability analyses
- Decision psychology
- Task analyses (including, for example, decision support)
- Validation of ICT tools
- Human-computer interface design
- Human performance evaluation

S23. Second Session on Data Mining as a Tool for Environmental Scientists (S-DMTES-2010)
Organizers: Karina Gibert, Miquel Sanchez-Marre, Joaquim Comas, Ignasi Rodriguez-Roda, Antonio Ciampi, Ioannis Athanasiadis, Joaquin Izquierdo,

This session is strongly linked with W-DMTES-2010, third iEMSs workshop, and aims to approach and to promote the interaction between the Environmental Sciences community to the Data Mining community and related fields, such as Artificial Intelligence, Statistics or other fields to discuss the contribution of Data Mining techniques to Knowledge Discovery in Environmental Sciences, as well as to make data mining techniques more accessible to environmental modellers and to give data miners
and developers a better idea of the needs and desires of the environmental community. The workshop will introduce interested parties to a range of data mining techniques and a selection of software packages. We also invite presentations of interesting applications of data mining to environmental problems. New or improved techniques or methods are welcome, as well as innovative applications.

**S24. Intelligent Environmental Decision Support Systems**

Organizers: Miquel Sànchez-Marrè, Virginia Brilhante, Uilises Cortés, Joaquim Comas, Karina Gibert, Andrea Emilio Rizzoli, Ignasi Rodríguez-Roda, Rick Sojda, Jean Philippe Steyer, Peter Struss, Manel Poch

**Topics**

- Methodologies and frameworks for the development of IEDSSs
- Integration of AI and Statistical/Mathematical Models in IEDSSs
- Model recommendation in IEDSSs
- Benchmarking and validation of IEDSSs
- Relevant Applications and Case studies of IEDSSs
- Other open issues in IEDSSs: spatial reasoning, temporal reasoning, uncertainty modelling and management

The session establishes a discussion platform for Artificial Intelligence (AI) and Environmental researchers involved in the development of applications in the Intelligent Environmental Decision Support Systems (IEDSS) area. Nowadays, AI techniques such as Rule-based reasoning, Fuzzy models, Case-based reasoning, Qualitative reasoning, Artificial neural networks, Genetic algorithms and programming, Model-based reasoning, Bayesian networks, and Multi-agent systems provide a solid basis for construction of reliable and real applications. IEDSS are present in the environmental management process at different levels such as hazard identification, risk assessment, risk evaluation and intervention decision-making, but there is neither a well defined methodology or framework for the development of IEDSSs nor for Model integration nor for Model recommendation techniques nor for Benchmarking and validation of IEDSSs.

Outstanding applications and case studies of IEDSSs with important contributions are also welcome. Other open issues can be addressed, such as the spatial reasoning, temporal reasoning, and uncertainty modelling and management in IEDSSs. These are the open challenges to be addressed by the session papers, and special emphasis will be given to Environment's sake issues.

**S25. Managing regional water resource systems under changing conditions**

Organizers: Julien Harou, Patrick Reed, Amaury Tilmant, David Rosenberg

Changing social, economic, climatic, and environmental conditions continuously challenge our preconceptions on the availability, use and management of water resources. Modeling and decision support innovations are needed to promote adaptation and resilience within our water resources systems given the nonstationarity of future hydrologic extremes and human needs. Paths forward will require mixtures of innovative management schemes that promote flexibility in water allocations, new infrastructure, alternative supplies and demand management programs. This session seeks to address the following questions: what modeling and planning methods work best for preparing integrated regional water resource systems for potentially severe changes in future conditions? How can such planning approaches be effectively deployed in decision support tools? Will they be stochastic? Multi-objective? Collaborative? Hydro-economic? And, now that online real-time data is available to water managers, is there an opportunity for short-term information to connect to mid and long-term risk-based planning applications? This session seeks to showcase modeling applications and software that helps better manage regional water systems in changing conditions.
S26. Modelling and support tools for management and optimization of the integrated wastewater system
Organizers: Peter Vanrolleghem, Joaquim Comas, Wolfgang Rauch, Xavier Flores Alsina, Joaquim Comas, Dirk Muschalla, Eduardo Ayesa, Hubert Colas, Ignasi Rodriguez-Roda, David Butler

The aim of this session consists in the creation of a discussion platform for researchers involved in the development and application of modelling and support tools for the integrated management and optimization of wastewater systems. The idea is not only to present the last trends in mechanistic integrated modelling and system analysis (e.g. uncertainty and sensitivity analysis, frequentist and Bayesian inference) but also to incorporate new techniques/tools such as intelligent decision support systems, soft computing, case-based reasoning, fuzzy control, qualitative reasoning, neural networks, evolutionary algorithms, benchmarking, agents, etc. The session is open to new paradigms and technologies to support integrated wastewater management and optimization, including aspects related to the interaction between the different sub-systems (sewer system, wastewater treatment and receiving media) and the elements of each sub-system relevant for this interaction. Thus, this session will favour contributions extending the classical modelling approach, going further to cover modelling and support tools for innovative processes and emerging technologies and multi-objective process optimization.

S27. Linkages in Human and Environmental Health Modelling
Organizers: Wilfred Cuff, Nick H. Ogden, Venkata R. Duvvuri

Human health threats in the Developed World (such as Western equine encephalitis, West Nile virus and H1N1 influenza) have, as a condition for their spread, a considerable environmental component. Many diseases (for example tuberculosis and malaria) thrive in conditions typical of environmental challenges. As the environment is degraded, so is the state of human health where this degradation, whether in air and water quality, lack of sanitation or overcrowding. Human health matters to us all, and the classical providers are medical doctors (MDs) who are important in health care. MDs are not particularly well educated in software technologies; as a result software is forced to appreciate the nature of the medical doctor. This session will focus on the provision of environmental information as it affects human health, by various means, from the perception of the governmental health agencies and environmental health information providers, and the potential of human health information to assist in environmental model application.

S28. Modelling extremes in climate variables
Organizer: L. A. Sanabria

This session will focus on mathematical tools to model extreme values in climate variables such as rainfall, wind speed, temperature, etc. using observations or climate-modelled data.

Papers on techniques to fit extreme value distributions, point process, etc. to location-based as well as regional (gridded) data are welcome. Also papers on application of extreme value analysis to practical problems in weather-related fields such as natural hazards, impact of extremes in agriculture, health, infrastructure, etc. are welcome.

S29. Multi-scale and multi-physics modeling of environmental flows
Organizers: Hansong Tang, Timothy Keen, Zhifeng Yang, G. Q. Chen

Environmental processes are commonly inter-related and efforts to simulate them must, therefore, span a range of spatial and temporal scales. It is often necessary to
incorporate feedback mechanisms in these modeling studies as well. In order to accommodate these multi-scale and multi-physics (MSMP) requirements, new strategies in coupled/nested/adaptive modeling are emerging in numerical simulation of environmental flows for a range of problems. MSMP approaches, however, are challenging in view of complicated interactions between different physics and scale phenomena. This session will provide researchers with a forum to present their successful results as well as problems from their own research, and to discuss common issues and future developments. The session invites papers on MSMP strategies as well as applications.

Topics include, but are not limited to:

- Numerical analysis on modeling coupling
- Numerical methods for coupling different models
- Integrated wave, current, morphology modeling
- Sediment transport and its interaction with current
- Algae blooming and lake/estuary flow
- Surface water and ground water
- Atmosphere and ocean flow interaction
- Wetland flow modelling
**W1. Effects of climate change: landscape modelling, providing decision support, and understanding uncertainties**  
Organizers: Richard S. Sojda, Thomas C. Edwards, Karina Gibert, Mark Borsuk, Carmel Pollino, Tony Jakeman

A current dilemma facing the world is that we need interdisciplinary, multi-scale science to effectively cope with and adapt to the effects of climate variability and change, but our current diverse capabilities across the landscape make it difficult to realize this integrated capacity. Many agencies are involved in collecting, analyzing, and using climate data and services. Uncertainty associated with both climate data, itself, as well as the impacts of changing climate on natural resources make providing such climate services to resource managers a challenge. Many researchers are building and empirically evaluating models and decision support tools for understanding climate variability and change. Can advanced methodologies in computer science, such as those based in artificial intelligence, lend insight? The kinds of climate issues that fisheries and wildlife biologists, foresters, water managers, park superintendents, agricultural producers, and other natural resource managers are asking are diverse and wide-ranging. Therefore, it is imperative to delineate their management decisions being supported as a basis for validation.

We invite presentations that present scientific findings, delineate new approaches, review existing literature, or demonstrate active models or decision support tools. The intended outcome of this workshop is a jointly authored article among interested attendees that (1) describes the challenges of modelling the effects of climate change as a basis for providing decision support, (2) provides a framework for how scientists are currently propagating uncertainties in their decision support models and tools, and how they are describing those uncertainties to the end user, and (3) suggests future directions for effectively providing decision support tools related to climate change.

**W2. Linking ecologic performance measures to hydrologic models for improved water management**  
Organizers: D. P. Loucks, Carmel Pollino, Wendy Merritt

Water managers are increasingly being asked to consider and provide for environmental flows as they make decisions regarding water allocations to water users. This increased interest in environmental flows stems from the steady degradation of many streams and rivers as the flows have been dammed, diverted and polluted. These hydrological alterations have led to widespread degradation of aquatic ecosystems. During the past decade the increasing concern about the impact of such interventions has motivated the development of numerous methods for assessing environmental flow. These can be grouped into different categories depending on their level of detail. The result of the application of a flow assessment method is one or more descriptions of possible future flow regimes for a river, each linked to an objective which this achieves in terms of the condition of the aquatic ecosystem.

What is the current state of practice in determining environmental flow requirements and implementing them? What evidence do we have that these methods work? What research is needed, and what institutional measures should be considered to enable their implementation? This workshop will address these questions and present some case studies of successes and failures.

**W3. Modeling and deciding with stakeholders**  
Organizers: Alexey Voinov, Nigel W. T. Quinn, Jaroslav Mysiak, Francois Bousquet

Goals: Stakeholder engagement, collaboration, or participation, shared learning or fact-finding have become buzz words and hardly any environmental assessment or modeling effort today can be presented without some kind of reference to stakeholders and their
involvement in the process. Stakeholder involvement became almost a "must". This is clearly a positive sign, however in far too many cases stakeholders are given only lip service and their engagement turns out to be quite formal. The linked session and workshop will explore the expanding field of participatory modeling and participatory decision analysis.

We intend to contribute to the emerging theory of stakeholder involvement, come up with some classifications and categorization of the multiple efforts in this area, and then focus on various applications and case studies. In particular we will consider the various modeling tools and frameworks that are available and will decide what are the benefits that they present for participatory efforts.

**W4. Integrating surface water quality models at the basin scale**  
Organizers: David Swayne, Yongbo Liu, William Booty, David Lam, Isaac Wong

Innovative and complex models have been developed for watershed models at field scale and for watersheds up to several hundred km². While some improvements have still to be made, the process models are increasingly accepted as reliable and realistic, if somewhat cumbersome in implementation and calibration.

All of these problems grow with the size of the area under observation. Currently, there is a "space barrier" beyond which the only practical estimators are loading estimates and crude spreadsheet models. Transport ability of models can be a problem within a large basin that straddles more than a few ecozones (eg. the Lake Winnipeg Basin in North America).

This workshop seeks input for the best practices for watershed modelling, when applying small scale models to large scale basin transport and fate calculations.

**W5. Modelling for Northern Environment's Sake: the models in need and the models in deed**  
Organizers: Georgii Alexandrov, Kaz Higuchi

The feature of Northern Environment is the lack of human domination. Indigenous population is relatively small and lives in balance with nature. This is a fragile balance, vulnerable to any technological intervention and climate change. Intensified development of the North may destroy this balance and make the indigenous population extinct. The purpose of this session is to identify the research lines in environmental modelling that may be essential for developing the knowledge that local communities will need to survive in a changing North.

Topics: Model use in the studies of Northern Environment (including case studies, decision models, and guidelines for involving stakeholders in model development)

- Environmental monitoring of the North (with respect to environmental impact assessments)
- Environmental management in the North (principles and conceptual frameworks)
- Ecological economics of indigenous lifestyle in the North
- Scenarios of the North development
- Scenarios of Northern climate change

**W6. The Future of Science and Technology of Integrated Modeling**  
Organizers: Gerry Laniak, Gene Whelan, Noha Gaber, Alexey Voinov, Vikas Kumar, Craig Aumann, Nigel Quinn, Kurt Wolfe

This workshop will provide participants with the opportunity to discuss advancing the
science and technology of integrated modeling for environmental assessment and decision making. Key questions for discussion include:

- What limitations currently exist, what new problems are emerging, what new technologies will be available?
- How important is re-use and interoperability of technologies?
- What will be the role of standards, ontologies and conventions?
- How to describe models, space, time, complexity?
- How will the QA associated with complex interdisciplinary information flow be achieved?
- How will uncertainty be conceptualized, quantified, and presented to decision makers?
- What issues will require long term research?
- How can the knowledge and products of research be more effectively (i.e., timely) transferred to the applied world and decision making?

In addition, the workshop will seek to identify software and computational technology trends and how these may impact the development of integrated modeling.

**W7. Web Portal for the Community for Integrated Environmental Modeling**
Organizers: Dan Ames, Noha Gaber

The USEPA has taken the lead on engaging the environmental modeling community to develop a community of practice to facilitate greater scientific collaboration and allow more efficient resources by limiting redundancies and duplication in technology development. The Community for Integrated Environmental Modeling was initiated December 2008 and focuses on multimedia and multidisciplinary integrated environmental modeling. The community of practice seeks to integrate and leverage the activities of existing organizations and communities to support and bridge these domain-specific communities but not duplicate their efforts. The goal of the Community's activities is to facilitate collaboration across domains and topical communities (e.g. hydrology, uncertainty, air quality, etc.). This "community of communities" will work with the existing communities to facilitate the flow of science and technology among these communities for the purpose of integrating the science/technology at higher levels for solving highly integrative problems. The Community for Integrated Environmental Modeling will also go beyond providing access to models, frameworks and tools but also serve as an important professional networking and collaboration resource to link researchers, modelers and model users.

The goal of this workshop is to present CIEM's collaborative web-portal to the wider modeling community, obtain their feedback and comments and encourage its use a key resource for the environmental modeling community.

**W8. Complexity reduction strategies for effective use of process based models in environmental decision making**
Organizers: Andrea Castelletti, Rodolfo Soncini-Sessa, Peter C. Young, Hoshin V. Gupta, Marco Ratto

Computational limitations remain a major barrier to the effective and systematic use of large-scale, process-based simulation models in rational environmental decision-making. Whereas, complex models may provide clear advantages when the goal of the modelling exercise is to enhance our understanding of the natural processes, they introduce problems of model identifiability caused by over-parameterization and may not be the best choice for control, management and planning purposes, i.e. when any kind of feedback control, optimization or real-time forecasting is required. Therefore, a combination of techniques for complex model reduction with procedures for data
assimilation and learning-based control could help to bridge the gap between science and practical decision-making.

This workshop will host discussions on the development and the application of new or improved approaches to effectively integrate process-based models and rational decision-making. The focus will be on, but not limited to, model complexity reduction techniques, such as dominant mode analysis, large model emulation and meta-modelling, response surfaces, diagnostic model evaluation, model structure and parameter estimation, model correction and, in the case of partial differential equations, methods such as stochastic collocation on sparse grids. Contributions on specific subtopics, such as design of experiments, sparse grids, feature extraction, etc. are also welcome. Applications might include large water system management, integration of water quality and quantity in surface and/or groundwater, management of large distribution networks, integrated management of surface and groundwater and real-time forecasting of flood inundation.

The discussion will be launched by a position paper that will be available in early March 2010. The workshop will be organized in two parts, consisting of a series of standpoint presentations (each five-minutes long) followed by a round table discussion moderated by the organizers and aimed at modifying, integrating and improving the position paper, which will result in a new collaborative paper. Potential contributors are invited to submit an extended abstract (max 2 pages) of their planned communication.

**W9. Feedbacks in socio-environmental systems**
Organizers: Joerg A. Priess, Nina Schwarz, Sven Lautenbach

The dynamics of socio-environmental systems are driven by exogenous forces and by the interaction of endogenous system components, both within the social and environmental realms, as well as between them. In recent years, the number of models and modelling frameworks explicitly representing feedbacks has increased. Thus, a synthesis of chances and pitfalls in including feedbacks in such models is needed. Land use systems are an important example of socio-environmental systems and will be used as the main focus of the analysis.

Land use changes are on the one hand caused by a complex interaction of human and/or institutional land use demands and the environment which limits human use in several aspects. But on the other hand, land use changes and their effects at least partly influence the respective driving forces and future land-use decisions, e.g. by affecting the productivity of agricultural land, in- or decreasing the quality of life in (residential) urban areas, increasing accessibility and thereby facilitating the economic development of areas and so forth. Accordingly, feedback loops of (i) land use changes or (ii) changes of the state of the environment on the socio-economic drivers and land-use decisions are crucial for capturing at least some aspects of the complex socio-environmental system.

The workshop will draw upon the presentations given in the corresponding session on land-use focused feedbacks. During the workshop, a preliminary version of a synthesising review paper on this topic will be discussed (to be published on the conference website beforehand). This review paper focuses on land-use related feedbacks in various socio-environmental systems and analyses

- feedbacks that are usually tackled in such models,
- possibly neglected, but important feedbacks
- scales of feedbacks (temporal, spatial)
- implementation issues (calibration, validation, model coupling, how generic is the implementation, effect of initial conditions, uncertainty, scaling issues)
Selected participants will be invited to contribute as co-authors to the synthesis paper to be published in a special feature of Environmental Modelling and Software (to be confirmed).

**W10. Impact Assessment (IA) for sustainable development – linking the IA research community with IA policy makers and practitioners**

Organizers: Jan-Erik Wien, Jacques Jansen, Onno Roosenschoon

The project LIAISE (Linking Impact Assessment Instruments with Sustainability Expertise) is a Network of Excellence starting end 2009 and funded by the EC in the 7th Framework Programme. The European Commission considers Impact Assessment (IA) as an important instrument for realizing the key objectives of the renewed Sustainable Development Strategy. IA must enhance the evidence base, transparency and effectiveness of decision making processes. However, existing research points out that the full potential of Impact Assessment is not being realized. Many tools to support IA have been developed, but are not yet fully employed by policy makers.

These missed opportunities are symptomatic of a large and deep gap between the broad communities of IA researchers and IA practitioners. Practitioners tend to look for tools that are simple, robust and transparent, while the researchers are more interested in the sophistication and innovative aspects of IA tools. The main purpose of LIAISE is to identify and exploit opportunities to bridge the gaps in a way that leads to an enhanced use of Impact Assessment tools in policy making. Its centerpiece will be a shared toolbox – simultaneously accessible and useful for policy makers as well as for the research community. A structured dialogue between IA researchers, practitioners and policy makers will be organized to develop and update the IA research agenda.

The main goal of the workshop is to mobilize the international community of IA researchers (modelers and tool developers) and IA users towards an enhanced use of Impact Assessment tools in policy making. To achieve this goal requires an exchange of information between LIAISE and similar initiatives outside the EU. Possible topics for a session are:

- Attracting IA-relevant contributions from other scientific communities to develop a shared vision and set of implementing steps
- International standards and practices of IA and new internationally recognized benchmarks
- A shared IA toolbox as an infrastructure that meets the demands of policy makers

**W11. W-DMTES-2010. Third Workshop on Data Mining as a tool for Environmental Scientists**

Organizers: Karina Gibert, Miquel Sanchez-Marre, Joaquim Comas, Ignasi Rodriguez-Roda, Antonio Ciampi, Ioannis Athanasiadis, Joaquin Izquierdo

This workshop (W-DMTES-2010, 3rd iEMSs Workshop) aims to approach and to promote the interaction between the Environmental Sciences community to the Data Mining community and related fields, such as Artificial Intelligence, Statistics or other fields to discuss the contribution of Data Mining techniques to Knowledge Discovery in Environmental Sciences, as well as to make data mining techniques more accessible to environmental modellers and to give data miners and developers a better idea of the needs and desires of the environmental community. The workshop will introduce interested parties to a range of data mining techniques and a selection of software packages. We also invite presentations of interesting applications of data mining to environmental problems. New or improved techniques or methods are welcome, as well as innovative applications.
W12. Breaking down disciplinary silos: what can environmental modellers in different domains learn from each other?
Organizer: Barbara Robson

Do modellers of terrestrial ecology use the same type of models that aquatic ecological modellers use? Do modellers of catchment hydrology employ the same criteria in evaluating model complexity as those who model estuaries and coastal systems? Do limnological modellers have something to learn from oceanographic modellers, and ocean modellers something to learn from atmospheric modellers? Do biophysical modellers have something to learn from economic modellers?

In short, are environmental modellers employing consistent modelling frameworks, validation techniques, evaluation criteria and approaches to handling uncertainty across disciplinary boundaries? If not, are there good reasons for this? Are the differences due to inherent differences in the systems being modelled and the questions being asked, or are they down to the differences in the training and expectations of scientists in each field? What should we be learning from each other?

The workshop will explore these issues, consider the lessons that emerge, and work towards a position paper for Environmental Modelling and Software.

W13. Expert knowledge in landscape ecological decision support tools: Benefits and cautions

Advances in remote sensing, GIS, and computing technology have made popular the development of decision support tools for conservation and management of terrestrial and aquatic landscapes. In most instances both the development and applications are aided by knowledge of professional experts, who impart their wisdom and insight on ecological patterns and processes. This contribution has remained informal in the past but attempts are now being made to formalize the process of eliciting and including expert knowledge in landscape ecological decision support tools. While the benefits that expert knowledge offers in this regard are many, it is necessary to ensure the same rigour and explicitness that would be associated with using conventional empirical input in developing such tools. Recent advances in science, especially in statistical methods are useful in this endeavour.

This workshop is intended to address the use of expert knowledge in landscape ecological decision support systems and explore the advantages and limitations of its use. We invite presentations that discuss experiences, in both research and/or implementation that may include many steps of using expert knowledge: eliciting and formalizing, assessing uncertainty, validating, and incorporating expert knowledge into the development cycle of decision support systems. Also, we welcome demonstrations of models or decision support tools and literature reviews or syntheses that involve expert knowledge and its use in landscape ecology.

The intended outcome of this workshop is a jointly authored article among interested attendees that (1) highlights the diversity of roles of expert knowledge in developing and implementing decision support tools, (2) summarizes methods employed in eliciting, formalizing, and incorporating expert knowledge, (3) examines the advantages and disadvantages of applied use of expert knowledge, (4) provides insight to the state of knowledge, and (5) suggests topical areas for further research related to use of expert knowledge in landscape ecological decision support tools.
Repeat after me: “I [state your name] am not going to get rich writing environmental modeling software.” The sooner you and I and the rest of our community accept this truism, the more quickly we can advance our science by breaking down walls of software secrecy – be they intentionally or unintentionally emplaced – and hence fostering collaborations at all phases of modeling software development, testing, and use. Indeed, a new spirit of software “openness” has sprung forth in some of the least likely of places. To wit: Microsoft now sponsors a fast growing open source software development community portal and has released all of its key development languages as free “express editions” – in part to support the development of open source software. This movement definitely follows the long standing scientific tradition of publishing one’s research methods and findings in the open literature; certainly the release of source code is the most fundamental form of publication in the field of environmental modeling and software.

There are many reasons why you may not be participating in the open source movement. For example: discomfort at the thought of other individuals viewing your spaghetti code, lack of a clear understanding of the different licenses available and what they mean, lack of time and energy to manage such an effort, or possibly delusional ideas about the fortune to be made from selling your latest groundwater model optimization code (if this last reason is yours, then be sure to review the opening mantra in this workshop summary).

The purpose of this workshop is to address these issues through presentations and discussion of 1) licensing options and implications, 2) shared code development tools and systems, and 3) shared/open source model software development case studies. Participation is sought from individuals with experience and success stories related to this topic. Also, individuals new to open source software development, or who are afraid that one day their code will be sitting in a doorstop (the final resting place of so much good code long since forgotten in an old worn out computer) are also highly encouraged to join this workshop.

W15. Tales of DSS adoption: How and why are DSS successful in environmental and related sectors?
Organizers: Marten Stavenga, Brian S. McIntosh, Serena Chen, Tony Jakeman

Concerns persist that DSS tools fail more often than succeed in being adopted by the intended end users. Contemporary environmental concerns including cross-border pollution and integrated resource management under conditions of climate change require larger, multi-scale, complex modeling efforts. Such concerns must be addressed but present significant challenges to the achievement of successful adoption of Decision Support Systems. This workshop will present a number of case studies of successful DSS adoption in environmental and related sectors, such as agriculture, fisheries, production and environmental health sectors. Each case study will review DSS adoption as a process rather than as a single event with the aim of teasing out the factors which influence adoption in context.

The workshop will be organized based on a matrix framework using two perspectives: a) type of DSS application, and b) subject area / context of application. The subject application areas in scope will include all relevant areas of environmental sciences, including, but not limited to waste, environmental pollution, water management, climate change, environmental planning, environmental chemistry. In addition, also DSS applications in bordering sectors are in scope, such sectors being, but not limited to,
forestry, oceans and fisheries, agriculture, environmental health, sustainable production and consumption.

What have been key factors for successful adoption of DSS in these various sectors? What have been pitfalls with adoption and application of DSS? What are key learnings on the adoption process across these various sectors? This workshop will aim to address these questions and issues.

The goal of this workshop is to deliver key insights for DSS developers to create a deeper understanding what the real uses and needs are for DSSs.

W16. Interoperability for Web Based Modeling
Hosted by the Open Geospatial Consortium, Inc (OGC)
Organizers: David Arctur, George Percivall, Phillip C. Dibner

The goal of this workshop is to develop familiarity among iEMSs participants with OGC open standards for web services, with particular attention to the utility of these and complementary standards to support improved interoperability of environmental modeling. The workshop will also be positioned to identify and prioritize challenges and issues for joint work by OGC and iEMSs members as part of the Alliance partnership established between these two organizations in late 2009.

Background: The OGC has developed a range of OGC Web Service (OWS) standards to improve the ease at which location or geospatial information can be discovered, accessed, fused and applied to increasingly complex problems facing decision makers worldwide. In the past several years, OGC members have emphasized standards development in support of broad geosciences objectives in the areas of hydrology, climate change, ocean observation, geology, and environmental science. Recently, OGC released a family of Sensor Web Enablement standards which provide rapid and real time access to a range of fixed and mobile sensors, and the ability to access, integrate, fuse and apply sensor information for decision making in a location and temporal context. This coupled with the release of the OGC Web Processing Service provides a significant level of standards-based capability to help advance the objectives of the modeling community.

Proposed focused work sessions and discussions

- Overview of OGC organization and process, and OGC service implementation specifications as they relate to location interoperability. Emphasis on user community benefits being realized in environmental, ocean, climate and other science communities of interest.
- Presentation of OGC standards and best practices relevant to environmental modeling. Discussions include emphasis on Web Processing Services, Workflow Management, Digital Rights management, Sensor Web Enablement and other standards.
- Group Discussion. Development of issues. Opportunities and challenges potentially ripe for OGC / iEMSs collaboration, areas for immediate use of OGC standards, iEMSs use cases for potential use in future collaborative project activities.
- Further discussions on potential collaboration between OGC and iEMSs.

W17. Low-carbon industry and multi-scale input-output modeling
Hosted by Beijing Development Area Co., Ltd. (BDA Ltd), Beijing, China
Organizers: G. Q. Chen, B. Chen, S. Y. Zhou, H. S. Tang,

As an urgent dilemma facing the human society, with all our strength we make efforts to reduce and mitigate greenhouse gas (GHG) emissions at distinctive industrial levels, but
as a whole the carbon emission is booming with rapid increase. This is due to the bottom end characteristics of conventional mitigation strategies necessitating intensive economic input at the end with massive carbon cost. The emerging trend to resolve the dilemma is based on the low-carbon systems engineering supported by databases and module packages associated with top-down and bottom-up integrated multi-scale multi-regional Leontief modeling.

We aim to identify the systems theory, methods, technologies, business modules and best practices in low carbon development that could be applied to reduce GHG emissions for nations, regions, sectors, and industries. To achieve the goal of the carbon reductions as reinforced by the Copenhagen Accord, specific efforts on environmental modeling will be instrumental to obtain more fruitful results. This may help make corresponding government policies to promote the energy efficiencies and reduce carbon emission intensity at each level of the concerned systems, provoke financial incentives in the form of domestic and international investments for low carbon projects, and spur development of multi-scale input-output tools and deployment of new energy, environmental and ecological technologies to realize carbon emission reduction.

The workshop will be structured but not limited to the following inter-dependent topics:

- Low-carbon industry park
- Low-carbon building and real estate
- Low-carbon supply chain
- Low-carbon logistics
- Low-carbon evaluation and consultant
- Multi-scale ecological input-output models
- Multi-scale multi-regional databases for direct and embodied carbon inventory
- Renewable and substitute energy
- Life cycle analysis
- Environmental accounting
- Greenhouse emission accounting
- Embodied energy accounting
- Carbon footprint
- Low-carbon wastewater treatment
- Carbon measurement
- Carbon capture
- Low-carbon technologies

This special workshop invites professionals from universities, enterprises, and administrative departments concerned with low-carbon projects and multi-scale input-output methods to make effective comparisons, to present and to share new ideas, innovations, trends, experiences, and concerns in the environmental modeling and systematic simulation. We also believe the workshop for low-carbon industry and multi-scale input-output modeling will become an important platform for the other participants of iEMSs 2010 to exchange knowledge, perspectives and ideas for the low-carbon economy and to discuss the most recent advances in simulation models and assessment methods from both theoretical and practical perspectives.
Workshop Report: W5: Modelling for Northern Environment's Sake
Summary of the workshop W5 held at iEMSs 2010

Participants:
Georgii Alexandrov, Kaz Higuchi, Rachel A. Hirsch, Glen Lesins, James I. MacLellan

The feature of Northern Environment is the lack of human domination. Indigenous population is relatively small and lives in balance with nature. This is a fragile balance, vulnerable to any technological intervention and climate change. Intensified development of the North may destroy this balance and make the indigenous population extinct. The purpose of this workshop was to identify the research lines in environmental modeling that may be essential for developing the knowledge that helps indigenous communities to survive in a changing North.

Social adaptation to climate change may occur over various scales. At the national scale, this may result in elaboration of adaptation strategies based on scientific knowledge. Since the available scientific knowledge is quite abstract and counterfactual, only generic adaptation strategies may be produced at this scale. Local institutions may develop a specific planned response based on the past experience of responding to extreme events. As to the indigenous communities, they may only accept the changes pragmatically and respond involuntary.

Institutional communications are essential for achieving the synergy of a nationwide adaptation strategy, a locally planned response and an involuntary response of indigenous communities (that is, response dictated by changing environment, not by a strategic goal). Understanding the indigenous usage of threatened ecosystem services is an important part of such communications. A locally planned response must address the way in which indigenous communities interpret their role on the land.

Recent advances in agent-based modeling make it possible to support institutional communications by a model simulating involuntary response of indigenous communities. Development of an "indigenous response simulator" was identified as a major research line in modeling for Northern environment’s sake. Not all environmental changes that seem economically safe from the national viewpoint are culturally acceptable for indigenous communities. The ultimate purpose of the proposed inter-disciplinary effort on developing the "indigenous response simulator" is to find a scientifically sound method for balancing economic and cultural targets.
Figure 1 Brainstorming map of the issues discussed at the workshop

Papers:
- Agent Based Modelling of Caribou Environmental Interactions in the Canadian Arctic  Glen Lesin
- Modelling Governance Structures and Climate Change Policy Communications on Community Resilience in the Canadian Arctic  Rachel Hirsch
- Modelling needs assessment for social adaptation to climate change in Siberia Georgii Alexandrov
- The Northern Global Climate Change Adaptation Dialogue  James Maclellan
Workshop Report: W12: Breaking Down Disciplinary Silos: What can Environmental Modellers in Different Domains Learn from Each Other?

Summary of the workshop W12 held at iEMSs 2010

Organizer: Barbara Robson

Participants: Tony Jakeman, Marit Kragt, Kit Macleod (Kit.Macleod@csiro.au), John Norton, Mark Borsuk, Sid Pendelberry, Rebecca Lester, Francesco Falgieri, Serena Chen, Ejaz Quereshi, Vincent Lyne by videoconference

Papers presented

This workshop provided a venue to discuss cross-disciplinary learning in environmental modelling. The discussion component of the workshop was preceded by a contributed paper session on the same topic. Several papers were presented, and are published in the conference proceedings.

The session began with a presentation by Vincent Lyne, who drew parallels between ecological and socioeconomic systems, demonstrating that some of what holds for ecological structures and typologies is also true of socio-economic structures. In developing models for one disciplinary field, we may have much to learn from the patterns and structures identified in another field, even if the processes and components involved are very different. The presentation was somewhat hampered by problems with the videoconferencing technology – by the time of the next iEMS symposium, the technology should be more trouble-free. Nonetheless, a very interesting paper is included in the proceedings.

The second speaker was Ejaz Quereshi, who presented work on integrating biophysical and economic data. The different spatial and temporal scales of biophysical data and economic data present a challenge for interdisciplinary work, and Dr Quereshi described how his group handled this challenge. Variation in spatial and temporal scales is an issue common to many environmental modelling problems, with scales dictated by the time- and space-scales of the processes involved, the available measurement technology or metrics, and the questions being asked of the system.

Third, Serena Chen discussed her work in combining Bayesian Network and GIS habitat modelling approaches. By combining two approaches to ecological modelling, Ms Chen and her colleagues were able to overcome some of the limitations of each approach: the spatial limitations of Bayesian Networks and the potential rigidity of rules-based habitat models. This presentation was well received and was subsequently awarded an iEMSs student prize.

Continuing the theme of ecological modelling in a multidisciplinary context, Rebecca Lester presented her work on overcoming data constraints to create meaningful ecological models. Dr Lester described how she was able to use poor-quality, sparse data from a variety of sources with mismatched time-scales to develop an ecological state model from which she could draw some meaningful predictions. The state-based statistical modelling approach presented in Dr Lester’s paper may be applicable to other multidisciplinary problems where sparse and ill-matched datasets are a problem.
The final session speaker was Kit Macleod, who presented an overview of systems based modelling, and made a case for the need for better integration of whole systems and better communication of systems modelling.

Discussion and follow-up

The presentations were followed by a fruitful discussion. The presence of Professor Jakeman (Editor in Chief of Environmental Modelling and Software) provided a focus for much of this discussion, which was, “How can Environmental Modelling and Software and iEMSs conferences facilitate cross-disciplinary knowledge transfer in environmental modelling?”

Suggestions included encouraging workshop contributors and others to write plain-language introductions to specialist topics, such as particular modelling approaches, issues, or disciplines. Such introductions should be targeted at those outside the specialist field who are interested in learning more about it, and perhaps adopting some of its methods or learning to their own work. The introductions should avoid jargon, include glossaries where necessary, and be reviewed by researchers who are naive to the specialist topic as well as by disciplinary experts. These introductions could perhaps be published on a website associated with the journal or society, if not as a short communication series in EMS itself.

Topics for such introductory essays might include, for example:
- Climate and weather models
- Hydrological models
- Ecological models
- Mechanistic and process-based modelling
- Statistical modelling (and introductory essay for various types of statistical modelling).
- Models for particular environments (e.g. Lakes, Estuaries and coastal waters, Wetlands, Floodplains, Oceanic waters, Deserts, Agriculture, Urban systems, Savannah and grazing land, Forests).

EMS could solicit more review papers. Postgraduate students writing literature reviews for their theses might be ideal authors for such reviews: as newcomers, they may more recognise the entrenched jargon of a field and recognise more easily which aspects of a topic may seem opaque to newcomers. Encouraging students to submit review papers may also be a way to enhance the career development of students, as good review papers are often highly cited. There was some disagreement on this point, with some feeling that more established scientists who have watched their discipline grow and develop over years may be better placed to write reviews that put work in their area in a broad context.

Some possible topics for review papers were discussed. These included “lessons learnt from climate modellers”, which could cover model communication, handling uncertainty and probability, when and how to get involved in policy and outreach, how to work across multiple scales, developing an interconnected modelling community, and model operationalisation. Some of these components might be good topics for review papers in their own right.
Journal editors could be more active in working with authors of reviews and commentaries to make sure the language is clear and more widely accessible.

Modellers and journals should be encouraged to publish more papers with negative results, so that we can learn not just from successes, but also from the failures of others. Such papers would provide more accessible information about the pitfalls of different modelling approaches, to help build an understanding of when and when not to use them. A thematic issue on the topic “stuff we really thought should work that doesn’t” might prove very instructive.

More generally, reviewers should encourage manuscript authors to analyse and openly discuss model failures and weaknesses rather than trying to downplay and hide them. Such discussion should include analysis of why the failure occurred.

Conference organisers could be encouraged to solicit keynote speakers from disciplines outside the primary skill base of the conference, to encourage cross-disciplinary knowledge transfer. Such speakers should, naturally, be outstanding communicators.

As workshop organiser, I’d like to thank all participants for such a positive and productive discussion. In coming months, I hope we will follow this up with review papers and action on these suggestions.
## iEMSs 2010 Conference Student Awards

**Student Awards Chair: Susan Cuddy**

## Recipients of Awards

<table>
<thead>
<tr>
<th>Student</th>
<th>Affiliation</th>
<th>Paper session and title</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Best papers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stefano Balbi</td>
<td>CEEM, Department of Economics, University of Venice</td>
<td>S07 – A spatial agent-based model to explore scenarios of adaptation to climate change in an alpine tourism destination</td>
</tr>
<tr>
<td>Lyubomir Halachev</td>
<td>School of Business Management, Ryerson University, Toronto, Canada</td>
<td>S23 – Analysis of the economic sustainability of companies in the water sector</td>
</tr>
<tr>
<td>Gert Everaert</td>
<td>Environmental Toxicology and Aquatic Ecology, Ghent University, Belgium</td>
<td>S23 – Development of data-driven models for the assessment of macroinvertebrates in rivers in Flanders</td>
</tr>
<tr>
<td>Anna Cord</td>
<td>Department of Remote Sensing, University of Wuerzburg, Germany</td>
<td>S12 – Remote sensing time series for modelling invasive species distribution: a case study of Tamarix spp. in the US and Mexico</td>
</tr>
<tr>
<td><strong>Highly recommended</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serena Chen</td>
<td>Australia</td>
<td>W12 – Bayesian network for modelling habitat suitability of an endangered species</td>
</tr>
<tr>
<td>Simon Hood</td>
<td>University of Guelph, Canada</td>
<td>S00 – A novel model calibration technique through application of machine learning association rules</td>
</tr>
<tr>
<td>Ioan Ferencik</td>
<td>Dept of Civil &amp; Environmental Engineering, Aalto University, Finland</td>
<td>S00 – On site environmental modelling and monitoring: the Nordic scenario in HYDROSYS project</td>
</tr>
<tr>
<td>Anas Altartouri</td>
<td>Dept of Civil &amp; Environmental Engineering, Aalto University, Finland</td>
<td>S00 – Spatiotemporal modelling of the spread of common reed on the Finnish coast</td>
</tr>
<tr>
<td>Monica Carvalho</td>
<td>University of Zaragoza, Spain</td>
<td>S4 – Modelling a simple trigeneration system: environmental costs</td>
</tr>
<tr>
<td>Lewis Gill</td>
<td>Dept of Computer Science, Sheffield University, UK</td>
<td>S14 – An interactive visual decision support tool for sustainable urban river corridor management</td>
</tr>
<tr>
<td>Javier Holguin</td>
<td>Environmental Toxicology and Aquatic Ecology, Ghent University, Belgium</td>
<td>S14 – Modelling the ecological impact of discharged urban waters upon receiving aquatic ecosystems. A tropical lowland river case study: city Cali and the Cauca river in Colombia</td>
</tr>
<tr>
<td>Sidney Pendelberry</td>
<td>Computing &amp; Information Sciences, Rochester Institute of Technology, New York</td>
<td>S11 – A taguchi-based method for assessing data center sustainability</td>
</tr>
<tr>
<td>Yi-Liang Kuo</td>
<td>School of Forestry and Resource Conservation, National Taiwan University</td>
<td>S00 – Unexpected Side-effects of Winter Feeding: Learning from Mahalanobis Distances Factor Analysis in the case of red-crowned Cranes in Hokkaido, Japan</td>
</tr>
</tbody>
</table>
Process

Formation of Committee

Several weeks prior to the conference, a call went out to members to volunteer for the student Award Committee. Susan Cuddy agreed to chair the Committee. Fourteen members volunteered and others were co-opted during the week.

Assessors

Members who volunteered for the Committee were wonderfully generous and supportive through the week. One hundred and sixteen assessments were completed.

Ari Jolma  John Norton  Peter Gisbers
Barbara Robson  Karina Gilbert  Rebecca Lester
Giorgio Guariso  Kristina Voigt  Rick Sojda
Ilias Pechlivanidis  Lucio Moreira  Susan Cuddy
Ioannis Athanasiadis  Mark Borsuk  Tony Jakeman
Jim Ascough  Paul Feikema  Vikas Kumar
International Sustainability Governance and the Science-Policy Nexus: Lessons from the Arctic

Karen Kraft Sloan

Climate Adaptation and Sustainability Governance Specialist
EcoNexus

Abstract: The presentation highlights key sustainability milestones and the role played by science and research to spur advances in international sustainability governance. The scientific community, by focusing attention on the threats of global environmental change, has galvanized political action required to create much needed international and domestic policy responses in areas related to ozone depleting substances, acid rain, endocrine disrupting chemicals, child environmental health, climate change, desertification and biodiversity loss.

The International Council for Science (ICSU) is calling for a global ‘Apollo Project’, an unprecedented, intense and focused decade of research to address global change. “Given the pace and magnitude of human-induced global change, immediate actions are needed to avoid dangerous outcomes for people and for the planet. In this context, science needs to deliver useful and reliable information that will directly and effectively inform and support the responses and actions of decision-makers and citizens in all regions of the world.”

The Arctic has a rich history of international scientific cooperation dating back to the first International Polar Year of 1881. Three more International Polar Years (IPY) followed, including the fourth most recent polar year (IPY 2007 – 2008). The preeminent Arctic institution, the Arctic Council has its roots in environmental and scientific initiatives and the council is responsible for overseeing environmental programs originally created under the Arctic Environmental Protection Strategy (AEPS). The eight Arctic nations came together in 1989 to create the AEPS, acknowledging the responsibility of the Arctic States to protect and preserve the Arctic environment. Lessons learned from the Arctic to inform our understanding of international environmental governance and the science – policy nexus will be explored, including current post IPY attempts to institutionalize international Arctic science cooperation and to strengthen linkages between the worlds of science and policy in order to better prepare for global and regional challenges.

---

A Bayesian perspective on climate policy modeling

Mark E. Borsuk*, Michael D. Gerst*, Richard B. Howarth

* Thayer School of Engineering, Dartmouth College, 8000 Cummings Hall, Hanover, NH 03755
b Environmental Studies Program, Dartmouth College, 6182 Steele Hall, Hanover, NH 03755

* Corresponding author: 603-646-9944 (phone), 603-646-2277 (fax), mark.borsuk@dartmouth.edu

Abstract: Bayesian decision theory requires consideration of the relative likelihoods (or probabilities) of all possible outcomes of decision alternatives, as well as relative preferences (or utilities) for these outcomes. While advancements have been made recently in characterizing the uncertainty in climate change and the effectiveness of alternative mitigation policies, there has been relatively little progress in characterizing society’s preferences for these uncertain climate futures. It is clear that conclusions regarding socially optimal climate policy are critically dependent on the choice of social preference parameters (i.e. the discount rate and risk aversion coefficient). Yet, there has been a hole in the modeling literature concerning the interplay of climate change uncertainty, economic risk, and description of social preferences.

Previous high-profile climate economic analyses have either: (i) adopted a wholly prescriptive approach to characterizing social preferences (e.g. the Stern Review and the PAGE model), based on moral argument rather than consistency with historical observation, or (ii) adopted a descriptive approach to social preference description (e.g. Nordhaus’s DICE, Tol’s FUND) by choosing parameter values that calibrate the deterministic Ramsey economic growth model to average returns on investments. The former approach has generally led to the conclusion that aggressive greenhouse gas reductions are socially optimal, while the latter has suggested much more moderate action on climate change. This has led to much debate over the relative merits of prescriptive versus descriptive social preference description.

Of course economic growth under climate change is not perfectly known, as assumed by the Ramsey model, but rather highly uncertain. Not only is such ignorance of economic risk logically inconsistent, but it also fails to address a basic observation of investment behavior – that investors demand substantially greater return on risky assets than safe assets. Referred to as the equity premium, this difference is a fact that the Ramsey model cannot explain, but has important implications for the appropriate choice of preference parameters.

To overcome this problem, we adopt a Bayesian approach to climate risk assessment and social preference description. As previously pointed out by economist Martin Weitzman, the near impossibility of ruling out the potential for economic disasters leads to a distribution on consumption growth under the Bayesian framework with relatively high probabilities of large losses. Such a distribution is referred to as ‘fat-tailed’, and Weitzman argues that society should incur any cost to avoid fat-tailed risks because they lead to infinite expected loss. We propose a sensible lower bound to the utility of consumption losses to avoid this problematic conclusion and use a large historical data set to show fat-tailed distributions on consumption growth can actually lead to reasonable values for preference parameters. This model accounts for the equity premium and has important implications for climate policy evaluation.

Specifically, our results suggest that the parameter values typically used for social preference description in climate policy models specify a society that puts too much weight on future consumption and is too tolerant of consumption risk. Consequently, our model
identifies the comparatively low CO$_2$ stabilization target of 450 ppm (from the range 400 to 1000 ppm) as providing the greatest net benefit. This result is unique in indicating that aggressive reductions in emissions of global warming gases are economically justified without appealing to a low discount rate, fat-tailed uncertainty, or a steeply convex damage function. Our future research will address the issue of ambiguity aversion in addition to risk aversion, as the characterization of probability distributions in risk analysis is itself often uncertain or ambiguous.
Gaps and missing links in water quality and ecological modeling for river basin management

Ann van Griensven

UNESCO-IHE Water Education Institute, Department of Hydroinformatics and Knowledge Management, P.O. Box 3015, 2601 DA DELFT, The Netherlands
(Tel: ++31-15-2151812, Fax: ++31-15-2122921
E-mail: A.vanGriensven@unesco-ihe.org)

Abstract: The Water Framework Directive requires a holistic and integrated assessment at river basin scale to predict the effect of programmes of measures on the ecological status of all waters (physical, chemical and biological). Integration of models was popular in the ‘80s and ‘90s. It was driven by the Integrated Urban Water Management community and led to new formulations of river water quality models (RWQM1) to overcome variable clashes. In the ‘90s, the link to hydrological models at river basin scale allowed the joint accounting for diffuse and point sources. Links have also been made to socio-economic and ecological models. Nevertheless, many gaps are still to be tackled:

- better inclusion of natural ecosystems, eg forests, wetlands, to account for the ecosystem functions and services, such as pollution removal or erosion control;
- better integration of groundwater-surface water quality, e.g. denitrification processes in the groundwater;
- improved accounting for spatial variability of processes (i.e., composite modelling);
- more generic approaches to link to other modeling paradigms, e.g. cellular automata, agent based models, Bayesian networks for ecological assessment;
- standardized methods for linking models to models/models to data;
- changed attitudes (dare to share);
- better linking to the social world, e.g. use of web-services, knowledge systems; and
- new methods to propagate uncertainty in linked models.
Table of Contents

Session Descriptions ........................................................................................................... 10
Workshop Descriptions ........................................................................................................ 23
Workshop Reports................................................................................................................ 32
  Workshop 5: W5: Modelling for Northern Environment's Sake ..................................... 32
Student Awards .................................................................................................................... 37
Plenary Speakers .................................................................................................................. 39
  International Sustainability Governance and the Science-Policy Nexus: Lessons from the Arctic Karen Kraft Sloan ................................................................. 39
  A Bayesian perspective on climate policy modeling Mark E. Borsuk .............................. 41
  Gaps and missing links in water quality and ecological modeling for river basin management Ann van Griensven ................................................................. 43
S 0. Open Session ................................................................................................................ 67
  A tree-based feature ranking approach to enhance emulation modelling of 3D hydrodynamic-ecological models A. Castelletti, S. Galelli, R. Soncini-Sessa .... 67
  Comparison of Upper Air Mixing Height Estimation Methods for Urban Air Pollution Modeling Afsana Khandokar, Abdullah Mofarrah, Tahir Husain ............ 75
  Effective Microorganisms (EM) Technology for Water Quality Restoration and Potential for Sustainable Water Resources and Management Zuraini Zakaria, Sanjay Gairola, Noresah Mohd Shariff ..................................................... 80
  Modeling for Uncertainty Assessment in Human Health Risk Quantification: A Fuzzy Based Approach Abdullah Mofarrah, Tahir Husain ................................. 88
  Response Surface approximation and Interactive Decision Maps for water quality planning A. Castelletti, D. Limosani, A.V. Lotov, X. Quach, R. Soncini-Sessa .... 96
  Towards the Adoption of Integrated Urban Water Management for Planning Shiroma Maheepala ......................................................................................... 104
  Unexpected Side-effects of Winter Feeding: Learning from Mahalanobis Distances Factor Analysis in the case of Red-crowned Cranes in Hokkaido, Japan Yi-Liang Kuo ........................................................................................................... 112
  Validation and Sensitivity Analysis of Catchment Runoff and Erosion Simulation Technology (CREST): A GIS-assisted Soil Erosion Model at Watershed Level Reynaldo L. Lanuza, Eduardo P. Paningbatan, Jr. .............................................. 117
S00. Tools and Techniques For Environmental Modelling and Software .................... 126
A Novel Model Calibration Technique through Application of Machine Learning Association Rules  
Simon Hood, David Swayne ................................................................. 131

Calibrating Environmental Models using ParaMES  
Daniel Princz, L. Shawn Matott ................................................................. 139

Developing a model for detecting growth pulses in the observations and scenarios of CO2 emissions  
Gleb Alexandrov..................................................................................... 149

Development of a methodology for Integrated Water Resources Management in Mediterranean phosphate mine areas  
Bru Kathy, Graveline Nina, Soulis Konstantinos, Dercas Nicholas, Stefopoulos Angeliki, Guesmi Anis, Karaouli Fatma, Slimani Mohamed, Guezennec Anne-Gwenaëlle ................................................. 153

Estimates of Seasonal Methane Emissions from Paddy-Rice Cropped at Pindamonhangaba Municipality, Brazil, Using DnDc Model  
Maria Conceição Peres Young Pessoa, Magda Aparecida de Lima, Omar VieiraVilella ............................................. 162

Guidelines for Good Practice in Bayesian Network Modelling  
Serena H. Chen, Carmel A. Pollino .............................................................. 170

Modeling and support tools for studying disease spread in livestock using networks  
Joel Francis, Greg Klotz, Neil Harvey, Deborah Stacey ................................................. 179

Modelling the environment using graphs with behaviour: do you speak Ocelet?  
P. Degenne, A. Ait Lahcen, O. Curé, R. Forax, D. Parigot, D. Lo Seen ............................................. 188

On site environmental modeling and monitoring: the Nordic Scenario in HYDROSYS project  
Ioan Ferencik, Tero Niemi, Ari Jolma............................................. 196

Parcel subdivision automation for agent-based land use modelling  
Rohan Wickramasuriya, Laurie Chisholm, Marji Puotinen, Nicholas Gill, Peter Klepeis.. 205

Spatio-temporal modelling of the spread of common reed on the Finnish coast  
Anas Altartouri, Ari Jolma, Päivi Korpinen ................................................................. 213

Use of hydraulic modelling to aid decision making in the management of Oakley Creek  
Achela K. Fernando, Bruck Gebreselasie, Jacinto Capiral ............................................. 224

Using NVIDIA GPU for Modelling the Lagrangian Particle Dispersion in the Atmosphere  
Jiye Zeng, Tsuneo Matsunaga, Hitoshi Mukai ....................................................... 232

Using satellite-derived optical data to improve simulation of the 3D light field in a biogeochemical model  
Barbara J. Robson, Nagur Cherukuru, Vittorio Brando .. 237

Waste water discharge impact modeling with QUAL2K, case study: the Zayandehrood River  
Nader Nakhaei, Amir Etemad Shahidi.................................................... 245

S01. Forest Fire Modeling and Software .......................................................... 253

S01A. Forest Modelling and Software.............................................................. 253

Modelling forest fires hydrological impact using spatio-temporal geographical data  
Soulis Konstantinos, Valiantzas John, Dercas Nicholas ................................................ 253

Modelling spatiotemporal variability of natural forest fire size class distribution in a boreal forest  
Wenbin Cui, Ajith. H. Perera.......................................................... 262

A Model-Based Quantitative Assessment of Ecosystem Services in the Scenarios of Environmental Management  
Peter A. Khaiter and Marina G. Erechtchoukova .. 272
Habitat models as a research gap in biodiversity conservation in tropical rain forest of southeast Asia Mona Nazeri, Kamaruzzaman Jusoff, Nima Madani ............ 280

From landscape-scale models of forest biodiversity change towards an integrated planning Tool – a GUI for building empirical ecological extrapolation models based on land cover data Tobias Lung, Gertrud Schaab ................................ 288

Using 3PG+ to simulate longterm growth and transpiration in Eucalyptus regnans forests Paul Feikema, Jim Morris, Craig Beverly, Patrick Lane, Tom Baker ...... 296

S02. Environmental Fluid Mechanics - Theoretical, Numerical and Experimental Approaches........................................................................................................ 304

3D computational model of external intrusion in a pipe across defects P. Amparo López-Jiménez, Jesús Mora-Rodríguez, Francisco José García-Mares, Vicente S. Fuertes-Miquel................................................................. 304

Distributed Moments of a Scalar PDF Partha Sarathi, Roi Gurka, Paul J. Sullivan, Gregory A. Kopp .......................................................................................... 313

Evaluation of odour impact from a landfill area and a waste treatment facility through the application of two approaches of a Gaussian dispersion model Úbeda, Y., Ferrer, M., Sanchis, E., Calvet, S., , Nicolas, J., López, P. A. ..................... 322

Experimental Observation Of Turbulent Structures In A Straight Flume Donatella Termini, Vincenzo Sammartano................................................................. 332

Experimental study of the critical criteria for incipient sediment movement using a simple model eddy R J Munro ................................................................. 340

Flow kinematic characteristics in the scour hole downstream of a grade-control structure Donatella Termini, Vincenzo Sammartano ............................................. 348

Interacting Environmental Interfaces: Synchronization in Substance Exchange between Environmental Interfaces Regarded as Biophysical Complex Systems D.T. Mihailović, M. Budinčević, D. Perišić, I. Balaž, A. Firanj ...................... 356

Land Ice Sea Surface Model: Short Description and Verification A.Vukovic, B. Rajkovic, Z. Janjic ..................................................................................... 364

Model of ventilated façade in buildings by using CFD techniques P. Amparo López-Jiménez, Miguel Mora-Pérez, Gonzalo López-Patiño, Miguel A. Bengochea Escribano ...................................................................................... 372


Modelling flow and concentration field in rectangular water tanks F. Javier Martínez-Solano, Pedro L. Iglesias Rey, Carlo Gualtieri, P. Amparo López-Jiménez ................................................................. 389

Numerical simulation of transition layer at a fluid-porous interface Carlo Gualtieri ............................................................................................................. 399

The Spatial Variation of the Maximum Possible Pollutant Concentration from Steady Sources N Mole, R J Munro ..................................................................... 410
S03. Modelling of dangerous phenomena and innovative techniques for hazard evaluation, mapping, and mitigation

DEM uncertainty and hazard analysis using a geophysical flow model  
E.R. Stefanescu, M. Bursik, K. Dalbey, M. Jones, A.K. Patra and E.B. Pitman ................................. 418

Impact of DEM uncertainty on TITAN2D flow model output, Galeras Volcano, Colombia  

Multilinear Diffusion Analogy Model for Stage Hydrograph Routing  
Muthiah Perumal, Tommaso Moramarco, Bhabagrahi Sahoo, Silvia Barbetta, Florisa Melone ................................................................. 436

Using a statistical model for the description of uncertainties associated with dispersal models  
Ali S. Gargoum ............................................................................................................ 444

S04. The European Environmental Information Space: News and Trends

Meta-Modelling the European Environmental Information Space  
Werner Pillmann, Karl-Heinz Simon.................................................................................... 452

A Pan-European Information System on Environmental Informatics Research Programmes and Projects  
Werner Geiger, Richard Lutz, Christian Schmitt ...... 460

Information System on Literature in the Field of ICT for Environmental Sustainability  
Martin Schreiber........................................................................................................ ....... 469

A Workflow-based Compliance Assistant for Facilities in EU Emission Trading System  
Philip Joschko, Christian Schmitz, Bernd Page, Nicolas Denz ............... 474

Modeling a simple trigeneration system: Environmental costs  
Monica Carvalho, Miguel A. Lozano, Luis M. Serra, Volker Wohlgemuth ............................................. 483

REACH-IT: The European regulation on chemicals and the impact of information technology  
Gerlinde Knetsch .............................................................................................................. 492

Theoretical Environmental Disciplines in Support of Sustainable Chemistry  
Kristina Voigt, Hagen Scherb, Rainer Brüggemann ................................................................. 499

S05. Modeling and deciding with stakeholders

An integration between Cognitive Map and Bayesian Belief Network for conflicts analysis in drought management  
R. Giordano, E. Preziosi, E. Romano ............... 507

Build collaborative models or capacity? Reflections from two years on  
J.L. Ticehurst, C. A. Pollino .............................................................................................. 516

Decision Support for Stakeholders  
William Silvert ............................................................................................................... 523

Effective engagement of stakeholders in Total Maximum Daily Load development and Implementation  
Erica Gaddis, Carl Adams, Alexey Voinov ................................................................. 530

Engaging stakeholders for a software development project: River Manager model  
Wendy D. Welsh, Dugald Black ...................................................................................... 539

From prediction to learning: the implications of changing the purpose of the modelling activity  
M. Brugnach ...................................................................................................................... 547
Gaming as the method to integrate modeling and participatory approaches in Interactive Water Management  Qiqi Zhou, Igor Mayer ................................. 554

Impacts of environmental factors and shellfish practices on the Mont-St-Michel Bay ecosystem (France): a modelling study involving scientists and local stakeholders Philippe Cugier, Caroline Struski, Karine Grangere, Michel Blanchard, Katia Frangoudès, Thierry Robin, José Pérez, Rémi Mongrue, Joseph Mazurié, Patrick Le Mao, Guy Fontenelle .............................................................................................. 562

Modelling the Complexity of the Cropping Plan Decision-making J. Dury, F. Garcia, A. Reynaud, O. Therond, JE. Bergez .......................................................... 569

Modelling trajectories of urban shrinkage – involvement and role of local stakeholders Haase, A., Haase, D., Kabisch, N., Rink, D., Kabisch, S. ............... 577

Participation in multi-criteria decisions: a software tool and a case study Eliot Laniado, Alessandro Luè, Simona Muratori ................................................................. 584

Participatory Planning for Climate Change Adaptation in the Brahmatwinn Project Carlo Giupponi, Valentina Giannini ................................................................. 593

Qualitative reasoning in participatory spatial planning: the use of OSIRIS in the Yellow River Delta P.J.F.M. Verweij, M. van Eupen, J. Roos-Klein Lankhorst, W. Nieuwenhuizen ................................................................. 601

The contribution of modelling and stakeholder engagement to transdisciplinary landscape research: an exploration of the Landscapes Toolkit Iris C. Bohnet... 609

The Use of Experts’ Judgements to Assess Agri-Environmental Policies in the Venice Lagoon Watershed: a Bayesian Network Approach Carpani Marta........... 617


S06. Ecosystem Services Concept for Environmental Management.............. 633

Mapping, Modelling and Managing Ecosystems Services in New Zealand Daniel T. Rutledge, John Dymond, Suzie Greenhalgh, Anne-Gaelle Ausseil, Robyn Sinclair, Alexander Herzig, Fraser Morgan, Robbie Andrew, Alison Collins ............... 633

Evaluating ecosystem services of Afforestation on Erosion-prone Land: A Case Study in the Manawatu Catchment, New Zealand Ausseil A.-G.E, Dymond J.R. 641

Management of Agricultural Nitrogen Losses in European Landscapes: An Ecosystem Modelling Case Study Tommy Dalgaard, Nick Hutchings and Chris Kjeldsen ................................................................. 650

Identifying optimum strategies for agricultural management considering multiple ecosystem services and climate change A. Holzkämper, P. Calanca, J. Fuhrer .. 658


Multi-Scale Modelling of Ecosystem Services – an Iterative Approach Castellazzi, M.S., Brown, I., Poggio, L., Gimona, A. ................................................................. 683

Quantifying Ecosystem Service Trade-offs Sven Lautenbach, Martin Volk, Bernd Gruber, Carsten F. Dormann, Michael Strauch, Ralf Seppelt ........................................ 692

S07. Spatial agent-based models for socio-ecological systems......................... 700

A Spatial Agent-Based Model to Explore Scenarios of Adaptation to Climate Change in an Alpine Tourism Destination Stefano Balbi, Pascal Perez, Carlo Giupponi.................................................................................................. 700

Agent-based model of the growth of an informal settlement in Dar es Salaam, Tanzania: An empirically informed concept Gina Felicia Young, Johannes Flacke ........................................................................................................ 708

An Agent Based Model of Climate Change and Conflict among Pastoralists in East Africa Atesmachew B. Hailegiorgis, William G. Kennedy, Mark Rouleau, Jeffrey K. Bassett, Mark Coletti, Gabriel C. Balan, Tim Gulden ........................................ 716

Agent-based modeling of a rental market for agricultural land in the Argentine Pampas Federico Bert, Guillermo Podestá, Santiago Rovere, Michael North, Angel Menéndez, Carlos Laciana, Charles Macal, Elkke Weber, and Pamela Sydelko ... 724

An Agent-Based Model of Coupled Housing and Land Markets Nicholas Magliocca, Elena Safirova, Virginia McConnell, and Margaret Walls .......................... 732

An Agent-based Model of Housing Search and Intraurban Migration in the Twin Cities of Minnesota Shipeng Sun, Steven M. Manson ......................... 740

An agglomeration payment for cost-effective biodiversity conservation in spatially structured landscapes Martin Drechsler, Frank Wätzold, Karin Johst, Jason F. Shogren ........................................................................................................ 748

Towards Adaptive Control of Landscape Biodiversity J. Gary Polhill, Andrew Jarvis, Alessandro Gimona, Nicholas M. Gotts .......................................................... 757

Exploring Forest Management Practices Using an Agent-Based Model of Forest Insect Infestations Liliana Pérez, Suzana Dragicevic ........................................ 766

Exploring the Choice of Decision Making Method in an Agent Based Model of Land Use Change A. R. Cabrera, P. J. Deadman, E. S. Brondizio and M. Pinedo-Vasquez ........................................................................................................ 774

Integrating land markets, land management, and ecosystem function in a model of land change Derek T. Robinson, Tatiana Filatova, Shipeng Sun, Rick L. Riolo, Daniel G. Brown, Dawn C. Parker, Meghan Hutchins, William S. Currie, Joan I. Nassauer .............................................................. 782

Knowledge-Brokering with Agent-Based Models: Some Experiences from Irrigation-Related Research in Chile Thomas Berger, Chris Schilling, Christian Troost, Evgeny Latynskiy ........................................................................ 791

MIRANA: a socio-ecological model for assessing sustainability of community-based regulations Sigrid Aubert, Jean-Pierre Müller, Julliard Ralihalizara .................... 801
Analysis of Incentive Schemes for Biodiversity Using a Coupled Agent-Based Model of Land Use Change and Species Metacommunity Model  J. Gary Polhill, Alessandro Gimona, Nicholas M. Gotts ................................................................. 809

Urban shrinkage: a vicious circle for residents and infrastructure? - Coupling agent-based models on residential location choice and urban infrastructure development  Nina Schwarz, Dagmar Haase ................................................................. 817

S08. Integrating surface water quality models at the basin scale .................. 825
Assessing the suitability of two model structures for basin resources estimation in the Mpologoma basin, within the Upper Nile  M. Kigobe, A. van Griensven, DL Shrestha, D. Solomatine, G. DiBaldassarre ................................................................. 825

AutoCalibration of Integrated Lake Water Quality Models for the Lake Winnipeg Basin Initiative  Markiyan Sloboda, David Swayne, William Booty, Craig McCrimmon, Isaac Wong................................................................. 833

CMLS94 Trends on Water Risk Contamination by Pesticide at Two Field Crops over Areas of Aquifers at Amazon and Southeast Regions, Brazil  Maria Conceição Peres Young Pessoa, Ricardo de Oliveira Figueiredo, Sonia Claudia Nascimento de Queiroz, Marco Antonio Ferreira Gomes, Anderson Soares Pereira, Eduardo Jorge Maklouf Carvalho, Vera Lucia Ferracini, Manoel Dornelasde Souza, Lilianne Maia Lima, Fábio Monteiro Cruz ................................................................. 841


Advances in the Integration of Watershed and Lake Modeling in the Lake Winnipeg Basin  Leon, L.F., Booty, W., Wong, I., McCrimmon, C., Melles, S., Benoy, G., Vanrobaeys, J................................................................. 860

Impact of land-use changes on the hydrological processes in the Elbow river watershed in southern Alberta  G. N. Wijesekara, A. Gupta, C. Valeo, J. G. Hasbani, and D. J. Marceau ................................................................. 868

Integrated river assessment by coupling water quality and ecological assessment models  Ine S. Pauwels, Gert Everaert and Peter L.M. Goethals ......................... 876

Modelling Climate Impacts on Hydrologic Processes in the Lake Winnipeg Watershed  Rajesh Shrestha, Yonas Dibike and Terry Prowse ................. 885

Modelling Connectivity Between Pollutant Source Areas & Streams  Moshirvaziri S., Sheridan G., Lane P., Jones O................................................................. 894

Modelling scenarios of agriculture changes on freshwater uses and water quality at a large watershed scale – the case of the Charente watershed (France)  Françoise Vernier, Paul Bordenave, Marie Chavent, Odile Leccia, Kevin Petit ................. 902

Scenarios to Investigate the Effect of Wetland Position in a Watershed on Nutrient Loadings  S.J. Melles, G. Benoy, B. Booty, L. Leon, J. Vanrobaeys, and I. Wong 911

Letsmap do Brasil - a web-based planning support tool for sediment management in Western Central Brazil  Lorz, Carsten, Fürst, Christine, Pietzsch, Katrin, Makeschin, Franz ................................................................. 919
The Climatic Change Impact in Water Potential Processes on the Albanian Hydrographic River Network  Niko Pano, Alfred Frashei, Bardhyl Avdyli........... 927

The application of a validated hydrodynamic model to improve the water management of an Egyptian shallow water coastal lake  M A Bek, I S Lowndes, D M Hargreaves ................................................................. 934

S09. Modelling for Northern Environment’s Sake: The Models in Need and The Models in Deed ................................................................. 952

Agent Based Modelling of Caribou Environmental Interactions in the Canadian Arctic  Glen Lesins, Kaz Higuchi ................................................................. 952

Modelling Governance Structures and Climate Change Policy Communications on Community Resilience in the Canadian Arctic  Rachel A. Hirsch ...................... 959

Modelling needs assessment for social adaptation to climate change in Siberia  G.A. Alexandrov, G. Inoue, T. Matsunaga ............................................................. 966

S10. Integrated Modeling Technologies ......................................................... 972

Modelling and control of a hybrid renewable energy system to supply demand of a green-building  H. Dagdougui, R. Minciardi, A. Ouammi, M. Robba, R. Sacile .. 972

A generic framework for multi-disciplinary environmental modelling  Rolf Hennicker, Sebastian S. Bauer, Stephan Janisch, and Matthias Ludwig ........... 980


Advances in integrated hydrological modeling with the LIQUIDR framework  F. Branger, S. Debionne, P. Viallet, I. Braud, S. Jankowfsky, O. Vannier, F.Rodriguez, S. Anquetin .................................................................1003

An Integrated Assessment approach to linking biophysical modelling and economic valuation tools  M.E. Kragt, J.W. Bennett, A.J. Jakeman .............................................................1011

An integrated, multi-modelling approach for the assessment of water quality: lessons from the Pinios River case in Greece  C. Makropoulos, E. Safiolea, S. Baki, E. Douka, A. Stamou, and M. Mimikou .................................................................1023

An interdisciplinary framework for the development and analysis of crop models  Sebastien Roux, Myriam Adam, Jacques Wery .............................................................1033

An open-source model platform for water management that links models to a generic user-interface and data-manager  Julien J. Harou, Didrik Pinte, Amaury Tilmant, David E. Rosenberg, David E.Rheinheimer, Kristiana Hansen, Patrick M. Reed, Arnaud Reynaud, JosueMedellin-Azuara, Manuel Pulido-Velazquez, Evgenii Matrosov, Silvia Padula, Tingju Zhu .................................................................1041

DeltaShell - an open modelling environment  Gennadii Donchyts, Bert Jagers 1050

Development of a suburban catchment model within the LIQUID framework  Jankowfsky S., Branger F., Braud I., Viallet P., Debionne S., Rodriguez F. ........1058

End-to-End Workflows for Coupled Climate and Hydrological Modeling  Kathy Saint, Sylvia Murphy .................................................................1066
<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>From OpenMI 1.4 to 2.0</td>
<td>Peter Gijsbers, Stef Hummell, Stanislav Vaneček, Jesper Groos, Adrian Harper, Rob Knapen, Jan Gregersen, Peter Schade, Andrea Antonello, Gena Donchyts</td>
<td>1081</td>
</tr>
<tr>
<td>Implementation of a Land Suitability Rating System as Open Web Services</td>
<td>Peter Schut</td>
<td>1089</td>
</tr>
<tr>
<td>Integrated Meta-Modelling for Decision Support in Integrated Catchment Management</td>
<td>Vikas Kumar, Annelie Holzkaemper, David Lerner</td>
<td>1105</td>
</tr>
<tr>
<td>Integrated Model of Municipal Waste Management of the Czech Republic</td>
<td>J. Hřebíček, Jana Soukopová</td>
<td>1113</td>
</tr>
<tr>
<td>Integrated Modeling for Source Characterization of Pathogenic Contamination in Watersheds</td>
<td>Michael Tryby, S. Thomas Purucker, Gene Whelan</td>
<td>1121</td>
</tr>
<tr>
<td>Integrated Modelling for Health and Environmental Impact Assessment of Air Pollution and Climate Change</td>
<td>Stefan Reis, Tim Oxley, Ed Rowe</td>
<td>1129</td>
</tr>
<tr>
<td>'Integronsters' and the special role of data</td>
<td>Alexey Voinov</td>
<td>1139</td>
</tr>
<tr>
<td>Linking Data, Models and Tools: An Overview</td>
<td>H.R.A. (Bert) Jagers</td>
<td>1150</td>
</tr>
<tr>
<td>Methods to Register Models and Input/Output Parameters for Integrated Modeling</td>
<td>James G. Droppo, Gene Whelan, Michael E. Tryby, Mitchell A. Pelton, Randel Y. Taira, Kevin E. Dorow</td>
<td>1158</td>
</tr>
<tr>
<td>Modeling freshwater uses in coastal areas – the case of Pertuis Charentais</td>
<td>Cédric Bacher, Johanna Balle-Beganton, Julien Neveu, Harold Rethoret, Françoise Vernier, Paul Bordenave, Rémi Monguel, José Perez</td>
<td>1166</td>
</tr>
<tr>
<td>OpenMI 2.0 - What's new?</td>
<td>Gennadii Donchyts, Stef Hummell, Stanislav Vaneček, Jesper Groos, Adrian Harper, Rob Knapen, Jan Gregersen, Peter Schade, Andrea Antonello, Peter Gijsbers</td>
<td>1174</td>
</tr>
<tr>
<td>Parallel computing of a large scale spatially distributed model using the Soil and Water Assessment Tool (SWAT)</td>
<td>S. G. Yalew, A. van Griensven, L. Kokoszkiewicz</td>
<td>1182</td>
</tr>
<tr>
<td>Rethinking modeling framework design: Object Modeling System 3.0</td>
<td>O. David, J.C. Ascough II, G.H. Leavesley, L. Ahuja</td>
<td>1190</td>
</tr>
<tr>
<td>TaToo: tagging environmental resources on the web by semantic annotations</td>
<td>Andrea E. Rizzoli, Gerald Schimak, Marcello Donatelli, Jiri Hrebíček, Giuseppe Avellino, Jose Lorenzo Mon, Ioannis Athanasiadis</td>
<td>1199</td>
</tr>
<tr>
<td>The use of ontologies in peer reviews of Integrated Assessment Models</td>
<td>M. de Vos, N. Koenderink, B. van Ruijven, J. Top</td>
<td>1207</td>
</tr>
</tbody>
</table>
Towards model component reuse for the design of simulation models – a case study for ICZM  Jean-Luc de Kok, Guy Engelen, Joachim Maes...............................1215

Using an Integrated, Multi-disciplinary Framework to Support Quantitative Microbial Risk Assessments  Gene Whelan, Michael E. Tryby, Michell A. Pelton, Jeffery A. Soller, Karl J. Castleton.................................................................1223

S11. Sustainability Appraisal – Concepts, Tools and Outcomes .........................1231

Farmers as water managers in a changing climate – can we turn sustainability research into outcomes? Matthews, K.B., Blackstock, K.L., Langan, S., Rivington, M., Miller D.G., Buchan, K. .................................................................1231

GIS-based Spatial Hydrological Zoning for Sustainable Water Management of Irrigation Areas  Yun Chen, Jianyao Chen, Emmanuel Xevi, Mobin-ud-Din Ahmad, Glen Walker ..........................................................................................1240

Modelling Urban Land Use Change Using Geographically Weighted Regression and the Implications for Sustainable Environmental Planning  Noresah Mohd Shariff, Sanjay Gairola, Anita Talib .................................................................1249

Informing Regional Planning in Alberta’s Oilsands Region with a Land-use Simulation Model  Matthew Carlson, Terry Antoniuk, Dan Farr, Shawn Francis, Karen Manuel, John Nishi, Brad Stelfox, Mika Sutherland, Cornel Yarmoloy, Craig Aumann, Daiyuan Pan ..........................................................................................1257

A new tool for integrated and interactive sustainability impact assessment of urban land uses change: the PLUREL iIAT  Dagmar Haase, Annette Piorr, Nina Schwarz, Ingo Zasada..........................................................................................1265

Lessons learnt on Requirement Analyses to establish new model systems: prerequisites to assure model use towards policy outcome  Stefan Sieber, Karen Tscherning, Klaus Müller, Peter Verweij, Torbjörn Jansson .................................................................1273

User interaction during the development of the Waikato Integrated Scenario Explorer  Hedwig van Delden, Derek Phyn, Tony Fenton, Beat Huser, Daniel Rutledge, Liz Wedderburn ..........................................................................................1283

Empathy: a Unifying approach to address the dilemma of 'Environment versus Economy’  A.Guergachi, O. Ngenyama, V. Magness, J. Hakim.................................................................1292

Formalizing knowledge on international environmental regimes for integrated assessment modeling  M. de Vos, P. Janssen, S. Frantzi, P. Pattberg, A. Petersen, F. Biermann, M. Kok..........................................................................................1300

World Sustainability in the climate policy sphere: a scenario assessment of different commitments for curbing CO2 emissions in a Sustainability Framework  Carlo Carraro, Marinella Davide, Elisa Lanzi, Ramiro Parrado .................................................................1310

A Taguchi-Based Method for Assessing Data Center Sustainability  Sidney L. Pendelberry, Sophie Ying Chen Su, Michael Thurston.................................................................1327

Evaluation as a Tool to Support Decision Making in Municipal Waste Recycling System for Thailand Sustainable Environment  Vilas Nitivattananon and Yuthtapong Wattanalapa ..........................................................................................1335
Sustainability, information needs and organisational change in UK water and sewerage companies Tanner, A., McIntosh B.S., Seth, A., Widdowson, D. .....1343

S12. Environmental Monitoring: Methods, Models, Designs and Criteria of Efficiency ..................................................................................................................1352

A Low-cost Automatic Water Sampler Equipment Lúcio Flávio Ferreira Moreira, Jean Oliveira de Paiva ..................................................................................................................1352

A Study of Variance Estimators for Material Sampling Using Computerized Models of Contaminant Heterogeneity in Soil Stockpiles B. Geelhoed, F.P.J. Lamé .....1357

Decision Support for Environmental Monitoring and Restoration: Application of the Partially Observable Markov Decision Process David Tomberlin ..............................1371

Assessing the Value of Environmental Observations in a Changing World: Nonstationarity, Complexity, and Hierarchical Dependencies Patrick M. Reed, Ruchit Shah, Joshua B. Kollat ..............................................................1379

Efficiency Criteria for Water Quality Monitoring Marina G. Erechotchoukova, Peter A. Khaiter ..................................................................................................................1387

Pitimbú River Lowland Portion Water and Sediment Monitoring Data, Natal Brazil Rafael Takeshi Kobayashi, Lúcio Flávio Ferreira Moreira, Herbert Tadeu de Almeida Andrade ..................................................................................................................1397

Remote Sensing Time Series for Modeling Invasive Species Distribution: A Case Study of Tamarix spp. in the US and Mexico Anna Cord, Doris Klein, Stefan Dech ..................................................................................................................1404

Using Generalized Additive Models to Assess, Explore and Unify Environmental Monitoring Datasets Russell G. Richards, Rodger Tomlinson, Milani Chaloupka ..................................................................................................................1412


Development of a land-use component for an integrated model of the German biogas system A. Schaldach, D. Lapola, U. Jessen, M. Thees, R. Schaldach .....1421

Sustainable forest management for bioenergy Giulia Fiorese, Giorgio Guariso ..................1430

S14. Modelling The Coupled Social-Environmental and Physical Systems of Urban Water ..................................................................................................................1438

An interactive visual decision support tool for sustainable urban river corridor management Lewis Gill, Vikas Kumar, Eckart Lange, David Lerner, Ed Morgan, Daniela Romano, Ed Shaw ..................................................................................................................1438

Bayesian Networks and Social Objectives: A Canoeing Case Study Ed Shaw, Vikas Kumar, David Lerner, Eckart Lange ..................................................................................................................1446

Modelling the ecological impact of discharged urban waters upon receiving aquatic ecosystems. A tropical lowland river case study: city Cali and the Cauca river in Colombia Holguin Gonzalez J., Goethals P.L.M. .........................................................................................1454

S15. Integrated Assessment Modelling of Air Quality and Climate Change across Different Spatial and Temporal Scales ..................................................................................1463
Modelling the seasonal climate effects on grapevine yield at different spatial and unconventional temporal scales  *Subana Shanmuganathan, Philip Sallis, Ajit Narayanan* .................................................................1463

**S16. Feedbacks in Socio-Environmental Systems**.................................1473

Deceiving Feedbacks: The challenge of policy design for inshore fishery activities in complex ecosystems. The case of the Ciénaga Grande de Santa Marta  *V. Otero, C. Olaya, C. Castro Sanguino* .................................................................1473

Dynamics of habitat banking under changing conservation costs and habitat restoration time lags  *Martin Drechsler, Florian Hartig* .................................1481

Feedbacks in socio-environmental land systems  *Joerg A. Priess, Nina Schwarz, Sven Lautenbach* ........................................................................1489

Irrigation management in Chile: integrated modeling of access to water  *T. Arnold, H. Uribe, C. Troost, T. Berger* .................................................................1499

Land use change modelling in an urban region with simultaneous population growth and shrinkage including planning and governance feedbacks  *Dagmar Haase* ..................................................................................................................................1507

Linking wildfire behaviour and land-use modelling in Northern Mongolia  *C. Schweitzer, J.A. Priess* ........................................................................1516

Modeling the adaptation of land-use decisions to landscape changes using an agent-based system: a case study in a mountainous catchment in central Vietnam  *Quang Bao Le, Roman Seidl, Roland W. Scholz* .................................................................1524

Risk management in an uncertain environment - a study from semi-arid grazing systems  *O. Jakoby, M. F. Quaas, B. Müller, K. Frank* ........................................................................1532

Towards the integration of economic and land use change Models  *Hedwig van Delden, Garry McDonald* ........................................................................1540

**S17. Scientific Workflow Tools for the Environmental Domain - Technologies and Applications** ........................................................................1549

A scientific workflow tool for biosphere modelling: Carbon Sink Archives  *G. A. Alexandrov and T. Matsunaga* .................................................................1549

Exposing the Kepler Scientific Workflow System as an OGC Web Processing Service  *Andrew Pratt, Chris Peters, Siddeswara Guru, Brad Lee, Andrew Terhorst* ........................................................................1554

Hydrologists Workbench - a hydrological domain workflow toolkit  *Susan M. Cuddy, Peter Fitch* ........................................................................1562

Hydrologists workbench: A governance model for scientific workflow environments  *Paul Box* .................................................................................................1570

On the appropriate granularity of activities in a scientific workflow applied to an optimization problem  *Perraud, Jean-Michel, Qifeng Bai, David Hehir* ........................................................................1578

Opportunities and limitations of DelftFEWS as a scientific workflow tool for environmental modeling  *Peter Gijsbers* ........................................................................1587
Uncertainty of a biological nitrogen and phosphorus removal model  A. Cosenza, G. Mannina, G. Viviani .......................................................... 1753

A comparison of Fuzzy, Bayesian and Weighted Average formulations of an in-stream habitat suitability model  C. J. Brookes, Vikas Kumar, S. N. Lane .......... 1762

Limitations in Interpretable Top-Down Effective Rainfall-Runoff Modelling  J. P. Norton .................................................................................. 1770

Robust estimation of the total unit hydrograph  F. T. Andrews, B. F. W. Croke, K. W. Jeanes ......................................................... 1779

The Use of Entropy as a Model Diagnostic in Rainfall-Runoff Modelling  I. G. Pechlivanidis, B. Jackson, H. McMillan .......................................................... 1787

Propagating Data Uncertainty and Variability into Flow Predictions in Ungauged Basins  Andrew D. Gronewold, Ibrahim M. Alameddine ...................... 1795

S21. Intelligent and Collaborative Engineering of Environmental Knowledge: Software Platforms, Agents and Semantics – A Special Session Organized by I-Seek, International Workshop on "Intelligent Systems For Environmental (Knowledge) Engineering and Ecoinformatics" .................................................. 1812

Predicting Polycyclic Aromatic Hydrocarbon Concentrations in Soil and Water Samples  Geoffrey Holmes, D. Fletcher, P. Reutemann ...................... 1812

Forest Fire Sensing and Decision Support using Large Scale WSNs  Evangelia Kolega, Vassilios Vescoukis, Christos Douligeris ................................................. 1820

Modelling Environmental Impact with a Goal and Causal Relation Integrated Approach  He Zhang, Lin Liu ................................................................. 1828

Reasoning about Actions for the Management of Urban Wastewater Systems using a Causal Logic  Dario Garcia-Gasulla, Juan Carlos Nieves and Ulises Cortés .... 1841

Semantic mediation of an integrated assessment tool for agricultural systems  Sander Janssen, Ioannis N. Athanasiadis, Erling Andersen, Andrea E. Rizzoli, Jan-Erik J.F. Wien and Martin K. van Ittersum ............................................. 1849

Towards a service-oriented e-infrastructure for multidisciplinary environmental research  Ayalew Kassahun, Ioannis N. Athanasiadis, Andrea E. Rizzoli, Arno Krause, Huub Scholten, Marek Makowski, Adrie Beulens ........................................... 1858


A framework for integrated modeling using a knowledge-driven approach  Jean-Pierre Müller ............................................................................. 1877

S22. Interaction Design for Environmental Information Systems ..................... 1885

A review of user interface conventions in web applications for climate change information  Markus Wrobel, Luís Costa, Tabea Lissner, Marta Moneo Lain, Tobias Weiss, Jürgen Kropp ................................................................. 1885
Agilists and the Art of Integrated Assessment Tool Development  Rob Knapen, Peter Verweij, Sander Janssen ........................................................................................................................................1894

Aquisys – A Computer System to Support Good Practices of Management for Brazilian Tilapiaiculture  Maria Conceição Peres Young Pessoa, Ana Flávia Rodrigues Seixas, Marcos Eliseu Losekann, Julio Ferraz de Queiroz, Mariana Silveira Guerra Moura e Silva, Carlos Pazzianotto ........................................................................1902

Design and Development of S.O.L.E. Software  Daryl H. Hepting ..........................................................1910

Social Shopping Using Food Spimes  Timothy Maciag, Daryl H. Hepting, JoAnn Jaffe, Katherine Arbuthnott, Darryl Dormuth ..................................................................................................1917

S23. Second Session on Data Mining as a Tool for Environmental Scientists (S-DMTES-2010) ..........................................................................................................................1925

An approach to water supply clusters by semi-supervised learning  M. Herrera a, S. Canu, A. Karatzoglou, R. Pérez-García and J. Izquierdo ..........................................................1925

Analysis of the Economic Sustainability of Companies in the Water Sector  Lyubomir Halachev, Aziz Guergachi, Yashodhan Athavale, Sri Krishnan ........................................1933

Choosing the right data mining technique: classification of methods and intelligent recommendation  Karina Gibert, Miquel Sánchez-Marrè, Víctor Codina ..........................1940

Development of data-driven models for the assessment of macroinvertebrates in rivers in Flanders  Gert Everaert, Ine S. Pauwels, Peter L. M. Goethals ................................1948

Integration of Statistical and Machine Learning Models for Short-term Forecasting of the Atmospheric Clearness Index  L. Mora-Lopez, M. Piliougine, J.E. Carretero and M. Sidrach-de-Cardona ...........................................................................................................1957

Renyi’s-entropy-based Approach for Selecting the Significant Input Variables for the Ecological data  Can-Tao Liu, Bao-Gang Hu ........................................................................1965

The Application of Artificial Neural Network for Forecasting Dam Spillage Events  Anita Talib, Yahya Abu Hasan .................................................................................................1973

The tasks of pre and post-processing in Data Mining applied to a real world problem  José Luis Díaz, Manuel Herrera, Joaquín Izquierdo, Rafael Pérez-García ........................................................................................................1980

S24. Intelligent Environmental Decision Support Systems .................................................................1990

A multi-agent framework for an IEDSS in urban water management  Joaquín Izquierdo, Idel Montalvo, Rafael Pérez-García, Joanna A. Gutiérrez-Pérez .........1999

Development of a DSS for the generation of WWTP configuration alternatives  M. Garrido, A. Riu, X. Flores-Alsina, I. Rodríguez-Roda, M. Poch ........................................1999

Development and validation of a decision support system for the integrated operation of membrane bioreactors  Joaquim Comas, Hèctor Monclús, Giuliana Ferrero, Ignasi Rodriguez-Roda, Luis Sancho and Eduardo Ayesa .............................2007

Evolving GESCONDA to an intelligent decision support tool  Miquel Sánchez-Marrè, Karina Gibert, Beatriz Sevilla ..............................................................................................2015

An integrated model to study environmental, economic, and energy trade-offs in intermodal freight transportation  J. Scott Hawker, Bryan Comer, James J. ........................2020
Corbett, Arindam Ghosh, Karl Korfmancher, Earl E. Lee, Bo Li, Chris Prokop, James J. Winebrake ............................................................... 2025
Agent Swarm Optimization: a paradigm to tackle complex problems. application to water distribution system design Idel Montalvo, Joaquín Izquierdo, Silvia Schwarze, Rafael Pérez-García ........................................................ 2035
A new data-based modelling method for identifying parsimonious nonlinear rainfall/flow models V. Laurain, M. Gilson, S. Payraudeau, C. Grégoire, H. Garnier ........................................................................................................ 2044
CARMA: scalability with approximate-model-based adaptation John D. Hastings, Alexandre V. Latchininsky, Scott P. Schell ......................................................................................... 2053
Providing intelligent decision support systems with flexible data-intensive case-based reasoning Beatriz Sevilla Villanueva, Miquel Sánchez-Marrè ............ 2063
Fuzzy logic based IEDSSs for environmental risk assessment and management Zabeo A., Semenzin E., Torresan S., Gottardo S., Pizzol L., Rizzi J., Giove S., Critto A., Marcomini A. ................................................................................................... 2073
Automatic interpretation of classes for improving decision support K. Gibert, A. Pérez-Bonilla ........................................................................................................ 2079
The value of using Bayesian networks in environmental decision support systems to support natural resource management W.S. Merritt, J.L. Ticehurst, C. Pollino, and B. Fu ........................................................................................................ 2087
S25. Managing Regional Water Resource Systems Under Changing Conditions... 2096
Assessing hydrological response to change in climate: statistical downscaling and hydrological modelling within the upper Nile M. Kigobe, A. van Griensven .... 2096
Comparison of climate change impacts and development effects on future Mekong flow regime Chu Thai Hoanh, Kittipong Jirayoot, Guillaume Lacombe, Vithet Srinetr ........................................................................................................ 2114
Coupling hydrological models and weather forecast for improved real-time management of water resources A. Salvetti, F. Pianosi, E. Weber, R. Soncini-Sessa ........................................................................................................ 2123
Development of a Decision Support System (ElmaaDSS) for the integrated water management of Mediterranean phosphate mining areas Soulsis Konstantinos, Dercas Nicholas, Bru Kathy, Graveline Nina, Stefopoulou Angeliki ............... 2131
Integrated modelling for the conservation of river ecosystems: progress in the South Australian River Murray Bryan, B.A., Overton, I., Higgins, A., Holland, K., Lester, R.E., King, D., Nolan, M., Connor, J.D., Hatton MacDonald, D., Oliver, R., Lorenz, Z., Bjornsson, T., Waanders, P. ........................................................................................................ 2139
Intercomparison of generic simulation models for water resource systems Giovanni M. Sechi, Andrea Sulis ........................................................................................................ 2150
Many-objective Management of Population and Drought Risks: A Case for De Novo Programming  
Patrick M. Reed, Joseph R. Kasprzyk, Brian R. Kirsch, Gregory W. Characklis ................................................................. 2160

A procedure for the quantitative assessment of water resources management under climate change and the design of adaptation measures  
D. Anghileri, F. Pianosi, R. Soncini-Sessa and E. Weber ................................................................. 2166

Optimal water allocation in the Zambezi Basin  
A. Tilmant, W. Kinzelbach, L. Beevers, D. Juizo ................................................................. 2174

Simulating the Thames water resource system using IRAS-2010  
Evgenii S. Matrosov, Julien J. Harou ................................................................. 2184

Sustainable watershed management by fuzzy game optimization  
Chih-Sheng Lee, Chun-Yun Yang, Shui-Ping Chang, Yi-Chao Lee ................................................................. 2196


Analysis of rising sludge risk in Activated Sludge Systems: from operational strategies to clarifier design  
Xavier Flores-Alsina, Joaquim Comas, Ignasi Rodriguez-Roda ................................................................. 2204

Integration of mathematical models in a decision support system for control of priority pollutants in urban catchments  
Natasa Atanasova, Matej Cerk, Primoz Banovec, Mateja Skerjanec, Webbey de Keyser, Lorenzo Benedetti ................................................................. 2212

Multi-criteria evaluation of control strategies in WWTP removing organic carbon, nitrogen and phosphorus  
Xavier Flores-Alsina, Lluís Corominas, Peter A. Vanrolleghem ................................................................. 2221

Nonlinear optimization for improving the operation of sewer systems: the Bogotá Case Study  
J. M. Giraldo, S. Leirens, M. A. Díaz-Granados, J. P. Rodríguez ................................................................. 2229

PCA intelligent contribution analysis for fault diagnosis in a sequencing batch reactor  
Alberto Wong Ramírez, Joan Colomer Llinás, Marta Coma, Jesús Colprim ................................................................. 2237

Uncertainty estimation of a complex water quality model: GLUE vs Bayesian approach applied with Box – Cox transformation  
Gabriele Freni, Giorgio Mannina ................................................................. 2245

S27. Linkages in Human and Environmental Health Modelling ................................................................... 2257

A geosimulation tool to assess intervention scenarios in relation to the spread of infectious disease and environmental factors  
Mondher Bouden, Bernard Moulin, Pierre Gosselin ................................................................. 2257

Evaluating the impact of climate change on global HPAI H5N1 outbreaks  
Zhijie Zhang, Dongmei Chen, Wenbao Liu, Lei Wang ................................................................. 2265

Modeling dynamical temperature influence on tick Ixodes scapularis population  
Xiaotian Wu, Venkata R.S.K Duvvuri, Jianhong Wu ................................................................. 2272

Modelling human health by levels of biological organization  
Wilfred R. Cuff, Clifford Clark, Paulette Richard, Venkata R. S. K. Duvvuri ................................................................. 2288
Modelling the spread of influenza-like illnesses in an urban environment  R. Neighbour, M. Borkowski, S. Mukhi, M.R. Friesen and R.D. McLeod ..........2299

Multi-disciplinary development of an EarlyWarning and Automated Response System (EWARS) for epidemic prevention Ajit N. Babu, Engelbert Niehaus .......2307

S28. Modelling Extremes in Climate Variables ...............................................2316

Dust Storms over the Arabian Gulf: An optimistic vision toward climate change consequences Waleed Hamza, Mohamed Rizk, Huda Al-Hassini, Jan Stuut, Dirk de-Beer...........................................................................................................2316

Extreme value analysis for gridded data Sanabria L.A., R. P. Cechet.........2317

Fast variable selection for extreme values A. Phatak, C. Chan, H. Kiiveri......2326

Modelling the effects of daily extreme weather on grapevine and wine quality Subana Shanmuganathan, Philip Sallis, Ajit Narayanan..........................2334

Modelling Extreme Events in a Changing Climate using Regional Dynamically-Downscaled Climate Projections Christopher J. White, Augusto Sanabria, Stuart Corney, Michael Grose, Greg Holz, James Bennett, Robert Cechet, Nathaniel L. Bindoff .................................................................................................2343

S29. Multi-Scale and Multi-Physics Modeling of Environmental Flows ............2351

CFD and GFD hybrid approach for simulation of multi-scale coastal ocean flow H. S. Tang, X. G. Wu ...........................................................................................................2351

A dynamic model of primary production and plant coverage in an oligotrophic tropical river Barbara J. Robson ..............................................................2359

Numerical modeling of the salinity distribution and environmental flows assessment in the Yellow River Estuary, China T. Sun, Z. F. Yang .................2366

Environmental flow requirements (EFRs) related to preference of phytoplankton in the Yellow River Estuary (YRE) based on an ecohydrodynamic model R. Zhao, Z. F. Yang, T. Sun .................................................................2372

Numerical simulation of the three-dimensional flow field in a reservoir using energy equation Yanfang Wang, Minquan Feng, Shiping Fan, Fangmin Zheng 2378

Parameter inversion model for two dimensional parabolic equation using Levenberg-Marquardt Method Tao Min, Zhulin Hao, Minquan Feng, Hansong Tang ...........................................................................................................2385

W01. Effects of climate change: landscape modelling, providing decision support, and understanding uncertainties .................................................................2393

Farm decisions under dynamic meteorology and the curse of complexity Thorsten Arnold .................................................................................................2393

Functional linear models to test for differences in prairie wetland hydraulic gradients Mark C. Greenwood, Richard S. Sojda, Todd M. Preston ...............2401

Risk assessment and decision support tools for the integrated evaluation of climate change impacts on coastal zones Torresan S., Zabeo A., Rizzi J., Critto A., Pizzol L., Giove S., Marcomini A. ......................................................................................2409
New software methods in radar ornithology using WSR-88D weather data and potential application to monitoring effects of climate change on bird migration
Reginald Mead, John Paxton, Richard S. Sojda ......................................................2416

W03. Modeling and Deciding with Stakeholders ..............................................2425
The contribution of modelling and stakeholder engagement to transdisciplinary landscape research: an exploration of the Landscapes Toolkit Iris C. Bohnet.2425
Modelling trajectories of urban shrinkage – involvement and role of local stakeholders Haase, A., Haase, D., Kabisch, N., Rink, D., Kabisch, S* ..............2434

W05. Modelling For Northern Environment’s Sake: the Models in Need and the Models in Deed .........................................................................................2441
The Northern-Global Climate Change Adaptation Dialogue J. I. MacLellan ......2441

W09. Feedbacks in Socio-Environmental Systems ..........................................2454
Feedbacks in socio-environmental land systems Organisers: Joerg A. Priess, Nina Schwarz, Sven Lautenbach .................................................................2454

W11. W-DMTES-2010. Third Workshop on Data Mining as a Tool for Environmental Scientists .................................................................................................2455
Integral support to environmental decision-making through GESCONDA Miquel Sàncchez-Marrè, Karina Gibert, Beatriz Sevilla ........................................2455
Choosing the right data mining technique: classification of methods and intelligent recommendation Karina Gibert, Miquel Sàncchez-Marrè, Víctor Codina ..............2457
Multi-Objective Optimization in Data Mining: An ASO Approach J. Izquierdo ..2465
Pre and post-processing on a Water Supply System Database J. Izquierdo ....2466

W12. Breaking down disciplinary silos: what can environmental modellers in different domains learn from each other? ..................................................2467
Integrating biophysical and economic data to support water allocations in the Murray Basin, Australia M. Ejaz Qureshi, Stuart Whitten ..........................2467
Overcoming data constraints to create meaningful ecological models Rebecca E Lester, Peter G. Fairweather .................................................................2476
What can we learn from systems based approaches: from systems biology to Earth systems science? C. J. A. Macleod .........................................................2485
Bayesian networks for modelling habitat suitability of an endangered species Serena H. Chen, Carmel A. Pollino .................................................................2493
A typological study on stability of structures in systems: case studies from socio-economics and ecology Vincent Lyne, Peter Last, Seu Keow Cheng, Tim Skewes, David Brewer, Donna Hayes, Warwick Jones ..............................................2503
W15. Tales of DSS adoption: How and why are DSS successful in environmental and related sectors? .................................................................2513

Decision support system for large-scale hydropower system operations  Cheng Chun-tian, Shen Jian-jian , Liao Sheng-li, Wu Xin-yu .........................2513

DSS and MAF (Multi-agencies framework) for sustainable water management Amgad Elmahdi, Don McFarlane .......................................................2521

Environmental information management systems as templates for successful environmental decision support  N. W. T. Quinn .............................2533

Solar wind and energy resource assessment (SWERA): a usability case study Lucas Michels, Omar El-Gayar, Matt Wills ........................................2541

The GPFARM DSS for Agroecosystem Sustainability: Past, Future, and Lessons Learned J.C. Ascough II, G.S. McMaster, G.H. Dunn, and A.A. Andales ..........2549

W17. Low-carbon industry and multi-scale input-output modeling ...............2557

The cumulative effects of dam project on river ecosystem based on multi-scale ecological network analysis  Chen Shaoqing, Yang Jin, Chen Bin ................2557

The comparison among four carbon footprint estimation boundaries of the sectors using Chongqing as a case  L.P. Ju, B. Chen ........................................2564


Ecological and economic analysis of planting greenhouse cucumbers with anaerobic fermentation residues  N. Duan, C. Lin, R.Y. Gao, Y. Wang, J.H. Wang, J.Hou .................................................................2578

Study on Integrated Simulation Model of Economic, energy and environment safety system under the low-carbon policy in Beijing  Liu Dacheng, Yang Xiaou, Tan Xianchun, Wu Ruihao, Wang Li .................................................................2586

Study on energy strategy of Chinese Capital Region under the New National Policy of Reducing Carbon Dioxide Emissions  Liu Dacheng, Li Ning, Tan Xianchun, Yang Xiaou, Wang Li, Liu Jianbing .................................................................2595

Application of Life Cycle Analysis (LCA) in evaluating energy-consuming of integrated oxidation ditch (IOD)  Xianchun Tan, Junyi Xu, Sheng Wang, C.Y. Miao .................................................................2603

CO2 embodied in China’s foreign trade 2007 with global policy implications  Lei Liu, Xiaoming Ma .................................................................2611

Study on the effect of biogas project on the development of low-carbon circular economy — A case study of Beilangzhong eco-village  N. Duan, C. Lin, X. D. Liu, Y. Wang, X. J. Zhang, Y. Hou .................................................................2619

Energy-based ecological economic evaluation of Beijing Urban Ecosystem  Geng-Yuan Liu, Zhi-Feng Yang, Bin Chen .................................................................2627
Evaluation of the changed properties of aquatic animals after dam construction using ecological network analysis  Chen Shaoqing, Liao Xueqin, Yang Jin, Chen Bin .................................................................................................................................2635


Integrated water resource security evaluation of Beijing based on Grey Relation Analysis and TOPSIS  Jing Dai, Jingjing Chi, Bin Chen, Liping Ju .......................2651

Is there causality from investment for real estate to carbon emission in China: a cointegration empirical study  Jinqiu Xu.................................................................2660

Limiting factor analysis of urban ecosystems based on energy—a case study of three cities in the Pearl River Delta in China  Meirong Su, Zhifeng Yang, Bin Chen ..................................................................................................................2667

Soil Total Organic Carbon, δ13C values and their responses to the soil core transferring experiment from high- to low-elevation forest along natural altitudinal transect of old temperate volcanic forest Soils  Zhang Xin-Yu, Meng Xian-Jing, Fan Jin-Juan, Gao Lu-Peng, Sun Xiao-Min .................................................................2675

AUTHOR INDEX ........................................................................ 943, 1803, 2683
Agent Based Modelling of Caribou Environmental Interactions in the Canadian Arctic

Glen Lesins\textsuperscript{a} and Kaz Higuchi\textsuperscript{b}
\textsuperscript{a}Department of Physics and Atmospheric Science, Dalhousie University (glen.lesins@dal.ca)
\textsuperscript{b}Faculty of Environmental Studies, York University (kazh@yorku.ca)

Abstract: Agent or individual based modelling (IBM) is recognized as an important tool in biological modelling for simulating animal behaviour and interactions with the environment at the level of the individual while maintaining a collective perspective. The method allows rules to be programmed for the interactions between like and unlike species, responses to geographic factors and impacts from human activities. By emphasizing the development of appropriate rules for the individual, one can avoid imposing ad hoc assumptions concerning the population as a whole and instead allow emergent phenomena to unfold at the larger scale. Here we report on preliminary results using IBM to simulate the movement and population changes of an idealized caribou herd in a Northern Canadian Arctic setting. A wolf population is included in the simulation to study how the predator-prey relations impact the fluctuations in total population. The simulations use GIS data to account for a varying landscape including topography, water bodies and vegetation. The challenges of developing the software components will be discussed including the relative merits of using C\# and NetLogo as programming languages. Future inclusion of Inuit hunter agents will be discussed.

Keywords: individual-based; modelling; agent; caribou; NetLogo

1. INTRODUCTION

Multi-agent or individual based modelling (IBM) simulates complex systems by applying rules for the interaction of agents or individuals with each other and with the environment. It is well-suited for applications where variation in the individuals’ attributes is required to better understand the behaviour of the population as a whole. It captures the richness associated with individual members evolving according to their own unique experiences and interactions, instead of imposing an average effect uniformly to the entire population or sub-population.

In an agent approach the programmer inputs rules of interaction at the level of the individual. This more closely simulates the way natural systems develop or evolve in time. Instead of imposing preconceived notions directly on the population at large, the programmer designs empirically-based rules for the individual and then allows the interactions to determine the consequences for the whole population. This is referred to as emergent phenomena which are common place for complex or nonlinear systems, including those in biology.
The IBM approach in computational ecology and biology is a well-established discipline (Grimm and Railsback, 2005). Early successes include simulating the social hierarchy of the green woodhoopoe (Neuert et al., 1995), the Beech forest model that reproduced spatial patterns (Neuert, 1999) and individual fish behaviour using a stream trout model (Railsback et al., 1999). Recent work has closely critiqued and improved the agent based approach in behavioural ecology (Lehmann, 2009). The method is well suited to predator-prey simulations where interactions at the individual level such as predation and reproduction determine the population periodicities and spatial pattern formations in large-scale populations (Wilson, 1996; Gui-Quan et al., 2008). The limitations of the classical Lotka-Volterra predator-prey model have been demonstrated using IBM comparisons (Donalson and Nisbet, 1999; Law et al., 2003).

Predator-prey interactions take place in a heterogeneous landscape with varying topography, vegetation, water supply and climate. These environmental factors can be included as parameters in the behavioural rules for the individual animals resulting in an important source of variability that needs to be captured in ecological simulations. Rohner and Demarchi (2000) developed an IBM as a tool for comparing forest policy options in which individual caribou were simulated in a GIS application. Metsaranta (2008) used an IBM to show that caribou’s preference for home ranges in mature coniferous forests appears driven by a refuge need from predation.

The current work has a long-term goal to better understand how changes in caribou movements and numbers may impact the Inuit peoples in the Canadian Arctic. Wild caribou is an important food and income source for the indigenous Peoples living in parts of Nunavut, Northern Quebec, Northwest Territories and Yukon. Environmental change forced by global warming and local human activities have the potential to disrupt the caribou herds. Berman et al. (2004) assessed various scenarios associated with economic and climate change by using an agent-based model of individuals and households in a hypothetical community in Old Crow, Yukon. The model appears to be able to project changes in economic, demographic and hunting conditions that is consistent with observed trends.

Predicting changes in the herd characteristics is challenging for a number of reasons. The predation from wolves and bears involve complex interaction biology, the strategies of the Inuit hunt can vary between settlements and even individuals, the landscape and food supply for the caribou have large geographic variations, and changes in weather patterns associated with global warming are already being observed. We are beginning a research project to test the IBM approach as a tool to simulate complex processes that have strong individual variability in a heterogeneous landscape.

In this presentation we report on a preliminary project to evaluate the tools that might be used to accomplish the longer-term objectives. The simulated system is simplified to a single predator and single prey, namely wolves and caribou, interacting in a tundra landscape typical of Northern Canada. The software must allow easy design of agent objects that can interact with a GIS (Geographic Information System) database. It must be able to handle many thousands, preferably hundreds of thousands of individuals, moving in a landscape 1000 km or more in horizontal extent. A time step of less than one hour is preferable to handle the movement of hunting wolves and migratory caribou. Total simulation time should ideally extend for one year to encompass an entire annual cycle with seasonal variations included. All these requirements put an enormous demand on the speed and memory of the software and hardware components.

2. METHODS

In this work two software languages will be considered: C# and NetLogo. C# is a modified version of C++ which is distributed by Microsoft as part of their .NET framework of languages. Its advantages include that it is type-safe and automatically deals with garbage collection when an object is deleted. C# design of the classes needed for predator-prey
simulations gives the programmer total control of a fully object-oriented language yielding fast code when using good programming practices. The disadvantage of C# is that its target platforms must be Windows based, a reason why many developers opt for a JAVA based environment. Although one can use traditional Windows system graphics it is preferable to attach the Microsoft XNA library which is designed for fast graphics and animation. Both C# and XNA program design can be done within the Microsoft Visual Studio IDE. All these software components have free versions that are downloadable from the web.

NetLogo (Wilensky, 1999) is a multi-agent programming platform developed by the Center for Connected Learning at North-Western University in Illinois that can be downloaded for free. It is a closed source, multi-platform, JAVA based high level programming language that comes packaged with its own IDE and graphing capabilities. A great advantage of NetLogo is its ease in learning and intuitive use of commands. Novice programmers can get fairly sophisticated code up and running in a day. Some compromises in speed and versatility can be expected.

Between the extremes of building applications from scratch using C# and the pre-designed interface and data structures of NetLogo are a range of multi-agent libraries and IDEs that are typically JAVA based. These include MASON, RePAST, SWARM and many others, some of which are free to download and some proprietary. Some might find these alternative products a better compromise between software development and program execution times, but we will not be examining them.

A simple simulation, using a single predator type (wolf) with a single prey type (caribou), is coded in both C# and NetLogo. The animals move pseudo-randomly with no herd or pack behaviour. Caribou health is maintained by consuming fixed vegetation and they reproduce at a fixed rate. Wolf health is maintained by consuming caribou and they reproduce at the time and location of the caribou predation. Predation occurs when a wolf and caribou happen to be co-located in the same grid box.

For the purposes of evaluating the performance of similar codes in C# and NetLogo the predator-prey simulation is a modification of the Wolf Sheep Predation sample model included in the NetLogo library (Wilensky, 1997; 1999). A uniform domain of 701 x 481 grid points with wrap-around boundaries is used. Initially 500 caribou and 250 wolves are positioned randomly with a random direction vector for movement. Caribou move 0.1 grid length per time step with a random directional change of up to 45 degrees while wolves move 1 grid length per time step with a directional change of up to 20 degrees. Animal age is predetermined randomly with caribou living for a maximum of 1000 time steps (ticks) and wolves 600 time steps. A caribou’s life is ended if it is collocated on the same grid box as a wolf. At this predation event the wolf gives birth to one new wolf. Caribou births occur with a probability of 0.3 % per caribou per time step.

Another simulation was done to experiment with more realistic environmental conditions using only NetLogo. A digital elevation map downloaded from Geobase (http://www.geobase.ca/geobase/en/index.html) but originating from NRCan’s Centre for Topographic Information (Region 116O10 in Northern Yukon) was input using the GIS extension of NetLogo. The domain extended from 138.5 to 139.0 degrees West longitude and 67.50 to 67.75 degrees North latitude using 1202 x 1202 grid points which results in a grid size of around 23 metres. Vegetation growth was simulated by a constant regeneration rate for the whole domain while vegetation loss occurred on patches with caribou. Initially 3000 caribou and 2000 wolves are randomly positioned in the domain, with a maximum life span of 1990 and 600 ticks for caribou and wolves, respectively. Caribou moved 0.24 grid lengths per tick while wolves moved 1 grid length per tick using the same directional randomness as before. Caribou and wolf births and deaths were simulated as before. Caribou were allowed to quickly cross the river while wolves were not. Caribou were prevented from climbing a slope greater than 3 metres in elevation over a 20 metre span, whereas wolf movement was not restricted by topography. Caribou moved faster through terrain with low vegetation density, and both caribou and wolves are prevented from crossing the domain boundaries.
3. RESULTS

Using NetLogo it took just a couple of hours to write the entire program in about 60 lines of code for the software comparison experiment, whereas C# required over 400 lines of code and took a couple of weeks to complete and debug. Although these numbers can vary widely depending on the programmer’s experience, there is a substantial savings in development time using NetLogo. C# requires the extra overhead in designing the various classes from scratch, dealing with inputting parameters and outputting the display and results and writing type-safe code. Program execution speeds are compared in Table 1. The comparison tests were performed on an Acer 4810T notebook with a 1.3 GHz Intel U7300 processor using 4 GB of RAM under Windows 7.

<table>
<thead>
<tr>
<th>Table 1. Comparison of execution speeds in time steps per minute for the cases with and without display updates every time step.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Update display each time step</td>
</tr>
<tr>
<td>-------------------------------</td>
</tr>
<tr>
<td>300</td>
</tr>
<tr>
<td>No display</td>
</tr>
</tbody>
</table>

The performance bottleneck for NetLogo is in the display update. Fortunately it is easy to modify the frequency of display updates or to turn it off entirely. It is surprising that without display updates NetLogo was over 5 times faster than C#. This likely reflects inefficiencies caused by a non-expert C# programmer and the handling of the XNA time clock in C#. It is remarkable how small the graphics overhead is for C# when using the XNA library.

Figure 1 shows a screenshot during a NetLogo simulation of the caribou-wolf model used in the C# comparison study. The domain in green is populated by individual caribou in white pixels and wolves in black pixels. To the left are slider boxes that allow the numerical parameters of the simulation to be adjusted. In the lower left is a graph showing how the total population of caribou and wolves change with each time step. The phased-shifted periodicities typical of simple predator-prey differential equations are evident. An interesting result with the parameters used is the clustering of caribou in groups while the wolves have sufficient speed to move between the caribou groups during their lifetime. This is an example of an emergent phenomenon that was not imposed in the algorithm. This scenario was run for about an hour of actual time without changes in the basic oscillatory and spatial patterns.

Figure 1. Display from the caribou-wolf simulation using Netlogo.
In Figure 2 a screen shot of the domain is shown from the NetLogo simulation with topography and vegetation included. The elevation is depicted by the brightness with the highest elevation of around 575 m appearing nearly white and the lowest land elevation of around 260 m appearing dark. The river depth is shown in shades of blue with the lighter blues corresponding to deeper water. Vegetation density is tracked by an integer and displayed in 3 categories on the map: brown for the lowest third in vegetation density, yellow for the middle third and green for the highest third. Individual caribou and wolves are plotted as white and orange pixels, respectively. The simulation with elevation and vegetation was performed using a PC tower with an Intel i7-940 processor at 2.93 GHz with 8 GB RAM under 64 bit Windows 7 Pro.

Even though the simulation started with the same vegetation density (middle of the yellow category) for all land grid points, the spatial variability of the caribou’s grazing resulted in coherent patterns of enhanced and reduced vegetation. As in the C# comparison simulation the spatial distribution of the caribou and wolves does not appear random with well defined areas completely void of caribou. Regions void of wolves are smaller since wolves move faster than caribou. Also the tendency of caribou to avoid climbing steep terrain results in their paucity in the higher elevations.

**Figure 2.** A screenshot of the domain after 593,307 time steps into the simulation with elevation and dynamic vegetation included.
4. SUMMARY

We compared the performance of NetLogo and C#/XNA in developing and running a simple caribou-wolf interaction in a spatial domain. NetLogo was easy and intuitive to use which greatly reduced development time compared to C#. Surprisingly the execution speed of NetLogo outperformed C# when display updates were turned off. With display updates turned on every time step NetLogo was much slower than C#. Based on these limited tests we conclude that NetLogo is an efficient tool to use in a research mode for predator-prey type modelling. Memory limits should only be dictated by the available RAM in the hardware configuration.

Preliminary simulations were successful at including topography from a real elevation dataset and dynamic vegetation. Future work needs to explore the sensitivity of results to initial conditions, topographic features and the details of the dynamic vegetation processes. NetLogo is also capable of importing vector GIS data (SHAPE files) although a full-fledged GIS application is usually needed to prepare arbitrary GIS files for input into NetLogo. Future model development will include the role of the Inuit with their interactions with the caribou, and the impact of climate change on the behaviour of the caribou and the condition of the environment. NetLogo appears to be well-suited for these ambitious tasks.

REFERENCES


Gui-Quan, S., J. Zhen, Quan-Xing and L. Li, Pattern formation induced by cross-diffusion in a predator-prey system, *Chinese Physics B*, 17(11), 3936-3941, 2008.


Modelling Governance Structures and Climate Change Policy Communications on Community Resilience in the Canadian Arctic

Rachel A. Hirsch, PhD

Faculty of Environmental Studies, York University, 4700 Keele Street, Toronto, Ontario, Canada, M3J 1P3 (rhirsch@yorku.ca)

Abstract: The Government of Canada recognizes that climate change is now unavoidable making it imperative to help Northern Canadians adapt. There is evidence of adaptive capacity as local communities are forced to adjust; however, these adaptive strategies tend to be reactive and localized in nature. There is therefore an increasing need to develop integrated regional climate change policies. I argue that an ecological health approach based on complexity modelling can inform climate change policy in the Canadian arctic. I propose a two-phase institutional mapping procedure in order to identify and understand multiple stakeholder perspectives or ways of knowing about climate change resilience. A key problem is to understand how – in particular, indigenous knowledge is currently integrated into climate change policy-making about the Canadian Arctic. I propose two interrelated projects: 1. systematic ‘mapping’ of institutional structures with climate change policy agendas aimed at enhancing community resilience in the Canadian Arctic; and 2. a communication model illustrating how indigenous knowledge and climate change policy information ‘flows’ between members of these climate change policy networks.

Keywords: indigenous knowledge; institutional mapping; communications modelling; climate change policy; Canadian Arctic

1. INTRODUCTION

Circumpolar regions, including the Canadian Arctic, are facing accelerated warming due to climate change. Northern dwellers and in particular, Inuit, are faced with fundamental shifts in a ‘way of life’ that has been synonymous with adaptability to the cold. The Government of Canada [2010] recognizes further climate changes are now unavoidable making it imperative to help Northern Canadians adapt. There is evidence of adaptive capacity as local communities are forced to adjust, for example, by switching from traditional to store foods [Ford et al., 2010]; however, these adaptive strategies tend to be reactive and localized in nature. There is therefore an increasing need to develop integrated regional climate change programs that work across scales to incorporate both traditional and scientific approaches to climate change adaptation [Adgera, et al., 2005].

Increasingly, researchers are seeking out indigenous sources of knowledge about what
adaptation strategies are best suited to a changing northern environment. For example, Krupnik and Jolly’s [2002] edited volume, “The Earth is Faster Now”, is dedicated to reporting indigenous observations of Arctic environmental change. What is unclear is how findings from these indigenous knowledge studies are being integrated back into climate change policy-making, especially at the regional level. An ecological health approach (Ecohealth), promoting overall human and environmental wellbeing, is adopted as the main conceptual framework for this proposed study.

In collaboration with a growing network of climate change policy stakeholders interested in Inuit knowledge, I propose a two-phase research project aimed at modelling regional to territorial to national institutional communications about indigenous knowledge, climate change, and health. This project will therefore focus geographically on the four regions of the Inuit Nunangat including: the Inuvialuit Settlement Region (Northwest Territories), Nunavut, Nunavik (Northern Quebec), and Nunatsiavut (Northern Labrador) [Inuit Tapiriit Kanatami (ITK), 2010]. Understanding how to include indigenous knowledge in policy-making about the Canadian Arctic will advance help close the Arctic climate change adaptation research-policy-practice gap.

1.1 Conceptual Framework

An Ecohealth approach offers a lens through which complexities inherent in climate change policy-making can be better understood by considering the interplay between social, political, natural, and any other relevant systems (see Figure 1) [Berkes et al., 1998]. Kendrick [2003] explains how Ecohealth, as an action-learning approach, encourages the integration of multiple ‘ways of knowing’ through co-management of environmental issues such as climate change. A main goal of co-management is to identify and navigate between discrepancies over core values and beliefs about sustainability in human-environment systems. In this paper, sustainability refers to the health or resilience of a social-ecological system in terms of its vulnerability (i.e., ability to ‘bounce back’) to surprises such as extreme weather events. Parkes et al. [2008] argue that understanding and encouraging ecosystem health from a multi-stakeholder perspective will produce corresponding health benefits in human communities.

Identification and understanding of various stakeholder perspectives is necessary in order to consider how different ways of knowing about climate change resilience might be acted upon. For example, it is important to consider how different ‘epistemic communities’ [Chilvers, 2008] adopt ‘types of knowledge’ (e.g., scientific, indigenous, or local) [Gilligan et al., 2006]. Views about how natural and human worlds are situated (i.e., worldviews) influence how the realities (and implications) of climate change impacts are interpreted and therefore what policy initiatives are advanced.

It is therefore useful to define several commonly referred to ‘types of knowledge’. Gilligan et al., [2006] differentiates between traditional, local, and scientific ways of knowing:

- **Traditional knowledge**: “a knowledge system based on tradition that is created, preserved and dispersed” [Gilligan et al., 2006, p. 3]. This type of knowledge tends to be passed down through generations and is influenced by many factors including

![Figure 1. One possible depiction of an Ecohealth worldview](image-url)
spirituality, relationships with natural/human communities, oral traditions and cultural beliefs. A key requirement of this epistemology is that connection within and between human-natural systems be acknowledged (e.g., Inuit Qajimajatuqangit); however because this terminology can be confused with scientific knowledge – traditional knowledge will be referred to as indigenous knowledge in this paper.

- **Local knowledge:** “is possessed by a particular group of people [i.e., community members] and generated through first hand experience of one’s surroundings” [Gilligan et al., 2006, p. 4].
- **Scientific knowledge:** “refers to the product of a Western or European approach to empirically studying, researching, and recording observations of phenomena” [Gilligan et al., 2006, p. 4].

So then, we come to the crux: why is indigenous knowledge important for enhancing ecological health? Kendrick [2003] explains the Dene concept of inkonze where humans can only ever know ‘a little bit’ about nature. These people who reside near the Athabasca River in the Canadian sub-arctic feel that our descriptions of the world around us will never be complete and are always prone to reinterpretation. In this sense, co-management becomes an adaptive process where human needs must be continually reassessed as changes occur in the surrounding natural system. Sustainability in the face of changing global environment is therefore one of perpetual learning. Indigenous communities are uniquely situated in that they have direct experience with and historical knowledge of their immediate environment [Gilligan et al., 2006]. As such, it is important to consider how indigenous knowledge is integrated into institutional communications aimed at enhancing community resilience.

### 1.2 Research Question and Objectives

This study, determining how indigenous knowledge is integrated into climate change policy-making about the Canadian Arctic, is informed by the above literature on an Ecohealth approach to adaptive co-management. The stakeholders, in this case, represent regional, territorial, and federal government institutions in Canada with explicit policies aimed at enhancing community resilience to climate change that incorporate indigenous knowledge. The core research question and main objectives of this proposed study follow.

**Research question:** How is community-level indigenous knowledge being integrated into climate change adaptation and resilience policy-making about the Canadian Arctic?

**Objectives:**

1. Create a vertical institutional communications map that: i) identifies regional, territorial, and national governing institutions (nodes) and channels of information (flows) focused on enhancing community adaptation/resilience in the Inuit Nunangat Region of the Canadian Arctic by incorporating indigenous knowledge; and ii) offer a preliminary assessment of the role indigenous, local, and scientific knowledge types play in these communications.
2. Develop a vertical institutional communications model in collaboration with key government, research, and community stakeholders using narrative policy analysis of the ‘policy stories’ told about how indigenous knowledge has been or is currently being integrated into climate change policy-making.

### 2. DISCUSSION

#### 2.1 Proposed Methods

I propose two phases of research for this study. The aim of the first phase is to produce, in
collaboration with climate change policy stakeholders, an institutional communications map identifying key governmental agencies concerned with incorporating indigenous knowledge into Arctic climate change policy in Canada. The aim of the second phase is to develop a communications model explaining how indigenous knowledge is integrated into climate change policy communications by using narrative policy analysis to identify a larger story or meta-narrative that incorporates the varied viewpoints of government, academic, nongovernmental, and community stakeholders.

2.1.1 Institutional Analysis: Mapping Communication

In the area of risk communication, messages about potential threats are sent to various receivers (e.g., emergency managers, risk analysts, the news media) by way of ‘risk signals’. The social amplification of risk framework (SARF) [Kasperson et al., 1996] is one of the models used to understand this process of communicating risk signals. In Figure 2, a simplification of the SARF, any policy is based on information that can be transmitted to one or more receiver(s).

A key facet of this framework is that value is either added or subtracted from these messages as signals are received (i.e., amplified or attenuated). The value type depends on the content of the message about the policy so that not all information communicated will be about, for example, risk. Signals sent to policy-makers might also contain information about the relative priority of particular policy agendas. Both the priority of various policy issues as well as the integration (use) of indigenous knowledge can be mapped according to ‘signal priority’ as part of institutional communications about climate change policy.

I propose the following participatory approach to creating a vertical institutional communications map for each Inuit Nunangat Region in the Canadian Arctic (i.e., four maps). Please note that preliminary networking with academic, government, and nongovernmental stakeholders has provided the starting point for this procedure.

1. 1. Identify key national institutions (e.g., Health Canada, Indian and Northern Affairs Canada, Inuit Tapiriit Kanatami) with directives that aim to enhance climate change resilience in the Canadian Arctic and/or integrate indigenous knowledge into climate change policy-making.

2. 2. With the assistance of key contacts (see Appendix A) from these national institutions, identify regional institutions and stakeholders concerned with climate change adaptation policies that incorporate indigenous knowledge (e.g., Nunavut Research Institute or Sachs Harbour Community Corporation).

3. 3. Pending ethics approval, invite regional stakeholders outlined in Appendix A to collaborate on a pilot mapping exercise. Visits will be made this summer to discuss/begin the territorial to national mapping in Ottawa (June/July) and regional to territorial mapping in one of the four Inuit Nunangat regions (July/August) (see Appendix A).

4. 4. The mapping will be an iterative, participatory process consisting of a short questionnaire asking stakeholders to rate: the priority (no priority = 0; low priority = 1; medium priority = 2; high priority = 3) they ascribe to various issues including food security, climate change adaptation, contaminant exposure, northern development, and any other issue of concern to them; and their use of indigenous, local and scientific
knowledge (no use = 0; low use = 1; medium use = 2; high use = 3) in communications about each of the above issues. As well, participants will be asked to suggest other institutional stakeholders who might have an interest in climate change adaptation policy-making in their region at any level of government (regional to national). Once all relevant stakeholders are identified, the map will be returned to each participant for validation.

8. Apply the final, validated procedure from the pilot-mapping project to three case studies in each of the other Inuit Nunangat Regions.

9. Prepare four validated vertical climate change adaptation communications maps highlighting the priority of various issues in the Canadian Arctic and the usefulness ascribed to each knowledge type for narrative policy analysis based on in-depth interviews with stakeholders from the institutions identified through the mapping procedure.

2.1.2 Narrative Policy Analysis: Modelling Communication

Narrative policy analysis will be used to understand how various knowledge types are, or could better be integrated, maximizing the ability of communities to adapt to climate change impacts. Narrative policy analysis is useful for understanding why various stakeholders might want different types of policies to be enacted and to tell a larger story or meta-narrative which addresses a range of stakeholder concerns [Roe, 1994; Bridgman et al., 2002]. Hirsch [2010] adapted Roe’s [1994] narrative policy analysis to include the following three steps: 1. identify the main story/narrative developed by each opposing coalition (i.e., those with a stake in the issue at hand); 2. consider any alternative stories, alterations to the current stories, or potential counter-narratives (i.e., opposing viewpoints); and 3. consider how any of the alternatives arrived at in step two may be coalesced into a larger meta-narrative that combines the opposing narratives from step one.

I therefore propose the following procedure in order to model institutional communications about how to integrate indigenous knowledge into policy-making about the Canadian Arctic. In-depth interviews will focus on how to best integrate multiple types of knowledge into climate change adaptation policy. Questions will focus on, for example, how to define indigenous, local and scientific knowledge, similarities and differences between knowledge types, and perceived barriers to or opportunities for knowledge translation in the policy-making process. Thus, there is potentially much to learn from a narrative policy analysis approach [Roe, 1994] that considers how stakeholders prioritize and use different types of knowledge [Young et al., 2008].

3. CONCLUSIONS

The main goal of this study is to determine how community-level indigenous knowledge is being integrated into climate change adaptation and resilience policy-making about the Canadian Arctic. Two phases of research are proposed including preliminary mapping to describe key institutional communications between various stakeholders followed by in-depth narrative policy analysis [Roe, 1994] of these communications in order to identify opposing policy arguments, underlying issues, and suggestions about how communications might be improved within a particular policy context [Hirsch, 2010]. Indigenous knowledge adds an important component to climate change policy-making because it recognizes that both human and natural systems are complex and ever changing requiring continual re-evaluation of our needs and priorities [Kendrick, 2003]. This is especially the case in the Canadian Arctic where policy-makers [Health Canada, 2010] and researchers [Berkes et al., 2001] have been interested in working with communities in order to gather evidence about local level adaptive capacity.
This two-phase study has the potential to advance a larger interdisciplinary project with substantial theoretical, methodological and ethical contributions. This research contributes theoretically to an Ecohealth approach by exploring how the integration of various types of knowledge (i.e., indigenous, local, and scientific) facilitates or constrains communications about environment and health policies [Kendrick, 2003]. Methodologically, this study uses a two-step approach to enhance rigour by first scanning for organizational structures and stakeholders by creating an institutional map and then by delving into the processes of information exchange and valuation by modelling communication between these stakeholders. Further, there is an ethical imperative to ensure that communities benefit from research done by communicating traditional knowledge in a manner that remains true to the goals and interests of participants. Synthesis of these 'policy stories' will uncover opportunities and obstacles for communicating about indigenous knowledge and will therefore facilitate policies aimed at enhancing community resilience. Overall, this project will help link research done, policies currently in place, and the actual practices of communities all aimed at promoting ecological resilience in the face of a changing environment.

REFERENCES

Ford, J.D., and T. Pearce, What we know, do not know, and need to know about climate change vulnerability in the western Canadian Arctic: a systematic literature review, *Environmental Research Letters*, 5, 1-9, 2010.


Modelling needs assessment for social adaptation to climate change in Siberia

G.A. Alexandrov1, G. Inoue2, and T. Matsunaga1
1National Institute for Environmental Studies, Onogawa 16-2, Ibaraki 305-8506, Japan
g.alexandrov@nies.go.jp
2Research Institute for Humanity and Nature, Motoyama, Kyoto, Japan

Abstract: Despite the common expectation that climate change may make Siberia a warm heaven, this would not be good for local people. Surprisingly the people of Siberia do rely on the winter cold. Warming may be a serious challenge here. Climate change threatens Siberian forests, lakes and permafrost. In the region where landscapes are unstable this implies the risks of losing the traditional means of subsistence, at least for a while. Social adaptation to climate change is becoming an urgent issue, as the first signs of warming are becoming noticeable to local people. This article reports an attempt to explore the gaps between the state-of-the-art and desired progress in environmental modelling with respect to the integrated assessment of the climate change impacts on Siberia.

Keywords: climate change, environmental modeling, Siberia, reindeer herding, permafrost modelling

1. INTRODUCTION

For the past two decades, Siberian ecosystems were intensively studied to understand how they are functioning in present conditions and how they will be functioning in a warmer climate [Anisimov et al., 2002; Arzhanov et al., 2008; Demchenko et al., 2002; Maximov et al., 2008; Romanovskii and Hubberten, 2001]. These studies resulted in a relatively good understanding of major biophysical processes and their sensitivity to climate change. These studies also led to a conclusion that environmental changes induced by both socio-economic development of the region and climate change are dangerous. The rate of the changes would be too high to cope with their consequences in a business-as-usual way [Iijima et al., 2010].

There is an increasing understanding that a focus of research must be shifted to determine strategies for social adaptation [Ford and Furgal, 2009]. This challenge requires dramatic improvement of environmental models. Models should be able to forecast not only the environmental changes but also the potential impact of those changes on human well-being. Models also should respond to the questions of people potentially affected and provides direct support to political and management decisions.

This article is to emphasize the demand for the models that can adequately connect the scientific vision of future environmental changes to the real consequences for Siberian people.
2. DISCUSSION

2.1. How warm will Siberia be?

The answer to this question depends on the scenario of carbon dioxide emissions and the climate model in use.

The emissions scenarios that are used to run climate models are listed in the corresponding Special IPCC report [Nakicenovic and Swart, 2001]. This report was released in 2001, and scenarios were not updated since that time, although the growth rate of global emissions in 2000-2008 exceeded the conservative expectations. This does not undermine the confidence in the ability of these scenarios to predict the future, because the increase may be a short-term trend.

The large number of scenarios, however, makes it difficult to forecast the future climate. To forecast the future climate, we should first forecast the long-term trend of carbon dioxide emissions. Scenarios were not developed with this purpose in mind. They display the options for world development, and do not take into account the modern political tendencies towards stabilization of carbon dioxide emissions. The models that are used to create scenarios are not exactly the same that are needed to make a forecast of the long-term trend of carbon dioxide emissions and to update the forecast regularly proceeding from the observed rate of emissions and the state of emissions regulation efforts.

The SRES B1 (SRB1) scenario, that was used to run climate models for the Fourth IPCC Report, is perhaps the best candidate to the role of the emissions forecast. This scenario assumes the world to be more integrated and more ecologically friendly. The rapid growth will be accompanied with the rapid changes towards a service and information economy, towards the introduction of clean technologies, and towards global solutions for achieving socio-environmental stability.

Under this scenario, almost all parts of Siberia will be warmer by more than 2°C in sense of mean annual temperature, whereas northern parts will be warmer by more than 3.5°C (Fig 1). Projections for a particular region are model dependent, but in most cases vary in the same range.

The analysis of the 16 climate model outputs available at IPCC Data Distribution Center show significant discrepancies [IPCC 4th Assessment Report (2007), 2007]. Therefore, for constructing a model independent projection, one has to combine model outputs. This can be done either by simple averaging or by weighted averaging. In the latter case one need to develop criteria for assigning weights, which is not an easy task. Another option is to use a recently proposed method [Alexandrov and Matsunaga, 2008; Alexandrov and Matsunaga, 2009], which is based on the premises of evolutional epistemology [Bradie, 1994]. Since the number of climate models is quite large, development of methods for constructing model-independent projections become more important issue than improvement of any single model.
2.2. Will permafrost disappear?

Permafrost, in a broad sense, is ground remaining below freezing point of water for at least two years. The mean annual ground temperature (MGT) is usually higher than the mean annual air temperature (MAT), and therefore permafrost regions are the regions where mean annual air temperature is below -5°C. In these regions the temperature of ground at some depth remains below 0°C during the whole year, because the amplitude of seasonal variations decreases with the distance from ground surface.

The long-term records of ground temperature in permafrost regions show that the trends in MGT generally follow the trends in MAT. Both of them are increasing. However, the rates are not the same. MGT grows 1.5 times slower [Pavlov, 2008]. It would increase by 2°C when MAT would increase by 3°C. The area of the permafrost regions (that is, the regions where MAT<-5°C) would decrease by the end of the century, but they would still cover most of Siberia.

Since water keeps ground warm, the area of permafrost in permafrost regions may be sensitive to changes in hydrological regime, especially where the ice content in the ground is high. A depression in permafrost table tends to grow, because it accumulates water that makes ice melting. Hence, permafrost table is locally unstable. This property manifests itself in patterned landscape which is typical for permafrost regions.

Even in the regions of so-called continuous permafrost, permafrost covers about 80% of land. The other land is covered by taliks - year-round unfrozen ground. Taliks occur beneath lakes and rivers where deep water does not freeze in winter.

The percentage of taliks (and water bodies) at a given watershed reflects the water balance of this watershed, and may respond non-linearly to the warming. Perhaps permafrost may really disappear at some watersheds, even if MAT will remain below - 5°C. Thus, to assess changes of permafrost area within permafrost region one needs a spatially-distributed model of heat and water balance within a small watershed.

2.3. What happens to reindeer herders?

The modern Siberian “reindeer civilization” was formed somewhere in the 18th century, when the rapid expansion of semi-domestic reindeer population occurred [Krupnik, 2000]. This was supposedly the result of favorable environmental change that improved the quality of summer pastures and thus opened the “window of opportunity” for the socio-economic transformation.

Pasture quality is conventionally evaluated in terms of the amount and nutritional value of forage that the pasture produces. However, reindeer herders tend to evaluate summer pastures proceeding from other criteria [Istomin, 2004]. Herding reindeer flocks at the pastures covered with tall vegetation (or having complicated relief) is not as easy as at the pastures covered with short vegetation (or having flat relief). First, a flock more often breaks into small groups. Second, herders need more time to react on the changes in the flock behavior. The size of flock that can be managed by 1-2 herders depends on the pasture relief, the height of vegetation and some other factors having
no direct relation to the pasture productivity. The best pastures allow 1-2 herders to keep under control the flock of 3000-3500 reindeers.

The environmental conditions determine two parameters of the reindeer civilization – the size of total semi-domestic reindeer population and the size of the flock that can be handled by a family. The latter determines how wealthy the reindeer civilization may be. In the case of indigenous way of life, the minimum size of a flock is 25-30 reindeers per capita [Galanin and Galanina, 2004; Yoshida, 1997]. In the modern social-economic conditions, a herder needs a flock of 250-300 reindeers to ensure his family’s well-being. On the south margins of the reindeer civilization (i.e., in the forest zone), the typical size of the flocks is much smaller, 30-50 reindeers, and therefore the other traditional sources of subsistence (like fishing and hunting) become more important [Yoshida, 1997].

Although climate change would make summer pastures more productive, this would not be good for herders, because warming may seriously reduce pasture quality. The summer pastures are vulnerable to the breaks in the vegetation cover and the upper layer of soil. The vegetation cover and upper layer of soil protect ground from thawing and permafrost table from melting. Due to instability of permafrost table even relatively small disturbances may turn a flat pasture into a patterned piece of land. Herders keep flocks moving to reduce the risk of losing summer pastures. The larger the flock the faster it should move. Warming may seriously decrease efficiency of herder’s labor – that is, make herders working more hard or being poorer.

### 2.4. Modelling needs

The carrying capacity of a summer pasture may be defined as the number of reindeer that would stay there within one day without causing long-term damage to pasture productivity. The carrying capacity depends not only on the productivity of the pasture but also on the intensity of trampling damage to vegetation and soil cover. Disturbing local heat and water regime, the trampling damage to vegetation and soil cover may initiate thermokarst development (or differential thawing) and turn a flat pasture into a patterned piece of land not suitable for herding large flocks.

The ability of the pasture to withstand trampling damage depends on whether heat and water balance at the watershed scale may “diffuse” the local disturbances of the heat and water regime. Since local disturbances are of small size, one needs a 3-dimensional permafrost model of fine spatial resolution to address such a question.

The climate change studies have been concentrated on the thickness of active layer (i.e., the depth of permafrost table) [Arzhanov et al., 2008; Lunardini, 1996; Malevsky-Malevich et al., 2001; Romanovsky et al., 1997; Smith and Riseborough, 1996; Stendel et al., 2007]. Since this research question may be adequately addressed by one-dimensional models simulating the vertical profile of ground temperature, the development of models needed for simulating differential thawing and formation of patterned-ground landscapes was not a priority. Nevertheless, there were efforts in this direction [Nishimura et al., 2009; Pohl et al., 2009; Riseborough et al., 2008; Smith and Riseborough, 2010; Thomas et al., 2009; Woo et al., 2004; Woo et al., 2008], and therefore development of a 3-dimensional permafrost model mentioned above is a feasible need.

### 3. CONCLUSION

It is not an easy task to develop models responding to the questions of people potentially affected by climate change. Most of models developed in connection to climate change studies were to provide discovery of the changes that might take place. Today, in a world that is aware of its responsibility for planetary changes and planning globally concerted actions, environmental modelling is expected to provide forecasts of changes that would take place and affect human well-being. Therefore, modelers have to mobilize themselves and develop new kind of models.

This article emphasizes the demand for coming from plural scenarios of regional climate change to a single “official” forecast that can be used as the basis for developing adaptation strategies. Will Siberia become permafrost free by the end of this century? There are worries that this might happen, but there is no forecast that this would happen.
The next problem is how to evaluate the consequences of permafrost retreat for Siberian people. This article concentrates on reindeer herders whose well-being depends on the availability of land suitable for pastures, identifies constraints for future development of the Siberian reindeer civilization, which may be imposed by permafrost melting, and formulates the direction in which permafrost models should be improved.

The recent social studies [Forbes and Stammmler, 2009; Takakura, 2010a; Takakura, 2010b] demonstrate that one should employ a holistic approach to understand the adaptive capacity of the Siberian reindeer civilization: “...the daily practice of tundra nomadism involves permanent processes of negotiating one’s position in a changing environment, which is why “adaptation” is woven into the society, and cosmology as a whole, rather than being separable into distinct “bodies” of knowledge or Western-designed categories”, [Forbes and Stammmler, 2009]. The resilience of the reindeer civilization is based on the complex use of land area and freshwater resources. For example, fish always remains an important component of dietary culture of reindeer herders despite the size of the flock [Yoshida, 1997]. In other words, to simulate the future development of Siberian reindeer civilization we need not only models of environmental processes but also an agent-based model reflecting the way in which reindeer herders perceive their role on the land.

ACKNOWLEDGEMENTS

The paper was prepared in connection with the on-going research project “Global Warming and the Human-Nature dimension in Siberia” of the Research Institute for Humanity and Nature (RIHN) and improved by taking into account the comments made by Tetsuya Hiyama, Hiroki Takakura and Atsushi Yoshida.

REFERENCES

Demchenko, P.F., A.A.Velichko, A.V.Eliseev, I.I.Mokhov, and V.P.Nechaev, Dependence of permafrost conditions on global warming: Comparison of models, scenarios, and paleoclimatic reconstructions, Izvestiya Atmospheric and Oceanic Physics, 38, 143-151, 2002.
Ford, J.D. and C.Furgal, Foreword to the special issue: climate change impacts, adaptation and vulnerability in the Arctic, Polar Research, 28, 1-9, 2009.


Pavlov, A.V., Trends of contemporary changes of soil temperature in Northern Russia (in Russian), Earth Cryosphere, 12, 22-27, 2008.


Smith, S.L. and D.W.Riseborough, Modelling the thermal response of permafrost terrain to right-of-way disturbance and climate warming, Cold Regions Science and Technology, 60, 92-103, 2010.


Modelling and control of a hybrid renewable energy system
to supply demand of a green-building

H. Dagdougui\textsuperscript{a,b}, R. Minciardi\textsuperscript{c}, A. Ouammia\textsuperscript{a}, M. Robba\textsuperscript{a}, R. Sacile\textsuperscript{a}

\textsuperscript{a}Department of Communication, Computer and System Sciences (DIST), University of Genova, Italy.
\textsuperscript{b}MINES ParisTech, CRC, Sophia Antipolis Cedex, France.
\textsuperscript{c}Energetic Laboratory, Sciences Faculty, Abdelmalek Essaadi University, Tétouan, Morocco

Abstract: Renewable energy sources are an “indigenous” environmental option, economically competitive with conventional power generation where good wind and solar resources are available. Hybrid plants can help in improving the economic and environmental sustainability of renewable energy systems to fulfil the energy demand. The aim of this paper is to present the architecture of a Decision Support System (DSS) that can be used for the optimal energy management at a local scale through the integration of different renewable energy sources. The integrated model representing a hybrid energy generation system connected to the grid is developed. It consists of PV and solar thermal modules, wind turbine and biomass plant. Moreover, a framework is presented for the optimization of the different ways to ensure the electrical and thermal energy demand of the microgrid as well as the water demand, with specific reference to two main cases for the real time energy optimal control: the presence/absence of a storage system. Finally, the optimization model has been applied to a case study.

Keywords: DSS, optimization, wind energy, PV, smart grid.

1. Introduction

The sustainable security of energy supply, led both developed and developing countries to make and implement new policies to improve efficiency in energy consumption, to adopt new alternatives like renewable energy systems. To face the economic, social, technological and environmental challenges, the need for energy conservation as well as for developing renewable technologies becomes ever more critical. As reported by Juan et al. [2010], in the EU and US, buildings energy consumption has even exceeded the energy consumption of the industrial and transportation sectors. Renewable energy utilization is one of the most important aspects of green buildings. The wind and solar energy are freely available and environmental friendly. One common disadvantage of these resources, is their unpredictable behaviour, in addition, the variation of these sources may not match with the time distribution demand (Yang et al [2008]). Furthermore, the wind energy systems may not be technically viable at all sites because of low wind speeds and being more unpredictable than solar energy. As to solve this drawbacks, the complementary combination of each component characteristic may lead to enhancement of system efficiency and reliability. In addition, combined utilization of these renewable energy sources are therefore becoming increasingly attractive and are being widely used as alternative of oil-produced energy (Nema et al. [2009]). Hybrid renewable energy systems are becoming popular for remote area power generation applications due to advances in renewable energy technologies and subsequent rise in prices of petroleum products, and due to the possibility of attenuating fluctuations in produced power. Economic aspects of these technologies are sufficiently promising to include them in developing power generation capacity for developing countries (M.K. Deshmukh and S.S Deshmukh [2008]). Hybrid systems can be considered as a reasonable solution, capable to support systems that cover the energy demands of both stand-alone and grid connected consumers. Commonly, it consists of a mix of two or more energy sources used jointly to provide increased system efficiency as well as greater balance in energy supply. In literature, several papers have studied the design and planning of hybrid renewable energy systems (for example: Paska et al. [2009]; Ashok [2007]; Ekren and Yetkin Ekren [2008]).

The aim of this paper is to present the architecture of a Decision Support System (DSS) that can be used for the hourly energy management of a mix of renewable energy systems. Specifically, an integrated model representing a hybrid energy generation system (characterized by solar plate collector, PV, biomass, wind and battery storage) connected to the grid is developed. The approach is based on mathematical modelling of each component, then an optimization problem is solved in order to better manage and control the energy flow so to ensure reliable supply of demand.

2. The system model
The hybrid energy generation system for a general building connected to the electrical network (that is considered in this work) is reported in Figure 1. This hybrid system consists of one PV module, one wind turbine, one biomass plant, one solar flat plate collector, and one battery to store electricity. The optimization of such system aims to generate energy satisfying the demands in real time (i.e., heating demand, electrical demand and water demand) thus taking into account the available renewable energy resources in each time interval. The flat plate collector and biomass plant are uniquely used to ensure heating demand. Either energy produced by the wind turbine or the energy produced by the PV module can be directly used to satisfy a part of the electrical demand as well as the water demand through pumping, and/or can contribute to supply the heating demand. The electricity surplus from wind turbine and the PV module can be sent to the storage battery or/and sold to the network. The battery storage system can receive electrical energy from PV module and wind turbine and can provide free energy for heating, electricity, and pumping needs in cases of deficit in electricity. Furthermore, the network connection offers the possibility to purchase electricity in case of failure of the storage system.

In particular, the following entities are defined:

- $E^f_X$: The output energy produced from the wind turbine/PV/Biomass/FPC [kWh] in time interval $(t, t+1)$, $t=0,..,T-1$;
- $E^f_{Y,h}$: The energy provided from renewable energy system [kWh] for heating in time interval $(t, t+1)$;
- $E^f_{Y,e}$: The energy provided from the renewable energy system and the grid [kWh] for electric supply [kWh] in time interval $(t, t+1)$;
- $E^f_{Y,p}$: The energy from the renewable energy system and the grid [kWh] for supply electricity for pumping water in time interval $(t, t+1)$;
- $E^f_{Z,Net}$: The surplus produced from wind/PV [kWh] sent to the network in time interval $(t, t+1)$;
- $\hat{E}^f_{Net}$: The overall energy that is sent to the network from wind and PV module [kWh] in time interval $(t, t+1)$;
- $\Delta E_{Net,eff}$: The overall energy effectively sent to the network [kWh] in time interval $(t, t+1)$;
- $E^f_{Z,b}$: The energy from wind turbine/PV that is sent to the battery [kWh] in time interval $(t, t+1)$;
- $Q^f_w$: The amount of water pumped [m³/h] in time interval $(t, t+1)$;
- $CH^f_e$: energy provided from the battery to the electricity [kWh] in time interval $(t, t+1)$, $t=0,..,T-1$;
- $CH^f_h$: energy provided from the battery to the heating system [kWh] in time interval $(t, t+1)$;
- $CH^f_p$: energy provided from the battery to the pumping [kWh] in time interval $(t, t+1)$;
- $CB^f$: the level of battery charge [kWh] at time instant $t$, $t=0,..,T-1$.

![Figure 1. The considered hybrid system](image)

2.1 Energy from the wind turbine
The power output from the wind turbine unit is expressed as a function of the wind speed. In fact, a linear wind model \cite{16} assumes a linear (affine) dependence (within the interval \([v_c, v_r]\)) of the wind turbine power output, \(P_{wt}'\), on the current wind speed at the hub height \(v', t=0,...,T-1\), being \(T\) the time horizon in hour. The following equation is used to simulate the electrical power output of the wind turbine (Nutton et al. [2001]):

\[
P_{wt}' = \begin{cases} 
0 & v' < v_c \\
\frac{v' - v_c}{v_r - v_c} & v_c \leq v' \leq v_r \\
v_r & v_r \leq v' \leq v_f \\
0 & v' > v_f 
\end{cases} \quad t=0,...,T-1
\] (1)

where \(p_r\) is the rated electrical power, \(v_c\) is the cut in wind speed, \(v_r\) is the rated wind speed, and \(v_f\) is the cut off wind speed.

In general, the wind speed measurements are given at a height different than the hub height of the wind turbine, the following equation is used to evaluate the wind speed at the desired height (Rodolfo et al. [2008]):

\[
v' = \frac{\ln(H_{hub}/z_0)}{\ln(H_{data}/z_0)} v_{data} \quad t=0,...,T-1
\] (2)

Where \(v'\) is the wind speed at the height of the hub, \(H_{data}\) is the height of the measurement, \(H_{hub}\) is the hub height, \(z_0\) is the surface roughness length, and \(v_{data}\) is the wind speed at the height of the measurements.

\subsection{2.2 Energy from the PV module}

The output power generated from the PV module, with respect to the solar radiation, can be calculated using the following formula (Hocaoglu et al. [2009]):

\[
P_{pv}' = S_{pv} \eta_{PV} p_f \eta_{pc} G' \quad t=0,...,T-1
\] (3)

where \(S_{pv}\) is the solar cell array area, \(\eta_{PV}\) is the module reference efficiency, \(p_f\) is the packing factor, \(\eta_{pc}\) is the power conditioning efficiency, and \(G'\) is the hourly irradiance.

\subsection{2.3 Energy from the biomass heating plant}

Energy provided by the biomass heating plant depends on the used biomass quantity \(u_t\) [m$^3$ hour$^{-1}$], the biomass volumetric mass, \(VM\) [kg m$^3$] (i.e., the ratio between the dry mass [kg] and the volume [m$^3$]), the lower heating value, \(LHV\) [MJ kg$^{-1}$]. The LHV assumes that the latent heat of vaporization of water in the fuel and the reaction products is not recovered. Then, it can be calculated once higher heating value (HHV) and moisture content (MC) are known. The HHV is the total energy release in the combustion with all of the products at 273 K in their natural state when water has released its latent heat of condensation. In the considered work, the HHV is evaluated from the basic data analysis of biomass. The biomass MC [%] represents the water amount present in the biomass and it can be expressed as a percentage of the dry weight. As regards production plant, the plant is supposed to operate at the maximum productivity level. The following equation states that the plant developed energy \(E_B'\) [kWh]:

\[
E_B' = 0.2778 \eta_{bh} LHV u_t VM \quad t=0,...,T-1
\] (4)

Where \(\eta_{bh}\) is the plant efficiency and 0.2778 is a conversion factor.

\subsection{2.4 Energy from the flat plate collector}

The useful thermal energy extracted from the water collector depends on the instantaneous incident solar irradiation, the plate area, and its efficiency (El Fadar et al [2009]). That is,

\[
E_{FPC}' = \eta_{fpc} A_{fpc} G' \Delta t \quad t=0,...,T-1
\] (5)

Where \(\eta_{fpc}\) is the efficiency of the solar flat plate collector, \(A_{fpc}\) [m$^2$] is the area and \(G'\) [kW/m$^2$] is the solar irradiation.

\subsection{2.5 Produced thermal and electrical energy}
The hourly energy that can be used in time interval for heating, $E_{h}^{t}$, is expressed by:

$$E_{h}^{t} = E_{WT,h}^{t} + E_{PV,h}^{t} + E_{FPC,h}^{t} + E_{b,h}^{t} + E_{Net,h}^{t} + CH_{h}^{t} \quad t=0,...,T-1$$  \hspace{1cm} (6)

The hourly electrical energy $E_{e}^{t}$ that can be used in time interval is expressed by:

$$E_{e}^{t} = E_{WT,e}^{t} + E_{PV,e}^{t} + E_{Net,e}^{t} + CH_{e}^{t} \quad t=0,...,T-1;$$  \hspace{1cm} (7)

The hourly energy $E_{p}^{t}$ that can be used in time interval for pumping water is given by:

$$E_{p}^{t} = E_{WT,p}^{t} + E_{PV,p}^{t} + E_{Net,p}^{t} + CH_{p}^{t} \quad t=0,...,T-1;$$  \hspace{1cm} (8)

The amount of pumped water is proportional to the energy used for this purpose, that is:

$$Q_{p} = \frac{E_{p}^{t}}{\rho g h \eta_{PS}} \quad t=0,...,T-1$$

Where $\rho$ is the water density [kg m$^{-3}$], $g$ is the gravity constant acceleration [m s$^{-2}$]; $\eta_{PS}$ is the pumping system efficiency and $h$ is the height of pumping.

### 2.6 Energy flow exchange with the network:

The energy that is sent to the network is composed by the surplus produced by the wind turbine and the surplus produced by the PV module, i.e.,

$$\bar{E}_{Net}^{t} = E_{WT,Net}^{t} + E_{PV,Net}^{t} \quad t=0,...,T-1$$  \hspace{1cm} (10)

The energy that is taken from the network is given by

$$E_{Net}^{t} = E_{Net,h}^{t} + E_{Net,e}^{t} + E_{Net,p}^{t} \quad t=0,...,T-1$$  \hspace{1cm} (11)

The energy productions by the wind turbine and the PV module in each time interval can be used for different supply purposes: electrical demand, heating demand, water demand, sent to the battery or sold to the network. Thus, the following equations hold

$$E_{WT}^{t} = E_{WT,h}^{t} + E_{WT,e}^{t} + E_{WT,p}^{t} + E_{WT,Net}^{t} + E_{WT,b}^{t} \quad t=0,...,T-1$$  \hspace{1cm} (12)

$$E_{PV}^{t} = E_{PV,h}^{t} + E_{PV,e}^{t} + E_{PV,p}^{t} + E_{PV,Net}^{t} + E_{PV,b}^{t} \quad t=0,...,T-1$$  \hspace{1cm} (13)

Instead, as regards biomass and flat collector plant, the produced energy can be less or equal to the available potential. In fact, biomass may not be used (also because if it is summer heating does not work and heating cannot be sent to the network), and water for heating passing through the plate collector may be stopped. This implies the following relations:

$$E_{b,h}^{t} \leq E_{b}^{t} \quad t=0,...,T-1$$  \hspace{1cm} (14)

$$E_{FPC}^{t} \leq E_{FPC}^{t} \quad t=0,...,T-1$$  \hspace{1cm} (15)

Finally, $E_{PV,Net}$ and $E_{WT,Net}$ are known because they are the surplus of electrical energy. That is,

$$E_{PV,Net}^{t} = \max \left[ 0, E_{PV}^{t} - E_{PV,p}^{t} - E_{PV,e}^{t} - E_{PV,b}^{t} - E_{PV,Net}^{t} \right] \quad t=0,...,T-1$$  \hspace{1cm} (16)

$$E_{WT,Net}^{t} = \max \left[ 0, E_{WT}^{t} - E_{WT,p}^{t} - E_{WT,e}^{t} - E_{WT,Net}^{t} - E_{WT,b}^{t} \right] \quad t=0,...,T-1$$  \hspace{1cm} (17)

### 2.7 The battery storage state equation

The battery works as an inventory for the electrical energy that can, in this way, be stored. Specifically, a state equation for the battery storage can be formalized. That is,

$$CB^{t+1} = CB^{t} + E_{WT,b}^{t} + E_{PV,b}^{t} - CH_{b}^{t} - CH_{h}^{t} - CH_{e}^{t} \quad t=0,...,T-1$$  \hspace{1cm} (18)

### 3. The optimization problem

The decision variables of the optimization problem are $E_{WT,h}^{t}, E_{PV,h}^{t}, E_{FPC,h}^{t}, E_{b,h}^{t}, E_{Net,h}^{t}, E_{WT,e}^{t}, E_{PV,e}^{t}, E_{Net,e}^{t}, E_{WT,p}^{t}, E_{PV,p}^{t}, CH_{h}^{t}, CH_{b}^{t}, CH_{e}^{t}$, while $CB^{t}$ are the state variables of the overall system.

The objective function (to be minimized) is characterized by the weighted sum of the deviation from the various demands, as well two terms to be maximized related to the energy sent to the storage system and the electrical network. That is,
\[ Z = \sum_{t=1}^{T-1} \left( E_{WT,t} + E_{PV,t} + E_{FPC,t} + E_{Nat,t} + CH_{Net} - E_{Dhw} \right)^2 + \]
\[ \alpha \sum_{t=1}^{T-1} \left( E_{WT,t} + E_{PV,t} + E_{Nat,t} + CH_{Net} - E_{Dhw} \right)^2 + \]
\[ + \beta \sum_{t=1}^{T-1} \left( E_{WT,t} + E_{PV,t} + E_{Nat,t} + CH_{Net} - E_{Dhw} \right)^2 - \chi \sum_{t=1}^{T-1} E_{Net} - \epsilon \sum_{t=1}^{T-1} CB \]

Where \( \alpha, \beta, \chi, \) and \( \epsilon \) are weighting factors, and \( E_{Dhw}, E_{Dve}, Q_{Dhw} \) are respectively the effective heating, electricity and water demands.

The constraints of the optimization problems are represented by equations reported in section 2. Moreover, there is a constraint related to the battery capacity. That is,
\[ CB \leq C_{max} \]

where \( C_{max} \) is the size of the battery.

The problem is here solved using mathematical programming techniques through a commercial optimization package. Dynamic programming could also be used to reduce the overall problem complexity, decomposing it in sub-problems. In the case of absence of battery energy storage, since there are no state equations and available energy varies in each time interval, the optimization problem can be run at each time interval in a separated way. In the following, the optimization problem is solved for a specific case study, in which there is not the presence of the battery storage (thus, the variables related to the battery storage are known and equal to zero).

4. Application to a case study

The proposed DSS has been tested using the real data obtained from Capo Vado site, which is the windiest site in the region of Liguria, in Italy [Ouammi et al (2010)]. The data of the First November 2008 have been used. They consist of the hourly wind speed, recorded at the height of 10m, and hourly solar irradiation (see figure 2). The optimization problem has been solved using the optimization tool Lingo (www.lindosystems.org).

Figure 2: Hourly wind speed and solar radiation in the site of Capo Vado

The wind model described by equations above has been applied to the specific case study, using the wind turbine G-3120 35 kW with the following parameters www.endurancewindpower.com: \( v_c = 3.5 \) [m/s], \( v_r = 11 \) [m/s], \( v_f = 25 \) [m/s], \( P_r = 35 \) [kW], \( H_{hub} = 30.5 \) [m], \( H_{data} = 10 \) [m], \( z_0 = 0.03 \) [m]. For the PV module, the features that has been used consist of: \( S_{pc} = 100 \) m\(^2\), \( \eta_{pv} = 0.11 \), \( \eta_{pc} = 0.86 \) and \( P_{f} = 0.9 \).
Figure 3: Hourly energy produced during the first November 2008.

Figure 3 shows the hourly energy produced by the wind turbine, as well as the hourly energy produced by the PV module. In fact, the figure shows that the energy range produced by the wind turbine is included between 14.3 kWh at 3:00 and a maximum value of 36.5 kWh at 16:00. However, for the solar module, the energy production reaches its maximum 4.5 kWh at 12:00. It seems that the energy production coming from the wind turbine is higher than the energy produced from the PV module. This fact is mainly due to the high wind speed in November and the low solar radiation.

Figure 4: Optimal electrical energy control

The optimal energy management that is addressed to satisfy the electrical energy needs of the household is displayed in Figure 4. The electrical energy demand reaches its higher hourly peak value of 2.4 kWh at 18:00 and a minimum of 1.3 kWh at 04:00. The hourly periods have been defined throughout the day: period A (1<t<8), period B (8≤t<17) and period C (17≤t≤24). During period A and C, the electrical energy needs were totally provided by the wind turbine owing to the absence of the solar radiation. During the period B, both the wind turbine and solar PV module provide sufficient energy to meet the hourly electrical energy demand of the household with a dominant participation of the solar PV module.

Figure 5: Optimal heating energy control

Figure 5 shows the contribution of biomass unit and the flat plate solar collector through the daily hours, we mention that these systems are responsible in addition to energy given by PV and wind turbine to provide a part of
heating energy needs. In one hand, it can be shown that the energy delivered by the biomass system goes between a maximum that equals to 2 kWh and a minimum equal to 0.5 kWh. On the other hand, as regard the flat plate solar collector (FPSC), the energy produced from the sun goes directly to contribute in the satisfaction of the heating energy requirements, reaches a maximum of 1 kWh at 12:00. We must notate that in this typical day, no wind and PV contribution is remarked for the supply of heat needs, the Biomass as well as the FPSC are the only one that furnish the heating needs of the household,

As regard the water demand, its availability is depending on the working pumps which must use the electrical energy coming from the wind and PV systems in order to pump water. As displayed by Figure 6, three peaks are perfectly met during the hours of the day 6:00AM, 12:00PM and 19:00PM, thus mainly using electrical energy from the wind turbine. By comparing energy produced by each system, it appears that the wind turbine provides the higher amount of energy. This behavior is mainly due to the high wind potential available in Capo Vado site compared with solar energy.

In order to ensure the sustainability of the hybrid system, and addressing the mismatch between the intermittencies of wind and solar irradiation on one part and of demand in another part, the system presented in this paper integrates both: an internal storage system and a connection with the electrical grid, the main goal of implementing these two systems are:

- Once the electricity generated by the wind and the solar modules exceed the total demand of the household, electricity will be “stored” at the battery and/or “sold” to the network.
- Once the battery is fully charged, the electricity excess will be sold directly to the electrical Network.
- Once the electricity generated by the wind, solar modules, biomass and FPSC is not sufficient, the deficit of electricity and/or the heating and/or pumping water system will be compensated by the electricity cumulated in the battery or purchased from the grid.

The choice between these two electrical systems (battery and Network) will be guaranteed by the objective function and its constraints, thus by making the system more reliable as possible.

The cumulative storage battery level can be well seen from Figure 7, in fact, the battery has a large transitory regime until it reaches its higher storage value. This storage is effectuated during the day and at each hour, thus assuming that the battery was empty at t=0 which correspond to the initial time. In figure 8, the hourly electricity sold to the network is displayed. It can be shown that this transfer occurs during all the hours of the day, where the wind turbine is the one that send the excess of energy.

Figure 6: Optimal control of the water demand

Figure 7: wind and solar energy sent to the battery
5. Conclusion

A DSS for real time energy management is here proposed to define the optimal energy flows in a building characterized by a mix of renewable resources (solar plate collector, PV, biomass, wind and battery storage) to satisfy different energy demands. The model is applied to the case of Capo Vado (Liguria Region) and optimal results to satisfy all the energy demands are found for a testing day in the month of November.

REFERENCES


A Generic Framework for Multi-Disciplinary Environmental Modelling *

Rolf Hennicker a, Sebastian S. Bauer a, Stephan Janisch a, and Matthias Ludwig a

aInstitut für Informatik, Ludwig-Maximilians-Universität München, Germany
{hennicker, bauerse, janisch, mludwig}@pst.ifi.lmu.de

Abstract: We present a generic framework for computer-based environmental modelling which supports the coupling of simulation models from various sciences to perform integrative simulations. The framework is, in principle, applicable to any kind of model which simulates spatially distributed environmental processes with an arbitrary, but discrete, time scale. During an integrative simulation the framework coordinates the coupled models which run in parallel exchanging iteratively data via their interfaces. For proving the correctness of the coordination, formal methods of software development have been applied. The framework provides a developer interface for the implementation of natural science and socio-economic simulation models. For the latter, a framework specialisation has been developed which supports the modelling of agent-based social simulation models. The framework design was driven by the idea of enforcing common rules for integrative simulations, which must be respected by all simulation models, while leaving as much freedom as possible for discipline-specific implementations. Within the GLOWA-Danube project, the framework has been successfully applied to construct the distributed simulation system DANUBIA which integrates up to 15 simulation models from various disciplines, like meteorology, hydrology, plant physiology, glaciology, economy, agriculture, tourism, and environmental psychology. Actually, DANUBIA is already in use as a tool for decision makers to support the sustainable planning of the future of water resources in the Upper Danube basin.

Keywords: modelling framework, coupled simulations, agent-based social simulations, software development.

1 INTRODUCTION

Climate change has an increasing impact on our natural and social environment which reveals more and more the need of interdisciplinary research to better understand the complex, mutually dependent processes occurring in nature and in socio-economic systems. In this article, we report on a generic framework for computer-based environmental modelling which supports the runtime coupling of various simulation models from natural science and socio-economic disciplines. The framework is generic in the sense, that it is, in principle, applicable to any kind of model which simulates spatially distributed environmental processes on an arbitrary, but discrete time scale. During an integrative simulation for some simulation period, the framework coordinates the coupled models which run in parallel exchanging iteratively data via their interfaces. For the development of the modelling framework best practices of software engineering have been applied like abstraction, separation of concerns and formal methods based on precise mathematical

*This research has been partially supported by the GLOWA-Danube project 01LW0602A2 sponsored by the German Federal Ministry of Education and Research.
notations. For instance, the framework provides abstractions of simulation models and thus facilitates the development and integration of concrete simulation models of particular disciplines. Separation of concerns involves the aspects of information exchange between simulation models, consistent modelling of simulation space, coordination of concurrently running simulation models, and agent-based social simulations.

Technically, for the development of the simulation framework the Unified Modeling Language UML (cf., e.g., Rumbaugh et al. [2005]) has been used in the requirements and in the design phase, while the framework implementation is programmed in Java making use of Java’s Remote Method Invocation interface (RMI) for communication between distributed components. Formal methods have been applied to specify the requirements for the general life cycle each simulation model must obey and for specifying and proving the correctness of the coordination of distributed simulation models, for which the framework is responsible. For this purpose, the process algebra Finite State Processes FSP (cf. Magee and Kramer [2006]) has been used. For the specification of requirements concerning correct simulation configurations, invariants have been stated and formalised in terms of the Object Constraint Language OCL; cf. Warmer and Kleppe [2003]. The framework design follows a component-oriented approach which allows to plug in arbitrary simulation models as long as the common requirements concerning, e.g., the simulation space and the life cycle of a model are satisfied. Thus, the integration of different simulation models into a coupled simulation is considerably facilitated and (most) errors occurring in simulation models caused by not meeting common requirements can be detected already at compilation time.

The framework has been developed and successfully applied to construct the integrative simulation system DANUBIA within the interdisciplinary research project GLOWA-Danube1 (Ludwig et al. [2003]), which is part of the German national initiative GLOWA (Global Change in the Hydrological Cycle) running from 2001 to 2010. Within GLOWA-Danube, a group of researchers from various natural science and socio-economic disciplines have teamed up to investigate the impact of climate change on the water cycle within the Upper Danube watershed and to support the development and evaluation of regional adaptation strategies. Actually 15 simulation models have been developed by the research groups of GLOWA-Danube, such that various simulation configurations can be built and run by our modelling framework within the DANUBIA system.

A number of other frameworks supporting integrated environmental modelling emerged since the GLOWA-Danube project started in 2001. In the field of integrated water resource management there are, e.g., the Object Modelling System OMS (cf. Kralisch et al. [2005]), ModCom (cf. Hillyer et al. [2003]), The Invisible Modelling Environment TIME (cf. Rahman et al. [2003]), and the Open Modelling Interface OpenMI (cf. Gregersen et al. [2007]). While TIME is a platform for the development of standalone modelling tools, OMS, ModCom, and OpenMI are frameworks which support the independent development of models and allow for execution of coupled simulations. In particular, OpenMI is designed to extend existing standalone models by standard interfaces for coupling. However, to our knowledge, none of these frameworks is tailored towards distributed and parallel execution of coupled simulation models, which is a main characteristic of our approach.

The outline of this paper is as follows. In Sect. 2, we identify common requirements for integrative environmental simulations. Then, in Sect. 3, we summarise the design of our framework in accordance with the given requirements. Sect. 4 introduces briefly the integrative simulation system DANUBIA obtained by framework instantiation within the context of the GLOWA-Danube project. Finally, in Sect. 6, we summarise our main results.
2 REQUIREMENTS FOR INTEGRATIVE ENVIRONMENTAL SIMULATIONS

We consider in the following as a simulation model a computer program that simulates an environmental process over a certain time span, called the simulation time, with regard to a certain geographical area of the environment, called the simulation space. In an integrative simulation system several simulation models are coupled in order to analyse dependencies and feedbacks of the simulated processes. It is obvious that, besides the technique of coupling itself, the consistent treatment of simulation time and simulation space is crucial for the integration of different simulation models. Moreover, for a comprehensive environmental simulation not only processes occurring in nature but also processes reflecting human behaviour must be taken into account. Therefore we have identified four major requirements. The framework should support

1. data exchange between simulation models at runtime,
2. coordination of simulation models according to simulation time,
3. consistent treatment of simulation space, and
4. agent-based social simulations.

In the following we elaborate more on each of the four requirements.

2.1 Data Exchange between Simulation Models

The coupling of simulation models is based on interfaces. Interfaces for data exchange specify data queries. We distinguish between provided interfaces specifying queries for data that is provided by a simulation model, and required interfaces specifying queries for data that is needed by a simulation model for its own computation. The general requirements concerning data exchange are modelled in the UML class diagram in Fig. 1. It says that a simulation may involve arbitrarily many models, which play the role of the participatingModels for the simulation, and that a model may have arbitrarily many interfaces, playing the role of provided or required interfaces. For the reader, who is not familiar with the UML notation, a brief introduction of the modelling elements used hereafter is given in Sect. 7.

![Figure 1: Requirements model for data exchange](image)

A concrete example of a provided and required interface is given later on when we illustrate the framework instantiation in Fig. 9. The following invariant expresses a consistency requirement for data exchange which must be satisfied for any integrative simulation.

**Invariant for data exchange**

- In an integrative simulation, for each required interface of each participating model there exists exactly one participating model which provides that interface.
This invariant can be formalised in terms of the following OCL-expression:

```oclintext
context Simulation inv:
    self . participatingModels . forAll (m |
        m . required -- forAll (r |
            self . participatingModels -- one (n |
                n . provided -- includes (r)))))
```

All invariants described in the sequel of this paper can be formalised in a similar way in terms of OCL invariants (cf. Warmer and Kleppe [2003]) which will, however, be omitted here.

### 2.2 Coordination of Simulation Time

An important characteristic of our problem domain is the concurrent execution of different simulation models which iteratively exchange information at runtime via their interfaces. In order to guarantee the consistency of data exchange during a simulation run, the single simulation models must be appropriately coordinated with respect to the progressing simulation time. The correct coordination is a non-trivial task since, in general, simulation models have different, individual time steps determining the model time between two consecutive computations. Model time steps depend, of course, on the simulated processes which typically range from minutes or hours, like in natural sciences, to months, like in social sciences. Hence a precise, unambiguous specification of the coordination problem is mandatory. We first describe the general life cycle which a simulation model must follow:

- **provide** initial data at the model’s provided interfaces
- while not at simulation end
  - **get data** from the model’s required interfaces
  - **compute** new data for the next time step
  - **provide** newly computed data at the model’s provided interfaces

For the formalisation of a model’s life cycle we use the process algebra Finite State Processes FSP; cf. Magee and Kramer [2006]. The following FSP process `MODEL` specifies the general behaviour of a simulation model. In order to be generally applicable the process is parameterised with respect to the model’s time step. The sequence of actions in line 5, `getData[t] -> compute[t+Step] -> provide[t+Step]`, is iteratively performed with increasing time `t` and thus formalises the iteration in the informal description of a model’s life cycle given above. Note that the computation of new data for time `t+Step` relies on data obtained for time `t`. This time difference avoids deadlocks of concurrently running models (in the case of feedback loops) but it may also lead to imprecisions whose relevance must be analysed in concrete cases and, if necessary, can be resolved by using smaller time steps.

```fstop
range SimTime = SimStart .. SimEnd

MODEL(Step) = (start -> provide[SimStart] -> M[SimStart]),
M[t:SimTime] =
if (t+Step <= SimEnd)
then (getData[t] -> compute[t+Step] -> provide[t+Step] -> M[t+Step])
else (finish -> STOP).
```

When several simulation models are executed in parallel, it is essential that only valid data is exchanged, i.e. data that fits to the local model time of the participating models. To specify this
requirement we consider only two simulation models at a time, one, say \( U \), acting as a user of data, and the other one, say \( P \), acting as a data provider. From the user’s point of view we obtain the coordination condition \((U)\), from the provider’s point of view the coordination condition \((P)\).

\[(U)\] \( U \) gets data expected to be valid at time \( t_U \) only if the following holds:
The next data that \( P \) provides is valid at time \( t_P \) with \( t_U < t_P \).

\[(P)\] \( P \) provides data valid at time \( t_P \) only if the following holds:
The next data that \( U \) gets is expected to be valid at time \( t_U \) with \( t_U \geq t_P \).

Condition \((U)\) ensures that the user does not get obsolete data while condition \((P)\) guarantees that data, available at the provider’s interface, will not be overwritten if it is not yet considered by the user model. If one can show that all (pairwise) combinations of all models participating in an integrative simulation considered in both roles, as user and as provider of data, satisfy the two coordination requirements, then the whole integrative simulation is coordinated correctly. We have again used FSP to formalise the coordination conditions in terms of so-called property processes. Then, in a next step, we have constructed an explicit coordination process with FSP and we have verified by model checking techniques that the coordination process provides a solution for the coordination problem of integrative simulations (which later on can be implemented in Java). For more details see Hennicker and Ludwig [2005].

### 2.3 Modelling of Simulation Space

In an integrative environmental simulation the consistent treatment of the underlying simulation space is crucial. It is obvious that in spatially distributed simulations one needs geographical units, which in the following will be called proxels. The term proxel (cf. Tenhunen and Kabat [1999]) stems from process pixel and suggests that a proxel does not only model a structural element of the simulation space, but it shows also dynamic behaviour by simulating the environmental processes on this particular geographical unit. The entire simulation area is then modelled by a set of (non-overlapping) proxels.

![Diagram of Simulation Space](image)

**Figure 2:** Requirements model for simulation space

The spatial requirements of an integrative simulation are described by the UML class diagram in Fig. 2. It says that a simulation concerns always exactly one simulation area which, in turn, consists of a set of proxels. The class **Proxel** requires that each proxel has a unique identifier **pid** and an operation **computeProxel()** to compute the next state of a proxel in each time step. Moreover, each proxel can have a number of properties which must be common to all simulation models (like, e.g., geographical coordinates, elevation, etc.). On the other hand, each simulation model has a set of proxels, on which it operates. The following invariant requires that the models...
participating in an integrative simulation agree on the set of proxels determined by the area of the simulation.

Invariant for the simulation space

- In an integrative simulation, all participating models operate on proxels which belong to the simulation area of the simulation.

Besides the basic properties, a proxel may store domain-specific properties as illustrated for groundwater proxels in Fig. 9 later on.

2.4 Support for Agent-Based Social Simulation

Another fundamental aspect of integrated environmental simulation is the explicit support of societal issues concerning human behaviour. A common modelling approach within socio-economic disciplines is agent-based social simulation which uses concepts from the field of multiagent systems; cf. Weiss [1999]. While models for social simulation integrate into coupled environmental simulations like any other simulation model by interface-based data exchange, their local computations are conceptually different from natural science models. At the core of such simulation models are actors which model deciding entities such as individuals (e.g. farmers, tourists, households, etc.) or organisations (e.g. water suppliers, companies). The class diagram in Fig. 3 shows the fundamental concepts and relationships for agent-based social simulations.

An actor model is a specialisation of a simulation model that refers to an arbitrary number of actors. An actor administrates a set of plans, each consisting of a number of actions. Moreover, an actor can be linked to a number of sensors to observe the environment and it may use its history to retrieve information on the success of previously executed plans. Concerning the dynamic behaviour, an actor model is supposed to follow the same life cycle as known from ordinary simulation models; cf. Sect. 2.2. However, the computation step of an actor model is specialised in the sense, that each actor referred by the model must perform (possibly concurrently) the following steps:

- query sensors for the state of the environment
- decide which plan should be executed by performing the following steps:
  - select relevant plan options
Hennicker et al. / A Generic Framework for Multi-Disciplinary Environmental Modelling

- filter the options to decide which plan should be implemented
  
  - implement the selected plan by executing its associated actions
  
  - inform the actor model on the effects of the plan implementation

Finally, the actor model must update its provided interfaces accordingly.

Our conception to support agent-based social simulation follows the principles of reactive agent architectures (cf., e.g., Wooldridge [1999]) characterised by simple behaviours, which directly map perceived states to actions to be executed, together with the parallel execution of these behaviours. How the agent-based simulation principles of our modelling framework work in a concrete example of (inter)acting water suppliers and households is described in Barthel et al. [2010].

3 SIMULATION FRAMEWORK

The requirements and concepts described in the previous section are realised in a component-based simulation framework. The framework defines a general data interface and generic components that implement common structure and behaviour, thus imposing general rules which must be respected by concrete simulation models when it comes to framework instantiation. D’Souza and Wills [1999] use the notion of plug-point, provided by a generic component, and plug-in, provided by some extension which completes the generic component to an executable implementation. In this section we focus on the framework and plug-points while, in Sect. 4, we show how the framework can be instantiated by simulation models with plug-ins. The framework itself is split into two layers, the framework core and the developer interface. To explain the principle ideas behind these layers we consider the framework excerpt shown in Fig. 4 (without taking into account components yet).

![Diagram](image)

Figure 4: Two-layered framework architecture

The framework core implements all features that can be handled by the framework itself like, e.g., the time coordination, the common properties of a simulation model (ModelCore), and the management of the spatial distribution (ProxelTable). Concerning time coordination, there exists at runtime exactly one instance of the class TimeCoordination which is a monitor object that is called by each model core instance before data is fetched from or provided to a data exchange...
interface. Each model core instance itself will be linked at runtime to exactly one concrete simulation model which must implement the abstract operations, given by the plug-points `getData`, `compute` and `provide`, of the class `AbstractModel` of the developer interface. The UML sequence diagram in Fig. 5 illustrates the sequence of interactions implemented in the run method by taking into account the life cycle of simulation models as described in Sect. 2.2.

Figure 5: Interactions of the run method

Let us now consider how the two layers explained above fit into the component architecture of the simulation framework shown in Fig. 6.

The component `SimulationAdmin` is the central component for managing integrative simulations and hence provides an interface for a (graphical) user interface. This interface offers operations for starting, observing and aborting a simulation. Before starting a simulation the component `SimulationAdmin` checks the simulation configuration for consistency, i.e. if the invariant concerning interfaces as stated in Sec. 2.1 is fulfilled. At the beginning of a simulation the component `ModelLinker` establishes the links between simulation models over their interfaces for data exchange, including distribution over a network. The component `TimeCoordination` is responsible for the correct time coordination of the single models during a simulation run. The component `BaseData` reads and stores initialisation data for the basic properties of the simulation area for all simulation models. `Model` is a generic component which contains framework core classes as well as the classes constituting the programming interface for model developers as discussed above.

Concerning agent-based social simulations a particular so called DEEPACTOR framework has been developed providing a common architecture for socio-economic simulation models. The design of the DEEPACTOR framework follows again our general pattern which distinguishes core classes and (abstract) base classes for model developers, exemplified by the classes `ActorModelCore` and `AbstractActorModel` in Fig. 7. The remaining components of the DEEPACTOR framework are implemented similarly along the requirements described in Sect. 2.4 and we omit further details here.
Figure 6: Component-based design of the simulation framework

Figure 7: Integration of the DEEPACTOR framework

The integration of the DEEPACTOR framework into the general simulation framework is achieved by extension of the developer interface of the overall framework. As illustrated by the excerpt in Fig. 7, the core part of the DEEPACTOR framework specialises (solely) the developer interface of the general framework. Concrete actor models then use the developer interface of the DEEPACTOR framework and, for basic functionality, the developer interface of the general framework such as AbstractProxel. The manual Janisch [2007] provides detailed descriptions of the modelling features available in the DEEPACTOR framework and Barthel et al. [2010] shows its application for the development of DEEPACTOR models in the context of GLOWA-Danube.

We have used UML 2.0 for the complete framework design, in particular for the detailed documentation of the developer interfaces. The framework is implemented in Java SE 6 and contains approximately 25,000 lines of code.

4 APPLICATION OF THE FRAMEWORK: THE DANUBIA SYSTEM

Within the GLOWA-Danube project (Ludwig et al. [2003]) our simulation framework has been applied to construct the integrative simulation system DANUBIA which integrates up to 15 sim-
ulation models for natural processes (like hydrology, plant physiology, groundwater, glaciology etc.) as well as socio-economic models. The latter have been developed to model the behaviour of the involved actors in the areas of agriculture, economy, water supply, private households, and tourism based on the structure of societies and their interests. The purpose of DANUBIA is to serve as a tool for decision makers from policy, economy, and administration for the sustainable planning of water resources in the Upper Danube basin under global change conditions. DANUBIA is a distributed simulation system – the simulation models are executed in parallel on a computer cluster periodically exchanging data during a simulation run. DANUBIA was validated with comprehensive data sets of the years 1970 to 2005. It is actually in use to run and evaluate coupled simulations which are driven by climatic as well as societal scenarios for the next 50 years. Concerning performance, integrative simulation runs with DANUBIA for typical configurations over a 50 year period actually take (at least) 36 hours computing time. An overview of the system architecture of DANUBIA is given in Fig. 8.

![System architecture of DANUBIA](image)

Figure 8: System architecture of DANUBIA

How a concrete simulation model is integrated in the framework is shown in Fig. 9. The upper layer of Fig. 9 depicts the part of the developer interface known from Fig. 4, and the lower layer shows parts of a sample Groundwater model. One can see that all model classes (and interfaces) of the groundwater model extend the base classes (the base interface `DataInterface` resp.) of the developer interface by certain domain-specific properties, like the proxel attributes `gwWithdrawal`, `gwLevel` etc., and by providing implementations for the plug-in operations like, e.g., `compute` and `computeProxel`. Thereby the framework’s core functionality concerning runtime coordination, management tasks and the like is completely hidden for model developers.

Fig. 9 shows that the Groundwater model uses the interface `WatersupplyToGroundwater` for importing (getting) required data and it realises the interface `GroundwaterToWatersupply` by implementing the interface operations, e.g., `getGroundwaterLevel`. In each time step, when the (plug-in) operation `getData` is executed, the model imports a spatially resolved table of required data and distributes it on the corresponding attributes, e.g., `gwWithdrawal`, of its local proxels. Then a computation step is performed proxelwise and finally, when the (plug-in) operation `provide` is executed, the newly computed data on the corresponding attributes of the Groundwater proxels, e.g., `gwLevel`, are collected from the single proxels (which can be accessed using the
query proxel inherited from the model’s superclass AbstractModel) and stored, again spatially resolved, in a corresponding export table. For example, the attribute values of gwLevel of each proxel are stored in the export table gwLevelTable of type LengthTable. When the operation getGroundwaterLevel is called by another model, the corresponding table is returned.

To ensure that models agree on the units of exchanged data each predefined data table type specifies a common unit which applies to all data in the table. Of course, a model can internally operate with different units, but then the values must be appropriately converted before data exchange. A further provision for data consistency is achieved by the possibility to enhance data interfaces by specifications which define the range of expected and guaranteed values when data is imported and provided via an interface. The model developer is then responsible to comply with the specifications during runtime.

While the framework is primarily intended for the development of new simulation models, legacy models can yet be integrated into the framework as long as their computation steps are controllable from the outside. In this case the legacy model is surrounded by a wrapper which must implement the (plug-in) operations like any other model. The concrete computation steps of the legacy model are then initiated within the compute operation of the wrapper, e.g. by using the Java Native Interface.

A concrete application of the DEEPACTOR framework for the simulation of water-related issues in the Upper Danube basin is reported in Ernst et al. [2007], describing a simulation model for the water consumption of households, and in Barthel et al. [2008b] who developed a model for water supply companies as an important link between socio-economic and natural science simulation models. An agent-based simulation model for the diffusion of environmental innovations is described in Schwarz and Ernst [2009]. From the natural science point of view, Barthel et al. [2008a] focus on the groundwater model within DANUBIA, which is based on MODFLOW and hence an example for the integration of legacy models, and, in Lenz-Wiedemann et al. [2010] the crop growth model within the landsurface component of DANUBIA is described.
5 Performing an Integrative Simulation

Fig. 10 summarises the essential steps which are necessary to perform an integrative simulation with the DANUBIA system. We assume that the available simulation models – in terms of source code and executables, as well as metadata, initialisation data and other data which is required to run the model – are already checked-in at the project repository.

At first the available models and the framework must be downloaded and installed on either a computer cluster, a local area network, or on a single computer, depending on the simulation configuration to be executed. In the next step a simulation configuration is created which defines the simulation time, the simulation space, the data resources for a consistent initialisation, some project specific parameter settings and, most importantly, the set of models to participate in the simulation. To assist the user in defining a simulation configuration the framework provides a graphical user interface which allows for setting the necessary global parameters and for selecting the participating models from a list of available models. In particular, each participating model must be allocated to a computer resource on which it will be executed.

Before starting a simulation, its configuration is checked for reasonableness of parameter settings and compatibility of interfaces for data exchange. When the simulation is started the simulation environment is initialised on each computer and the models are distributed and concurrently started on their allocated computer resources. Data interfaces are linked automatically by matching the names of corresponding import and export interfaces. During the simulation run, the progress of the simulation and of each participating model can be observed via the graphical user interface. After the simulation has finished, result data is available in a binary format which is optimised with respect to the required hard disc space. From this point, the data can be visualised and post-processed with appropriate tools which are also available in the project repository.

6 Conclusions

In this paper we have described requirements, design and implementation of a generic framework for environmental modelling. The framework supports the development and the coupling of simulation models from various disciplines to perform integrative simulations.

Applying the framework paradigm to integrative environmental modelling provides several advantages: common routines and services like network support, time coordination or space initialisation can be separated from the scientific code of the simulation models. The model developer only has to implement distinguished extension points of the framework, thus one can be sure that generally valid rules are respected by each model. The simulation framework is generic in the sense that it is independent from actual simulation models. In fact, it scales up to an arbitrary number of simulation models and can be applied to any simulation area as long as the requirements of the framework are satisfied.

The framework has been successfully instantiated in the interdisciplinary research project GLOWA-Danube by the implementation of the distributed simulation and decision support system DANUBIA. With DANUBIA a number of scenarios concerning changes of climate and society
have been simulated which shall show their impacts on the future of water resources thus giving hints for sustainable planning.

REFERENCES


Hennicker et al. / A Generic Framework for Multi-Disciplinary Environmental Modelling

7 APPENDIX: UML NOTATIONS

We give a short summary of the graphical notations of the unified modelling language UML used in this paper. For more details see Rumbaugh et al. [2005].

A class diagram is a graphical description of system entities, in UML called classifier, and the relationships among them. Classifier and relationships are further distinguished and provide a number of concrete modelling features. Fig. 11 shows an overview of those modelling elements which are used in this paper.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Graphical Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>class</td>
<td>Simulation</td>
</tr>
<tr>
<td>abstract class</td>
<td>AbstractModel</td>
</tr>
<tr>
<td>interface</td>
<td>DataInterface</td>
</tr>
<tr>
<td>(req, prov)</td>
<td></td>
</tr>
<tr>
<td>component</td>
<td>TimeCoordination</td>
</tr>
</tbody>
</table>

Figure 11: Relevant UML2 classifier and relationships

A class models a concept of some application domain. Objects are instances of classes. Abstract classes represent (abstract) concepts which can not be instantiated. They are used to describe base classes which are generalisations of more concrete classes. In our framework description we use abstract classes to model generic concepts that are reified within concrete simulation models. Interfaces gather operation signatures required (req) or provided (prov) by a class or component. Providing an interface means to provide an implementation of the particular operations, requiring an interface means to use the operations as described in the interface and implemented by a third party unknown (decoupled) from the user. This kind of decoupling is crucial for component-based system design. A component is a classifier which can encapsulate (sub-)classifier and which, usually, has required and/or provided interfaces. Therefore, a component defines a part of a system which may be implemented and used only in accordance with its interfaces. Implementation proceeds without knowledge of concrete usage, and usage proceeds without knowledge of the concrete implementation.

The most fundamental kind of relationship in UML is association used to specify connections between classifier instances. To emphasise the role one classifier plays for an associated classifier various features may be used. Among them is the distinction between unidirectional (unid) and bidirectional associations (bid), role names and multiplicities. Directionality expresses the possibility to navigate along the association in order to retrieve information about the connected entity. Role names provide an identifier to refer to that entity, and multiplicities are used to describe constraints on the number of associated classifier instances. A generalization relates a general concept, represented by a “parent” classifier, to a more specific “child” element which adds (reifies) information given by the parent classifier. It is graphically represented by an arrow from the child to the parent classifier. The relationship realization expresses that a classifier, usually a class or a component, implements a (provided) interface. A dependency relationship expresses that one model element makes use of some feature of another model element and hence is dependent on that.
A modelling platform for complex socio-ecosystems: an application to freshwater management in coastal zones

J. Ballé-Bégantona, M. Lamplab, C. Bacherc, A. Fiandrinoa, V. Guillard, T. Laugierd, R. Mongruela, J. A. Pérez Agúndez

a Ifremer, UMR Amure, Département d’Economie Maritime, BP70, 29280 Plouzané Cedex, France (Johanna.Beganton@ifremer.fr)
b Université de Brest, UEB, UMR Amure, 12 rue de Kergoat, CS 93837, 29238 Brest Cedex 3, France
c Ifremer, Centre de Brest, B.P. 70, Z.I. Pointe du Diable, 29280 Plouzané, France
d Ifremer, Laboratory of Environmental Resources, Bd Jean Monnet, BP 171, 34203 Sète Cedex, France
e 1point2, 12 rue des Pies, 38360 Sassenage, France

Abstract: Providing deliberation support tools for freshwater management in coastal zones requires complex socio-ecosystem modelling in a joint effort from research scientists, software engineers and stakeholders. The SPICOSA System Approach Framework provides guidelines for the building of dynamic models for a better comprehension of the variety of processes and issues regarding coastal management. A modular platform was developed using the ExtendSim® software. We explore how the SPICOSA/Extend platform supports participatory modelling by offering the necessary tools for the integration of multidisciplinary knowledge and how the use of a common platform facilitates dialog between research scientists of varied backgrounds. Graphical tools developed in ExtendSim allow for a synthetic and user-friendly representation of the system processes and, in particular, of performances of management options in prospective scenario simulations.

We discuss how an economical and social approach of the issues influences the choice of processes and variables to be modelled as well as the level of complexity needed to represent the ecological and physical system. The illustration of this approach will be supported by two case studies. The first concerns the freshwater use competition in the Charente River on the French Atlantic coast. The second refers to water quality management applied to micro-biological contamination from watershed runoff in the Thau lagoon on the French Mediterranean coast.

Keywords: Integrated assessment tools, socio-ecosystem modelling, ICZM, ExtendSim.

1. INTRODUCTION

SPICOSA (Science and Policy Integration for Coastal System Assessment) is an integrated European research project in support of ICZM (Integrated Coastal Zone Management). The project started in February 2007 and ends in January 2011 and encompasses 18 study site applications. The SPICOSA project develops a System Approach Framework (SAF), which aims at incorporating the ecological, social and economic dimensions in order to support decision making in complex coastal systems.

Models for integrated assessment deal with extremely varied processes, approaches, scales and scientific domains. In this context, building predictive models is a highly time-consuming undertaking and often lacks the necessary data, knowledge and flexibility. Following Brugnach et al. [2008] model typology, system approach models must be built: (1) for exploratory purposes: to allow managers to test a high variety of policy options, (2)
for learning purposes: to integrate and share knowledge to deal with the complexity of issues facing the decision-makers in coastal areas [Westmacott, 2001].

We here present system approach guidelines for participatory modelling supported by our experiment in two case studies: (1) the Pertuis Charentais on the French Atlantic coast deals with fresh water use competition, (2) the Thau Lagoon, on the Mediterranean coast, refers to water quality management applied to micro-biological contamination from watershed runoff. The on going development of those two assessment models using ExtendSim software led us to formalise a building methodology in three main stages: cognitive system representation, model components development and outputs construction.

2. DEVELOPMENT TOOLS AND STUDY SITES

2.1 The ExtendSim modelling software

The SPICOSA project chose the ExtendSim software (www.extendsim.com) for its system modelling capacities inherited from STELLA type platforms (www.iseesystems.com) with an additional and essential feature of hierarchical decomposition of processes. It is designed for discrete or continuous modelling. It uses independent blocks set in drag-and-drop libraries and offers numerous modelling, graphical, input/output basic pre- programmed blocks. Data are passed from block to block using user designed connections. A very attractive feature of the software is the possibility of building “custom” blocks in ModL language (similar to C language) with a graphical interface for icon view, a dialog box for parameters and a “help” box for comments and documentation. Custom blocks are stored in user defined libraries allowing for modular development. A data base has been embedded in the software in version 7.0: a very useful commodity for parameter readability, storage of data and of outputs. The data base and the blocks can communicate with the EXCEL software. However, queries are heavy-handed and we had to establish user-friendly functions and nomenclature.

2.2 Quality protocol

Integrated socio-ecosystem modelling must be creative, disciplined and systematic and improves with participation of multiple actors: stakeholders, local managers, scientist and modellers [Jakeman et al., 2008]. All along the development process of the model, it is recommended to apply a quality control methodology. In the SPICOSA framework, we use a Design-Formulation-Appraisal (DFA) protocol. This iterative process should be applied at all levels and be approved (appraisal step) by consensus of the experiment participant group thus promoting a strong participatory involvement.

2.3 Two SPICOSA study sites: the Pertuis Charentais and the Thau Lagoon

Fresh water competition in the Pertuis Charentais:
Along the French Atlantic coast, the Pertuis Charentais coastal waters are strongly impacted by the Charente River. Two important industries in the area are dependent on freshwater availability: agriculture and shell fish farming. However, the river suffers from low flow in summer and shows a risk of failing the EU Water Framework Directive in sustaining drinking water for households and tourism as well as the local industries and a good ecological status of the coastal ecosystems. The policy issue for the study site was decided with a stakeholder group of six local managers and concerns the quantitative management of the freshwater in the Charente river watershed and its impact on the coastal waters. The objectives must be addressed following the regional management plans: hierarchy of freshwater uses and a target of Reachable Discharge Thresholds at chosen monitoring stations on the river in summer over the next 10 years.

Microbiological contamination of the Thau Lagoon:
The Thau Lagoon, situated in the west of the Mediterranean French coast, sees its water quality strongly impacted by a rapid demographic growth and by an important seasonal
influx due to tourism and failures in the water treatment system. The current political debate regards the ways the reduction of microbiological contamination of the lagoon could be translated into an operational policy objective, which will combine the sanitary classification of the Lagoon (A or B) and the occurrence of commercial bans for the shellfish farming industry. The SPICOSA Thau lagoon team has entered into a partnership with the local public organisation in charge of the Thau lagoon water management (“Syndicat Mixte du Bassin de Thau”, SMBT) and is expected to supply scientific advice for the implementation of the local water management plan (SAGE).

3. SYSTEM APPROACH MODELLING GUIDELINES

The construction process of the integrated model system is separated in three main stages:

- Cognitive system representation: a participatory investigation of the best representation of the system.
- Model components development: all the model components are mathematically formulated and implemented in ExtendSim with an objective of a minima level of complexity.
- Outputs construction: development of adapted visualisation and documentation tools for exploratory, learning and communication purposes.

3.1 Cognitive system representation

Participatory development of assessment tools for ICZM can be time consuming as it involves numerous people: stakeholders of different management agencies, scientists and engineers of different domains and institutes. It is however essential that stakeholders be strongly involved in the first design step [Voinov et al., 2008] to insure that the question asked and so the focus of the model corresponds to the need of end users [Westmacott, 2001] and will be useful as a deliberation support tool.

Figure 1. The Pertuis Charentais model

In a first step, using a DPSIR causal methodology, the different “actors” –modules– are identified within the three ecosystem services approach categories: “governance”, “uses” and “resources”. The ExtendSim software interface allows building of empty container hierarchical blocks, each with its own user defined icon representation. As shown on the Pertuis Charentais model interface (Figure 1), the blocks are organised according to the
following classification: (1) on top, governance and regulation e.g. irrigation restrictions based on river water flow monitoring, (2) in the middle, environmental resources uses e.g. agriculture, drinking water for households, recreational fishing and shell fish farming, (3) on bottom, the resources systems e.g. the Charente river hydrology, wetlands and coastal water productivity.

Once the “actors” are identified, the second step consists of defining the vertical and transversal interactions with their corresponding indicators. Stakeholders are usually comfortable with defining their resources needs; however their possible downstream impact on other actors of the system—often unknown by the stakeholders themselves—must now be formulated. In the Pertuis Charentais system, the shell fish farming industry claims that summer low water flow from the Charente river in the coastal waters due to excessive irrigation diminishes their production and their spat collection activity. The model must therefore represent the claimed link between agriculture and shell fish farming through the hydrological system: the agriculture module collects water for irrigation from the river hydrology which is connected through a nutrient supply flux to a coastal productivity compartment. The shell fish farming module consumes the produced phytoplankton and feedback on coastal waters through this food assimilation (see Figure 1). For the governance modules, input link for monitoring of “resources actors” and output link for application of governance rules on “uses actors” are established.

In the third step of the cognitive representation, the modelling team has to look for internal feedback loops. Indeed, for the socio-ecosystem model to be dynamic, it must present such a feature. If no internal feedback loop can be established, the scope of the system or the “actors” definition must be broadened. This can be tested using Forrester [1971] representation of system dynamics (Figure 2).

![Figure 2. Representation of internal feedback loops (Pertuis Charentais system)](image_url)

In the final step, the system representation has to be appraised by the participant group. At this stage, the common feedback we received was that the model clearly visualised the connections between the different components and showed each participant how their management sector was situated and interacted in the global coastal system. Furthermore, when showed to a panel of scientists specialised in the Marennes-Oleron bay ecosystem for review and discussion of existing knowledge and data, the model opened an integrative scientific discussion about the impact of the Charente river inputs for oyster growth and juvenile collection, and again highlighted the links between different highly specialised areas of research.

3.2 Model components development
At this stage, we have a representation of the system with its network of connections between the different components. In the next stage of the model building, we found a helpful sequential procedure:

1. first, express connection indicators in quantifiable variables.
2. then program each component *a minima* in order to model only the state variables required by the directly connected modules.

The formulation of each “actor” sub-model is determined by the indicators defining the connections to the other “actors”. For example, in the Pertuis Charentais, the governance module needs water flow daily values at regulatory monitoring stations on the Charente river watershed (see Figure 1). The linked hydrology module must therefore be able to simulate these flows. In the same way, the socio-economic impact of the fresh water competition on shell fish farming is estimated through gains and losses of oyster sales due to lack of nutrient input from the river in coastal waters and subsequent loss of plankton productivity for oyster feeding. The oyster growth and production module is consequently designed to link phytoplankton input to graded oyster weight output (see Figure 3). Each sub-model is thus formulated *a minima* following the connection model network.

In terms of scales, time and space range from regional and yearly for economic information to less than hourly micro-level ecological processes [Westmacott, 2001]. For exploratory purpose, time and spatial scales must be detailed enough to be relevant to managers: monitoring networks must be simulated (flow values for the Charente River; microbiological levels in oyster for the Thau Lagoon) and management areas must be represented (hydrological sub-catchments and counties, urban and industrial and areas). At the same time, the model must not become so complex that it looses its communication and learning qualities. For the two models, we chose a common daily time step which is appropriate for most ICZM processes. When necessary, it is possible to use a smaller time step e.g. an hourly step for the Primary Productivity module.

Environmental computer models, apart from multi-agent platforms, mostly deal with the physical or biological part of the system. A common bias observed in a number of SPICOSA study sites is for modellers to start building the integrated model from existing and often complex environmental models and then try to establish links to socio-economic modules. This often leads to a poorly integrated ESE model (Ecology-Social-Economy). Our experience on the two French study sites showed that it is crucial for integrated ESE modelling to approach the system with a top-down formulation from the socio-economics uses and governance components leading to the choice of the appropriate physical and biological resources models. Quoting Westmacott [2001] on ICZM modelling: “*There is little point in having a complex three-dimensional hydrological model when the economic model […] is a simple cause-effect relationship.*” The finality of the ESE model -in order to answer the questions asked by the stakeholders- constrains the necessary level of complexity for the physical and biological sub-models.

For the resources modules, complex models existed for some components in each study site previous to the SPICOSA experiment, e.g. SWAT model for agriculture [Arnold and Fohrer, 2005] or MARS-3D three-dimensional hydrodynamic model for the lagoon [Fiandrino et al., 2003]. Rizzoli et al. [2008] justly advocate reuse of existing tools. However, except for PC-Raster interactive raster GIS environment [Deursen and Wesseling, 1993], SPICOSA prescribed reformulating models in ExtendSim rather than developing coupling interfaces. Our experience showed that reformulating in a common software language and structure, which was at first felt as a constraining loss of time, highly improved communication and common understanding of the system in the end, particularly in a multidisciplinary experiment with scientists and modellers from different background and methods.

In an objective of flexibility and transparency, formulating *a minima* the sub-models requires thoughtful downscaling to target the necessary processes. In the Pertuis Charentais model (Figure 4), we translated in ExtendSim the parts the hydrological model used by the Water Agency managers needed to monitor the flow levels of the Charente river. However, in some cases, the system complexity bans downscaling e.g. the hydrodynamic transport in the Thau lagoon waters asks for a 3D model. However, the contribution of watershed outlet
fluxes to micro-biological concentration at monitoring station in the lagoon are additive and allowed us to use transport transfer functions calculated off-line with MARS-3D.

For learning and communication purposes, ExtendSim software allows for a didactic modular and hierarchical organisation of the model. For the shell fish farm module a conceptual schematic presentation was used (Figure 3), with a block for each process: grow-out start, DEB (Dynamic Energy Budget) growth model, mortality, harvest, economical assessment, etc…. For the hydrology module, each sub-catchment is simulated by an identical reservoir model with specific data for each and time-lag connections between sub-catchments (Figure 4). Using a back-ground image over which to organise the different sub-catchments blocks conveys the spatial dimension of the system.

![Figure 3. The Shell fish farm module](image1)

![Figure 4. The Hydrology module](image2)

All the sub-models being at this stage formulated, appraisal is the following step in the DFA quality protocol. Considering the variety of processes to be modelled and the paucity of data, ICZM models are difficult to appraise and new assessment criteria must be established and used [Jakeman et al., 2006]. The models presented being highly modular, each component can be run separately with data in place of upstream connection and be validated against available data. This modular validation paired with sensitivity tests of each module leads to assessing if the chosen level of complexity is appropriate. For example, in the Pertuis Charentais model, the shell fish farming module (Figure 1) is forced by chlorophyll a and sea temperature calculated by the primary production module. A highly simplified Nitrate-Phytoplankton model was chosen for the production module which has been calibrated and validated on the Marennes-Oléron bay. Sensitivity tests on the shell fish farming block will determine if the coastal ecosystem representation is detailed enough or if other trophic compartments must be added.

### 3.3 Output construction

Classical scientific models deal with a high numerical production intended for an academic audience and are not necessarily adapted to decision support. Integrated socio-ecosystem models should not be designed for predictive or operational purposes. *A contrario* their outputs must highlight their exploratory, learning and communication aspects. The outputs must be carefully designed so as not to be misinterpreted and used out of context by end users. These tools should favour visual and animated indicators rather than classical graphs. The strong hypothesis taken in the *a minima* formulation step must be advertised an clearly documented.

Compared to physical sciences, models in biology, ecology and *a fortiori* in socio-economics require analysis of complex systems but lack data and process knowledge [Voinov et al., 2004 ; Voinov et al., 2008] leading to ‘‘‘black-box’’ much of the underlying complexity’’ [Voinov et al., 2004]. For exploratory models, formulation becomes a creative process [Jakeman, 2008] where uncertainty must not necessarily be avoided, but accepted
and transparently advertised and documented [Brugnach et al., 2008]. Therefore, special care is given to documentation which can be called anywhere in the model in PDF format through “information” blocks (Figure 5).

**Figure 5.** Example of model documentation (E.Coli transport on the watershed)

ExtendSim offers animation features that can be run during simulation, with a dynamic choice of speed. An animation tool was developed to visualise, over a chosen background image, any model variable with dots alternately coloured in green, orange or red depending of its value: under, between or over two selected thresholds. At the end of the simulation, histograms are plotted next to each dot visualising the number of days. This tool is used for the Thau lagoon to display the contamination level at monitoring stations. It is combined to a traffic light tool that visualises the “open/closed” sanitary status of the lagoon for shell fish farming commercialisation (Figure 6). This tool is used in the Pertuis Charentais model, to display the Charente river flow levels compared to regulatory thresholds at stations of the surveillance network. These tools were appreciated by stakeholders as well as by the scientific teams as they help to envision the dynamics of the system in a spatial representation.

**Figure 6.** Animated visualisation of contamination level at monitoring stations. The traffic light visualises the open/closed status of the lagoon for shell fish commercialisation.

4. **Conclusion**

The experience acquired while developing the models of the two French study sites, highlights the need of a detailed methodology in order to build successful integrated socio-ecosystem modelling platforms for deliberation support. Our approach covers a certain number of Westmacott’s [2001] check-list for decision support tools: it incorporates multiple objectives and views, covers multi-disciplinary areas, deals with limited data and information, collates ICZM data and information and plays a learning role in the participant
group. It also involves end users in the model development, offers an easy to use interface and visual displays of results.

At this stage of the project, we were pleased with the positive feedbacks from the participant group. The water management agency EPTB for the Charente river foresees a high potential for exploration of different management scenarios and is now highly involved in the continuing development of the model. They are also very interested by the communication possibilities of the tool for future discussions and negotiations with other management actors and local farmers. For the Thau lagoon, the water management agency (SMBT) representative was particularly interested in the potential for testing the numerous water treatment improvement options with an integrative outlook.

By the end of the SPICOSA project we will be able to offer detailed e-handbooks with the SAF integrated model building methodology as well as reusable ExtendSim block libraries from all the SPICOSA study sites with the ultimate objective of developing a virtual community of ICZM socio-ecosystem integrated assessment modellers.

REFERENCES


Advances in integrated hydrological modelling with the LIQUID® framework

F. Branger\textsuperscript{a}, S. Debionne\textsuperscript{b}, P. Viallet\textsuperscript{b}, I. Braud\textsuperscript{a}, S. Jankowfsky\textsuperscript{a}, O. Vannier\textsuperscript{c}, F. Rodriguez\textsuperscript{d} and S. Anquetin\textsuperscript{c}

\textsuperscript{a}Cemagref, UR HHLY, 3bis quai Chauveau CP220, 69336 Lyon Cedex 09, France (flora.branger@cemagref.fr)
\textsuperscript{b}HYDROWIDE, 1025 rue de la Piscine, 38600 Saint-Martin d’Hères, France
\textsuperscript{c}LTHE, Université de Grenoble (CNRS, INPG, IRD, UJF), BP 53, 38400 Grenoble Cedex 9, France
\textsuperscript{d}LCPC, Nantes, Division Eau et Environnement, BP 4129, 44341 Bouguenais Cedex, France

Abstract: Environmental modelling frameworks are valuable and increasingly used tools for building customized models, in a context where the complexity of management issues and the availability of data require much flexibility. The LIQUID® framework has been developed since 2005. It is mostly dedicated to hydrological modelling. It aims at easily integrating hydrological processes while preserving their characteristic temporal and spatial scales. LIQUID® allows the user to build and run integrated models on the basis of reusable and exchangeable modules. It provides templates for easy development of new modules, connections to databases and GIS for data input and output, and module coupling mechanisms, that synchronize different time steps and handle irregular geometries. LIQUID® suits a wide range of applications, involving various spatial scales and process conceptualisations. The framework is also able to simulate complex interactions between modules, in particular including feedback. The paper will present the recent advances of LIQUID®, in terms of concepts as well as in terms of technical specifications, and a brief overview of the ongoing main applications. Those deal with the assessment of landscape management impact on hydrology in agricultural and suburban areas, and with the analysis of hydrological responses in the context of flash floods.

Keywords: LIQUID®; integrated modelling; framework; distributed hydrological models

1 INTRODUCTION

For several years, environmental modelling frameworks have been increasingly developed and used. They offer the possibility to build and run customized integrated models on the basis of reusable and exchangeable components. Simulation models are built by connecting these components together. The models are run through a communication system that allows components to exchange data during simulation progress. This modular approach suits particularly well water management issues as the setup of hydrological models requires much flexibility, depending on the study objectives, the size of the target catchment, the dominant processes and the availability of input and validation data.

Along with other projects for model integration such as the OpenMI [Gregersen et al., 2007], JAMS [Kralisch et al., 2007], or WaterCAST [Argent et al., 2009], LIQUID® is such a framework. It specifically aims at integrating hydrological processes while preserving their characteristic temporal and spatial scales. It has been developed since 2005 by Hydrowide [Viallet et al.,...]
2006], while many components and models have been built and used by Cemagref and Grenoble University laboratories [Branger et al., 2010]. The objective of this paper is to present the recent advances of LIQUID®. An overview of LIQUID® concepts and technical specifications is given in Sec. 2. Sec. 3 illustrates how LIQUID® is currently used in practice by hydrologists and Sec. 4 gives an insight into past and ongoing applications.

2 OVERVIEW OF LIQUID® CONCEPTS

LIQUID® implements the classical features of environmental modelling frameworks. It manages a library of model components, called modules, that contains conceptualizations for various hydrological processes. Models can be constructed from these modules through a build system. A discrete event simulator, the Scheduler, handles the time course of simulations and allows the modules to communicate at runtime. Tools are available to make the development of new modules easier, such as additional libraries (for example for numerical analysis or geometry computations such as mesh generation), but also a test framework (for unit tests but also more complex module tests, such as comparisons of module outputs with analytical solutions), a system for automatic generation of code documentation and a collaborative development environment that includes a version control system, a website, mailing lists and forums. LIQUID® is developed in C++ and is mostly used under the Windows operating system, although the use of ANSI C++ makes it portable. The modules are mostly implemented in C++, but since recently the integration of code in other languages (e.g. Fortran) is also possible. More detail about LIQUID® concepts and implementation can be found at www.hydrowide.com/liquid.

2.1 Modules and models

A module simulates one or several hydrological processes with specific time and space scales. All types of process representations are allowed, from simple empirical conceptualizations to numerical solutions of partial differential equations.

In LIQUID®, each module is autonomous. The structure of a module consists of a data scheme, a pre-processor and a solver (Fig. 1). The data scheme contains the parameters and initial values that are required for the computation, as well as the description of the simulation domain (geometry) which the module is applied on. Geometries are described as vector data (lines or
polygons, depending on the module), following the principle of space discretization in hydrological response units and hydro-landscapes [Flügel, 1995; Dehotin and Braud, 2008]. According to this data scheme, empty tables are created in a PostgreSQL/PostGIS database, that is connected to LIQUID® through an ODBC connection. Once the user has defined the parameter values, the pre-processor reads the data and initializes the solver with the parameters and initial values. For each record in the data scheme tables, a solver instance is created and initialized.

The solver is where the hydrological computation takes place. Each solver is responsible for its own time step, which is estimated again at each execution according to the solver’s internal state. The solver is also able to interrupt the computation and shorten its time step if required. The communication with the outside world is managed through the solver slots (inputs) and signals (outputs). Slots also indicate the solver response to the inputs. For example, this response can be an interruption in order to take into account a new input value immediately. There is one slot per input variable. Signals send the output values at each solver execution. There is also one signal per output variable.

A model is simply a collection of modules that are connected through their slots and signals.

2.2 Simulation progress

The Scheduler manages the time course of simulations. It can be defined as a shared calendar where the different module solvers schedule their executions. At a given time step during a simulation, when a module solver has estimated its next time step, it schedules its next execution as a new call in the Scheduler. At the scheduled date, the Scheduler then acts as an alarm-clock and calls the solver. Thus simulation progress is possible with a variable time step.

Combined with the slots and signals, this system allows the modules to communicate during a simulation, as represented in Fig. 2. In this example, a signal of Solver1 is connected to a slot of Solver2. Fig. 2b shows what happens if this slot specifies that the next planned execution of Solver2 should be cancelled and rescheduled immediately when a new value is received. This is how solver executions are synchronized in LIQUID®. With appropriate definitions of slots and signals, all sorts of couplings between solvers can be simulated, from the weakest one-way coupling to strong couplings involving feedbacks.
2.3 Spatial connections between modules

In most LIQUID® models, the application catchments are discretized into hydrological response units or hydro-landscapes, with irregular polygon or line geometries. A module instance is then applied on each hydro-landscape. When there are process interactions between the spatial units (e.g. routing of overland flow, groundwater flow...), they must be accounted for through the slot / signal connections. This means that the slots and signals must carry information not only on the exchanged values and dates, but also on which hydro-landscapes they are associated with. The first solution that was implemented in LIQUID® was to connect the slots and signals of the different spatial units according to the identification numbers of the solvers (one to one connection, see Fig. 3a).

Figure 3: Examples of spatial connections managed by the LIQUID® framework: (a) one to one connection between two adjacent hydro-landscapes; (b) n to one connection between several hydro-landscapes and one river subdivided in reaches.

However, with the setup of more complex catchment models, this solution soon appeared to be restrictive as several other connection types may be encountered. As shown in Fig. 3b, there are cases where one solver has several slots / signals for one variable (e.g. for a river solver, one slot / signal for each reach). Therefore \textit{n to one}, \textit{one to n} and more generally \textit{m to n} connections must also be considered. Improvements to the definitions of slots and signals were done recently to handle this problem. Inside a given solver, the slots and signals now have their own identification numbers. The connections between the spatial units are made directly through these numbers.

3 LIQUID® IN PRACTICE

3.1 How to develop a module

A module consists of a set of C++ code files. LIQUID® provides templates and examples, so that a non-expert C++ programmer does not have to bother with the file structuring and just needs to complete a pre-written canvas. The first task is the definition of the module data scheme, and the solver main computing variables, slots and signals. Then the module developer must implement the slots, the pre-processor and the solver main computing functions. LIQUID® provides now default ready-to-use slot implementations for the most common cases. In the solver, the developer can choose freely the most convenient equations and solving methods. The only rule to follow is that the implementations must be time step independent, so that the solver can be run with a variable time step. Therefore, in comparison with more classical approaches, it may be necessary to question the modelling concepts and formulations more in-depth when developing a module.
3.2 How to build a model

In order to build a model, the user must simply select the appropriate modules and connect their slots and signals. Input and output modules are also available. The input module forces the other modules with time series (e.g. rainfall series) stored in a database. The output module prints simulation results in ASCII files. The model assembly task is done through a file with a .model extension (Fig. 4). The LIQUID® build system reads the .model file, translates it to C++ code, compiles the appropriate modules and generates an executable file. As represented in Fig. 4, the user specifies only the names of the modules and those of the slots and signals in the .model file. The effective connections of the slots/signals for the different spatial units of the model are made automatically by LIQUID® at the pre-processing step.

3.3 How to run a model

The model executable file can be run independently from LIQUID®. An ODBC connection to a PostgreSQL/PostGIS database is all that is needed. The run of a model takes three steps. First, a set of empty tables is generated in the user’s database, according to the data schemes of the different modules involved in the model. For each module, the data scheme consists usually of one to five tables. Once all these tables are completed by the user, the second step is the pre-processing of the modules and the initialization of the module solvers. The last step is the model run itself, according to the start and end dates indicated by the user. All these steps can be done in a command line environment and since recently through a simple graphical user interface (Fig. 5). Model outputs are stored in text files, so that the simulation results can be visualized and processed by the user’s favorite software. We are currently using mostly the R1 software for which we have developed scripts for plotting LIQUID® models outputs.

4 ONGOING APPLICATIONS

Since 2005 more than 20 modules have been developed, representing a wide range of hydrological processes, conceptualizations and scales. The module library currently contains several modules for soil infiltration and surface runoff, with either conceptual (reservoir-based) or physically-based (Richards equation) approaches. Some of them are designed for specific environments (agricultural fields, hedgerows, urbanized areas). There are also a set of modules for evapotranspiration, including plant growth and root extraction, two modules for river flow, three modules for groundwater flow. Specific modules were developed for agricultural drainage, including a conceptual

1http://www.r-project.org/
Figure 5: The graphical user interface used to run LIQUID® models

pesticide transport module. A lumped rainfall-runoff module is also available.

LIQUID® has been used for local scale applications, with the development of a conceptual model for pesticide transport in a tile-drained field [Branger et al., 2009], and a more detailed study of the influence of pipe pressurization on the discharge of a tile-drained system [Henine and Nedelec, 2009]. At the catchment scale, LIQUID® models were built and used to investigate:

- the effect of landscape management practices (drainage, hedgerows, ditches) on the hydrology of a small agricultural catchment [Branger et al., 2008]
- the sensitivity of models to the rainfall distribution and to the description of soil properties variability in the context of flash floods on mid-size catchments [Manus et al., 2009]
- the sensitivity of long-term water balance to modifications of land use on large catchments [Dehotin, 2007; Dehotin et al., 2010].

At the moment, we are mainly working in two directions. First, the CVN model for simulation of flash floods is being improved, with the inclusion of evapotranspiration processes in order to study the influence of initial conditions on catchment response [Manus, 2008; Vannier, 2009]. Second, a model for small suburban catchments, PUMMA, is being developed in order to assess the influence of urban extension on the hydrological regime of small suburban rivers. The objectives, development and first results of the PUMMA model are presented in detail in a companion paper [Jankowfsky et al., 2010].

5 CONCLUSIONS AND PERSPECTIVES

LIQUID® is an environmental modelling software tool developed since 2005. It is dedicated to hydrological modelling and allows the user to build custom-made models on the basis of hydrological process modules available from a module library. The modules are applied to irregular geometries. A discrete event simulator called Scheduler and a callback mechanism through slots and signals enable simulations with variable time steps that respect the characteristic time and space scales of the hydrological processes. The strong points of LIQUID® are the full freedom that is given to module developers, who can develop modules with any process representation, and also the efficiency of the coupling system that is able to synchronize time steps and data exchange of modules that have very different process conceptualizations.
LIQUID® has been continuously improved since 2005. The improved definition of slots and signals now make module programming and spatial connections easier and transparent to module developers. For model users, the addition of a graphical user interface simplifies the launch of simulations, although it is still necessary to master several software tools (in particular related to PostgreSQL/PostGIS).

All these characteristics make LIQUID® a flexible tool that is well suited to research applications for a wide range of spatial and temporal scales, from the most detailed physically-based study to more conceptual meso-scale model applications. Next developments will be oriented towards more computation efficiency and user (and developer) friendliness. In particular, challenging issues will be the addition of facilities for the display of simulation results, the inclusion of parameter optimization mechanisms and the improvement of the user interface.

ACKNOWLEDGMENTS

The work presented in this paper has been supported by Hydrowide, Cemagref and CNRS fundings. Fundings were also obtained from the European FP6 project HYDRATE under the n° GOCE 37024 and from the French National Research Agency (ANR) project AVuPUR under contract ANR-07-VULN-01.

REFERENCES


An Integrated Assessment approach to linking biophysical modelling and economic valuation tools

M.E. Kragt\textsuperscript{a,b}, J.W. Bennett\textsuperscript{a} and A.J. Jakeman\textsuperscript{b}
\textsuperscript{a} The Crawford School of Economic and Government, The Australian National University, Canberra
\textsuperscript{b} Integrated Catchment Assessment and Management Centre, The Fenner School of Environment and Society, The Australian National University, Canberra
(marit.kragt@anu.edu.au / jeff.bennett@anu.edu.au / tony.jakeman@anu.edu.au)

Abstract: Catchment natural resource management (NRM) involves complex decisions that affect a wide variety of values, issues and stakeholders. Designing efficient NRM policies requires assessments of the environmental impacts, and costs and benefits of management interventions in an integrated manner. However, despite the need for integrated assessment (IA), there are few comprehensive frameworks that integrate biophysical models with economic valuation. Cost-benefit analysis (CBA) is a framework that can support efficient NRM by assessing and comparing the total social costs and benefits of management interventions. However, the environmental modelling that has underpinned CBA has typically been poor, reducing the credibility of the framework to assist in the formulation of policy efficiencies. IA provides an approach to integrate the several dimensions of catchment NRM by considering multiple issues and knowledge from various disciplines and stakeholders. In this paper, we demonstrate how IA can be used to consistently integrate economic information with environmental data in a systematic framework to guide economically efficient decision making. We develop a Bayesian Network (BN) model that can be used as a decision support tool to evaluate the welfare impacts of NRM actions.

Keywords: Bayesian Networks; Catchment Management; Cost-Benefit Analysis; Environmental values; Integrated Assessment and Modelling; Non-market valuation

1. INTRODUCTION

Natural resource management (NRM) typically entails complex decision problems that involve a variety of issues and evolve in a dynamic social context [Ritchey, 2004]. Integrated assessment (IA) provides an approach to analyse the various dimensions of NRM by considering multiple issues and sharing scientific and stakeholder knowledge drawn from multiple disciplinary backgrounds [TIAS, 2009]. IA can support the development of economically efficient catchment NRM if all the marginal social costs and benefits associated with the impacts of alternative NRM actions are assessed. However, despite an increased interest in IA, there are few integrated modelling studies that combine biophysical modelling tools with economic valuation techniques in a robust framework to guide NRM decisions [Kirkpatrick and Lee, 1999; Croke et al., 2007].

Environmental impacts and financial costs and benefits of NRM changes may be relatively easy to estimate. However, changes in catchment environments will also impact non-market (intangible) values that people derive from ecosystem goods and services [Hanley and Barbier, 2009: 40]. Non-market economics valuation tools can be used to obtain estimates of the non-market costs and benefits of NRM policies. Although there are challenges involved in estimating non-market values [Hanley and Barbier, 2009: 55-61, 67-70 and 91-93], not accounting for non-market values of environmental impacts may lead to a misallocation of resources and less efficient decision making [Bennett, 2005]. The decision framework for economic valuation is based on cost-benefit analysis (CBA). The CBA decision rule is that if the marginal benefits of a policy change exceed its costs by a larger amount than any other management alternative, then the proposed policy should be adopted.

In this paper, we demonstrate how economic non-market valuation tools can be integrated with predictions of biophysical changes in one Bayesian Network (BN) modelling framework, to support an IA of catchment NRM changes. The BN integrates a process-based water quality model, ecological assessments of native riparian vegetation, estimates of management costs and non-market values of changes in riparian vegetation for the George catchment, Tasmania. The modelling approach illustrates

1011
how different tools can be combined in one framework to evaluate the environmental and economic impacts of NRM actions, as well as the uncertainties associated with the estimated welfare effects. The evaluation of impacts in a CBA framework can provide more economic rationality to NRM decisions. The next section introduces the IA approach that underlies this study. The various tools that were used to predict impacts of NRM actions on catchment water quality, native riparian vegetation and non-market environmental values are briefly described in Section 3. The model development process and the techniques used to integrate information about multiple systems in the BN model are described in Section 4. The results are illustrated by a model scenario in Section 5. The final section concludes the paper.

2. INTEGRATED ASSESSMENT

In this study, an IA approach is used to assess environmental and socio-economic changes resulting from catchment NRM options (Figure 1). Different tools can be employed to inform the various stages of the assessment process [De Ridder et al., 2007].

An IA process starts with developing an understanding of the issues and system variables under consideration [Jakeman et al., 2006]. An iterative IA participatory approach that involves multiple stakeholders can strengthen a shared understanding about the economic, environmental and social issues of concern. Stakeholders in the process may include different scientific disciplines, model developers, natural resource managers, and/or local landholders, who will typically have different (and sometimes conflicting) ideas about the issues at stake. Conceptual influence diagrams or cognitive mapping techniques may be used to describe the system variables and their interrelationships.

The aim of the second phase is to identify the range of alternative future management scenarios that may be undertaken to address the issues identified in the first phase (including a ‘business as usual’ scenario) [De Ridder et al., 2007]. It is important that the scenarios match the (scientific, political and socio-economic) context of the system and are relevant to the stakeholders involved. The identification of alternative courses of action can be aided by, for example, surveys, focus group discussions or other tools such as General Morphological Analysis [Ritchey, 2004].

All the potential bio-physical impacts of the alternative management actions need to be assessed. Science-based modelling tools are useful to characterise environmental processes, and to predict changes in a range of (biophysical) indicators.

Figure 1. Analytical steps in an IA process to policy analysis.

1 In this paper, the term ‘environmental’ refers to natural systems and impacts on biophysical indicators.
IA modelling studies often focus on natural systems, with a sparse representation of socio-economic costs and benefits [Ward, 2009]. Economic valuation tools are needed to estimate the cost and benefit impacts of NRM actions in monetary terms.

The impacts of alternative policies on multiple system variables need to be evaluated to provide support for decision making. The use of multiple indicators in biophysical modelling means that impacts are measured in disparate units, which does not allow for a comparison of impacts in a meaningful way [Brouwer et al., 2003: 32]. CBA provides a decision making framework to consistently compare NRM impacts across different systems (such as water quality and biodiversity changes), by measuring all impacts in identical (monetary) units² [Ward, 2009]. This enables an analysis of the trade-offs between the marginal costs and benefits of alternative policy proposals and can aid decision making to evaluate the economic efficiency of management changes.

3. METHODS

A variety of tools were used to inform an IA of catchment management changes in the George catchment, Tasmania. This study engaged multiple academic disciplines along with public and other stakeholder representatives, to develop an integrated BN model that incorporates water quality changes, riparian vegetation condition and economic costs and benefits.

3.1 Water quality modelling

A physically based, semi-distributed catchment model was developed for the George catchment, based on the Catchment Scale Management of Diffuse Sources framework [CatchMODS - Newham et al., 2004; Drewry et al., 2005]. The CatchMODS framework was adapted for the George catchment to integrate a range of process-based hydrologic, erosion and economic sub-models that simulate the impacts of different management actions on river flows, sediment delivery and nutrient loads, calculated as steady-state averages [Kragt and Newham, 2009].

3.2 Choice Experiments

Information about the non-market value impacts of changed catchment NRM was elicited using choice experiment (CE) techniques. In a CE survey, respondents are presented with a series of choice questions describing the outcomes of alternative hypothetical policy scenarios [Bennett and Blamey, 2001]. These outcomes are described in terms of different levels of a monetary attribute (costs) and several non-marketed attributes. Respondents are asked to choose their preferred option in each choice question. This allows an analysis of the trade-offs that respondents make between attributes. If cost is included as one of the attributes, these trade-offs can be used to estimate the marginal value of each environmental attribute in monetary terms.

For the present study, a CE survey was developed using a combination of literature review, biophysical modelling, interviews with science experts and regional natural resource managers and feedback from focus group discussions [Kragt and Bennett, 2008]. An example choice question is shown in Figure 2. The survey was administered in various regions in Tasmania between November 2008 and March 2009.

Figur**e 2. Example choice question in the George catchment CE [see: Kragt and Bennett, 2009].

---

² Note that an assessment of physical impacts remains an essential prerequisite to environmental valuation.
3.3 Bayesian Networks

A major challenge in this study was the integration of knowledge from different sources into a logically consistent modelling framework. A process-based model provided predictions of water quality changes. Literature values and expert judgements were used to assess changes in ecosystem variables. CE survey data provided information about non-market value impacts. BN modelling techniques were used here to combine the information from various assessments in a single integrated model for decision support.

BNs (sometimes called belief networks) are probabilistic, graphical models consisting of a directed acyclic graph of variables (called ‘nodes’). The values that each variable can assume are classified into discrete, mutually exclusive ‘states’. BNs can incorporate different data sources, including expert opinion when observational data is not available [Pearl, 1988]. The propagation of information between variables is described by conditional probability distributions, thus incorporating the uncertainties in relationships between [Borsuk et al., 2004]. BNs are widely used for knowledge representation and reasoning under uncertainty in NRM and have been applied to different catchment issues [see, for example, Bromley et al., 2005; McCann et al., 2006; Castelletti and Soncini-Sessa, 2007]. There are, however, few BN applications that focus on non-market economic impacts of environmental changes [an exception is described in Barton et al., 2008].

4. THE GEORGE CATCHMENT MODEL

The George catchment, in North-Eastern Tasmania, was chosen as a suitable study area for both the biophysical modelling and the economic research. Land use is dominated by native vegetation, forestry, plantations and agriculture. The catchment has significant socio-economic values through its environmental assets, recreational values and aquaculture production in the estuary. Although the catchment environment is currently in good condition [Davies et al., 2005], there are significant concerns that land use changes are affecting ecosystem conditions [Sprod, 2003; BOD, 2007].

The first phases of the IA process were aimed developing a conceptual influence diagram to define the scale and scope of the George catchment system. Natural scientists, policy makers and community stakeholders were involved in the conceptual model development3, to ensure that the considered variables and links between variables matched the scientific and policy context of the system. The geographical scale of the system was based on the contours of the George catchment, delineated using digital elevation models. A twenty-year period was considered an appropriate temporal scale, as it is thought to be long enough for management changes to have a demonstrated biophysical impact on the George catchment environment, and short enough to be pertinent to policy makers and CE survey respondents.

Further model development was an iterative process, aimed at identifying a parsimonious model that would represent the interactions between catchment management actions and environmental variables that impact human welfare [Kragt et al., in press]. The resulting conceptual model for the George catchment (Appendix A) incorporated four local management changes that were assumed to impact catchment ecosystem conditions: stream-bank engineering works; riparian zone management through limiting stock access to rivers and establishing buffer zones; changed catchment land use, and; vegetation management through weed removal. Some of these actions are already being implemented in the George catchment on a small scale, which increases the plausibility of the modelled management scenarios. Three indicators of George catchment environmental conditions were considered in the conceptual model: native riparian vegetation, number of rare native species and the area of seagrass in the estuary. The ecosystem component was integrated with the CE survey by using expert predictions and modelling results of changes in the ecosystem indicators as environmental attribute levels in the CE survey.

There was not enough information about changes in all the variables included in the conceptual model to develop a fully functioning BN for the George catchment. To adequately populate the conditional probability tables for all variables, one needs to know the probability that a certain state is observed at every possible combination of the input variables. Within the time frame of this study, it was not feasible to collect data about all the variables in the conceptual model and specify the relationships between them as probability distributions. Research efforts therefore focused on a sub-section of the conceptual model. A BN was developed that combines the costs of management actions with predictions of river water quality, native riparian vegetation length and non-market values (Figure 3). Each of the model variables is

---

3 The consultation process involved three workshops with Tasmanian scientist between November 2007 and September 2008, 31 structured interviews with experts on river health, threatened species, bird ecology, forestry management, riparian vegetation, estuary ecology and local natural resource managers and eight focus group discussions with members of the public in Hobart, St Helens and Launceston in February and August 2008.
described in more detail in Appendix B. The different techniques used to predict the levels of the variables and the ways in which they were integrated in one BN model are described below.

4.1 Predicting management costs

The main focus of this research was the integration of biophysical modelling and non-market valuation techniques. However, in order to demonstrate how the integrated model can be used in a CBA, the financial costs associated with implementing and maintaining management actions were included in the model. Assumptions about the costs of NRM in the George catchment were based on literature values (see Appendix B). The impacts of land use changes were represented as the change in aggregate land use returns for different catchment land use scenarios. The marginal costs of establishing riparian buffer zones and stream-bank engineering works were calculated as the present value (PV) of the summed one-off implementation costs and discounted maintenance costs over a twenty year period. A discount rate of three percent was used in the PV calculation.

Notwithstanding efforts to obtain accurate information, the knowledge about management costs in the George catchment remains limited. Uncertainties arise from, for example, knowledge gaps about the returns to land use, the types of materials used and the labour time involved in implementation and maintenance. These uncertainties are represented in the BN model by estimating a range of costs, rather than a single value. Given the limited number of data-sources and the high levels of uncertainty in knowledge, the predicted costs should be seen as illustrative rather than accurate estimates for a CBA.

4.2 Predicting water quality changes

George-CatchMODS was used to predict the impacts of management changes on steady-state mean annual river flow (MAF in ‘000 ML/year), and total suspended sediment (TSS in tonnes/year), total phosphorus (TP in tonnes/year) and total nitrogen (TN in tonnes/year) loadings to the George catchment streams and estuary. Monte Carlo simulations of George-CatchMODS were run that combined scenarios of land use changes with varying lengths of stream-bank engineering works and riparian buffer. The results from these Monte Carlo simulations were used to define the conditional probability distributions for the water quality variables. Uncertainties in the predictions arise from uncertainty in the model parameters and were specified as an uncertainty bound around the deterministic predictions from the George-CatchMODS model. Note that George-CatchMODS is integrated with the other components by considering the same management scenarios in each model.

4.3 Predicting impacts on native riparian vegetation

The impacts of NRM actions on native riparian vegetation were predicted based on information collected through literature reviews and expert interviews [Kragt and Bennett, 2008]. The most important management actions assumed to impact native riparian vegetation in the George catchment are land use changes, establishing riparian buffer zones and weed management (Figure 3). An intermediate node (‘Native Veg in riparian zone given different land uses’) was included to measure the length of native vegetation in the riparian zone under alternative land use scenarios. Assumptions about the proportion of the riparian zone that is likely to be vegetated under each land use, and the ‘naturalness’ of that vegetated riparian zone were based on Tasmanian digital vegetation mapping [DPIW, 2005a; DPIW, 2005b] and expert review [Daley, 2008]. It was assumed that areas with native vegetation for non-production purposes have a fully vegetated riparian zone with at least 80 percent native vegetation. Native production forests and forestry plantations were assumed to have a respectively 90 and 80 percent vegetated riparian zone, with 70 and 30 percent native vegetation respectively. The base case assumption was that agricultural and urban areas did not have any vegetation in their riparian zones, but that the establishment of riparian buffers and weed management could increase this.

The ‘Length of Native Riparian Vegetation’ variable in Figure 3 measures the total length of rivers in the George catchment with healthy native vegetation along both sides of the river. The intermediate node ‘Native Veg given land use’ was assumed to contribute directly to the total Length of Native Riparian Vegetation in the George catchment. The ‘naturalness’ of newly established riparian buffers was assumed to depend on the extent of weed management in the riparian zone: (i) ‘low’ weed management was assumed to result in 15% of healthy native vegetation; (ii) ‘medium’ weed management was assumed to

---

4 The review included regional, State and National documents about the impacts of catchment management on native vegetation conditions, and previously developed models of vegetation changes in river catchments. Structured interviews were conducted with Tasmanian experts on river health and riparian vegetation.

5 Note that establishment of riparian buffer does not change catchment land use in our model.
result in 50% of healthy native vegetation; and (iii) ‘high’ weed management was assumed to result in 85% of healthy native vegetation in the established riparian buffer zones [Daley, 2008]. These assumptions mean that if, for example, six km of riparian buffer is established with ‘medium’ weed management, the contribution to the total Length of Native Riparian Vegetation in the George catchment is three km (in addition to the native vegetation in the riparian zone under the given land use scenario). Uncertainty in the assumptions was accounted for by imposing a 95% uncertainty bound on the calculated values.

The riparian vegetation model was used to predict the length of native riparian vegetation in the George catchment under a ‘best case’ and ‘worst case’ scenario. The predictions ranged from 40km (the ‘worst case’ scenario) to 81km (the ‘best case’ scenario) and were integrated with the attribute levels in the CE survey [Kragt and Bennett, 2009].

4.4 Estimating non-market values

Results from the CE study were used to estimate non-market values of native riparian vegetation in the George catchment. Marginal willing to pay (WTP) estimates indicated that Tasmanian households were, on average, WTP $3.57 for every km increase in native riparian vegetation, over the presented base case scenario of 40km of native riparian vegetation [Kragt and Bennett, 2009]. The CE results also provided information about the uncertainty range in the WTP distribution, with an estimated standard deviation of 0.532.

Household marginal WTP was aggregated over the total numbers of households in the ‘relevant’ population to calculate the total non-market values of changed native riparian vegetation condition in the George catchment. What constitutes the ‘relevant’ population and which proportion of this population has a positive WTP is subject to debate [Morrison, 2000]. To reflect the aggregation issue, an additional variable ‘Aggregation assumptions’ was included in the BN. This variable represents three alternative assumptions for aggregating the household WTP estimates: (i) Only the survey respondents have a positive WTP = 832 households; (ii) 64 percent of all households at the sample locations has a positive WTP = 35,799 households; and (iii) 64 percent of all Tasmanian households have a positive WTP = 116,418 households [ABS, 2006a].

5. RESULTS

The process-based water quality model and native riparian vegetation assessment predicted the state of the environmental conditions in the George catchment, given a certain management input. The economic estimates of marginal values required predictions of changes in environmental conditions resulting from implementing new management actions. In the integrated model, this was achieved by using predictions from the biophysical models before and after the management change (Figure 3).

5.1 Scenario analysis

To illustrate how the model enables an integrated impact assessment of NRM actions on a range of system variables, results of an example scenario are presented in Figure 3. In this scenario, land use in the George catchment is as currently observed, and no stream-bank engineering works are undertaken. The top part of the figure illustrates the predicted environmental conditions before implementing new management actions. For example, the model predicts a 73.3 percent probability that TSS loads are between 6900 and 8000 tonnes/year. The bottom part of Figure 3 illustrates the impacts of establishing ‘between six and twelve’ kilometres of additional riparian buffers combined with ‘medium’ weed management actions. TSS loads are now predicted to be between 6100 and 6900 tonnes/year. The direct costs of establishing additional riparian buffers are approximately $149,000 (Figure 3). Uncertainty in the predicted costs is represented in the model by predicting a 92.3 percent probability that costs are somewhere between $100,000 and $200,000.

In the presented scenario, the length of native riparian vegetation is most likely to increase from between 45 and 67 kilometres (‘before’) to between 67 and 78 kilometres (‘after’). Note that uncertainty in the model still leads to a 32.4 percent probability that the length of native riparian vegetation remains between 45 and 67 kilometres.

If we assume that 64 percent of the population at the sample locations has a positive WTP for riparian vegetation changes, there is a 32.4 percent probability that the total non-market value (NMV) of the change in native riparian vegetation is between two and five million dollars. However, uncertainty in the

---

6 The average survey response rate was 64 percent (Kragt and Bennett, 2009b).
predicted length of native riparian vegetation, and uncertainty in household WTP results in a predicted probability of 24.3 percent that the total NMVs are between one and two million dollars, and even a 21.9 percent probability that there is no change in NMVs at all. Hence, although the length of native riparian vegetation is likely to increase as a result of establishing riparian buffer zones in the George catchment, there remains a probability that the benefits will not outweigh the costs.

5.2 Sensitivity analysis

Sensitivity analyses were also conducted, to assess which variables have the largest influence on the uncertainty in predicted length of native riparian vegetation and total non-market values. These analyses revealed that, in our model, establishing new riparian buffer zones, land use changes, and the assumptions about native vegetation under different land uses have the largest impact on uncertainty in the predicted total Length of Native Riparian Vegetation. Uncertainty in the predicted total non-market values is mostly affected by uncertainties in the predicted Length of Native Riparian Vegetation, establishing riparian buffers and land use changes.

Figure 3. Scenario analysis of establishing between 6-12 km new riparian buffers with ‘medium’ weed management in the George catchment, assuming that 64 percent of the population at the sample locations have a positive WTP and keeping land use and stream-bank engineering constant.
6. DISCUSSION AND CONCLUSION

The research described here aimed to assess the biophysical and economic impacts of catchment NRM actions in the George catchment, Tasmania, in an integrated manner. Following an IA approach to model development, various academic disciplines, policy makers and community stakeholders were engaged in the project. The iterative consultation process provided valuable information about different stakeholder perspectives, to be included in the final integrated model. Probabilistic modelling techniques were used to integrate results from deterministic models, expert interviews and survey data into one BN model. The integrated process to developing the biophysical models and the economic non-market valuation survey tailored the information exchange between separate model components and ensured that the outputs of the different tools were compatible with each other.

The integrated BN can be used to assess the impacts of NRM actions on a range of indicators, including water quality parameters, native riparian vegetation condition and non-market environmental values. Including the management costs of NRM actions as well as non-market benefits allows a CBA to determine which management investments deliver the greatest net returns to society. Contrary to traditional CBA studies, the BN modelling approach used here accounts for uncertainties in the relationships between NRM actions, environmental impacts and economic consequences in a probabilistic way. The wide probability distributions in the scenario predictions show the large uncertainties in predicted costs and benefits. The explicit recognition of these probabilities enables an assessment of the risks associated with implementing new management actions.

Some challenges related to using BN modelling should also be mentioned here. The experts involved in the model development process found it difficult to express their knowledge about relationships between variables as probability distributions. Another limitation of BN models lies in its use of discrete states, rather than continuous probability distributions. Information losses arise from discretisation of probability distributions, which may affect modelling outcomes.

The model development was based on limited information about management costs and ecosystem changes in the George catchment. This means that model predictions of the net welfare impacts should not be considered as accurate inputs into a CBA. Results from the sensitivity analysis indicated that future research is needed to more accurately predict the impacts of riparian buffers or land use changes on native riparian vegetation. It is also recommended that the estimated management costs undergo further peer review to reduce the uncertainty in the model predictions.

7. REFERENCES


DPIW, TASVEG, the Tasmanian Vegetation Map. Department of Primary Industries and Water, Information and Land Services Division: Hobart, 2005b.


Appendix A. Conceptual model for the George catchment, incorporating four management actions (stream-bank engineering, creating riparian buffer zones, land use changes and weed management) and three environmental attributes (seagrass, rare native species and native riparian vegetation)
### Appendix B. Variables in the integrated model for the George catchment model

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>States</th>
<th>Variable type</th>
<th>Data/information sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Costs of undertaking stream-bank engineering works</td>
<td>Present value of the one-off implementation costs of stream-bank engineering works plus the discounted maintenance costs</td>
<td>0, 0-50, 50-100, 100-150, 150-200, 200-400 (’000$)</td>
<td>Utility, continuous</td>
<td>Literature values from NLWRA [2000], Liff [2002], Freeman and Dumsday [2003], Sprod [2003], ABS [2006b], FPA [2007], Thorn [2007] and ABARE [2009].</td>
</tr>
<tr>
<td>Changed returns to catchment land use</td>
<td>Difference in total returns to land use in the George catchment between alternative land use scenarios.</td>
<td>&lt;-10, -10to-5, -5to-2, -2to0, 0, 0-2, 2-5, 5-10, &gt;10 ($m)</td>
<td>Utility, continuous</td>
<td></td>
</tr>
<tr>
<td>Costs of established riparian buffer zones</td>
<td>Present value of the one-off implementation costs of establishing a riparian buffer zone plus the discounted maintenance costs associated with continuing weed management in the riparian buffer zone</td>
<td>0, 0-40, 40-100, 100-200, 200-500, 500-2,500 (’000$)</td>
<td>Utility, continuous</td>
<td></td>
</tr>
<tr>
<td>Stream-bank engineering</td>
<td>Length of stream-bank engineering works undertaken in the George catchment to reduce stream-bank erosion</td>
<td>none, 0-3, 3-7, &gt;7 (km)</td>
<td>Management action, continuous</td>
<td>Observed length of actively eroding sites from George Rivercare Plans [Liff, 2002; Sprod, 2003].</td>
</tr>
<tr>
<td>Changing catchment land use</td>
<td>Changes in the total catchment area under alternative land uses (native vegetation non-production, native production forest, forestry plantations, grazing pastures, irrigated agriculture, urban area)</td>
<td>Current land use, loss native vegetation, expanding native vegetation, expanding production forest, expanding plantation forest, expanding agriculture, urbanisation (low, medium, high)</td>
<td>Management action, discrete</td>
<td>Modelling assumptions</td>
</tr>
<tr>
<td>Establishing riparian buffer zones</td>
<td>Length of riparian buffers established on agricultural and urban lands to reduce stream-bank erosion and trap sediment runoff from hill-slope erosion</td>
<td>none, 0-6, 6-12, &gt;12 (km)</td>
<td>Management action, continuous</td>
<td>Modelling assumptions</td>
</tr>
<tr>
<td>Weed management</td>
<td>Weed control measures and planting native vegetation to improve the naturalness of the riparian zone</td>
<td>low, medium, high</td>
<td>Management action, discrete</td>
<td>Australian National Resource Atlas [NLWRA, 2000]</td>
</tr>
<tr>
<td>River total suspended sediment (TSS)</td>
<td>TSS loads into the Georges Bay at St. Helens under alternative management scenarios</td>
<td>4500-5500, 5500-6100, 6100-6900, 6900-8000, 8000-12300 (tonnes/year)</td>
<td>Nature, continuous</td>
<td>Modelled in George-CatchMODS water quality model</td>
</tr>
<tr>
<td>River total phosphorus (TP)</td>
<td>TP loads into the Georges Bay at St. Helens under alternative management scenarios</td>
<td>2.4-3.6, 3.6-4.1, 4.1-4.6, 4.6-5.7, 5.7-12 (tonnes/year)</td>
<td>Nature, continuous</td>
<td>Modelled in George-CatchMODS water quality model</td>
</tr>
<tr>
<td>Variable</td>
<td>Description</td>
<td>States</td>
<td>Variable type</td>
<td>Data/information sources</td>
</tr>
<tr>
<td>---------------------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>---------------------------------------------</td>
<td>---------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>River total nitrogen (TN)</td>
<td>TN loads into the Georges Bay at St. Helens under alternative management scenarios</td>
<td>66-80, 80-90, 90-100, 100-120, 120-220 (tonnes/year)</td>
<td>Nature, continuous</td>
<td>Modelled in George-CatchMODS water quality model</td>
</tr>
<tr>
<td>Mean annual flow (MAF)</td>
<td>Total river flows into the Georges Bay at St. Helens under alternative land use scenarios</td>
<td>178-183, 183-188, 188-191, 191-203, 203-230 ('000 ML/year)</td>
<td>Nature, continuous</td>
<td>Modelled in George-CatchMODS water quality model</td>
</tr>
<tr>
<td>Native Veg in the riparian zone given different land uses</td>
<td>The total length of native vegetation in the riparian zone under alternative land use scenarios</td>
<td>&lt;60, 60-65, 65-70, &gt;70 (km)</td>
<td>Nature, continuous</td>
<td>Calculated in the model, based on modelling assumptions</td>
</tr>
<tr>
<td>Length of Native Riparian Vegetation</td>
<td>The total length of native riparian vegetation given land use changes, creation of riparian buffers and weed management</td>
<td>&lt;45, 45-67, 67-78, &gt;78 (km) (equivalent to &lt;40%, 40-60%, 60-70%, &gt;70% of total catchment stream length)</td>
<td>Nature, continuous</td>
<td>Calculated in the model, based on expert assumptions</td>
</tr>
<tr>
<td>Aggregation assumptions</td>
<td>Assumptions on the total number of households in Tasmania with a positive marginal willingness-to-pay</td>
<td>Only sampled households ( = 832), RR at sample locations ( = 35,799), RR at all TAS ( = 116,418)</td>
<td>Nature, discreet</td>
<td>Modelling assumptions based on CE response rate and total number of households in Tasmania</td>
</tr>
<tr>
<td>Household WTP for changes in native riparian vegetation</td>
<td>Household marginal willingness-to-pay for every additional km of native riparian vegetation, compared to the base case scenario (= 40km of native riparian vegetation left in the catchment)</td>
<td>&lt;2, 2 to 3, 3 to 4, 4 to 5, 5 to 6, &gt;6 ($)</td>
<td>Nature, continuous</td>
<td>CE survey results</td>
</tr>
<tr>
<td>Total NMVs of changes in native riparian vegetation</td>
<td>The total non-market value of increased length in native riparian vegetation in the George catchment, compared to the base case scenario</td>
<td>0, 0-0.5, 0.5-1, 1-2, 2-5, 5-10, 10-20, 20-45, &gt;45 (m$)</td>
<td>Utility, continuous</td>
<td>Equation combining parent nodes ‘WTP’, ‘Aggregation assumptions’ and ‘Native Riparian Vegetation’</td>
</tr>
</tbody>
</table>

^ Discounted at three percent over a twenty year period.
An integrated, multi-modelling approach for the assessment of water quality: lessons from the Pinios River case in Greece

C. Makropoulos*, E. Safiolea*, S. Baki*, E. Douka*, A. Stamou*, and M. Mimikou*

*Dept. of Water Resources and Environmental Engineering, School of Civil Engineering, National Technical University of Athens (cmakro@mail.ntua.gr)

Abstract: Major factors influencing water quality along a river are land use practices, seasonal hydro-meteorological conditions, groundwater interactions and wastewater discharges. These complex water quantity and water quality aspects demand integrated solution approaches. The study links hydrologic, hydraulic and water quality models using the OpenMI standard to evaluate water quality in the Pinios River in Greece. OpenMI allows data to be exchanged at run time, between models from different providers, thus facilitating integrated modelling. The Pinios River was selected due to its high intensity of cultivation with water demanding crops. The objectives of the overall project, of which this study forms a part, were to assess water quality during extreme events and identify areas where any further pollution could be critical. A multi-modelling approach was utilized, where two separate integrated models were developed by two different research groups, each combining, using OpenMI, commercial and academic model components, thus creating a form of modelling ensemble. The assumptions and results are compared and critically discussed. The study’s conclusions also address generic integrated modelling issues such as the benefit of bi-directional links and integrated model stability. They also identify challenges in model comparison, within a multi-modelling framework in view of differences in conceptualization, discretisation and solving schemes chosen by different researchers, which become apparent once the barriers for direct comparison are alleviated, with the use of approaches such as OpenMI.

Keywords: decision support; integrated modelling; multi-modelling; pollution; uncertainty; water quality

INTRODUCTION

The water environment in Europe (and overseas), faces increasing pressures from multiple facets of human activities, including water abstractions and wastewater generation from urban developments and agriculture. Current environmental legislation (of which the Water Framework Directive (2000/60/EC) is the more often quoted example), sets ever more stringent requirements for water quantity and quality objectives in water bodies, in the presence of natural variability and uncertainty. This trend necessitates the use of better tools (Gourbesville, 2008) able to capture the interactions between interventions, natural processes and environmental objectives. Such tools, modelling the environment (preferably) at the catchment scale, require data availability (due to the number of processes that need to be modelled), processing power (due to the complexity of individual models) and a way of seamless interaction between them (due to the requirement of interaction between models). This paradigm of focusing on process interactions through the coupling of multiple individual models, which has been known for some time as “integrated modelling” (Beven, 2007) is now moving even further away from large all-encompassing models, towards a more model-component based approach (Argent, 2005) where each model is seen as a component of the integrative whole.
This approach also enables the exploration of the effect of changing model representations within the integrated model (or model ensemble) by swapping one model with another (Figure 1).

Such a capability presents multiple benefits:
- It facilitates the improvement of integrated models, whenever a better modelling component becomes available for some part of the system (something valuable to software developers (see Argent, 2005))
- It allows a direct comparison of the effect of different system conceptualisations on the uncertainty of the results (which is crucial in situations where the modelled system is highly complex – e.g. in climate model ensembles (Fowler and Ekstrom, 2009) but also in complex water management problems (Makropoulos et al., 2008)). This approach is often termed multi-modelling (Stranger, 2000).
- It allows for a seamless switching between data sources, by changing modelling components that link the integrated model to specific databases (e.g. online data source, Fotopoulos et al (in press)).
- It allows a harmonious, synergistic co-existence between model components from different developers (incl. for example commercial models and research prototypes)

Figure 1. A component-based multi-modelling approach

However, to allow for such a seamless (drag n drop) “swapping” of model components there is a need for modelling mechanisms (standards, frameworks etc (Argent, 2004)) that handle the (possibly bi-directional) data exchange between models, in runtime. In this work we used, one of the most promising technologies available for model linking: the OpenMI (Open Modelling Interface) Standard (Moore and Tindall, 2005). OpenMI is a software component interface definition for the computational core (the engine) of the hydrological and hydraulic models (Gregersen et al., 2007).

The OpenMI Standard dictates the way models can be linked to other models and exchange information in real (run) time (Gregersen et al., 2007) without using external files. In this way, integrated models can be created using OpenMI compliant models (or model components) from different providers, thus enabling the end-user to select those models that are best suited to a particular problem. OpenMI supports two-way links where the involved models mutually depend on results calculated by each other (Gregersen et al., 2007). Linked models may run asynchronously with respect to timesteps and data represented on different geometries (grids) can be exchanged seamlessly (Figure 2).

The paper compares two OpenMI integrated models (with a number of common components) developed by different research groups which were applied to the same case: a part of the Pinios River basin in Greece.
Both integrated models explore water quality issues and specifically the influence of diffused pollution sources on the river quality. First a description of the problem is included to facilitate the understanding of the approach. Then the methodology is explained, by focusing on the model ensembles and some results are presented and discussed. The main discussion of the paper is in the presentation of insights from the comparison between these integrated models and the role of OpenMI in enabling this comparison.

2. STATING THE PROBLEM

The Pinios River catchment is the Greek pilot basin for the Water Framework Directive. The catchment drains an area of approximately 10,500 km² and the area of interest in this study is the upper part of the river (until the Ali Efenti bridge) where hydrologic, hydraulic, and water quality are modelled to investigate source and impact of pollution (Figure 3). Agriculture is the main source of income for the Thessaly Water District and the Pinios catchment is intensively cultivated with water demanding crops. Most of the pollution in the area arises from non-point sources (including for example livestock, which is the most important source of BOD pollution in the area).

The increased irrigation of the basin has seriously decreased groundwater levels and river flow. Additionally, the fertilizers and pesticides used for the agricultural activities resulted to water quality degradation. Furthermore, untreated industrial waste and municipal
wastewater discharged into Pinios have added to the local water pollution issues. Illegal damping areas are flushed in the river during storm events. The aim of the study was to create a model of the hydrology (rainfall-runoff) and couple it with a river hydraulics model and a river water quality model so that sources and fate of pollutants in the catchment could be modelled as a basis for prioritising interventions and assessing their performance.

![Figure 4. Two alternative integrated model representations for the same case](image)

### 3. METHODOLOGY

To achieve the case study objectives and to investigate the validity of the perceived benefits of multi-model component-based representations, identified in the introduction, two alternative integrated model representations (Figure 4) set up by different research groups were used. The model components (ie. the individual process models) in each case were migrated into OpenMI and linked using the OpenMI configuration editor. The schematizations of the models (reach lengths, inflow nodes, pollution nodes) are shown in Figure 5. It is interesting to notice that Integrated Model 1 was designed to evaluate water quality along the whole length of Pinios river. However, the comparison study between the two integrated models was conducted at selected locations along the first 75 km of the main Pinios channel.

![Figure 5. The two integrated model schematizations](image)
3.1 Integrated Model #1

An integrated 1-D linked scheme was developed using the OpenMI standard to simulate hydrodynamics and water quality in rivers and streams. The integrated scheme consisted of the rainfall runoff NAM module of MIKE 11, the hydrodynamic model RISH–1D and the water quality model RISQ–1D (Figure 7a); the latter two models were developed in NTUA (Stamou and Douka, 2010). The MIKE 11 NAM module is a deterministic, conceptual and lumped rainfall-runoff model. The main inputs for the NAM module include precipitation and potential evaporation for each sub-catchment, as well as measured discharge for the validation procedure (DHI, 2009). RISH-1D is based on Saint–Venant equations, which are discretized using the implicit weighted four-point Priesmann scheme and solved by the iterative method of Newton–Raphson. RISQ–1D consists of mass balance equations for the water quality variables of interest that are discretized using the second order accurate implicit Crank-Nicolson method and solved via the Thomas algorithm. RISH-1D and RISQ-1D were calibrated and validated for the total length of the Pinios River. The available field measurements, used in the computations, had a high level of uncertainty; moreover, there was also significant uncertainty in the estimation of pollution loads. Despite these limitations, the separate and linked model run successfully, simulating flow and water quality at selected locations during low (June 1998) and high (December 1998) flow conditions (Figure 6).

![Figure 6. Calibration and validation of integrated model 1 (June and December 1998)](image)

3.2 Integrated Model #2

In this model the hydraulic and water quality component were substituted with MIKE-11’s hydraulic model and OTIS respectively. This configuration retains the NAM model (which
is part of MIKE11) as the rainfall runoff component (Figure 8). The MIKE11 hydrodynamic module (HD), which represents the core of MIKE11, uses an implicit, finite difference scheme for the computation of unsteady flows in rivers and estuaries. The computational method is based on the numerical solution of the St. Venant equations on the conservation of continuity and momentum. Main inputs for the HD model include cross-sectional information at various locations along the main river network, measured stage or discharge for model calibration, stage-discharge relationship, etc.

MIKE11 includes an advanced graphical user interface in order to facilitate data input and editing. Besides, generating stage and discharge for the points of the computational grid that represent the main model output, MIKE11 calculates additional outputs, including velocity and cross-sectional area. Discharge and cross-sectional area at selected locations are important inputs for the water quality model OTIS that utilised them for the calculation of BOD levels along the main river. The solute transport model OTIS (One-Dimensional Transport with Inflows and Storage) started being developed in the beginning of the 1990’s by USGS (Runkel, 1998). OTIS simulates the fate and transport of water-borne solutes in streams and rivers. Specifically, OTIS calculates the solute concentrations that result from hydrologic transport and chemical transformation when providing the catchment loading. It is a one-dimensional model using the main assumption that pollutant concentration varies only in the longitudinal direction and not with width or depth. The model simulates various hydrologic and chemical processes including advection, dispersion, lateral inflow, transient storage, first-order decay and sorption using the advection-dispersion equation with additional terms to account for all the processes. All equations used within the model are solved numerically using a Crank-Nicolson finite-difference solution for each segment of the river. The model has the ability to simulate the fate and transport of both conservative and non-conservative pollutants (Runkel, 1998). Pollution loads from both point and non-point sources, flows and cross sectional area in the main river represent the main inputs for OTIS. Water quality measurements along the river are necessary for model calibration (Figure 9). At the end of the simulation OTIS generates the solute concentration at various selected locations along the modelled river and in the case of the particular application the concentration of BOD (Figure 10).

![OpenMI Configuration for representation #2](image)

**Figure 8.** OpenMI Configuration for representation #2

![BOD concentration along river](image)

**Figure 9.** Calibration and validation of integrated model #2 (October and December 1993)
It can be seen that the results of the individual models (run in sequence) match well with the results of the integrated model, exchanging data though OpenMI. As before this is an indicator of a successful migration and linking.

It should be noted however, that this is because the links established within this and the previous configuration, are one-directional. In the presence of bi-directional links, where data supplied by the “upstream” model are influenced by data provided by the “downstream” model, it is expected that differences would be observed between individual models (run in sequence) and the integrated model, without this being an implication of faulty migration and linking. On the contrary: provided that sufficiently detailed data sets are available to support model set up and calibration to capture (in some detail) interactions between different processes, the results of the integrated model would be expected to be more realistic, as this is exactly the type of problem that OpenMI enables to tackle.

4. RESULTS AND COMPARISON

The two integrated modelling configurations were developed by separate research teams to answer different water resources management questions and were therefore based on different assumptions and originally run for different time periods. To allow for a meaningful comparison, a common time period was selected (February 1998) to run both configurations. The same pollution loads were assumed for both configurations (for both diffused and point source) and the same BOD decay coefficient and dispersion coefficient were used in both RISQ-1D and OTIS. However, some assumptions remained different:

- The river schematisation used by OTIS does not match exactly the one used by RISH/RISQ. This results in a difference between the exact locations of the point sources which in turn affects BOD distribution along the river.
- Due to a simplifying assumption in OTIS, diffuse pollution loads enter only from the major four tributaries of Pinios with pollution loads from the smaller tributaries required to be added to these. This assumption however has an effect on the distribution of the BOD concentration along the river.
- The cross sections in the two configurations slightly varied in width along the main channel, with an average cross section width of 30 meters for all models. These (+/-3 meters) cross section variations along Pinios river improved the individual model calibrations and the stability of the linked runs. However, flow and stage changes fed in the water quality models (RISQ-1D and OTIS) inevitably varied in space and time.

Figure 11, below shows the results of both configurations at selected locations.
Makropoulos et al. / An integrated, multi-modelling approach for the assessment of water quality...

Figure 11. Comparative results of the two integrated model configurations at two locations

The two configurations were built with slightly different topographic characteristics, observed points, initial conditions and assumptions. Many of these differences were a direct result of individual model characteristics, strengths and limitations and the availability of segmented data sets. Furthermore, the input data discrepancies were enhanced by the different modelling techniques followed by different research groups. These variations are clearly depicted in the comparative modelling results presented for two river nodes (14 and 26) in Figure 11. At both locations, there is a significant lag between the BOD concentration peaks as well as between the BOD concentration valleys. It also appears that configuration #1 overestimates, while configuration #2 underestimates BOD concentrations at every location. This difference increases as we move downstream.

It is suggested that this variation is in effect a measure of the uncertainty related to this multi-model simulation. The visualisation of this (significant) uncertainty may be very important for decision making, including, but not restricted to, the identification of the required level of water treatment for local communities. If water at node 26, for example, was used for recreational purposes, that would necessitate a BOD<4 mg/l. This threshold, however, is in the middle of the uncertainty band produced by the multi-modelling approach – while each integrated model on its own would produce a more definitive (and perhaps misleadingly authoritative) answer.

5. CONCLUSIONS

The OpenMI standard was successfully used to link rainfall-runoff, hydraulic and water quality models, as components of an integrated model for the Pinios River catchment. Two alternative configurations of an integrated model were developed, by two different research groups, by substituting modelling components (ie. in this case whole domain models). Three models (RISH-1D, RISQ-1D, OTIS) developed by different providers (two separate research teams at NTUA and the OTIS developers in the USGS) were successfully migrated to OpenMI. An additional, commercial model (MIKE-11) which was already OpenMI compatible was used, albeit in different roles, in both configurations. The individual models were set up, calibrated and validated in the study area. The individual and OpenMI linked model runs matched well in both configurations, proving that linking models in OpenMI does not compromise the accuracy of model runs.

The comparative analysis of the two configurations illustrated the significant differences between model components and (consequently) modelling results, even when OpenMI is used as the integrating medium and model schemes are set up for the same study area by collaborating modelling teams. The comparison also introduces a series of questions regarding the extent to which alternative modelling configurations (or members of a multi-modelling ensemble) can be “objectively” compared to each other.

Comparisons between models are nothing new, and have been traditionally based on such criteria as computational speed (e.g. Ji et al., 1996), ease of model use (Saloranta et al., 2003), accurate forecasting of analytical solutions (Rossman and Boulos, 1996) or observed values (Habets et al., 2010) to name a few. However, all these metrics, with the possible exception of the last two, are significantly biased by the expertise of the modeller,
who more often than not is the developer of one of the models and the (reluctant) user of the other. The last two comparisons are unfortunately often difficult to perform, in exactly these situations where it is most needed: either in complex problems where analytical solutions are not possible or where field observations are scarce and modelling results are required exactly to substitute them. The ease with which alternative configurations of integrated models can be build using OpenMI however, change this picture, by often putting real end-users (as opposed to model developers) in the position to setup, run and ultimately assess these configurations.

We suggest that although by carefully selecting parameter values and identifying equivalent schematisations, models do become more comparable, this may not be a very productive exercise, as significant difference will inevitably remain. Perhaps a more productive way of working, would be to exploit these different assumptions, schematisations and parameterisations (assuming of course they are all well developed and supported by the evidence (Beck, 2002)) and use the ensemble of results to analyse, assess and visualise the underlying uncertainty (Potempski and Galmarini, 2009).

It could be argued that such an approach is becoming increasingly feasible through the use of OpenMI and coupled with techniques such as Evidence Theory (Bicik et al., 2009) could open new avenues of investigation on issues of epistemic uncertainty that have been lying “under the surface” of much of the modelling endeavour.

ACKNOWLEDGMENTS

The work presented in this paper was undertaken within the OpenMI LIFE project, which was funded by the European Commission under the Life Programme (Contract no: LIFE06 ENV/UK/000409).

REFERENCES


An interdisciplinary framework for the development and analysis of crop models

Sébastien Roux a, Myriam Adam b c d, Jacques Wery e

a INRA, UMR System, Montpellier, France, (rouxs@supagro.inra.fr)
b Plant Production Systems Group, Wageningen University, Wageningen, The Netherlands
c CIRAD, UMR System, Montpellier, France
d Plant Research International, Wageningen University and Research center, Wageningen, The Netherlands
e Montpellier SupAgro, UMR System, Montpellier, France

Abstract: Development and use of crop models has become an interdisciplinary field with the increase on one hand of the biophysical knowledge and the development of numerous dedicated software platforms on the other hand. This trend calls for formalizing the integration of the disciplinary contributions of scientists having different expertises needed in crop modelling: agronomists, mathematicians, software scientists.

We propose an interdisciplinary framework to: (1) describe the model with different expert domain knowledge, (2) identify bridges to help scientists of different domains to work together on the same model, (3) give a methodological basis (engine) for crop model studies. This theoretical framework is thus based on the integration of: (1) the agronomic domain describing the simplified functioning of soil-plant systems and the associated hypotheses; (2) the mathematical domain giving the system of equations and its properties and (3) the software domain defining the software components and their relationships.

We show how these views can be specifically defined for the description and studies on crop models, how they are related to each other, and how this framework integrates some relevant concepts on integrated modelling. We finally illustrate the use of the framework for building a new crop model (a pea crop model) by changing an existing one (a wheat crop model).

Keywords: modelling framework, interdisciplinary framework, crop model, agronomic model, mathematical model, software model.

1 INTRODUCTION

The expression crop models hides an interdisciplinary perspective behind the use of the word model. It is often reduced to the software tool used to simulate the crop biomass production depending on different parameters and input variables: soil, climate, management practices, etc. However when an agronomist analyzes the results of a crop model, he/she starts to question whether the choice of crop processes selected for their specific application corresponds with their expert knowledge. An applied mathematician on the other hand questions how the systems of equations is solved, how the system parameters have been estimated, how the results are sensitive to uncertainty. A computer scientist concentrates on the software design and architecture, and looks for bugs. Each of these different specialists questions the model from their own perspective. Beyond this example of model evaluation, we want to stress that all three disciplines-agronomy, mathematics and software engineering science- have a major role to play when dealing with crop models development.
Today dedicated software platforms (i.e., integrated modelling platforms) make the development of a crop model potentially quicker and easier than in the past, especially when they integrate graphical interfaces and graphical programming functionalities. Some of the most sophisticated tools, such as Simile (Muetzelfeldt and Massheder [2003]) propose a visual modelling environment to specify graphically the model and to reduce the effort required to implement it. This is undoubtedly a progress and a gain in time that enables the modeler to keep more energy for the other aspects of crop modelling process such as the statement of hypotheses and the mathematical model formalization. These aspects remain irreducible when dealing with crop modelling, since the increase of knowledge does not reduce the difficulties and limitations of biological prediction at macroscopic spatial and temporal scales (crop scale and daily time step) with limited calibration and validation data. These issues were already pointed out in Monteith [1996] and Sinclair and Seligman [1996]: it is impossible to identify all possible factors influencing crop performances. This agronomic model validity is therefore tightly linked to the hypotheses stated on local relevant crop processes and on the data available to parameterize and evaluate the model. We add another sense of mathematical validity associated with a crop model. The translation of the selected processes into a mathematical dynamical system is not unique, and the properties of the system once built can play a major role to interpret the simulation results: typically a very unstable system of equations can propagate uncertainties of the input data in such a way that any prediction based on experimentally calibrated parameters can not be reasonably accurate. Finally the software validity relates to code quality and to software testing. The latter ensure that the simulation software runs properly with software components consistently translating the concepts and equations of the agronomic and mathematical models.

These different aspects of model validity are often interlaced for a model developed in a platform. Hence it is more difficult to identify the origin of the wrong results of crop models as they can originate from each of the three domains. This fact stresses the need of a triple domain -agronomy, mathematics and software engineering- disciplinary framework. More generally, we designed an interdisciplinary framework with the purpose to: (1) describe the model with different expert domain knowledge, (2) identify bridges to help scientists of different domains to work together on the same model, (3) give a methodological basis (engine) for crop model studies.

2 AN INTERDISCIPLINARY FRAMEWORK BASED ON A TRIPLE DOMAIN VIEW

We introduce our framework based on views from three disciplinary domains and their respective activities. We also describe the associated links between these views.

2.1 Description of the triple domain view framework (Figure 1)

**Agronomic Modelling.** We start the description by considering the agronomic view of the crop model. We refer to it as the “agronomic conceptual model” as introduced by Wery et al. [2009], Rapidel et al. [2006] and Lamanda et al. [2010] by analogy with the conceptual models related to Information Systems. Its main properties are described in the following. The main outcome of this agronomic modelling process is to formalize a scientific question into a conceptual model with an abstraction level not directly linked to mathematical or software representations. The principles governing its construction are those of the system analysis theory (Rabbinge and De Wit [1989]) and the different steps proposed in Wery et al. [2009] to design it are:

- Formulation of objectives
- Definitions of the limits of the system
- Conceptualization of the system
- Quantification of input relations
This agronomic conceptual model expresses the scientific question to be solved by modelling. It is the initial explicitation and specification step of the model development in the framework. As a disciplinary tool, it can also be used independently of the numerical model development: for instance it can help agronomists to develop an in-field comprehensive analysis of yield components (Rapidel et al. [2006]). It might also be the correct initial level of abstraction for a collaborative model design when agronomists and other scientists such as hydrologists, biologists or economists gather knowledge from their expertise field.

This property explains why its representation is more general than the diagrams used in some visual programming approaches. For instance, it contains the definition of the main basic crop processes (Adam et al. [2010b]) needed for the model application. But it also states explicitly the hypotheses and includes a detailed description of the validity domain of subsystem model and parameters. It clarifies whether the model parameters have to be measured only once or for each new soil-climate-genotype combination. An example of a diagram extracted from a conceptual crop model based on crop processes is given in section 3.

Inside the framework, it plays the double role of product of the initial crop system analysis to be used by agronomists and communication tool between scientists involved in the whole modelling process.

**Mathematical Modelling.** The mathematical part of the triple disciplinary framework comprises the description in the mathematical form of the dynamical system and some associated activities such as dynamical system definition and analysis.

The role of mathematical expertise in the modelling process applied to the crop modelling shall indeed not be underestimated, even if the classical structure of crop models, which are classically composed by a set of explicit difference equations, leads to straightforward resolution of the system by direct time integration. This structure is also characterized by a very large number of parameters coming from the detailed accumulative description of crop processes (for example more than one hundred in the STICS Model (Ruget et al. [2002])). In that sense, despite the simple structure of the dynamical system of equations, the mathematical object resulting from numerous biophysical processes can be complex. The need to work on this complexity with expert knowledge makes necessary the mathematical representation of a crop model inside our framework.
A disciplinary activity related to the mathematical field is the study of model structure to reduce the number of its parameters or more generally to improve its mathematical properties. Another example is the statistical analysis that focuses on the response of the system to variations of its input. This is commonly addressed by sensitivity and uncertainties analysis of a model, and which may require expert skills when the system is large and its resolution time consuming (Wallach et al. [2006]).

**Software Engineering and Modelling.** The third view including a disciplinary model representation and some associated activities deals with the software engineering domain. Here we do not want to enumerate the different possible ways to implement a crop simulator with or without a software platform. The software model of this view (see Figure 1) denotes different objects depending on the simulator implementation: (1) in the most sophisticated integrated tools, it is the graphical model representing the model components as in Simile (Muetzelfeldt and Massheder [2003]), STELLA (Richmond [2001]); (2) in a standalone application coded using an object oriented language it can be the UML class diagram of the software application, and (3) in its most crude form the model is reduced to the source code.

Even if we have chosen not to point out one method rather than another, the need of interdisciplinary communication required by our framework promotes tools having explicit software models underlying the source code. This objective is closely related to the paradigm of Model Driven Architecture, applied to crop modelling in Papajorgji and Pardalos [2006]. Its goal is to separate the platform independent conceptual models of the software components written in an UML-based approach, from the executable code that can be at least partly automatically generated by an engine using the UML model. It is also the same objective of the so called Declarative Modelling approaches implemented in Simile and STELLA : the model structure is stored in an open text file and is thus separated from the rest of the source code needed to run simulations. The benefits of such approaches are recognized in the software engineering field for the gain of reliability, productivity and modularity (Papajorgji and Pardalos [2006]). Moreover it fits with our framework objectives of interdisciplinary exchanges, as we will see in the coming section.

### 2.2 Links between framework domain views (see Figure 2)

The three disciplinary representations of the crop model have to be connected in order to make the framework operational, as suggested by the double sided arrows in figure 2. The detailed formulation of the dynamics of crop processes is made in the mathematical domain view, as well as the formalization of the relations (influence, regulation) that are mentioned in the agronomic conceptual model. The connection between the mathematical model and the software model is made by the numerical methods needed to solve the equations of the mathematical system. Finally the software model is created from the agronomic and the mathematical models, and should be as much as possible a consistent transcription of the agronomic conceptual model into a software architecture.

Like the agronomic conceptual model, abstraction in the software architecture helps the communication between experts from the three disciplines and the sharing of their own model definition: this is necessary to strengthen the global model and to sort out problems specific to each domain when a simulation gives wrong results. Such software models (sometimes called software conceptual models) however shall not be identified with the two others model representations even if there are links as we already mentioned. It is indeed very unlikely that the agronomic expert knowledge can be embedded in the software view of a model. The validity hypotheses attached to the agronomic conceptual model are not building block of the software construction. The same remark holds with the fine mathematical properties of a crop mathematical model : they are unlikely to be better expressed than in the mathematical language.

Going one step further towards the formalization of the relations among the disciplinary domain views leads to the concept of component approach. Classical component approaches are related to software elements, which typically wrap a crop process in a software unit. The latter can be switched in the same software modelling solution and in theory exchanged between different scientists (Donatelli et al. [2009], Wang et al. [2002], Jones et al. [2001], Bergez et al. [2009]. This approach is necessary to help code re-use but often entails a common software platform or simu-
3 AN EXAMPLE OF THE FRAMEWORK APPLICATION

The framework can be used for a wide class of problems on crop models that require interdisciplinary expertise. In the dashed box in Figure 2, we mentioned the general examples of model construction, model evaluation and model integration. In this section we detail the application of the proposed framework on the following question: how can we build a new crop model from an existing one? More precisely, this example takes a wheat simulation model and aims at deducing a pea growth simulation model (Adam et al. [2010b]).

The framework gives a straightforward procedure based on the successive use of the three domain views to get an implementation of a pea growth simulator from an existing wheat model. This is a one loop process because at this level we do not consider the data validation of the new simulator, which would require an iterative procedure between the different models as suggested in the central loop in Figure 2. The framework based procedure that we have successfully used (Adam et al. [2010b]) can thus be summarized as follows:

- at the agronomic level: analyse the conceptual model with expert knowledge to identify which basic crop processes and relations must be added or removed and modified through for instance the change of the species specific parameters.
• at the mathematical level: define or update the mathematical definition of the processes and the new relations

• at the software level: propagate these changes into the software components through the software platform (CROSPAL (Adam et al. [2010a]) in this example)

This process is illustrated in figure 3, with the comparison between the diagrams of two conceptual crop models and the required changes coming from the previous analysis. The agronomic analysis of the basic crop processes to be modelled leads to the addition of the new component Nitrogen Fixation. This change implies adding a new module in the software model. On the other hand, the change in the Leaf Development process only requires to change the mathematical formulation in an existing software module.

![Diagram showing the changes needed to develop a pea crop simulator from a wheat crop simulator.](image)

Figure 3: Illustration of the changes needed to develop a pea crop simulator from a wheat crop simulator. Using the framework, the changes are firstly analysed on the conceptual model. The associated modification are shown on the right side with different colours depending on the models to be mostly affected: light grey for the mathematical model, dark grey for both the mathematical and software models.

4 CONCLUSION AND DISCUSSION

We have introduced a framework for the interdisciplinary development and analysis of crop models. We have shown how to define the three main domain views and what are the related disciplinary activities. We have explained the relations between the different roles and shown how
the representations are connected. How to go one step further and formalize semantically these relations is still an open question in our framework. To this aim, evaluating the possibilities to establish formal links between the disciplinary representations through ontologies as in Villa et al. [2009] seems promising. This framework does not rely on an unique integrated software tool for several reasons. Promoting a software tool would unbalance the equilibrium toward the software aspects, and be incompatible with the necessary freedom for experts to use the most adapted tool required by their activity. For instance, applied mathematicians will be keen to use popular scientific software like Matlab or R. Moreover, such integrated tools would necessarily limit the framework to a concrete definition. Conversely an original benefit of our framework is its ability, as a theoretical tool, to suggest methodologies to tackle new problems on a crop model.

The complete assessment of the operational value of our framework is an ambitious task that will probably reveal some limitations. Nevertheless, as we learned from the presented example (pea crop model) and from other case studies currently under progress, the triple disciplinary framework already appears to be a stimulating tool for crop models formalization between agronomists and a facilitation tool to initiate collaborations between agronomists, applied mathematicians and computer scientists. This is why we believe that separating the three domains representations can ease international collaboration on disciplinary questions (for instance exchanges between agronomists on concepts and processes that can be made on conceptual models) while allowing an improved interdisciplinary collaboration around crop models based on the same software platform or environment such as Simile (Muetzelfeldt and Massheder [2003]) APSIM (Keating et al. [2003]), SEAMLESS (Van Ittersum et al. [2008]), RECORD (Bergez et al. [2009]).

REFERENCES


Adam, M., P. Leffelaar, F. Ewert, H. Van Keulen, and J. Wery. Developing new crop models within a flexible crop modelling framework: use of crop physiological knowledge to change a wheat model to a pea model. In Agro2010, the international scientific week around agronomy (submitted), Montpellier, 2010b.


An open-source model platform for water management that links models to a generic user-interface and data-manager

Julien J. Harou1, Didrik Pinte2, Amaury Tilmant3, David E. Rosenberg4, David E. Rheinheimer5, Kristiana Hansen6, Patrick M. Reed7, Arnaud Reynaud8, Josue Medellin-Azuara5, Manuel Pulido-Velazquez9, Evgenii Matrosov1, Silvia Padula1, Tingju Zhu10

1 University College London (UK) (j.harou@ucl.ac.uk), 2 Enthought (BELGIUM), 3 IHE Delft (Holland) & ETH (Switzerland), 4 Utah State University (USA), 5 University of California, Davis (USA), 6 University of Wyoming (USA), 7 Pennsylvania State University (USA), 8 Toulouse School of Economics (INRA-LENA) (France), 9 Polytechnic University of Valencia (Spain), 10 International Food Policy Institute (USA)

Abstract: Water management models that lack a user interface, data management, reporting and visualization capabilities are less likely to make an impact on real world water systems. Model platforms can improve this situation; they allow researchers and model developers to link model codes to a generic user interface and data management system. We present an open-source model platform for network-based models called HydroPlatform. The platform links users and their data to water management models through generic import/export functions or custom add-ins. This paper focuses on export functionality, and specifically on a link to the Generalized Algebraic Modeling System (GAMS). HydroPlatform is currently being used for several large-scale modeling projects and add-ins have been built for two generic water management models. Benefits, limitations and future developments of the software are discussed.

Keywords: water management models, modeling frameworks, decision support systems (DSS), simulation and optimization models, model platforms, open-source.

1. INTRODUCTION

Although researchers propose and apply many innovative models to help manage water resources, only a few models are regularly used in practice. ‘Real-world’ modeling studies are data intensive and have time constraints. In addition to the software related to modeling water movement and allocations, real-world water studies require ways to efficiently create, manage and visualize model inputs, scenarios and results. In this paper we call this ‘user-software’, i.e. software that helps users apply a mathematical model to a real case study. To meet this need, practitioners increasingly use regularly–updated decision support system (DSS) software for water resources studies rather than models developed in research institutions which typically lack user interfaces, data management, reporting and visualization capabilities. State-of-the-art research models are less likely make a significant contribution to water management globally unless significant resources are applied to building user-friendly model interface.

HydroPlatform provides water management model developers a generic user-interface and data management system. Any model with a network structure (node and links) can be linked to the platform regardless of the language or software used to create the model. HydroPlatform allows model developers to quickly and inexpensively link network models to a modular and flexible open-source platform.

To better explain the context for why such a linkage is useful, we briefly describe the structure and objectives of water management models. We then list existing software
options, including model platforms, a term introduced here. We describe the structure and organization of the HydroPlatform software and show how the platform can be used to connect to commercial models and modeling systems such as the Generalized Algebraic Modeling system (GAMS). Finally, limitations, project structure and future developments are briefly discussed.

1.1 Water resource management models

Since Maass et al. (1962), water resource systems have been modeled as networks to represent hydrology, infrastructure (storage, conveyance, treatment, etc.) and diverse water demands. The network structure is straightforward, flexible, parsimonious and works for both simulation and optimization models (Letcher et al. 2007).

Management models use simulation or optimization methods to model the use, flow and storage of water throughout the system under different scenarios, infrastructures or policies. Simulation models predict system performance under a determined set of conditions and rules. Optimization models search through a range of alternatives to find which system solution best satisfies mathematically stated objectives taking into account various physical, environmental, social, political and institutional constraints. The purpose of management models is to support planning and management decisions for system design, expansion and operation.

2. SOFTWARE APPROACHES

There are many software options currently available to run water management models. These options include dedicated water resource decision support systems (DSS), modeling frameworks, commercial modeling systems, and stand-alone software. The boundaries between these approaches sometimes overlap. We describe each type before introducing the model platform approach.

2.1 Water resource decision support systems (DSS)

A DSS is a user-interface built around one or more custom models. The user-interface manages a particular model or modeling process and provides a system to enter, store, and retrieve model data, run the model(s), and view results. Disciplinary water resource DSS have been historically the most popular choice for water resources management studies; some systems have been in use for over 2 decades. Established water resource management DSS include WATHNET (Kuczera 1992), AQUATOOL (Andreu et al. 1996), OASIS (Randall et al. 1997), MISER (Fowler et al. 1999), MODSIM (Labadie and Baldo 2000), RIVERWARE (Zagona et al. 2001), MIKE BASIN (Jha and Das Gupta 2003), CALSIM (Draper et al. 2004), REALM (Perera et al. 2005), WEAP (Yates et al. 2005), WRAP (Wurbs 2005), HEC-ResSim (USACE 2007), WaterWare (Cetinkaya et al. 2008), RIBASIM (WL Deflt Hydraulics 2004), AQUATOR (Oxford Scientific Software 2008), and WARGI (Sechi and Sulis 2009).

Advantages of this approach are its conceptual simplicity, the ability to accommodate popular legacy codes and full customization of the user experience around a specific model or set of models. However, this approach forces developers to build, maintain, and extend (as new models are included) the user interface to accommodate existing and new models. Only a handful of DSS exist because software development and maintenance costs are high. As more models or functionality are included to allow integrated assessment, it becomes more effective to use the component-based paradigm (next section).

2.2 Modeling and Component-based frameworks

Modeling frameworks (Argent 2004) or application hosting systems (Rizzoli and Argent 2006) enable building environmental management systems by assembling different components. Such frameworks are designed to integrate models, where modular
components can be assembled to form a custom application. A component-based approach is ideal for building interdisciplinary models as it provides a structured way for different models to interact.

Rizzoli and Argent (2006) and Argent et al. (2006) review modeling frameworks that can be used to build integrated water resource management tools. Examples include the Interactive Component Modeling System (ICMS) (Argent 2004), TIME, the Integrated Modeling Toolkit, TwoLe (Soncini-Sessa et al. 1999) and others.

One potential disadvantage of using comprehensive modeling platforms is they may require that model components be substantially rebuilt to be compatible with the framework. This task has been simplified by the open modeling interface (OpenMI) that allows building wrappers around legacy codes allowing them to be connected to other models (Gregersen et al. 2007). If modelers are willing to adopt a modeling framework and learn the skills required to modify their existing tools for compatibility with a new platform, the rewards in terms of usability and future integration of other models can be large.

2.3 Commercial modeling systems

When more model formulation customization is required than what is offered by existing water management DSS, another option is to use generic commercial modeling systems. Both simulation and optimization variants exist. Optimization modeling systems integrate model data, formulation, solution and results definition. These systems separate generically posed mathematical model formulation and the data related to a particular application or “instance” of the model. The ability to flexibly define the data sets for which model equations are applied enable relatively quick development of models. Examples include GAMS, AMPL, LINDO and AIMMS, each of which link model formulations written in high level languages to commercial optimization algorithms called solvers. These systems are flexible, transparent, self-documenting, provide simple links between model formulation and solver solution, and are widely used by researchers and systems analysis practitioners to build custom models. Simulation modeling systems (also referred to as ‘systems dynamics’ software) solve user-defined simulation models with a commercial solver engine. Examples of commercial systems with simulation capability include STELLA, Goldsim, Powersim, AnyLogic, MATLAB Simulink, and others. Some modeling systems allow building custom graphical user interfaces (GUI), while others (e.g. GAMS) do not.

2.4 Stand-alone programs

Traditional stand-alone programs were intended for well conscribed analyses with pre-defined inputs and outputs. Alternatively the term ‘monolithic’ is used by Rizzoli and Argent (2006) to underline the static and less flexible nature of these programs, particularly when it comes to linking them to other programs. Many legacy stand-alone programs are prevalent and effective due to their simplicity and familiarity. Some common examples include MODFLOW, HSPF, SWAT, HEC-RAS, etc.... An effective software solution should be able to accommodate such programs.

2.5 Model Platforms

Models built using commercial modeling systems, stand-alone programs and modeling frameworks often lack easy-to-use interfaces. Model platforms manage, store and display model input and output data. Unlike a modeling system, model platforms do not support model formulation and solution; models are external and built independently (Figure 1). The goal is a software platform to enter, manage and visualize model inputs and results and exchange these with models through export/import functions or add-ins. The term ‘model’ rather than ‘modeling’ platform underscores that model coding does not occur within the platform. Rather the platform hosts existing, external models. The external model no longer requires its own user interface and data management software. If tighter coupling between the user-interface and model is preferred, platform ‘add-ins’ can provide seamless integration (the user need not know that several applications are running).
Figure 1: Model platform concept: user-software (interface and database) connect to external model(s) via generic functions or add-ins.

The challenge of efficiently building effective but flexible user interfaces and data management systems for existing water management models is not new. Existing solutions include creating a custom interface from scratch for a specific model or using a modeling system that already has user-interface capabilities. An intermediate solution is using a Geographic Information System (GIS) to store and visualize data that is then passed on to a specific model. These approaches and others are reviewed by Argent (2004). All approaches have advantages and limitations but they all require significant investment by modelers beyond building the model itself. Most water management DSS create a specific user interface to link to a particular model. This technique can be effective but costly. Below we describe HydroPlatform, a model platform software that allows modelers to link a generic user interface to various models thereby decreasing development costs.

3 HYDROPLATFORM

HydroPlatform is an open-source, generic model platform for water management network-based models. The model platform manages and displays spatial and temporal model data. Models are run either independently, uncoupled from the platform’s user-interface and database, or from within the platform using add-ins.

Project goals:
- Make research models easier to use for real problems
- Help modelers focus on model development, not data management

Software objectives:
- Allow modelers to create, manage and visualize network model data efficiently
- Provide an intuitive graphical user interface (GUI) using object-oriented code
- Maintain generality and independence from any single model or modeling system
- Use open-source components for maintainability, transparency and lower costs

HydroPlatform is an initiative by and for the water management community and its name has been selected for that audience. However, the tool itself is generic and can serve as an interface for any mathematical network model (transport, energy, trade, etc.).

3.1 Capabilities

The completed first phase of HydroPlatform development focused on input, storage, display and export of model data. Node-link networks are built by dragging and dropping node and link objects over a map. The map can be a georeferenced raster or vector data or a non-georeferenced background image for spatial reference. Custom network objects (nodes or links) can be built by defining and adding custom data fields to the object. These fields include parameters, seasonal parameters, time-series, and tables of any dimension.
A database made for one model can be reused for a different model. Imagine an initial database was prepared for Model A but now needs to be used for a similar but different Model B. Model B possesses extra parameters and functions to represent desalination plants. In this case the user would save the Model A database as Model B (using 'Save As'), then go to the Object Type Editor and define a new Object Type called for example ‘DesalNode’. Then, new DesalNodes can be added to the existing network, populated with data, and the new dataset can be exported to Model B.

Model data is stored in a database residing either on the analyst’s local computer or on a network. This flexibility enables collaborative model building over the internet so that modelers in different locations can work concurrently on a shared system.

Export Functions allow networks defined and data entered in HydroPlatform to be easily exported to external models. Exported data also includes connectivity matrices showing how network nodes in a complex water system are linked. These matrices are automatically generated when the export function is run. External models can be run on any system such as spreadsheets, optimization modeling systems (e.g. GAMS, AMPL, AIMMS, etc.), or custom applications built using programming languages (Fortran, C++, MATLAB, etc.).

Extensions or ‘add-ins’ allow custom export functions (e.g. model-specific input data export functions) or to ‘host’ external models as modular add-ins. Add-ins enable tighter coupling between the platform and model if desired. Decision Support System features like special interfaces for managing model use or processing results can be developed as add-ins. This paper does not describe the use of add-ins but focuses on the loose coupling link which uses of export/import functions.

3.2 Projects, Networks, Scenarios, Object Types and Objects

HydroPlatform’s organizational structure is based on Projects, Networks, Scenarios, Object Types and Objects. Upper-case letters denote words with specific meaning in HydroPlatform. A Project has characteristics such as modeled time horizon, time-step, preferred units, geographic projection, and set of Object Types. A Network is a collection of node Objects connected in a certain way by link Objects (topology). A Project may have more than one Network (e.g. hundreds). Scenarios share a Network but have different data values, i.e. one or more parameters, tables or time-series for one or more node or link Objects are unique. A Network may have any number of Scenarios. When HydroPlatform exports data to an external model, it exports the data content of a particular Scenario.

Each Project has a group of Object Types available for use in that project; each Network is built of individual Objects. In object-oriented programming parlance Object Types are classes whereas Objects are instances of Object Type classes. For example in HydroPlatform ‘Reservoir’ is an Object Type, whereas ‘Reservoir A’ and ‘Reservoir b’ are Objects, i.e. ‘instances’ of the ‘Reservoir’ Object Type (class). The instances have the same data structure (data fields) but will differ from each other in their location in the network and possibly the data values in one or more fields. Object Types exist in two main categories (‘superclasses’): node (a point) or link (connection between 2 nodes).

Custom definition of Object Types with an easy user-interface is HydroPlatform’s most essential contribution to modeling networks. Custom Object Types allows the user to define and HydroPlatform to store and manage data for any network model. Custom Object Types can include any number or grouping of data fields. Data fields can be either parameters, tables, time-series or node references.

Figure 2 shows a simple network implemented in HydroPlatform. The network linking several Objects is drawn over a satellite image. Eight Node Object Types exist in this Project (Figure 2). When new Object Types are deleted or created, they (dis)appear from the Object Type toolbar.
3.3 Software implementation

HydroPlatform links an open-source interactive geographic data viewer called Thuban (extensible, multi-platform, multi-lingual) with an open-source database. It is programmed entirely in Python (an object-oriented, multi-platform, open-source scripting language). Data storage is implemented using proven open-source components. It uses SQLAlchemy, the Python SQL toolkit and Object Relational Mapper, and thus supports most database systems such as MySQL, PostgreSQL, SQLlite, …). HydroPlatform uses a set of powerful Python libraries such as Chaco, shapely, networkx, Enthought Traits, etc. for data management to ensure state-of-the-art data manipulation, organization and visualization.

HydroPlatform is free and open-source available under a General Public License (GPL). Add-ins must also be GPL but external models can use any licensing scheme.

3.4 Working with a modeling system

Here we describe how HydroPlatform exports model data to GAMS, one of several modeling systems. GAMS consists of a language compiler and numerous open-source and commercial optimization algorithms (‘solvers’). The user enters model equations defined over specified indices and data, then GAMS interprets the code and converts the model to matrix-form for solution by a user-selected solver.

A GAMS model consists of 4 ‘classes’. A class in object oriented parlance is an abstract definition of a group of things with common characteristics. For example, the Object Types described in HydroPlatform (e.g. Desal nodes) could be considered a class. The GAMS classes are: SETS (the indices over which the model is defined, e.g. for a daily model there will be a set $t$ of all days or a network will have a set of all nodes $i$), SCALARS (fixed constants), PARAMATERS (vector of fixed constants for a set), TABLES (matrices of fixed constants for multiple sets), VARIABLES (decision and state variables, that are unknown and whose values are determined by the solver), and EQUATIONS (mathematical expressions using sets, data and variables).

These classes can have further nested subtypes. For example, the SET object type could include the index $t$ for time and the index $i$ for nodes, both of which are themselves sets.
These classes can have further subclasses, e.g. \textit{desal(i)} denotes the subset of nodes \(i\) that are desalination nodes. Once subtypes are defined, the instances of subtypes needed for the model can be declared. For example, instances of sub-type \(t\) could be the years 1990 to 2000, or for the \textit{desal(i)} subclass, the instances of subtype \(i\) could be desalination plants A, B and C.

In a similar way, the SCALAR, PARAMETER, and TABLE object types can have instances declared for each data item the model requires. For example if an instance of node subclass \textit{junction(i)} named ‘Junction A’ requires an inflow time-series, then Junction A will be associated with a table instance that accepts time-series data.

\textsc{variable} class instances are variable declarations. Variables are defined using previously defined sets. For example if the model has a storage variable ‘\textit{Stor}’ that represents the storage at node \(i\) during time period \(t\), it could be defined using the \(i\) and \(t\) sets as indices, e.g. \(\textit{Stor}(i,t)\).

The object-oriented structure of GAMS easily accommodates water resource network models, for example flow along a network link is defined as \(Q(i,j,t)\), the flow from node \(i\) to node \(j\) during time-step \(t\). A network connectivity matrix summarizes the network topology (i.e. which nodes connect to each other) for use in model equations.

HydroPlatform organizes and exports all relevant data in text files or Excel sheets readable by GAMS. The HydroPlatform export function turns all (i) node and link Objects, (ii) the data fields associated with each object, (iii) data for each instance of the objects, and (iv) the network connectivity matrix into forms available for use by GAMS: namely (i) SETS, (ii) PARAMETERS and TABLES indexed by those SETS, (iii) data values for each indexed element in a PARAMETER and TABLE, and (iv) a network connectivity matrix (\(i \times i\)) who’s columns and rows are the SET of all model network nodes. A connectivity matrix value of 1 indicates the associated nodes are linked; 0 indicates other-wise.

In summary, HydroPlatform allows users to build a database customized for the data needs of any network model. The custom HydroPlatform database content is then exported to GAMS using generic export functions where it is read-in, linked to a particular model, and run. The HydroPlatform GAMS export function is a good example of how HydroPlatform works with modeling systems in particular and external programs more generally.

4. **BENEFITS AND LIMITATIONS**

There are benefits and limitations of the separation of models and interface in model platforms. One benefit of loose coupling is that the modeling and data management software can use different licenses. For example a commercial model (proprietary license) vendor can use HydroPlatform (open-source license) to commercialize a new model or prototype new model formulations. Generally modularity facilitates software project management; model and data manager can evolve separately.

Disadvantages include a potentially less smooth user-experience than with tightly integrated DSS. If export functions are used, HydroPlatform and the external model must be used separately. With add-ins, the model can run from within HydroPlatform, but this setup requires python programming skills. Previously implemented add-ins can serve as templates. For models that are already tightly integrated into an existing proprietary DSS, options for linking to HydroPlatform will be limited unless the model can be isolated.

5. **PROJECT STRUCTURE, CURRENT STATUS AND FUTURE**

The project is an open-source, international collaboration among numerous researchers and practitioners. We are developing and using HydroPlatform as the single user interface and data management system for our varied and diverse set of existing water management models and models under development.
A working version is available for early adopters. The software download will be distributed through www.hydroplatform.org. The project website contains documentation and will eventually work like an app store, showcasing the open-source, freeware, or proprietary water management models that link to HydroPlatform. The project is managed with an online subversion (SVN) repository that stores the wiki, code and bug reports.

The completed phase 1 focused on building network data sets and exporting them to external models using export functions or add-ins. Two generic export functions exist, exporting to text files or to a single spreadsheet; both formats can be read by GAMS. Several GAMS models currently under development have been linked to HydroPlatform including a water supply infrastructure capacity expansion model (South England, UK), a hydro-ecological allocation model for a bird refuge (Utah, USA), a hydro-ecological model investigating hydropower impacts under climate change (California, USA) and a hydro-economic model studying water trading (California, USA). Two add-ins have been completed, generating input files for IRAS-2010, a generalized water resource simulation model (Loucks et al. 1995; Loucks 2002; Matrosov 2009), and the AquaPlan hydro-economic optimization model (Tilmant et al. 2008).

Phase 2 will focus on two tasks: 1) importing model results back into HydroPlatform for storage, analysis and display and 2) upgrading data structure to accommodate stochastic data (i.e. ensembles of inputs and outputs including parameters, tables or time-series, performance measures, or resulting time-series). Phase 2 will also introduce Scenarios; the current proto-type uses only Projects and Networks.

6. CONCLUSIONS

A model platform allows users to input, visualize and manage network model input data and results. Users can link their model to a generic user-interface and data management platform and easily apply a model to real systems. This setup contrasts with other prevalent water management software setups such as DSS, modeling systems, or stand-alone programs that typically either provide limited user-interface capabilities or have sophisticated user interfaces but force use of the built-in model.

We have developed a generic model platform called HydroPlatform as an open-source, international collaborative effort. HydroPlatform can be used with any network-based model as formulating and solving models is done external to and independent of HydroPlatform. HydroPlatform uses standard data export functions or customized add-ins to send data entered in the platform to the external model. By helping to input, visualize, and manage network model data, HydroPlatform allows model developers to focus their time and effort on model rather than software development.

References


DeltaShell - an open modelling environment

Gennadii Donchyts¹, Bert Jagers²
¹ Deltas, Delft, The Netherlands gennadii.donchyts@deltares.nl
² Deltas, Delft, The Netherlands bert.jagers@deltares.nl

Abstract: Over the last two years an open modelling environment has been developed by the Dutch research institute Deltares. Main goal of the development was to create a solid software platform for running 1D, 2D and 3D environmental models. The system allows creating, storing and visualizing different types of environmental data, like time series, networks, GIS features, and other data types defined in the spatio-temporal domain. The architecture of the system, due to its modular and open design, can also support a wide range of other applications such as GIS data management, management and analysis of the hydrological data. The present version of the system makes extensive use of the open-source libraries such as SharpMap, GeoAPI, NetCDF, NHibernate, SQLite, PostSharp; some of those libraries were significantly extended and will be made freely available to the community.

Keywords: integrated modelling; environmental modelling; software framework; domain-driven design.

1. INTRODUCTION

Integrated environmental modelling involving multiple models becomes more and more popular. Multiple efforts have been made trying to develop frameworks and standards allowing linking various components provided by the different vendors and allowing them to exchange data between each other (OpenMI, Gregersen [2007], ESMF, Collins [2005]). An integrated modelling environment is another area of growing interest Ames [2009], Donchyts [2008], Kralisch [2005], Rahman [2005]. However in most cases such frameworks mainly target the integration aspects and rarely consider modelling the whole application domain, including all aspects associated with it. We believe that in order to create a good integrated modelling system it is necessary to perform a detailed object-oriented analysis of all domains associated with the system. Such domains in our case are:

- Model specific domains, e.g. hydrology, hydraulics, water quality, morphology
- Specialized libraries required to describe all mathematical, physical and other aspects of the data types used by the models, for example units of measure, model variables defined in a discrete way, integration and interpolation, etc.
- GIS domain, containing such OGC standards as geometry, feature, coverage specifications
- Data storage using high-quality, scalable and fast technologies, this may include relational databases and/or multi-dimensional file formats such as NetCDF, HDF.
- Graphical user interface including all required components and all related usability issues.
- Modern computer science software development techniques such as domain-driven design, aspect-oriented programming, use of Inversion of Control / DI principles, continuous integration, test-driven development, etc.

After application of the object-oriented analysis Booch [1991] and domain-driven design Evans [2003] methods to the above listed application domains a new modelling framework
and modelling environment called DeltaShell was developed. A few examples of the graphical user interface of the DeltaShell are shown on the two figures below. Figure 2 shows integration of the 1D river flow model SOBEK being developed in Deltares. Figure 1 demonstrates OpenMI plug-in, which allows use of OpenMI-compliant components in the DeltaShell.

![Figure 1](image1.png) **Figure 1** 1D water flow model integrated into DeltaShell (Po River, Italy)

![Figure 2](image2.png) **Figure 2** DeltaShell with a prototype of the OpenMI integration, hosting HEC-RAS and MODFLOW models

In the next chapters we will present some of the design solutions created during the development of the DeltaShell modelling environment. In the first chapter an overview of the DeltaShell architecture is presented. The following chapters focus on a number of more specialized topics such as possibilities to extend DeltaShell framework with plug-ins for new graphical user interface components, data types and models.
2. ARCHITECTURE

The architecture of the DeltaShell is presented on Figure 3. As can be seen from the figure, DeltaShell consists of a set of common class libraries, mainly representing different domains and software standards such as OGC-based geometries and features, hydrologic network objects, mathematical and physics libraries and general application framework class libraries (Core, Gui). These general framework class libraries define on a very high level how a typical console or graphical user interface application is constructed. They also define how different parts of the environment can be extended by means of plugins. This includes both non-gui plugins providing new data types or computational models to the system, but also graphical user interface components used to visualize different data.

![Figure 3 Overview of the DeltaShell architecture](image)

On the next page the main components of the DeltaShell framework are presented (see Figure 4 and Figure 5). The red rectangles highlight parts which need to be implemented by the developer when developing a new plug-in. The following parts of the modelling framework can be extended by the third-party plug-ins:

- Data types
- 0D, 1D, 2D and 3D models of any type
- Data import and export components
- Graphical user interface components such as data editors, tool windows, menu items, toolbar items.

After a plug-in is implemented and deployed, all its custom menus and toolbars may be configured using a plug-in XML file.

DeltaShell uses a project class as single container of all data. The project is a working document of the user and is organized in a hierarchical way. A project may contain data items such as time series, raster grids, hydraulic networks, but also more general items such as models, maps, charts. Items provided by third-party plugins should also usually be stored in the project. All data items contained in the projects are by-default wrapped with a meta-class called DataItem (similar to ExchangeItem in OpenMI but much more generic). This is required in order to allow linking of the data used in the project but also to provide generic way to store data contained in the project.

Figure 4 Core Framework UML class diagram

Figure 5 Gui Framework UML class diagram
3. **CUSTOM DATA TYPES**

In order to extend DeltaShell with a new data type it is required that developer at least implements an interface called IDataProvider. DeltaShell searches all implementations of that interface in all plug-ins during start-up and exposes all data types provided by them to the user. Data types exposed in such a way can be any custom entity defined by the user such as time series, rasters, tables, etc. DeltaShell also provides a set of default data types which can be shared between plug-ins. After data types become available in the system, they can be used in the project together with all other data types.

4. **MODEL INTEGRATION**

Very often models are implemented using programming languages such as FORTRAN, C++. DeltaShell allows integration of the external models on a lowest level, allowing exchange of data between model engine and modelling environment practically on every time step. The main advantage of this approach is that all model data can be edited, visualized and stored by the system within the system using generic components available as a part of the system or provided by developers of the plug-ins.

![Application programming interface of the flow model engine](image)

**Figure 6** Application programming interface of the flow model engine

An example of a typical model integrated into the DeltaShell is shown on Figure 6. Principles to integrated models into DeltaShell are similar to those required by OpenMI software standard. So for those models which are already OpenMI-compliant an implementation of the DeltaShell plug-in would be a relatively easy procedure. The main difference in approaches between OpenMI and DeltaShell regarding model integration is that the latter requires that all input and output data items used by the model have to be exposed using either standard DeltaShell data structures or custom data structures (as described in the previous chapter).

5. **SCIENTIFIC DATA MANAGEMENT**

DeltaShell provides a class library used to define model data variables. The main interfaces used in the class library are shown on Figure 7. This class library is similar to the Common Data Model (CDM) developed by UCAR which allows defining and storing discrete variables defined using multi-dimensional data structures. However the present class library also models relations between different variables.
Figure 7 Functions and variables

This class library practically implements a vector function of a multiple arguments and multiple components. As can be seen on figure, a function is defined as an entity composed of its argument and component variables. At the same time a variable is defined as an entity derived from the function. This definition makes it look very similar to the mathematical definition of the vector function. This class library is used in many places of the system to define model variables used by the models. For example the following functions are used by a 1D water flow model:

- Time series such as: \( Q = Q(t), y = y(t) \)
- Rating curves: \( Q = Q(H) \)
- Model initial conditions: \( Q = Q(s) \), where \( s = (\text{branch, offset}) \) - location on a network
- Model results: \( Q = Q(s, t) \)

The next figure demonstrates how these classes can be used in order to define a two-dimensional function \( f = f(x_1, x_2) \).

```java
// construct f(x1, x2) and set values to 10.0 for x1 = 1.0 and x2 = 2.0
[Test]
public void SetValues()
{
    Variables<float> f = new Variables<float>("f");
    Variables<float> x1 = new Variables<float>("x1");
    Variables<float> x2 = new Variables<float>("x2");

    Function function = new Function("OneComponentTwoArguments Test");
    function.Components.Add(f);
    function.Arguments.Add(x1);
    function.Arguments.Add(x2);

    function.SetValues(new float[] { 10.0f },
                       new VariablesValueFilter(x1, 1.0f),
                       new VariablesValueFilter(x2, 2.0f));
}
```

Figure 8 Defining 2d function using Functions class library

In order to store values of the functions an IFunctionStore interface is used. Currently we support the following implementations:

- MemoryFunctionStore
- NetCDFFunctionStore
- GdalFunctionStore

The first one simply keeps a set of multi-dimensional array in the memory for every argument and component variable of the function which is being used. The second one stores functions in the NetCDF files, wrapping UCAR Java implementation converted to .NET on a byte-code level. The third implementation is used to access raster data stored in GDAL file formats. Additionally to the simple data structures, the IFunction is also used as a basis for our custom implementation of coverages in the GIS class library. Currently the following coverages are supported (extending IFunction):

- ICoverage
- IRegularGridCoverage
- IDiscreteCurveCoverage
- IDiscreteGridPointCoverage
6. HYBRID DATA STORE

DeltaShell uses a mixed SQLite + NetCDF file format in order to store its project on the file system. Based on the type of data it decides where to store values of the objects used in the project. For example, most of the objects are stored in the relational database using NHibernate object-relational mapping library. But some of the data types (mainly those based on IFunction) are stored in NetCDF and only meta-information is stored in the relational database. In this case NetCDF file names always remain in sync with the corresponding records in the database. All storage logic is completely encapsulated in the data access plug-in, so a developer does not need to know anything about storage method. But he/she is required to provide a database mapping XML file once the plug-in introduces a new data type.

![Figure 9 Project states and hybrid data store of the DeltaShell](image)

7. GIS LIBRARIES

With increasing availability of detailed geospatial information, GIS becoming a crucial part of any integrated modelling system. In some cases it is used only as a data visualization tool, but it’s increasingly also used to generate and store model schematizations used by numerical models, including all topological relations between different features of that schematization, for example ArcHydro, Maidment [2002]. However, ArcHydro is defined using ArcGIS as a platform which makes it hard to reuse directly as a part of an integrated modelling system.

The goal of the DeltaShell development in relation to modelling of the HydroNetwork was to provide a lightweight class library which is based only the open-source GIS class libraries and which can be reused by third-party developers working in object-oriented programming languages like C#/Java/C++.

DeltaShell is based on the following open-source GIS libraries:

- GeoAPI.NET – definition of OGC simple geometry specifications
- NetTopologySuite – implementation of OGC simple geometry specifications
- Proj.NET – coordinate system transformation library
- SharpMap – mapping library allowing to work with vector and raster layers
While these libraries were sufficient to implement a lightweight GIS subsystem, there were lacking some very important GIS concepts like Feature, Coverage, see OGC Abstract Specifications\(^1\) for details.

As a part of DeltaShell development these libraries were significantly reworked and extended with the following features:

- **IFeature** – OGC simple feature specifications
- **Coverages** – OGC-like implementation, based in Function class discussed earlier
  - Regular grid coverages
  - Discrete curve coverages used to define data on 1D network
- **Network library**, used as a basis of the HydroNetwork
- **Mapping functionality**
  - Advanced map control supporting parallel layer rendering
  - Geometry and network editing functionality supporting snapping, topologies
  - Vector and raster layer symbology editors

8. CONCLUSIONS AND FINAL REMARKS

Main components and a few examples of the DeltaShell were presented. A very brief description of the major components was given. The goal of the authors was not to discuss every component of DeltaShell in detail but only to give a brief overview of the main functionality available there.

DeltaShell is currently used at Deltares as the basis for a new generation of graphical model user interfaces, and it is provided upon request as an open environment to others.

REFERENCES


Kralisch, S., P. Krause, O. David, Using the object modeling system for hydrological model development and application, *Advances in Geosciences*, 4, p. 75-81, 2005.


---

Abstract: The world-wide trend towards a growing urbanization mainly affects suburban areas. These areas are subject to rapid modifications such as an increase of impervious areas, concentration of runoff in sewer systems, river regulations, but also a decline of agricultural areas causing a forest increase. These changes have an impact on the local hydrology and can induce floods, pollution or decrease of groundwater resource. Distributed hydrological models are useful tools for water management in these areas. They can simulate floods or the impact of land use scenarios on the water balance. This paper presents the PUMMA model (Peri-Urban Model for landscape MAnagement), dedicated to the hydrology of suburban catchments. The model is built within the LIQUID modelling framework. LIQUID facilitates the development of PUMMA by providing a set of modules for different hydrological processes, templates for easy development of new modules and module coupling mechanisms. PUMMA follows an object-oriented approach. The landscape is discretized into cadastral units in urban areas and irregular hydro-landscapes, resulting from intersection of land use, geology, soil and sub-basin maps in rural areas. Each model unit represents one implementation of a module. The considered hydrological processes are infiltration in natural soils and hedgerows; overland flow in urban zones and over roads, storage in storm water retention basins and flow routing within networks consisting of ditches, natural rivers and sewer pipes. Several drainage networks can coexist and interact, which allows the modelling of complex suburban drainage systems. The paper presents the model structure and a first application to a simple test case.

Keywords: Suburban; LIQUID; distributed hydrological modelling; sewer system; ditches;

1. INTRODUCTION

All over the world people migrate towards cities, causing an encroachment of built-up areas in their peripheries and a change in agricultural practices. This has an impact on land use, like an increase of impervious areas and a decline of agricultural areas for the benefit of forests. The change in land use influences the hydrological processes in a catchment. Furthermore, urbanization is often followed by installation of sewer systems, river regulations or more recently construction of retention basins. Often, a faster catchment response is the result of these anthropogenic modifications, causing floods, a decrease of groundwater resource and a higher risk of water contamination. The optimization of water management in these areas can be obtained by application of distributed hydrological models. They offer the possibility to take into account land use or climate change, and more particularly to consider urbanized areas [Franczyk and Chang, 2009; Semadeni-Davies et al. 2008]. Regarding suburban areas, several questions arise: How to handle the landscape heterogeneity in suburban areas? How to deal with different time scales with fast hydrological response in urban zones and slower response in natural zones? How to simulate different parallel drainage systems with multiple outlets and interactions, such as overflow sewer devices? How to integrate the fast changes in suburban areas into the model? Regarding all these questions, an adaptable modelling approach seems appropriate. Using a modelling framework instead of a single model has the advantage that different
processes can easily be coupled and that new processes/objects can be added without much difficulty. We propose an object oriented approach based on the LIQUID modelling framework [Viallet et al., 2006]. Among all the existing modelling frameworks (WaterCAST [Cook et al., 2009], JAMS [Kralisch and Krause, 2006], OpenMI [Gregersen et al., 2007], etc.) LIQUID was chosen due to several reasons: It can handle irregular geometries, physically based and conceptual approaches can easily coexist, each process has its own time step, feedback between process modules is easily possible, water flow paths do not have to follow topography, all code is written in C++ and new process modules can be added straightforwardly. Furthermore, the computation time can be minimized by use of irregular time steps, which allows computations to be performed only when required. The LIQUID framework is presented in more detail in a companion paper [Branger et al., 2010b]. This paper presents the Peri-Urban Model for landscape MA/Nagement (PUMMA) built within the LIQUID framework and adapted to small suburban catchments. First a description of the model discretization, retained for suburban areas is given, then the structure of the model and the model components are presented. In section 3 the model is applied to the Rez catchment, France and the results are discussed.

2. DESCRIPTION OF THE PUMMA MODEL

2.1 Model mesh

The model mesh of the PUMMA model consists of irregular hydro-landscapes [Dehotin and Braud, 2008] representing hydrological response units (HRUs). In the rural area the HRUs are created by intersection of detailed land-use maps, geology, soil and sub-basin maps. The sub-basins are calculated using the stream-burning technique [Hutchinson, 1989] in order to obtain one sub-basin for each river and artificial ditch reach. The urban area is discretized into urban hydrological elements (UHE) [Rodriguez et al., 2008] encompassing one cadastral unit (typically house plus surrounding garden) and half of the adjoining street. The UHEs are connected to the closest drainage network, which can consist of a sewer pipe or a natural ditch.

2.2 Model

The Peri-Urban Model for landscape MA/Nagement (PUMMA), intended for the detailed simulation of water fluxes in suburban areas, is under development within the A/VuPUR (Assessing the Vulnerability of Peri-Urban Rivers) project [Braud et al., 2010]. It is an extended combination of the BVFT model [Branger et al., 2008], which was developed for the assessment of the hydrological impact of landscape management practices and of the URBS model [Rodriguez et al., 2008] for urban areas, which was integrated into the LIQUID framework. As in the BVFT model the choice of the modules in PUMMA depends mainly on the land-use properties. Five main modules are involved: URBS for urban hydrological elements, FRER1D for natural surfaces, HEDGE for hedgerows, SIMBA for retention basins and lakes and RIVER1D for the drainage network. Figure 1 shows a vertical landscape profile with the applied modules and modelled water flow paths. Each model unit (HRU/UHE) represents one module instance. We can see that infiltration is the main process in forests and fields, modelled with the FRER1D module. By setting the soil hydraulic conductivity to zero, FRER1D can also model overland flow, for example on streets. URBS, which is applied on each UHE, models overland flow and infiltration. The natural river and the artificial drainage network are modelled with RIVER1D. HEDGE is applied to vegetated field borders and riparian zones. The subsurface flow exchange is modelled with interfaces, explained in part 2.4. The next two paragraphs explain the modules in more detail, their connections, and the construction of the final model.
2.3 Modules

The URBS module [Rodriguez et al., 2008] calculates the components of the rainwater flux at the UHE scale. The UHE is divided into three land-use classes: street, impervious area and natural area. Each land-use class is represented by four reservoirs (tree, surface, vadose zone and saturated zone), which are connected due to vertical water fluxes. With this representation it is possible to describe the following processes: interception by trees, evapotranspiration, surface runoff, soil infiltration and infiltration into sewer pipes due to sewer defects. The ideal drain approach [Gustafsson et al., 1996] is used for the calculation of the network infiltration, which can be observed in most sewer systems due to cracks. The FRER1D module represents vertical infiltration over natural surfaces using a 1D resolution of the Richards equation [Ross, 2003; Varado et al., 2006]. As input of the FRER1D module, the evapotranspiration is calculated in an extra set of modules, integrating the effects of crop rotation and root extraction. Following the approach of Viaud et al. [2005], the conceptual HEDGE module calculates the water table dynamics and evapotranspiration processes below hedgerows. Lateral flow in the saturated zone is considered in HEDGE, FRER1D and URBS modules due to source/sink terms. The SIMBA (SIMulation of storage Basins) module, developed especially for the PUMMA model, uses a simple linear reservoir equation to simulate retention basins or natural lakes. It has a lower outflow depending on the hydraulic head and an overflow. Evaporation, precipitation, irrigation and water exchange with the groundwater table are taken into account due to source/sink terms. Natural river network, ditches and sewer system are modelled with the RIVER1D module [Branger et al., 2008], which calculates a solution of the one-dimensional kinematic wave approximation [Moussa and Bocquillon, 1996]. The flow velocity is calculated with the Manning-Strickler equation. The drainage network is divided into several reaches that can receive lateral surface and subsurface flows. Each reach can have a different cross section (rectangular, trapezoidal or triangular) and Manning coefficient value. Input modules allow the distribution of rainfall and potential evapotranspiration over the model units. Nearly all the current state variables such as soil moisture content, river reach discharge and actual evapotranspiration can be extracted as distributed output.

2.4 Spatial coupling

The spatial coupling between different modules is realised by means of interfaces (Figure 2). They simulate the subsurface flow between adjacent model units. The Water Table Interface (WTI) uses the Darcy law to calculate the subsurface flow between agricultural
fields, hedgerows and urban cadastral units. The water exchange between river, or lake and adjacent agricultural field/hedgerow is computed in the Water Table River Interface (WTRI) based on the Miles [1985] approach. The calculated flow direction depends on the hydraulic head inside the model units and is thus bidirectional. Depending on the flow direction the computed discharge, which is taken into account in the receiving modules due to source or sink terms, is positive or negative.

Figure 2. Structure of the PUMMA model and couplings between the modules.

At the moment, the surface outflow and network infiltration of the URBS module are directly injected into the drainage network. Several rivers (implementations of RIVER1D module) can coexist and exchange water via source and sink terms. This facilitates the representation of an artificial drainage network besides a natural river. Sewer overflow devices are simulated by an interface between sewer network and river streams. The overflow is calculated with a weir or orifice equation based on a control water depth threshold value. The connection between river and lake is modelled by interrupting the river network at the lake inlet and by starting a new river at the outlet.

3. MODEL APPLICATION AND RESULTS

3.1 Test case

The model is assembled gradually by adding new modules and connections. A simple test case was designed for this first version of PUMMA, consisting of a combination of the URBS, RIVER1D and SIMBA modules.

In order to compare these first results to measured data, the Rez catchment [Berthier et al., 1999], located in the suburbs of Nantes, France was chosen (Figure 3). It consists of 70 cadastral parcels connected to either street or sewer network. As the pre-processing of the geographic data has still to be done manually, the catchment was conceptualised into ten equally parameterised cadastral parcels connected to ten drainage reaches for this first model test. The outflow of each cadastral parcel (sum of surface runoff and network infiltration) was multiplied by 7 and added as source term to the adjacent drainage reach. The parameters of the URBS module correspond to the retained parameters in Rodriguez et al. [2008]. The depth of the drainage pipe was set to 1.2m and the saturated zone depth was 0.7m. The drainage network was represented by a rectangular cross section of 50cm width with a slope of 0.5% and a Manning value of 0.011. The precipitation is the average of three rain gauges [Berthier et al., 1999] and has a time step of 5 minutes. The evapotranspiration data come from Météo France. All state variables were initialized to zero. The results were compared to measured flow data, see Figure 3 for its location. An hypothetical planning
scenario has been tested, by introducing a retention basin downstream of the last drainage reach in order to highlight the simulation results of the SIMBA module. No measured data were available for comparison. The retention basin of 2m depth had a bottom outflow. Two simulations were run with retention parameters of 500 and 1000 s for a rain event on the 3rd of January 1991 including a three day warm-up period to eliminate the effect of the parameter initialisation. No lateral subsurface exchange was modelled, except the infiltration into the sewer pipes.

3.2 Results

This first test allows the assessment of the correct functioning and interaction of the modules. Each implementation of URBS (UHE) signals the sum of the surface runoff and network infiltration towards the closest drainage reach. RIVER1D routes the received water from one reach to the next and transmits then the discharge to the hypothetical retention basin.

Figure 4 shows the rain event and the simulated discharge at the outlet of the Rez catchment versus the measured discharge. The total amount of precipitation during the simulation period was 21.3 mm. The modelled discharge corresponds to the main peaks of the observed discharge, however it fluctuates too much. This is probably caused by the conceptual representation of the catchment, due to which the correct simulation of the flow routing is not possible. The full distributed application of the model to the Rez catchment, which is in process and which will be presented during the conference, should improve the model results. Another important point is the variable time step of the model. Here, we plotted the instantaneous values. Using averaged values for 5 minutes intervals should equally even the flow curve. The flow volume during the simulated period could be reproduced to 98.8% for this specific rain event.

Figure 5 presents the different flow components of one UHE. A short response of surface runoff to the incoming precipitation can be observed for the UHE, which is typical for urban areas. The built surface contributes to 40.6% to the runoff, the road area contributes to 29.1% and the network infiltration contributes to 30.3%. The response of the road surface is delayed in comparison to the built area, which is due to the parameterization of the UHEs (maximum size of surface reservoirs). The natural area does not contribute to the surface runoff because the soil is not yet saturated. However, the rising water content in the natural soil reservoir provokes an increasing infiltration into the sewer pipes. This response is much slower than the surface response. The discharge of each UHE is added as source term to the closest drainage reach and routed downstream.
Figure 4. Simulated (black) versus measured (red) flow data and the precipitation (green) during the simulation run (3/1/1991). The simulated discharge corresponds to the RIVERID output at the catchment outlet.

Considering the hypothetical planning scenario using the SIMBA module, a retention effect can be simulated. Depending on the retention parameter chosen for the module, the retention effect of the hypothetical basin is more or less strong. A retention parameter of 500s results in a peak reduction of 36% and a retention parameter of 1000s in a reduction of 49% combined with retardation and dispersion effects.

Figure 5. The discharge of one UHE divided in network infiltration (blue) and surface runoff from built areas (black), the road (green) and natural areas (red).

Figure 6. Discharge from the hypothetical retention basin with 500s (black) and 1000s (red) as retention parameter.
4. CONCLUSIONS AND PERSPECTIVES

PUMMA is a distributed model under development describing hydrological processes in suburban areas. It is constructed using the LIQUID modelling framework. The model is composed of different interacting modules (URBS, FRER1D, HEDGE, SIMBA and RIVER1D). Each of them represents a different hydrological process depending on the land use property. Each module has its own time step, facilitating the simulation of different time scales regarding the hydrological processes. The model mesh consists of a mix of HRUs and UHEs with irregular geometries, which allows to represent the landscape heterogeneity in suburban areas. Different parallel drainage systems can be modelled by repeated implementation of the RIVER1D module. As the flow does not depend directly on topography, anthropogenic features as ditches and sewer pipes with pumped water can easily be simulated. Due to the modelling framework, new modules describing anthropogenic features as for example sewer overflow devices can be added straightforwardly.

A simplified version of the model was applied to the Rez catchment and compared to measured data. Despite a very simple representation of the model mesh and a homogenous model parameterization, it was possible to reproduce the measured flow peaks. However, the simulated flow fluctuates too much because of the conceptual representation of the model mesh and the variable time step. This will be improved in a model application using the real Rez model mesh. This application allowed to verify the correct functioning of the modules and their interactions. The results are encouraging as the model reflects well the involved hydrological processes. Due to the distributed character of the model and the great number of necessary input parameters, the model application remains restricted to small catchments of several square kilometres dedicated to research purposes. This is also partly explained by a still time consuming pre-processing of geographical data. An automation of the pre-processing is under development, using the GRASS geographical information system. The application to the two suburban sub-basins Chaudanne (4.1 km$^2$) and Mercier (6.8 km$^2$), located in the periphery of Lyon, France, is under preparation [Braud et al., 2010]. The complete model can then be validated with measured data. With this kind of model a detailed description of the processes in suburban areas is possible. The impact of soil impermeabilization, planting of hedgerows, installation of ditches, sewer systems or retention basins on catchment hydrology can be quantified. The results of these models can be used to understand the major hydrological processes and derive simplified approaches for use in larger catchments.

ACKNOWLEDGMENTS

The development of the PUMMA model is part of the AVuPUR project funded by the French National Research Agency (ANR) under contract ANR-07-VULN-01. This work has also been supported by the Cemagref and Hydrowide fundings. The data for the Rez catchment were provided from the LCPC and Mto France. Special mention should be made of Yvan Paill who created the UHEs for the Rez catchment.

REFERENCES


Dehotin, J. and I. Braud, Which spatial discretization for distributed hydrological models? Proposition of a methodology and illustration for medium to large scale catchments, Hydrology and Earth System Sciences, 12, 769-796, 2008.


End-to-End Workflows for Coupled Climate and Hydrological Modeling

Kathy Sainta and Sylvia Murphy

aSGI Inc., 17141 Primavera Cir. Cape Coral, FL 33909 (ksaint@sgi.com)
bCooperative Institute for Research in the Environmental Science, University of Colorado Boulder, Box 216 UCB Boulder, CO 80309-0216 USA (sylvia.murphy@noaa.gov)

Abstract: In order to examine the effects of environmental changes on local watersheds, it has become increasingly important to be able to interoperate between diverse models and document and share the resulting data. To this end, we are prototyping an end-to-end workflow that executes multiple, distributed models, loosely couples the model components, and disseminates the model results and metadata using a data portal. This project is made up of three major components and is implemented in multiple phases. The main components include the interface to the models for configuration and execution, the interface to the portal for data dissemination, and the workflow that provides the glue to make it all work together. The models being coupled are a high performance atmospheric model (currently represented by a stubbed-out ESMF Component) and the hydrological model, SWAT (Soil and Water Assessment Tool). Each model is accessed via standard interfaces (OpenMI, ESMF), set up as web services, and in the final product, the resulting data will be published to the ESG (Earth System Grid) or other data portal. Managing the model execution and coupling is initially handled by the OpenMI Configuration Manager, but may eventually be replaced by a more extensive workflow application, such as Kepler. The resulting product is an end-to-end, self-describing and repeatable workflow that demonstrates one-way interactive systems involving climate and other models can be created to address emerging questions about climate impacts.

Keywords: Frameworks; Hydrology; Climate; Modelling

1. INTRODUCTION

Climate change is one of the most pressing scientific questions of the day and climate models have become an important tool in the examination of climate change. While most climate models attempt to resolve the hydrological and water balance cycles, their resolution is usually insufficient to fully resolve climate impacts upon water resources. A separate hydrological model is usually required [Graham et al., 2007]. Studies show, however, that these separate hydrological models perform better when they are coupled in some way to a global or regional climate model [Zeng et al., 2003; Yong et al., 2009]
There are two paradigms for the coupling of climate and hydrology models. The traditional method is to directly force the hydrology model with output from the climate model [Graham et al., 2007]. Attempts are also being made to integrate standalone hydrological models into climate modeling systems so that two-way feedback is possible.

Advances in the direct forcing method are hindered by the technologies in use within the climate and hydrological communities. Climate models run in high performance computing (HPC) environments while many hydrological models run on personal computers. Both communities have, however, a standardized framework, that if utilized, can bridge this technological gap.

Standardized model frameworks have another key advantage in that they remove the barriers to interoperability on a broad scale. Developers are no longer limited to models they have laboriously integrated. Instead, any model that is wrapped in the framework may be used. This greatly expands scientific flexibility.

This paper describes a project that improves the direct forcing of hydrology models by climate models through the integration of the Earth System Modeling Framework (ESMF), and OpenMI. ESMF (http://earthsystemmodelling.org) is a widely used framework for geoscience HPC applications while OpenMI (http://www.openmi.org/) is a hydrological community standard. Both systems, as well as our modeling workflow, will be described in detail in the following sections.

It should be noted that this paper describes the results from the first phase of an ongoing effort to couple a hydrological model with an active atmosphere and to archive and describe the data from both sides of the simulation to a science gateway. The final phase of the project will build upon both of these milestones to achieve two-way communications between the models.

2. SYSTEM DESCRIPTION

This section describes the main pieces of the coupled climate/hydrology system (Figure 1). A high level description of the entire system is provided first. Descriptions of the individual components follow.

At the highest level, this coupled system is comprised of three main components: a hydrological model, an atmospheric model, and a driver application. The atmospheric model is implemented as an ESMF component and is wrapped with an OpenMI interface, which facilitates the communication with the OpenMI-compliant hydrological model. The hydrological model, driver, and the two OpenMI wrappers physically exist on a personal computer running Windows. The ESMF component itself exists on a multi-threaded Linux computer. An ESMF web services interface is used to bridge the two computer platforms.

The data files represented in Figure 1 can exist at any network accessible location. In phase 1 of the project, the files are copied to the PC by the atmospheric model wrapper. In the final phase of the project, this runtime data transfer will occur exclusively by the web services interfaces. At that point the arrow between the PC and the data files in Figure 1 will become uni-directional, representing the archival of model output only.
2.1 The Frameworks

2.1.1 Earth System Modeling Framework (ESMF)

ESMF is a high-performance, flexible software infrastructure that increases the ease of use, performance portability, interoperability, and reuse of Earth science applications. ESMF provides an architecture for composing complex, coupled modeling systems and includes data structures (e.g. representing physical fields) as well as utilities for developing individual models. The software includes tools to describe and organize models and the ability to interface with model components via web services [Hill et al., 2004].

**ESMF components**

The terminology “ESMF component” has a particular meaning that is important to define. ESMF components are systems that have been broken into initialize, run, and finalize segments. Each of these segments is called using a specific, and standardized interface.

**ESMF data structures**

ESMF data structures include Arrays, ArrayBundles, Fields, FieldBundles, and States. ESMF Arrays store multi-dimensional data. They can also store halo points. The ESMF Field represents data defined within a continuous region of space. It includes the physical grid associated with the data and a decomposition that specifies how data associated with points in the physical grid are distributed in computer memory. ArrayBundles and FieldBundles are groupings of Arrays and Fields respectively. States can contain any of the aforementioned data structures and are used to transfer data between components.

**ESMF web services**

ESMF web services (http://www.earthsystemcurator.org/projects/web_service.shtml) consist of a socket interface library that allows any networked ESMF component (defined in section 2.1.1) to be available as a web service. This is accomplished through a series of Java classes that provide the following: a) access to the networked ESMF component via a java socket client, b) a set of SOAP interfaces that select and execute the component.
services, and c) a registrar application that the components use to register themselves with the web service.

2.1.2 OpenMI

OpenMI Software Development Kit (SDK) is a software library that provides the hydrological community with a standardized interface that focuses on time dependent data transfer. It is primarily designed to work with systems that run simultaneously, but in a single-threaded environment. Regridding and temporal interpolation are also part of the package [Gregerson et al., 2007].

OpenMI data structures

Components in OpenMI are called LinkableComponents. They perform a variety of functions. The primary data structure in OpenMI is the ExchangeItem, which comes in the form of an InputExchangeItem and an OutputExchangeItem. The former describes what can be accepted at each location while the latter describes what can be provided at each location [Gregerson et al., 2007].

OpenMI GetValues

Data transfer begins in OpenMI when a LinkableComponent requests data of another LinkableComponent via the GetValues method. In a two-way system, the data provider does not run forward in time until it receives this data request. Once it does, the component runs forward in time, stops, and converts its data onto the grid or location of the requesting LinkableComponent [Moore and Tindall, 2005]. Often in the hydrological community, these data requests are for point data.

2.2 System Components

2.2.1 The Hydrological Model

The hydrological model chosen for this project is the Soil Water Assessment Tool (SWAT). It is a river basin scale model developed to quantify the impact of land management practices in large, complex watersheds [Gassman et al., 2007]. It was chosen for this project because it is widely used, is open source, and runs on a Windows platform. The OpenMI version that has been made into a LinkableComponent, was downloaded with permission from the United Nations Educational, Scientific and Cultural Organization-Institute for Water Education (UNESCO-IHE). Inputs to SWAT include maximum and minimum temperatures, daily precipitation, relative humidity, solar radiation data, and wind speed data [Gassman et al., 2007].

2.2.3 The Atmospheric Model

In this phase of our project, an ESMF web services prototype containing a stub component representing an atmospheric model was used to send data from ESMF to OpenMI and SWAT. The next phase of the project will replace this system test with an atmospheric model. Using stubs in this first phase allowed us to focus on the OpenMI to ESMF communications and postpone the complexity of dealing with a HPC atmospheric model until those communications were in place.

The ESMF web services code exercises a full component lifecycle (initialize, run, finalize) through a socket interface. In the run portion of the code, the stub component reads in atmospheric data, converts that data into ESMF Arrays, and then outputs the same data, mimicking history output. This output file served as the source file for SWAT climatic inputs.
2.2.2 **The System Driver**

The system driver controls the application flow and is implemented using OpenMI’s Configuration Editor. It loads the models as defined in configuration files (OMI files) and uses a trigger to start the run. The Configuration Editor provides a graphical user interface for setting up and configuring the workflow execution. When a model is loaded into the Configuration Editor, its input and output exchange items are defined. The user can then define how models exchange data by mapping output exchange items in one model to input exchange items in the other model.

The OpenMI Configuration Editor was a convenient tool for the testing of the OpenMI implementations and model interactions used in this project. Since the Configuration Editor is part of the OpenMI Graphical Users Interface, it allowed for the running of the OpenMI wrapped model codes without additional coding. This was a significant developmental timesaver. In the final phase of the project, however, the goal is to include this system within a larger workflow that includes the archival of history files to a science gateway. This final workflow will also contain post-processing steps such as data visualization. The Configuration Editor is unsuited to this task and therefore may eventually be replaced by a more versatile workflow package (e.g. Kepler [https://kepler-project.org/]).

2.3 **Logical Workflow**

The OpenMI interface standard defines a serial approach to communication between models. Figure 2 provides a detailed view of the method calls for the system. An OpenMI implementation will follow these fundamental steps of execution: initialization and configuration, preparation, execution, and completion. These steps correspond to methods in the LinkableComponent interface: Initialize, Prepare, GetValues, and Finish/Dispose. The Initialize method is where the input and output exchange items are defined and the connection to the ESMF web services is created. In the current system, the Prepare method is where the atmospheric model does most of its work; the model is run and the output data is exported into a file. Once the models are prepared, the driver performs a time step on the SWAT model, which then triggers a request to the atmospheric model (via the GetValues method) to exchange data. The atmospheric model retrieves the data from the data file and returns it to the SWAT model. The time step process repeats until the end time is reached, after which the models are completed using the Finish and Dispose methods.

![Figure 2. The sequential method calling sequence for the entire system.](image-url)
2.4 The Data flow

The format of the data in this coupled system changes several times during execution. The data initially starts out as an ESMF Array contained within an ESMF State. That State is exported to an external NetCDF file. Once SWAT requests data via OpenMI’s GetValues method, the atmospheric OpenMI wrapper accesses the NetCDF file and converts the data into an OpenMI OutputExchangeItem. This then gets passed into the SWAT OpenMI wrapper where it is converted to an InputExchangeItem. The data is now available for use by SWAT (Figure 3).

![Figure 3. The flow of data through the system.](image)

3. IMPLEMENTATION

To reiterate, this paper represents a snapshot of an ongoing effort to create a two-way coupling between SWAT and an atmospheric model and to archive and describe the data from this simulation on a science gateway. The following sections describe the implementation details of the current phase and then the plans for migrating to a more complete implementation.

3.1 Current Implementation Details

This project is being implemented using an incremental prototype approach, and the current phase focuses on the interfaces between each of the system components. Some of the major differences between the current implementation and the final system design include the following: a stubbed-out ESMF component is used to represent the atmospheric data model; data is transferred between this component and the OpenMI wrapper using NetCDF files instead of streamed across the interface; and output data is not currently published to a data portal. Detailed instructions to reproduce this work can be found on our open source repository at [http://esmfcontrib.cvs.sourceforge.net/viewvc/esmfcontrib/HydroInterop/](http://esmfcontrib.cvs.sourceforge.net/viewvc/esmfcontrib/HydroInterop/).

3.2 Future Directions

The remaining tasks involved in this project will be implemented incrementally, and will involve several releases of the prototype system. By using an incremental approach to implementing this project, we can also learn from the early implementations, and use these lessons learned to modify and improve the final system architecture.

The major tasks to be implemented in the near future include replacing the stubbed out ESMF component with the Community Atmospheric Model (CAM) that is part of the...
Community Climate System Model (CCSM), or another atmospheric model. In addition, the NetCDF data transfer will be replaced by an online data transfer through the ESMF web services. Finally, the Earth System Grid (http://earthsystemgrid.org) or another science gateway will be brought into the system to host and describe the resultant coupled simulation data.

4. CONCLUSION

The first phase of this project has successfully demonstrated a coupled ESMF/OpenMI modeling system. Both SWAT and the ESMF atmospheric stub component were initialized and run, and data was transmitted, on request, to SWAT via an external climate data file. The code for this system and instruction to reproduce our results is publically available at http://esmfcontrib.cvs.sourceforge.net/viewvc/esmfcontrib/HydroInterop/.

Enabling OpenMI and ESMF to interface is an important milestone. These two infrastructures are both widely used within their respective communities and their integration has the potential for enabling diverse geoscience and hydrological modeling systems to interact in a more flexible way.

REFERENCES


Environmental Modeling Framework Invasiveness: Analysis and Implications

W. Lloyd a, O. David a, J.C. Ascough II b, K.W. Rojas c, J.R. Carlson e, G.H. Leavesley a, P. Krause d, T.R. Green b, L.R. Ahuja b

a Dept. of Civil and Environmental Engineering and Dept. of Computer Science, Colorado State University, Fort Collins, CO 80523 USA (e-mail address: odavid@colostate.edu)
b USDA-ARS, ASRU, 2150 Centre Ave., Bldg. D, Suite 200, Fort Collins, CO 80526 USA
c USDA-NRCS, 2150 Centre Ave., Bldg. A, Fort Collins, CO 80526 USA
d Department of Geography, Friedrich-Schiller-Universität Jena, Jena, Germany

Abstract: Environmental modeling frameworks support scientific model development by providing an Application Programming Interface (API) which model developers use to implement models. This paper presents results of an investigation on the framework invasiveness of environmental modeling frameworks. Invasiveness is defined as the quantity of dependencies between model code and the modeling framework. This research investigates relationships between invasiveness and the quality of modeling code. Additionally, we investigate the relationship between invasiveness and two common framework designs (lightweight vs. heavyweight). Five metrics to measure framework invasiveness were proposed and applied to measure invasiveness between model and framework code of several implementations of Thornthwaite and the Precipitation-Runoff Modeling System (PRMS), two hydrological models. Framework invasiveness measurements were compared with existing software metrics including size (lines of code), cyclomatic complexity, and object-oriented coupling with generally positive correlations being found. We found that models with lower framework invasiveness tended to be smaller, less complex, and have lower coupling. In addition, the lightweight framework implementations of the Thornthwaite and PRMS models were less invasive than the heavyweight framework model implementations. Our initial results suggest that framework invasiveness is undesirable for model code quality and that lightweight frameworks may help reduce invasiveness.

Keywords: Component-based modeling; Environmental modeling frameworks; Invasiveness; Frameworks; Software metrics.

1. INTRODUCTION

Environmental modeling frameworks support model development through provisioning of libraries of core modeling modules or components, component interaction/communication, time/spatial stepping/iteration, up/downscaling of spatial data, multi-threading/multiprocessor support, and cross language interoperability, as well as reusable tools for data analysis and visualization. Environmental modeling frameworks provide structure for models by supporting the disaggregation of modeling functions into components, classes, or modules. In this paper, we refer to functional units of model code as components. Components are able to be reused in other models coded to the same framework with little migration effort. One advantage of using an established environmental modeling framework is they often provide pre-existing libraries of components to help facilitate model development (Voinov et al., 2004; Argent et al., 2006). In this paper, we define the degree of dependency between an environmental modeling framework and model code as “framework invasiveness.” This is the degree to
which model code is coupled to the underlying framework. Framework to application invasiveness occurs from the following:

- Use of a framework Application Programming Interface (API) consisting of data types and methods/functions which developers interface with to harness framework functionality;
- Use of framework specific data structures (e.g., classes, types, constants);
- Implementation of framework interfaces and extension of framework classes;
- Boilerplate code (“non-science” code required for model to run under the framework); and
- Framework requirements including language, platform, and libraries.

Framework to application invasiveness is a type of code coupling; object-oriented coupling (i.e., coupling between classes in an object-oriented program) has been shown to correlate inversely with software quality (Briand et al., 2000; Basil et al., 1996). One goal of this research is to explore relationships between environmental model code quality and the degree of invasiveness between model code and environmental modeling frameworks. There are many dimensions to model code quality, often referred to as quality attributes. Quality attributes that may be impacted by framework invasiveness include understandability, maintainability, and portability/reusability.

Modeling frameworks can be classified as either heavyweight or lightweight (Richardson, 2006). Framework type characteristics are described in Table 1. A primary difference is how frameworks present functionality to the developer. Heavyweight frameworks provide developers with a large application programming interface (API) and developers typically spend considerable time becoming familiar with it before writing model code. Lightweight frameworks provide functionality to developers using techniques aimed at reducing the API’s overall size. Programming annotations capture metadata and are used to identify points in the model code where framework functionality should be integrated. Framework integration can also be accomplished using external XML files. Wherever possible, “convention over configuration” is favored in that system defaults are assumed and developers only specify unconventional details in model code. Non-default behavior may include unique component data input/output requirements, pre-conditions, post-conditions, etc. in model code. Framework specific data types which take the place of system data types are avoided in lightweight framework designs. A second goal of this research is to explore the relationship between the framework type (i.e., heavyweight vs. lightweight) and the degree of framework to application invasiveness.

Table 1. Heavyweight versus lightweight framework design classification.

<table>
<thead>
<tr>
<th>Heavyweight Frameworks</th>
<th>Lightweight Frameworks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Components under the framework:</td>
<td>Components under the framework:</td>
</tr>
<tr>
<td>- bound statically at compile time</td>
<td>- bound dynamically at run time by use of language annotations/dependency injection</td>
</tr>
<tr>
<td>- tightly coupled to the framework by extension of framework classes,</td>
<td>(inversion of control software design pattern)</td>
</tr>
<tr>
<td>implementation of framework interfaces,</td>
<td>- loosely coupled and framework independent</td>
</tr>
<tr>
<td>use of framework data types, and use of framework functions</td>
<td>- Convention over configuration: developers only specify unconventional details in code</td>
</tr>
<tr>
<td>- Provides specialized versions of native language data types</td>
<td>as defaults are assumed</td>
</tr>
<tr>
<td>- Have a “large” programming interface (API)</td>
<td>- Uses native language data types</td>
</tr>
<tr>
<td>- Use may depend on many libraries</td>
<td>- Have a “small” programming interface (API)</td>
</tr>
</tbody>
</table>

The broad objectives of this research are to investigate the implications of framework invasiveness on model code quality and to investigate the framework invasiveness characteristics of both lightweight and heavyweight frameworks. Specifically, we seek to answer the following research questions: 1) what is the impact of framework to model code invasiveness on model code quality, and 2) do the design characteristics of lightweight frameworks enable model development resulting in lower framework to model code invasiveness? Previously, environmental modelers have developed models with only a vague understanding of how the design of environmental modeling frameworks impact modeling efforts. A better understanding of the phenomenon of framework to application
invasiveness can help modelers in choosing and designing modeling frameworks to improve the quality of scientific models throughout their entire software life-cycles.

To investigate the above questions, we performed a case study using two environmental models, a monthly water balance model (Thornthwaite) and a complex watershed-scale model (the Precipitation Runoff Modeling System, PRMS). A set of software metrics was devised and applied to quantify the invasiveness between the framework and model code. Several traditional software quality metrics were used to assess the quality of the environmental model implementations in terms of size, complexity, and object-oriented coupling. An analysis of results was performed to identify relationships between model code quality and invasiveness, and also between framework type (e.g., heavyweight/weight) versus invasiveness.

2. ENVIRONMENTAL MODELING FRAMEWORKS

For this framework invasiveness study, the ESMF 3.1.1, CCA 0.6.6, OpenMI 1.4, and OMS 2.2 and 3.0 environmental modeling frameworks were used to implement the Thornthwaite (Thornthwaite, 1948) and PRMS (Leavesley et al., 2006) environmental models. Additionally, three non-framework based implementations of Thornthwaite were implemented in Java, C++, and FORTRAN to assist in developing framework-based versions. ESMF is an open source framework developed by the National Center for Atmospheric Research (NCAR) for building climate, numerical weather prediction, data assimilation, and other Earth science software applications (Collins et al., 2005). CCA was developed by the members of the Common Component Architecture Form, and is a component architecture for high performance computing (Armstrong et al., 1999). OpenMI is sponsored by the European Commission LIFE Environment program and is a software component interface definition for developing models in the water domain (Blind and Gregersen, 2005). OpenMI Thornthwaite model implementation in this study was performed using a Java-based implementation of OpenMI, although a .NET/C# version exists and is generally considered more popular. The Object Modeling System (OMS) versions 2.2 and 3.0 are developed by the USDA – Agricultural Research Service (ARS) in cooperation with Colorado State University. OMS facilitates component-oriented simulation model development in Java, C/C++ and FORTRAN (David et al., 2002), and version 2.2 provides an integrated development environment (IDE) with numerous tools supporting data retrieval, GIS, graphical visualization, statistical analysis and model calibration (Ahuja et al., 2005). The ESMF 3.1.1, CCA 0.6.6, OpenMI 1.4, and OMS 2.2 frameworks can be considered as heavyweight frameworks where modeling code is coupled to the framework through dependencies on a framework's API, i.e., components must use specific data types and functions to interface with the framework. OMS 3.0 has been developed with a “non-invasive” lightweight framework design for model development. That is, modeling components have been decoupled from the framework API wherever possible so that they exist as plain classes implementing only model specific logic. In addition, boilerplate code has been re-factored out using language annotations.

3. ENVIRONMENTAL MODELS

For this study, we investigated several implementations of Thornthwaite and the Precipitation-Runoff Modeling System (PRMS) hydrological models. Thornthwaite is a monthly water balance model which simulates water allocation among components of a hydrological system (Thornthwaite, 1948). The model was selected since it has a typical structure for a hydrological simulation model and its size and complexity were manageable for this study. All model implementations were coded to produce identical numeric output, programming language specific formatting functions were not used, and only framework support for component aggregation and component interaction/communication were utilized. The average code size of the framework based Thornthwaite model implementations was 754 lines of code (LOC).

The Precipitation-Runoff Modeling System (PRMS) is a deterministic, distributed-parameter model developed to evaluate the impact of various combinations of precipitation, climate, and land use on stream flow, sediment yields, and general basin hydrology (Leavesley et al., 2006). PRMS was implemented in Java using the OMS 2.2
and 3.0 frameworks as time and resources were lacking to implement the model under additional frameworks. The PRMS implementations utilized framework support for component aggregation, interaction and communication as well as model time stepping. The average implementation size of the PRMS models studied was 13,580 LOC.

4. FRAMEWORK INVASIVENESS MEASURES

Research in object-oriented software evaluation has produced numerous metrics which help to measure attributes such as the coupling, cohesion, and inheritance among classes in an object-oriented program (Chidamber and Kemerer, 1994). However, the existing metrics were not designed to specifically quantify the dependencies between framework and modeling. The measures in the following sections were applied to quantify invasiveness between environmental modeling frameworks and model code.

4.1. Framework Data Types (FDT) and Framework Functions (FF)

We quantify usage of two primary framework constructs in a model: framework data types and framework functions. We count the total number of framework data types (classes, data structures, types, etc.) used (FDT-used), and the total number of uses of these framework data types in modeling code (FDT-uses). The total number of framework functions (functions, methods, subroutines, etc.) used (FF-used), and the total number of uses of these framework functions (calls) appearing in the modeling code (FF-uses) are counted. Three variations of the framework metrics were calculated: a raw count, a count of framework construct usage weighted per 1000 lines of code (KLOC) (e.g. FDT/FF-used/-uses per KLOC), and the percentage of usage relative to all framework constructs used/uses in the application code (e.g. % FDT/FF-used/-uses).

4.2. Framework Dependent Lines of Code (FDLOC)

To measure the invasiveness between model code and framework code, we counted the total number of lines of code which depend on the framework. A framework dependent line of code is defined as a line of code which depends on the framework such that if the framework libraries were removed the line would not compile. This implies that a framework dependent line of code contains at least one framework reference. In this study, we calculated two variations of FDLOC: raw count and a percentage relative to the total lines of model code (% FDLOC).

4.3 Software Quality Measures

As a surrogate for measuring model quality, we used three measures in this study: 1) size, measured by counting lines of code (LOC); 2) complexity, measured by determining cyclomatic complexity; and 3) coupling, measured using efferent coupling (fan-out), and afferent coupling (fan-in). Cyclomatic complexity (CC) counts the number of linearly independent paths through a program's source code. This is a surrogate for measuring code complexity and has been a widely used in computer science. To measure coupling, we used both efferent and afferent coupling measures because they can be collected on programs in both procedural and object-oriented languages. Efferent coupling is the number of classes which make reference to a class. This can be thought of as the number of uses “outside” of the class. This can be thought of as the classes used “inside” the class. Size, complexity and coupling measures generally inversely correlate with code quality (Basil et al., 1996; Briand et al., 2000; and Briand et al., 1999).

5. RESULTS

Static analysis tools were used to support analysis of the model implementations. SLOCCOUNT (SLOCcount, 2009) was used to count lines of code. Understand 2.0 Analyst (Understand, 2009) was used to collect the LOC, cyclomatic complexity, coupling between objects (CBO), and fan-in/fan-out coupling software metrics. Function and data type usage reports produced by Understand 2.0 were parsed using a custom program to generate data for the FDT and FF usage measurements. FDLOC were determined manually by counting lines of code.
5.1 Thornthwaite Model

The Thornthwaite model implementations were coded to provide identical output given the same inputs to allow us to attribute differences observed between the implementations to the invasiveness incurred from the different frameworks. Size and complexity measurements of the Thornthwaite model framework implementations are shown in Table 2.

<table>
<thead>
<tr>
<th>Language/Framework</th>
<th>Total LOC</th>
<th>Average CC/method</th>
<th>Total CC</th>
</tr>
</thead>
<tbody>
<tr>
<td>FORTRAN only</td>
<td>244</td>
<td>3.33</td>
<td>40</td>
</tr>
<tr>
<td>OMS 3.0 Java</td>
<td>295</td>
<td>2.38</td>
<td>31</td>
</tr>
<tr>
<td>Java only</td>
<td>319</td>
<td>2.85</td>
<td>37</td>
</tr>
<tr>
<td>C++ only</td>
<td>405</td>
<td>2.41</td>
<td>41</td>
</tr>
<tr>
<td>OMS 2.2 Java</td>
<td>450</td>
<td>1.18</td>
<td>103</td>
</tr>
<tr>
<td>ESMF 3.1.1 C</td>
<td>583</td>
<td>1.97</td>
<td>65</td>
</tr>
<tr>
<td>ESMF 3.1.1 FORTRAN</td>
<td>683</td>
<td>1.44</td>
<td>56</td>
</tr>
<tr>
<td>OpenMI 1.4 Java</td>
<td>880</td>
<td>1.61</td>
<td>116</td>
</tr>
<tr>
<td>CCA 0.6.6 Java</td>
<td>1635</td>
<td>2.25</td>
<td>276</td>
</tr>
</tbody>
</table>

The OMS 3.0 framework was the only framework which enabled a smaller model (in LOC) than the implementation in the equivalent native language, i.e., the OMS 3.0 Thornthwaite implementation was 295 LOC compared to 319 for Java-only. Ideally, a framework-based model implementation should have a smaller code size than a plain-language implementation with the reduced model code size reflecting code reuse where some aspects of the model implementation are provided by framework code.

Coupling measures for the Thornthwaite model framework implementations are shown in Table 3. For the Thornthwaite model framework implementations, measurements for size, complexity and coupling were positively correlated. Total LOC and cyclomatic complexity had a correlation coefficient of $r = 0.94$ (df = 4, $p < 0.01$), total LOC and total fan-in had a correlation coefficient of $r = 0.92$ (df = 3, $p < 0.05$), and total cyclomatic complexity with total fan-in had a correlation coefficient of $r = 0.95$ (df = 3, $p < 0.02$).

<table>
<thead>
<tr>
<th>Language/Framework</th>
<th>Total Fan-In (Afferent)</th>
<th>Total Fan-Out (Efferent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OMS 3.0 Java</td>
<td>116</td>
<td>70</td>
</tr>
<tr>
<td>OMS 2.2 Java</td>
<td>116</td>
<td>70</td>
</tr>
<tr>
<td>ESMF 3.1.1 C</td>
<td>100</td>
<td>155</td>
</tr>
<tr>
<td>ESMF 3.1.1 FORTRAN</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>OpenMI 1.4 Java</td>
<td>126</td>
<td>177</td>
</tr>
<tr>
<td>CCA 0.6.6 Java</td>
<td>195</td>
<td>215</td>
</tr>
</tbody>
</table>

Detailed invasiveness measurements for the Thornthwaite model framework implementations are shown in Table 4. For the framework invasiveness measures, the OMS 3.0 Thornthwaite model framework implementation appeared to be the least invasive, i.e., this implementation had far fewer framework dependencies than others. The Thornthwaite scientific code was essentially the same for all of the environmental modeling framework implementations, with the observed differences resulting from various framework-specific requirements to implement the model. The large variations in the metrics suggest that variations in framework design likely impact the modeling code. Table 4 also shows framework invasiveness metrics scaled to a percentage. The percentage scaling shows how much of the overall percentage of an attribute is framework dependent. Overall, a model implementation with low framework invasiveness should have a low percentage of data type, functions, and LOC dependence on the underlying framework.

The final invasiveness measurement scaling shown in Table 5 is a scaling of attribute occurrences per 1000 lines of code (KLOC), i.e., this scaling represents the expected number of occurrences if there were 1000 lines of code. Since the model implementations
varied in size, this scaling provides a method for a side-by-side comparison. FDLOC, FDT-used, FF-used correlated with model size (df = 4, p<.05); however, none of the percentage or scaling invasiveness measures correlated with size. For complexity, three invasiveness measures (FDLOC, FDT-used, and FF-used) were shown to correlate with total cyclomatic complexity (df = 4, p<.05). A correlation existed between FF-used/KLOC and average method cyclomatic complexity; however, correlation coefficients for other measures with average CC/method seem almost random so it is possible the FF-used/KLOC relation is spurious. Correlation coefficients between invasiveness and total complexity were generally positive though they varied in magnitude. Total fan-in (afferent) and fan-out (efferent) coupling correlated significantly with FDLOC, FDT-used, and also %FF-used (fan-in only) (df=3, p < 0.05).

Table 4. Framework invasiveness detailed measurements.

<table>
<thead>
<tr>
<th>Implementation</th>
<th>FDLOC</th>
<th>FDT-used</th>
<th>FDT-uses</th>
<th>FF-used</th>
<th>FF-uses</th>
</tr>
</thead>
<tbody>
<tr>
<td>OMS 3.0 Java</td>
<td>44</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td>21</td>
</tr>
<tr>
<td>OMS 2.2 Java</td>
<td>147</td>
<td>5</td>
<td>72</td>
<td>7</td>
<td>33</td>
</tr>
<tr>
<td>ESMF 3.1.1 C</td>
<td>178</td>
<td>10</td>
<td>122</td>
<td>13</td>
<td>77</td>
</tr>
<tr>
<td>ESMF 3.1.1 FORTRAN</td>
<td>280</td>
<td>3</td>
<td>109</td>
<td>11</td>
<td>148</td>
</tr>
<tr>
<td>OpenMI 1.4 Java</td>
<td>338</td>
<td>8</td>
<td>73</td>
<td>20</td>
<td>280</td>
</tr>
<tr>
<td>CCA 0.6.6 Java</td>
<td>533</td>
<td>15</td>
<td>135</td>
<td>48</td>
<td>215</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Implementation</th>
<th>FDLOC (%)</th>
<th>FDT-used (%)</th>
<th>FDT-uses (%)</th>
<th>FF-used (%)</th>
<th>FF-uses (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OMS 3.0 Java</td>
<td>14.84</td>
<td>4.67</td>
<td>1.35</td>
<td>26.67</td>
<td>40.38</td>
</tr>
<tr>
<td>OMS 2.2 Java</td>
<td>32.67</td>
<td>41.67</td>
<td>64.29</td>
<td>50.00</td>
<td>73.33</td>
</tr>
<tr>
<td>ESMF 3.1.1 C</td>
<td>30.85</td>
<td>30.30</td>
<td>49.59</td>
<td>46.43</td>
<td>76.24</td>
</tr>
<tr>
<td>ESMF 3.1.1 FORTRAN</td>
<td>41.42</td>
<td>27.27</td>
<td>51.90</td>
<td>78.57</td>
<td>96.10</td>
</tr>
<tr>
<td>OpenMI 1.4 Java</td>
<td>38.41</td>
<td>23.53</td>
<td>32.30</td>
<td>37.74</td>
<td>79.10</td>
</tr>
<tr>
<td>CCA 0.6.6 Java</td>
<td>32.60</td>
<td>46.88</td>
<td>49.82</td>
<td>70.59</td>
<td>69.58</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Implementation</th>
<th>FDLOC/KLOC</th>
<th>FDT-used/KLOC</th>
<th>FDT-uses/KLOC</th>
<th>FF-used/KLOC</th>
<th>FF-uses/KLOC</th>
</tr>
</thead>
<tbody>
<tr>
<td>OMS 3.0 Java</td>
<td>148</td>
<td>3.39</td>
<td>3.39</td>
<td>27.12</td>
<td>71.19</td>
</tr>
<tr>
<td>OMS 2.2 Java</td>
<td>327</td>
<td>11.11</td>
<td>160.00</td>
<td>15.56</td>
<td>73.33</td>
</tr>
<tr>
<td>ESMF 3.1.1 C</td>
<td>309</td>
<td>17.15</td>
<td>209.26</td>
<td>22.30</td>
<td>132.08</td>
</tr>
<tr>
<td>ESMF 3.1.1 FORTRAN</td>
<td>414</td>
<td>4.39</td>
<td>159.59</td>
<td>16.11</td>
<td>216.69</td>
</tr>
<tr>
<td>OpenMI 1.4 Java</td>
<td>384</td>
<td>9.09</td>
<td>82.95</td>
<td>22.73</td>
<td>318.18</td>
</tr>
<tr>
<td>CCA 0.6.6 Java</td>
<td>326</td>
<td>9.17</td>
<td>82.57</td>
<td>29.36</td>
<td>131.50</td>
</tr>
</tbody>
</table>

5.2 PRMS Model

The invasiveness metrics were applied to evaluate the PRMS model implementations under the OMS 2.2 and 3.0 frameworks. Size and complexity metrics are shown in Table 5. PRMS model code size was reduced 40% in the OMS 3.0 framework implementation. Much of the size reduction can be attributed to the elimination of component getter and setter methods. Getter and setter methods are accessor methods which intercept read/write access to data variables in an object-oriented program. These constructs are encouraged to provide data encapsulation to prevent unintentional changes to variables. The average complexity per method increased significantly from OMS 2.2 to OMS 3.0. This is because the total number of methods dropped significantly through elimination of the getter and setter methods. A reduction in model complexity is reflected in the more than three-fold reduction in total cyclomatic complexity observed in the OMS 3.0 PRMS model implementation versus OMS 2.2 (Table 5).

Table 5. PRMS model size and complexity metrics.

<table>
<thead>
<tr>
<th>Framework Implementation</th>
<th>Total LOC</th>
<th>Average CC/method</th>
<th>Total CC</th>
</tr>
</thead>
<tbody>
<tr>
<td>OMS 3.0 Java</td>
<td>10163</td>
<td>9.75</td>
<td>702</td>
</tr>
<tr>
<td>OMS 2.2 Java</td>
<td>16997</td>
<td>1.37</td>
<td>2575</td>
</tr>
</tbody>
</table>

Table 6. PRMS model coupling measures.

<table>
<thead>
<tr>
<th>Framework Implementation</th>
<th>Total Fan-In (Afferent)</th>
<th>Total Fan-Out (Efferent)</th>
<th>Avg. Number Methods/Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>OMS 3.0 Java</td>
<td>1232</td>
<td>755</td>
<td>3.6</td>
</tr>
<tr>
<td>OMS 2.2 Java</td>
<td>3517</td>
<td>1428</td>
<td>85.41</td>
</tr>
</tbody>
</table>
Coupling measures for the PRMS model implementations under the OMS 2.2 and 3.0 frameworks are shown in Table 6. Reductions in both total fan-out (efferent) and total fan-in (afferent) coupling are observed in the OMS 3.0 PRMS implementation. Coupling was likely reduced in relation to the reduction in code size attributed by removing getter and getter methods. The average number of methods per component dropped from 85 to 3.6 in OMS 3.0. Framework invasiveness measures for the PRMS model implementations are shown in Table 7. A significant reduction is seen in the use of framework data types and functions in the OMS 3.0 lightweight framework implementation.

<table>
<thead>
<tr>
<th>Implementation</th>
<th>FDT-used</th>
<th>FDT-uses</th>
<th>FF-used</th>
<th>FF-uses</th>
</tr>
</thead>
<tbody>
<tr>
<td>OMS 3.0 Java</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>OMS 2.2 Java</td>
<td>16</td>
<td>1788</td>
<td>15</td>
<td>2854</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Implementation</th>
<th>FDT-used (%)</th>
<th>FDT-uses (%)</th>
<th>FF-used (%)</th>
<th>FF-uses (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OMS 3.0 Java</td>
<td>5</td>
<td>0.19</td>
<td>5.5</td>
<td>2</td>
</tr>
<tr>
<td>OMS 2.2 Java</td>
<td>50</td>
<td>65.9</td>
<td>19.2</td>
<td>91.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Implementation</th>
<th>FDT-used/KLOC</th>
<th>FDT-uses/KLOC</th>
<th>FF-used/KLOC</th>
<th>FF-uses/KLOC</th>
</tr>
</thead>
<tbody>
<tr>
<td>OMS 3.0 Java</td>
<td>0.09</td>
<td>0.29</td>
<td>0.49</td>
<td>0.69</td>
</tr>
<tr>
<td>OMS 2.2 Java</td>
<td>0.94</td>
<td>105.2</td>
<td>0.88</td>
<td>167.9</td>
</tr>
</tbody>
</table>

6. DISCUSSION

To investigate relationships between model code quality and invasiveness, we used the approach recommended by Briand et al. (1999, 2000) and Basil et al. (1996) to use Chidamber and Kemerer’s (1994) object-oriented software metrics as indirect measures of software quality. Using fan-in/fan-out coupling as an inverse surrogate of software quality, we found that framework implementations for both models having the lowest invasiveness measures for FDT-uses and FF-uses also had the lowest values for fan-in/fan-out coupling ($p = 0.002, 0.011; df = 5$). Framework implementations for both models with higher fan-in/fan-out coupling used more framework functions and data types. This relationship suggests that more invasive model implementations may exhibit lower code quality. We also found that model framework implementations with low invasiveness measures for FDT-uses and FF-uses also had the smallest code sizes (LOC) ($p = 0.024, 0.024, df = 5$) and total cyclomatic complexity (total CC) ($p = 0.0007, 0.0007, df = 5$). Models with larger LOC and total CC also used more framework functions and data types. The results indicate that framework-based model implementations which used the most framework functions and data types had larger size, complexity and more coupling. These relationships suggest a negative relationship between framework invasiveness and software quality.

The Thornthwaite and PRMS model implementations under the lightweight OMS 3.0 framework had lower framework to model invasiveness (Tables 4 and 7). Additionally, the Thornthwaite and PRMS model implementations under the OMS 3.0 framework had lower overall code size (LOC), cyclomatic complexity, and afferent (fan-in) and efferent (fan-out) coupling (Tables 3 and 6). It appears that a lightweight framework based modeling approach produces both smaller and simpler model implementations.

7. SUMMARY AND CONCLUSIONS

This paper presents a unique comparison of environmental modeling framework invasiveness using the Thornthwaite and PRMS hydrologic models. Our results showed that less invasive model implementations tended to have higher code quality as observed in terms of code size, complexity and coupling. Models implemented using the OMS 3.0 had the lowest invasiveness scores and the smallest size, complexity, and coupling. For the Thornthwaite model, the OMS 3.0 implementation was on average 40% as large as the heavyweight framework implementations and about 30% as complex. For the PRMS model, the OMS 3.0 implementation was 40% smaller and about 30% as complex as the heavyweight OMS 2.2 framework implementation. Overall, the OMS 3.0 framework produced less invasive model implementations when compared to the heavyweight
framework implementations using OMS 2.2, ESMF 3.1.1, OpenMI 1.4, and CCA 0.6.6. In conclusion, the lightweight framework approach to environmental modeling appears to produce smaller, less complex models with less coupling and framework-to-model invasiveness. Based on this result, a lightweight framework approach to environmental modeling appears to help modelers develop higher quality and more concise model implementations. For environmental modeling, this lightweight framework design approach seems further attention.

ACKNOWLEDGMENTS

We would like to acknowledge Cecelia DeLuca from UCAR, a member of the ESMF support team, who provided a code review of the ESMF FORTRAN Thornthwaite model implementation and also Dr. Andrea Antonello who provided a code review of the OpenMI 1.4 Java Thornthwaite implementation.

REFERENCES

Ahuja, L.R., Ascough II, J.C., and David, O., Developing natural resource modeling using the object modeling system: feasibility and challenges. *Advances in Geosciences*, 4, 29-36, 2005.


From OpenMI 1.4 to 2.0

Peter Gijsbers1,2, Stef Hummel1,3, Stanislav Vaneček1,4, Jesper Groos1,5, Adrian Harper1,6, Rob Knapen1,7, Jan Gregersen1,8, Peter Schade1,9, Andrea Antonello1,10, Gena Donchyts1,11

1 The OpenMI Association, Lelystad, The Netherlands
2 Deltares USA Inc., Silver spring, Maryland, USA, Peter.Gijsbers@deltares-usa.us
3 Stichting Deltares, Delft, The Netherlands, Stef.Hummel@deltares.nl;
4 DHI, Prague, Czech Republic, S.Vanecek@dhi.cz
5 DHI, Horsholm, Denmark, jgr@dhigroup.com
6 MHW Soft, Wallingford, UK, Adrian.Harper@wallingfordsoftware.com
7 Alterra, Wageningen, The Netherlands, Rob.Knapen@wur.nl
8 Hydroinform, Kirke Hyllinge, Denmark, Gregersen@hydroinform.com
9 Bundesanstalt für Wasserbau, Hamburg, Germany, Peter.Schade@baw.de
10 University of Trento, Italy, Andrea.Antonello@ing.unitn.it
11 Stichting Deltares, Delft, The Netherlands, Gena@deltares.nl

Abstract: The Open Modelling Interface (OpenMI) was launched end of 2005 with the aim to become a global standard for linking models and tools in the environmental domain with focus on water. Over the past few years, the user and development community has grown substantially and various well known models have become compliant. Some of the uses did not adopt the OpenMI standard interfaces completely, but used a slight deviation to achieve their goal in a similar style. Improvements would be necessary to become a true global interface standard instead of a style for developing new model codes. Starting in 2007, a core group of six institutes has worked on an upgrade of the OpenMI towards version 2.0. A long list of deficiencies was composed, having a few use cases as general guidance for improvement. The proposed redesign, based on similar leading concepts and a similar data model, required however a non-backward compatible upgrade of the interface standard to remove the weak points from the first version. This decision allowed the OpenMI to become more suitable for a larger range of applications, from non-time dependent Geographical Information Systems (GIS) towards e.g. master-slave controlled modelling frameworks. The OpenMI 2.0 standard has been open for review early 2010. Once completed and processed, the release of OpenMI 2.0, in both C# and Java is expected late 2010. This paper will discuss the reasons for change in more detail and highlights how the proposed solution meets the needs in a better way.

Keywords: Open Modelling Interface; linking; model interoperability; integrated modelling

1. THE USER COMMUNITY OF OPENMI EXPANDS

The Open Modelling Interface (OpenMI) was launched end of 2005 with the aim to become a global interface standard for linking models and tools in the water domain. A scope that later was expanded to the environmental domain with focus on water. Over the past few years, the user and development community has grown substantially and various well known models have become compliant. Some of the uses did not adopt the OpenMI standard interfaces completely, but used a slight deviation to achieve their goal in a similar style. Improvements would be necessary to become a true global interface standard instead of a style for developing new model codes. Starting in 2007, a core group of six institutes has worked on an upgrade of the OpenMI towards version 2.0. A long list of deficiencies was composed, having a few use cases as general guidance for improvement. The proposed redesign, based on similar leading concepts and a similar data model, required however a non-backward compatible upgrade of the interface standard to remove the weak points from the first version. This decision allowed the OpenMI to become more suitable for a larger range of applications, from non-time dependent Geographical Information Systems (GIS) towards e.g. master-slave controlled modelling frameworks. The OpenMI 2.0 standard has been open for review early 2010. Once completed and processed, the release of OpenMI 2.0, in both C# and Java is expected late 2010. This paper will discuss the reasons for change in more detail and highlights how the proposed solution meets the needs in a better way.

Keywords: Open Modelling Interface; linking; model interoperability; integrated modelling
in the field the impression that OpenMI is another framework, similar to e.g. TIME, CSDMS, ESMF or OMS. The OpenMI Association has not the intention to provide a framework, but deems an open source software implementation necessary to reach world wide adaptation of the standard, given its current development and acceptance phase.

To take the OpenMI from a research output to an operationally viable technology, the OpenMI Life started under the EU-Life programme (LIFE06 ENV/UK/000409). In this project, water management agencies in Belgium, the Netherlands and Greece applied the OpenMI to connect their models from different suppliers and conduct crosscutting analysis on real world issues that require integrated assessments.

In order to support this process, a legal association was established to take ownership of the OpenMI and stimulate the use and support for the OpenMI: The OpenMI Association. After some initial hurdles to understand what it takes to deliver an open source standard and SDK, the confidence grew outside the core group that the OpenMI was a highly potential technology for the larger community. Other parties started joining the Association and contributed to the development. Alterra (part of Wageningen UR, NL) adopted the OpenMI in various EU-projects (e.g. SEAMLESS, SENSOR), developing and launching an open source Java implementation of the OpenMI, with support from Univ.Trento (I). The Bundesanstalt für Wasserbau (BAW, GER) ported the C# implementation from Microsoft .Net to the Mono platform and tested this implementation on a Linux based proprietary file based database to exchange initial and boundary data with the (Windows based) three-dimensional flow model Delft3D [Schade et al. 2008]. New public models have become OpenMI compliant (e.g. SWAT, migrated by UNESCO-IHE) while the USACE-HEC and USGS are collaborating towards an OpenMI compliant implementation of HEC-RAS and Modflow. CUAHSI has demonstrated that its WaterOneFlow web service can be connected to OpenMI compliant models [Goodall et al. 2007] and are keen to explore model-based web services using OpenMI as an important vehicle. CSDMS, has adopted the OpenMI as its model interoperability interface using the Common Component Architecture (CCA) as implementation framework for its HPC-model platform [Peckham, et al. 2009]. USDA has implemented the OpenMI interface on top of its Object modelling System (OMS) [David et al. 2009]. CSIRO is experimenting with OpenMI while US-EPA has announced that the new version of its multi-model multi-media framework, Frames 3, will be using the OpenMI interface. Finally, the Open Geospatial Consortium (OGC) shows a keen interest in OpenMI as a means towards model interoperability services. In addition to those known users, i.e. known to the Association, many other users exist. Bulatewitz et al. [2009] demonstrates that, even without questions to the help forum, people take the OpenMI concept and build integrated multi-disciplinary modelling applications with it.

Some of those uses did not adopt the OpenMI.Standard interfaces completely, but used a slightly deviation to achieve their goal in a similar style. This indicated room for improvement, an observation that was supported by experiences in the core group as well. This paper will take those experiences as a starting point to discuss in general terms the changes that have been made to make OpenMI more flexible, more efficient and more intuitive for integrated modelling applications in the wider environmental domain.

2. USER EXPERIENCES EXPRESSING A DESIRE FOR IMPROVEMENT

2.1 User experiences with OpenMI 1.x

OpenMI 1.x was developed with an extensive focus on numerical models, primarily in the field of hydraulics and groundwater-surface water interaction. Gregersen et al. [2010] demonstrate such application. Although notice was taken of the needs of other domains, e.g. economics, this hardly influenced the outcome of OpenMI version 1.

After OpenMI 1.x was released and OpenMI was being used in real practice, many successful applications have been developed. Simultaneously however, a few design decisions became apparent as being not well equipped to meet the desired flexibility and performance. An important nuisance appeared to be the link object, i.e. the interface that
defines what data is exchanged between two components. The link had no clear ownership, thus creating a confusing situation when adding data operations. Another less favourable feature of the link was the need of a full interface implementation on both sides of a link, also for potentially simple components such as data viewers. The link object also stimulated memory resource intensive implementations, as each link would typically buffer the source data it needed to support the data request from its target. Viewing the data being passed over a link to a model actually required instantiation of a second similar link, often resulting in the same data being duplicated in memory to support the second link. Especially with larger grids, the memory consumption could become substantial. The link also played a prominent role in the request for data (the GetValues method). Unfortunately, this call referred to a link by its linkId, introducing a performance hit at each call when the internal list of links became large. Finally, the link design also stimulated the use of a non-intuitive trigger object, an object which often was added to create a ‘trigger’ link for starting the data exchange and computation process.

Within OpenMI, data is exchanged on a request-reply basis, in other words: data is pulled to a component. This pull-concept enabled Goodall et al. [2007] to implement OpenMI as a web-service, hence demonstrating its applicability as a service-oriented architecture. However, the implementation as a modelling web-service was not intuitive as OpenMI 1.x only allowed values being exchanged per timestamp, and not as a time series object in one go [Goodall et al. in press]. Such capability would be highly desirable for web-service implementations to reduce the number of calls needed over a network.

Prototype applications with OpenMI 1.x illustrated that data assimilation [Seymour, 2005] and decision support applications [Dirksen et al., 2005, Knapen et al., 2009a, b] could be developed with OpenMI, but the pull-based concept was not very fit for purpose. These applications typically conduct multiple model runs, redefining the input after evaluation results from a previous run. A ‘set value’ approach would be natural, but OpenMI 1.x required an unnatural call back construct to pull new input into the engine.

OpenMI 1.x provides optional interfaces to implement state management. Unfortunately, the interface could not guarantee support for persistent state management. So far, few applications have been using state management functions, partly since its application was mostly limited to engines iterating back and forth to arrive at numerical stable conditions. Generation of multi-model restart files, desirable to prevent numerical instability of complex interacting models at start-up, could not be facilitated. The same applies to forecasting systems requiring frequent model restarts from ‘warm states’.

Application in the SEAMLESS and SENSOR projects [Knapen et al. 2009a, b, Wien et al. 2009, Verweij et al. 2007] illustrated that the OpenMI data model was not very suitable for interdisciplinary modelling applications requiring exchange of non-numerical data. Furthermore, its tight connection with the concept of time as well as its over elaborated spatial data model made OpenMI 1.x not very attractive for combination with Geographic Information Systems or databases holding non-transient data.

For above reasons, various OpenMI users deviated from the Standard specification in the applications. Given these experiences, the OpenMI Association Technical Committee (OATC) proposed to the OpenMI Association Executive Board to develop a non-backward compatible upgrade to overcome some of the issues identified. This redesign, conducted over the past three years, was guided by the user experiences and an integrated modelling domain analysis to highlight the needs that had not sufficiently been met by OpenMI 1.4.

2.2 Domain analysis

The OATC started the development of the next version of the OpenMI with an analysis of experiences, focussing on domain modelling characteristics in (to be) integrated modelling disciplines. Its results are expressed in an abstract perspective focussing on spatial and temporal characteristics of various component types as illustrated in Table 1.
Table 1 Characteristics of selected component types

<table>
<thead>
<tr>
<th>Component Type</th>
<th>Spatial characteristic</th>
<th>Temporal characteristic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numerical model</td>
<td>ID-based and/or geo-referenced (x, y, z)</td>
<td>proceeds in time, may step backwards for numerical stability</td>
</tr>
<tr>
<td>Constant value provider</td>
<td>constant in space</td>
<td>constant in time</td>
</tr>
<tr>
<td>0-D time series (boundary condition) provider</td>
<td>ID-based</td>
<td>data varies over time</td>
</tr>
<tr>
<td>Geographical Information Systems</td>
<td>geo-referenced</td>
<td>constant in time</td>
</tr>
<tr>
<td>Tools for Optimization, Calibration and Scenario Management / Decision Support</td>
<td>with/without geo-reference</td>
<td>proceed in time</td>
</tr>
<tr>
<td>Analytical functions</td>
<td>exact geo-referenced (without interpolation)</td>
<td>exact in time (without interpolation)</td>
</tr>
<tr>
<td>Database or reader</td>
<td>with/without geo-reference</td>
<td>values of all time steps are available from the start</td>
</tr>
</tbody>
</table>

As indicated, OpenMI 1.4 was relatively strong in support of numerical models, but required considerable improvement for the other types. The major changes are:

- The link object is replaced by a direct reference to the consumers respectively producers of data.
- Exchange items get a new role as the port for data provision and consumption.
- Time has become an integrated part of the query for data.
- Progress in time is less tightly bounded to the call for data transfer.
- Enhanced support for non-numerical data.

3. CHANGES FROM OPENMI 1.4 TO 2.0

3.1 Replacing the link - A new role for exchange items

In OpenMI 1.4 the link object was used to connect components, containing a reference to the source component and the target component, as well as to the source output (exchange) item and the target input (exchange) item. Exchange items had been introduced to describe the data being exchanged in terms of ‘what’ (IQuantity) and ‘where’ (IElementSet). To make output items compatible with input items, a series of data operations (IDataOperation) was specified in the link object. Experience has shown that the link object was not very efficient and intuitive for many applications. ‘Ownership’ of the object was unclear, therefore creating confusion how additional data operations could be implemented.

Adopting a port concept would make the addition of data operations much more intuitive, especially if a chain of data operations could be implemented on top of a port. In addition, a port concept would make connecting a (non-compliant) data viewer to an OpenMI component much easier. Many model providers used the internal exchange item object for holding the data that the IExchangeItem interface described. This interface thus would offer an intuitive place to act as a port for data provision and consumption. Two types are identified with different tasks. InputItems provide an entry point for data input, while OutputItems provide an exit point for output. Data viewers only need to implement the input interface such that they can register themselves as a consumer to output of another component. Data viewers thus do not need to implement all data providing capabilities.

In OpenMI 2.0 a connection between a source component and a target component is realized by a direct reference between an output item and an input item. The GetValues method, called for actual data transfer, has been moved from the LinkableComponent level to the exchange item level. The link, as a separate interface, has been removed from the specification. If the values of an output do not exactly fits the required input, an ‘adapted’ output needs to be provided. The adapter pattern (similar to a decorator pattern) is used for
this purpose, enabling the addition of data-operations in between the original output and input item, realizing a truly piping and filtering pattern.

The adapted outputs take over the role of data operations in OpenMI 1.4. Typically, such intermediate adapted outputs are spatial and time interpolations and unit conversions. The sequence of adapted outputs defines explicitly in which order the operations are carried out. The adapted output can, via the Refresh() method, be notified that the values of its parent output may have been changed.

An important advantage of the ability to ‘stack output adaptations’ is the ability to reuse adapted outputs for various consumers (Figure 1). E.g. a model requiring point input data at different locations from a grid model, forced the grid model - under OpenMI 1.4 - to hold two internal copies of the grid data. With version 2.0, the source model only needs to hold an internal copy of the interpolated points. Similarly, data viewers monitoring a connection do not enforce a duplication of the link (and its data) anymore, as they can directly tap the same adapted output as the direct model-to-model connection.

Finally, the adapted output mechanism enables the user to define the order of a series of operations. By performing the operation with the greatest data reduction first, the subsequent operations can significantly be sped up.

![Figure 1 Re-use of adapted outputs](image)

### 3.2 Reformulating the query for values

In OpenMI 1.4, values were asked from a linkable (model) component for a certain time and for a certain link through the GetValues call. The component had the responsibility to return the values asked as specified by the target quantity and target element set of the link. Components that had no temporal knowledge still were enforced to deal with time. This annoyance has been removed.

In OpenMI 2.0, the link identifier is gone, as a direct relation between producers and consumers defines the connection between components. The query specifies the value definition and either the element set (space), the time set or both. The query is passed as an InputItem argument in the GetValues call on the OutputItem. In this manner, all variations in component types - as identified in the domain analysis - can be supported with their specific query needs. Components dealing with time can be informed beforehand at what time stamps the consumer requires data. This allows a providing component to prepare itself e.g. by creating appropriate internal buffers or by computing ahead if possible.
3.3 Redefining the data model

Users requests have also resulted in some adjustments and extensions of the data model in order to accommodate non-numerical data, to improve interoperability with OGC-feature types and to enable retrieval of time series objects in one go.

To support non-numerical data, the values have changed from doubles - representing a numerical scalar or (xyz) vector - to objects. The value definition has been extended from numerical quantities to both quantities and so-called qualities, the latter being categorized items enabling representation of e.g. soil types, land use, seasons and fuzzy values.

The spatial domain of an OpenMI component is described by the element set. IElementSet is an overarching interface allowing high-performance access to individual elements of a specific element type. The list of element types is GIS-oriented, but the separation between the horizontal and vertical dimension, as defined in OpenMI 1.4, showed to be overdone. Both the GIS-world and the numerical modelling world did not fit into this design. Hence, the number of element types has been reduced and is harmonized with the feature types as common in the GIS-world. For numerical modellers, both Z-coordinates (geo-referenced) and M-coordinates (model layer reference) are supported.

The actual data values being exchanged are held in a so-called value set. For OpenMI 2.0, the IValueSet interface has been redesigned as well. New methods have been introduced to retrieve the value objects for all element values for a specific time or for all time series values of a specific element. In addition, the interface provides a method to set values into the object. This method, combined with the accessibility of a value set as a property of an input exchange item, allows external components to set a value if needed. Tools used for data assimilation, calibration and optimization will benefit from this capability, while the dominant data exchange mechanism remains pull-based.

3.4 Extending control flow options

In OpenMI 1.4, the GetValues call served various purposes: data transfer, progress in time and triggering calculations at other components when needed to progress. The control flow was purely pull driven, since linkable components were triggered to perform an action when there was a request to produce data.

In OpenMI 2.0 this is pull mechanism is still the same, with components actively triggering other components when they need data. However, a second control flow option, called the ‘loop’ approach, has been added to accommodate progress control from the outside. An external controller can invoke each linkable component separately, by calling its Update() method. This will let the linkable component progress to its next step, but only if all necessary data for that component is available. If the data is not available yet – e.g. because another component has not computed it yet - the component specifies its state as “waiting for data”, and returns the control flow. During the next Update() call, the component checks whether the required data is available, (and if so) computes the new value and sets its state to “updated”. A typical application of this approach could be master-slave computational system where a master component controls the computation sequence of its various computational domains enabling parallel computing.

Note that the loop approach doesn’t offer wider usage of linked components, it is only an alternative way of controlling the data flow. Implementing the loop approach is optional. A linkable component that does not support it is still OpenMI compliant.

3.5 Enabling persistent states

While OpenMI 1.4 allowed models to manage their state and roll back to a previous state, it was up to the developer to implement this in memory or with so-called restart files. Various use cases would benefit if OpenMI could enforce a persistent state to be saved and restored. A complex model combination could reduce its numerical instability problems at start-up if the various components could save a restart file on request when the shared system is in balance. In addition, operational forecasting systems need this ability to keep the latest
model state update-to-date. OpenMI 2.0 therefore accommodates conversion of a model state from a memory object to a persistent byte stream and back.

3.6 Improving component status information

OpenMI 1.4 contained a message mechanism to broadcast component status. This design has shown to impact performance as the composition of messages could take substantial time as compared to the calculation time. In OpenMI 2.0, status is handled as property of the linkable component. The states can be used in both the pull-based as the loop-based control flow to make decisions. The available states can be defined in two types: action states (e.g. updating) and idle states (e.g. updated). States are related to the call sequence conducted on the component.

4. CONCLUSIONS AND FINAL REMARKS

The redesign presented in this paper, supported with more technical details in Donchyts et al. (2010), is expected to be a major step forward towards improving the applicability of OpenMI as an Application Programming Interface for interoperability between computational models, databases, geographic information systems and web-services. Its updated interface specification is much more in line with the actual implementation at computational cores, while it simultaneously has improved capabilities for utilization in GIS or web service environments.

However, not all requests and wishes are supported yet. Additional funding and resources will be needed to test the OpenMI 2.0 interface design in a High Performance Computing environment in order to decide if extensions are essential. Similarly, more experience needs to be gained in the utilization of ontology in combination with OpenMI 2.0, based on the initial steps taken in the SEAMLESS project [Knapen, 2009a,b]. Furthermore, new web-service experience needs to be gained to leverage the capabilities of the OpenMI in this field if needed.

The expectation is that extending the available or implementing new Software Development Kits can facilitate most of these requests. Most likely, some extension will be required to the interface standard to accommodate specific needs. However, the current interface design is foreseen to be sufficiently mature, such that any of those requests does not need to result in another non-backward compatible upgrade.

ACKNOWLEDGEMENTS

The development of OpenMI version 2.0 has been funded by the members of the OpenMI Association as well as the European Commission through various 6th Framework Programs and the LIFE Environment program.

REFERENCES


Goodall, J.L., Robinson, B., and Castronova, A.M. (in press), Modeling complex water resource systems with service-oriented computing. *Submitted to Environmental Modelling & Software*


Schade, P.; Lang, G. and Jürges, J.; (2008) OpenMI Compliant Import of Initial and Boundary Data into a numerical 3D Model. *In: Proceedings of the iEMSs Fourth Biennial Meeting: International Congress on Environmental Modelling and Software (iEMSs 2008) Vol. 2, pp 1086-1093*


Abstract: The Land Suitability Rating System for Canada (LSRS) assesses climate, landscape, and soil factors in order to rate the suitability of land for growing crops. Currently, the list of crops includes alfalfa, brome, canola, corn, soybeans, and spring-seeded small grains. LSRS has been under development since the mid 1980’s, and implemented in a variety of desktop software. Recently, the model was rewritten and deployed as suite of web services, based in part on recent and emerging standards from the Open Geospatial Consortium. The standards include the Web Processing Service (WPS) and the Table Joining Service (TJS). The use of these data exchange and web service standards has made it easy to feed climate change scenarios into LSRS, and to use outputs from the LSRS model as inputs for other models designed to assess socio-economic effects. This paper will give a high-level overview of the LSRS, introduce the modelling framework and web service standards that are used, and demonstrate the benefits of this approach using examples from the LSRS implementation. It describes a generic standards-based approach to implementing environmental models in such a way that they can be easily integrated into more complicated systems.

Keywords: modelling; web services; XML; WPS; TJS; LSRS.
The rest of this introduction describes the LSRS and the two OGC standards that are used in its implementation, namely WPS and TJS.

1.1 The Land Suitability Rating System (LSRS)

Development of the LSRS was started by AAFC’s Land Resource Research Center in 1987. By 1995 it had lead to the production of the basic processing approach and a manual for assessing the suitability of land for growing spring-seeded small grains such as wheat barley and oats (see Pettapiece [1995]). The calculation was subsequently extended to other crops, including alfalfa, brome, canola, corn, and soybeans.

The LSRS calculation rates the suitability of a particular location on the earth’s surface for growing crops, using an expert system approach to assess climate, soil, and landscape factors on a crop-specific basis. The system interprets commonly available environmental measurements in terms of their limiting effect upon crop production, and ranks land according to a scale similar to the Canada Land Inventory (CLI) ratings that have been used for land use zoning decisions in Canada since the 1960’s.

LSRS was initially programmed in dbase. Later, it was rewritten as an extensive set of spreadsheets controlled by Visual Basic program. Performance was an issue in both cases, as was scalability, but more importantly the management of data inputs and outputs was quite awkward in both of these prototypes.

1.2 The Web Processing Service (WPS)

In 2007, the OGC released a new web service interface specification designed to enable spatial calculations and modelling on the Internet. This specification is known as the Web Processing Service (OGC [2007]). A WPS can be used to enable access to any sort of calculation or model, although it is primarily intended to operate on spatially referenced data. The data required by the service can be available locally, or delivered across a network. There is no limit on the complexity of model which can be delivered via WPS.

WPS specifies a standardized way to make a model available as all three conventional types of web services, namely KVP GET, XML POST, and as SOAP. As such, it allows the client to choose their preferred method of implementation. WPS essentially defines a way to wrap a model so that it can be executed over a network. It has been implemented in JAVA, Python, and Ruby, and has recently started to become popular in the geospatial GRID computing community. Since WPS is purely a web interface wrapper, the underlying model can be retained in its original programming language.

A model in WPS is known as a “process”. A single WPS instance can be used to implement multiple processes. WPS defines three operations. Service discovery is provided through the GetCapabilities operation, while complete documentation of the supported inputs and outputs is obtained via the DescribeProcess operation. Models are run via the Execute operation.

WPS can handle inputs in a variety of formats that are specified by the individual implementation. Inputs can be local to the server, embedded in the execution request, or available from remote servers. It provides a way to specify the form of output required, a way to view the completion status of a process, and a way to store the outputs of a process.

The response from a WPS execute request contains an optional “Lineage” section, which encodes the request parameters sent to the service. This means that outputs from the WPS contain sufficient metadata to be useful long after processing has been completed.
1.3 The Table Joining Service (TJS)

The OGC is currently working on the Table Joining Service (TJS). This candidate standard was published in a Request for Comments last year under the name Geographic Linkage Service (GLS) (OGC [2009]).

TJS specifies an XML encoding of tabular polygon attribute data that is known as GDAS (Geographic Data Attribute Set). This encoding includes a substantial amount of metadata that completely describes the data contents. GDAS is thus useful as a way to store and exchange data suitable for polygon-based modelling.

2. APPROACH TO WEB-ENABLED MODELLING

This section describes the principles and general approach for web-enabling environmental models that has been used to implement the LSRS and related models and data services at AAFC.

2.1 Goals

The implementation of models as networked services must:

1. Simplify the running of environmental and economic models
2. Enable model chaining (i.e. facilitating the feeding of outputs of one model to the next in a series of models)
3. Store model outputs for later examination or reuse
4. Minimize effort in terms of development, testing, and operating the model.

2.2 General Approach

In order to meet the aforementioned goals, the following general approach to implementing models was used.

1. Make each model available as a separate service over HTTP
2. Store all data inputs and outputs at web-accessible URLs
3. Identify data to be used in for model runs via these URLs
4. Encode or wrap data to be exchanged between models using XML
5. For batch runs, wrap data to be exchanged between models via self-documenting XML (GDAS)
6. Use native code where models are already programmed and well tested. Develop new code only where absolutely necessary.

2.3 Web Service Interfaces

Models and associated data services are implemented according to the following guidelines:

1. Use simple web service interfaces (preferably HTTP GET/POST with KVP)
2. Control model runs using scripting languages (e.g. Ruby)
3. Provide an interface description for the web service interface for each model, according to the following conventions:
   a. service descriptions are obtained by a call to the service root (e.g. http://foo.bar/foo)
   b. service descriptions are either HTML or WPS Process Description documents in XML
   c. services are found at http://foo.bar/foo/service
   d. clients are found at http://foo.bar/foo/client

2.4 Model Controllers
For each model, implement both a single calculation and a batch calculation capability.

The single calculation capability is used for debugging and validating the model. It includes:
1. A web page that allows a user to specify the model inputs for a single polygon (or small region of cells)
2. A web service that takes those inputs and runs the model for that polygon
Make the level of detail provided in the output from the service selectable via the service interface. For debugging purposes, the output should list the inputs, as well as interim results and error conditions encountered. When called by a batch job, the output is pared down to the bare minimum necessary to return the model results.

The batch calculation capability is used for running a model on multiple polygons. It includes:
1. A batch client web page that allows a user to specify the model inputs for the run. It normally
   - sends a request to the batch web service
2. A batch web service that prepares the model run. It normally
   - creates a control file for the batch processing service
3. A batch processing service that will run the model on the specified collection of polygons (or a large region). It normally
   - obtains the model inputs via web or database requests
   - runs the model on each identified polygon
   - updates a batch status document both during and after processing (using WPS status documents)
   - stores the model output(s) as web accessible resources (using GDAS or some other format as appropriate)
4. A batch job controller web service that allows a user to
   - examine the contents of all batch control files
   - view the status of pending, current, and completed batch jobs
   - view results of a batch job
   - delete a batch job

Chaining of multiple models in sequence is done either by a custom script, or, in the case of a widely accessible service, by setting up a separate batch controller for the purpose of running the integrated model.

3. LSRS IMPLEMENTATION DETAILS

The LSRS algorithm was implemented in Ruby, according to the principals outlined in section 2. Details of that implementation are detailed below.

3.1 The LSRS Calculation

The LSRS individual rating calculation was implemented as a simple HTTP GET process. The LSRS calculation was deliberately not wrapped in WPS because there was no apparent benefit to doing so. The calculation itself is not supposed to be discoverable or bindable by other models, and the process executes so quickly there is no reason to ever want to store the outputs.

The model itself is implemented as a simple KVP GET service - it was simpler that way, and the LSRS model uses a large number of relatively static inputs, including soils and landscape data, for each calculation. It would have been prohibitively slow to pass all the inputs via a web interface - instead, it uses a direct database connection to get the bulk of its data inputs. The exceptions are the parameters for which the calculation should be run, including the soils dataset name, the polygon number, the type of crop being assessed, and
the climate inputs. With the exception of the climate inputs, these other inputs are all included as part of the request to the LSRS calculation service.

3.2 Climate Inputs to LSRS

The climate inputs to the LSRS are obtained from a custom web service that provides climate data. This service is again a custom KVP GET service interface that includes an SLC number as part of the request, and responds with a custom XML document containing the climate data for that SLC polygon. A web service interface was used in order to make it possible to connect to different climate scenarios produced by models running on other servers. This approach makes it possible to chain together climate change models with LSRS into a super model that can evaluate the effects of climate change scenarios on the suitability of land for growing crops.

3.3 The LSRS Batch Processor

The batch calculation of LSRS is the manner in which other models are expected to interact with LSRS. For this reason, the batch calculation was implemented using OGC standards including the WPS standard and the emerging TJS candidate standard. The batch process is itself is based on WPS, while the output from the LSRS calculation requests is collated as a TJS GDAS document and returned to the user, or, in the case of larger runs, stored on the server. The GDAS response is rendered into human readable HTML using XSL.

The LSRS batch system takes advantage of WPS capabilities, because for small runs it returns the output directly, while for larger runs it store the output and updates a status page. Instead of implementing the entire WPS specification, just the core of WPS was used: namely the Execute response behaviour and the response document encoding.

4. CONCLUSIONS AND RECOMMENDATIONS

4.1 Web service based environmental modelling is feasible and advantageous

This implementation of LSRS has demonstrated that it is feasible to wrap environmental models as web services. Significantly, the use of web services made it easy to develop the different modules that comprise LSRS, and to have users directly engaged in the debugging of the model with just the use of a web browser. Although other models have not yet been chained to LSRS, their presence has been simulated through the use of static web-based data sources, and there appears to be no reason why the approach described in section 2 above is not eminently scalable and effective.

The implementation of models using web services forces the implementation of code as reusable components, and it also forces the testing of such components. Furthermore, once such components have been developed and tested, there may not be any need to further modify the code. This isolation of tested and deployed code inherent to web services appears to confer a significant advantage over the development of systems based on a more conventional monolithic single-server approach, because it reduces the opportunity to introduce inadvertent changes to functioning code.

4.2 The implementation of WPS saved design work

The WPS standard was designed to work as a general purpose wrapper for spatial calculations and models. The experience of implementing LSRS showed that it works extremely well. Although the batch processor could have been implemented with completely custom interface, the choice of WPS meant that the behaviour was well-defined, and the coding simply involved following the WPS recipe, thus saving a significant amount of design work. The self-documenting nature of the WPS response documents meant that it
was trivial to generate human-readable documentation from the results of past processing runs.

4.3 Partial implementation of WPS saved work

WPS is a highly flexible and standardized way of wrapping a web service. It brings significant coding advantages even without the use of the “GetCapabilities” and the “DescribeProcess” operations that would be required for its complete implementation. Not implementing these aspects of WPS meant that the batch service was not discoverable, but that was of little consequence in this case because the users are well known and can be directed to the service without the use of a registry.

4.4 WPS could be simplified

For really simple implementations, the OGC should consider simplifying the KVP GET interface such that its construction does not require the presence of the “Service”, “Request” and “Version” parameters, and the values of the process inputs and outputs are provided as separate key value pairs and not double-encoded as in the current “DataInputs” and “ResponseForm” parameters.

4.5 Web service calls are surprisingly fast

When the processing time of a routine is measured in a few tenths of seconds or more, the additional overhead of implementing a web service wrapper can be essentially negligible. In the case of LSRS, the combination of a fast network and small packages of information, and a computing-intensive algorithm meant that the response time delays incurred by the use of a web service were essentially negligible. This would not necessarily hold true for systems built on networks with higher latency or with less demanding computations.

4.6 Fixed inputs should be delivered from a local datastore

Initially, access to the contents of the soils database was set up as an HTTP web service which returned XML, but performance was an issue. Performance remained an issue even when the lower overhead of a direct connect via Oracle SQLnet to a remote server was implemented. Ultimately, the data was delivered to the model via a MySQL instance housed on the same server that processed the LSRS calculation and called directly by the calculation. This configuration provided substantially better performance than a call to a separate server. In the case of the LSRS accessing soil layer records, network latency becomes a critical bottleneck, because each polygon calculation requires tens of calls to the soil database. Developers should be aware that static database content that is called frequently may not be suitable for delivery via a networked service.

5 REFERENCES


Schut / Implementation of the LSRS in a web-based open modelling framework


Integrated Agricultural System Modeling
Using OMS 3: Component Driven Stream Flow and Nutrient Dynamics Simulations

J.C. Ascough IIa, O. Davidb, P. Krausec, M. Finkc, S. Kralischc, H. Kipkaa, and M. Wetzelc

a USDA-ARS-NPA Agricultural Systems Research Unit, Fort Collins, CO 80526 USA (e-mail: jim.ascough@ars.usda.gov)
b Depts. of Computer Science and Civil and Environmental Engineering, Colorado State University, Fort Collins, CO 80523 USA
c Dept. of Geography, Friedrich-Schiller University, Jena, Germany

Abstract: This study reports on the integration of the European J2K-S model (a component-based system for fully distributed simulation of water balance and N dynamics in large watersheds) under the Object Modeling System 3 (OMS3) environmental modeling framework and subsequent evaluation of OMS3/J2K-S performance on the Cedar Creek Watershed (CCW) in northeastern Indiana, USA. Uncalibrated model performance for daily and monthly stream flow response was assessed using Nash-Sutcliffe model efficiency (ENS) and percent bias (PBIAS) model evaluation coefficients. Simulations for nitrogen (N) loadings to Cedar Creek were also performed; however, the OMS3/J2K-S N dynamics sub-model is still undergoing testing so a formal statistical evaluation of this component was not performed. Comparisons of daily and average monthly simulated and observed stream flows for the 1997-2005 simulation period resulted in PBIAS and ENS coefficients ranging from -18.6% to -8.6% for PBIAS and 0.46 to 0.68 for ENS. These values were similar or better than others reported in the literature for uncalibrated stream flow predictions at the watershed scale. The results show that the prototype OMS3/J2K-S watershed model was able to reproduce the hydrological characteristics of the CCW with sufficient quality, and should serve as a foundation on which to better quantify water quality (e.g., N dynamics) at the watershed scale.

Keywords: Hydrologic modeling; Watershed; Stream flow; Model evaluation; OMS3.

1. INTRODUCTION

The Object Modeling System 3 (OMS3) currently being developed by the USDA-ARS Agricultural Systems Research Unit and Colorado State University (Fort Collins, CO) provides a component-based environmental modeling framework which allows the implementation of single- or multi-process modules that can be developed and applied as custom-tailored model configurations (David et al., 2002). The value of continuous watershed simulation models like the Soil and Water Assessment Tool (SWAT) (Arnold et al., 1993) is reflected by programs like the Conservation Effects Assessment Project (CEAP) in the United States and the EU-Water Framework Directive (WFD) in Europe. The ARS CEAP Watershed Assessment Study (WAS) Project Plan (USDA-ARS, 2004) provides detailed descriptions of ongoing research studies at 14 benchmark watersheds in the United States. In order to satisfy the requirements of CEAP WAS Objective 5 (“develop and verify regional watershed models that quantify environmental outcomes of conservation practices in major agricultural regions”), a new watershed model development approach was initiated to take advantage of OMS3 modeling framework capabilities. The European J2K-S model (Krause et al., 2006), a component-based system for fully distributed simulation of the water balance and N dynamics in large watersheds and catchments, was selected to provide the initial process-based model components. Specific objectives of this study were to: 1) implement J2K-S hydrological and N
dynamics components under the OMS3, 2) assemble a new modular watershed scale model for fully distributed transfer of water and N loading between land units and stream channels, and 3) evaluate the accuracy and applicability of the modular watershed model for estimating stream flow and N dynamics. The Cedar Creek watershed (CCW) in northeastern Indiana, USA was selected for application of the OMS3-based watershed model.

2. THE OBJECT MODELING SYSTEM 3 (OMS3)

OMS3 closely resembles the generic architecture for environmental integrated modeling frameworks as presented by Rizzoli et al., 2008). It contains four primary foundations (Figure 1): modeling resources, the knowledge base of the system (e.g., metadata and ontologies), development methods and tools (e.g., science models), and various modeling products. The core consists of an internal knowledge base and development tools for model and simulation creation. OMS3 is based on the Java platform; however, it is highly interoperable with C, C++, and FORTRAN on all major operating systems and architectures. Model development under OMS3 is component-based and there are only minimal requirements for plain objects (POJOs) to be represented as an OMS3 component. In OMS3, as well as most other modeling frameworks, the term component refers to a concept in software development which extends the reusability of code from the source level to the executable. Components are context-independent, both in the conceptual and technical domain, and represent self contained software units that are separated from the surrounding framework environment. Furthermore, the OMS3 modeler does not have to learn and use framework data types or an extensive Application Programming Interface (API).

To allow for scalable models and processing with complex data sets, the execution of components under OMS3 is always multi-threaded (i.e., components are executed in parallel if the data flow allows it and no explicit thread coding is required). OMS3 utilizes the Domain Specific Language (DSL) concept to provide for a concise, robust, and flexible representation of model simulations. With easy to setup DSLs, simple simulations (e.g., model calibration, sensitivity and uncertainty analysis setups) can be created and executed in OMS3 from different environments such as IDEs, the OMS3 modeling console, the command line, or any application that embeds an OMS3 runtime version. Finally, OMS3 is a non-invasive modeling framework as there are no framework interfaces to implement, no classes to extend and polymorphic methods to overwrite, and most importantly no need to replace common native and custom language data types with framework-specific data types.

3. THE OMS3/J2K-S WATERSHED MODEL

The J2K-S modeling system (Krause et al., 2006; Krause et al., 2009) integrated in OMS3 was used for the simulation of the hydrological and N dynamics of the CCW in Indiana. J2K-S is a modular, spatially distributed system which implements hydrological and N processes as encapsulated process components and operates at various temporal and
spatial aggregation levels throughout the watershed. The J2K-S model was developed in the JAMS (Jena Adaptive Modeling System) modeling framework which is derived from an earlier, more API-based OMS Vers. 1. Therefore, the JAMS J2K-S model was already componentized and its classes followed the general modeling framework IEF (Init/Execute/Finalize) lifecycle-approach. For component integration into OMS3, framework-specific JAMS data types for component exchange-data were substituted with generic native data types and super class dependencies became obsolete. Moreover, JAMS-specific annotations had to be converted into more generic OMS3 annotations. The overall migration process was automated using scripts with regular expression string substitutions that automated roughly 90% of the JAMS to OMS3 conversion. The remaining 10% of the conversion was performed manually, consisting mainly of transforming JAMS model XML representations into Java code and DSL simulation files, introducing classes for complex (i.e., spatial) types such as hydrologic response units and stream reaches, and finally optimizing HRU processing for parallel execution under OMS3. Since JAMS and OMS are closely related with respect to component conceptualization and implementation, the migration process was limited to the source statement level, i.e., no structural change in components was necessary.

The OMS3/J2K-S hydrological model contains components for climate data regionalization, evapotranspiration, interception, snow accumulation and ablation, water and N balance in the unsaturated zone, water and N balance in the saturated zone, surface runoff and N concentration, and explicitly computed lateral surface and subsurface water/N routing and stream channel/river network flood routing in catchments (Krause et al. 2006). The N dynamics model contains components that are mainly adopted from the Soil Water Assessment Tool (SWAT) model (Arnold 1998) and coupled to the hydrologic components (Figure 2). The N dynamics modules include process components for simulating soil temperature, crop growth and N turnover according to Neitsch et al. (2002) and Williams et al. (1984) with some minor adaptations. Five different soil N pools are considered in order to allow modeling of different N inputs (e.g., mineral fertilizer, organic manure) and N transformations between these pools. N flows are modeled by a dynamic crop growth module and subsequent N uptake of plants (residues and yield), as well as through denitrification and volatilization. The land use management routines include modules for fertilizer management, tillage, and harvest operation (Krause et al., 2009).

![Figure 2. OMS3/J2K-S science component structure (adapted from Krause et al., 2009).](image)

After calculation of HRU surface and subsurface runoff and N dynamics, runoff and N routing is performed based on topological interconnections of the single HRU polygons, i.e., water and N flows are passed to a receiving HRU defined by its topological position (derived by GIS analysis), or to a receiving stream reach if the HRU is connected to one. Runoff and N routing inside the stream network is simulated by connecting the reach storages, receiving the runoff and N from the topologically connected HRUs by a hierarchical storage cascade approach, and calculating flow velocity inside the stream channel using the Manning-Strickler equation. The outflow of each specific stream reach is then transferred as inflow to the connecting downstream reach.
4. MATERIALS AND METHODS

The Cedar Creek Watershed (CCW) is located within the St. Joseph River Basin in northeastern Indiana, USA (41°10'10" to 41°32'38" N and 84°53'49" to 85°19'44"W). The CCW drains two 11-digit hydrologic unit code (HUC) watersheds, the Upper (04100003080) and Lower Cedar (04100003090), covering an area of approximately 700 km². The DEM data used in this study were obtained from the USGS at 10-m elevation resolution, 1/3 arc second, and a map-scale of 1:24,000 quadrangle sheet. The average land surface slope of the watershed is 2.6%, and the predominant soil textures are silt loam, silty clay loam, and clay loam with six STATSGO soil associations represented. The annual mean precipitation in the watershed area from 1989 to 2005 was 962 mm. For this study, a land use map from the National Agricultural Statistics Survey (NASS) was used collected between the dates of April 29, 2001 and September 5, 2001 with an approximate scale of 1:100,000 and a ground resolution of 30 x 30 m. Both standard ArcGIS 9.2 (ESRI, 2008) geoprocessing tools (e.g., overlay) and customized Avenue scripts for deriving HRU flow connectivity were used for HRU delineation which consisted of partly reclassifying and combining (by overlay analysis in ArcGIS 9.2) DEM topographical parameters (e.g., elevation, slope, aspect) with STATSGO soil and NASS land use GIS layers. The delineation of HRUs for the entire CCW resulted in 4,174 HRU polygons featuring areas between 0.02 to 2.5 km². Figure 3 shows the stream channels and HRU polygons of the CCW, together with topological connections as red arrows draped over the HRU polygons.

The OMS3/J2K-S simulation period in this study was 1997 through 2005. Daily precipitation, solar radiation, wind speed, relative humidity, and maximum/minimum air temperatures for these years were obtained from the NOAA National Climate Data Center (NOAA-NCDC, 2004) for the Garret and Waterloo weather stations within the CCW. Regionalization pre-processors in OMS3/J2K-S automatically distributed the climate data from the two gauges over the watershed. Historical measured data for Cedar Creek stream flow Gauge 04180000 (41°13'08"N, 85°04'35"W) were supplied by the USGS for the 9-year period from January, 1997 to December, 2005. Initial model parameter values were taken from simulation studies successfully applying J2K-S to watersheds in Germany and elsewhere exhibiting physical characteristics (e.g., topography, size, and agricultural land use) very similar to the CCW.

Figure 3. Routing topology with overland flow routing vectors for the CCW including an expanded view of flow routing vectors with HRU and stream channel flow linkages.

Nash-Sutcliffe Efficiency coefficient (E_{NS}) and percent bias (PBIAS) statistical evaluation coefficients were used to evaluate the overall correspondence of simulated output to measured values. E_{NS} indicates how well the plot of observed versus simulated values fits a 1:1 line. PBIAS is a measure of the average tendency of the simulated flows to be larger or smaller than their observed values. The optimal PBIAS value is 0.0; a positive value indicates a bias toward overestimation, whereas a negative value indicates a model bias toward underestimation. In this study, E_{NS} and PBIAS values were computed for both daily and average monthly stream flow.
5. RESULTS AND DISCUSSION

Historical measured data for Cedar Creek stream flow from the USGS for a 9-year period from January, 1997 to December, 2005 at Gauge 04180000 (41°13'08"N, 85°04'35"W) near Cedarville, IN was compared with daily and average monthly OMS3/J2K-S noncalibrated stream flow. In general, the OMS3/J2K-S model underestimated stream flow on a daily time-step as shown in the 1:1 plot in Figure 4 where all data points are included for the 9-year simulation period. The negative value for PBIAS (-18.55%) indicates that the model underestimated stream flow, and the $E_{NS}$ value (0.46) is considered unsatisfactory according to Moriasi et al. (2007) (although the PBIAS value is acceptable since it is under 25%).

![Figure 4. Daily CCW stream flow 1:1 plot of OMS3/J2K-S initial parameter set simulated values versus observed (Jan. 1997 to Dec. 2005).](image)

Average monthly observed and OMS3/J2K-S simulated stream flow from January, 1997 to December, 2005 are presented in Figure 5. This figure shows that the trend in simulated average monthly stream flow followed the observed values much more closely than the simulated daily stream flow results. Furthermore, it is extremely easy to discern that simulated average monthly stream flow in Figure 5 was significantly underestimated for nearly all of the 9-year simulation period. The $E_{NS}$ coefficient increased to 0.60 for average monthly stream flow (as compared to 0.46 for daily stream flow) with the average monthly PBIAS value remaining essentially the same as daily stream flow. The initial uncalibrated simulation results exhibited a rather large overprediction of ET on the watershed (data not shown) in addition to a systematic underprediction of stream flow.
Figure 5. Monthly CCW stream flow for observed and OMS3/J2K-S initial parameter set simulated values (Jan. 1997 to Dec. 2005).

across all time scales. Land use on the CCW is quite diverse, furthermore, the simplistic representation of evapotranspiration dynamics in OMS3/J2K-S may not adequately capture complex soil-water-plant interactions occurring on the watershed. Therefore, the soilLinRed coefficient was increased. This coefficient controls the partitioning of PET to AET, i.e., increasing soilLinRed decreases the amount of PET partitioned to AET. In addition, an attempt was made to account for areas of tile drainage on the Cedar Creek Watershed. A logical way to represent the effects of tile drainage in OMS3/J2K-S was to increase both the amount of water available in the LPS and the rate of outflow from LPS. Therefore, the soilDistMPSLPS and soilOutLPS coefficients were both decreased. Decreasing soilDistMPSLPS increases the amount of infiltrated water available for LPS; decreasing soilOutLPS increases the outflow rate from LPS. These adjustments approximate the more rapid removal of water from tile drains than what would normally be expected with the absence of tile drainage. All OMS3/J2K-S CCW simulations were then re-run using the modified values for soilLinRed, soilDistMPSLPS, and soilOutLPS.

All statistical evaluation coefficients for daily stream flow improved substantially for the modified parameter set, in particular, the ENS coefficient increased from 0.46 to 0.58 and PBIAS decreased from -18.55% to -8.59%. The OMS3/J2K-S model still underestimated stream flow on a daily time-step (as shown in the 1:1 plot in Figure 6). The ENS coefficient for average monthly stream flow improved for the modified parameter set (ENS = 0.68) as compared to the initial parameter set (ENS = 0.60). Average monthly improvement was of similar magnitude as the improvement in daily stream flow. Average monthly observed and OMS3/J2K-S simulated stream flow from January, 1997 to December, 2005 for the modified parameter set are shown in Figure 7. This figure shows

Figure 6. Daily CCW stream flow 1:1 plot of OMS3/J2K-S modified parameter set simulated values versus observed (Jan. 1997 to Dec. 2005).
that the trend in simulated average monthly stream flow for the modified parameter set followed the observed values much more closely (both in trend and in better estimation of peak stream flow events) than the simulated monthly stream flow results for the initial parameter set shown in Figure 5. Even with stream flow prediction improvements using the modified parameter set, OMS3/J2K-S underestimated stream flow at all time scales. Additional possible explanations for the underprediction may be attributed to using inappropriate values for recession coefficient parameter that govern simulated flow through the shallow and deep groundwater storage. Other studies (e.g., Krause, 2002) have shown the OMS3/J2K-S hydrologic model to be particularly sensitive to the recession coefficients used for groundwater storage calculations. Underprediction of monthly stream flow may be due to the lack of measured data for solar radiation and wind speed which are needed to estimate potential ET based on the Penman-Monteith equation in OMS3/J2K-S. Furthermore, the lack of available measured ET data for the study period makes it difficult to validate simulated ET results. Under or over estimates of ET could thereby affect the overall water balance, particularly during the summer months when ET demand is higher. Simulations for N loadings on each HRU and runoff N loading to Cedar Creek were also performed; however, the OMS3/J2K-S N dynamics sub-model is still undergoing testing so a formal statistical evaluation of this component was not performed. Figure 8 shows N pools simulated by OMS3/J2K-S averaged across all HRUs for the CCW.

In summary, we chose to evaluate noncalibrated stream flow results considering that OMS3/J2K-S was developed for applications on ungauged watersheds. More importantly, however, is the potential for formal model calibration to introduce a level of bias that
could ultimately mask or eliminate the impact of the simulated runoff generation processes.

6. CONCLUSIONS

The long-term continuous hydrologic simulations of OMS3/J2K-S performed reasonably well in predicting daily, monthly, and annual average flows on the Cedar Creek (Gauge 04180000) near Cedarville, IN. For initial and modified parameter sets, OMS3/J2K-S underpredicted the majority of the peak flows during the 9-year simulations of the Cedar Creek Watershed, with some individual storm events underpredicted by many orders of magnitude. Despite the underprediction, the majority of the evaluation statistics for $E_{NS}$ and PBIAS for both uncalibrated and manually adjusted parameter sets were within the range of other evaluation results reported in the literature for various watershed models such as SWAT. It was unclear whether OMS3/J2K-S needs enhancements in storm event simulations for improving high and peak flow predictions, or whether the distribution of rainfall over the entire watershed was misrepresented due to the use of only two climate stations.

The results show that the OMS3/J2K-S prototype watershed model was able to reproduce the hydrological dynamics of the Cedar Creek Watershed with sufficient quality, and should serve as a foundation on which to build a regionalized model for the CEAP initiative that is able to quantify the impact of conservation practice implementation on water quantity and quality at the watershed scale. In particular, the topological routing scheme employed by OMS3/J2K-S (thus allowing the simulation of lateral processes important for the modeling of runoff concentration dynamics) is much more physically based and robust than quasi-distributed routing schemes used by other watershed scale natural resource models (e.g., SWAT). The largest advantage of the OMS3/J2K-S routing approach is a process-oriented view of spatial watershed characteristics that drive hydrological behavior. With a fully-distributed routing concept (Figure 3), the dynamic spatially distributed character of the OMS3/J2K-S watershed model that separates it from other watershed models (e.g., SWAT) becomes apparent. Furthermore, higher spatial resolution in combination with the lateral transfer of water between HRUs and stream channel reaches can be considered a very important advancement (in hydrological modeling) towards deriving suitable conservation management scenarios for CEAP.

REFERENCES


Integrated Meta-Modelling for Decision Support in Integrated Catchment Management

Vikas Kumar\textsuperscript{a}, Annelie Holzkaemper\textsuperscript{b}, David Lerner\textsuperscript{a}
\textsuperscript{a}Catchment Science Centre, Kroto Research Institute, University of Sheffield, Broad Lane, Sheffield S3 7HQ, UK. (vikas.kumar/ d.n.lerner@sheffield.ac.uk)
\textsuperscript{b}Agroscope FAL Reckenholz, Swiss Federal Research Station for Agroecology and Agriculture, Air Pollution/Climate Group, Reckenholzstr. 191, CH-8046 Zurich, Switzerland (annelie.holzkaemper@art.admin.ch)

Abstract: Developing supporting models for multidisciplinary, uncertain and complex Integrated Catchment Management (ICM) is a highly challenging task. Knowledge from multiple disciplines must be integrated, and the process is compounded by significant uncertainty. The key gap that provides the research context is the need for a holistic modelling framework to support ICM, able to capture system complexities and interrelationships, and identify long-term solutions to catchment management problems. In this paper, we present the feasibility study of a new framework for developing an integrated meta-model for decision-support in ICM. The study undertaken by the Catchment Science Centre at the University of Sheffield in a project called the Macro-Ecological Model (MEM) in collaboration with the Environment Agency of UK. The MEM is developed as a consistent framework for the integration of knowledge and information about environmental, social and economic processes and process-interactions that are affected by management actions and have impacts on multiple management objectives. The MEM combines the advantages of “soft” techniques of stakeholder participation for problem structuring, interdisciplinary communication and negotiation with the “hard” predictive capabilities for analysing the likely outcomes of different management scenarios. The meta-model could serve as a learning and decision-support tool to be applied within a group of decision-makers and stakeholders.

Keywords: Bayesian Network; Decision-support; Integrated catchment management; Meta-model; Water Framework Directive.

1. INTRODUCTION

Integrated Catchment Management (ICM) is a complex interdisciplinary area, with scientific and socio-economic sides to it, which intersects various knowledge domains, and acts at various levels of abstraction. ICM takes a holistic approach to all the interconnected water-related issues, while also considering the activities of various stakeholders [Holzkaemper et al., 2010a, 2009]. The practical applicability of integrated models to support ICM depends on many factors. The many different kinds and sources of information have to be condensed to support decision makers in handling the management problem. Effective decision making depends on the power of knowledge abstraction: the ability to draw back one or more layers from the complex system around us, and then assemble piece by piece different processes, system linkages, management hypothesis into a desired level of abstraction. A framework for effective abstraction can help to organizing, analysing, and ultimately “operationalizing” information—integrating system functioning and decision-making processes into an analytical tool. The process of analysis obviously involves thinking, but while the human brain is capable of highly complex thought patterns, there is a limit to how far out a person can abstract, or how many concepts one can juggle.
in his mind at once. An integrated meta-model can be considered as a conceptual analytical tool for multi-source information integration and abstraction which can be used to help human abstraction processes for the purpose of decision making or Meta scale analysis.

Over the last two decades increasing numbers of model-based decision support tools have been developed to support Integrated Catchment Management. Two fundamentally different approaches to integration have been used namely model couplings and conceptual network models. These differ in their primary focus of integration. The main focus of conceptual network models is to facilitate interdisciplinary communication and structure the management problem based on the knowledge and views of different stakeholders. Whereas the model coupling approach can be described as a “hard” approach to integration, aiming to integrate quantitative process descriptions. Its main focus is on analysing different management scenarios to rank the possible solutions. A detail discussion has been provided in our paper or project report Holzkaemper et al. [2010a, 2009], which also reviewed the requirements of such tools in the context of ICM and concluded that they need to: (i) integrate a wide range of objectives and processes, (ii) aggregate system complexity to the level that is appropriate in the decision-making context, (iii) represent and communicate prediction uncertainties, (iv) be fast and easily applicable as exploratory learning tools, and (v) be easily transferable to new regions.

Different methods for knowledge abstraction or facilitating analysis have emerged, for example hierarchical based methods that create and harnesses different layers of abstractions. Other methods frequently used are multi-attribute utility analyses, Bayesian Networks, decision trees and influence diagrams, stochastic optimal control theory, partially observable Markov decision processes, neural network, rule-based cognitive architectures etc. In the last decade Bayesian networks (BNs) have increasingly been applied to environmental management problems under uncertainty, and recently also to integrated water management issues [Ames et al., 2005; Barton et al., 2008; Kumar et al., 2008]. BNs are graphical probabilistic approaches, based on Bayesian probability theory, which are commonly used as a decision analysis framework. BNs allow the integration and abstraction of qualitative and/or quantitative information with exclusive consideration of uncertainties associated with that information. The BN approach involves describing a system in terms of variables and linkages, or relationships between variables, at a level appropriate to the decision maker. This is achieved through representing linkages as conditional probability tables and propagating probabilities through the network to give the likelihood of variable outcomes. BNs are increasingly being used as decision support tools to aid the management of the complex and uncertain domains of natural systems [Barton et. al., 2008]. Common to these studies is the use of BNs to integrate probabilistic information derived from data sets, model simulations and expert opinion in the study of water allocation or pollution problems. As an alternative to extensive scenario analysis using deterministic models [e.g. Hein, 2006], BNs hold the promise of a more complete accounting of integrated model uncertainty. In some cases, BNs are used as data testing or analysis framework for example to study the properties of integrating a number of sub-models for purposes of targeting data collection or joint risk analysis [Borsuk et al., 2004]. In other cases, BNs are extended to include uncertainty regarding the scenario analysis of management decisions in what is known as influence diagrams [Barton et. al., 2008].

The main aim of this paper is to report the outcome of the feasibility study of a new framework for developing an integrated meta-model for decision-support in ICM based on BNs approach which we recently developed at Catchment Science Centre (CSC) at the University of Sheffield.

2. INTEGRATED META-MODEL DEVELOPMENT FRAMEWORK

We have recently completed a scoping study [Holzkaemper et al. 2009] of developing an integrated meta-model for decision-support in ICM, called the MEM (Macro Ecological Model). A number of design principles were adopted which make the MEM a novel and ambitious approach for developing a decision-support model. We sought to integrate across
traditional domains (e.g. linking ecology, hydrology, water quality and socio-economics) and across sources of knowledge (e.g. combining empirical data with the outputs of existing models). The final model was designed to be simple enough to be used by non-experts within decision-making and negotiation meetings, and to explicitly communicate the uncertainties associated with predictions. Bayesian Networks and model coupling techniques were used, and the prototype was built from a combination of numerical models, data analysis, and expert knowledge in a way that, we believe, has rarely been attempted before. The framework and techniques are currently being used to build similar tools in a research project on urban river corridors (www.ursula.ac.uk).

Detail of proposed framework has been discussed in Holzkaemper et al. [2010a, 2009] and summarised in Figure 1. The framework combines the advantages of “soft” techniques of stakeholder participation for problem structuring, interdisciplinary communication and negotiation, alongside the “hard” predictive capabilities for analysing the likely outcomes of different management scenarios. The integrated modelling cycle has been envisioned as a four phase model development framework with close engagement of three different groups of people. It makes a broad distinction between three groups of people with different roles and expertise: stakeholders, domain experts and system modellers. The stakeholders, including policy-makers and various interest groups, make the final decisions and have responsibility for the outcome. The domain experts are subject knowledge experts and help in model development from the problem from Phase 1 into an integrated conceptual model (Phase 2) and finally into a functional integrated model (Phase 3). The system modellers provide the continuity between the phases, have the expertise to elicit the knowledge and assemble the various models with help from the other groups and finally to create the simplified meta-model (phase 4) for use by the stakeholder or user group.

![Figure 1](image.png)

**Figure 1.** The four phases of participatory model building framework developed in MEM project.

The model development starts with an iterative procedure of specification of problem definition (a description of the nature and scope of a specific problem that needs to be addressed) with close engagement of decision-makers and stakeholders (phase 1). The cognitive mapping approach is applied to identify the objectives and management actions that represent the major interests and activities of the stakeholders and decision-makers. In phase 2, an integrated conceptual model is developed, linking the identified management actions to the objectives under consideration. The cognitive mapping approach is applied on a more detailed level in collaboration with domain experts to integrate knowledge from different disciplines. Data requirements and availabilities according to conceptual model are also analysed at this stage. Once the scope for the decision-making problem is agreed, and knowledge on possible system interactions is integrated, an operational model is developed in phase 3 to quantify the impacts of the management interventions on
management objectives as specified in the conceptual model. This integrated model would couple existing process-based models with empirical and knowledge-based models. This allows for system components to be included in the integrated model even if process-based models are unavailable for these components. In this way, important interactions, non-linear effects or spatial or temporal dynamics are not neglected. Engagement of domain experts is required, especially for the development of knowledge-based sub-models. In phase 4, the integrated operational model developed in the third phase is abstracted to result in an integrated meta-model that resembles the behaviour of the complex coupled model and adapts the information about the complexity of the system to the decision-making level. Multiple simulations with the coupled model are usually required to generate outputs for various management scenarios and derive uncertainty estimates. The BN approach provides a promising possibility for implementing the meta-model and is especially interesting in the context of representing uncertainties. Abstract information along with prediction uncertainties would be quantified and represented. The operationally simple meta-model developed could be used to inform decision-making in ICM. Decision-makers and stakeholders could test their ideas for management strategies and explore synergies and trade-offs between different objectives.

The stakeholders, including policy-makers and various interest groups, make the final decisions and have responsibility for the outcome. They will be most interested in framing the problem (Phase 1) and having access to an easy-to-use model with which to explore the options (Phase 4). The domain experts have the knowledge of the cause-effect relationships in the catchment. They can help translate the problem from Phase 1 into an integrated conceptual model (Phase 2) and assist in converting the conceptual model into a full model (Phase 3). The system modellers provide the continuity between the phases, have the expertise to elicit the knowledge and assemble the various models with the help from the other groups, and finally to create the simplified meta-model for use by the stakeholder group (Phase 4). Once the stakeholders have reached a decision, the more complex model developed in phase 3 can be used for more detailed analyses.

3. CASE STUDY

The framework developed in the MEM project was implemented for the Don catchment in North East England, UK. The Don catchment comprises an area of ~1800 km² and it is located in the Humber River Basin District (RBD). Alongside eight other RBDs, the Humber represents the administrative unit for which River Basin Management Plans are developed as part of the implementation of the WFD in England and Wales. The upland and downstream rural parts of the catchment contrast with the middle reach which contains the previously heavily industrialised urban conurbations of Sheffield, Rotherham and Doncaster, now undergoing regeneration and redevelopment. It incorporates a rich mixture of geological, topographical, soil and land use types, and is representative of many catchments in the UK.

The development process and outcomes of the prototype MEM have been described in detail by Holzkaemper et al. [2010b, 2009]. The stakeholders were a mixture of policy, operations and science staff in the Environment Agency of England and Wales (EA). The system modellers were researchers from the CSC. The domain experts were academics and researchers in the CSC and science staff in the EA. There was some overlap between the stakeholder and expert groups.

In Phase 1, agreement was reached that the model would investigate a selection of catchment management options related to agricultural land, urban drainage and flood management, and that options would be compared using the objectives of water quality (with phosphate and BOD as indicators), biological quality (biological general quality assessment score, GQAbio, as indicator) and flood risk (flood damage costs as indicator). This is a relatively small set of options and indicators compared to those that would be needed to populate a full model. They were selected as examples which were sufficient to show whether the method could link indicators for the Water Framework Directive with
those for other objectives such as flood risk, and whether the framework and technical solutions being proposed would be suitable. In Phase 2, an integrated conceptual model was constructed of the relationships between the management options and indicators, as shown in Figure 2. In simple terms, four sub-models, one for each indicator, were developed, integrated together and the resulting network simplified by eliminating unimportant or unspecified links; this phase was iterative and collaborative with the domain experts.

In Phase 3, the conceptual model was translated into a fully functional model of the system. Three different types of information (numerical models, data analysis, and expert knowledge) were used to populate the links shown in Figure 3 in order to create the model. The existing numerical models Psychic and SIMCAT were loosely coupled and used to simulate phosphate loads and transport from diffuse and point sources, while SIMCAT also simulated inputs of BOD from point sources. The effects on water quality caused by changes in travel times due to flood management were simulated in SIMCAT. Results from an existing flood modelling study [Hankin, 2008] were used to estimate flood damage. Expert knowledge was used to define the relationships between the flood management options and both the flood protection standard and the travel time changes which affected water quality. A combination of data analysis and expert knowledge was used to define the biological quality sub-model. A complex integrated model was the product of this third phase of the modelling building framework. This was the longest phase of the work, with many tasks to be completed, including acquiring, coupling and calibrating models; acquiring, cleaning up and analysing datasets; eliciting knowledge from experts; and finally coupling all the different aspects together.

The model produced within phase 3 is too complex to run in planning meetings or by non-experts such as the original stakeholder group. In Phase 4, a meta-model was created to emulate the full model in a way that was rapid, easy to use, and retaining the confidence of stakeholders. It relates closely to the conceptual model (Figure 2), and was created as a Bayesian Network derived from a set of runs of the full model. Finally the BN-based meta model was packed in a user friendly graphic user interface (GUI) to make it more intuitive to operate for non-specialist users (i.e. managers and stakeholders; Figure 3). The meta-model predicts indicator values and their uncertainty for combinations of management options. Estimates are considered accurate enough for decision-making; once a scenario has been chosen, the more detailed model of phase 3, or the initial specialist models, would be used for detailed design.
4. EVALUATION OF PROPOSED FRAMEWORK AND PROTOTYPE MODEL

At the end of the prototyping study, a workshop was held with the potential users of the tool (EA staff involved in the development of River Basin Management Plans for the WFD and members of the project board). Detail discussion on evaluation has been provided in Holzkämper et al. (2010b). The perceived benefits of the MEM were:

- It provides a structured approach to address complex planning issues and integrate knowledge from different domains.
- The visualisation of cross-benefits between management measures enhances the effectiveness of planning.
- The presentation of uncertainties allows for systematic review and identification of robust measures.
- Limiting the level of complexity and detail enhances the applicability for non-specialist users.
- The tool could support communication and social learning in participatory planning.

However, there are some limitations to the prototype model and development framework. Technical and institutional challenges that were encountered during the prototyping study are discussed in detail in Lerner et al. [2010] and Holzkämper et al. [2010b, 2009]. Major technical challenges in building such a tool include the difficulties of coupling disparate models of components of the catchment system, and the conversion of a complex coupled model into a simplified meta-model. The model will have to be customised to each basin in which it is used. Validation of an integrated catchment model is difficult because there are few, if any, suitable sets of field observations which could be used to test it against. Moreover, the prototype model does not represent all uncertainties. For example, uncertainties about the efficiency of buffer strips were not considered in the prototype model, as this uncertainty analysis would have required a large effort because the source code of the underlying model (Psychic) was not accessible which prohibited a tight coupling and automated uncertainty analysis. Limited access to data and models because of issues surrounding intellectual property rights presents significant barriers to any integrated
modelling exercise. This issue will remain a significant barrier to future development of integrated decision-support tools.

The institutional challenges are more severe, starting with the question of whether individuals and institutions accept the overall concept of integrated catchment management through collaboration and new ways of open and participatory working. Developing integrated model at large scales will have major resource implications within the sponsoring organisations under the participatory route we propose. Continuity and availability of personnel is a key challenge when developing a framework that involves long-term, intense engagement between scientists and potential users. Identifying relevant individuals within organisations that have complex structures, particularly individuals with sufficient power to promote frameworks such as the one described in this paper, is also a significant challenge. Will a model such as the one discussed here be acceptable to all the groups and vested interests involved, and will the project team get the cooperation over data, current models and input of stakeholder time that it needs? Such a project is ambitious and risky enterprise, but has potentially high rewards for sustainable development of catchments in the UK and beyond.

5. CONCLUSION

Within the Macro-Ecological Model project we developed a consistent framework for integrated meta-model for decision-support in ICM based on a BN knowledge integration approach. The framework seeks to reduce some of the current limitations and promote the development of modelling tools that can support ICM both by providing an integrated scientific evidence base and by facilitating communication and learning processes. The results of the scoping study suggest that this approach, although challengingly difficult in its own right, could help to support the implementation of river basin planning and other activities based on philosophies such as ICM.

As such, it is designed to be a tool for high-level decision support in integrated catchment management, which brings together knowledge from different disciplines to support a more holistic evaluation of planning alternatives. The BN approach is well suited to integrating knowledge from different resources. It also provides the opportunity to perform rapid scenario analyses, which makes it a very practicable tool to be applied in a planning context. The possibility to take modelling uncertainties explicitly into account enables robust decision support [Schlüter & Rüger, 2007]. Previous research in the area of decision-support systems has pointed out that decision-support tools are only accepted by their potential users if these users are involved in the model development from the beginning [Borowski & Hare, 2007]. Therefore, a close interaction between the model developers and the potential users is promoted in the development of the Macro-Ecological Model. The intuitive model structure of the BNs and the integration of information from trusted sources (e.g. EA data, models and expert knowledge) should support the acceptance of the model amongst its potential users.

ACKNOWLEDGEMENTS

The Macro-Ecological Model project (SC060039) is funded by the Environment Agency and the University of Sheffield and URSULA (EP/F007388/1) is the UK Engineering and Physical Sciences Research Council funded project. We are grateful to all researchers at CSC, stakeholders and domain experts who took part, but the opinions expressed are those of the authors alone.
REFERENCES


Holzaemper, A., V. Kumar, B. Surridge, A. Paetzold, S. Lautenbach, R. Harris, and D.N. Lerner, How to develop a simple tool for a complex task – Decision support for integrated catchment management, Submitted, 2010a.

Holzkaemper, A., V. Kumar, B. Surridge, A. Paetzold, and D.N. Lerner, Testing the development of a meta-model tool for decision-support in integrated catchment management, Submitted, 2010b.


Lerner, D. N., V. Kumar, A. Holzkaemper, B. Surridge, and R. Harris, Challenges in developing an Integrated Catchment Management Model, *Water and Environment Journal*, 2010 (Accepted)

Integrated Model of Municipal Waste Management of the Czech Republic

J. Hřebíček, Jana Soukopová
1Masaryk University, Institute of Biostatistics and Analyses, Kamenice 125/3, 625 00 Brno, Czech Republic, hrebicek@iba.muni.cz
2Masaryk University, Faculty of Economics and Administration, Lipová 41a, 602 00 Brno, Czech Republic, soukopova@econ.muni.cz

Abstract: An integrated model of municipal waste management of the Czech Republic was developed as a balanced network model for a set of municipal solid waste (MSW) sources (mostly municipalities) connected with a set of chosen waste treatment facilities processing their waste. The model involves composting, energy utilisation, material recycling, and land-filling. It is implemented as a combination of four sub-models (GIS transport sub-model, quantification and composition of MSW sub-models, cost economic sub-models of all facilities and sub-model optimizing greenhouse gases expressed as CO₂ equivalent) and it is easily scalable. It enables the optimisation of environmental impacts (land-filling and greenhouse emissions). Its application is demonstrated in the case study as a decision support tool for the planning allocation of potential facilities of waste recovery instead of land-filling.

Keywords: integrated waste management model; municipal solid waste; waste cost modelling.

1. INTRODUCTION

Waste is an unavoidable by-product of human activities. Economic development, urbanisation and improved living standards in cities increase the quantity and complexity of generated municipal solid waste (MSW). The decisions in the area of MSW management are not only very capital-intensive, but also difficult from environmental and social points of view. There is the need to develop, master and implement simple but reliable tools that will help decision-makers to analyse waste management processes. This paper discusses an integrated model of municipal solid waste management to assist in identifying alternative MSW management strategies and plans that meet cost, material, energy, and environmental emissions objectives of European Union (EU).

The MSW is considered like all waste generated within the community (as well as the source MSW) by the activities of inhabitants (households) and businesses (e.g. trade waste), which is separated into its components and transported to waste treatment facilities. The MSW normally contains the remains of food and vegetables, paper, plastic, glass and metal containers, printed matter (newspapers, magazines, and books), destroyed products, ashes and rubbish, used or unwanted consumer goods, including shoes and clothing. Chosen components of MSW are collected separately, and thus, they are balanced separately. The waste (or its components) can be composted, used as raw material (paper, plastic, glass, and metals), used in bio-gas, energy recovery (incineration) plants or land-filled. The separation of its components may take place at the source (separate collection in the municipalities) or in the facilities. So we define the individual waste streams, which are mass balancing.

Earlier this decade, the development of models of waste management began moving towards the integrated model waste management (IMWM), which is designed to minimise
the economic costs and/or environmental impacts, see Berger et al. [1999], Wang [2001], Haigh, Shmelev nad Powell, Yeomans [2006].

Consider the IMWM discussed by Hřebiček et al. [2009], [2010] which consists of the set of MSW sources (municipalities) of the Czech Republic connected by the road network with the set of waste treatment facilities (composting, bio-gas, mechanical-biological treatment (MBT) and pre-treatment of recyclable waste plants, incinerating plants with energy recovery (ERP) and landfills), where MSW (or its components separated at source) is transported to chosen facilities for recovery or final disposal. The material balance is examined in terms of material flows between MSW sources and waste treatment facilities. The production of MSW in the Czech Republic is approximately 3.2 million tons of MSW annually (in 2008), and most (85 percent) is land-filled.

In developing an IMWM for the Czech Republic, we started from the models available in literature. Since the early 1990s, a number of IMWMs have been developed, which were based on life cycle analysis (LCA), i.e. materials and energy balances, see McDougall et al. and Solano et al. [2002]. Most available models are static, respectively deterministic and quantify the uncertainty of estimates due to random nature of input values. Another disadvantage of models based only on the LCA is that they do not allow optimising the allocation of waste treatment facilities from sources and/or quantifying the transport emissions. We tried to reduce the greatest uncertainty of our model by the estimation of the composition of municipal waste, waste separation, varying the proportion of resources, varying quantities of trade waste and the like.

The developed IMWM for the Czech Republic consists of the combination of four sub-models, where we used following tools:

a) The geographic information system (GIS) ArcMap, which computed a transport matrix linking the sources MSW and waste treatment facilities and the simple model, which generated emissions from the transport of MSW and enable to find the closest facility.

b) The sophisticated model of Hejč et al. [2008] for the determination of the quantity and composition of MSW at every source (municipality) or the database of collected data from annual waste reports of municipalities regarding the quantity and separated components of MSW.

c) The cost economic sub-models of every type of waste treatment facility including the generation of the emissions of MSW treatment.

d) The software LINGO for the carbon emissions optimisation of allocated waste treatment facilities with the choice of either the economic or the environmental point of view.

The above IMWM requires criteria (prioritisation) from decision-makers (government regulators), which may involve an acceptable level of pollutant emissions and costs, as well as a reduction of landscape and biodiversity or prevent a pollution of groundwater and surface water. Practically, such optimisation comes into consideration for regulators when deciding on localisation of a new facility (technology and capacity) and/or closure of existing facilities, the regulation of their capacities and the like. Some chosen feasible minimum is usually acceptable for regulator without optimisation.

2. TRANSPORT NETWORK MODEL

Consider the MSW flows at the Czech Republic among all sources (municipalities) $S_i, (i=1…N), N = 6245$ and all waste treatment facilities $F_j, (j=1…M), M = 307$, where $ML = 237$ is the number of landfills. Consider these MSW flows in a continuous manner and mass balance between sources and facilities carry out over a longer period of time (annual reporting). We built the transport matrix $D = \{d_{ij}\}, (NxM)$, of real transport distances $d_{ij}$ (e.g. road maps) among all sources $S_i$ and all facilities $F_j$ and the vector of the distance $dc = \{dc_i\} (NxI)$ of the source $S_i$ from its closest landfill $F_c, c \in \{1, …, M\}$. 
We have used the GIS program *ArcMap* 9.2 with its extension *Network Analyst* 9.2 from ESRI for the analysis of the closest facility (e.g. landfills) to the individual sources (municipalities).

![Map of landfills of the Czech Republic.](image)

**Figure 1.** Map of landfills of the Czech Republic.

The *Network Analyst* program enables us to implement networking analysis - finding the shortest path between two points, finding time to travel between two points, etc. Users can create and maintain network data sets in shape file, personal geo-database, and enterprise geo-database formats. By using *ArcGIS Network Analyst*, we created simple applications that provide us transport distances among all $M$ sources and $N$ facilities, find closest facilities, and create the distance matrix $D$ and the vector $dc$.

We used municipalities and roads layers of the Czech Republic for *ArcGIS Network Analyst* from the open-source project *FreeGeodataCZ data package*. It incorporates an advanced connectivity transport model that can represent complex scenarios, such as multi-modal transportation networks.

### 3. MODEL OF QUANTITY AND COMPOSITION OF MSW

The developed model of quantity and composition of MSW is based on the production of MSW in each municipality of the Czech Republic and was published by Hejč and Hřešiček, Hejč et al. [2008]. They described formally the simple model of MSW production as the function of appropriate variables taking into account specific waste production, and local demographic, socio-economic influences:

$$ P = inh \cdot spec \cdot std \cdot sz \cdot unemp \cdot hsg \cdot heat, $$

where:

- $P$ is the amount of the MSW production of municipality per year in tons [t];
- $inh$ is the number of inhabitants of municipality;
- $spec$ is the specific waste production coefficient (reference values of other coefficients), measured in tons [t];
- $std$ is the standard of living coefficient;
- $sz$ is the size of the community coefficient;
- $unemp$ is the unemployment rate coefficient;
- $hsg$ is the type of housing (recreation, blocks of flats, empty houses, etc. coefficient and
- $heat$ is the type of heating coefficient.

The model (1) came with a finer division of demographic, socio-economic impacts on production and treatment of MSW at the level of individual municipalities. It enables us to meet the conditions required by the Ministry of Environment (MoE) to hit the regional dimensions (at least at district level) and, therefore, to meet different impacts on relative
prices of waste management in different regions of the Czech Republic, see Hřebíček et al. [2010]. This model was investigated, calibrated and verified for three years in the South Moravian region of the Czech Republic, where some of the above variables were optimised by Hejč et al. [2008], Hejč and Hřebíček [2008a] with the simple expression:

\[ x = \frac{act}{ref} \cdot cx, \]

where \( x \) means a variable from \{std, sz, unemp, hsg, heat\} and

\( ref \) means a reference value from three year investigations;

\( act \) an actual value from given year and

\( cx \) is the compensator (given by optimisation process) of the considered variable \( x \).

The above model (1) calculates the production \( P_i \) of MSW in each municipality \( S_i \) based on the adjusted number of inhabitants \( inh \), the specific waste production coefficient \( spec \) and specific demographic data reflecting the population behaviour with respect to MSW management (i.e. the type of housing \( hsg \), and other variables \( std, sz, unemp, heat \), of municipality \( S_i \), \( i = 1 \ldots N \)). These data are downloaded from publicly accessible registers of the Center for Regional Development of the Ministry for Regional Development of the Czech Republic and the Czech Statistical Office. They are updated annually from all municipalities in the Czech Republic; therefore, the model enables us to calculate the production of MSW for the given year with actual variables in (1) and predict waste production with using the linear model of the Waste Management Plan of the Czech Republic. We were able to calculate waste production MSW for the year 2008 and predict the increase of the production of MSW in 2016 and 2020 years.

The validation and optimisation of the model (1) outputs - the production \( P_i \) of MSW - was done by Hejč and Hřebíček [2008a] with the available data from the annual reports of municipalities \( S_i \) of the South Moravia region; however, annual reports of MSW of \( S_i \) bear some error, which arose from different data qualities. The process of the improvement of the data quality of municipalities \( S_i \) of the South Moravian region lasted several months. The data from the annual reports of all municipalities about their waste production are collected by the Information System of Waste Management (ISWM) of the Czech Republic. We used these, but we had to solve the problem without complete data, because more than 500 municipalities of the Czech Republic did not report their annual MSW production to ISWM. So we had to use the model (1) for the calculation of their missed MSW production in 2008 and the prediction of their MSW production in 2016 and 2020 years.

We have to estimate the MSW composition for the calculation of the amount of separated components of MSW at each municipalities \( S_i \) to obtain the rest \( PD_i \) of MSW \( P_i \) after the separation of recyclable components. We used for this estimation values listed in the Table 1, which are based on the results of research of Benešová et al. [2009] and Vrana et al. [2010].

### Table 1. MSW composition at the Czech Republic (weight %)

<table>
<thead>
<tr>
<th>Material</th>
<th>The share of material groups in waste (% by weight), average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Housing estates of big cities</td>
</tr>
<tr>
<td>Paper</td>
<td>22.7</td>
</tr>
<tr>
<td>Plastics</td>
<td>13.8</td>
</tr>
<tr>
<td>Glass</td>
<td>8.7</td>
</tr>
<tr>
<td>Metals</td>
<td>3.4</td>
</tr>
<tr>
<td>Bio-waste</td>
<td>18.2</td>
</tr>
<tr>
<td>Textile</td>
<td>5.6</td>
</tr>
<tr>
<td>Energy recovery waste</td>
<td>12.4</td>
</tr>
<tr>
<td>Under 20 mm</td>
<td>9.7</td>
</tr>
<tr>
<td>Other</td>
<td>5.5</td>
</tr>
<tr>
<td>Totally</td>
<td>100.0</td>
</tr>
</tbody>
</table>

We considered values from Table 1 to estimate real quantity of a disposable production \( PD_i \) of MSW from the municipality \( S_i \) to waste treatment facilities (new ones or available ones).
after separation of recyclable components of MSW, which was estimated by Hejč et al. [2008], Hřebíček et al. [2010]:

$$PD_i = (1 - sep_i \cdot will_i) P_i,$$

where $sep_i$ is the ratio of separation at source $S_i$, $will_i$ is willingness to separate MSW (paper, glass, metals, textile and bio-waste) at municipality $S_i$ ($i = 1 \ldots N$). Coefficients $sep_i$ and $will_i$ came from data of the investigation of the MoE and were validated in the South Moravian region by Hejč and Hřebíček [2008a], Hřebíček et al. [2010].

The model (2) helped us to solve some uncertainties stemming from the different state of population awareness about MSW management and estimate the amount of disposable production $PD_i$ of MSW from the municipality $S_i$ to appropriate waste treatment facilities. The MoE has used this model since 2009 after several months reviewing process by experts using Vrana et al. [2010] approach.

4. ECONOMIC MODELS FOR FACILITIES

We developed cost economic models for all types of facilities $F_j$ ($j = 1 \ldots M$), i.e. composting, biogas, MBT and ERP plants and landfills, Hřebíček et al. [2010]. These models are similar and we introduced this economic model for a generic facility $F$.

Calculate the price $p$ of one t of the waste treatment at a new composting, bio-gas, MBT and ERP plant $F$. This calculation is based on the financial and economic analysis and financing methods for the measuring the efficiency of investment, see Valach [2006], etc. We used the Net Present Value (NPV) as the basic calculation method for the price $p$.

$$NPV = -I + \sum_{i=1}^{n} \frac{CF_i}{(1 + r)^i}$$

where
- $I$ means an investment expenditures in facility $F$,
- $CF_i$ means a cash flow generated in the period $i$,
- $r$ means the discount rate and
- $n$ means the lifetime of facility.

To calculate the price $p$ is assumed that the $NPV$ must be at the time of return positive. Thus the basic assumption was that we set the maximum acceptable payback period of investment $I$ in the facility $F$. Then $n = lifetime = payback$ in formula (2). If we assume that

$$CF_i = pK + B_i - C_i - u_i - j_i - E_i - T_i$$

$$T_i = t(pK + B_i - C_i - u_i - j_i - E_i - O)$$

Then price $p$ is defined as

$$p = \frac{I}{(1-t)\sum_{i=1}^{n} \frac{1}{(1+r)^i}} - B + C + \sum_{i=1}^{n} \frac{1}{(1+r)^i} - \frac{rO}{1-t}$$

where
- $B_i$ is the total revenue generated from the facility in the period $i$,
- $C_i$ is the total operating costs arising from the facility during the period $i$,
- $K$ means the capacity of the facility,
- $T_i$ means a tax on income arising from the facility during the period $i$,
- $u_i$ means the interest due on loans for the period $i$,
- $j_i$ means repayment of principal on loans for the period $i$ and
- $E_i$ means the costs of emission allowances for the period $i$.
- $i$ means the period (year) from 0 to $n$
- $n$ means lifetime and also payback of the facility.
It is clear that different facilities will have different costs, incomes, investments, etc. For each-mentioned facility $F_j$ (composting, bio-gas, MBT and ERP plants, and landfills) we developed economic sub-models for the construction of the price $p_j$ of the given facility $F_j$ ($j=1\ldots M$). These models were based on the real level of investment, operating expenses, operating incomes, interest on loans, capacity of facility and emissions, Hřebíček et al. [2010]. The economic model of landfill was evaluated based on the average price of all landfills in the Czech Republic because the standard deviation of prices was less than 10 percent.

5. CARBON EMISSION MODEL

Reducing greenhouse gas emissions is an important social topic in the Czech Republic—particularly the suppression of landfill methane emissions. Total emissions of CO$_2$ equivalent will have to be significantly reduced in the waste management sector. In developing the carbon emission model, we have confined ourselves to minimise greenhouse gas emissions in the transportation, composting, incineration and land-filling MSW. Modelling of these emissions is a standard part of the LCA models of MSW management, so that in Solano et al. [2002], there are emission factors. This means that the emission factors, unit fuel consumption, energy prices, waste categorisation and other parameters are fixed set according to the Czech Republic where the IMWM was constructed. Moreover, users are not accessible to balance relationships, the models developed above do not optimise traffic and are strictly deterministic (do not take into account random variations of input data and uncertainty of adjustable parameters). Besides the mass-flow models, there are also above economic models that can describe the system of unit costs and to examine the impact of economic instruments; therefore, the carbon emission model was simply transformed into the above economic model by replacing the unit cost of emission factors. Because the model allows us to insert individual emission factors, which depend on the waste treatment technology and its optimal use, it is possible by analogy to conduct economic optimisation with regard to the cost of waste treatment facilities; however, the data for new facilities are not available to the regulator (MoE) and can be obtained only from the operators (or potential investors at prepared facilities) or expert estimations.

6. INTEGRATION OF SUB-MODELS INTO INTEGRATED MODEL OF WASTE MANAGEMENT

The above chapters introduced shortly four developed different sub-models needed for the regulation of waste management of the Czech Republic and a decision support of the allocation of subsidies from EU. We used properties of the MS Excel spreadsheet for the integration of above sub-models into one of the IMWMs for the Czech Republic to evaluate cost and price relationships for the municipal waste management of the country.

This implementation of IMWM enables the central option of the set of the input economic parameters of the model at the single control sheet of the MS Excel with interconnected sheets, where we implemented above sub-models:

a) the sheet (table) of socio-demographic variables $inh$, $std$, $sz$, $unemp$, heat, $hs$, $sep$, will, of all municipalities $S_i$ of the Czech Republic needed to calculate the outputs $P_i$ and $PD_i$, ($i=1\ldots N$), of the model (1), (2),

b) the sheet with the dynamically calculated the vector $dc$ of the distance $dci$ of source $S_i$ from the closest landfill $F_i$, by Network Analyst program, and the cost $CTF_i$ of waste treatment of $PD_i$ at the landfill $F_i$ together with the cost $CTE_i$ of transport to this facility including carbon emissions cost, ($i=1\ldots N$),

c) the sheets of economic models (6) of (planned and current) waste treatment facilities $F_j$ with dynamically calculated prices $p_j$ including costs of carbon emission, ($j=1\ldots M$),

d) the sheet with dynamically calculated a potential amount of MSW from “the collecting waste area” of the facility $F_j$ ($j=1\ldots M$), where the collecting waste area consists of the municipalities, where are cheaper costs ($CTF_i + CTE_i$) to the closest appropriate facility than ones to the closest landfill,
the sheet of main communication interface the IMWM, where the input economic variables together with the allocation of new facilities are set up with further options required for the model.

Decisions makers of the MoE were able to use this IMWM to allocate subsidies from EU to investors of potential facilities to decline MSW from landfills to new facilities (ERP and MBT). They could choose inputs: the list of K planned facilities \( F_s \) (where they are connected with their economic models); their common payback; common value-added tax; chosen percentage of subsidy; charge of landfilling and landfill reclamation. They obtained outputs of this model, where were prices \( p_s \) of waste treatment at planned facilities \( F_s \), and calculated prices \( C_{Ti} = (C_{TF_i} + C_{TE_i}) \) for all municipalities \( S_i \) (where \( i = 1…N \)) of the Czech Republic which will pay for the treatment of MSW.

### 7. CASE STUDY FOR ALLOCATION OF NEW FACILITIES

We will illustrate the IMWM application to monitor total cost and pricing relationships in waste management for the Czech Republic. The IMWM was applied to estimate price load per capita at every MSW source \( S_i \) (where \( i = 1…N \)) and total cost and pricing relationships in the waste management of the Czech Republic, depending on planned EU subsidies to new allocated facilities (ERP and MBT) including the total amount MSW declined from landfills to these facilities. The IMWM was used for 36 different scenarios of subsidy schemes to split the amount of subsidy of EU structured funds for the investment of 12 possible allocations of mechanical biological treatment (MBT) plants and incineration plants with energy recovery (ERP). We modelled: total planned EU subsidies to new allocated facilities (ERP and MBT); the quantity of MSW which is available for each facility in comparison with its planned capacity; and the price load per capita (average and maximum) of MSW treatment for the Czech Republic.

The Table 2 shows the outputs of the model, i.e. prices (in CZK – Czech Crowns) of waste treatment of 1 ton MSW at planned facility (ERP and MBT) of given capacity with respect to considered EU subsidies.

<table>
<thead>
<tr>
<th>EU subsidy</th>
<th>ERP capacity per year</th>
<th>MBT capacity per year</th>
<th>Price per capita</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>100 kt</td>
<td>200 kt</td>
<td>80 kt</td>
</tr>
<tr>
<td>20%</td>
<td>1 565 CZK</td>
<td>1 363 CZK</td>
<td>1 742 CZK</td>
</tr>
<tr>
<td>30%</td>
<td>1 328 CZK</td>
<td>1 137 CZK</td>
<td>1 638 CZK</td>
</tr>
<tr>
<td>40%</td>
<td>1 091 CZK</td>
<td>912 CZK</td>
<td>1 539 CZK</td>
</tr>
</tbody>
</table>

### 8. CONCLUSIONS

The integrated waste management model of the Czech Republic is introduced in the paper. The model was implemented with using MS Excel as the combination of the tools: GIS Network Analyser, sophisticated sub-model calculating MSW production and separation of its component at all municipalities of the Czech Republic, cost economic sub-models of all facilities including carbon emission.

It enables to optimise environmental impacts (land-filling and greenhouse emissions). Its application was used as the decision support tool of the MoE in the case study of optimising EU subsidies to the planning allocation of new waste treatment facilities (ERP and MBT) with respect to expenses per capita of waste management of the Czech Republic.

### REFERENCES

Integrated Modeling for Source Characterization of Pathogenic Contamination in Watersheds

Michael Tryby a, S. Thomas Purucker a, and Gene Whelan a

aUS EPA, Ecosystems Research Division, 960 College Station Road, Athens, GA 30605, USA
(Tryby.Michael, Purucker.Tom, Whelan.Gene @epa.gov)

Abstract: The US EPA’s regulatory framework for recreational waters has protected public health for decades. Pathogenic contamination of these waters, however, remains a frequent cause of impairment. Integrated modeling is being leveraged to advance the agency’s understanding of pathogen fate and transport processes in watersheds and improve its ability to predict the consequences of exposure. This paper describes integrated modeling research focusing on source characterization techniques for pathogen transport scenarios in watersheds. Source characterization is a hidden requirement of Quantitative Microbial Risk Assessment, a method for estimating infection risks being evaluated across several programs within the EPA. A hybrid source characterization approach is described and demonstrated that utilizes integrated and inverse modeling methodologies to determine pathogen source allocations.

Keywords: Integrated Modeling; Multimedia Modeling; Source Characterization; Inverse Modeling; Receptor Modeling; Pathogens; Fecal Contamination; Recreational Waters; Risk Assessment

1 INTRODUCTION

The most frequent cause of impaired recreational waters in the United States is pathogen contamination. Pathogen releases are often associated with rainfall events and occur in receiving waters influenced by various sources such as septic systems, sewage discharges, combined sewer overflows, and agricultural land use. Current recreational water quality criteria established by the EPA are meant to protect public health from acute illness associated with exposure to pathogens. These criteria, however, are more than 20 years old. During the ensuing time, research has improved our fundamental understanding of the problem. In the near term, the agency seeks to apply this knowledge as it promulgates new recreational water criteria in 2012. Looking beyond, the agency seeks to advance pathogen fate and transport simulation in watersheds and exposure science with the development of tools to support a quantitative framework for assessing pathogen exposure risks in recreational waters.

Multimedia modeling is central to the approach being taken because pathogen fate and transport in watersheds encompasses several cross media processes. For example, consider a manure land application scenario as illustrated in Figure 1. The manure applied there to fertilize crops is a potential pathogen source. Transport processes acting on the manure-laden field would depend on meteorological conditions such as rainfall intensity and watershed topological conditions such as tillage practices, and crop and soil properties. These processes typically include infiltration, overland flow, in-stream dilution, advection, dispersion, and the complex mixing patterns occurring in the receiving water. Thus, the multimedia modeling domain encompasses atmospheric, surface and subsurface hydrologic and environmental hydrodynamic modeling. Furthermore, the exposure potential at the receiving water body is a complex stochastically driven function of these processes acting on multiple point and non-point pathogen sources across media boundaries.
Figure 1: A hypothetical watershed impacted by agricultural activities. The scenario consists of three agricultural operations located in a watershed adjacent to each of the three reaches which make up the drainage. Farms A, B, and C represent animal feed and land application operations. Fecal contamination from Farms A, B, and C enters each adjacent reach as runoff, then flows downstream to Receiving Water R where it is observed as a mixture.

This work focuses on agricultural sources, although fecal contamination in watersheds can originate from virtually anywhere and from several source categories, including humans and native wildlife, in addition to domesticated animals. Hundreds of different pathogens can be transmitted via exposure to fecal-contaminated water. Most pathogens infect the intestines and result in gastroenteritis. With the potential for so many viral, bacterial, and parasitic pathogens to be present in contaminated waters it is impractical and cost prohibitive to directly monitor all of them. Thus, broad spectrum surrogate organisms such as Escherichia coli in fresh waters and Enterococcus in marine waters [Wade et al., 2003] are used to indicate the presence of fecal contamination and the risk of illness associated with pathogens.

A future regulatory framework for recreational waters may adopt risk based criteria that build on the historical pathogen indicator paradigm. Hunter et al. [2004] describes Quantitative Microbial Risk Analysis (QMRA) as taking the well-developed chemical risk paradigm and modifying it to access microbial risks. Using QMRA an estimate of exposure risk is calculated based on the types of pathogens present, their infectivity, concentration distribution, the exposure pathway, dosage, and characteristics of the population being exposed. Thus, quantitative characterization of pathogen sources (i.e. estimates of pathogen concentrations broken down by type) is a necessary step in conducting a QMRA. A host of confounding factors, however, complicates practical application of QMRA.

The main difficulty is that a loose relationship exists between indicators and pathogens, but indi-
Indicators are used as the basis for rules and regulations; furthermore, observations tend to focus at the receptor. For example, refocusing attention on Figure 1, samples I are taken at location R to detect the presence of fecal contamination. The results indicate their presence, and their relative abundance correlates with the total fecal load. Conventional methods for detecting indicator organisms only provide data on the abundance of specific indicator organisms associated with fecal contamination. The results do not provide any characterization of the fecal source, such as pathogen host type, nor do they provide any direct information on pathogen presence or abundance. Therefore, it is not possible to estimate with accuracy exposure risks at location R using conventional analysis of Samples I. This paper describes how an integrated modeling approach can be used to quantitatively characterize pathogenic contamination sampled at location R.

A hybrid source characterization approach is proposed that integrates receptor, forward, and inverse modeling techniques to improve risk estimates for recreational exposures to pathogens occurring in natural waters. The first phase of this approach adapts receptor modeling techniques developed for air pollution management to provide source apportionment estimates — the relative contributions from each fecal source. Fecal sterol / stanol profiles derived from multiple fecal sources (e.g. humans, cows, ducks, dogs) [Shah et al., 2007] are used as the basis for the receptor modeling calculations. Two receptor modeling approaches have been investigated for the application — deterministic and stochastic chemical mass balance receptor models. These receptor modeling approaches, however, are limited by a linear input / output assumption between source loadings and concentrations observed at the receptor; they do not capture the complex fate and transport dynamics present in watersheds, constituent reactivity, microbial reproduction and die off, although source term uncertainty and measurement errors are represented.

A forward model developed using multimedia modeling techniques, can simulate the release, fate, and transport of microbial constituents and thereby account for these unrepresented processes. A forward model is currently under development and being coupled with a QMRA model within FRAMES [Whelan et al., 2010] that can translate source term loadings into exposure risks. Future work will investigate, general inverse modeling techniques that utilize a forward model for the input / output relation, and therefore, incorporate these missing processes into source allocation estimates. We seek to integrate these source characterization approaches and the QMRA model within FRAMES to improve pathogen exposure risk estimates. The intent of this paper is to present the integrated approach currently under development.

2 Methodology

Source characterization is a generic term describing methodologies that infer the characteristics of a pollutant source from observational data. In practice, source characterization can play an important role in environmental monitoring and remediation activities. To date, the work on source characterization conducted in hydrologic contexts has focused on nutrients (nitrogen and phosphorous) and heavy metal contamination. The solution of source characterization problems includes the identification of source locations, compositions, allocations, concentrations or some combination thereof. A key objective of this work is the development of a pathogen source allocation capability; specifically, quantifying the contribution of each different source species type to the total fecal load. The composition of fecal contamination is important for QMRA, however, making this determination is difficult. Host specific markers are difficult to identify and use because similar markers are present in many different species varying only by degree. The mixing of these markers due to transport phenomena in the watershed further confounds analysis. The paragraphs that follow describe each of the modeling techniques being integrated.

2.1 Receptor Modeling

Given observational data at a receptor, the goal of Receptor Modeling (RM) is to identify pollutant sources and apportion loadings to them. RM is a type of inverse model that assumes a linear input / output relationship between source loadings and concentrations observed at a receptor. RM is difficult due to complexity of transport processes, reactivity of constituents, and poor delineation
of source characteristics. Different approaches for formulating a receptor model depend on the characteristics of the system being analyzed and the observational data available; they include mass balance and multivariate models. Mass balance models are predicated on the fundamental principle of mass conservation. A chemical mass balance, therefore, is assumed, meaning that the observed sample can be described mathematically as a linear combination of source components. The mass balance assumption is limiting since it means reactive contaminants cannot be used as the basis for RM. It does, however, make the problem mathematically tractable. The assumption of mass conservation complicates the selection of markers since they must be quantitative and nonreactive.

**Receptor Modeling Example.** To illustrate the type of analysis possible using this hybrid approach, a receptor modeling example has been prepared using data from the published literature. The data set consists of source profiles for humans, cows, dogs, and ducks (eight fecal sterol/stanol compounds per profile) and 39 different contrived fecal mixtures [Shah et al., 2007]. A stochastic receptor modeling formulation using bio-chemical marker profiles, specifically fecal sterols/stanols, has been developed to allocate an observed mixture of fecal contamination across a set of profiled fecal sources (see Figure 2). One important aspect of receptor modeling with bio-chemical marker signals is the high degree of natural variability present in the source profiles. A linear mass balance equation is the mathematical basis of the formulation combined with the sophisticated treatment of error, natural variability, and uncertainty possible using Bayesian inference.

The stochastic approach presented here is a modification of the Bayesian source apportionment
receptor model developed by Keats et al. [2009]. Using it, source fecal profiles, source allocations, and the observed fecal contamination are considered stationary stochastic variables. The estimate of the posterior distribution for the source allocation is proportional to the likelihood of the observations, given the source allocations and profiles, and their a priori estimates (see 3 for one of the mixtures). The formulation is solved by constructing Markov chains for each of the observed 39 fecal mixtures (with a common set of source profiles), then sampling from the domain of the prior distributions, allowing for inference on the source contribution and profiles. A Metropolis-Hasting algorithm is used to determine acceptance and construct the posterior distribution that reflects the observed mixtures. At convergence, the Markov chain approximates the solution of the source allocation problem of interest an apportionment probability distribution for each source type (i.e., the gray histograms in 3).

Even in a laboratory setting with the Shah data set and an artificially restricted set of predictors, this is a high dimensional problem. There are 35 parameters in each of the 39 simulation sets corresponding to the observed mixtures, resulting in > 1K parameters. Field applications will be expected to deal with an even greater set of potential predictors. Dimensionality causes problems across the board: for chemical mass balance approaches to the receptor problem (summarized in Christensen and Gunst [2004]), fate and transport model calibration, as well as for conventional Bayesian Markov chain Monte Carlo applications. In addition, due to correlation in the Markov chains, simulations may run significantly longer than for direct Monte Carlo sampling, with runs > 100K simulations necessary to have a significant chance of convergence. Therefore, solution approaches must be able scale well for both parameters and sampling. A great advantage of the Keats et al. [2009] approach that we adopt is its use of Hamiltonian Monte Carlo
techniques [Hajian, 2007] that increase the rates of acceptance and allow for larger jumps in the multi-dimensional parameter space, leading to much more efficient parameter estimation, compared to traditional techniques, as dimensionality increases.

The linear chemical mass balance receptor model is appropriate for this laboratory scenario where fate and transport processes do not reduce or transform the sterol/stanol compounds between the source and receptor. However, the dynamics of fate and transport processes requires additional algorithms and parameters for effective inference in QMRA applications. Therefore, a research question that the authors are addressing is replacing the linear mass balance assumptions with more realistic fate and transport process algorithms. Current multi-media forward models cannot be calibrated on the observations of simultaneous multiple predictors necessary for QMRA and typical mass balance receptor modeling approaches do not faithfully represent known environmental processes. Combination of these two approaches is therefore necessary and would be a significant advance for QMRA, source identification, allocation, and remediation.

2.2 Multimedia Forward Modeling

Multimedia modeling can leverage existing modeling capabilities to address complex cross-media problems. Linking the contamination source to a receptor using multimedia modeling is a new tool for the risk assessment process that may improve the correlation between indicator and pathogen predictions. A multimedia model is currently being developed to express pathogen loading, transport, and fate (source-to-receptor modeling) and will provide source modeling capabilities that enable QMRA. The Framework for Risk Analysis in Multimedia Environmental Systems (FRAMES) will support pathogen fate and transport model development and QMRA [Whelan et al., 2010]. The flexibility and extensibility of integrated multimedia modeling enables the forward modeling approach outlined by Whelan et al. to be extended, thereby providing a more comprehensive understanding of pathogen fate and transport through an integrated analysis with receptor and general inverse modeling techniques. Furthermore, the powerful analysis capabilities of mature integrated modeling frameworks like FRAMES can aid in the solution of the source characterization approaches discussed and improve handling of uncertainties through application of sophisticated tools such as optimization algorithms, Monte Carlo methods, and uncertainty analysis. The proposed inverse modeling approach is briefly described in the paragraphs that follow.

2.3 General Inverse Modeling

Solution of source characterization problems often involves formulating and solving an inverse problem, since the monitoring data is sparse and contains only faint signatures of the desired system information. Inferring source characteristics, such as source locations, pathogen release, and total fecal loading from the observational data collected at the receptor (scenario illustrated in Figure 1), is a canonical inverse problem. Problems such as this are difficult to solve since they are frequently ill-posed and computationally intensive.

There are several approaches for solving inverse problems encountered as part of a source characterization analysis. Deterministic and stochastic inverse problem formulations are powerful tools that infer input characteristics, given sufficient output data. These formulations, however, rely on a forward model, usually a system of partial differential equations, that describes the dynamic processes of the environmental system and defines the relationship between model inputs and outputs. Development of a forward model would allow us to formulate and solve more general inverse problems than the receptor modeling formulations discussed previously. For example, an integrated pathogen fate and transport model-based inverse problem could incorporate more diverse sources of information — such as chemical kinetics, transport delays, and flow paths — and thus produce better solutions. Solving inverse problems with integrated forward models may prove particularly challenging due to the characteristics of the coupled system of model components that describe transport processes in watersheds. The solution of inverse problems is also several orders of magnitude more computationally intensive than solution of the corresponding
integrated model, since thousands of forward model evaluations are typically required.

Deterministic inverse problem formulations couple the forward model with formal mathematical or heuristic search procedures to determine the model parameters or boundary conditions that best approximate the observed data. Gradient-based search techniques represent the state-of-the-art search procedure for solving inverse problems using the optimization approach. Using an integrated forward model, however, may be difficult as gradient computations would occur across model boundaries, which may lead to solution instabilities or excessive computational effort. Furthermore, gradient-based search techniques, while efficient for well formed problems, are often poorly suited for ill-posed environmental problems with multiple optima and non-linear, discontinuous, and discrete features; and thus, can converge erroneously to local optima, missing the global optimal solution of the problem. One alternative is the use of global optimization techniques, such as evolutionary algorithms, which can provide a more robust search of the decision space [Tryby et al., 2009]. This alternative, however, has a high computational cost that require the application of distributed computing techniques for tractability.

Another potential approach is statistical inversion. Statistical and deterministic inverse models are related and differ in their treatment of information and error. Deterministic inversion is a special case of statistical inversion that takes a frequentist view of probability and assumes errors are normally distributed and additive. Statistical inversion represents information generically using probability distributions and is rooted in a Bayesian view of probability. Bayes’ theorem makes it possible to move back and forth between the conditional and marginal probabilities that are used to represent the observational data and the source model. The forward relationship between the model and the data is captured in the likelihood function that is also a conditional probability. Problem structure dictates the solution technique employed and ranges from analytical solutions to demanding Monte Carlo simulations. Statistical techniques are better able to handle some types of ill-posed problems than deterministic inversion. Given the non-linearities and uncertainties associated with the pathogen source characterization problem, investigation of statistical inversion methodologies is planned.

3 Conclusions

Utilizing source characterization methodology that combines bio-chemical MST markers with receptor modeling techniques, the authors have demonstrated the potential for pathogen source apportionment. Indeed, these results are promising and motivate the authors to continue development of the integrated approach we have described. A multimedia pathogen fate and transport model is currently under development; once it is completed more general inverse modeling methodologies can be explored that will allow the assumption of linear input/output relations to be relaxed and alternate sources of information to be utilized to eliminate false allocations and reduce the uncertainty associated with allocation estimates.

Although receptor, forward, and inverse modeling are different approaches they share the same objectives; integrating them will provide insights not otherwise possible. For example, integrating modeling approaches can inform the QMRA process by improving source term estimation and illuminating the differential fate and transport of indicators and pathogens. Combining source allocation estimates generated using receptor and general inverse modeling with pathogen fate and transport forward modeling and QMRA, the authors have described an approach by which the quality of pathogen exposure risk estimates may be improved.

Acknowledgments

The views expressed in these Proceedings are those of the individual authors and do not necessarily reflect the views and policies of the United States Environmental Protection Agency. These Proceedings have been reviewed in accordance with EPA’s peer and administrative review policies and approved for presentation and publication.
REFERENCES


LIST OF ABBREVIATIONS

bS $\beta$-Sitosterol
Chnl Cholestanol
Chrl Cholesterol
Cm Campesterol
Cp Coprostanol
Ecp Epicoprostanol
Stnl 24-ethylcoprostanol
Strl Stigmasterol
Integrated Modelling for Health and Environmental Impact Assessment of Air Pollution and Climate Change

Stefan Reis¹, Tim Oxley², Ed Rowe³

¹ Centre for Ecology & Hydrology, Bush Estate, Penicuik, EH26 0QB, Edinburgh, United Kingdom, srei@ceh.ac.uk
² Centre for Environmental Policy, Faculty of Natural Sciences, Imperial College London, South Kensington campus, London, SW7 2AZ, United Kingdom, Oxley, t.oxley@imperial.ac.uk
³ Centre for Ecology & Hydrology, Environment Centre Wales, Deiniol Road, Bangor, Gwynedd, LL57 2UW United Kingdom, ecro@ceh.ac.uk

Abstract
Modelling the impacts of air pollution and climate change on human health and ecosystems in integrated assessment models (IAMs) has emerged as a key tool to inform policy decision making, where simplistic solutions are unlikely to deliver efficient and sustainable pathways for future development.

Model integration is facing a complex set of challenges in different dimensions, as integrated models have to be:

- Spatially explicit and of sufficiently high spatial resolution for their respective domain, with nesting approaches providing the integration across different spatial scales.
- Temporally dynamic to model system responses and recovery e.g. pollutant accumulation, time-lag (e.g. of measure implementation) and time-bomb effects. Due to different temporal horizons for different processes (e.g. days-years for air pollution, decades-centuries for climate change, centuries-millennia for accumulation of heavy metals/POPs in soils), integrated models also need to nest models with different temporal resolution.
- Sectorally detailed to model trade-offs and synergies and to allow for the representation of paradigm-shifts (e.g. in energy systems) and behavioural changes (e.g. non-technical measures).
- Accessible, providing clear illustrations of inter-sectoral synergies and trade-offs (e.g. ammonia emission reduction vs. nitrate leaching in agriculture) using visualisations and multi-media.

In addition to the aforementioned requirements, integrated models need to be flexible and scalable to be able to provide answers to varying problems. This paper discusses current challenges faced by IAMs and emerging developments based on a literature review.

Keywords: integrated modelling; integrated assessment; air quality; climate change; human health; ecosystems.
1. INTRODUCTION

1.1 The situation of integrated assessment modelling in Europe today

In recent years, a number of anniversaries and milestones could be observed in relation to European air pollution control activities. To begin with, the UNECE Convention on Long Range Transboundary Air Pollution (CLRTAP) celebrated its 30 year anniversary, having spawned 8 protocols since it was established in 1979 (Sliggers and Kakebeeke, 2004). And while the CLRTAP was initially conceived and driven by a single purpose, to combat transboundary air pollution identified as the main cause for the Waldsterben (forest dieback) in the 1970s, it had soon evolved into a comprehensive hub for monitoring and modelling of air quality and action to reduce the burden of air pollution for human health and ecosystems in the UNECE region.

Within the CLRTAP, especially the 1999 Protocol to Abate Acidification, Eutrophication and Ground-level Ozone (known as the Gothenburg Protocol), the target year 2010 was selected as an intermediate step towards closing the gap between critical levels and loads and observed exceedances of for the protection of health and ecosystems. Having now reached this target year, it is timely to reflect upon the development so far and the challenges ahead. This is of particular importance, as the Gothenburg Protocol – as well as the EC National Emissions Ceilings Directive – are currently under revision and aim at setting new targets for the year 2020 and beyond, with aspirational targets are discussed for 2050 (see http://ec.europa.eu/environment/air/pollutants/rev_nec_dir.htm).

As another relevant anniversary, it has been 20 years since the publication of Alcamo et al. (1990), which marks a starting point for many years of continuous development and application of the RAINS (Regional Air Pollution INformation and Simulation). The RAINS model has been an essential tool for the analysis of alternative strategies to reduce acidification, eutrophication and ground-level ozone in Europe, underpinning evidence-based policy decision making. In the past 20 years, its focus has evolved as well towards integrating air pollution control and climate change aspects into the latest version of the GAINS (Greenhouse Gas and Air Pollution INteractions and Synergies) model.

Finally, the past decades have been marked by a strong trend towards appreciating the complex relationships and connections between different drivers of global environmental change. The year 2010 has been declared by the United Nations as the International Year of Biodiversity, and current research into the influence of the human perturbation of the nitrogen cycle has identified many areas where air pollution control, climate change and the protection of vulnerable ecosystems cannot be separated (Erisman et al., 2008). It is thus timely to conduct a review of the current understanding of, the state-of-the-art in science and the availability of tools for the integrated assessment of air quality and climate change.

1.2 What is Integrated Assessment (Modelling)?

The term Integrated Assessment (Modelling), in short, IA(M), is widely used, however, in different contexts and within a variety of scientific disciplines and policy applications. For the purpose of this paper, we will briefly discuss two definitions, firstly that of CIESIN (1995):

"An assessment is integrated when it presents a broader set of information than is normally derived from a standard research activity. Because integrated assessments bring together and summarise information from diverse fields of study, they are often used as tools to help decision makers understand very complex environmental problems."

With regard to the modelling aspect, CIESIN (1995) further defines:

"Integrated assessment modelling is a tool for conducting an integrated assessment."

The two activities, however, are not identical even though the terms are often confused and used interchangeably. Integrated assessment models (IAMs) are mathematical computer
models based on explicit assumptions about how the modelled system behaves. The strength of an IAM is its ability to calculate the consequences of different assumptions and to interrelate may factors simultaneously, but an IAM is constrained by the quality and character of the assumptions and data that underlie the model.

A similar definition is available from The Integrated Assessment Society (TIAS, 2010) and states:

“Integrated assessment (IA) can be defined as the scientific "meta-discipline" that integrates knowledge about a problem domain and makes it available for societal learning and decision making processes. Public policy issues involving long-range and long-term environmental management are where the roots of integrated assessment can be found. However, today, IA is used to frame, study and solve issues at other scales. IA has been developed for acid rain, climate change, land degradation, water and air quality management, forest and fisheries management and public health. The field of Integrated Assessment engages stakeholders and scientists, often drawing these from many disciplines.”

Both definitions highlight the interdisciplinary nature of IA(M) and its role to provide evidence for policy making, as a tool to inform and support policy decisions. It is interesting to note that while the CIESIN (1995) definition has emerged from a climate change background, the work of TIAS has strong roots in water management and sustainable development.

1.3 Aims and scope

A global review of all aspects of IA(M) would by far exceed the limits of a single paper; hence this review will focus on the following key aspects:

- Integrated assessment modelling of air quality and climate change in general and
- Specifically, how IA(M) has informed policy making in the context of the UNECE CLRTAP and the EC NECD.

In addition to that, the following discussion will highlight:

- Major achievements in the evolution of IA(M)s in the past 20 years, as well as
- Key future challenges for IA(M)s with regard to their development, application and policy support role.

2. HISTORIC DEVELOPMENT OF IAM – EUROPE AND BEYOND

2.1 Existing models and their evolution

A wide range of IA(M)s has been developed over time and has been, respectively is currently, applied in scientific research projects and for direct policy support. These can, for instance, be distinguished by their spatial coverage, by environmental compartment, by topic, their degree of integration and so on:

- Global climate change modelling (e.g. IMAGE, ICAM , MERGE , IGSM )
- European air pollution control and greenhouse gas reduction strategies (e.g. RAINS/GAINS, see Höglund-Isaksson L. and Mechler R., 2005; Klaassen et al., 2005; Tohka et al., 2005; Winiwarter et al., 2005; Hordijk, L. and Amann, M., 2007)
- National modelling of air pollution control and greenhouse gas emission reductions (e.g. UKIAM for the UK, MINNI for Italy)
- Models for integrated assessment of water resource allocation and contamination (e.g. Letcher et al., 2007)
Furthermore, we can distinguish between models developed primarily for policy decision support (applied, operational) and models for scientific research (often process based, technical). While the latter are often more closely linked to new scientific developments, the former reflect the demands of policymakers. Transforming a model suite from its scientific development stage into an operational, applied stage is not straightforward, and funding for such work is often elusive since the effort is unlikely to be rewarded in the conventional scientific “currency” of peer-reviewed publications.

As indicated above, the abundance of different models does not allow making general statements for the whole domain of IAM, yet, a few common trends can be observed:

1) Many models have evolved to extend their scope, e.g. incorporating additional topics (air quality based models extending to climate change, multimedia models covering air, water, soils and working across environmental compartments)

2) Spatial integration, both up- and downscaling and/or nesting is frequently being conducted (e.g. regional models downscaling to national/local applications)

3) Addition of further process detail to existing stand-alone models and extended coupling/linking of specialised models takes place (e.g. the integration of process-based agricultural models into RAINS/GAINS)

4) Development of object-orientated and semantic frameworks to wrap and link component models (e.g. the OpenMI framework, http://www.openmi.org)

These developments reflect the growing understanding of the complex connections between a variety of environmental problems on local to global scale and represent a trend towards a systems approach to problem solving, in contrast to the single-problem, one-purpose strategies that had marked the initial stage of (environmental) research and policy development.

2.2 Emerging challenges

As indicated in the previous section, the development and application of IAMs face quite a few challenges. In the field of extending models to include additional topics - for instance when integrating impact assessment of air pollution and climate change - different time scales and time steps need to be taken into account. While air pollution effects typically occur in a matter of hours or days where human health is concerned, or within days to years for ecosystem effects, climate change effects are likely to occur in decades and centuries. This is of particular relevance when trying to assess the ration between the costs of action and the benefits, as arriving at a monetary evaluation of effects that may occur over a period of decades or centuries to compare on equal terms is not a trivial task.

A similar challenge occurs when trying to integrate across different environmental media, e.g. air pollution control and effects of the deposition of air pollutants on water and soil quality. Last, but not least, research into ecosystem effects has been moving towards more dynamic approaches (see for instance Joint Expert Group on Dynamic Modelling under the Working Group on Effects, http://nora.nerc.ac.uk/8658/). As current IAMs typically operate on an annual scale and deliver annual average values as output, integrating dynamic modelling results on the effects side will likely require multi-year assessment runs and more detailed temporal profiles within annual assessments.

The issue of integrated assessments across different time scales is closely related to that of dealing with varying spatial scales. Climate change is a global phenomenon, whereas air pollution effects typically occur on regional to local scale, with distinct hotspots due to differential deposition onto and damage to different landscape components. But not only effects are spatially explicit; the sources of pollutants and precursors are highly spatially variable, and the location of observed effects does not always coincide with source locations. With regard to the spatial representation in models, again two trends can be observed: on the one hand, applying dedicated models at different scales, with the potential
of nesting or linking model input and output across scales and on the other hand, integrating for instance local indicators into regional models, using derived functional relationships. The latter approach has been taken e.g. to include an urban increment to air pollution exposure of the urban population in the European-scale RAINS model (Cuvelier et al., 2007). Examples for a nested/scaled approach can be found in the national implementations of IAMs, as they have been developed in Italy (RAINS-ITALY, Zanini et al., 2005) or the UK (UKIAM, Oxley et al., 2004). The accuracy of the spatial results of IAMs is of particular importance for the development and implementation of national policies.

2.3 Different trends in integration and towards complexity

The previous section has highlighted two main trends emerging with the development of European IAMs, increasing levels of integration and complexity on the one hand, and a modularisation or disaggregation into individual models for specific tasks/scales on the other hand. Both developments have advantages and caveats.

Any increased complexity of models may render the interpretation of results and the assessment of uncertainties substantially more difficult (Warren, 1999; ApSimon, 2002; Krysanova, 2007). In addition to that, the relationships between changes in parameters and the response observed in model results are often not straightforward to predict. In contrast, applying different models for different purposes often provides robust, individual results, yet faces the difficulty of combining or integrating results based upon very different model formulations into policy relevant scientific evidence.

3. CURRENT STATE AND FUTURE DEVELOPMENTS

3.1 What next?

As it has been discussed in the previous sections, IAMs are widely applied in providing policy decision support in particular in the development of integrated air pollution control strategies. There is a trend towards extending models that were primarily developed for air pollution control into the realm of climate change, both with regard to greenhouse gas emission reductions and the quantification of changes in radiative forcing (e.g. Dentener et al., 2005). Further to that, a growing community of national scale IAM developments in Europe is fertilising the ground for a drive towards a larger knowledge base both regarding model development and application for policy decision support on different levels.

The extension of European IAM to further include climate change aspects is reflected by an increasing interest in longer time scales, for instance regarding energy and emission scenarios and aspirational targets for the year 2050 and beyond. Such long time scales have typically not been relevant for the assessment of air pollution alone.

Interactions between air pollution and climate change have primarily focused on CO₂, but the global nitrogen (N) cycle also strongly interacts with global climate processes, via effects on primary production and on trace greenhouse gas production. Recent developments involve the more complex perturbation of the global nitrogen cycle and feature biochemical process models that allow for a quantification of nitrogen input and losses at different stages of the cycle. Nitrogen species are closely linked to air pollution effects as well as contribute to climate change, yet the nitrogen cascade spans not only air pathways, but affects soils, freshwater and marine ecosystems through biochemical transformation and physical transport processes. The key difference between modelling carbon and nitrogen in IAMs, however, is the relevance of spatiotemporal aspects for the representation of N effects compared to a more simplistic mechanism that is sufficient to quantify the effect of CO₂ equivalents on e.g. global temperature increases.

IAM in Europe has covered both health impacts and ecosystem effects from an early stage. However, it has to be stated that due to the comparatively more advanced knowledge on the monetary evaluation of health effects (and the lack of a comprehensive approach for a similar valuation of ecosystem effects to date), health impact costs have been the main
driver for assessment results in recent years. Particulate Matter (PM) has thus had a strong influence on the priority setting for air pollution control, while acidification, eutrophication and ground level ozone have been of less importance until recently. The revision of the Gothenburg Protocol and the EC National Emissions Ceilings Directive will lead to an inclusion of PM for instance. Emerging evidence on the relevance of ecosystems for carbon sequestration, as well as the concept of ecosystem services as a means of quantifying the benefits from natural ecosystems have somewhat changed this again recently.

Another challenge can be identified regarding the use of IAMs for ex-ante or ex-post cost-benefit assessment (CBA) of environmental policy, as it has for instance been conducted by Kelly et al. (2010). The quantification of health impacts in monetary terms is – even acknowledging the substantial uncertainties in this field – more advanced as it is the case for ecosystem impacts. Thus, a full scale comparison of all costs and benefits of a policy measure is – at this time – not feasible. Instead, most often policies designed to achieve compliance with individual directives or protocols are evaluated, lacking a full and meaningful integration and the quantification of co-benefits and spill-over effects of potentially conflicting (environmental) policy targets. A full-scale integrated assessment in this area requires further advances in evaluation methodologies and a consistent framework for a monetarisation of ecosystem effects in a similar fashion as it is being done for human health effects. For both areas, health and ecosystems, however, the underlying scientific evidence for the quantification is currently scarce and larger scale empirical studies in a European realm are needed.

Apart from these specific topical aspects, a potentially greater challenge is the lack of a common framework or concept for the development of integrated assessment models. The following two sections will briefly discuss two key issues arising from this.

3.2 Degree of integration

The first question to tackle is how to achieve integration, and what measure or indicator can serve to distinguish integrated from partial models. This is not just of academic relevance, as IAMs, as alluded to previously, are widely used in support of policy development and as it is unlikely that one single model will satisfy all policy needs, a way of determining and describing the level of integration a model represents can serve as a core selection criterion.

It is difficult to find a comprehensive indicator for the degree of integration that current models are reflecting. One angle that could be taken is to measure the components of IAMs to in how far they cover a full-chain impact assessment, as for instance represented by the DPSIR framework (EEA, 2010), an extension to the OECD Pressure-State-Response (PSR) model (OECD, 1993). The DPSIR framework describes the causal chain from the origin of an environmental problem to its outcome, covering the following stages:

- Driving forces
- Pressures
- States
- Impacts
- Responses

While the DPSIR model has been modified and applied in different research areas (e.g. Morris et al., 2006). However, for the purpose of analysing levels of integration, the original DPSIR framework is well suited. A similar approach has been developed and applied in the frame of the ExternE (http://www.externe.info) project series developing a framework for impact assessment and external costs of energy, transport etc. and was termed the “impact pathway methodology” (see Mensink et al., 2007).

However, any indicator or concept of “integration” on its own account is not sufficient or suitable to assess the quality of an IAM. Specialist models representing only selected parts of the DPSIR chain can equally be marked by a high coverage of all aspects relevant for a
specific problem or task. There remains a need for methods to evaluate the quality and suitability of an IAM for answering a given set of questions.

Current applications of integrated assessment models have most often emerged from a specific area of research or with a well-defined policy question to answer. With an increasing understanding of the complex relationships of these specific issues integration has then occurred by extending the system boundary of the models and including additional parameters, datasets and modules. This approach requires either the scalability of a modelling concept from the start; alternatively, substantial conceptual rethinking and redesign of existing models are needed.

In the field of Earth System Modelling (ESM) approaches exist to design and apply common frameworks to support the development of integrated models. An example of this approach can be found with the Earth System Modelling Framework (ESMF, http://en.wikipedia.org/wiki/ESMF).

3.3 Methodological challenges

With regard to the assessment methods implemented in current IAMs, one could easily classify or categorize along a vast number of different concepts or topics, for instance by

- the timing of the assessment (ex-ante or ex-post)
- design (simple one-dimensional vs. complex multi-dimensional)
- application (decision support system or optimisation tool)
- spatio-temporal resolution (short term to long term, local to global)
- topic (air quality, water quality, catchment modelling, climate change, ...), etc.

What needs to be kept in mind, however, is that most IAMs that are currently applied for policy decision support have not (or not entirely) been designed and implemented strategically for this purpose, but have often evolved over extended periods of time, reacting to emerging policy needs. In this process, models have at times begun their life cycle as a specialist, scientific tool and matured to more easily accessible tools that may be operated by non-expert users. However, the gradual evolution of models often results in legacies which can seriously affect the performance and flexibility of their application, e.g. due to programming or hardware restrictions imposed on previous versions that have been long overcome by current technological progress.

4 CONCLUSIONS AND OUTLOOK

4.1 Conclusions

The previous sections have highlighted a few of the many issues marking the complexity of the field of integrated (assessment) modelling in its current state. This complexity exists in two different, but closely related, dimensions: the complex design and structure of IAMs as a challenge to the methodological and conceptual development of models on the one hand, and the difficulty to represent complex biochemical, physical or economical/social processes in assessment models on the other hand.

With regard to the latter, we can observe a substantial improvement in the understanding of the interactions between different environmental problems and a strong drive towards a more integrated approach in solving them. This is for instance the case in tackling air quality and climate change in combination, taking full account of the co-benefits and potential spill-over effects of individual measures in a common framework (see for instance Pleijl et al., 2009).

While understanding the need for integration helps to focus research into the interactions and dependencies of the underlying processes, the modelling community has been actively
discussing concepts, model linkages and interactions. In this context, Harris (2002) sees 
IA(M) as an essential and systematic way forward, in connection with Earth System 
Modelling (ESM), Natural Resource Management (NRM) and Ecological Sustainable 
Development (ESD). Around the same time, Jakeman and Letcher (2003) derive common 
features of IA starting from an example in catchment management. This list of features 
contains, among others, the “Connection of complexities between natural and human 
environment; recognition of spatial dependencies, feedbacks, and impediments; an 
iterative, adaptive approach.” (Jakeman and Letcher, 2003, p. 492). More recently, 
Jakeman et al. (2006) propose “Ten iterative steps in development and evaluation of 
environmental models”, which could form a basis for a comprehensive framework for IAM 
development that is currently lacking (see Section 3.2). A more systematic and formalised 
approach towards the development and implementation of IAMs could not only be 
beneficial for knowledge transfer and collaboration between modellers, but as well help to 
inform the users of IAM output with regard to uncertainties (quantified) and a “better 
qualitative understanding of the system” (Jakeman et al., 2006). The aspect of the usability 
of tools for policy-relevant research is highlighted as well by McIntosh (2007).

From the general trend of discussions in literature, a development to a more systematic, 
methodological approach towards IAM design and application has consolidated in the last 
five to ten years.

Looking forward, there is no lack of emerging topics and some of these have been the 
subject of recent publications. D’Elia et al. (2009) focus on the integration of non-technical 
measures (NTM) into IAM, which has been discussed vividly in the European IAM 
community for some time. This reflects a growing concern that more ambitious 
environmental targets are likely not achievable using technological control measures alone. 
At the same time, the integration of behavioural and structural change (which marks most 
of those non-technical measures) is a non-trivial task as most models have been built with a 
focus on end-of-pipe control options and established energy systems, which cannot be 
easily overcome. Closely related is the conceptual integration of external costs into the 
IAM process, which has been extensively done for human health effects (see Section 3.2), 
but has gaps when it comes to ecosystems or climate change effects. Kosugi et al. (2009) 
describe an approach linking an IAM with a model for life cycle assessment (LCA) to fully 
internalise external costs of air pollution and climate change, as well as land-use and land 
cover change.

Another trend could be the modularisation of IAMs, as described by Hinkel (2009) for a 
specific model. Making IAMs modular does not yield improved models per se, but could 
enable their linkage using concepts such as OpenMI and increase the interoperability and 
flexibility of the IAM by allowing to select different modules for specific tasks. Yet, for 
this to work, models and modules need to be described and documented in a consistent way 
that is accessible across disciplines and scientific domains. Janssen et al. (2009) elaborate 
on the use of a common ontology to achieve such a level of integration in a large-scale 

4.2 Outlook

The development of IAMs faces a lot of challenges, but at the same time has the potential 
to mature into an essential and indispensable tool to provide underpinning scientific 
evidence for informed policy decisions to address the critical issues of today’s global 
environmental change.

Among the various challenges highlighted in this paper, the authors see two emerging areas 
as key to achieve progress in integrated assessment modelling: on the one hand, a 
conceptual framework and advances for a better spatio-temporal representation of cause 
and effects in both health and ecosystem impact assessment is required, while on the other 
hand a more comprehensive integration across different environmental media and 
environmental pressures has to be realised. Both are not trivial and the necessary 
complexity of models may be a limiting factor. For this purpose, an overarching concept of 
modularisation and linking of models and modules may be seen as the best way forward.
This would as well further aid the application of methods for in-depth, quantitative uncertainty assessment, which will be vital to provide robust scientific results to underpin evidence-based policy development.

5 REFERENCES


'Integronsters' and the special role of data

Alexey Voinov
University of Twente Faculty of Geo-information Science and Earth Observation (ITC)
P.O. Box 6, 7500 AA, aavoyn@gmail.com

Abstract: In many cases model integration treats models as software components only, ignoring the fluid relationship between models and reality, the evolving nature of models and their constant modification and re-calibration. As a result, with integrated models we find increased complexity, where changes that used to impact only relatively contained models of subsystems, now propagate throughout the whole integrated system. This makes it harder to keep the overall complexity under control and, in a way, defeats the purpose of modularity, when efficiency is supposed to be gained from independent development of modules. Treating models only as software in solving the integration challenge may give birth to 'integronsters' - constructs that are perfectly valid as software products but ugly and useless as models. We argue that one possible remedy is to learn to use data as modules and integrate them into the models. Then the data that are available for module calibration can serve as an intermediate linkage tool, sitting between modules and providing a module-independent baseline dynamics, which is then incremented when scenarios are to be run. In this case it is not the model output that is directed into the next model input, but model output is presented as a variation around the baseline trajectory, and it is this variation that is then fed into the next module down the chain. The Chesapeake Bay Program suite of models is used to illustrate these problems and the possible solutions.

Keywords: Modularity; Calibration; Integrated modeling; Chesapeake Bay; Module linking; Components

1. INTRODUCTION

As our impacts on the environment become more dramatic more interest is drawn to analysis of systems that span over several traditional scientific disciplines. These systems are more complex and have to be described by models that may have various components characterized by different scales, resolutions, and developed under different assumptions and paradigms coming from different scientific traditions and backgrounds. Integrated modeling is the method that is developing to bring together diverse types information, theories and data originating from scientific areas that are different not just because they study different objects and systems, but because they are doing that in very different ways, using different languages, assumptions, scales and techniques.

There may be two ways of doing integrated modeling. One is to build the model as a whole. Here the modeling team collects data and information from various scientific fields, processes it, and translates it into one formalism. This is how some of the well known integrated models were built. For example, the Club of Rome models resulted in the famous World 3 model by Meadows et al. (1979), which integrates information about agriculture and food production, industry, demographics, non-renewable resources, and pollution (http://www.whole-systems.org/world3.html). Another example is the Global Unified Metamodel of the Biosphere (GUMBO) (Boumans et al., 2002, http://ecoinformatics.uvm.edu/projects/the-gumbo-model.html). Here knowledge about geology, global climate, sociology, economics, atmospheric processes, and ecosystems is integrated in one model, designed to run various scenarios of global change. A significant feature of these models is that they are developed and maintained by the same team as a whole.
A. Voinov. ‘Integronsters’ and the special role of data

The other approach to integrated modeling is to put together already built models. For example, the modeling suite developed by the Chesapeake Bay Program (Cerco, 2000; Linker et al., 2000; Wang and Johnson, 2000) consists of three major parts:

- The atmospheric transport model that produces atmospheric deposition predictions for nutrients and other constituents. The Airshed Model is based on the Community Multi-Scale Air Quality modeling system (CMAQ). The latest CMAQ (2009) code runs on a 12 km fine grid in the Chesapeake region, with a 36 km grid used for the continental scale boundary conditions;
- The watershed model, a highly-modified version of HSPF (Linker et al. 2000, Bicknell et al. 1996), that produces loadings that come from the land into the estuaries;
- The estuary model, which is, itself, a combination of three linked models, a hydrodynamic model (Johnson et al. 1993), a eutrophication model (Cerco and Cole 1993), and a sediment diagenesis model (DiToro 2001). This Water Quality and Sediment Transport Model (WQSTM) is a three dimensional model of the tidal Bay comprised of 57,000 cells. Currently it represents transport processes, eutrophication processes, and living resources such as submerged aquatic vegetation and benthos.

This modeling system has been developed over the past 25 years and went through many phases. The models are linked loosely, so that there is no formal software mechanism involved. Mostly output from one model is sent as input into the other model as a data file. Yet still, the models work in concert and are used for decision making purposes as a suite. Here integration is performed at the level of existing models that are taken from different fields of science as modules.

Since there are already numerous legacy models that have been carefully designed and tested in numerous applications by skilled researchers who are specialists in their corresponding fields of science, it makes perfect sense to try to reuse their products as building blocks for more complex systems. We only need to make sure that they can be linked together in a meaningful way matching the variables, scales and resolutions. Currently there are several efforts to develop the standards and software tools that would provide for this kind of integration. For these purposes, models are merely treated as software components that are to be made to work together and talk to each other.

The two approaches to integrated modeling bear their own caveats and complications, which we consider below:

2. INTEGRATED MODELS

Building integrated models is a rather conventional approach as we can see from their history. In most cases such models were built by integrating knowledge from different fields of science in the form of data, concepts, functions, approximations, etc. The models themselves were built from scratch and little care was given to future reuse of the pieces that went into the model formulation. The models were not designed for reuse or modification by others outside of the team.

At some point it became clear that the model components could be reused or modified for future applications, and models started to be packaged as modeling systems. The Modular Modeling System (MMS, Leavesley, et al., 1996) was one of the first attempts to make modules available for reuse. Another example of modular modeling is the Library of Hydro-Ecological Models (LHEM, Voinov et al., 2004). Here modules were developed as a mix of Stella and C++ code, together with the Spatial Modeling Environment (SME) used to stitch the modules together and take care of all the input and output.

A more recent and sophisticated development is the Object Modeling System (OMS) by the US Department of Agriculture (David, et al., 2002; Kralisch, et al., 2004; Ahuja, et al., 2005). OMS requires rewriting modules in Java or C# to be then inserted into the system library.

Another framework in this same category is the MIMOSA developed by CIRAD (Müller, 2009, http://mimosa.sourceforge.net/). It is a modeling and simulation platform, which can
support the whole model building process, starting from the stage of conceptual models and up to running and analyzing simulations. The specification uses ontologies and an extensible set of formalisms to present system dynamics, initialize and visualize the model. The simulation kernel is based on DEVS principles (Discrete Event System Specification; Zeigler, 1976).

There are also proprietary platforms that bear promise in supporting integrated modeling such as Extend and Simile.

Note that all these frameworks are not really designed to integrate legacy code as is, but rather they serve as tools for future model development, offering input, output and project support tools and common library standards, into which different modules are invited to be contributed. The future of these frameworks will very much depend upon how wide their standards will be accepted within the modeling community and whether there will appear a critical mass of contributed models. So far, unfortunately, there are very few examples of frameworks being used beyond their respective development teams. Since each module has to be rewritten in the language of the framework, there is less worry about the consistency of models that these frameworks are producing. All the scaling and linkage issues are expected to be solved when modules are tailored, adjusted and added to the framework libraries.

3. INTEGRATING MODELS

The alternative approach is to support model integration at the level of existing models, allowing these models to talk to each other directly.

Some of the most advanced attempts of such model integration come from physics and engineering. For example the Common Component Architecture (CCA) is developed by the Department of Energy and Lawrence Livermore National Lab teams (Bernholdt, 2004). The design places minimal requirements on components and facilitates the integration of existing legacy code into the CCA environment by means of the Babel (2004) language interoperability tool, which currently supports C, C++, Fortran 77, Fortran 90/95, and Python. The CCA is being applied in a variety of disciplines, including combustion research, global climate simulation, and computational chemistry. At this time CCA is adopted as the integration engine by the Community Surface Dynamics Modeling System (CSDMS; Peckham, 2010).

The US Environmental Protection Agency (EPA) has been developing the FRAMES (2009, 2009a) (Framework for Risk Analysis in Multi-media Environmental Systems) system to manage the execution and data flow among the science modules. 3MRA (2009) (Multi-media, Multi-pathway, Multi-receptor Risk Analysis) (Babendreier and Castleton, 2005) is a collection of 17 modules that describe the release, fate and transport, exposure, and risk associated with various contaminants. FRAMES was developed as the framework that would allow these modules to communicate with each other.

The Open Modeling Interface and Environment (OpenMI, 2009) developed by a consortium of European universities and private companies, is a standard for model linkage in the water domain (Moore, et al., 2005). The OpenMI standard defines an interface that allows time-dependent models to exchange data at runtime. When the standard is implemented, existing models can be run in parallel and share information at each time-step.

All these systems are primarily about software issues and the major concern is how to link models as software components. There always seems to be a software solution that we can find, and the promise is that we will be able to make models work together and exchange information at runtime. Actually, models are more than just software and there are issues that may not be easy to deal with at the software level. This takes root from the long lasting discussion about modeling as science or art. Let me try to explain what I mean by using some metaphors.

3.1 “Ugly” constructs
There is certainly some beauty in the most well-known models from physics. Agren and Bosatta (1990) have called for ecologists to use similar esthetic criteria to guide them towards useful theories in their domains. They asked: “Are we like construction engineers who mainly care about getting the house built, or like artists who try to capture the essence of the house in our paintings?” (p.213). I would argue that similar principles should apply to model coupling. In Fig.1A we can see an example of a somewhat esthetically pleasing effort in module integration. Coupling modules in a different way can produce results that are way less attractive for an average eye (Fig.1B). It may be hard to explain why one coupling is nicer than the other one: technically the linkage process in both cases has followed quite similar rules. By following only technical principles we may be giving birth to ‘integronsters’ - constructs that are perfectly valid as software products but ugly and useless as models.

3.2 Skewed geometry

Geometry and spatial resolution of a system is very important. When building a model we tend to imply much about the spatial representation that we use. In many cases the system boundaries are not quite carefully spelled out and the spatial resolution and system topology are assumed as a given. This works nicely as long as the modelers themselves are the ones who use the model. If the model is packaged as a module and offered for reuse and recombination, all these hidden and unsaid features become crucial. We can imagine a geometrically consistent system as in Fig.2A, combined with another geometrically consistent system as in Fig.2B. The result of this integration shown in Fig.2C makes little sense in our normal Euclidian space. Even while the connection was made in a reasonable way: a lower leftmost column was connected to the upper leftmost column and so on. The overall orientation of the object may have not been part of the module description. The result, while technically correct, becomes meaningless.

3.3 Mismatched scales

In environmental sciences we find various combinations of processes and systems, and each one is characterized by its own specific scale and resolution in space, time and complexity. While linking in time and space is mostly a technical problem and can be resolved by appropriate software and documentation, it is less clear how to communicate information from one complexity level to another, how to estimate system responses across scales or levels. The problem is that each model is built for its particular goals and these goals and scales are also specific for the field of application and discipline. For example, economic models mostly operate in terms of global equilibria, which may remain unchanged for months or even years. They also deal with variables that may be averaged across regional scales. It would be inappropriate to blame the economists that their models do not get even close to the resolution and complexity of, say, hydrologic models. If we need to make these two types of models work in concert as components, we find that the complexity of these components is in different dimensions, which would...
be hard to compare. In a way one model component in this case can be compared to a bull in a china shop: it simply does not belong there. Yet for a transdisciplinary system analysis, we need to integrate the two modules together.

There have been many attempts to couple models from different levels of organization ranging, for example, from leaf to ecosystem (Anderson et al., 2003). One approach connecting such models is to extrapolate or average the results obtained from a detailed level to a higher level (Ewert et al., 2006). In other cases we may want to focus on the critical thresholds, when the more detailed model will be analyzed beforehand for conditions that can be like a red flag for the rest of the system. For example, the hydrology model may be needed only to generate the extreme, flood events, which are important for the economy model. By estimating various important relationships from model simulations at lower levels, and then using the derived parameters as inputs for higher-level models, we can reduce the computation time. However, we may be missing some complex interactions and feedback mechanisms (Ewert et al., 2006).

3.4 Confusion of tongues

Different scientific disciplines speak different languages. This makes it only harder to integrate the models that they produced. There is much hope that the ontological approach (e.g. Hřebíček, Kisza, 2008; Rizzoli, et al., 2008; Athanasiadis, et al., 2008; Janssen et al., 2008; http://www.apesimulator.org/OntologyBrowser.aspx) can help to find a common way for communicating information. But ontologies emanating from different scientific fields are likely to be different and additional effort will be needed to synchronize them. Otherwise we are still likely to experience the problems described in the story about the Confusion of Tongues (Fig.4). This transdisciplinary synchronization of ontologies is hard to automate and standardize, especially when we are trying to use legacy models.

3.5 Do it together

Most of the above listed problems can be resolved if the modeling process is conducted as a community effort (Voinov et al., 2010). A community for model building, such as the CIEM (Community for Integrated Environmental Modeling - http://groups.google.com/group/commiem), which is based on open source and follows some principles similar to the ones developed by OGC (Dibner, Arctur, 2008), is certainly a very promising endeavor.

According to wiki-answers it is Geoffrey Haley who said that “When a collection of brilliant minds, hearts and talents come together ... expect a masterpiece”. This may be true but it is still hard to imagine how a painting like Mona Lisa could be painted by Leonardo working together with Michelangelo and Raphael in a community effort. More likely they would have produced some kind of a modular product as in Fig.5. Do we really expect masterpieces from community efforts or rather they should be focusing on routine work, serving the needs and goals of particular applications?
We see that model integration and integrated modeling may need efforts that go beyond software development and require research in community building, social networking, semantics and modeling methodology.

4. MODEL INTEGRATION AND DATA

Another important issue that distinguishes models from mere software is how they relate to data. This is especially evident for environmental and socio-economic applications, where in fact models are rarely based on sound theory and mostly represent empirical generalizations. Even in hydrology, where there is a solid theoretical background, the calibration phase turns out to be essential. No matter how undesirable calibration is, especially when we need to run scenarios for global change and explore the domains that are not well covered by data, using models without calibration is risky, to say the least.

Jetten et al. (1999) compared the performance of some 14 catchment models and found that uncalibrated use of models is not advisable, especially for small and medium scale catchments. Similar conclusions were drawn by Bormann et al. (2007): when calculating scenarios that were out of the calibration domain, different models produced quite different results.

The promise of module integration has been most prevalent for process-based models. The idea is that by representing the processes correctly in a model, we can then couple models of these processes together and produce simulations of more complex systems, where processes are linked and exchange material and information. This becomes less attractive, if we need to deal with empirical models, which are always dependent on data. The problem is that in environmental sciences there are hardly any really process-based models. What is a truly process-based model as against an empirical, regression one? In any process-based model we actually utilize certain empirical generalizations, rather than true process description. For example, there is hardly an adequate detailed biophysical molecular description of the photosynthesis process to be found among the models of vegetation growth, instead some variations of Michaelis-Menten kinetics are applied, which are already empirical generalizations of the process. Nevertheless these models claim to be process-based. As we go to larger systems, such as landscapes, we will need to employ even more generalized formalizations (Voinov, et al., 1998). We tend to forget that what we assume to be descriptions of processes are in fact only empirical generalizations of these processes. For almost all so-called process-based models, they actually describe processes only at a certain level of abstraction, and become empirical beyond that. This is probably why most of the process-based models still need to be recalibrated when applied to new areas and study cases. The more different the environmental conditions – the more recalibration needed.

Silberstein (2006) claims that “modellers who focus on their model without continual reference to real data are not really scientists but artists. They have their place, and indeed their ideas may well turn out to be useful, but their activities are not science if they are not base on observations”. But if calibration is still an essential task when integrating models, the promise of building new models of complex systems from modules, like from Lego blocks, becomes less feasible. If we still need to do calibration, we need to remember that calibration of complex models is even more complex. As more components are brought together, the calibration of the whole model becomes only more difficult.

An integrated model is made out of two or more relatively independent components. Each of these components can operate on its own and in many cases has been developed
independently by separate groups of researchers. The promise of integration in this case is that legacy code and models can be reused to analyze more complex systems, while analysis is simplified since the overall system can be studied and modeled in portions. Suppose we have model A for one sub-system and model B - for another. If a system is a composite of these two sub-systems then instead of building a whole new model C to present it, it should be possible to use an integration of the two existing models and model the system with an AB model, where the two components A and B exchange information as they run. For simplicity let us assume that there is only a one-way flow of information from A to B.

What should be the calibration process in this case? If A and B existed before, they were most likely calibrated previously with observation data. If not - still it should be much easier to calibrate smaller and simpler components, therefore we may assume that A and B are calibrated separately. However, obviously, after integrating A and B we should not expect that the resulting output from the AB model will match data as well as output from B. Note that for our system it is the output from B that matters, since B generates output for the integrated model AB.

The reason that AB will generate somewhat different results than the calibrated standalone B model is that now the forcings for B come from the calibrated model A, instead of being taken from data. No matter how well A is calibrated, its output is likely to deviate from the data, and will steer AB away from what B previously was generating. The results can be still improved by some further calibration of B as part of the integrated system. In this case this will look like a calibration of the AB model as a whole. This will also likely include some refinement of overall model performance by model B calibration, in fact, compensating for some deficiencies in the calibration of model A. Moreover, we may be even tempted to further tweak some of the parameters in A to get a better match. This is exactly what is done in integrated models: they need adjustment and re-calibration after the components are put together. However by doing this we lose much of the advantages of the modular architecture in the integrated AB model. We now need to deal with the calibration of the full model, and whenever any changes are made to any of the components, the other components need to undergo new calibration as well.

For example, for the Chesapeake Bay decisions are mostly based on predictions for the future state of the system in terms of such indicators as the area of hypoxia, or suitability of habitat for living resources, while most of the management and mitigation is carried out for the watershed, where the nutrient and solids loads are generated. Therefore the integration of the three models is crucial. But whenever the watershed model gets updated, it produces different output. As a result, every time the watershed model is changed, the estuary model needs to be re-calibrated. In theory, as the watershed model moves from one phase to the next one, its output is supposed to become ‘better’ in terms of matching the data, the model is supposed to be better calibrated. Sometimes this is indeed the case, in other cases the upgrades are driven by the need to include new features important for management, or by the need to change the resolution in the model. The output in this case becomes ‘different’ but not necessarily ‘better’ in terms of a better fit to data. Again and again the estuary model needs to undergo tedious re-calibration, which is entirely caused by developments that are external to it, and do not necessarily have anything to do with improvements of the estuary model itself. Much effort and time is spent with the only benefit of keeping the components working together.

Voinov and Cerco (2010) point out that the fact that two components are integrated into one model, does not have to make the available calibration data no longer relevant. When the downstream model, the estuary model in our example, was first calibrated, it was forced with observed loads, and there is no reason to think that the output from the up-stream model, the watershed model, is more accurate than these observations. If the data was still embedded in the integration process, then there would be no need for model re-calibration every time modifications are made ‘up the modeling stream’. Indeed, when new and ‘better’ modules are developed, the data are still invariable, and there is no reason why the same data set that was used for calibration of an up-stream model A cannot be used as forcing functions for the ‘down-stream’ model B.
When integrating models, the data can stand between the output of the upstream model and the input of the downstream model (Fig.6). When the first model is modified, the second model does not need to be re-calibrated, because still the first model is expected to represent the same data as before, and it is that data that feeds into the second model. In a way we employ a version of data assimilation, when data become part of the overall integrated modeling structure. Certainly, in many cases the output of the up-stream component is much more data-rich than what the observed data sets offer. For example, in case of the Bay model, the watershed model generates inputs for every single tributary, while data sets are available only for some. Yet still that is not a reason to exclude the available observations and replace them entirely by ‘artificial’ data. Wherever and whenever observed data are available they should be used and should have precedence over model generated data. What we suggest is a data-model fusion when integrating components.

But how do the two model components run in concert, if the data component sitting between them is invariant to the changes in the forcing functions that drive the upstream components? After all, most models are built to run scenarios, where certain parameters and forcing functions are modified to answer the ‘what if’ questions or make predictions for the future. How do we treat scenarios that produce results that are not available as part of the existing data sets?

The idea is to describe a scenario as relative to the calibrated base run, as a perturbation of this base run, which is then fed into the next component as an increment to the same data set that was used to calibrate that downstream component before (Voinov and Cerco, 2010). At any time step, the upstream component output is corrected based on the data set available, before it is fed into the next component as input. What is most important, we no longer need to undergo the tiresome re-calibration process, when changing the components ‘above’. Only the scenario runs will need to be rerun, since, obviously, a modified upstream component will likely produce different set of perturbations, which will change the performance of the components ‘below’.

Data, in fact, should be treated as an intrinsic part of any model, or even as models on their own. Data modules will exist along with model modules and when designing integrated systems we will choose which modules to use in which cases. This may simplify the calibration problem for cascade linking, when output from one module is fed further down the chain into the next module. Here data may serve as a constant ‘reality check’ between modules.

The situation becomes more complicated when we have circular linkage and modules exchange information in both directions. It is not clear where to place data in this case, and how the whole calibration exercise should be performed. I would argue that in such cases integration is becoming the less feasible the more empirical the modules are. CCA, DEV5 and other software linking technologies are more suitable for linking modules that are process-based and less dependent on data and calibration. For more empirical models it should be safer to build integrated models as a whole, using some of the integrated frameworks described at the beginning of this paper. After all with empirical models it is always hard to draw the line for the best level of complexity. A simpler model can perform much better than a complicated one, and many processes and parameters that come with pre-existing components may turn out to be unimportant.
5. CONCLUSIONS

A lot has changed during the last 20 years since Agren and Bosatta’s (1990) plea for more artistry and less construction-engineering or more science and less modelling in ecology. They were pointing out that it is not trivial to synthesize a larger system from pieces and then infer the system properties from the properties of the parts. With the modern module coupling tools it is getting only easier to put such larger systems together, but it is still questionable whether our products will be meaningful and useful.

While Silberstein (2006) insists that more data are needed and that models without data are not science, Andersen (2008) claims that there is already a deluge of data that can change the way science is done. Google and other similar efforts produce evidence that by analyzing huge arrays of data only, we can actually build new theories based only on correlation, ignoring causation. The new analytical tools that are going to be developed for petabyte computing in the “clouds” will require entirely different approaches to integration than the types of model integration that we were considering so far.

There are a lot of gains that can be made if instead of mechanistically plugging modules together we take into account the specific goals and features of the system, and approach the problem with some creativity. The considerations presented here by no means devalue the importance and need for module linking software tools, component interfaces and specifications. We are only reminding that models are more than just pieces of software and they require more to make them work together efficiently. For example, incorporating data sets in appropriate places within the integrated modeling system can help to keep components somewhat separate to avoid the propagation of perturbations and changes down the stream from one model to another. This can substantially increase the efficiency and accuracy of the integration process. We need data sets to be recognized as components on the same level as models. Such data components can then enter the integrated frameworks at various places, not only at the top, as input to drive the whole integrated model, and at the bottom, to compare with the output and to calibrate the model. Data components can be also used between components to test, adjust, and correct the data flows inside the integrated model. This will help to keep components more independent, and reduce the overall complexity of the calibration task for the whole integrated model.

ACKNOWLEDGEMENT

I have drawn inspiration and ideas from various personal communications and talks, in particular, by Hank Shugart, Marc Voltz, and Michael Schrage. My thanks are due to them.

REFERENCES

http://www.epa.gov/athens/research/projects/3mra/.
Ahuja, L.R., J.C. Ascough II and O. David, 2005. Developing natural resource models using the object modeling system: feasibility and challenges, Advances in Geosciences 4 pp. 29–36
http://www.wired.com/science/discoveries/magazine/16-07/pb_theory
A.Voinov. / 'Integronsters' and the special role of data


Cerco, C., 2000, Phytoplankton kinetics in the Chesapeake Bay model, Water Quality and Ecosystem Modeling 1 (1–4), pp. 5–49. Full Text via CrossRef


Linking Data, Models and Tools: An Overview

H.R.A. (Bert) Jagers

Abstract: Complex questions trigger researchers around the world to link data, numerical models and tools of different origins together for integrated modeling of the environment and related socio-economic fields. Different researchers have, however, chosen different ways to link such resources. As a result a seemingly wide range of interfaces and frameworks have been defined: some have used low level interfaces others more abstract and object oriented ones; some systems may require little or no code changes whereas others promote to fundamental rewriting of your code. So, if we all try to link data, models and tools, why are we then using such different environments? The basic idea is that the various approaches address supposedly conflicting demands like generality, flexibility, ease of use, accuracy and performance. What are the benefits of the various approaches? This paper addresses these questions by looking at a.o. common component architecture (CCA), earth system modeling framework (ESMF), FRAMES, object modeling system (OMS) and OpenMI. Are they really conflicting or do they to a large degree complement each other?

Keywords: integrated modeling, frameworks, interfaces, interoperability

1 INTRODUCTION

Environmental research integrates a large number of disciplines including atmospheric sciences, hydrology, geomorphology, geology, chemistry and ecology. A wide variety of models across these and other disciplines needs to be coupled to solve the challenges of today and tomorrow. Traditional (sub)mono-disciplinary numerical models have generally not been developed with interoperability in mind (especially not beyond the scope of the initial developers). The interaction of researchers, data, models and tools across (and to some degree even within) disciplines brought with it a range of ‘new’ challenges in the fields of information technology and semantics, such as: what data to exchange, how to exchange it and what does it actually mean? Many research communities around the world have tackled these questions in (relative) isolation such that different habits, approaches, conventions, interfaces and standards have emerged. As international co-operation increases (especially triggered by a limited number of researchers and common research models and tools) we reach a tipping point [Gladwell, 2000] and the need for interoperability among those communities becomes obvious. So, how different are those approaches?

In the following chapter I’ll describe a non-exhaustive list of component coupling technologies that seem to be competing and mutually exclusive since they are all aimed at improving the reusability through componentization, at reducing software development cost and at increasing the effectiveness of integrated research. However, it turns out not to be an either/or decision. To avoid confusion, I’ll use the following definitions which have been loosely based on the corresponding wikipedia descriptions.

Architecture is a general description of the structure of all generic parts of a framework.

Component is a software package or a module that encapsulates a set of related functions. Science components generally represent a coherent subset of the physical processes for the whole (or part of the) simulation domain.
Environment is a collection of central software services (infrastructure) used to initialize, start and finalize the components of a simulation. The environment may or may not play an important role in the communication between individual components.

Framework is a reusable implementation of a software architecture. It includes one or more of the following parts: a run-time environment, support libraries, components, their interfaces and conventions.

Interface is a formal, abstract definition of the functions/methods to be exposed/used by a component such that it can interact with the run-time environment and other components.

Implementation is a realization of an architecture or abstract component.

Coupling refers to (sequential or parallel) data transfer between components at run-time. This can happen in memory or via intermediate data files/repositories.

2 THE PLAYERS

This chapter gives a brief alphabetical overview of model coupling technologies developed by various groups. More information can be found in/on the referenced papers and websites.

2.1 CCA – Common Component Architecture

The CCA Forum\(^1\) was founded in 1998 to define a standard for a scientific, high-performance component architecture that includes HPC features not available in other generic component architectures such as CORBA, COM, .NET and JavaBeans. The CCA specifications were designed to (1) maintain the performance of components, (2) be non-exclusive with respect to inter-component communication mechanisms, (3) allow for parallelization across components, and (4) allow for configuration of components both prior and during execution.

In CCA the Scientific Interface Definition Language (SIDL) is used to describe the language specific component interface and to define ‘uses’ and ‘provides’ ports for the input and output arguments of the routines; these ports may represent scalars, arrays or functions. A CCA-compliant framework should provide (1) SIDL support to generate actual component interface wrappers, (2) services concerning communication, security, thread creation and management, memory management and error handling, (3) a configuration API to instantiate and couple components, and (4) a repository API to access component repositories. The Babel tools are the de facto standard for generating the wrapper ‘glue’ code to make routines written in Fortran, C, C++, Java or Python interoperable on a plug-and-play basis as described by Kumfert [2003].

The main CCA-compliant framework Ccaffeine, which has been developed for parallel computing, can be configured via a command line tool and via a graphical user interface. CCA has been shown to be interoperable with ESMF and MCT, and the CSDMS developers\(^2\) have successfully combined the Ccaffeine framework with the OpenMI 1.4 Java implementation. The CCA developments [Bernholdt et al., 2006] are currently led by TASCS\(^3\); this virtual organization is funded through the SciDAC (Scientific Discovery through Advanced Computing) program of the US Department of Energy.

2.2 CHyMP – Community Hydrology Modeling Platform

CHyMP\(^4\) is an initiative by the Consortium of Universities for the Advancement of Hydrological Science (CUAHSI) to develop, provide and support advanced simulation models to the academic community within a community-based ‘development-user-feedback’ framework. This development is still in its infancy, but its long term goal is to stimulate the development and uptake of a number of large community model components in the field of hydrology (contrary to the wide

---

\(^1\)CCA Forum website: http://www.cca-forum.org
\(^2\)Community Surface Dynamics Modeling System website: http://csdms.colorado.edu
\(^3\)Center for Technology for Advanced Scientific Component Software website: http://tascs-scidac.org
\(^4\)CHyMP website: http://www.cuahsi.org/chymp.html
variety of smaller, incompatible and unsupported models that exist today). This CHyMP initiative is closely related to the hydrology focus group of CSDMS.

This initiative should, however, not be confused with the similarly named development of the Community Hydrology Prediction System (CHPS) by the US National Weather Service (NWS) that uses Delft-FEWS\(^5\) for a nationwide early warning system. Such an operational system requires a robust approach and uses mostly one-way data streams with file-based data exchange.

### 2.3 ESMF – Earth System Modeling Framework

ESMF\(^6\) is a high-performance framework aimed at improving the software interoperability and reuse in climate, numerical prediction, data assimilation, and other Earth science applications. ESMF provides both a coupling superstructure and a utility infrastructure; it supports both MPI and openMP for parallelization. The component code sits between these two layers, making calls to the infrastructure libraries beneath it and being scheduled and synchronized by the superstructure above it. The utility layer includes type definitions, time, clock and alarm functions, parallel data communication and regridding routines, message logging tools, etcetera.

ESMF distinguishes between gridded components (physics and dynamics) and coupler components (interpolation and mapping). All components must define initialize, run and finalize methods: the components’ input and output arguments are bundled together in inputState and outputState data structures. Components may optionally use an internal state. These states can store (bundles of) Arrays, (bundles of) Fields, and other States. An Array is a distributed, multi-dimensional array that can carry information such as its type, kind, rank, and associated halo widths. To adopt ESMF, components may wrap their existing Fortran/C arrays into ESMF Array structures. A Field represents a physical scalar or vector field: it contains a data Array along with grid information and metadata. Gridded and coupler components may be nested in gridded components that can be coupled to other components at a higher hierarchical level. Generally, the drive module and all components are linked together into one executable.

ESMF is supported on UNIX, Linux and Windows HPC platforms. ESMF developments include (a) integration with workflow management and visualization services to create modeling environments, (b) automatic generation of couplers, executables, and metadata, (c) web services, and (d) support for a wider variety of grids and numerical methods. Saint and Murphy [2010] use web services for a one-way link from ESMF to OpenMI. The ESMF project is sponsored by the US Department of Defense (DoD), NASA, NSF and NOAA; as of November 2009, it is part of NOAA Environmental Software Infrastructure and Interoperability (NESII) group.

### 2.4 FRAMES – Framework for Risk Analysis of Multi-Media Environmental Systems

FRAMES\(^7\) is an operational modeling environment in use by the US Environmental Protection Agency (EPA) within which collections of models and modeling tools (e.g., data retrieval and analysis) can communicate with each other. FRAMES applies predefined connection schemes and dictionaries to guarantee correct coupling of the components by end users. 3MRA\(^8\) is an actual set of 17 modules placed within FRAMES that collectively simulate the release, fate & transport, exposure, and risk (human and ecological) associated with wastestream contaminants deposited in various land-based waste management units (e.g., landfills, waste piles). Because of the many processes and parameters involved, model results are based on ten thousands individual simulations\(^9\). To keep the total simulation time within limits, the components of 3MRA are based on highly simplified formulations for each domain which is quite the opposite of the climate

---

\(^5\)Delft-FEWS website: [http://public.deltares.nl/display/FEWSDOC](http://public.deltares.nl/display/FEWSDOC)

\(^6\)ESMF website: [http://www.earthsystemmodeling.org](http://www.earthsystemmodeling.org)

\(^7\)FRAMES 1 and 2 information: [http://mepas.pnl.gov/FRAMESV1 and /FRAMESV2](http://mepas.pnl.gov/FRAMESV1 and /FRAMESV2)

\(^8\)Multi-media, Multi-pathway, Multi-receptor Risk Analysis website: [http://www.epa.gov/athens/research/projects/3mra](http://www.epa.gov/athens/research/projects/3mra)

model components coupled by frameworks like ESMF. However, the demand for more complete representations is increasing.

For FRAMES v3 the current one-way file-based communication method will be (optionally) replaced by a faster in-memory two-way communication method. This is needed for more complex model components (resulting in more data exchange) as well as for more complex component interaction. This new coupling method will be based on OpenMI.

2.5 HLA – High Level Architecture

The High Level Architecture (HLA) was developed by the Defense Modeling and Simulation Office (DMSO) of the US DoD as a general purpose architecture for distributed real-time training/simulation environments. Typically, this concerns tightly coupled networks in which data exchanges are frequent, but usually small. The HLA baseline definition was completed on 1996; in 2000 it was accepted as general IEEE 1516 standard.10 It defines the general architecture (federation rules), the interfaces of components (federates) and the environment (RTI: run-time infrastructure) and an object model template (OMT) that provides a common method for recording information and that defines a.o. a federation object model (FOM: data exchanges during a simulation) and a simulation object model (SOM: description of component/federate and its possible exchanges) that can be queried at run-time. Data exchange occurs via the RTI.

HLA is not an implementation, it just provides an architectural sketch. The existing HLA RTI implementations (by a.o. Raytheon, MÁK, Pitch, and http://sourceforge.net/projects/ohla/) are not 100% compatible because the IEEE standard contains some errors and doesn’t fully prescribe the interface implementation. A dynamic link compatible (DLC) API has been defined (SISO-STD-004.1-2004) to make the implementations more consistent. All compatibility issues should have been resolved by the new HLA Evolved IEEE 1516-2010 standard.

2.6 Kepler

Kepler11 offers a general purpose (nested) workflow framework that supports continuous time, discrete event, and dynamic or parallel data flow concepts for application across a broad range of scientific and engineering disciplines. It’s a Java-based application that allows the user to visually assemble workflows by means of directed sequence graphs of components (‘actors’). Kepler comes with a default library of some 350 actors for a wide range of tasks including numerical integration, image processing, accessing webservices, reading and writing standard file formats, plotting, and running external command line applications. Development of the open source Kepler software is led by a team from the universities of Davis, Santa Barbara, and San Diego. Taverna12 is a similar Java-based workflow system developed in the context of the myGrid project led by Carole Goble of the University of Manchester, now part of OMII-UK.

2.7 MCT – Model Coupling Toolkit

MCT13 is an MPI-based library of Fortran90 modules which can be used to create parallel integrated grid-based models (both structured and unstructured). Version 2 of the toolkit was developed for and used in building the cpl6 coupler [Craig et al., 2005] of Community Climate System Model CCSM3 as described by Larson et al. [2005] and Jacob et al. [2005]. It uses an AttributeVector type to store the local data to be exchanged in a 2D parameter-location array; it uses a GlobalSegmentMap type to describe the global partitioning of a numerical grid across multiple processes. Based on these data types, MCT supports efficient parallel MxN intercomponent data transfer and MxM intracomponent data redistribution, intergrid interpolation using matrix-vector

11 Kepler website: https://kepler-project.org
12 Taverna website: http://www.taverna.org.uk
13 MCT website: http://www.mcs.anl.gov/mct
multiplication, spatial integration and time averaging. MCT can be used in single or multiple executable systems and allows sequential or concurrent execution. Much of what was learned and implemented in MCT and cpl6 has been included in ESMF which will be used for CCSM4.

2.8 OASIS and PALM

Earth system modelers organized in ENES\(^{14}\) initiated the Partnership for Research Infrastructures in earth System Modelling (PRISM) project to develop a common software infrastructure. Building on their earlier work, CERFACS\(^{15}\) led the development of the open source OASIS3 and OASIS4 frameworks\(^{16}\) [Valcke and Morel, 2006]. The OASIS frameworks consist of a driver, transformer and a MPI-based PRISM System Model Interface Library (PSMIle). The individual OASIS3 component executables include initialization, variable definition, get and put, and finalization calls to the statically linked PSMIle. The driver component initializes and connects the components at run time based on configuration files\(^{17}\). It also calls the data transformer to repartition and/or regrid the data; data that doesn’t require either of these actions will be passed directly from the providing to the using component. Although OASIS3 components may be multithreaded, the driver and transformer are single threaded (they are parallel in OASIS4). OASIS3 (transformer) requires grid coordinate data to be specified using netCDF files and it supports scalar 2D grid data only. The OASIS4 PSMIle interface includes support for vector quantities and grid definition (1D, 2D and 3D); hence it no longer requires data files for grid information.

Concurrent with these developments, CERFACS developed for the MERCATOR project the closely related proprietary PALM\(^{18}\) framework for oceanographic data assimilation applications [Valcke and Morel, 2006]. Contrary to the OASIS couplers, PALM supports the dynamic addition and removal of components during the execution by means of MPI2 features. However, it currently lacks the parallel interpolation features of OASIS4.

2.9 OMS – Object Modeling System

The OMS\(^{19}\) is a domain-specific, reusable framework with a set of interdependent Java classes developed by the US Department of Agriculture (USDA) in collaboration with other agencies and organizations involved with agro-environmental modeling. OMS provides an integrated programming, simulation and analysis environment. Individual components are plain Java classes with an execution method and optional initialization and finalize methods. Component methods, input and output variables as well as are identified and described by means of Java annotations such as @In, @Out, @Unit, and @Execute. Both time and spatial loops are moved out of the components such that most input and output arguments are scalars; individual components are relatively simple. OMS and specific model developers together have migrated (or are in the process of migrating) various watershed models, such as the Soil and Water Assessment Tool (SWAT), J2000 [Krause, 2002], and Precipitation Runoff Modeling System (PRMS) into OMS modules.

2.10 OpenMI – Open Modeling Interface

Version 1.4 of the OpenMI standard\(^{20}\) was developed within the EU-funded HarmonIT project by researchers from a.o. Delft Hydraulics (now part of Deltas), DHI and Wallingford Software (now part of MWH Soft). This interface standard enables end users to couple components created by different developers without recompilation. It requires simulation engines to be imple-

\(^{14}\)European Network for Earth System Modelling website: http://www.enes.org

\(^{15}\)Centre Européen de Recherche et Formation Avancées en Calcul Scientifique: http://www.cerfacs.fr

\(^{16}\)OASIS stands for: Ocean Atmosphere Sea Ice Soil

\(^{17}\)Each component has a PMIOD (potential model input and output description) and SMIOC (specific model input and output configuration). The SMIOC files combined with the SCC (specific coupling config) of the simulation determine the run time exchanges.

\(^{18}\)Projet d’Assimilation par Logiciel Multi méthode: http://www.cerfacs.fr/globc/PALM_WEB

\(^{19}\)OMS development website: http://honeycomb.javaforge.com/project/oms

\(^{20}\)OpenMI website: http://www.openmi.org
mented as ‘linkable components’ that provide methods to (1) initialize the component, (2) query the component for (providing and accepting) ‘exchange items’, (3) define links, (4) get values from the component, and (5) finalize the component. An exchange item is a quantity defined on an ‘element set’ which is a location specification (set of either labels or coordinates). Providing components are responsible for the interpolation of data to the element set of the requesting component. This first version of the standard is strictly based on a pull-based approach as it doesn’t include a method for setting values. After initializing all components and their links, a simulation starts by asking the ‘final’ component in the workflow for data. Subsequently, that component will start the necessary computation and, as needed, it will request values from the linked components without intervention of a central framework. Deadlocks in cyclic workflows are prevented by requiring that a component must return ‘best guess’ values without calling other components if it’s already waiting for data due to a previous GetValues() call. A first reference implementation of the standard was created using C#\textsuperscript{21}. A Java implementation was later provided by Alterra; they combined OpenMI with formal ontologies in the SEAMLESS\textsuperscript{22} Integrated Framework. Although both reference implementations use only a single execution thread, this not strictly required by the OpenMI standard. OpenMI has been shown to be compatible with remote and multithreaded engines, and webservises.

In 2007 the OpenMI Association was founded to take formal ownership of the standard and the associated reference implementation(s). The upcoming version 2.0 of the standard (see Gijsbers et al. [2010] and Donchyts et al. [2010] for details) adds a SetValues() method and separates time progress from the GetValues() method; these changes will make OpenMI easier to use when coupling with (geospatial) databases and data assimilation. In this context it is worthwhile to mention the Dutch OpenDA\textsuperscript{23} initiative to develop a flexible open source calibration and data assimilation framework based on a component interface that is consistent with OpenMI.

2.11 TIME – The Invisible Modelling Environment

TIME\textsuperscript{24} supports developers in creating, testing and delivering environmental simulation models; it has been developed by CSIRO [Rahman et al., 2005]. It is a .NET framework that includes object libraries for standardized data IO, GIS operations, data visualisation, uncertainty assessment and non-linear optimisation. A GUI specific for the model is automatically generated based on metadata tags to component variables. E2/WaterCAST [Cook et al., 2009] extends the TIME framework with functionality to link sub-catchment hydrological and constituent models along streams down to the tidal limits in estuaries.

3 COMPARING DEVELOPMENTS AND CONCLUSIONS

FRAMES and CHyMP differ from the rest by their focus on the end-user and collaborative side. The same holds for other (local and web-based) modeling environments like CSDMS, DeltaShell [Donchyts and Jagers, 2010], OpenWEB\textsuperscript{25} and iemHUB\textsuperscript{26}. Instead of developing another coupling technology, these initiatives tend to adopt (and adapt) such technologies; hence, I’ll ignore them in this comparison. Furthermore, it should be noted that CCA, HLA and OpenMI in essence only define architectures and interfaces, whereas the other initiatives have the actual implementation as objective. The standard body for HLA doesn’t even provide a reference implementation, whereas the CCA and OpenMI developers are at least creating one.

The interfaces defined by CCA, ESMF, OASIS, OMS, OpenMI and TIME are similar in the sense that they all use initialize, run, finalize, get and set concepts. However, a comparison by Lloyd et al. [2009] showed that the amount of code needed varies significantly: OMS 3.0 requires the

\textsuperscript{21}Runs on Linux via the open source mono environment: \url{http://www.mono-project.com}.
\textsuperscript{22}System for Environmental and Agricultural Modelling: Linking European Science and Society website: \url{http://www.seamlessassociation.org}
\textsuperscript{23}OpenDA association website: \url{http://www.opendata.org}
\textsuperscript{24}TIME website: \url{http://www.toolkit.net.au/Tools/TIME}
\textsuperscript{25}https://openweb.uk.net uses the OpenMI-based Pipistrelle C# environment by HR Wallingford
\textsuperscript{26}iemHUB website: \url{http://iemhub.org}
Table 1. Comparison of coupling technologies

<table>
<thead>
<tr>
<th>topic</th>
<th>CCA</th>
<th>ESMF</th>
<th>HLA</th>
<th>Kepler</th>
<th>MCT</th>
<th>OASIS</th>
<th>OMS</th>
<th>OpenMI</th>
<th>TIME</th>
</tr>
</thead>
<tbody>
<tr>
<td>defines framework</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>defines interfaces</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>provides (reference) implementation</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>defines object model</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>code invasiveness [Lloyd et al., 2009]</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>plug &amp; play (and graphical coupling)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>support for HPC environment</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>C/FORTRAN support</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Java support</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>.NET support</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

least due to its plain Java approach and the strategic use of annotations, but CCA generates a lot of language interoperability code for each component. CCA, HLA and Kepler don’t impose any particular (spatial) object model. ESMF, MCT and OASIS use numerical grids as spatial data representations, whereas OpenMI takes an approach consistent with OGC27 conventions. OMS removes in general the spatial loop from the model components, such that the component variables are scalars and no complex objects are needed. The same holds for TIME, but it supports raster and network data types. OMS and TIME object types may be extended using plain Java or C#.

CCA, HLA, Kepler, OpenMI and TIME allow the end user to couple pre-compiled components from different developers at run-time and (except for HLA) provide a graphical environment for that. ESMF, MCT and OASIS focus on the high-performance user community with primary programming languages Fortran and C. Kepler, OMS, OpenMI and TIME support the Fortran and C users only indirectly through wrappers. Only OpenMI and TIME provide support for the .NET platform. Which languages HLA supports depends on the implementation (generally Java or C). CCA is the only architecture that really addresses the issue of language interoperability through the use of SIDL and Babel.

There is a trend to guide users in coupling components. Ontologies and other metadata conventions are important to guarantee the validity of links and for the automatic discovery of possible links. Frameworks with an operational focus, like FRAMES, already apply such conventions to some degree. In climate research, US Earth System Curator28 and European METERFOR29 projects work together with ESMF and OASIS (component) developers to adopt Climate & Forecast conventions.30

Although the technologies described here were all developed with a common vision of a componentized architecture, they have addressed different parts of the integrated modeling challenge (generality, flexibility, ease of use, accuracy and/or performance). One might be able to use the similarities in their interfaces to implement a generic wrapper generator (see e.g. the Bespoke Framework Generator by Armstrong et al. [2009]). This would improve the portability of components among compatible frameworks.

REFERENCES


27Open Geospatial Consortium website: http://www.opengeospatial.org
28Earth System Curator website: http://www.earthsystemcurator.org
29Common Metadata for Climate Modelling Digital Repositories: http://metaforclimate.eu
30CF metadata conventions website: http://cf-pcmdi.llnl.gov


Methods to Register Models and Input/Output Parameters for Integrated Modeling

James G. Droppo, Gene Whelan, Michael E. Tryby, Mitchell A. Pelton, Randel Y. Taira, and Kevin E. Dorow

“Pacific Northwest National Laboratory, Richland, Washington 99354 (James.Droppo@pnl.gov); U.S. Environmental Protection Agency, Athens, Georgia 30606 (Whelan.Gene@epa.gov); U.S. Environmental Protection Agency, Athens, Georgia 30606 (Tryby.Michael@epa.gov); Pacific Northwest National Laboratory, Richland, Washington 99354 (Mitch.Pelton@pnl.gov); Pacific Northwest National Laboratory, Richland, Washington 99354 (Randel.Taira@pnl.gov); Pacific Northwest National Laboratory, Richland, Washington 99354 (Kevin.Dorow@pnl.gov)

Abstract: Significant resources can be required when constructing integrated modeling systems. In a typical application, components (e.g., models and databases) created by different developers are assimilated, requiring the framework’s functionality to bridge the gap between the user’s knowledge of the components being linked. The framework, therefore, needs the capability to assimilate a wide range of model-specific input/output requirements as well as their associated assumptions and constraints. The process of assimilating such disparate components into an integrated modeling framework varies in complexity and difficulty. Several factors influence the relative ease of assimilating components, including, but not limited to, familiarity with the components being assimilated, familiarity with the framework and its tools that support the assimilation process, level of documentation associated with the components and the framework, and design structure of the components and framework. This initial effort reviews different approaches for assimilating models and their model-specific input/output requirements by considering four case study examples: 1) modifying component models to directly communicate with the framework (i.e., through an Application Programming Interface), 2) developing model-specific external wrappers such that no component model modifications are required, 3) using parsing tools to visually map pre-existing input/output files, and 4) describing and linking models as dynamic link libraries. Most of these examples are derived from assimilation efforts for the widely distributed modeling system called Framework for Risk Analysis in Multimedia Environmental Systems (FRAMES). The review concludes that each has its strengths and weaknesses, the factors that determine which approaches work best in a given application.

Keywords: Integrated Modeling; Legacy Model; Data Wrapper; Model Linkage; Data Transfer; Database; Multimedia Modeling

1. INTRODUCTION

Integrated modeling systems combine model and database components into a single modeling framework. Often both new and old components are included. Both time and cost can be significantly reduced with well-vetted “off-the-shelf” models. A framework designer faces the challenge of including capabilities that will meet the assimilation needs for the components to be used in the framework. The trade-off is between investing more in the framework model assimilation capabilities and requiring more resources when constructing an integrated modeling system within a framework. This paper addresses assimilation approaches used for legacy components and new externally developed components.
It is potentially costly and labor-intensive to construct an integrated modeling system within a framework. A typical development effort requires assimilating dissimilar components (e.g., models and databases) created by different developers, often using a different software language or database system. The framework’s functionality needs to allow the framework’s user to assimilate components often based on a limited understanding of the details of a component’s characteristics. The framework needs to provide the framework’s user with the ability to assimilate a wide range of component-specific input/output requirements in a manner that accounts for their associated assumptions and constraints. The challenge of assimilating disparate components into an integrated modeling framework varies in complexity and difficulty. A key factor that can be a showstopper is the compatibility of the spatial and temporal design of the framework with the external component. Other major factors can be the user’s familiarity with the components and framework along with the level of component documentation available for components. The capabilities of the framework component assimilation tools need to match the input/output functionalities of the components being assimilated.

This paper addresses some processes that can be used for model assimilation with a framework. The discussion is based on the experience of the authors in model assimilation efforts for integrated modeling systems. Most of these approaches are illustrated using the widely distributed modeling system called Framework for Risk Analysis in Multimedia Environmental Systems (FRAMES). In FRAMES, logical groups of parameters and metadata are formally defined as DIctionary (DIC) files. These files, which are used by the FRAMES application programmers interface (API) to store and obtain data, have a highly organized, structured design [Gelston et al. 2004].

2. MODEL ASSIMILATION APPROACHES

A major challenge in the development of a functional integrated modeling framework is being able to design an effective system for incorporating models. There is, of course, no one best approach. Instead, there are a variety of approaches that have various strengths and weaknesses. Many different ways are used for registering models and parameters with frameworks. Major considerations for designing a system-specific model incorporation system are related to the origin of models, the approach for linking models, the desired flexibility for model implementations, and limitations related to obtaining the desired level of performance for the integrated modeling system.

A basic tenet of integrated modeling systems is that computations are performed that require the models to transfer data. An output from one model is an input to another model. This data exchange may be accomplished by using files, static databases, or dynamic databases. Whatever data-exchange methods are used, the model must be registered to implement those exchanges of model input and output data. The data that define the connectivity of the models are referred to as model boundary conditions. When integrating (or connecting) two models together, the challenge is in accurately passing output data from the first model such that it can be used as input data for the second model. Even in cases where models are compatible in terms of data being transferred (i.e., parameters and units match), it is extremely rare that the first model’s output file format identically matches the input file format of the second model.

3. MODEL REGISTRATION EXAMPLES

Integrated modeling frameworks provide different levels of support for implementing models. The input and output parameters must be defined and registered in the framework before a model can be implemented. As described in Whelan et al. (2010), FRAMES uses a formal parameter definition procedure that automatically matches parameter properties and handles unit conversions. Many of the examples presented below use FRAMES development tools (i.e., sets of subroutines and functions) to communicate with the
FRAME API. These tools support data transfers, obtaining data properties and other functions in a number of computer languages, and versions of computer languages. These examples also use the FRAMES generic data entry/viewer capability (DCE editor) (Whelan et al. 2010). Models can be registered within frameworks by modifying the code to work within the framework or by creating a model wrapper outside the framework. The first assimilation example modifies the source code to work within the framework. The second two examples create model wrappers either by writing code or using a data-parsing wrapper wizard. The fourth example creates model wrappers to conform to a dynamic link library (DLL) standard.

3.1 Example 1 – Modifying Source Code

This example modifies the model source code to input and output data from a model. These modifications are run as an integral part of the model.

Background

As noted above, when integrating (or connecting) two models together, the challenge is in accurately passing output data from the first model such that it can be used as input data for the second model. A schematic of the implementation and linkage of legacy models in FRAMES is shown in Figure 1. With this approach, a model is modified such that the model can 1) read input data from various data sources and 2) produce output data in a form that can be consumed by downstream models. Note that the downstream module output data source becomes the upstream module input data source for the next linked model.

Application

The U.S. Nuclear Regulatory Commission (NRC) model for codifying potential routine air effluent rates for an operating boiling water reactor (BWR) is referred to as GALE BWR-GE. As part of an effort to update this model whose original development was in the 1970s, a revised FORTRAN source code was developed (Droppo and Pelton 2010). The revised GALE BWR-GE model source code now includes calls to the FRAMES API for the model’s 1) input and output data exchanges and 2) initialization of model parameters. This model implementation in FRAMES was conducted after the development of wrappers for unmodified legacy versions of this model (described below). The process was very straightforward: 1) the read statements for inputs were replaced with API calls to obtain data from dictionary-based input data files and 2) code was added to write the desired output parameters through the FRAMES API to dictionary-based output files. Issues were encountered related to the transfer of data between the model’s FORTRAN code and the FRAMES C codes. Typically, additional code was needed to address these issues. Overall, we found the time to implement the model with this approach was much less than
for the wrapper-development approach. We also found the linkages to be cleaner and faster using the model modification approach.

Summary

Modifying the model source code provides a direct link between model and framework. Using this approach requires a good understanding of the selected model. Quality assurance/quality control (QA/QC) testing should be performed on the modified model to verify that the modifications did not introduce errors. The model modification approach has the potential of often being the least labor-intensive model assimilation approach.

3.2 Example 2 – Writing Code to Create Model Wrappers

This approach uses “wrapper” computer programs to feed input data to and obtain output data from a model. These wrapper programs run separately from the model—allowing the model to be used in an essentially unmodified form. In many situations, particularly for legacy models, it is highly desirable to use models in their unmodified state to verify the maintenance of the inherent functionality of a model.

Background

Registering a model with wrappers requires creating wrapper programs and handling the logistics of data transfers and model execution. Typically, pre-run wrapper programs create the input files and/or databases, and a post-run wrapper program imports the model output data from a model output file. When such model wrappers are used on an unmodified model, then the QA/QC testing needs be performed only on the wrappers. Two distinct methods were used to create the model wrappers: 1) write custom source code for each model wrapper and 2) use generic model wrapper programs to automate the model-wrapping process. These tasks are easiest for models with “well-behaved” input and output files—in which locations of needed data can easily be uniquely defined. Writing model-specific wrappers using the FRAMES API calls provides a high level of flexibility in creating the model wrappers. However, this approach can be the most labor intensive of the various wrapper approaches. The user must address the details of model data transfers plus language/compiler programming data-transfer issues. Using generic software wrappers can be an effective approach for implementing models. A large part of the model registration process can be quickly performed. Programming-related data-transfer issues can be resolved one time in generic wrapper codes. The use of a generic wrapper program in a new application is limited by its inherent capabilities.

Application

Model wrappers for NRC’s reactor emissions codes, the GALE codes, were developed by using a combination of custom and generic model wrapper programs (Droppo and Pelton 2010). A model run involves a series of operations (Figure 2). Step 1 starts with selecting a list of radionuclides to be addressed. Step 2 is to select which GALE code is to be run. In step 3, the user enters the code-specific input data using the FRAMES generic data editor (DCE). These input data are stored in a FRAMES DIC database file. Next, in steps 4 to 7, the wrapper programs and the model are run using batch files. Each of the GALE codes reads a wrapper-created text file for input and produces a text file read by a wrapper to get model results.

Generic wrapper programs were written for mapping the data input files (Pmod) in step 4 and results in the output files (Rmod) in step 7. These wrapper programs exchange data between FRAMES DIC input/output files and flat GALE input/output files. These generic wrapper program codes also were used to implement two other models (NRC GASPAR
and LADTAP-II models). A custom wrapper program was required for matching GALE and FRAMES names for radionuclides (Cmod) in step 6. Step 8 occurs as part of the data input process for the downstream model. In step 9, custom data wrappers were needed for creating the FRAMES DIC database files required by “downstream” air and water models (Dmod). An alternative approach would be to have the transport-pathway specific facility data be part of the facility inputs, so the required downstream FRAMES DIC database files could have been created in Step 7.

Figure 2. Wrapper-based “MODEL” Implementation of the GALE Codes

Summary

Creating model wrappers worked well for assimilating the GALE codes. Using generic wrapper programs for “well-behaved” data exchanges greatly simplified those efforts. However, the creation of custom wrapper codes for model-specific requirements was quite labor-intensive. The assimilation of unmodified GALE codes greatly reduced QA/QC testing requirements compared to the model-modification approach.

3.3 Example 3 – Assimilation of a Model Using a Data-Parsing Wrapper

This approach emulates the model wrapper approach discussed above using an interactive framework-based development environment for 1) creating model “wrapper” functionalities needed for input/output model data exchanges and 2) handling the logistics of running the model.

Background

Using an interactive data parsing approach can greatly reduce the resources needed to register a model using model wrappers. Assuming that the pertinent model parameters are defined in dictionaries, Dorow et al. (2007) describe parsing techniques used to map the model parameters. The concept is to have the user defining the data-mapping specifications through visual inspections, using a graphical user interface (GUI), of the data files. In practice, the model’s files must be in (or converted to) a readable text format.

Application

A data-parsing wizard was developed for a FRAMES application (Dorow et al. 2007) for mapping model output parameters. “Text File Tables” are used to define tables of data
within the output file, “Text Spans” are used to define areas of discrete values, and “Transforms” are used to parse and concatenate data formats such as dates, as necessary.

To register a discrete value, the user defines the exact location of the value by row and column or by some unique identifier that precedes the value (e.g., unique descriptor). Once mapped, any value that shows up in that location is registered with the system and passed along to downstream models requiring that DIC. Tables are more complicated—only a portion of the table might be needed. One must uniquely identify the exact data locations in tables (rows and columns) accounting for shifts in data location that may occur. To address this issue, mapping may reference absolute row numbers or indicate relative row location using unique static information (Dorow et al. 2007).

**Summary**

The data parsing approach can be used effectively in applications where the location and formats of the data in the mapped files can be uniquely identified. For this approach to work, a successful base run file needs to exist as a template for mapping the data values. The advantage is that once this base run case is registered, new simulations based on this case can be run. The limitation is that a re-mapping and re-naming of the model often will be required to address different runtime options. A major challenge with the data parsing approach is providing the functionality of addressing the many permutations for the content, format, and location of data in model input/output files.

3.4 **Example 4 – Model Wrappers to Implement a Linkage Standard**

**Background**

Unlike the linkage cases considered above, the final approach discussed here does not use a framework to manage data exchange; rather, linkage and data exchange occurs directly between model components communicating via a standard.

**Application**

OpenMI is a software standard that facilitates the linkage of individual models into integrated modeling systems (OpenMI 2010). It defines data structures and protocols for data exchange and has facilities for handling the spatial and temporal mismatch between model domains. The standard is open source and computing-platform independent. OpenMI places the responsibility for implementing its runtime capabilities on the component developer. Thus, the model integration process requires intermediate to advanced software development skills, development of new software, and in some cases, significant revisions of existing computational cores. OpenMI promotes a “Wrapper” pattern for model integration not unlike the model integration strategy described above.

Converting existing computational cores into OpenMI linkable components requires several well-defined steps. The first step, however, is the most difficult. The existing computational core must be converted from an executable to a DLL and must expose entry points through the following APIs: 1) initializing, running, time-step control, finishing, and cleaning-up, 2) setting initial conditions, and 3) accessing the computational core data model (e.g., setting and getting values for input and output exchange variables). Once the DLL is created, the process of conversion to an OpenMI Linkable component is straightforward.

**Summary**
OpenMI is an emerging standard; the future implications of which are uncertain. Many organizations have made significant investments developing integrated modeling frameworks that work well for their needs but are non-compliant with this standard. Short of outright adoption of the standard, OpenMI may facilitate integrated modeling by 1) providing a means of creating OpenMI linkable components that can be embedded in existing and new frameworks, 2) provide a common API for linking with integrated modeling tools and making them interchangeable between frameworks, and 3) defining an API for linking disparate integrated frameworks together.

4.0 CONCLUSIONS

In summary, Table 1 lists the general model implementation functionalities that need to be provided. The approaches discussed above illustrate ways of meeting the listed functionalities. In all methods, some level of software programming, compiling, testing, and documentation is normally necessary. The “best” assimilation approach for a given model may be some combination or hybridization of these approaches.

<table>
<thead>
<tr>
<th>Functionality</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Definition of Model-Specific Input Data</td>
<td>In addition to data from an “upstream” model, a model often requires that run-time parameters be defined. Implementing systems for handling these model-specific data represents a special challenge for generic model registration systems.</td>
</tr>
<tr>
<td>2. Framework Support of Model Import of “Dark Data”</td>
<td>A model often requires definition of run-time input parameters that do not need to be assessed by the modeling framework. These data are referred to as “dark data” because the framework does not provide a means of accessing or modifying these data.</td>
</tr>
<tr>
<td>3. Model Access to Framework Databases</td>
<td>It is often necessary to access global framework databases when implementing a model. For example, frameworks often have GIS and constituent property databases.</td>
</tr>
<tr>
<td>4. Match Linked Data Parameter Properties</td>
<td>The exchange of a parameter value between models needs to match the data properties (units, time, spatial and temporal average, etc.) from an “upstream model” with the properties of data expected by the “downstream model.”</td>
</tr>
<tr>
<td>5. Logistics Support for Running the Model</td>
<td>Registering a model requires that the logistics for running the model are in place. In addition to the applicable executable and batch files, the model implementation must include defining the paths and files for the various model data transfers between files and databases as well defining the model run status.</td>
</tr>
<tr>
<td>6. Source Code-independent Communication and Data Transfers</td>
<td>Model and framework software are often written in different computer languages. The model registration process must allow for variations in data structures, formatting conventions, and other differences in language procedures.</td>
</tr>
<tr>
<td>7. Means of Accessing Model Results</td>
<td>The model registration needs to include definition of the means that the framework will use to access the results generated by a model.</td>
</tr>
</tbody>
</table>

Our experience with FRAMES is that wrapper programs can easily be built to handle “well defined and well behaved” input and output file structures—but the task of developing wrapper programs to cover all possible file structures is prohibitively complex. The data-parsing wizard worked well in its original application. However, it has not found wide use yet in other applications, as was expected. The main impediment has been the need for additional file-mapping capabilities. Although this approach should require no new coding, we have found that in practice, some coding may be required to address model-specific mapping issues such as non-standard constituent naming conventions and non-unique data-mapping locations in model files.
The OpenMI standard is not seen as perfect and will not eliminate all the technical difficulties associated with integrated modeling. The foundation of a standards-based organization for integrated modeling is, however, an important and significant step forward for the integrated modeling community. Each approach considered works well in certain situations and not so well in other situations. Modifying the model source code to exchange data through the framework API is normally the easiest method for assimilating legacy models. Writing model wrappers works well for models whose source codes need to be used in an unmodified form. Creating a model wrapper wizard works well to perform specific functionalities required by a specific application. Creating specifications for implementing a model as a DLL has the potential to provide the best linkage performance.

The views expressed in these Proceedings are those of the individual authors and do not necessarily reflect the views and policies of the United States Environmental Protection Agency. Scientists in EPA have prepared the EPA sections, and those sections have been reviewed in accordance with EPA’s peer and administrative review policies and approved for presentation and publication.

5.0 REFERENCES


Modeling freshwater uses in coastal areas – the case of Pertuis Charentais (France)

Cédric Bachera, Johanna Balle-Begantona, Julien Neveuc, Harold Rethoretd, Françoise Vernierc, Paul Bordenavef, Rémi Mongruelg, José Perezh

aIfremer (cedric.bacher@ifremer.fr)
bIfremer (johanna.beganton@ifremer.fr)
cEAUCEA (julien.neveu@eaucea.fr)
dEPTB Charente (harold.rethoret@fleuve-charente.net)
eCEMAGREF (francoise.vernier@bordeaux.cemagref.fr)
fCEMAGREF (paul.bordenave@bordeaux.cemagref.fr)
gIfremer (remi.mongruel@ifremer.fr)
hIfremer (jose.perez@ifremer.fr)

Abstract: The SPICOSA European Project aims at implementing a system approach framework as a tool bridging ecosystem management, definition of policy options and system dynamics. An ExtendSim numerical modeling platform is being developed to share and integrate economic, ecological and governance modules, and to assess coastal systems functioning and changes. The implementation of a hydrological model is presented as an example of generic module. It is tested and applied on the Charente catchment study site in France where the freshwater share is the core issue of coastal zone management strategy in relation to several ecological services – e.g. supply of drinkable water, support of aquatic biodiversity, carrying capacity of coastal areas for shellfish farming, water demand for agriculture. The definition of hydrological units was based on sub-basins features, management rules and agriculture demand for freshwater. A typology of agricultural activities was built to define relevant spatial units based on the type of physical environment, technical parameters, types of crops and uptake of water. Irrigation rules and decision rules regarding Crisis Minimum Flow and Target Minimum Flow are included in the model. The model will be used to simulate freshwater shortage using climatology observed during the past 20 years and scenarios of agriculture changes. Several indicators have been constructed to assess the vulnerability of the catchment area and the coastal zones to freshwater shortage.

Keywords: hydrology, coastal zone management, system approach, ExtendSim

1. INTRODUCTION

In many aspects, sustainable development requires the integration of information on economic, environmental and social factors (Kelly, 1998). The structure of the information is therefore a key issue and several authors have emphasized the need for a system approach to select and combine the appropriate information to be used in the decision making process. Kelly (1998) identifies four main reasons to use System Approach: i) understanding the system dynamics due to the interactions between several components, ii) identification of knowledge gaps regarding some relationships, iii) learning about the system which results in changes of system perception by the decision makers, iv) support of an interdisciplinary approach by using a common language. In the frame of such a system approach, pioneer works on system dynamics start with Forrester (1971), Meadows et al. (2004) and were followed by many applied examples addressing sustainable issues (see
Kelly, 1998) and considering the dynamics of systems derived from interactions and feedback loops. On this basis, the SPICOSA project has been funded by the European Commission in 2007 with the aim of improving the sustainability of coastal systems by developing and combining System Approach for integrated assessment and deliberative tools to support decision-making. The implementation of system approach follows four logical steps. First, System Design step identifies the structure, function, and dynamics that should be studied to address a policy issue. Second, System Formulation step aims at represent the functioning of the system in both quantitative and qualitative terms by describing all the processes and interactions in the environmental and socio-economic spheres. Third, System Appraisal step assembles the components of the system and evaluates each functional parts. Fourth, System Output step involves the organization of the information for policy deliberations, scientific publication, and for dissemination the non-science end-user community. This procedure has been applied to 18 study sites to demonstrate how science and policy can be integrated (www.spicosa.eu).

One of the study sites is the Pertuis Charentais (France) where the policy issue is related to the quantitative management of the freshwater in the Charente river basin. This basin is characterized by low freshwater flows in summer and some needs related to human activities may not be satisfied – e.g. availability of drinking water for households and tourists, good ecological status of the coastal ecosystems (rivers, salt-marshes, nurseries, coastal water productivity) which support several services and environmental amenities. In addition, two private industries of the primary sector depend on Charente river flow: agriculture uptakes water for irrigation during summer for crop cultivation (mainly irrigated maize) and shellfish farming needs freshwater for spat production and river nutrients for oyster growth. This policy issue has been addressed by the regional plan for water management, which includes a “Water shortage Management Plan” (WMP). The implementation of SPICOSA methods is conducted jointly with a public agency (Territorial Public Agency for the Management of the Charente River) in charge of the implementation of the WMP (http://www.fleuve-charente.net/) in order to address the users conflicts generated by the freshwater scarcity in relation with ecosystem services provided by freshwater and knowing that the occurrence of conflicts will change due to long term trends (climate, agriculture activity, demographic changes). We developed a numerical model to compare different management options, under several assumptions regarding the forcings of the system and the behaviours of users. This paper focuses on the modelling of three components of the system which are tightly connected: hydrology of the Charente river, irrigation for the main agriculture crop, coastal zone primary productivity. It describes their implementation in the modelling tool used within SPICOSA project and shows some preliminary simulation results.

2. MODEL STRUCTURE

2.1 Modelling tool

The need of a common language for describing and modelling systems has been outlined by several authors. Costanza and Ruth (1998) mention the STELLA software as a tool capable to involve a group of relative modelling novices with the help of facilitators. ExtendSim (www.ExtendSim.com) belongs to this family of modelling software such as Vensim (http://www.vensim.com/), Stella (http://www.iseesystems.com/), PowerSim (http://www.powersim.com/), Simile (www.simulistics.com) which have a graphical user interface facilitating the construction of a model, the setup of simulations, the visualisation of the model outputs and the communication with non modelers (Odum and Odum, 2000; Ruth and Hannon, 1997). Compared to other softwares, ExtendSim has some appealing capabilities which make it a good candidate to simulate complex systems and which are extensively used in our modelling approach: i) ExtendSim handles databases which can be either built by importing external databases (Excel, Access) or generated with ExtendSim own tools. Databases are used to exchange data with other applications, store model results, exchange data between the model components; ii) the architecture of the model components...
(called blocks) is built by using extensive libraries of predefined objects which are assembled and connected graphically to one another in order to simulate discrete time or continuous time processes. For more complex modules, ExtendSim allows to develop customized blocks with its own powerful programming language very close to the classical C language and augmented with a lot of specialised functions – e.g. access to local database functions; iii) libraries of blocks are commonly used to share blocks with other modelers. Each block has its own user interface and contains the description of its functionalities. Libraries can be distributed and used for other purposes (a good example in Ecology is given by Odum and Peterson, 1996); iv) to help simplify and clarify models, ExtendSim makes possible to create hierarchical blocks that group several blocks together into one block while still allowing the user to drill down into the lower levels to access the individual blocks.

2.2 General model structure

The general architecture of the Pertuis Charentais system is displayed in Figure 1. The Hydrology box represents one Natural Component (NC) which interacts directly or indirectly with other NC (Coastal Productivity), Economic Components (EC - e.g. uses of freshwater, namely Agriculture/Irrigation, Biodiversity, Aquaculture) and Social Components (SC, e.g. Regulation). Arrows show the flow of information between these modules. Since the main issue is related to freshwater uses and production, much effort has been spent on the development and test of the Hydrology module.

![Figure 1. The simulation platform: user interface.](image)

Variables are passed from one block to the other by using the database capabilities, thus keeping a history of all relevant variables. Each module has its corresponding database that also records all model parameters. A specific effort was made in the development of graphical tools in order to visualize relevant indicators of the model and in a user friendly organization of the different blocks and modules. Scenarios, display and outputs blocks have been set to regroup all the model control panels.

2.3 Hydrological model

The watershed area of Charente river is around 10,549 km². Agriculture activity covers about 60% of this area and about 11% of the cultivated area is irrigated. The annual water supply to human activities is 4 millions m³, with 34% dedicated to drinking water and 57% to irrigation which are therefore the two main human activities of concern in the freshwater issue selected for this study site. Water supply is related to the hydrology of Charente river which depends on daily rainfall, soil characteristics and evapotranspiration (function of temperature and demand by plants, forests and agriculture). The hydrological model therefore simulates daily changes of river flow at the scale of the whole catchment area.
The watershed was discretized in a dozen of hydrological sub-basins which offered several advantages: i) it accounted for some specificity and spatial differences between soil features (e.g. karstic vs non karstic areas); ii) it was consistent with the spatial scale of the governance and water regulations (no shown here); iii) the mathematical equations were easily translated into ExtendSim software; iv) the model was already implemented in EXCEL®, calibrated, operational and used by managers to make decisions on restrictions of water supply to agriculture during drought events.

This model was initially developed by Eaucea, a private company which agreed on translating the equations and test datasets into ExtendSim. The Hydrology box was then built as a hierarchical block in ExtendSim to allow for more flexibility. Opening this block yields the structure shown in Figure 2 and composed of sub-basin models represented by a new series of blocks. Connectors were defined in each block to pass the value of the river flow calculated for each sub-basin to downstream blocks.

Following the ExtendSim procedure to pass the information between blocks, connecting lines were drawn between the connectors defined in each block. The name of each block is based on the location of its most downstream point. In our model terminology, this block is referred to as a watershed block. The watershed block is composed of three components to facilitate the model implementation. Thanks to ExtendSim programming facilities, data management and mathematical equations were coded in custom blocks which are generic blocks contained in the Hydrology library. The mathematical equations describe the dynamics of water levels in a series of reservoirs, the levels of which depend on rainfall, soil properties and evapotranspiration and which communicate with one another (Figure 3). Hydrological processes are therefore used to compute the river flow and the water available for crops and natural plants. Each reservoir is associated to one state variable and the time step used to simulated water levels in each reservoir is equal to one day.
For each watershed one Database has been created with the same name as the watershed hierarchical block. It contains several distinct tables for parameters, forcing time series, observed river flow time series, and state variables (see Figure 4). Such databases give much flexibility to the model, help in organizing the enormous quantitative information and facilitate the exchange of data between all the model components.

Figure 4. Examples of tables defined in the Database corresponding to one watershed and containing parameters, forcing variables, observed river flows, etc.

2.4 Irrigation module

The agriculture-irrigation module is closely linked to the Hydrology module. Since the hydrological model is based upon an existing hydrological model which has already been calibrated, it was chosen to preserve the structure of this model and to use the same set of sub-basins. The Agriculture module calculates, in a spatialised way -for each sub-basin- the values of the areas of irrigated and non irrigated crops, depending on the chosen scenario. It defines all the values of technical parameters needed for the calculation of the water demand by crops, which is done in the irrigation module. An output of this module is an indicator of an economic impact depending on the variations in irrigation water supply and the chosen scenario.

A typology has been built to define relevant spatial units and to describe the Charente watershed and agricultural activities in a simple way. With 40% of the usable farm area covered by cereals, the Charente basin is part of the second French pole of cereal production. Maize is the main irrigated crop and represents almost 80 % of irrigated area. Furthermore, 60% of the maize area is irrigated. For crops, it is important to evaluate maximum water loss under certain climatic conditions and under unlimited water availability at the root system level, i.e. the "maximum evapotranspiration" (ETM). The value of the "potential evapotranspiration" (PET) is taken from institutional data. The irrigation module allows to compute the daily volume needed for the irrigation of the irrigated crop area on the sub-basin. This volume is then subtracted from the river flow calculated in the Hydrology module. There is also a link with the Governance module: the thresholds defined by the local policy constraints have to be compared to the values of the estimated flow in the river. The comparison between the volume needed by the crops for irrigation and the available volume (either due to policy constraints or to climatic conditions) is used for the calculation of a stress indicator and losses in crop production (not detailed here).

The current version of the sub-basin irrigation block follows the same rule as the watershed hydrology sub-block and combines a custom block containing the equations and a database keeping track of the outputs (daily water demand for irrigation) and parameters. It is a simplified version of the Irrigation module based on a single culture and water uptake directly from the river.

2.5 Costal Productivity model

The Coastal Productivity component is designed to link the freshwater inputs from the Hydrology block and the Shellfish farming block. It aims at simulating primary
productivity of Marennnes-Oléron bay, which depends on nutrient fluxes from the watershed, ambient water temperature and light, and amount of cultivated oysters. The Coastal Productivity component is built as a hierarchical block which contains a few modules belonging to the same Coastal Productivity library (Figure 5). The main module is a generic custom block where the mathematical equations are written using the MODL programming language. As for all the other components, databases were designed to store parameter values, forcing variables and model outputs and are part of the model implementation in ExtendSim. The Coastal Productivity block also accesses the Hydrology database where it can retrieve the value of riverflow needed to compute nutrient inputs into the bay. In the same way, the block provides the phytoplankton concentration to the Shellfish Farm component through the History table used to save model outputs.

3. MODEL TESTS

Since the model is based upon an existing model which has already been calibrated, we checked that both models are producing the same results. This work is under progress, and we only show the output of one watershed, compared to the simulation produced by CycleauPE. Some slight differences between the two simulations were found, they are mainly due to differences between initial conditions and differences between the sequence of equations in EXCEL and ExtenSim, but they do not affect the overall good agreement between the 2 models (Figure 6).

Regarding the coastal zone, a huge amount of environmental data have been collected since 1977. Monthly values of the main environmental descriptors from 1977 until 2003 clearly show a strong seasonal pattern and, to test and calibrate the primary production model, we averaged annual variations and built an annual cycle for Salinity, Phytoplankton, Temperature, Nutrients. We also considered an average annual riverflow derived from available database, in order to test the model with real values of river inputs. The test
simulation shows a good agreement for salinity and nutrient concentration (Figure 7), though the phytoplankton seems to be underestimated.

![Figure 7. Comparison between observations and simulations a) Nutrients, b) Salinity, c) Phytoplankton.](image)

**4. DISCUSSION**

Our modelling group involved hydrological modelers, specialists of system approach and decision-makers in charge of advising local authorities with respect to irrigation for agriculture and water level in the watershed reservoirs. We debated in details the modelling strategy, e.g. the spatial scale and resolution of the watershed model. Alternative strategies have already been described or compared in other works (Voinov et al., 1999; Mathevet, 2005; Reynaud et al., 2008) and, technically, it is possible to interface ExtendSim with other softwares. For instance, some SPICOSA partners have considered the coupling between ExtendSim with PCRASTER (http://www.pcraster.nl/), a free community based software simulating spatially distributed systems. In our case, the subdivision in several catchment areas was preferred to a more detailed spatial discretization for several practical reasons, the most important being that this level of aggregation was consistent with the decision making scale. The model will therefore include a governance module which incorporates the rules regarding the access rights to water and their restriction during summer crisis. Such rules are based on freshwater availability and water demands. They are described in technical documents which will be translated into ExtendSim through decision rules which will act as feedback mechanisms on the irrigation component.

Communication with stakeholders and decision makers is essential in the System Approach framework. Heemskerk et al. (2003) shows how this way of thinking resulted in a common and transparent representation of systems during a multidisciplinary workshop where groups of participants analysed and represented different systems with a common set of symbols. Costanza and Ruth (1998) also emphasize the importance of supplementing mental models with dynamic models of systems which are too complex to be shared by different stakeholders and used to build a consensus upon a policy issue. A step further, Voinov and Gaddis (2008) describe in details the lessons learned from a participatory approach of watershed modelling and, among these, emphasize the role of stakeholders at all levels of the modelling process. In our case, stakeholders and decision makers have been involved from the very beginning of the project to identify policy issues, discuss and bring information on the appropriate temporal and spatial scales and boundaries of environmental, social and economic processes. Working groups have been held on the hydrology/agriculture components with several objectives: define appropriate spatial scales, construct system representations (mental models), identify policy options, provide input data (parameters, forcing functions, validation datasets), select scenarios of change. It is a continuous process, along the line defined by Voinov and Gaddis (2008) since the lifespan of the model is not limited to the project duration and expectations by decision makers and stakeholders will increase with the demonstration of model functionalities. As part of the system approach framework, other meetings with stakeholders were also dedicated to the definition of scenarios. The future of the system was divided into three domains: trends (climate, demography), management options and changes in uses and practices. Stakeholders have been asked to vote on combinations of assumptions and to select and refine the main indicators. In addition, scenarios of agricultural activities have
also been assessed through specific surveys among local authorities and professionals (experts, advisers, farmers) about the evolution of agriculture during the next ten years: evolution of the irrigated crop area, substitution of irrigated crops by other irrigated systems or other farming systems, implementation of specific policies, increase in the number of reservoirs or dams.

The amount of water in the watershed reservoirs and the river is used to build indicators of water shortage which show the locations and periods of time where and when regulation of water supply must be activated (this is the role of the Governance component, not shown here, Figure 1). It is clear that the availability of freshwater depend on several factors: climate (rainfall, evapotranspiration), agriculture activity (amount of irrigated crop), requirements of nutrients for shellfish farming. Simulations of scenarios will therefore allow to assess how such changes will increase or decrease the risk of conflicts between users.

5. LITERATURE


OpenMI 2.0 - What's new?

Gennadii Donchyts\textsuperscript{1,2}, Stef Hummel\textsuperscript{1,3}, Stanislav Vaněček\textsuperscript{1,4}, Jesper Groos \textsuperscript{1,5}, Adrian Harper \textsuperscript{1,6}, Rob Knapen\textsuperscript{1,7}, Jan Gregersen\textsuperscript{1,8}, Peter Schade\textsuperscript{1,9}, Andrea Antonello\textsuperscript{1,10}, Peter Gijsbers\textsuperscript{1,10} \\
\textsuperscript{1} The OpenMI Association, Lelystad, The Netherlands \\
\textsuperscript{2} Stichting Deltares, Delft, The Netherlands Gennadii.Donchyts@deltares.nl; \\
\textsuperscript{3} Stichting Deltares, Delft, The Netherlands Stef.Hummel@deltares.nl; \\
\textsuperscript{4} DHI, Prague, Czech Republic, S.Vanecek@dhi.cz \\
\textsuperscript{5} DHI, Hørsholm, Denmark, jgr@dhigroup.com \\
\textsuperscript{6} MHW Soft, Wallingford, UK Adrian.Harper@wallingfordsoftware.com \\
\textsuperscript{7} Alterra, Wageningen, The Netherlands Rob.Knapen@wur.nl \\
\textsuperscript{8} Hydroinform, , Denmark Gregersen@hydroinfrom.com \\
\textsuperscript{9} Bundesanstalt für Wasserbau, Germany Peter.Schade@baw.de \\
\textsuperscript{10} University of Trento, andrea.antonello@gmail.com \\
\textsuperscript{10} Deltares USA Inc., Silver spring, Maryland, USA Peter.Gijsbers@deltares-usa.us \\

Abstract: The first version of the OpenMI standard was developed as a joint effort of several European research organizations. OpenMI stands for Open Modelling Interface and aims to deliver a standardized way of linking environmental models at run-time. In the new version of the standard several new goals were defined based on experience obtained during migration and use of the OpenMI-compliant models. This includes on one side different IT aspects such as better object-oriented design of the standard and re-use of well-known engineering practices and patterns. On the other side, after successful implementation of OpenMI in many environmental models it was also decided to extend scope of the OpenMI standard to a broader set of applications such as GIS data types, monitoring databases, running models in parallel (versus sequential pull-driven approach), improved workflow management and many others. This paper gives the details of OpenMI 2.0.

Keywords: Open Modeling Interface; linking; model interoperability; integrated modeling

1. INTRODUCTION

The first version of the OpenMI standard was developed with an extensive focus on coupling of the numerical models, primarily in the field of surface and ground water hydraulics. Although notice was taken of the needs of other domains, e.g. economics, this hardly influenced the outcome of the OpenMI version 1.4. However, usage of OpenMI in these other domains, e.g. in the SEAMLESS project (Knapen et al. [2009]), or during application of OpenMI for web-based systems (Goodall [2007]) have shown that OpenMI requires considerable improvements to fit more properly to these type of applications (see Gijsbers et al. [2010], for details and argumentation).

Also for other reasons OpenMI turned out not to be perfect yet. While the list of OpenMI-compliant models grew\textsuperscript{1} it became clear that there were more issues to be considered then the domain area only. Better support was needed for time independent models, for algebraic models, and for GIS based applications.

Another important reason to adjust OpenMI was more technical: improve the quality of the standard by applying more object-oriented design patterns (Gamma et al. [1994]), and by using language and platforms specific features such as events, properties and standard collections, as available in both the .NET Framework and java. Making the standard more

\textsuperscript{1} See the list of OpenMI-compliant models on http://openmi.org
intuitive and self-explaining while at the same time expanding its scope was the main goal of the OpenMI Association's Technical Committee when developing OpenMI version 2.0. An important additional aspect taken into account was the fact that the implementation of an OpenMI 1.4 component usually relies quite heavily on the OpenMI 1.4 Software Development Kit. For computational cores this SDK provides more intuitive interfaces than the ones defined in the standard itself, so often component developers implemented the SDK's IEngine/IRunEngine interfaces instead of the more abstract ILinkableComponent in the standard. In the OpenMI 2.0 this will not be the case anymore, since ILinkableComponent (Figure 1) now better covers handling of the computational cores.

Figure 1 ILinkableComponent class diagram

The goal of this paper was to provide a technical overview of all major changes in OpenMI 2.0. The list below summarizes some main code changes of the OpenMI 2.0 which will be discussed in the present article:

- Combine all three concepts of Where, When and What in the IExchangeItem, whereas in the previous version Where, When in the previous version. What was part of the ILinkableComponent in the OpenMI 1.4.
- Remove the ILink interface and apply the Observer design pattern by connecting IOutput and IInput items by means of a provider/consumer relationship.
- Separate the concepts of Perform Computation and Retrieve Values. In OpenMI 1.4 they were combined in the GetValues() call.
- Make ILinkableComponent behave like a workflow activity, changing its status depending on operation performed.
- Use of the Adapter design pattern for data operations.
- Introduce language and platform-specific features into standard by allowing use of events, properties and standard collections Introduce loop-driven approach to run components
- Introduce, in addition to the pull driven control flow, a loop-driven control flow approach.
- Improve the management of a component's persistent state(s)
- Extend the exchangeable types of values with categorized variables that can represent either nominal or ordinal value categories
- Simplify time-related interfaces.
- Make OpenMI more OGC-friendly

2. NEW FEATURES

This paper discusses the technical details related to the mentioned changes, and provides argumentation on why they were applied.

2.1 Combining concepts of What, Where and When within IExchangeItem
From its beginning OpenMI used a concept of *What, Where* and *When* in order to describe values exchanged between different components. Probably the main change in implementations of these concepts in OpenMI 2.0 is that IExchangeItem now holds all meta information describing all of them (see Figure 2).

Typical steps required to exchange values in the OpenMI 1.4 are:

- Query information about element sets (*Where*) and quantities (*What*) from the exchange items defined in components.
- Create links between source and target components, element sets and quantities and add these links to the corresponding components.
- Prepare and initialize components.
- Perform the call *GetValues*(ITime time, string linkID) on a component, which on its turn can also pull values from other components by performing the *GetValues* call. The time for which values are required is provided as an argument to the *GetValues* method.

Note that *Where* and *What* are mainly used at configuration time and are defined in IExchangeItem, while *When* is used at a runtime as an argument to ILinkableComponent.GetValues call. IExchangeItem in OpenMI 1.4 did not provide any information about time, nor did it allow querying values available for an exchange item.

In OpenMI 2.0 all three aspects: *What, Where* and *When* are defined on IExchangeItem level, which results in a better separation of concepts. Figure 2 shows the interfaces used, and gives an example of an exchange item, e.g. water level, on an element set containing 5 elements (e1-5) and time set containing 2 time steps (10:00, 12:00). So in OpenMI 2.0 IExchangeItem is fully responsible to hold all meta-information plus values being exchanged while ILinkableComponent is responsible for performing an operation, e.g. query database, convert raster data to something else, and perform a computation for a few time steps, etcetera.

Note that in the new version IExchangeItem uses a property called ValueDefinition instead of Quantity in the previous version. The value definition allows the usage of either a Quantity or a Quality, in order to define values for nominal or ordinal variables, e.g. land use types, concentration level (low, medium, high). Quantity or Quality are defined as interfaces extending IValueDefinition.
In the new version we do not use ILink anymore. For several reasons the link turned out to be more a burden than a benefit: confusion arises when more than one link has been connected to one input item; even for simple value retrieval a full link has to be created; the link needs a target component, so the values retriever always has to be a linkable component itself. In the new version we use the Observer design pattern defined on IExchangeItem level, see Figure 3. The figure shows the IInput and IOutput interfaces which represent different types of exchange items. The main difference from the previous version is that exchange items are now fully self-contained, responsible to keep all information which can be exchanged for a single variable, including values currently available. The linkable component is responsible to fill in values in all these exchange items. Implementations of the exchange items can be used and tested separately from the components.

A single output exchange item may provide values for multiple input exchange items. At configuration time output exchange items must be connected to input exchange items by means of the Consumers and Provider property. These properties define all established links, and therefore can be used to check how exchange items are connected. Once all exchange items are connected, components owning these exchange items may generate values by means of the Values property. Usually this happens after the component has performed some work (at the end of the Update() call, see chapter 6 for more details).

Another way to retrieve data, more known to users of OpenMI 1.4, is to perform a GetValues(IExchangeItem query) call on output exchange item. In OpenMI 2.0 this method is defined on the output exchange item instead of on ILinkableComponent, as it was in the previous version.

![Figure 3 Output and Input exchange items](image)

### 2.3 Use of Adapter design pattern for data operations

OpenMI 1.4 allowed user to define various data operations when linking exchange items. However, they were not allowing performing different conversions in a chain, and the use of the data operations was not intuitive enough. They were defined on IOutputExchangeItem, and then had to be passed to the ILink implementation. OpenMI 2.0 simplifies this logic by introducing another type of output exchange item, called IAdaptedOutput, which can wrap any output exchange item (adaptee) in order to perform a certain conversion.
2.4 Separating compute and accessing values logic in the ILinkableComponent

The new version of OpenMI provides much better control over the workflow by means of separation of “Perform Calculation” and “Query Values” steps. In the previous version there was no way to query already computed values from the component unless component itself provided some kind of buffering (e.g. using buffer classes available in SDK). In the new version this mechanism was completely reworked so that current values can be queried at any time. As well as buffered values, if buffering is implemented using IAdaptedOutput. Additionally the new standard allows checking the status of the component, which might be required in order to know if values available on exchange items are up-to-date. Otherwise the Update() method must be called on the component. The table below summarizes differences between steps required to perform computation and query values in both versions.

<table>
<thead>
<tr>
<th>Version</th>
<th>Perform operation</th>
<th>Query values</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.4</td>
<td>IValueSet ILinkableComponent.GetValues(ITime time, string linkId) – returns set of values provided by the source component of link defined between 2 components, quantities and element sets.</td>
<td>IOutput.Values – property which can be used after component was updated and values were set on all required output exchange items.</td>
</tr>
<tr>
<td>2.0</td>
<td>ILinkableComponent.Update() – updates component to the next valid state. Usually calling this methods performs a time step, queries data from a database or performs any other activity required to generate values in component output exchange items.</td>
<td>IValueSet IOutput.GetValues(IExchangeItem query) – returns set of values for a given value definition, times, element set provided by query. If necessary – it may call update of the component where this exchange item is used, and perform GetValues calls on other components.</td>
</tr>
</tbody>
</table>

As can be seen from the table, OpenMI 2.0 provides 2 ways to query values from exchange items. The Values property can be used to get values which were generated previously, mainly after the Update() call. Additionally, the GetValues() method can be used to get a specific set of values from the exchange item. This is very useful for instance when target component / exchange item is not interested in all values available in the output exchange item but only in a subset. When specifying a query time that lies in the future, the providing component will have to propagate itself to the requested time; this leads to exactly the same pull driven control flow as in OpenMI 1.4.

2.5 Making ILinkableComponent behave like a workflow activity

Before version 2.0 OpenMI was mainly used in a single threaded environment. Components were not supposed to be used from another thread during e.g. GetValues() call. The new version does not have this limitation anymore. ILinkableComponent provides a new property LinkableComponentStatus Status { get; } which allows to determine what the component is doing right now. This opens new possibilities to use of OpenMI. However developers of the components should implement their components as thread-safe, if these components are expected to be used in a multi-threaded environment. In fact it means that
IOutput.GetValues() and Values property of IInput and IOutput interfaces must be thread-safe. The OpenMI 2.0 does not imply that all its methods need to be thread-safe, for example linking components to each other most probably will happen in a single-threaded environment.

![Linkable component state chart diagram](image)

**Figure 5** Linkable component state chart diagram

### 2.6 Pull and Loop Driven Approach — the actual data exchange between components

In OpenMI 2.0 components can support two ways of control flows, called *pull-driven* and *loop-driven*. In the *pull-driven* control flow (default) components work like in OpenMI 1.4. Which means that a component may call its own Update(), as well as the GetValues() call on output items of the other components.

Another way to run components (*loop-driven*) assumes that components or exchange items should never trigger their own Update() call, nor the Update() and GetValues() of other components and their output items. Propagating the system by means of the Update() call should happen somewhere outside, in the control program containing these components. The *loop driven* approach requires more careful implementation, but it also opens a new ways to run components. For example, it allows control program to run components in the different threads, processes of even machines. In order to specify if a component supports *loop-driven* way of work a new property was introduced on ILinkableComponent:  

```csharp
bool ILinkableComponent.CascadingUpdateCallsDisabled { get; set; }
```

The default value is false, which means that component is allowed to trigger other components. If this property is set to true, the component must never trigger other components. Usually this means that when a component requires certain input for its input exchange items that is not available on the related output exchange item of a connected component, the component should change its status to *WaitingForData* and check if the data is available during the next Update() call.

### 2.7 Improvements of the component’s persistent state management.

OpenMI 2.0 facilitates a way to handle persistent states of linkable components. In the previous version it was only possible to trigger a component to remember its state, resulting in a string identifying the remembered state. The new version allows external programs to remember component states. In this case component must implement IByteStateConverter interface in addition to IManageState, see Figure 6.
2.8 Simplification of the time-related interfaces

After review of the time-related interfaces in OpenMI 1.4 a few interfaces were removed from the standard. It showed that a single interface ITime can provide sufficient functionality for specifying a time stamp or a time span. ITime represents a time interval by defining the start of that interval as a time stamp and by specifying a duration. Duration equal to 0 simply means a time stamp. This approach simplifies the use of this interface in the Times property of the ITimeSet, a new interface in the OpenMI 2.0. ITimeSet in this case works similar as IElementSet: the elements set defines the spatial properties of an exchange item, the time set defines the time frame properties.

2.9 Making OpenMI more OGC-friendly

OpenMI uses a single interface IElementSet in order to define geometry of the exchange items. This interface was improved in order to simplify interoperability with OGC Simple Geometry Specifications standards. Since there are no standard C# and java versions of the OGC standards available yet, except GeoAPI.NET open-source project effort, it was decided to keep IElementSet as it was, and only extend it with a few properties: SpatialReferenceSystemWkt, HasM, HasZ.

2.10 Language and platform specific features

After careful consideration OpenMI was extended with platform-specific features available in C#/.NET and in java. These features include:

- Events
- Properties (in java still represented by get/set methods)
- Collections

It was decided that the benefits that developers will gain during use of these features will make the standard more acceptable in the .NET and the java communities. And finding a proper alternative to those features is always possible when OpenMI has to be used in any other nowadays language. This allows the use of linkable components in a way as shown in code listing on Figure 9.

![Figure 9 Use of events in OpenMI 2.0](image)

CONCLUSIONS

The list of the major changes to the OpenMI standard has been presented. OpenMI 2.0 is certainly a major step forward in the field of environmental modeling and we hope that the new features discussed here will simplify its application and result in better interoperability between different components. It will take a while before the new standard will fully replace the previous version. However, taking into account the new possibilities which OpenMI 2.0 opens to the developers, we hope that this paper stimulates the migration of existing components to the new standard and the development of new OpenMI components.

ACKNOWLEDGEMENT

The development of OpenMI version 2.0 has been funded by the members of the OpenMI Association as well as the European Commission through various 6th Framework Programs and the LIFE Environment program.

REFERENCES

Gamma Erich et al., Design Patterns: Elements of Reusable Object-Oriented Software. *Addison-Wesley Professional* (ISBN 0-201-63361-2), 1995

Gijsbers P., From OpenMI 1.4 to 2.0. *iEMSS 2010, in review*, 2010


Parallel computing of a large scale spatially distributed model using the Soil and Water Assessment Tool (SWAT)

S. G. Yalewa, A. van Griensven, L. Kokoszkiewicz

aUNESCO-IHE Institute for Water Education, Department of Hydroinformatics and Knowledge Management, 2601 DA Delft, The Netherlands (a.vangriensven@unesco-ihe.org)

bCERN - European Organization for Nuclear Research email: lukasz.kokoszkiewicz@cern.ch

Abstract: The Soil and Water Assessment Tool (SWAT) has been used widely for large scale applications, reaching entire continents. Within the EU funded EnviroGrids project, a detailed application of SWAT on the Black Sea Basin is envisaged using high resolution data. In order to support the computation, the model is run on a computer grid. The use of the SWAT allowed for such computations with little adaptations to the source. A 3-step procedure is needed. In the first step, a program is run in order to split the model into several sub-models. Afterwards, the sub-models are run in parallel. In a last step, the outputs of the sub-basins are collected at a central computer and the routing is performed. High computations are also needed when simulations have to be repeated, such as for sensitivity, calibration and uncertainty analysis. In these cases, the simulations are repeated for different parameter sets. In this paper, we discuss the gridification of the algorithm “LH-OAT” that performs sensitivity analysis and has been linked to the SWAT model. The results show a clear improvement in calculation time. Nevertheless, it is concluded that the parallel computing of a distributed model is mainly beneficial for large scale applications with high resolution, while running the sensitivity analysis algorithm has more general and obvious benefit. In a next step, the gridification will be optimised depending on the application and the overheads that are due to submission and receiving of files, as well as potential waiting times for executions on the grid.

Keywords: Parallel computing, grid computing, hydrological model, SWAT

1. INTRODUCTION

The Soil and Water Assessment Tool (SWAT) is an increasingly popular large scale watershed modelling software. The scale of the problems that use this modelling software seems to have increased to a remarkably large degree over time. The whole of continental US and the whole of Africa were modelled on separate occasions using SWAT with commendable results.

The effort being carried out by the EU funded ‘EnviroGRIDS@Black Sea Basin’ project is another of such pushes that require large scale SWAT models to be run and deliver at near real-time. The ‘EnviroGRIDS@Black Sea Basin’ is a project among whose aims are linking, gathering, storing, managing and distributing key private or hard-to-find environmental data in an easily accessible, shareable, and process-able format. Developing
the access to real time data from sensors and satellites, and an early warning system that will inform in advance decision makers and the public about risks to human health, biodiversity and ecosystems integrity, agriculture production or energy supply provoked by climatic, demographic and land cover changes are among the goals of the project. The project is working to make such tools and services to be available on a grid infrastructure. It intends to use the Soil and Water Assessment Tool (SWAT) to create high resolution hydrological models of the entire Black Sea Catchment.

This push for large scale modelling, or the need for near real-time output, however, comes at the expense of a huge computation power. This expense will prove even costlier whenever there is a need to iteratively run such very large scale models for common practices in modelling such as calibration and uncertainty analysis. Several endeavors for the reduction of computation time in large scale processes have been made, both related to SWAT and other model codes.

This study describes two types of gridification for hydrological modelling: one where a single simulation is run on a grid by splitting the model into sub-models. This concept is applicable for the computation of large scale spatially distributed SWAT models, thereby reducing the computation time, among other advantages, by parallelization of components or processes. In a second category of the grid application, multiple simulations of a single model are split over the computation grid, as may be needed for model parameter calibration, sensitivity and/or uncertainty analysis tools. In such types of applications, 1000 CPUs may be made to compute 1 simulation each instead of having to do 1000 simulations on a single PC.

1.1 Grid computing

Since its inception in the late 1990s, grid computing has advanced so rapidly to take an appealing acceptance by major institutes and projects worldwide. In recent years, Grids have emerged as wide-scale distributed infrastructures that support the sharing of geographically distributed, heterogeneous computing and storage resources.

Coordinated by CERN (the European Organization for Nuclear Research) and funded by the European Commission, the EGEE project (Enabling Grids for E-sciencE) aims to set up a worldwide grid infrastructure for science. The EGEE Grid initially focused on two well-defined application areas, Particle Physics and Life Sciences, due to the fact that these communities were already Grid aware and ready to deploy challenging real applications at the beginning of the project. A range of applications are supported by this grid currently.

Many grids are used for e-science: enabling projects that would be impossible without massive computing power. Scientists can now share data, data storage space, computing power, and results easier than ever before via the grid which enables researchers to tackle bigger questions: from disease cures and disaster management to global warming and the mysteries of the universe.

1.2 EGEE Grid infrastructure

The EGEE Grid (Enabling Grid for E-sciencE) is currently the largest multi-science Grid infrastructure in operation which federates over 250 resource centres world-wide, providing more than 68,000 CPU’s and more than 20 Petabytes of storage. This infrastructure is used by several thousands of scientists distributed in over 200 virtual organizations with access to EGEE resources that is managed by the gLite middleware.

The EGEE grid, like any other grid infrastructure, has its own middleware and access protocols. It demands the software to be run on the Grid to be developed in specified (parallelized) formats for better computational efficiency on the grid environment. Whereas new software tools might be developed with these requirements in mind, existing scientific legacy codes or applications, the majority of which are not grid-aware yet, often require major code restructuring before they could run on the grid.
Gridification is adapting applications to include new layers of grid-enabled software. An application that ordinarily runs on a stand-alone PC must be "gridified" before it can run on a grid. Just like "webifying" applications to run on a web browser, grid users need to "gridify" their applications to run on a grid (GridCafe2). Once gridified, thousands of people will be able to use the same application and run it on interoperable grids.

For example, a gridified data analysis application will be able to:

- Obtain the necessary authentication credentials to open the files it needs
- Query a catalogue to determine where the files are and which grid resources are able to do the analysis
- Submit requests to the grid, asking to extract data, initiate computations, and provide results.
- Monitor progress of the various computations and data transfers, notifying the user when analysis is complete, and detecting and responding to failures (collective services).

In this study, we will gridify SWAT on the EGEE grid and demonstrate it with an auto-calibration routine for a water quality model.

2. GRIDIFYING A SWAT MODEL

2.1 Soil and Water Assessment Tool (SWAT)

The Soil and Water Assessment Tool (SWAT) is a physically based model that can simulate water quality and quantity at the watershed scale (Neitsch et al., 2005). SWAT computes continuous on a daily time step. The model is primarily designed to predict impacts of management on water, sediment, and agricultural chemical yield in gauged and un-gauged basins. It allows for the watershed delineation into a large number of sub watersheds. The major modules in the model include hydrology, erosion/sedimentation, plant growth, nutrients, pesticides, land management, stream routing, and pond/reservoir routing. Input climate (precipitation, maximum and minimum temperature, relative humidity, wind speed, solar radiation) are required on a daily temporal resolution, although the most recent version of the model allows also for hourly input files.

A watershed modelled using SWAT is partitioned into different required and/or optional objects of subunits such as sub-basins, reaches/main channel segments, impoundments/reservoirs on the main channel network and point sources. First, a watershed in SWAT is subdivided into sub-basins which are, thus, the first level of the subdivision. Sub-basins are defined by geographical positions in the watershed and are spatially related to one another (SWAT User’s Manual, Version 2000). All sub-basins drain into the river network where water is routed from upstream to downstream reaches. Sub-basins will contain at least one, and theoretically as many as unlimited number of, HRUs (Hydrologic Response Units) and a main channel or reach. Furthermore, impoundments, a pond and/or a wetland, may also be defined in a sub-basin.

Secondly, the land area in a sub-basin may further be divided into hydrologic response units, so called HRUs. Hydrologic response units are portions of a sub-basin that possess unique land use, management, or soil attributes (SWAT User’s Manual, Version 2000). Unlike in the case of sub-basins, no spatial relationship or interaction can be specified among HRUs. Sediment, chemicals or nutrient loadings from each HRU are computed independently and then summed up to determine the total load from the sub-basins.

A watershed model should also incorporate one reach or main channel associated with each sub-basin. This channel carries loadings from the sub-basin or outflow from the upstream reach segments into the network of the watershed in the associated reach segment.
The parallelization of the Soil and Water Assessment Tool (SWAT) models for Grid execution is analyzed with various approaches. The general principle of this study might simply be stated as: divide-compute-merge. In other words, it follows the general principle of dividing or splitting a large scale, high resolution SWAT model into several small model components or subcomponents and computing all in parallel on the Grid, after which outputs from all these components or subcomponents are collected and merged for presenting results similar to that of the original, undivided model. Methodological questions such as "On what level can one split SWAT models for optimal efficiency: sub-basin or HRU level?" are among the first that require some analysis to answer. This study focuses on sub-basin level splitting for speculative reasons that HRU level splitting might increase overhead while splitting and merging, and communication overheads while computing on the grid. Further questions that require precise analysis are: From which files do we retrieve all the information for independent computations of such sub-basin level splits? Do we have to split all sub-basin specific information into independent sub-basin files or could we still maintain some common input files for all the sub-basins as they are in the original model? What is the advantage or disadvantage of splitting or maintaining those common input files? This study will try to give answers to these and other questions in the following sections.

Figure 1: The “Fig file” (a) for the SWAT model of the Nzoia catchment (Kenya) (b) and the corresponding conceptual schematization for this file (c), triangles represent the sub-basin and the lines represent the reaches (c).

Within a watershed, the sub-basin processes are computed independent from each other and within the sub-basin, the HRUs are computed independently as well. On the other hand, the computations of the reaches depend on the outputs of the sub-basins computation, but not the other way around. In other terms, reaches in sub-basins are supposed to incorporate loads from sub-basin processes, and thus are dependent on them. The reaches also depend on each other (downstream reaches depend on the outputs of the upstream reaches). For that reason, the parallelization is performed at the level of the sub-basins only, and not for the reaches. All HRUs within a sub-basin share the same weather input file, which are the largest model input files in SWAT. An HRU level splitting, thus, would mean the processing of these big input files during splitting, which is computationally costly. It was chosen not to go up to HRU level splitting of SWAT models in this study for the simple reason that doing so would reduce the gain in computation time. The level of sub-basins is for that reason much more efficient in exchanging of input- and output files.

Watershed configuration in SWAT models is stored in a file called “fig-file”. This file consists of a number of “command lines” for each time step to be computed in sequence. Among the most important command lines are “sub-basin” that computes the sub-basin processes, “route” for the computation of the processes in a reach and “add” that sums outputs from previous command lines. A typical structure of the file is given in the example of figure 1a, where the first block represent lines for the computations of the sub-basins (7...
in total), followed by a block with several “route” and “add” commands. The latter represents the river network.

![Diagram](image)

Figure 2: The original configuration “sequence” (a), the split SWAT models “paral1”(b) and “paral2” (c)

In the original SWAT code, the sequence of computation is as in Figure 2(a), all sub-basins, and all HRUs within the sub-basins, are computed first followed by the main channel networks/reaches. These computations are iterated for every day of every year stated for the simulation period. Two approaches of parallelization were devised in this study. In a first parallelization configuration, all sub-basins are computed in parallel (“paral1” in figure 2b), followed by all reaches in series. In a second configuration, the reaches of upstream sub-basins were computed during the parallelization process of the sub-basins in sequence to the sub-basin they belong to, followed by the reaches that depend on upstream reaches “paral2”, as presented in figure 2c.

2.3 Software tools used for parallelization of the SWAT model

In total 5 programs are used:

1. SWAT2005.exe: the original SWAT model code
2. SWATgrid.exe: adapted SWAT model code
3. Splitter1 or Splitter2: written for the splitting SWAT models (python/java)
4. SWATmerger.exe: written for merging sub-basin outputs (python/java)
5. Ganga: A job submission tool for the EGEE grid (python)

SWAT2005.exe is the original SWAT code and is used for the computation of the sub-basins (and the routing of the upstream reaches in the second configuration). SWATgrid.exe is an adapted version of the SWAT model that reads in the output files from the models of the sub-basins run on the grid, and initiates the add/route commands. The major change with the adapted SWAT code used for this purpose from the original is that it allows the routing of stream networks by skipping sub-basin computations.

Splitter1 splits the model into several sub-models. It creates a model for each sub-basin. The following files needed to be adapted:

- Fig.fig: this file is reduced to one command line for the sub-basin computation and one to save the outputs to a file.
- The precipitation (*.pcp) file contains only the rainfall gage data that is used in the sub-basin
- Only the input files that are use, or copied to the folder represent the sub-basin, including the executable program swat2005.exe
An adapted configuration file is created for the model that will read the outputs of the sub-basins that performs the routing. Splitter2 is similar to Splitter1, except that it includes a routing command and a routing input file in the configuration for routing of upstream tributaries. SWATmerger.exe copies the output files of the sub-basins into a single folder, and calls the swatgrid.exe code.

Ganga (a python script) is a grid user interface for specifying and for submitting jobs to the grid which connects to the grid, submits the files to the grid machines and copies the results back to the PC of the user.

3. Parallelization of the LH-OAT Sensitivity Analysis Algorithm

3.1 The LH-OAT algorithm

Sensitivity analysis is the process of determining the rate of change in the model output with respect to changes in the model input parameter. It is needed to determine how the hydrological model outcomes are sensitive to their parameter. Furthermore, most hydrological models are very complex and over-parameterized. The identification of the most important parameters so as to reduce time and complexity, before carrying out time-intensive activities such as model calibration, requires a faster and efficient way of sensitivity analysis tool and technique.

In this study, parallelization of the LH-OAT software tool which implements the Latin Hypercube-One factor At a Time method (van Griensven et al, 2005) is carried out.

The LH-OAT tool performs a Latin Hypercube (LH) sampling followed by One-Factor-At-a-Time (OAT) sampling (Figure 3). It starts with taking N Latin Hypercube sample points for N intervals, and then varying each LH sample point P times by changing each of the P parameters one at a time.

The method operates iteratively. Each loop starts with a latin-hypercube point. At each Latin Hypercube point j, a partial effect $S_{i,j}$ for each parameter $e_i$ is calculated as (in percent):

$$S_{i,j} = \left(100 \times \frac{\left(M(e_1,\ldots,e_i^*,(1+f_i),\ldots,e_p) - M(e_1,\ldots,e_i^*,\ldots,e_p)\right)}{\left(M(e_1,\ldots,e_i^*,(1+f_i),\ldots,e_p) + M(e_1,\ldots,e_i^*,\ldots,e_p)\right)/2}\right)$$

where $M(.)$ refers to the model functions, $f_i$ is the fraction by which the parameter $e_i$ is changed (a predefined constant) and j refers to a Latin Hypercube point. In equation 1, the parameter was increased with the fraction $f_i$, but it can also be decreased since the sign of the change is defined randomly. Therefore a loop requires $P+1$ runs.

A final effect is calculated by averaging these partial effects of each iteration for all Latin Hypercube points (thus for N loops). The method is very efficient, since N intervals (user defined) in the LH method require a total of only $N*(P+1)$ runs.

The final effects can be ranked with the largest effect being given rank 1 and the smallest effect being given a rank equal to the total number of parameters analysed. Oftentimes during a sensitivity analysis for a particular dataset, some parameters have no effect on model predictions or performance. In this case they are all given a rank equal to the number of parameters.

This method combines the robustness of the Latin Hypercube sampling method which ensures that the full range of all parameters has been sampled with the precision of a One-factor-At-a-Time (OAT) sampling method. The OAT sampling method assures that the changes in the output in each model run can be unambiguously attributed to the parameter altered in the input.

3.2 Parallelization of the LH-OAT algorithm
A LH-OAT application requires a number of simulations (equal to number of intervals for the Latin Hypercube times the number of parameters +1). These simulations are independent from each other and can thus easily be parallelised in order to perform the simulations, and to collect the objective function or a particular model output of interest, for each of these simulations. In our example we look at the average flow and the average sediment load at the outlet as an output of interest.

The parallelisation is performed by splitting the file with the parameter values into several files. Each of these files controls a batch execution that can be done on a particular processor. All the result files, containing the average flow and sediment load value for the simulations, are merged to one file (named sensobjf.out).

3.3 Software Tools used

The following software tools are used for the parallelization of the LH-OAT algorithm:

1. **LH_OAT_sampling.exe**: a program that samples the parameter sets, based on parameter ranges and the number of intervals in the Latin Hypercube
2. **SensSplitter.py**: a python script that creates subfolders, each of which containing the model and a file containing parameter files. The folder also contains a file with specifications of how the parameters should be changed (changepar.dat). In addition, the “ICLB” variable within master watershed file of SWAT models (file.cio) should get a code-value ‘5’ indicating that a batch file (batchpar.dat) with parameter sets should be performed.
3. **swat2005.exe**, that the original SWAT model
4. **SensMerger.py**: a python script that merges the files from the sub-parameter sets into one file (sensobjf.out)
5. **LH_OAT_analysis**: an algorithm that produces sensitivity rank for parameters, based on the parameter sets (senspar.out) and the objective functions (objpar.out)

After splitting the SWAT models and the LH-OAT parameter sets on a local machine using the respective splitter tools developed in this study, each sub-component is submitted to multiple high performance computing machines on the EGEE grid. The results are collected and presented using the respective merger tools developed for this same purpose in this study. The outcomes of the tests are depicted and discussed below.

4. RESULTS

4.1 Parallelization of SWAT Models

The computation times for the model are depicted in tables 1 to 4 below. In table 1, a small model, presented in figure 1, with 7 sub-basins simulating data for 43 years run on 7 CPUs on the grid, shows a speedup ratio of 2.96 in computation. A larger model built for the Balaton Lake basin, containing 204 sub-basins simulating data for 16 years, run on 204 CPUs, achieved a speedup of 1.4 in computation time. From these results, it seems valid to conclude that the speedup in computation times is higher when the number of simulation period is longer. This is because execution time for splitting and merging of the models is more sensitive to number of sub-basins than number of years of simulation.

<table>
<thead>
<tr>
<th>7 sub-basins, 7 HRU’s:</th>
<th>Computation time (s)</th>
<th>Number of CPUs</th>
<th>Speedup</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full model (”sequence”)</td>
<td>32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parallelisation Experiment</td>
<td></td>
<td>Approach I</td>
<td>Approach II</td>
</tr>
<tr>
<td>Splitting</td>
<td>1.2</td>
<td>1.4</td>
<td></td>
</tr>
<tr>
<td>Sub-basin</td>
<td>3.3</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Merging</td>
<td>6.3</td>
<td>4.4</td>
<td></td>
</tr>
<tr>
<td>Parallel computing</td>
<td>10.8</td>
<td>10.8</td>
<td>7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>204 sub-basins, 204 HRU’s:</th>
<th>Computation time (s)</th>
<th>Number of CPUs</th>
<th>Speedup</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full model (”sequence”)</td>
<td>1181</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parallelisation Experiment</td>
<td></td>
<td>Approach I</td>
<td>Approach II</td>
</tr>
<tr>
<td>Splitting</td>
<td>4.4</td>
<td>4.4</td>
<td></td>
</tr>
<tr>
<td>Sub-basin</td>
<td>5.3</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Merging</td>
<td>6.3</td>
<td>4.4</td>
<td></td>
</tr>
<tr>
<td>Parallel computing</td>
<td>1188</td>
<td>1188</td>
<td>1.4</td>
</tr>
</tbody>
</table>
Another interesting scenario is the same model with 7 sub-basins but a much higher resolution, that is, multiple HRUs per sub-basin. This scenario, with a speedup of 6.3 as shown in Table 2, demonstrates a clear gain in computation time for high resolution models, which is the objective of this study.

4.2 Parallelization of LH-OAT Sensitivity Analysis Algorithm

The computation time is equal to the total computation time divided by the number of sub-models that are created for the parallelization. As shown in the table below, splitting 330 parameter sets \([10 \times (32 + 1)]\) for independent simulations on 330 processing units on (CPUs) on multiple HPCs on the grid gives a speedup ratio of 7.3. Similar to the parallelization of high resolution SWAT models, this result shows a clear gain in computation time.

<table>
<thead>
<tr>
<th>Task</th>
<th>Computation for 1 CPU in s (m)</th>
<th>Speedup for gridcomputing (#CPU’s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7 sub-basins, 7 HRU’s</td>
<td>32</td>
<td>2.96 (7)</td>
</tr>
<tr>
<td>7 sub-basins, 1390 HRU’s</td>
<td>2671</td>
<td>6.9 (7)</td>
</tr>
<tr>
<td>204 sub-basins, 204 HRU’s</td>
<td>249</td>
<td>1.4 (204)</td>
</tr>
<tr>
<td>Sensitivity analysis</td>
<td>2835</td>
<td>7.3 (330)</td>
</tr>
</tbody>
</table>

5. CONCLUSION

The ‘gridification’ of the Soil and Water Assessment Tool allows for running the model on a grid environment. The results show an improvement in the performance in computation time, too. Nevertheless, there may be a considerable loss of time during the job submission procedure between the user’s PC and the grid which results in communication overhead, especially with very small models. It is therefore unlikely that grid computing would become very beneficial for small to mid-size cases. Nevertheless, for large cases, especially with multiple of HRUs per sub-basin, grid computing may gain a lot in computation time. In addition, it may also solve memory problems that may originate from running large models on a single computer. The applicability of gridified algorithms for multiple simulations, such as for sensitivity analysis, calibration and uncertainty analysis is more straightforward.

An important limitation of the present GRID-computations is that there are no interactions possible between the computer nodes. For that reason, we could not account for any relationships between the sub-basins or between rivers and sub-basins (eg water transfers). Also for integrated modelling, where model-dependencies are present by default, such grid computations may not be helpful in reducing the high computation load. In a next step, tests of the parallelization tools and techniques used so far in this study on multiple processors on the same HPC machine will be experimented.

ACKNOWLEDGEMENTS

The authors wish to thank the EU/FP7 EnviroGRIDS@BlackSeaBasin project for the financial support.

REFERENCES


Rethinking Modeling Framework Design: 
Object Modeling System 3.0

O. David\textsuperscript{a}, J.C. Ascough II\textsuperscript{b}, G.H. Leavesley\textsuperscript{c}, L. Ahuja\textsuperscript{b}

\textsuperscript{a}Depts. of Civil and Environmental Engineering and Computer Science, Colorado State University, Fort Collins, CO 80523 USA (e-mail: odavid@colostate.edu)
\textsuperscript{b}USDA-ARS-NPA, Agricultural Systems Research Unit, Fort Collins, CO 80526 USA
\textsuperscript{c}Dept. of Civil and Environmental Engineering, Colorado State University, Fort Collins, CO 80523 USA

Abstract: The Object Modeling System (OMS) is a framework for environmental model development, data provisioning, testing, validation, and deployment. It provides a bridge for transferring technology from the research organization to the program delivery agency. The framework provides a consistent and efficient way to create science components, build, calibrate, and evaluate models and then modify and adjust them as the science advances, in addition to re-purposing models for emerging customer requirements. OMS was first released in 2004 and version 3.0 represents a major milestone towards an easier to use, more transparent and scalable implementation of an environmental modeling framework. OMS3 development is the result of an in-depth analysis of successful framework designs and software engineering principles as provided by general-purpose modeling frameworks. Like any modeling framework, OMS3 is enabling technology for modeling. The main goal of OMS3 development is an easier integration of model source code based on language annotations while being flexible to adopt existing legacy models. In OMS3, the internal complexity of the framework itself was reduced while allowing models to implicitly scale from multi-core desktops to clusters to clouds, without burdening the model developer with complex technical details.

Keywords: Modeling framework; Software design; Component modeling; Noninvasive framework.

1. INTRODUCTION

Frameworks in general support efficient software development; environmental modeling frameworks (EMF) (Rizzoli et al, 2008) target modelers to help implement models that are state-of-the-art in science and software engineering, operate across domains, scale with problem complexity, and can be integrated with standard methods for calibration, optimization, and sensitivity/uncertainty analysis. Using an EMF for model implementation has advantages not just for the model developer(s) - organizations not initially involved in the model creation process can benefit from EMFs as well. EMF-based models are more interoperable (since they typically adhere to some programming standard), can be coupled and integrated, and new model uses can be implemented more quickly. In addition, EMF-based models have an enforced structure that is provided by the framework so they can be better understood by developers outside of the initial development team. All the above benefits are available once a model is implemented using an EMF, or a legacy model is adapted to an EMF.

An initial learning curve for EMF-based modeling always exists. Depending on the scope and comprehensiveness of the EMF, the modeler typically has to adopt programming techniques such as structured programming and modular decomposition (always a good practice), and be proficient in object-oriented programming or even architectural design patterns. If the framework targets high performance computing, a basic understanding is helpful towards dividing a complex problem in such a way that many CPUs can process models with measurable performance gain. The institutional support and acceptance of an EMF might help providing resource and in-kind support, but it is up to the modeler to
adopt an EMF in a way that is an integral part of the modeling workflow. Besides cultural and social barriers, there is also the technical challenge for the framework developer to lower the burden for successful framework adaptation by modelers. Web service or database framework projects outside the modeling community have demonstrated that model developers will adopt a framework if it is easy to approach, fits into an existing codebase and workflow seamlessly, and therefore does not fully invalidate existing practices of software development.

OMS in its previous versions (2.x and earlier) represented the traditional framework development approach, i.e., building a framework supporting object-oriented or more specifically component-based modeling. It provided, like many EMFs, a core Application Programming Interface (API) that featured framework data types, base classes, support classes for time/space management, process management, data manipulation and I/O in a library that the model used via inheritance, interface implementation, and plain calls.

The evolution of the OMS3 framework design is very much related to the design of model components, and is also driven by usage experience:

- OMS framework versions 1.x offered traditional library classes that a model or component had to either subclass or instantiate directly. The framework defined a limited set of data types that could be exchanged between model components. This was a simple approach, however, it limited modelers in their ability to use and share custom data types. For legacy code integration, this type of constraint is undesirable. Moreover, needed built-in flexibility for unit, type, and data transformation resulted in complex and large amounts of internal framework code. In addition, since model components had to be subclasses of the framework, blending into existing class hierarchies was impossible and a delegator/wrapper for existing components was always needed.

```java
/** Elevation.
 * @unit ft
 * @access read
 **/
private OMSDouble elevation;
```

- OMS 2.x simplified component design by allowing a model/component to use only interfaces instead of classes, i.e., no framework interface implementation was accessible from within the modeling component. As a result, the modeling component became more lightweight. This “design-by-contract” implementation allowed OMS 2.x to offer variants of framework data types with variable implementations for unit conversion, remote data access, data transformations, etc. while being fully transparent to the component. Switching to interfaces improved the overall quality and robustness of the framework and resulted in reduced API size. However, the data types being supported by the framework were still a fixed set:

```java
@Description("Elevation")
private Attribute.Double elevation;
```

For any OMS version prior to 3.0, framework data types for component data transfer objects had to be introduced in addition to the existing Java language counterparts. This represented a redundancy for types such as `double` and `Double` by providing an `OMSDouble` class implementation (OMS1.x) or an `Attribute.Double` (OMS2.x)

OMS 3.x departs from the traditional API-based framework approach for component design in favor of a more lightweight, non-invasive implementation. Software developers in general and model developers specifically, are in favor of lightweight frameworks in contrast to heavyweight frameworks. Heavyweight frameworks, e.g., traditional object-oriented frameworks, provide developers with an API that is often large and developers typically spend considerable time becoming familiar with framework APIs before writing model code. In contrast, lightweight frameworks offer functionality to the developer using a variety of techniques aimed at reducing the API’s overall size and developer dependence.
on the API. Programming language annotations which capture metadata are used to identify specific points in the model code where framework functionality is integrated. The following section introduces the component design and layout in OMS3.

2. MODELING COMPONENTS ARE PLAIN OBJECTS

Like in other modeling frameworks, an OMS3 modeling component implements a domain specific simulation concept as software code. The term component refers to a concept in software engineering which extends the reusability of code from the source level to the binary executable. Components are context-independent, both in the conceptual and technical domain. They represent self-contained software units that are separated from the surrounding framework environment.

OMS3 provides a radically simplified approach for model component design. It allows Plain Objects (Plain Old Java Objects) or POJOs that are annotated to be used within the framework. Annotated POJOs are easy to create since they are normal classes enriched with Java language level annotations. They provide the framework with hints to use the component as a model building block. Technologies that enable this type of framework design include:

- **Runtime introspection** on object and class structure, fields, methods, and classes - for exploring the internals of a component and finding framework entry points for data flow and execution.

- **Language level annotations** on classes, methods and fields – for describing dataflow fields and tagging IEF (Init/Execute/Finalize) methods.

- **Reflective access** on object fields and **reflective method invocation** – for component-to-component data transfer and indirect component method execution.

Any language or platform that supports the above features, such as Java or C#, should be sufficient to implement this type of architecture.

Listing 1 shows a simple component for lookup table computation in the OMS3/J2K-S watershed model (Ascough II et al., 2010). All annotations start with an (@) symbol.

```java
package climate;
import oms3.annotations.*;
import static oms3.annotations.Role.*;
@Author
  (name="Peter Krause, Sven Kralisch")
@Description
  ("Calculates land use state variables")
@Keywords
  ("I/O")
@SourceInfo
  ("$HeadURL: http://svn.javaforge.com/svn/oms/branches/oms3.prj.ceap/src/
climate/CalcLanduseStateVars.java $")
@VersionInfo
  ("$Id: CalcLanduseStateVars.java 1050 2010-03-08 18:03:03Z ascough $")
@License
  ("http://www.gnu.org/licenses/gpl-2.0.html")
@Status
  (Status.TESTED)
public class CalcLanduseStateVars {
  @Description("Attribute Elevation")
  @In public double elevation;

  @Description("Array of state variables LAI ")
  @In public double[] LAI;

  @Description("effHeight")
  @In public double[] effHeight;
```

1188
Annotations have the following features in OMS3:

- OMS3 package dependencies only exist for annotations (`oms3.annotations.*`).
  No API calls are needed (“Inversion of Control”).
- Annotations exist to document the whole component; single fields have them as well for data flow specification and documentation.
- A component adheres to the IEF cycle by tagging methods with the corresponding annotations. The computational method of this component is “tagged” with the `@Execute` annotation, the name of the method is not significant.
- Data flow indications are provided by using `@In` and `@Out` annotations.
- No explicit marshalling or un-marshalling of component variables is needed, i.e., an assignment is sufficient to pass them on to the receiving component.

Annotations provide a more integrated and context safe way for adding modeling specific metadata to code such as units, ranges, etc. Since there is API support within the Java platform, they are easier to comprehend, manage, and process that other metadata representation, for example XML.

Another core design aspect of components in OMS3 is the support of multithreading from the ground up. Since almost every new computer has multi-core processing, the inherent support of multiple execution threads via the modeling framework is required develop and deploy models which scale with processing resources.

Even for simple model applications, each OMS3 component is executed in its own separate thread which is managed by the framework runtime. Thread communication happens through data flow from the `@Out` field of one component to the `@In` field of another component. The component will execute if all inputs are present and satisfied. This is accomplished by native language synchronization features using `wait()/notify()`. OMS3 internally only mediates data flow in a producer/consumer-like synchronization pattern, and also protects itself from dead-lock situations caused by an incorrect Out/In setup. This ultimately leads to implicit parallelism at the component level, i.e., no explicit knowledge of parallelization mechanics and threading patterns is required for a model developer.

In addition to multi-threading, OMS3 also scales into cluster and cloud environments without any model recoding. Being able to implicitly scale models was also one of the major OMS3 refactoring accomplishments. Using Distributed Shared Objects (DSO) in Terracotta, geospatial models can share core model data structures (e.g., a hydrologic response unit, HRU) and process them in parallel within a model. Studies are being conducted to revaluate scalability for different models, model data sets, and setups on multi-core, cluster, and most interestingly in cloud infrastructures.
3. SIMULATIONS AS DOMAIN SPECIFIC LANGUAGES

OMS3 departs from the very GUI-centric design in version 2.2, by allowing more flexibility for integration into different development environments (IDEs) and general platform integration (e.g., JGrass and web-services stack). A flexible integration layer above the modeling components is provided by leveraging the power of a Domain Specific Language (DSL) for modeling and simulations. A DSL, in contrast to general purpose programming languages, is a programming language or specification language dedicated to a particular problem domain, a particular problem representation technique, and/or a particular solution technique. DSLs have the task of providing data (configuration or otherwise) to a program and let users write business rules for a particular task. This is usually motivated by the desire to allow coding without actually promoting it. OMS3 takes advantage of the builder design-pattern DSL, as provided by the Groovy programming language. This newly developed “Simulation DSL” allows the creation and configuration of simulations for OMS3; however, the Simulation DSL is not inherently bound to the framework.

What is a simulation in the OMS3 context? A simulation defines all the resources that are needed to run a model for a given purpose. A basic simulation in OMM3 consists of: (i) the model and component executables, (ii) model-specific parameter and other (e.g., climate) input data in files or databases, (iii) some strategy for handling output, and (iv) a method to evaluate model performance with simple graphing/plotting or formal evaluation statistics. There might be additional information and resources required if the simulation includes parameter estimation, sensitivity analysis, or uncertainty analysis.

Listing 2 shows a simulation using the Precipitation Runoff Modeling System (PRMS, Leavesley et al., 2006) Java-based model (PRMSHdJh) used for USDA-Natural Resource Conservation Service (NRCS) water supply forecasting in the western United States. The PRMS model has been set up for parameter estimation using the USGS Luca method. OMS3/luca is a multiple-objective, stepwise, automated procedure for model calibration that uses the Shuffled Complex Evolution global search algorithm to calibrate any OMS3-based model in multiple rounds and calibration steps. As shown in Listing 2, in addition to the standard simulation elements such as outputstrategy, resource, model, output, parameter, etc., a Luca DSL simulation (executable within OMS3) defines additional elements for the calibration parameter, model parameter bounds for each step, or objective function type. Besides luca, there are other OMS3 DSL simulation types, such as fast (FAST sensitivity analysis), dds (Dynamic Dimensioned Search parameter estimation), or esp (Ensemble Streamflow Prediction).

```groovy
/* Luca calibration. */

luca(name: "EFC-luca") {
    // workspace directory
def work = System.getProperty("oms3.work");

    // define output strategy: output base dir and
    // the strategy NUMBERED|SIMPLE|DATE
    outputstrategy(dir: "$work/output", scheme:NUMBERED)

    // for class loading: model location
    resource "$work/dist/*.jar"

    // define model
    model(classname:"model.PrmsDdJh") {
        // parameter
        parameter(file:"$work/data/efc/params_lucatest.csv") {
            inputFile "$work/data/efc/data_lucatest.csv"
            outFile "out.csv"
            sumFile "basinsum.csv"
            out "summary.txt"
            startTime "1980-10-01"
            endTime "1984-09-30"
        }
```

/* Luca calibration. */

```groovy
luca(name: "EFC-luca") {
    // workspace directory
def work = System.getProperty("oms3.work");

    // define output strategy: output base dir and
    // the strategy NUMBERED|SIMPLE|DATE
    outputstrategy(dir: "$work/output", scheme:NUMBERED)

    // for class loading: model location
    resource "$work/dist/*.jar"

    // define model
    model(classname:"model.PrmsDdJh") {
        // parameter
        parameter(file:"$work/data/efc/params_lucatest.csv") {
            inputFile "$work/data/efc/data_lucatest.csv"
            outFile "out.csv"
            sumFile "basinsum.csv"
            out "summary.txt"
            startTime "1980-10-01"
            endTime "1984-09-30"
        }
```
Listing 2. Luca DSL parameter estimation example in OMS3.

Our experience in developing and working with simulation DSLs has shown that they are easy to adjust to new simulation types (e.g., parameter estimation or uncertainty analysis) and provide the model user with a high degree of freedom in setting up complex simulations (e.g., batch processing of multiple watersheds for stream flow or water quality prediction). DSLs look very much descriptive (and in fact they are), and can also contain any Java language statements. This type of behavior makes them very attractive for flexible integration of simulations and superior over “data-only” and static representations such as XML.

4. OMS3 MODEL APPLICATIONS

Using OMS3, a number of science model implementations and applications are currently being tested and applied. All the models are open source and available on the OMS Javaforge project site (http://oms.javaforge.com).

4.1. OMS3/J2K-S Watershed Model

The OMS3/J2K-S watershed model is based on the European J2K-S model (Krause, 2002; Krause et al., 2006). OMS3/J2K-S (Ascough II et al., 2010) is a modular, spatially distributed hydrological system which implements hydrological processes as encapsulated process components. OMS3/J2K-S operates at various temporal and spatial aggregation levels throughout the watershed. For example, runoff is generated at the HRU level with subsequent calculation of runoff concentration processes (through a lateral routing scheme) and flood routing in the stream channel network. Research is currently being performed within the USDA-Agricultural Research (ARS) to develop, improve, and evaluate this model for selected watersheds within the Conservation Effects Assessment Program (CEAP) initiative.

4.2 NRCS Water Supply Forecasting

Seasonal water supply forecasts are an important function of the NRCS National Water and Climate center (NWCC). The forecasts are produced in cooperation with the National
Weather Service, are developed for hundreds of basins in the western United States, and are used by the agricultural community to optimize water use during the irrigation season. The NWCC has historically developed seasonal, regression-equation based forecasts of estimated seasonal stream flow volume. To address the agricultural community’s requests for more information on the volume and timing of water availability and to improve forecast accuracy, the NWCC is now developing the capability to use distributed-parameter, physical process hydrologic models to provide forecasts of daily, weekly, as well as seasonal stream flow using an Ensemble Streamflow Prediction (ESP) methodology. The primary objective of this model-based forecast effort is the development and implementation of an OMS3 PRMS-based model family (and associated methods and tools) to enable the provision of timely, improved, and more frequently updated water supply forecasts.

4.3. Conservation Delivery Streamlining Initiative (CDSI)

USDA- NRCS field consultants currently use an array of analytical tools when providing technical assistance, including the Web Soil Survey, Revised Universal Soil Loss Equation, Wind Erosion Equation, Nutrition Balance Analyzer, Soil Quality Index, Pesticide Screening Tool, Phosphorus Index, Energy Estimators, Cost and Returns Estimator, among several others. They will increasingly use more comprehensive process-based modeling tools, such as the Agricultural Policy Extender (APEX) and the Agricultural Nonpoint Source (annAGNPS) models, or at least the technology contained within them, for on-farm system analysis. Unfortunately, each model comes with its own data provisioning requirements, unique user interface, and processing requirements. Currently, field consultants have hit a wall of complexity and resource constraints, and the existing tools are not used to their full potential (Carlson, 2010).

To remedy this problem, the NRCS has initiated a Conservation Delivery Streamlining Initiative (CDSI) to integrate technology components with the workflows of the field consultant. CDSI will provide the framework and common user interface for the field consultant. The USDA-ARS is leading the development of models in OMS3 to integrate science components across multiple models and tools into various model bases, one of which will integrate with the CDSI framework. The purpose of the OMS-CDSI model base is to deliver science deployed as services available to the CDSI workflow.

4.4 JGrass – Horton Machine and NewAge

Recently, OMS3 was used as the modeling engine for JGrass through a collaboration effort between HydroloGIS, Colorado State University, and the University of Trento, Italy. JGrass is an open source front end for the Grass Geographic Information System. It is based on the uDig framework, which is maintained by HydroloGIS and CUDAM (University Centre for Advanced Studies on Hydrogeological Risk in Mountain Areas, University of Trento, Italy). Grass has a long history, emerging from a federal agency-sponsored project into an open source project supported by a community of researchers, consultants and the Open Source Geospatial Foundation (OSGeo). The OMS3 runtime engine was brought into JGrass to provide for the execution of any OMS3 model within this environment. The core components of the Horton Machine, a hydro-geomorphological tool box, were implemented under OMS3. Example Horton Machine components include a peak flow discharge model, watershed stream network delineation model, and an energy balance model. All components were integrated with core GIS processing through “Sprint”-like development sessions. The fact that data type handling for modeling components in OMS3 is open to Open Geospatial Consortium (OGC) types for raster and vector data allowed a straightforward and efficient implementation. In addition to the Horton Machine, the NewAge semi-distributed hydrological model was also ported into OMS3. NewAge consists of new conceptualizations where geographic features and modeling are mixed in an innovative way. Besides modeling the whole hydrological cycle, NewAge is also able to take into account human artifacts like reservoirs, intake and outtakes.

5. CONCLUSION
OMS3 is a lightweight and non-invasive modeling framework for component-based model and simulation development on multiple platforms. The simplified structure for components allows rapid implementation of new models and an easier adaptation of existing models and components. Recent studies have shown that this type of approach leads to models with less overhead and a more intuitive design. By fully embracing language annotations for modeling metadata in favor of traditional APIs, OMS3-based models keep their identity outside of use within the modeling framework. Annotations also enable multi-purposing of components which is hard to accomplish with a traditional API design. In OMS3, annotations provide for component connectivity, data transformation, unit conversion, and automated documentation model generation (Docbook5+). Agencies such as USDA-NRCS use those annotations for creating traceable simulation audits based on secure hash algorithm (SHA 256).

OMS3 also introduces implicit multithreading and an extensible and lightweight layer for simulation description that is expressed as a simulation DSL based on the Groovy framework. DSL elements are simple to define and use for basic model applications, or for more complex setups for calibration, sensitivity analysis, etc. This type of “programmable” configuration that eliminated core programming language “noise” turned out to be an extremely useful tool for current projects to support aspects such as automated, distributed setup of multiple batch models runs, integration into the open source GIS JGrass console, or IDE integration to name a few.

Ongoing work with several modeling applications in OMS3 has shown the efficacy and efficiency of a non-invasive framework approach. The OMS3 development goal is to provide features to the modeler to make it easy to create contemporary, inter-operable, scalable and lightweight models that take full advantage of computing resources, data stores, and infrastructure opportunities while keeping things simple and intuitive. Additional information about OMS3 and associated models applications can be found on http://oms.javaforge.com.

REFERENCES


TaToo: tagging environmental resources on the web by semantic annotations

Andrea E. Rizzoli\textsuperscript{a}, Gerald Schimak\textsuperscript{b}, Marcello Donatelli\textsuperscript{c}, Jiri Hrebicek\textsuperscript{d}, Giuseppe Avellino\textsuperscript{e}, Jose Lorenzo Mon\textsuperscript{f}, Ioannis Athanasiadis\textsuperscript{1}

\textsuperscript{a} IDSIA, Lugano, Switzerland, (andrea, ioannis @ idsia.ch)
\textsuperscript{b} AIT Austrian Institute of Technology, Austria (gerald.schimak@ait.ac.at)
\textsuperscript{c} JRC, Ispra, Italy (marcello.donatelli@jrc.it)
\textsuperscript{d} Masaryk University, Brno, Czech Republic
\textsuperscript{e} Elsag Datamat, Italy
\textsuperscript{f} Atos Origin, Spain

Abstract: The web is rapidly evolving and its traditional role of repository of static information is changing into a hub for collaboration among people. Web resources tend to become more and more complex, and to offer services that include access to remote databases, and computational power. All of this becomes very interesting not only for the common user, but especially for scientists and researchers which actually see their computers "disappear" into the web “cloud”, getting back an unprecedented access to services and computational resources.

Yet, to exploit these new facilities new tools are needed. The TaToo project aims at exploiting a common practice among web user: search, discovery and tagging of interesting resources. The practice of tagging allows user groups to label and classify resources enabling aggregators to display the most relevant ones according to the context. TaToo aims to take the core idea of tagging and adding the ability to add valuable information in the form of semantic annotations, thus facilitating future usage and discovery, and kicking off a beneficial cycle of information enrichment. Thus, the production of semantic meta information will improve the discovery process, but also interpretation in a larger sense (verification that its the information I was looking for, assessment of usefulness for a given situation, understanding of how to use the information correctly etc.).

Keywords: semantic annotation; semantic tagging; model search and discovery; web services; environmental information enrichment.

1. Introduction

Technological progress has been constantly providing new tools and techniques for environmental scientists. If we look back at the not so distant past, geologists, geographers, hydrologists, ecologists and other environmental scientists they all had to manually and painstakingly collect data by surveys, which were expensive, labour intensive, and required lots of time. Data were then stored in paper based repositories and archives, and processing data was also limited by the availability of data. Perhaps the most brilliant use of spatial data and analytic reasoning was the discovery of the source of a cholera epidemic in
London by Dr John Snow. In 1854 he identified the water pump of Broad Street in London by plotting deaths on a map of the borough, as represented in Figure 1.

![Figure 1](image)

**Figure 1.** A section of the original map published by Dr Snow in 1854. The black bars represent deaths at a given house number.

This type of representation would require only a few minutes work using a modern GIS tool, but it took Dr Snow several days of work to put together the above map, and a considerable effort in collecting the data.

Nowadays data can be automatically collected by remote sensing or by sensor networks; they are stored in information systems relying on advanced database techniques and data storage facilities. Data representation is facilitated by desktop Geographical Information Systems (e.g. ArcGis, MapWindows) or even web-based GIS-like applications (Google Maps, Bing Maps). Data processing is provided by sophisticated and complex models, supported by advanced computer architectures, exploiting distributed and parallel processing.

Given the state of things, the outlook for the environmental scientist should be particularly bright and rosy: vast amounts of data to process, a great variety of models available to process and elaborate data, and powerful visualisation tools. While we must regard with awe what we have achieved in the past 30/40 years, we should also be wary of the threat posed by having access to too much information, which we cannot neither discern nor make sense of.

Scientists and researchers have been aware of such a threat for a long time, and various approaches and mitigation measures have been devised. We can enumerate a few: data catalogues, metadata, model bases and model repositories (such as EIONET, UN FAO, etc.). At the same time, the pressure towards the integration and exchange of data pushed towards the creation of standards for data representation, such as HDF (hierarchical data format), the OGC standards for GIS data, and standards for model integration, such as OpenMI as described in Gregersen [2007].
Standards and metadata are part of the cure, but they are still too little to front the exponential increase in data availability provided by earth observation initiatives such as INSPIRE⁴, Google Earth², OpenTopography³, GEOSS⁴ and many others.

We claim that the expressivity of glossaries, dictionaries, thesauri and schemata is too limited for the demands that we expect to be posed to environmental information systems in the near future.

The Internet is growing in a non-coordinated manner, where different groups continuously publish and update information, adopting a variety of standards, according to the specific domain of interest: from agriculture to ecology, from groundwater to climate change. This unconstrained and unregulated growth has proven to be very successful, as more information is made available, even more is being added, in a virtuous cycle of information accrual. At the same time, modern search engines make looking for information rather easy, and despite some studies reprieve the presence of non-relevant hits in the result of most queries (e.g. Gordon and Pathak [1999]), we personally judge the overall performance more than satisfactory for most users (we don’t have data to support our claim, but if search engines were so bad, probably people would not use them...).

Yet, searching and discovering information requires a good deal of expertise and pre-existing knowledge. That may not be a problem when the user is searching its own domain of expertise, but what happens when the user is trying to gather environmental information on a trans-disciplinary study? Or when preparing and integrated assessment study? These type of research and development efforts require the detailed knowledge of multiple domains, from economy to ecology, from hydrology to social sciences. In our experience, we have seen situations where different groups of scientists labelled the same concepts with different terms, and also labelled similar (but distinct) concepts with the same names (Athanasiadis et al. [2009], Janssen et al. [2009]).

---

1 http://inspire.jrc.ec.europa.eu
2 http://earth.google.com
3 http://www.opentopography.org
4 http://www.earthobservations.org/geoss.shtml
To overcome this type of problems we need to add rich semantics, possibly axiom-based semantics, to our environmental resources, thus increasing the expressivity of information, but, at the same time, also increasing the complexity required to convey it, as shown in Figure 2 in Villa et al. [2009].

The recently started (beginning of 2010) TaToo project tries to fill this information and discovery gap, by providing a way to semantically annotate environmental resources on the web. The project idea is strongly inspired by existing social bookmarking initiatives, such as Delicious, reddit, StumbleUpon, Digg etc. Yet, TaToo aims to let user use semantics in their annotations, by accessing shared ontologies, thus enabling inference engines to process information and discover new facts and new relationships that are not explicitly stated in the body of knowledge.

In the remainder of this paper we will first review the state of the art in semantic annotations and tagging, then we will present our vision of how TaToo should work and operate, and we also describe the preliminary draft of the enabling software architecture. Finally we introduce some test cases, which will be developed during the project, and finally we will conclude by outlining our expectations for the future of this project, that is expected to end in three years time.

2. The semantic web and semantic annotations

The Semantic Web vision aims to provide mechanisms to augment the Web content with semantic annotations and retrieve concrete information contained in the Web, not only references to pages or resources where the requested information could be contained within. Semantic Web technologies have matured considerably in the last years and standardisation efforts have created standards for data exchange (Resource Description Format (RDF\(^5\))), ontology languages (e.g. Web Ontology language (OWL\(^6\))) and it is progressing in the standardisation of further components of the Semantic Web such as SAWSDL\(^7\), a language allowing Web services to be described with semantics; to be more amenable to automated processing on behalf of the users. Another relevant component is GRDDL/RDFa\(^8\): a way to embed RDF-based annotation into existing Web pages.

While the progress in standards for the Semantic Web is highly relevant to support the annotation of unstructured information contained in web pages, we are particularly interested in the annotation of web services (see the work of Hilbring and Usländer, [2006]), as most environmental resources will be made available under that form. Everything, from a model run, to a database query, or a GIS operation, can be served as a web service, and there are impressive efforts towards the standardisation of web services in the environmental domain, such as OGC’s OpenGIS web service common standard\(^9\).

For these purposes we refer to the WSMO\(^10\) conceptual framework for the description of Web Services currently realised in the WSML family of languages and the reference implementation WSMX. The WSMO Lite language realises the WSMO concepts using

\(^5\) [http://www.w3.org/RDF/](http://www.w3.org/RDF/)

\(^6\) [http://www.w3.org/TR/owl-features/](http://www.w3.org/TR/owl-features/)

\(^7\) [http://www.w3.org/2002/ws/sawSDL/](http://www.w3.org/2002/ws/sawSDL/)

\(^8\) [http://www.w3.org/2004/01/rdfx/spec](http://www.w3.org/2004/01/rdfx/spec)

\(^9\) [http://www.opengeospatial.org/standards/common](http://www.opengeospatial.org/standards/common)

\(^10\) [http://www.wsmo.org](http://www.wsmo.org)
RDF/S and SAWSDL. WSMO Lite can thus be used as the basis for a standard W3C ontology for modelling Web services.

Of interest are also DAML-S, OWL-S that standardise semantic web services OWL-S (formerly DAML-S) is a semantic mark-up language for Web services. It is an academic ontology proposal, submitted to the W3C, for describing Web services based on OWL. It provides a core set of constructs (machine-understandable) to provide descriptions of Web services properties and capabilities. Through OWL-S support, users or agents acting on behalf of the users (including software agents) are able to discover, invoke, compose, and monitor Web resources. Even if OWL-S has not reached yet the status of a standard, it is quite stable and adopted. There is still a good amount of research going on that involves OWL-S mainly because it is able to address workflow management issues.

While a thorough state of the art analysis cannot be included here for sake of the limited available space, we can simply state that a great number of projects and initiatives are working towards the semantic annotation of web services, such as SWWS\(^{11}\), DIP\(^{12}\), SOA4ALL\(^{13}\), while others are focusing more on the discovery and retrieval of annotated content, such as INFRAWEBS\(^{14}\), OWL-S Matchmaker (Sykara et al., [2003]), BREIN\(^{15}\) and again SO4ALL. The aim of TaToo is to join these lines of work and to contribute with the development of environmental-specific applications of the above-mentioned research topics and technologies.

3. The vision

Despite the great amount of work and resources currently deployed in the field of semantic annotation of web resources, there are some major hurdles to be overcome to make the TaToo vision become a reality.

TaToo is expected to work along the lines of one of those social tagging and bookmarking website. Here we focus on two major use-cases: finding and annotating a resource, and searching for and discovering an annotated resource.

In the first case, the user stumbles on an interesting resource during his/her work. Here we assume that the resource is a web service described in a web page. The user has simply to feed the URL of the web page to the TaToo server application. This can be done by simply dragging and dropping the URL in a modern browser such as Firefox. The TaToo server application recognises the URL and starts processing the page, automatically extracting the information regarding the web service and processing the text in the webpage. The TaToo server application then generates a web page where the various elements of the original webpage are presented to the user and offered for semantic tagging. The process is therefore a mix of automated semantic annotation, and manual annotation. The user-friendliness of the interface will be therefore a critical element for the success of the application.

In the second case, the user accesses the TaToo server to search for and discover environmental resources which have been semantically annotated. The advantage is the ability to come across the limitations imposed by specific domain jargons and semantic

\(^{11}\)http://swws.semanticweb.org
\(^{12}\)http://dip.semanticweb.org
\(^{13}\)http://www.soa4all.eu
\(^{14}\)http://www.infraWebs.eu
\(^{15}\)http://www.eu-brein.com
ambiguity. In section 5 we describe some case studies where this feature will be exploited. Also, in a possible future scenario, third-party web-services will be able to use TaToo discovery services to automatically chain web-services to answer complex and structured queries, requiring the integrated runs of multiple environmental resources.

Figure 3. Use cases for the TaToo vision.

4. The proposed software architecture

TaToo aims to deliver a web based system centred on a three main elements:

- a **clearinghouse**, which plays the role of organising the semantic information on environmental resources. The clearinghouse contains a list of semantic annotations of environmental resources, referring to the original content as available on the World Wide Web.
- a **semantic processor**, a core component of the TaToo system, since it uses a set of (pluggable) environmental ontologies to provide semantic services to the tagging tools of TaToo.

Figure 4. The TaToo overall architecture.
· a set of tagging tools (TaToos), which offer services such as tagging new environmental resources, including quality and uncertainty information, searching for and discovering tagged environmental resources, validating the results of a search.

The server side interacts with web based client components. These components can be interactively combined in order to enable human users to create, modify, delete and update TaToos. The composition is open, so that third party will be able to exploit the components to deliver their specifically targeted services. A schematic representation of the architecture is represented in Figure 4.

5. Expected results: the case studies

TaToo plans to validate the usability of its approach through the implementation of three different scenarios. All three scenarios are embedded in highly complex environmental domains and are therefore mainly addressed to domain expert groups and communities as well as to technically skilled users. The scenarios are tackling the following environmental domains: climate change, agriculture, and anthropogenic impacts of pollution.

In the climate change case study the aim is to be able to identify model regions, where the current climate matches with the expected future climate of the source region of interest. We call such region pairs with similar climate conditions (at different times) “Climate Twins”. A web-based “climate twins” exploration tool will identify those Climate Twins, where source grid-cell’s values representing future climate show high similarity with the current climate grid. To find climatic coincidence seems to be a simple exercise, but the accuracy and applicability of the similarity identification depends very much on the selection of climate indicators and uncertainty ranges. The TaToo platform can provide with tools to facilitate an improved, user-focused climate change resource search, through which end-users will be able to add tags and comment existing resources, reuse tags of other users, and eventually discover and retrieve climate twin-region data, through semantic rich, spatially explicit, user-tailored querying.

In the agro-environmental case study we will work in collaboration with the AGRI4CAST action of the Joint Research Centre focuses on the European Commission Crop Yield Forecasting System aiming at providing accurate and timely crop yield forecasts and crop production biomass. AGRI4CAST gets increasing requests for analyses to be run against the weather and soil database which require either new or modified modelling capabilities with respect to the set of models available in the operational system. To achieve this, software implementations of Crop Forecasting System model components target the objective of easy composition, extension and re-use. Though detailed model and software documentation is available, along with scientific papers and reports describing the application of the models, still the discovery of appropriate models to-be-employed for on-demand studies is a monotonous task that requires significant human expert efforts. TaToo will be put to the test as a tool to support the proper annotation of resources by defining attributes, such as description, maximum, minimum and default values, units, and URL. Then its search and discovery capabilities will be put to the test to find alternative modelling solutions, given that each component can make available alternate options for estimating/generating variables.

The anthropogenic impact of pollution case study will enable the synthesis of existing (air) pollution monitoring databases, with epidemiological data is required for identifying the effects of pollution on human health (anthropogenic impact). This task requires new, rich, data discovery capabilities within the bodies of knowledge available. Proper use of these
data requires contextual enhancements, which TaToo will deliver through tagging and enhanced meta-data information description embedded into the appropriate semantic environment.

6. Conclusions

In this paper we have introduced the aims and the vision of the TaToo project, a recently started EU-funded project, which aims at providing a collaborative platform for the semantic enrichment of environmental resources on the web. The main challenges of the TaToo project are the provision of an appealing user interface for the semantic annotation of environmental resources, more specifically web services, and the development of a set of tools to provide a preliminary semantic analysis of the content of web resources, with the ability to access different published ontologies that describe the available knowledge basis. Finally, the most critical factor will be the ability to involve the scientific and research community, which are expected to be the prime users of the platform. This is the main reason that drove us to prepare this work, in order to get an early involvement of the user community and to raise awareness about the scope and aims of our future work.

7. Acknowledgement

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement nr. 247893.

REFERENCES


The use of ontologies in peer reviews of Integrated Assessment Models

M. de Vosa, N. Koenderinkb, B. van Ruijvena, J. Topbc

a Netherlands Environmental Assessment Agency (PBL), P.O. Box 303, 3720 AH Bilthoven, The Netherlands [martine.devos@pbl.nl]
b Wageningen UR Food and Biobased Research, Wageningen, The Netherlands
c VU University Amsterdam, Amsterdam, The Netherlands

Abstract

Integrated assessment models are used to analyze global change issues and they allow a better understanding of our complex environment. It is crucial to be able to relate these models to their scientific basis, both for interpretation and validation purposes. Current model evaluation procedures focus on model behavior analysis; the conceptual knowledge and assumptions embedded in integrated assessment models are hardly tested. As such, current model evaluation procedures do not contribute to the understanding of the structure of the models and the selected mechanisms and assumptions. We submit that evaluation of the scientific basis of integrated assessment models should follow the standard procedures for evaluation of scientific theories, which implies that these models should be subjected to critical peer reviews. However, much knowledge is hidden in the source code and therefore not accessible to peers. In this paper we propose to use ontologies – explicit specifications of shared conceptual knowledge – to represent the knowledge encoded in integrated assessment models in a clear and transparent way. We show the proposed peer review evaluation procedure in a case study concerning a system-dynamics model on residential energy use in India. We found that the ontology helped peers to obtain more information on the model and to gain more insight in its structure. However, a better balance between different types of model documentation and explicit links between them are needed to improve the understanding by the peers. We believe that ontologies can be exploited further in a computational sense in order to achieve model transparency.

Keywords: ontology; integrated assessment; model evaluation; conceptual model

1. Introduction

Global change issues are analyzed using integrated assessment models, numerical models that integrate multiple (natural and socio-economic) academic disciplines. Integrated assessment models are often used for forecasting and management purposes, but their main value is heuristic, i.e. they allow a better understanding of complex phenomena [Oreskes et al., 1994, Rykiel Jr, 1996, Rotmans and Dowlatabadi, 1998, Risbey et al., 1996]. To allow proper interpretation of such a model, understanding the scientific knowledge captured by it is crucial. Unfortunately this knowledge is often poorly accessible to the users of the model, as it is hidden in the code or presented indirectly in publications and reports. It is clear that lack of
transparency can be detrimental for both policy makers and environmental scientists [Jakeman et al., 2006].

The need for validation of integrated assessment models is widely recognized. Integrated assessment models incorporate knowledge from a broad range of mono-disciplinary theories. In developing these models, modelers make choices on which parts of these theories to include in their models, which conditions to assume and how to translate all of these in model source code [Jakeman et al., 2006]. It is therefore important to test the conceptual knowledge and assumptions that are embodied in these models in addition to behavioral validation [Nguyen et al., 2007a, Rotmans and Dowlatabadi, 1998, Parker et al., 2002, Jakeman et al., 2006]. The literature on model validation is abundant and a confusing variety exists of used terminologies and methodologies [Nguyen and de Kok, 2007b]. But as current model evaluation procedures focus on quantitative tests for operational validation [Nguyen et al., 2007a], the conceptual knowledge and assumptions in integrated assessment models are hardly tested [Refsgaard and Henriksen, 2004, Nguyen et al., 2007a, Oreskes et al., 1994, Rykiel Jr, 1996]. As such, current model evaluation procedures do not contribute to the understanding of the structure and mechanisms of the modeled systems nor do they make integrated assessment models more reliable.

We believe that the knowledge embedded in integrated assessment models must be tested according to the standard procedures for confirmation or falsification of scientific theories, which implies that it should be subjected to critical peer reviews [Refsgaard and Henriksen, 2004]. Furthermore we believe that integrated assessment models can not be evaluated based on their scientific publications alone. These publications describe the underlying theories of a model and do not consider the knowledge that is actually captured in the model. A large part of the conceptual knowledge and assumptions is hidden in the model source code and has not been made explicit in a way that it can be understood and used without the modeler’s mediation [Villa et al., 2009]. It is important that the knowledge captured in these models is represented in a clear and transparent way so that peers are able to assess its quality and appropriateness.

In this paper we propose to use ontologies [Gruber 1993] – explicit specifications of shared conceptual knowledge – to represent the knowledge encoded in integrated assessment models in a clear and transparent way. Whereas traditionally the focus is on mathematical variables and equations, ontologies describe the concepts, their attributes and relations. The concepts represent the factors that have been taken into account in the model, thereby providing a link between the human understanding of the model context and the variables used in the source code. This is in the spirit of Villa et al. [2009], who argue for semantically aware environmental modeling by using an explicit and meaningful representation of scientific knowledge. In the present study we make a first step by employing the textual representation of an ontology as a means to complement traditional documentation of system-dynamics models. However, we believe that the formal character of ontologies can be exploited further in terms of computational support for enhanced model transparency. The constructed ontology provides an additional layer on top of an Integrated Assessment model.

In this paper, we assess the usefulness of such an ontological layer in a case study on one of the modules in the Residential Energy Model India (REMI), a system-dynamics model on energy use by Indian households. This paper is organized as follows. In the next section we explain how we constructed the ontology for our case study module. We also describe the experimental set up for testing the possible benefits of an ontology in a peer review. In Section 3 we present the results of the peer review experiment and in Section 4 we reflect on the process of constructing the ontology and its contribution to transparency of the module. We address our main findings in Section 5 and finally discuss novel scenarios for transparent modeling.

2. Materials and methods
2.1 Formalizing knowledge in ontologies

The knowledge in integrated assessment models is based on a conceptualization: an abstract and simplified representation of the real world. An ontology is an explicit formal specification of a conceptualization, which describes the entities in a knowledge domain and relations among them [Gruber 1993]. The development and use of ontologies originated from Artificial Intelligence but nowadays many disciplines use them to share and annotate information in their fields [Noy and McGuinness, 2001; Villa et al., 2009]. Besides sharing common understanding of domain knowledge, ontologies can be used to make domain assumptions explicit and to enable reuse and analysis of domain knowledge [Noy and McGuinness, 2001]. Ontologies are not unlike traditional (relational or object-oriented) data models and database schemas; however they are different in the sense that they are expressed in terms of a shared, system independent language. In this study we used the OWL (W3C Web Ontology Language) formalism to construct an ontology. As the OWL formalism can also be understood by computers, a future step could for example be to automatically check the consistency of the knowledge expressed in that ontology.

In an ontology the entities in a knowledge domain are called ‘concepts’. A concept in the domain of our case study model is for example Household. The concept Household does not refer to any particular household, but to the type of households that can be found in India. The actual meaning of the concept is defined through its ‘properties’ and ‘relations’. The concept Household is characterized by the fact that it has for instance properties stating the associated residential area and expenditure. The relationship between the concepts Household and Person is defined by the property ‘has person’; every household consists of a number of persons, defined by the concept Person.

Figure 1: Simplified representation of the flow chart of the CookWater module of the Residential Energy Model India [Van Ruijven, 2008]. Variable names in the flow chart are similar to those used in the source code.
Constructing an ontology for the Residential Energy Model India

The Residential Energy Model India [Van Ruijven, 2008] is a system-dynamics model that determines the energy use by Indian households. The model specifically addresses the socio-economic factors influencing energy uses. It includes knowledge on consumer behavior, public health and sustainable development. In this study we focus on one specific module in REMI on energy use for cooking and water heating. The REMI model has been developed using the integrated software environment M [De Bruin et al, 1996]. The model source code resembles the syntax of mathematics common in writing, describing the causal relations between model variables through difference and differential equations. Besides the source code, the model documentation consisted of a thesis chapter, describing the underlying theory of the REMI model and a flow chart (figure 1) giving a visual representation of the causal relations between the model variables.

We studied the model documentation to gain insight in the aim, structure and main processes in the module. To construct the ontology for this model we extracted the scientific concepts and relations hidden in the source code. We used the model source code to list and define all variables present in the module and grouped them according to subject. In consultation with the modeler we then defined the main concepts in the module and represented all module variables as properties of these concepts (figure 2). We did not use the variable names as used in the code for these properties but chose clear names to represent their function. From the thesis chapter we assumed several relationships between the concepts in the module, but these were difficult to identify in the source code. In consultation with the modeler we defined these relationships in our ontology, representing them as relations between the concepts. Finally we constructed a glossary that provided definitions of the terms used in the ontology diagram and linked those terms to variables in the REMI module (table 1).

Table 1: Glossary (example part) of terms in the ontology diagram and associated variables in the CookWater module of Residential Energy Model India [Van Ruijven, 2008]

<table>
<thead>
<tr>
<th>Term in Ontology</th>
<th>Definition</th>
<th>Associated model variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population group</td>
<td>Group within Indian population; there are 10 types of population groups according to different income level and residential area</td>
<td></td>
</tr>
<tr>
<td>has residential area</td>
<td>Area where people in a population group live in; can be either urban or rural</td>
<td></td>
</tr>
<tr>
<td>has expenditure</td>
<td>Expenditure level of a population group; can be either low, low-middle, middle, high-middle or high</td>
<td></td>
</tr>
<tr>
<td>has household</td>
<td>A population group consists of multiple households of type household</td>
<td></td>
</tr>
<tr>
<td>has nr of households</td>
<td>Number of households in a population group</td>
<td></td>
</tr>
<tr>
<td>has net yearly energy use for cooking + waterheating</td>
<td>Net yearly energy use of a population group for cooking and waterheating; per fueltype, in MJ</td>
<td></td>
</tr>
</tbody>
</table>

Application of the ontology in peer reviews

The aim of the constructed ontology was to give a clear and transparent representation of the knowledge captured in the REMI module, an intermediate between the source code and textual descriptions. We performed a peer review experiment as a proof of principle for ontological
model transparency, but also to receive qualitative feedback on our approach. We organized two peer review sessions with comparable peer groups consisting of four scientific researchers employed at the Netherlands Environmental Assessment Agency representing different disciplines, viz., economics, sustainable development, energy and public health. In both groups the modeling experience of peers ranged from researchers with no modeling experience to researchers who were familiar with the programming environment that had been used for the development of REMI. In session 1 experts were given only standard model documentation, which existed out of the thesis chapter describing the model, the flow chart (figure 2) and the source code. In session 2 experts were given the same standard documentation, plus the ontology diagram and the accompanying glossary.

After a short introduction on the aim and context of the module, we asked the experts to determine if the module consisted of the right elements to achieve its goal, or that elements should be added or deleted. During the sessions, which lasted 90 minutes, the experts were allowed to discuss these issues. They were not allowed to ask questions to the modeler or to the workshop supervisor. We observed the behavior of the experts during the sessions and registered group discussions on a voice recorder. At the end of each session we evaluated our experiences with the experts and the modeler.

![Figure 2: Ontology diagram (simplified version) of the CookWater module of Residential Energy Model India [Van Ruijven, 2008]](image)

3. Results

The following observations were made in the two groups. In group 1, the experts had access to the thesis, the flow chart and the source code. We noted that the thesis was the most frequently used source of information. Experts in group 1 questioned certain statements in the thesis documentation and stated that some information was lacking or could not be found. They assumed, for example, that “family size” and “distinction between urban and rural population groups” were important factors influencing residential energy use, but in the thesis they could not find whether these factors were included in the model. The experts indicated that it was
difficult to get an overview of all used factors in the model, and that it was therefore difficult to answer the evaluation question.

In the second group, all forms of documentation were frequently used: thesis, ontology diagram, glossary and source code. The experts also tended to start with the thesis as a source of explanation. In this group, the thesis documentation also appeared to lack certain information, but in about 50% of the cases the missing information was then retrieved from the ontology diagram and the glossary. Factors like “indoor air pollution” or “family size” were deemed to be important and could indeed be found in the ontology diagram. However, even with the ontology and the glossary given, the experts found it difficult to have a good overview of all factors used in the model, given the time available.

In both workshops experts showed a preference for the thesis as a source for model documentation. However, the thesis explains the theory behind the model, but not the knowledge as actually encoded in the model. Group 2 additionally used the ontology diagram and accompanying glossary, which are more true representations of the information included in the model. These documents helped the experts in group 2 to obtain information missing from the thesis chapter and to gain more insight in the structure of the model than group 1.

In both workshops we reflected on the experiment in a subsequent discussion with the modeler and the experts. All experts thought that 90 minutes was too short to get a good overview of the model. In group 1, experts found that they missed information as they did not have enough time to read the thesis documentation thoroughly. Experts in group 2 stated that they had too much information to choose from. The modeler observed that, in his opinion, the experts from both groups did not have a full overview of the main processes and assumptions in the module at the end of the sessions.

With regard to the provided model documentation, experts from group 1 indicated that they could not link terms in the thesis documentation to variables in the flow chart and source code as they all had different names. They also suggested that it would have been useful to have a schematic overview of the module that is less detailed and complicated than the flow chart (see Figure 2). Experts from group 2 found the ontology diagram and glossary useful, although they indicated that the definitions of the ontology terms in the glossary could have been stated more clearly.

4. Discussion

In our experiment the experts in both groups showed a preference for the thesis documentation as they assumed that the text in this document corresponded exactly to the information in the model. They also assumed that it was scientifically correct as it had been peer reviewed. Their confidence in the one-on-one correspondence of model and thesis was however too optimistic, since a thesis is not a living document and changes in the model do not lead to changes in the thesis.

The group discussion showed that the experts did not fully answer the evaluation question as they did not discuss the module at a suitable level of abstraction. As an explanation the experts mentioned lack of time, information overload and lack of recognizable connections between the different types of model documentation. As we aim for effective and efficient model evaluation procedures, it is not desirable to extend the duration of peer reviews. We believe that a more deliberate use of the different types of model documentation could enhance the experts’ understanding of the module. Moreover, links between the different types of documentation should be made more explicit. Ideally, model documentation gives a transparent representation of the knowledge captured in a model. Ontologies are useful in this process as they are
representations of the knowledge captured in models, thereby bridging the gap between source code read by computers and language read by humans. The ontology diagram as used in this experiment is a first step towards understanding a model in conceptual terms. We believe, however, that the use of ontologies to enhance model transparency is still in its early days.

The first step is to introduce ontologies as an additional conceptual layer of information for integrated assessment models. These systems-dynamics models are typically constructed around numerical variables that are related through mathematical relations representing causal dependency relations. However, a conceptual description explaining these concepts and relations is usually missing, even if careful model documentation is provided. An ontology provides a view that is orthogonal to the system dynamics view in the sense that it employs a linguistic rather than mathematical view. Our hypothesis is that this complementary (meta-level) perspective can contribute significantly to the much wanted model transparency in integrated assessment modeling.

Constructing an ontology for REMI was a time consuming activity for the knowledge modeler, as deducing concepts, properties and relationships from the list of module variables was not at all a trivial task. It turned out that the existing model documentation did not provide sufficient information to get a good overview of the structure and processes the module. In fact one could argue that good conceptual modeling practice [Refsgaard and Henriksen, 2004] requires proper data modeling already in the development phase. However, in practice this approach is not always followed due to time constraints and lack of modeling support [Jakeman et al., 2006]. Ontologies can help to fill this gap because they are formal representations that are amenable to computer processing. For example, model inconsistencies can be detected automatically. At present, no tools exists that assist systems-dynamics modelers in particular in building ontologies.

Another potential benefit of ontologies is that they are expressed in standardized languages. This allows world-wide exchange and reuse of (parts of) conceptual models through the web. In fact, ontologies have their highest impact on the world-wide web, turning it into a Semantic Web, or better the Web of Data. In the area of Integrated Assessment the quality and transparency of models would highly benefit from freely exchanged model data, concepts and mechanisms.

5. Conclusion and recommendations

The aim of our experiments was to assess the usefulness of ontologies as a link between the human understanding of an Integrated Assessment model in natural language and the knowledge captured in the source code of the model. In a peer reviewed model evaluation we compared the use of an ontology with the use standard model documentation only.

We observed that the ontology helped experts to obtain more information from the model and to gain more insight in the model structure. However, even with the ontology diagram, experts could not get a good overview of all used factors in the model and therefore had difficulties evaluating the model. We believe that a more deliberate use of the different sources and explicit links between different types of model documentation could enhance the experts’ understanding of a model. Ontologies are useful in this process, bridging the gap between source code read by computers and language read by humans. As stated above, they can enable a world-wide discussion between model experts, while staying close to the actual encoded models.

The ontology diagram as used in this study is a first step towards explaining a model in conceptual terms, complementary to the system-dynamics perspective. As a next step following our review experiments we plan to use the computational capabilities of ontologies by giving
experts access to the model documentation via a computer application. The core of this application is the ontology, which provides a formal description of the concepts, attributes, and relations between the concepts in the model. By hovering over or clicking on concepts in the ontology diagram, the experts are guided through corresponding elements of the model and model documentation. We expect that by giving the ontology a central position in the model evaluation session, experts are more likely to understand the model on a conceptual level.

References

Balci, O., Principles of simulation model validation, verification, and testing, *Transactions of the Society for Computer Simulation*, 14, 3-12, 1997


Towards model component reuse for the design of simulation models – a case study for ICZM

Jean-Luc de Kok, Guy Engelen and Joachim Maes

VITO NV - Flemish Institute for Technological Research, Unit Spatial Environmental Modelling, Boeretang 200, 2400 Mol – Belgium
ejeanluc.dekok@vito.be; guy.engelen@vito.be

European Commission - Joint Research Centre, Institute for Environment and Sustainability, Rural, Water and Ecosystem Resources Unit, Via E. Fermi, 2749, 21027 Ispra – Italy
joachim.maes@jrc.ec.europa.eu

Abstract: By itself scientific specialization does not serve the needs for integrated analysis of environmental problems required to improve the linkage between science and management. Models are often very case dependent, and lack the flexibility for rapid adaptation to new or different management or policy issues. The goal of the EU-FP6 research program SPICOSA is to develop a Systems Approach Framework (SAF) for Integrated Coastal Zone management. The project uses a modeling platform based on the integration of the dynamic simulation software ExtendSim and GIS software PCRaster. The approach is supported by model libraries in which reusable Model Building Blocks (MBBs) can be stored. The goal is to design these model components in a way allowing for reusability to design new simulation models for a different problem context from scratch. Such plug-and-play model library has several advantages: models are easier to build, adapt and maintain, but designing reusable MBBs for environmental modeling is not a straightforward task. Well-designed MBBs should be complementary to each other, self-explaining, encapsulated, robust, and meet the scientific standards. In addition, data management and the spatial dimension should be taken into consideration. The presentation will focus on the design principles of reusable MBBs and will provide some examples to explain the typical design issues and appropriate solutions.

Keywords: simulation, component-based design, model library, reusability, systems analysis

1. INTRODUCTION

Funded by the European Commission as an integrated project under the 6th framework program the SPICOSA project (www.spicosa.eu) aims to study the interaction between natural and socioeconomic processes that take place in Europe’s coastal zones by adopting a Systems Approach Framework (SAF). SPICOSA stands for Science and Policy Integration for Coastal System Assessment. The general goal is to develop a self-evolving, holistic research approach and support tools for the assessment of policy options for sustainable management through a balanced consideration of the ecological, social and economic aspects of coastal zone systems. Typical policy issues that are under study in the project are the socioeconomic problems arising from increased nutrient supplies to coastal
zone waters or the management of sustainable aquaculture in relation to freshwater supply and marine pollution. The project has a duration of four years and will last until February 2011. A key outcome of the project is a library of reusable model building components to standardize and facilitate the design and implementation of new, dedicated simulation models for addressing coastal zone management problems in the coastal areas of Europe. Such a component-based or “plug-and-play” approach for simulation modeling has several advantages. New models are easier to design and existing ones easier to modify, models become more transparent and easier to maintain, and scientific knowledge is easier to exchange and apply to different cases. Component-based design is the current-day standard in manufacturing and software engineering but few environmental modeling studies rely on reusable Model Building Blocks or MBBs. The APES framework for agricultural management (www.apesimulator.org) is an example of a component-based simulation environment. Due to its pan-European scale SPICOSA provides an excellent learning environment for examining the typical challenges and pitfalls of component-based design of multidisciplinary, environmental simulation models. All 18 SPICOSA case studies use the ExtendSim® simulation environment (ImagineThat, 2002; Krahl, 2002) to design, run, and test their coastal system models. In ExtendSim simulation models are constructed with standard library-based iconic blocks. In addition modelers can develop their own customized blocks and libraries. Each block describes a specific calculation or a step in a process, and is characterized by its in- and outgoing connectors, a user dialog and self documentation. An example is a block describing the temperature-dependent impact of light on phytoplankton growth. This paper takes the experience of SPICOSA to reflect on the pitfalls of designing reusable model building components and some actions to avoid these.

2. REUSABLE MODEL BUILDING COMPONENTS

The topic of reusable model building blocks for designing environmental simulation models receives more and more attention (Rizzoli et al., 1998; Daum, 1999; Argent, 2005; Papajorgji, 2005; Valentin and Verbraeck, 2002; Donatelli and Rizzoli, 2008; Verbraeck and Valentin, 2008). A model library allows models to be shared and (ex)changed in a modular fashion. Donatelli and Rizzoli [2008] emphasized the importance of framework independence of reusable model components and proposed a design approach based on component structure, interaction and quality. The design and population of a model library based on the outcomes of an ongoing project like SPICOSA is not straightforward. First of all, it is important to understand why a specific model block is reusable or not. Much of the success of component-based modelling depends on the quality and interoperability of the MBB’s. For SPICOSA the following requirements were formulated and communicated to the project participants (De Kok and Maes, 2009):

- reusable for new simulation models
- complementary like jigsaw puzzle pieces
- self-explaining and well documented (not only in a report)
- encapsulated, i.e. the implementation can be hidden
- easy to adapt, expand and maintain
- manageable in size
- robust and reliable (i.e. tested for e.g. foolproofness)
- scientifically correct and matured
One of the experiences of the project is that most scientists, in particular junior researchers, have great difficulty deriving reusable components from their models that can meet these requirements (see Section 4 for more details). In SPICOSA each system model consists of several sub models which are hierarchically structured around model components that allow better interpretation of the model (Figure 1).

![Figure 1. A framework for model block and hierarchical model design in SPICOSA.](image)

Some applications require the explicit spatial representation of processes. An example is the effect of the distribution of waste-water treatment facilities on the estuarine water quality. For this reason an interface was designed to link the dynamic simulation models in ExtendSim with PCRaster, a raster modelling environment (http://pcraster.geo.uu.nl). It includes sophisticated raster GIS functionality emphasizing on analytical capabilities. PCRaster uses a scripting language for constructing models describing processes through time. These dynamic simulation models can easily be constructed using the rich set of model building blocks and functions found in PCRaster (Van Deursen, 1995).

3. DESIGN AND ORGANIZATION OF THE MODEL LIBRARY

The reusable MBBs can either be designed as hierarchical blocks based on the standard graphical icon blocks provided in ExtendSim, or coded in ModL, the C-type modeling language available in ExtendSim. Although these MBBs are functionally dependent on the ExtendSim simulation framework the latter type of MBBs are easier to transfer to other simulation frameworks. To standardize the layout of the MBBs for the model library and ensure a satisfactory level of documentation the study sites have been provided with standardized model block templates (Figure 2).
In view of the scale and complexity of the project and the number and diversity of the deliveries to the model library, some additional measures were taken to ensure that the contents of the library will remain manageable and transparent for future users. First, the study sites were instructed to categorize their reusable MBBs in five categories, which are reflected by the templates: ecology, economics, physics, social processes, and a remainder category for non-domain oriented tools. Second, clear labelling and version number instructions were issued for all models, library files, and MBBs. For example, the name of each MBB is to indicate its functionality, the study site and version number. At a lower level similar instructions were given for the labelling of the connectors used to link the MBBs. A step-wise approach was recommended to guide the researchers with the design of their MBBs (Figure 3).
A separate work package in the project is responsible for the design and implementation of a model portal to be used for the storage, maintenance and exchange of both complete ExtendSim models as well as model libraries during and after the project (Figure 4). It can be used to upload and download models and model libraries, and search for specific model blocks depending on the requirements for its intended application.

An organizational aspect of the design of the MMBs is that a bottom-up approach was followed, i.e. the blocks were extracted after construction of the system models. This has
the advantage that the scientific experience gained with the case study applications could be used to the benefit of the MBBs. The drawback, however, is that the scientific content of the case applications limited the range of MBBs to be designed. Therefore, the model library can be improved in terms of completeness and complementarity of the MBBs.

4. PITFALLS AND CHALLENGES

Verbraeck and Valentin [2008] point out several potential problems with the use of a model library for constructing new models such as a limitation of the modelling scope, a lack of confidence in the MBBs, and wrong use of MBBs due to a lack of understanding of their applicability. In SPICOSA most effort has thus far been spent on familiarizing the modelers with the concept of knowledge reusability and providing guidelines and tools to facilitate the design of reusable model building components. In principle, the scale of SPICOSA should ensure a balance of the library components over the broad range of disciplines from marine ecology, pollutant transport and cost-benefit analysis to stakeholder participation concepts. Nevertheless, the social sciences and economics are less well represented in the work of the study sites, which will without doubt affect the consistency of the prototype model library. This in turn limits the applicability of the library contents to new problems or case studies. In SPICOSA the problem was addressed by means of a tentative list of model building blocks to obtain an overview of the missing or overlapping functionalities. The list helped avoiding overlapping functionalities and identifying key elements that were to be given a high priority. Another problem is inter-block consistency. When deciding on the in- and out-going connections the developers of a reusable model building block need to be aware of the operation of their block when linked to other model blocks, which may pertain to a scientific domain that is not theirs. This is a basic aspect of integrated systems modeling, but can be challenging as most researchers still take their education in a single domain. A general observation is that the focus of the project lies with quantitative simulation modeling, although the problem formulation and system conceptualization have explicitly been included in the design of the project. This raises the question to what extent qualitative concepts to describe processes, like for example stakeholder consensus, can be captured in a reusable MBB. An interesting technique that could be explored to clarify the underlying mechanisms behind a problem is that of conceptual mapping. Extracting reusable components from a model raised also less fundamental problems like that of the management of the data used by the model block. It was decided to make a distinction between three categories of data:

1. General constants and fixed parameters; these could be stored inside the block with the requirement to list them in the documentation of the block
2. User-defined parameters; for these dialogue boxes were designed to accompany the MBBs and present them in a more transparent manner
3. Medium and large-scale sets of data used by the blocks (as a rule of thumb exceeding more then 10 values); these were stored in ExtendSim databases. The name of the database could be set in the user dialogue, whereas the database itself should be included in the example model build illustrating the MBB.

Furthermore, some models were fine tuned during the calibration step, with model parameters that were strongly case specific and model output that was highly sensitive for changes in the value of these parameters. Taking a component out of its model context could easily lead to abuse of the block and unrealistic model behavior. To avoid this the MBB developers were advised to limit parameter ranges, document their model blocks very carefully and set up a sample model for the blocks. These sample models were also used on several occasions to clarify the principles of reusable model building components to the project teams. Other difficulties could be attributed to the heterogeneity in model structuring, layout, and documentation. This hampered the identification of library material and was expected to result in considerable quality differences of the contributions to the
model library by the different study sites. For this reason model block templates were developed and clear labeling and version number instructions were issued for the MBBs and library files.

5. CONCLUSIONS AND RECOMMENDATIONS

A bottom-up approach for the design of reusable model building blocks, based on existing simulation models, is more practical and can take advantage of the experience gained with the design of the integrated model. The drawback is that the scientific focus of the simulation studies will determine the scope of the model library. However, this can be expected to become less of a problem when (virtual) research communities start to exchange their model libraries more actively, which is one of the reasons for setting up the SPICOSA online model portal. Although the advantages of component-based design of environmental simulation models and model libraries of reusable components are clear to them, most scientists find it difficult to design model building blocks that meet the design criteria. Pre-constructed model block templates, tutorial blocks, and clear guidelines for a stepwise design of reusable MBBs can help overcome this problem. To ensure the robustness of the MBBs tests should not only be carried by means of plug-and-play replacement in the original simulation model, but need to be conducted also for new test applications. An inventory of potential candidate blocks for a model library can help avoid overlapping functionalities. On the other hand, the envisaged model library should leave room for different model concepts, even if these are closely related. To widen the support for component-based environmental modeling among the scientific community and facilitate the communication with end-users and stakeholders future work should also be aimed at developing tutorials and tools to make models more transparent and easier to control, improvement of the library completeness and inter-block consistency, testing blocks and setting up example applications which are scientifically convincing. The remainder of the SPICOSA project will be aimed at improving the quality of the model building blocks and complementing the library functionalities. The research consortium of the project, which includes over 50 institutes, provides a good starting point for this.
REFERENCES

Using an Integrated, Multi-disciplinary Framework to Support Quantitative Microbial Risk Assessments

Gene Whelan\textsuperscript{a}, Michael E. Tryby\textsuperscript{b}, Michell A. Pelton\textsuperscript{c}, Jeffery A. Soller\textsuperscript{d}, and Karl J. Castleton\textsuperscript{e}

\textsuperscript{a}U.S. Environmental Protection Agency, Athens, Georgia 30605 (Whelan.Gene@epa.gov); \textsuperscript{b}U.S. Environmental Protection Agency, Athens, Georgia 30605 (Tryby.Michael@epa.gov); \textsuperscript{c}Pacific Northwest National Laboratory, Richland, Washington 99354 (Mitch.Pelton@pnl.gov); \textsuperscript{d}Soller Environmental, Berkeley, California 94703 (jsoller@sollerenvironmental.com); \textsuperscript{e}Pacific Northwest National Laboratory, Richland, Washington (Karl.Castleton@pnl.gov)

Abstract: The Framework for Risk Analysis in Multimedia Environmental Systems (FRAMES) provides the infrastructure to link disparate models and databases seamlessly, giving an assessor the ability to construct an appropriate conceptual site model from a host of modeling choices, so a number of modeling analyses can be supported and reproduced. FRAMES is a Windows-based system that can incorporate and communicate with a array of software models and databases and that is uniquely designed to allow users, by themselves, to visualize the problem and add and link disparate models – even older legacy models – and databases to the system. Quantitative Microbial Risk Assessment (QMRA) is a modeling approach that integrates a wide range of disparate data, including fate/transport, exposure, and human health effects’ relationships, to characterize potential health impacts/risks from exposure to pathogenic microorganisms. Although QMRA does not preclude the use of source-term and fate and transport models, it most commonly has been applied where the “source term” is represented by the receptor location (i.e., exposure point), meaning that the full potential of a QMRA has not been realized traditionally. This paper describes unique attributes of FRAMES and demonstrates how open-system architecture can be used to link disparate models and databases to support a QMRA application, while addressing multiple microbial source types and organisms that impact downstream receptors.

Keywords: QMRA, Risk Assessment, Integrated Modeling, Pathogens, Watersheds

1. INTRODUCTION

Fecal bacteria are among the most common pollutants affecting rivers and streams. EPA [2002] revealed that 35% of impaired rivers and streams were polluted by fecal bacteria (generally indicated by fecal coliforms, enterococci, or \textit{E. coli}) which could indicate the presence of pathogens. Due to large numbers of farm animals and wildlife, animal fecal matter may be an important source of contamination in rural areas. Among various animal fecal sources, poultry are responsible for 44% of the total feces production in the United States, followed by cattle (31%) and swine (24%) [Kellog et al. 2000]. Quantitative Microbial Risk Assessment (QMRA) uses information on the distribution and concentration of particular pathogens and infectivity data to determine risk to public health [Hunter et al. 2003; Haas et al. 1999; ILSI 2000, 1996]. Although QMRA does not preclude the use of source-term and fate and transport models, as demonstrated by Ferguson et al. [2007] and Signor et al. [2007], it has focused most commonly on the receptor location (i.e., exposure point), dose-response relationships, and health impacts, as
prescribed in Haas et al. [1999]. Thus, the full benefit of the QMRA paradigm traditionally has not been realized. Gaber et al. [2008] define integrated modeling as “a systems analysis-based approach to environmental assessment. It includes a set of interdependent science-based components … capable of simulating the environmental stressor-response relationships relevant to a well specified problem statement.” This paper describes unique attributes of the Framework for Risk Analysis in Multimedia Environmental Systems (FRAMES) and demonstrates how open-system architecture can link disparate models and databases to support a QMRA application, while addressing multiple microbial source types and organisms that impact downstream receptors.

2. FRAMEWORK FOR RISK ANALYSIS IN MULTIMEDIA ENVIRONMENTAL SYSTEMS

FRAMES standardizes data exchange between models with different Input/Output (I/O) attributes [PNL 2010]. The operation of the system is controlled by the Application Programming Interface (API) and Framework Development Environment (FDE), which use standardized dictionaries to describe the metadata associated with all data recognized by the system. A DICTIONARY (DIC) file is a comma-delimited text file that contains each parameter’s metadata, including name, description, units, measure (i.e., groupings of units), data type, range, stochastic, and indices (dependency on other parameters) [PNL 2010]. The FRAMES API handles execution management and file I/O, and provides a series of editors that allows the user to register and operate components with and in the system, and helps facilitate the linking of disparate models. The editors help the user through the model and I/O registration processes. Editors and other tools include:

- DICTIONARY (i.e., DIC) Editor – Registers new or edits existing dictionaries.
- (Units) Conversion Editor – Registers additional or edits existing unit conversions supported by the system and allows legacy models to maintain their current use of units, relegating the responsibility of unit conversion between modules to the system.
- Module (DES) Editor – Registers attributes of the model in the system, such as the model’s icon pictogram; connection schemes with other models; input DICs consumed and output DICs produced by the model; folder location of executables, user interface, and related files; contact information; and software requirements.
- Domain Editor – Registers where the model fits in the system and is composed of a Domain, Group, and Subgroup. A Domain defines a grouping of components (e.g., models, databases, and related components).
- Simulation Editor – Allows the user to edit the Conceptual Site Model (CSM) work space, containing the drag-and-drop functionality of constructing a CSM, linking modules in a sequential order, and managing the sequential execution of the modules (Figure 1). The Simulation Editor is designed to be an intuitive interface for interaction with the CSM diagram and contains four user-interface areas. The upper left provides for a user-defined logo. The bottom left describes the Domain’s icon palette, from which the user chooses models, databases, viewers, or system tools. The top right is the “Global Workspace,” whose output can be accessed by all modules, and the bottom right is the “Local Workspace,” whose data flow is determined by physical connections.
- Data Client Editor (DCE) – Manipulates DICs. Because DICs represent the “monetary” exchange within FRAMES, the DCE can be used 1) as a user-defined module, allowing the user to provide input boundary conditions manually to any module; 2) to provide a relatively simple graphical user interface for those models whose UIs are not FRAMES-compliant; and 3) as a tabular viewer, providing the user with a simple means to view output in table form.
- Lock and Key Features – Allows “lock down” the CSM (i.e., picture containing linked icons), restricted access to certain models, or both, using password protection.
- Plus-Operator – Temporally combines multiple, like outputs, where appropriate, using linear superposition, to create a single input file for consumption by a downstream module (see Figure 1, icon titled “Sum SW Concentrations”).

1234
Figure 1. Simulation Editor, depicting the Conceptual Site Model for six source terms and their fate, transport, and health impacts at a receptor location.

- Simulation Packager/UnPackager – Packages the entire simulation, so it can be sent to any remote computer, unpackaged, and then executed on that computer containing a compatible version of FRAMES to reproduce the entire simulation.

- Synchronization Operator – Transforms the output from multiple modules and creates new input boundary conditions by ensuring that each time-varying profile for each module output contains values at the same time steps, using linear interpolation to fill times in between time steps.

- Dictionary Registration Tool (DRT) – Uses a spreadsheet formatted to import and register the DICs automatically, while displaying warnings and errors on the status of the process. The DRT is written in Visual Basic for Applications (VBA) for Excel.

- Sensitivity/Uncertainty Modules – Supports Monte Carlo analyses. This module accesses model inputs and allows the user to alias them and assign statistical characteristics (i.e., distribution, correlation, and/or equation) to each. New features are being added, like the statistical package R, maximum likelihood or least-squares model parameter estimation, and the parameter estimation tool PEST.

FRAMES was originally designed to cater to three types of users. Model developers build or import DICs, define units, build/import modules/models, set-up domains, and define connection schemes, and FRAMES supports developers with the Conversion, Dictionary, Domain, and Module Editors. Users/Analysts select Domains and icons, connect icons when building CSMs, select modules/models and databases, populate user interfaces, and execute the CSM, and FRAMES supports users with the Simulation, Data Client, and Module Editors. Database owners map database schema and develop database extraction plans. Another software program, D4EM, now shoulders these responsibilities.

3. EXAMPLE APPLICATION

FRAMES is a software structure for implementing an example QMRA that leverages and links disparate models in a unified framework for model integration. Source-to-outcome microbial exposure and risk modeling is demonstrated for an agriculturally contaminated runoff scenario in a conceptual watershed. The example consists of multiple adjacent fecal contamination sources, located within the same watershed.
3.1 Description

Six potential, but disparate sources of manure-based pathogen contamination, illustrated in Figure 2, are modeled in this multimedia example: tributary inflow (assumed boundary condition), grazing cattle on an open field (requiring overland runoff modeling), leaching from a waste storage basin (requiring subsurface modeling), pond overflow during precipitation events (inflow equals outflow directly to a stream), land application of disposal-pond contents (requiring overland runoff modeling), and cows directly shedding to the stream. Rainfall events drive contamination from sources related to runoff, while other sources are influenced by agricultural operations and practices. In each case, fecal contamination enters the appropriate stream segment and flows downstream to a recreational location; therefore, all sources require instream modeling. Figure 1 presents a CSM of the six potential disparate sources of manure-based pathogen contamination, routed from their sources to a receptor of concern, with each icon representing a separate model. Assumptions associated with the assessment are summarized in Tables 1 and 2.

3.2 Models

To illustrate the ability of FRAMES to support a QMRA application, a series of reduced-form models, designed and built by different developers, are linked to FRAMES; these cover several components of the QMRA paradigm: source terms (e.g., ponded release), watershed, stream, aquifer, and human exposure/risk. The groundwater and surface water models were previously linked to FRAMES. For this particular effort, new linkages associated with the exposure/risk model and watershed results were performed.

It is assumed that a leak occurs in the disposal pond over 1% of the area. A unit hydraulic gradient is assumed below the pond, resulting in constant outflow to the soil medium [Hillel 1971]. A watershed model, based on kinematic wave theory [Eagleson 1970], was constructed and employed to account for runoff from 1) land application of pond waste and 2) grazing cattle with daily loadings. The watershed model results were exercised, using Data Client Editors (DCEs) within FRAMES. The MEPAS saturated zone model, based on the three-dimensional dispersive, one-dimensional advective equation with inactivation and soil-water partitioning, forms the basis for microbial movement from an area source through a porous medium [Whelan et al. 1999], recognizing that research is still required to more fully understand and substantiate the transport of non-virus pathogens in porous media. The MEPAS surface water model is used in the analysis and is based on a vertically integrated, steady-state solution to the one-dimensional advective, one-dimensional (lateral) dispersive equation with inactivation [Mills et al. 1997].
Table 1. Microbial Characteristics

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Units</th>
<th>Salmonella</th>
<th>Ecoli0157</th>
<th>Cryptosporidium</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inactivation Rate (in soils)</td>
<td>1/d</td>
<td>0.23</td>
<td>0.16</td>
<td>0.04</td>
<td>Soller et al. [2009]</td>
</tr>
<tr>
<td>Inactivation (in surface water)</td>
<td>1/d</td>
<td>1.30</td>
<td>0.54</td>
<td>10</td>
<td>Soller et al. [2009]</td>
</tr>
<tr>
<td>Distribution Coefficient</td>
<td>mL/g</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>Pachepsky et al. [2006]</td>
</tr>
<tr>
<td>Prevalence</td>
<td>%</td>
<td>10</td>
<td>20</td>
<td>30</td>
<td>Soller et al. [2009]</td>
</tr>
<tr>
<td>Tributary Inflow</td>
<td>g/yr</td>
<td>6.00E-03</td>
<td>6.00E-03</td>
<td>1.39E+00</td>
<td>Assumed, 2-day lag</td>
</tr>
<tr>
<td>Excretion Density (Log10)</td>
<td>#/g manure</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>Soller et al. [2009]</td>
</tr>
<tr>
<td>Pathogen Pond Concentrations</td>
<td>mg/L</td>
<td>8.85E-03</td>
<td>1.77E-03</td>
<td>6.14E-01</td>
<td>Assumed, after Rogers et al. [2009]</td>
</tr>
</tbody>
</table>

Table 2. Source and Media Characteristics

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Units</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animal Characteristics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cow Density</td>
<td>5</td>
<td>cow/ha</td>
<td>Duhigg [2009], Butler et al. [2008a]</td>
</tr>
<tr>
<td>Number of Cows</td>
<td>360</td>
<td>#</td>
<td>Assumed</td>
</tr>
<tr>
<td>Shedding Rate of Cow</td>
<td>24</td>
<td>kg/d</td>
<td>Soller et al. [2009]; Duhigg [2009]; Butler et al. [2008a]</td>
</tr>
<tr>
<td># of Cows Shedding to Stream</td>
<td>36</td>
<td>#</td>
<td>Assumed 10%</td>
</tr>
<tr>
<td>Soil Characteristics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soil Type</td>
<td>Sandy Loam</td>
<td></td>
<td>Assumed</td>
</tr>
<tr>
<td>Land Bulk Density</td>
<td>1.58</td>
<td>g/cm³</td>
<td>Meyer et al. [1997]</td>
</tr>
<tr>
<td>Land Porosity</td>
<td>0.41</td>
<td>fraction</td>
<td>Meyer et al. [1997]</td>
</tr>
<tr>
<td>Saturated Hydraulic Conductivity</td>
<td>1.17E-03</td>
<td>cm/s</td>
<td>Meyer et al. [1997]</td>
</tr>
<tr>
<td>Overland Flow Characteristics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mannings Constant</td>
<td>1.49</td>
<td></td>
<td>Eagleson [1970]</td>
</tr>
<tr>
<td>Mannings Coefficient</td>
<td>0.20</td>
<td></td>
<td>Whelan [1980]</td>
</tr>
<tr>
<td>Friction Slope (Sf)</td>
<td>0.005</td>
<td></td>
<td>Assumed</td>
</tr>
<tr>
<td>Precipitation Intensity (i)</td>
<td>9.68</td>
<td>cm/d</td>
<td>NOAA [2009]</td>
</tr>
<tr>
<td>Mannings exponent (m)</td>
<td>1.67</td>
<td></td>
<td>Eagleson [1970]</td>
</tr>
<tr>
<td>Size of Overland Areas</td>
<td>72.8</td>
<td>ha</td>
<td>Assumed square</td>
</tr>
<tr>
<td>Precipitation Events per year</td>
<td>10</td>
<td>#/yr</td>
<td>Assumed</td>
</tr>
<tr>
<td>Pond and Land Application Characteristics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depth of Pond</td>
<td>3</td>
<td>m</td>
<td>Assumed</td>
</tr>
<tr>
<td>Area of Pond</td>
<td>1.44E+03</td>
<td>m²</td>
<td>Assumed square</td>
</tr>
<tr>
<td>Fraction of Pond that Leaks</td>
<td>0.010</td>
<td>fraction</td>
<td>Assumed</td>
</tr>
<tr>
<td>Storage Basin E coli Concentration</td>
<td>3.16E+06</td>
<td>MPN E. coli/100mL</td>
<td>Rogers et al. [2009]</td>
</tr>
<tr>
<td>Flow into/out of Pond/event</td>
<td>133</td>
<td>L/d/cow</td>
<td>Duhigg [2009], Butler et al. [2008a]</td>
</tr>
<tr>
<td>Pond Land Applications/yr</td>
<td>4</td>
<td>#/yr</td>
<td>Assumed</td>
</tr>
<tr>
<td>Groundwater Characteristics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soil characteristics</td>
<td></td>
<td></td>
<td>Meyer et al. [1997]</td>
</tr>
<tr>
<td>Darcy Velocity</td>
<td>1</td>
<td>cm/d</td>
<td>Assumed</td>
</tr>
<tr>
<td>Surface Water Characteristics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discharge</td>
<td>42.5</td>
<td>m³/s</td>
<td>Assumed</td>
</tr>
<tr>
<td>Width</td>
<td>30.5</td>
<td>m</td>
<td>Assumed</td>
</tr>
<tr>
<td>Velocity</td>
<td>0.91</td>
<td>m/s</td>
<td>Assumed</td>
</tr>
<tr>
<td>Tributary Characteristics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lag time</td>
<td>2</td>
<td>d</td>
<td>Assumed</td>
</tr>
<tr>
<td>Maximum Tributary Discharge</td>
<td>6.8</td>
<td>m³/s</td>
<td>Assumed</td>
</tr>
</tbody>
</table>
Microbial Risk Assessment Interface Tool (MRA-IT) is an open-source, MathCad- and event-based, integrated software tool for characterizing human-health impacts from ingestion of reclaimed water, based on the pathogen of interest, exposure, intake, and dose [Soller and Eisenberg 2008; Soller et al. 2007]. As a stand-alone, it does not contain a fate and transport component, nor does it accept inputs of microbial densities from multiple upstream models. FRAMES provides the upstream fate and transport models and the ability to combine multiple inputs, as illustrated by the “Plus-Operator” (see module titled “Sum SW Concentrations” in Figure 1), forming a time-varying input density curve for consumption by MRA-IT. A pathogen list (i.e., *Salmonella*, *Cryptosporidium*, and *E. coli* 0157) is supplied from the FRAMES Constituent Database Selection module. After calculations are complete and output written, the MRA-IT MathCad-based UI presents graphical and tabular results to the user, as with its stand-alone version.

### 3.3 Results and Discussion

Figure 3 presents typical time-varying pathogen densities for the first four rainfall events (i.e., peaks A, B, E, and F) associated with *Cryptosporidium* at the receptor location; the results account for contamination from all six sources, with point C being tributary inflow. These results are very similar to those exhibited by *Salmonella* and *E. coli* 0157 and are indicative of the entire one-year simulation. This is the density curve exiting from the Plus-Operator module. The only sources contributing to contamination at the receptor at all times are leakage from the Pond and cows shedding directly to the stream (D in Figure 3). Closer inspection of the results indicates that the manure application method (e.g., shedding, spreading, pond leakage, etc.); pathogen rate of release; timing of the manure loading; sequence and type of transporting media; pathogen characteristics (e.g., prevalence, excretion density, inactivation rate, and distribution coefficient); timing of rainfall events; duration and intensity of rainfall; antecedent moisture conditions; and landscape characteristics all play important roles in identifying which source contributes to the contamination and pathogen density at the receptor location, and to what degree.

![Figure 3. Time-varying Cryptosporidium Concentrations at the Receptor location for the First Four Rainfall Events](image)

MRA-IT is used to estimate risks where a receptor potentially is exposed to contaminated water (e.g., swimming for the day at a beach); hence, an event window needs to be defined. In this example, an exposure event was chosen during the recession limb of the pollution hydrograph after the second storm event, and it is assumed to last 1.46 days. Figure 4 presents the time-varying densities associated with *Cryptosporidium*, *E. coli* 0157, and *Salmonella* during the event window of 0.102 – 0.106 yr at the receptor. These data represent the input boundary conditions produced by the transport modeling for consumption by MRA-IT. The Monte Carlo-based risk assessment for the three pathogens indicates that the risk for infection to *Cryptosporidium* is slightly larger than that of *E. coli* 0157 and significantly larger than *Salmonella*. For example, there is a 50% probability of exceeding an individual risk of 1.4x10^{-4} and 8.0x10^{-5} for *Cryptosporidium* and *Salmonella*, respectively, and there is a 10% probability of exceeding an individual risk of 3.8x10^{-2} and 1.5x10^{-2} for *Cryptosporidium* and *Salmonella*, respectively.
4. SUMMARY

Quantitative Microbial Risk Assessment (QMRA) is a modeling approach that integrates fate/transport, exposure, and dose-response relationships, to characterize potential health impacts/risks from exposure to pathogenic microorganisms. FRAMES facilitates a user’s linkage of disparate models and databases to support a custom assessment and to provide a structure that better leverages the capabilities of QMRA beyond the point of exposure. A series of models and databases were linked to assess six potential sources of manure-based pathogen contamination, thereby simulating the fate, transport, and health impacts from three pathogens to a recreational receptor at a downstream exposure point. By combining fate and transport modeling with point-of-exposure calculations, an analyst can begin to evaluate importance of the components more holistically, including manure application method, pathogen rate of release, timing of the manure loading, sequence and type of transporting media, pathogen characteristics, timing of rainfall events, duration and intensity of rainfall, antecedent moisture conditions, and landscape characteristics. Past user experience (e.g., hazardous or radioactive risk assessments) has indicated that the mechanics of building Conceptual Site Models are straight-forward. Training though is advised for the more involved operations (e.g., registering models, parameters, and connection schemes, and constructing dictionaries), which is not an atypical requirement for a sophisticated modeling system. The user also needs to have some familiarity with the models, as the system takes no responsibility to train the users in their operation.

The views expressed in these Proceedings are those of the individual authors and do not necessarily reflect the views and policies of the United States Environmental Protection Agency. Scientists in EPA have prepared the EPA sections, and those sections have been reviewed in accordance with EPA’s peer and administrative review policies and approved for presentation and publication.

5. REFERENCES


Sollers, J., S. Petterson, and T. Bartrand, Development of a QMRA model to evaluate the relative impacts to human health risks from animal-impacted recreational waters, prepared by Soller Environmental, LLC, for the U.S. Environmental Protection Agency Office of Water, Washington, DC. (Draft), 2009.


Farmers as water managers in a changing climate – can we turn sustainability research into outcomes?

Matthews, K.B.1, Blackstock, K.L.2, Langan, S.3, Rivington, M.1, Miller D.G.1, Buchan, K.1

1 Integrated Land Use Systems Group, Macaulay Institute, Craigiebuckler, Aberdeen, U.K.
2 Socio-Economic Research Group, Macaulay Institute, Craigiebuckler, Aberdeen, U.K.
3 Catchment Management Group, Macaulay Institute, Craigiebuckler, Aberdeen, U.K.
Email: k.matthews@macualay.ac.uk

Abstract: How well does research-derived knowledge on sustainability translate into practical improvements in the sustainability of land and water management? This paper reflects on progress being made through the lens of the “Aquarius - Farmers as Water Managers” project part of Interreg IVb. This project seeks to use research-derived knowledge to aid the specification, delivery and evaluation of a flood alleviation scheme that uses natural flow management measures. The paper sets out previous research on the definition of outcomes and places these in the context of the specific criteria being used within the Interreg programme. Interreg provides an opportunity for researchers interested in translating research into outcomes to be part of large scale interventions that are beyond the scope of research studies. The paper details the research approach being adopted; a variant of adaptive management schemes intended for use in complex coupled social-ecological systems. The base line studies that characterise the bio-physical and socio-economic systems and the framework of stakeholder issues have been completed. These have emphasised the contested nature of the causes of, and responsibilities for, flooding and its future management in a changing climate. The paper also reports the central role of computer-based modelling in the statutory cost-benefit analyses. Going beyond these statutory processes the paper reports an analysis of the levels of compensation needed by land managers to offset the loss of income from temporary flooding. This amounts to <5% of the annual damage and presents an opportunity for financial solutions based on insurance or public support for regulating ecosystem-service provision. The paper concludes by reflecting on progress against the Interreg outcome criteria and notes that the research-practice-policy partnerships are working well and that international cooperation has been successful in promoting innovative engineering, financial, governance and policy options.

Keywords: farmers; water-managers; climate-change; Aquarius

1 INTRODUCTION

How well does research-derived knowledge on sustainability translate into practical improvements in the sustainability of land and water management? This paper reflects on progress being made through the lens of the “Aquarius - Farmers as Water Managers”
project, funded by the European Regional Development Fund (ERDF) as part of the Interreg IVb programme in the North Sea Region (NSR).

1.1 Defining outcomes

Outcomes are most simply defined as effects that occur beyond the walls of the research organisation. The authors have previously argued that it is useful to differentiate between process effects and outcomes [Matthews et al., 2010]. Process effects are changes to how things are done, for example improved efficiency or capacity. Outcomes on the other hand are changes to awareness, attitudes and actions of stakeholders. The former have the potential to be transformative beyond the research and development environment while the latter actually cause verifiable change. Outcomes can thus encompass a wide range of changes. In Interreg IVb the outcomes they are seeking are defined through the promotion of three ideals that together aim to improve to quantity of life and protection of the environment of the NSR. The three ideals are – innovation, inclusion and implementation. Innovation seeks to improve on current practice, (providing an opportunity to use existing research derived knowledge, tools and approaches). Basic or strategic research per se is explicitly excluded since the focus of Interreg is on process and the delivery of outcomes. Inclusion supports the building of cross-sectoral partnerships and networks: locally, regionally and internationally (research-practice-policy linkages). Finally implementation means demonstrating how good practice works through building and evaluating pilot projects (a form of action research). These ideals combine to provide opportunities for researchers who are interested in seeing their research result in improvements in the sustainability of land and water management.

1.2 Evaluating outcomes

Even when outcomes are sought, the criteria by which they are to be evaluated are often unclear. Within Interreg IVb, however, projects are evaluated against explicit outcome criteria, summarized as the 5 L’s.

- **Linkage**, building on existing knowledge (including research) from previous Interreg or research programs.
- **Longitude**, ensuring strong cross-sectoral partnerships (research, agencies, government, NGO and publics).
- **Latitude**, partnerships between regions to share expertise.
- **Locality**, or how to ensure that any intervention includes the issues of overriding importance to local stakeholders.
- **Legacy**, or how to create structures, capacity and outcomes that can be sustained beyond the lifetime of a particular program.

This paper reflects on how well the Aquarius project is progressing towards the outcomes sought by Interreg using the 5 L’s as criteria. The paper also seeks to draw some more generic conclusions on the nature of the challenges that confront researchers seeking to turn sustainability research in to sustainability outcomes.

1.3 **AQUARIUS**

Aquarius is a transnational project (7 partners in 6 countries across the region) and is part of the ERDF intervention “Adapting to and reducing the risks posed to society and nature by a changing climate” within the wider objective “Promoting the sustainable management of our environment”. The overall aim of Aquarius is to identify and overcome the barriers to farmers contributing positively to addressing water management issues (e.g. in flooding, water shortage and water quality)\(^1\). Farmers are seen as crucial actors for a wide range of water management issues (e.g. mitigating flooding, avoiding damaging low-flows and controlling diffuse pollution). Many of these issues are likely to be exacerbated by climate change.

\(^1\) The particular mix of interests varies between partners but all are concerned with multifunction land-water systems.
change and there is a recognition that existing approaches to their management are either inadequate, cannot be sustained financially or result in undesirable unintended consequences. Particularly important for Aquarius is the promotion of partnership between land managers, competent authorities and researchers. This made Aquarius an appropriate case study within which to develop and test research derived approaches to sustainable land and water management.

Each regional partnership is undertaking a pilot/demonstration project addressing the most important local issues (in addition to transnational or bilateral exchanges of knowledge between the Aquarius partners via expert networks, partner meetings and reports). For Scotland the partners are Aberdeenshire Council (as the competent authority) the Macaulay Land Use Research Institute (land, catchment management and socio-economic research groups) and a local environmental consultancy (LandCare NE). There are also a significant number of direct and indirect stakeholders.

The Scotland pilot is being undertaken in the Tarland Burn catchment (a sub-catchment of the River Dee in Aberdeenshire – see Figure 1). Aquarius is informing the specification and implementation of the Tarland Flood Prevention Scheme (TFPS), intended to relieve flooding pressure on the villages of Tarland and Aboyne. While the primary focus is on flood alleviation the pilot partners aspire to go beyond hard engineering approaches, to take a systemic, multi-scale approach and to include in the evaluation of intervention options a wider range of criteria so as to provide a more rounded assessment of their sustainability.

The intervention options being considered are “natural” in that they seek to restore or enhance the storage capacity within of river systems (e.g. re-meandering, (re)establishing wetlands or flood storage basins) rather than building defences that while locally effective simply move the problem downriver. For AC more natural flow management options are being considered since these have the potential to be undertaken as partnerships with land managers, rather than requiring the compulsory purchases of land (incurring substantial legal and other costs) and requiring subsequent maintenance. More conventional interventions, such as a by-pass channel for Tarland village are not being ruled out but are recognised as at best partial solutions as they simply move the locus of impacts.

The process of generating the Aquarius proposal and undertaking the project is providing a wealth of insights into the practical, institutional and research challenges that remain to be overcome in translating research derived knowledge, methods and tools into practical improvements in the sustainability of land and water management. This paper reports on progress to date in responding to these challenges and assesses how successful we have been in translating sustainability research in to improvements in outcomes for land and water management.

2 MATERIALS AND METHODS

2.1 Aquarius Tarland Case-Study

Direct stakeholders are involved in the project, beyond these stakeholders are consulted and/or informed.
Aquarius started in March 2009 and has four phases (A-D). Phase A (March 2010 to March 2011) has established a system-baseline: current conditions, climate change scenarios and the key land-water management issues as seen by stakeholders (farmers, local residents, land-owners, statutory agencies, fishery interests etc). The aim of this phase was to fully understand the barriers to farmers acting as water managers. Phase B (from March 2010 to 2011) will assess scenarios of potential interventions (online and offline storage) using empirical and simulation-model derived data and experience from transnational partners. This analysis will explore the series of trade-offs required e.g. between cost of the intervention, its effectiveness (for flood alleviation), and the impacts on the existing land management regimen. The analysis also seeks to highlight and where possible quantify synergies between water management and other services such as enhanced bio-diversity; water quality or landscape character. With stakeholders the intention is to investigate how payment for environmental services or for the maintenance of key public infrastructure could contribute positively to participating farm businesses. In Phase C a pilot demonstration and monitoring site will be built. Phase D will assess and evaluate the project in terms of its implications for planning processes, construction impacts, maintenance and acceptability.

2.2 Research Processes

The approach being adopted by the Aberdeenshire Aquarius partners draws its inspiration from those developed and applied by Kay et al. [1999]. These adaptive management approaches for complex systems recognise that there is the need to generate baseline system descriptions and a framework of visions and preferences for the future. The systems baseline includes both biophysical and socio-economic components and the choice of scale and detail is guided both by the nature of the system (e.g. a catchment for a water focused issue) and the preferences and concerns of the key stakeholders identified in the issues framework. It is particularly important to generate explicit (if not agreed) definitions of the issues otherwise the research will be neither salient nor credible and this will undermine the legitimacy of the whole process.

The information generated by the baseline analysis for the Tarland case study is intended to 1) increase the project Partners understanding of the nature of the land-water management issues; 2) underpin assessments of the feasibility and acceptability of possible interventions, 3) identify the key barriers to active land-water management by farmers (particularly financial and/or institutional) and 4) inform government of possible policy measures. Table 1 presents the baseline analyses undertaken by the Aquarius Scotland partners. It can be seen from the number and variety of activities considered necessary (and given the resources available to the Aquarius project the list is in no way exhaustive) that there is little prospect of tangible improvements to the sustainability of land and water management though single-disciplinary, science led initiatives alone.

<table>
<thead>
<tr>
<th>Baselines</th>
<th>Components</th>
<th>Data, methods and tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geographical</td>
<td>land cover, use, ownership, holding size &amp; tenure, stocking, employment</td>
<td>Field, holding and business databases, GIS mapping [Matthews et al., 2008a].</td>
</tr>
<tr>
<td>Climate</td>
<td>Agro-meteorological indicators for current and 2070-2100 period</td>
<td>UK Meteorological office datasets and regional climate models [Matthews et al., 2008b]</td>
</tr>
<tr>
<td>Economic</td>
<td>Output, added value, assets, liabilities &amp; subsidies,</td>
<td>Review of regional economic summaries [SG RERAD, 2009]</td>
</tr>
<tr>
<td>Ecological</td>
<td>Water quantity, quality and river morphology</td>
<td>WFD water body characterisation from 3DeeVision⁴</td>
</tr>
<tr>
<td>Sociological</td>
<td>Farmers as water managers Authorities and Advisers views Farmers and Factors views</td>
<td>Questionnaires, Interviews and Workshops.</td>
</tr>
</tbody>
</table>

⁴ 3DeeVision website http://www.3deevision.org/78/
Legal Applicable policies, support mechanisms, planning regulations | Review of published sources – governance mapping.

**Statutory processes**

The research being undertaken by the authors is designed to complement the work being undertaken as part of statutorily defined processes. These statutory processes include strategic environmental assessments (SEA, undertaken for the Council by environmental consultants) and site specific environmental Impact Assessments (EIA) for particular activities on chosen sites. The SEA has been ongoing since a strategic decision was taken by elected members of AC that there needed to be an investigation of intervention options following significant flooding in the village of Aboyne in 2002. The initial identification of potential sites suitable as temporary flood storage basins was undertaken by Aberdeenshire Council staff on the basis of visual survey of suitability (see Figure 2).

Hydrological and hydraulic modelling was conducted by consultants to Aberdeenshire Council. A 1-d hydraulic model for the catchment was constructed using InfoWorks4 (1000 node). Input hydrological data was available from four hydrographs. The overall topography of the catchment was defined using 1m resolution digital elevation model derived from LIDAR5. The model has 25 km of watercourses consisting of a main channel and principal tributaries. Cross channel and riparian profiles were defined by topographical survey at 399 locations. The model recognises 385 spill units (parts of floodplains). The model also contains 37 key infra-structures that affect flow e.g. bridges and culverts. Seven point inflows and nine lateral inflows provide inputs to the rainfall-runoff relationships defined using UK Flood Estimation Handbook (FEH) methods [Centre for Ecology and Hydrology, 2008].

The model was calibrated for four events between 2005 and 2008. Full calibration was restricted to the Scottish environmental Protection Agency (SEPA) Aboyne hydrograph at the bottom of the catchment, with the Aberdeenshire Council Tarland site used for stage and the Macaulay sites at Tarland (top of the upper catchment) and Coull (at the end of the Tarland floodplain) used only for the timing of the flood peaks. Particular issues were found in calibrating the model to reproduce observed events: the need for high values for the rainfall to runoff coefficient, the need for different soil wetness values across the catchment and timing of events (ensuring that the peak flow was not too slow to reach Tarland village in the upper catchment nor too slow to reach Aboyne). These issue were overcome and frequency and volumes of peak flows were then calculated using statistical estimation procedures of the FEH. These values were then used to scale the outputs of the InfoWorks model deriving maps of the median, 5, 10, 25,100 and 200 year inundation events and a 200 year + 20% event (the standard simulation of climate change).

The validation of the modelling process was undertaken by SEPA but this does not evaluate the quality of the results obtained. Informal validation has been carried out through contacts with land managers at Aquarius workshops backed by historical “trash line” surveys recorded after major events. Given the relative ease of identification using remote sensing (standing water being particularly distinctive) the Macaulay team intend to pursue this as a more systematic way of evaluating the effectiveness of the model for particular events. The issue of course is the availability of satellite coverage for the largest events that occur during periods of significant rainfall and thus cloud cover.

The inundation maps are the key factor used in the calculation of damage to property and infrastructure. The entities affected are identified by overlaying the inundation maps with asset registers for public and maps of housing. The depth of inundation is used to assess whether the waters have passed floor level (a key threshold) and the cost of damage assess [Penning-Rowsell, 2005]. Strictly the budget for any works is the damage assessment (discounted over 100 years) minus any limitation in the effectiveness of measures (e.g. they may only be effective 1 in 100 instead of the standard 1 in 200). Informally it is also

---

5 LIDAR – LIght Detection And Ranging derived elevation data.
possible to argue that if there are secondary benefits that can be assessed as resulting from the intervention then these maybe included in the assessment, particularly if they are likely to provide a resource stream for maintenance or capital build.

RESULTS

2.3 Overall Characterisation

The overall characteristics of the Tarland catchment are reported in the Baseline Report\(^6\) and it is only possible to briefly summarise key highlights here. The area of the catchment is 7300 ha, and has an elevation range of 620m to 109m. This means that the catchment is on the margins of the more intensive agriculture (cropping and rotational grass). Arable and rotational grassland makes up 42%, permanent grassland 18% and forestry 30%. Most land is rented on long term tenancies (76%), with median holding size of 38 ha. There is a complex mosaic of ownership and management (see Figure 3 where the candidate basins are overlaid with land ownership, note especially the woodland areas which are managed directly by the tenant’s landlord.

![Figure 2 - Storage Basins](image)

![Figure 3 Basins and Ownership](image)

The mix of enterprises means the farm types are dominated by mixed farming (41%) and upland sheep and cattle rearing (36%). Farm incomes have risen over the period 2003 to 2008 but are still heavily dependent on subsidy. Only 89 FTE are employed in agriculture in Tarland. Regionally agricultural employment is ~3% and 1.8% of regional Gross Value Added. The villages have a significant role as dormitory towns for Aberdeen and the provision of local services. The watercourses are designated under the Habitats’ Directive for fish and invertebrate species, and are currently failing to meet good ecological status because of morphology (excessive canalisation) and diffuse pollution (phosphorous).

2.4 Flood volumes, extents and cost-benefit analysis

The consultants’ modelling has indicated that that the likely volume that any intervention will need to deal with to alleviate flooding in Aboyne is 950,000 m\(^3\). The first option being considered is constraining the maximum flow permitted with the excess diverted into temporary flood storage basins. This means, assuming a mean depth of 1m is achieved, 95 ha of temporary flood storage beyond those areas already inundated. The budget for such an intervention is constrained by a cost-benefit analysis. Accepting that the outputs of the hydraulic model are an adequate representation of the flooding extent a preliminary cost-benefit analysis has been undertaken. For Tarland village, see Error! Reference source not found., the inundation mapping when combined with property mapping estimates that 21 properties are affected, 7 above floor level, in the 25 year event and 30 properties are affected, 14 above floor level in the 200+climate change year event. Using standard inundation-to-damage functions this means average annual damage of £25k to £30k and total benefits of £800k for 100 years. Initial assessment are that this is unlikely alone to justify constraint and basin works but a bypass channel may be added for local protection.

\(^6\) Aquarius website – Tarland report
with increased storage further down the catchment. It may be possible, however, to make the argument for other “natural” flow management measures due to their reduced capital and administrative costs (through cooperation with land managers). For Aboyne the flooding is more significant with the modeling and mapping identifying 72 properties as affected, 18 above floor level in the 25 year event and 101 properties affected, 36 above floor level in the 200 year + climate change event. Using the same inundation-to-damage relationships this means an average annual damage of £90k to £95k and total benefits of £2,400,000 for 100 years. This is a more substantial sum to justify alleviation works but the geography of the catchment makes it very difficult to get sufficient area of sites close to Aboyne since the river is much incised. This will limit the flow that can be captured by any interventions reducing effectiveness and thus the budget for the intervention.

Beyond the (largely) capital costs considered by the cost-benefit analysis is the issue of acceptability, and this will depend on the degree of compensation available for loss of income or provision of additional services. Table 2 shows a simplified analysis of the “value” of the main land uses present in the Tarland catchment.

<table>
<thead>
<tr>
<th>Land Use</th>
<th>Land Cover</th>
<th>GM/ha</th>
<th>Subsidy/ha</th>
<th>All/ha</th>
<th>% Use</th>
<th>£ per Use</th>
<th>% damage</th>
<th>Opp cost/y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finished Cattle</td>
<td>Grass &lt; 5 y</td>
<td>£155</td>
<td>£224</td>
<td>£379</td>
<td>24%</td>
<td>£3,498</td>
<td>25%</td>
<td>£874</td>
</tr>
<tr>
<td></td>
<td>Grass &gt; 5 y</td>
<td>£124</td>
<td>£179</td>
<td>£303</td>
<td>8%</td>
<td>£932</td>
<td>25%</td>
<td>£233</td>
</tr>
<tr>
<td>Store Cattle</td>
<td>Grass &lt; 5 y</td>
<td>£74</td>
<td>£224</td>
<td>£297</td>
<td>24%</td>
<td>£1,661</td>
<td>25%</td>
<td>£415</td>
</tr>
<tr>
<td></td>
<td>Grass &gt; 5 y</td>
<td>£59</td>
<td>£179</td>
<td>£238</td>
<td>8%</td>
<td>£442</td>
<td>25%</td>
<td>£111</td>
</tr>
<tr>
<td>Hill Sheep</td>
<td>Rough Grazing</td>
<td>£12</td>
<td>£14</td>
<td>£26</td>
<td>10%</td>
<td>£112</td>
<td>10%</td>
<td>£11</td>
</tr>
<tr>
<td>Cropping</td>
<td>Spring Barley</td>
<td>£317</td>
<td>£200</td>
<td>£517</td>
<td>12%</td>
<td>£3,639</td>
<td>25%</td>
<td>£910</td>
</tr>
<tr>
<td></td>
<td>Spring Oats</td>
<td>£307</td>
<td>£200</td>
<td>£507</td>
<td>12%</td>
<td>£3,528</td>
<td>25%</td>
<td>£882</td>
</tr>
<tr>
<td></td>
<td>Winter Barley</td>
<td>£443</td>
<td>£200</td>
<td>£643</td>
<td>1%</td>
<td>£510</td>
<td>25%</td>
<td>£128</td>
</tr>
<tr>
<td></td>
<td>Winter Wheat</td>
<td>£442</td>
<td>£200</td>
<td>£641</td>
<td>1%</td>
<td>£508</td>
<td>25%</td>
<td>£127</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>£14,830</td>
<td>£3,691</td>
<td></td>
</tr>
</tbody>
</table>

The income figures are 10 year average gross margins (GM) per ha per year. GM exclude rent, labour and other fixed costs but are a generally accepted metric for the income generating potential of land. For grassland there needs to be assumptions made on the nature of the livestock system and two alternatives are used (finishing and store cattle with a 50:50 mix). Note the significant proportion of income derived from subsidy that need not be affected (or could even be enhanced) by the use of the land of temporary storage of flood waters. The mix of land uses affected by a real scheme would depend on the fields chosen but for this analysis we are assuming the 95 ha has the land use/cover mix identified for all the basins. This gives a maximum opportunity cost to land managers (assuming loss of all GM income) of £14,830 per annum. The real opportunity costs will be less since frequency and degree of inundation will vary. This is reflected in a simple damage function (% damage per year). These initial estimates of damage are speculative but they do provide a

---

Note that in both cases the estimates are highly sensitive to the assumptions made on discount rates for future costs. Particularly when dealing with climate change historical assumptions on the appropriate rates may no longer hold, see Stern [2009].

[Scottish Agricultural College, 2009]
first order estimation of real opportunity costs ~£3,700p/a. It is important to note that opportunity cost is less than 5% of the annual damage estimated for Aboyne, but that particular businesses may suffer larger business viability effects beyond the loss of GM income for particular fields. This estimation opens up the possibility of cost effective public or private (insurance based) compensation since it is significantly cheaper to compensate farmers than house owners. The wider framing of the issues and the acceptability of the interventions is further discussed in the results of the Issues framework below.

2.5 Issues Framework

The sociological baseline, the adviser and authority workshop, the farmer and factor workshop and informal meetings have revealed flood alleviation using natural approaches as a complex and contested issue. There are strongly held and basic disagreements on what flooding is, what causes flooding and who should deal with it. Flooding to land managers is a drainage issue – particularly the restriction placed on them from dredging or deepening ditches. This is now a controlled activity and there is a presumption against as this can be detrimental to protected species of fish and invertebrates covered by the EU Habitats Directive. However, even were it not prohibited the increasing the capacity of the channel simply moves the problem elsewhere and indeed may increase the speed to and consequently size of the peak flow. This is recognised but seen as an issue of mismanagement elsewhere. There is particular criticism of the failure by the planning authority (Aberdeenshire Council) to prevent (in the past) development of housing on sites likely to be flooded. These sites may, however, have been preferentially sold precisely because of their wetness. The idea of farmers as water managers does however have strong acceptance. There is an idea that they have always been water managers, responding to previous policy and public pressures to increase production by draining and otherwise improving land. There is perhaps a growing recognition that in addition to accepting regulation to ensure negative externalities are avoided there is the potential to argue for mechanisms (market, insurance or publicly funded) that reward the provision of an ecosystem service of flooding alleviation.

3 DISCUSSION AND CONCLUSIONS

The Interreg ideals and evaluation criteria provide a useful framework against which to assess progress of projects that seek to achieve outcomes in terms of improving the sustainability of land and water management. These have shaped the Aberdeenshire Aquarius project and there has been significant progress made. There are, however, serious challenges that remain. In terms of innovation, Interreg provides an incentive for competent authorities to go beyond standard best practice, indeed this is a prerequisite for Interreg funding. This provides and supportive environment to demonstrate and test research based approaches and to refine these so that they can fit with the reality of the situated internal practice of the non-research partner(s) particularly in terms of project timelines and priorities. The Interreg funding model of 50:50 can, however, be challenging for some research organisations with limited matching budgets. Inclusion through partnerships and deliberative or participatory processes with direct and indirect stakeholders is highly rewarding by ensuring salience, building credibility and enhancing legitimacy. There is a significant cost, however, in the overhead of communication, both in learning about issues beyond a disciplinary specialism and in defining project governance. The opportunity to undertake action-research though implementing pilot or full-scale infrastructure gives researchers a unique opportunity to in effect experiment with large scale eco-social systems that would normally beyond the scope of academic research.

The expectation of linkage with previous and existing projects was easily met for the Aquarius project since the work builds on 3DeeVision and Interreg IIIIB project addressing

---

9 Indeed one could make the case that all the existing land that is inundated during flooding events is a service provided at no cost to householders/local authorities. If the land were further protected from inundation then flooding would be much worse in settlements.
diffuse pollution issues in a variety of circumstances. This linkage does, however, depend on the maintenance of institutional memory in key individuals and this can be challenging in the face of staff turnover, which can be more rapid in agencies and local government than in some academic organisation. Aquarius is seen as leading the way in generating a Scottish evidence base for natural flow management policy at both local and national government level and has fed back into the research agenda and funding – the locality criteria for Interreg. A particular challenge is the latitude criteria – for strong transnational partnerships. Since many of the issues that NSR Interreg seeks to address require interregional cooperation (e.g. marine pollution) trans-nationality in delivery is now insisted on. Transnational exchanges between researchers, authorities and stakeholders have proven highly influential in Aquarius but insistence on shared ways of working and common outputs in terms of guidance and best practice have been divisive and unproductive so latitude needs to be considered carefully. The longitude criterion has largely been covered by the ideal of inclusion but the emphasis here is on cross sectoral partnerships. As noted above these partnerships have the potential to be highly productive but do require significant upfront investment in team building. The final criterion of legacy is often seen as difficult to deliver since it aims for outcomes beyond the lifetime of the project. As researcher’s influence on continuing practice and policy is increasingly valued by some research funders and Interreg provides and explicit opportunity to invest time in ensuring that these sustainability outcomes occur.

REFERENCES

Matthews, K. B., Buchan, K., Miller, D. G., Renwick, A., and Barnes, A., Comparison of SAF and JAC holding/business populations. Research study to assess to what extent data from the Single Application Form could be used to meet the statistical requirements of the June Agricultural Census, Report, Scottish Government, 2008a.
Matthews, K. B., Rivington, M., Buchan, K., Miller, D. G., and Bellocchi, G., Characterising and communicating the agro-meteorological implications of climate change scenarios to land management stakeholders, Climate Research 35(1), 59-75, 2008b.
GIS-based Spatial Hydrological Zoning for Sustainable Water Management of Irrigation Areas

Yun Chen*, Jianyao Chen, Emmanuel Xevi, Mobin-ud-Din Ahmad, Glen Walker
CSIRO Land and Water, Canberra, Australia (yun.chen@csiro.au)

Abstract: This study identifies hydrological zones in terms of their hydrologic suitability for sustainable irrigated agricultural development and management in the Murrumbidgee Irrigation Area (MIA) of Australia. Spatial data, including soils, groundwater level and salinity, recharge to watertable, as well as aquifer hydraulic properties were analysed within a GIS framework. Critical threshold values used in zoning process were defined based on experts’ knowledge and literature review taking into account significant issues in irrigation management, e.g. root-zone depth, water quality for crops to maintain a sustainable yield, and target recharge for rice industry. Spatial datasets were processed, integrated and analysed in ArcGIS. An integrative spatial modelling approach was applied for delineating hydrological zones. The results can be used to manage landuse and irrigation practices to reduce accesses to the watertable thus minimising the risk of waterlogging and salinisation due to rising water table levels. Landuse data were incorporated to reveal irrigation occurred in each zone. Recharge potential maps under various landuses were also combined with the zones to determine areas that are well-suited for irrigation without incurring a high risk of waterlogging and salinisation. The analysis provides hydrological indicators to assess current hydrologic suitability, landuse, and potential opportunities for improvement and expansion of irrigated areas in the MIA.

Keywords: Salinity; Waterlogging; Suitability; Groundwater; Aquifer; MIA

1. INTRODUCTION

Waterlogging and salination of irrigated land are major obstructions to the sustainability of agriculture. They are eroding the valuable irrigated croplands and posing a life-threatening to Australia’s food security (National Land and Water Resources Audit, 2001). As one of the most evident adverse effects of irrigation development, the prime cause of these natural hazards is the rising watertable due to inadequate drainage, although strong climatic aridity is a pre-condition for salinity (Smedema, 1990). A combination of waterlogging and salinity stresses is a severe threat to crop growth, development and yield in arid and semiarid regions like the Murrumbidgee Irrigation Area (MIA). There is an urgent need to spatially define areas of varying suitability for irrigation to avoid salinity and waterlogging for a sustainable agricultural production. Therefore, the objective of this study is to develop a GIS modelling approach to classify these areas as zones with similar hydrological characteristics suitable for irrigation and other land management practices in order to reduce the risk of salinity and waterlogging in the MIA.

2. STUDY AREA

Situated in south-west New South Wales of Australia, the MIA is about 3,120 km² with its major town centre in Griffith (Figure 1). The area is a flat open plain with an elevation ranging from 100 to 135m. There are approximately 795km of supply and drainage channels, whilst Mirrool Creek and the Murrumbidgee River cover a distance of 255k
within the boundary of MIA (Khan and Abbas, 2007). The annual average rainfall and pan evaporation at Griffith are 406mm and 1797mm, respectively (Bureau of Meteorology, 2006).

The MIA is located in the Riverine Plain associated with the Murray Geological Basin. The aquifers in the area consist of semi-consolidated to unconsolidated sedimentary deposits. There are three major aquifer systems. The surface Shepparton Formation (late Pliocene-Pleistocene) is sandy-clays with some local areas of poorly sorted sands. It can be further subdivided into two layers: the Upper Shepparton and the Lower Shepparton. The late Miocene-Pliocene Calivil Formation is in the middle. It contains a high proportion of coarse quartz sand with some lenses of clay. This formation is generally the most productive aquifer system. The deepest aquifer is the late Eocene-early Miocene Renmark Group. It is predominantly clay but comprise extensive areas of medium-grained quartz sand. The regional hydraulic gradient is from east to west, aligned with the surface flow system (Xevi, et al., 2010). Figure 1 shows a 3D conceptual groundwater flow system in the MIA. Each flow system has an area of recharge, an area of through flow or transfer, and one of discharge. The zonation of these areas is basically determined by the geomorphology of the aquifers and by the direction of hydraulic gradient. Most of the Upper Shepparton is less than 20m thick. The Lower Shepparton ranges from close to zero thickness in an area in the northwest to 60m thick in the southwest. The majority of Calivil is between 40 and 70m. The thickness of the Renmark ranges between 200m thick in the southwest to close to zero in the north east where the aquifer is essentially non-existent.

3. METHODS

3.1 Spatial Data

The hydrological zones are defined in terms of their hydrologic suitability for sustainable irrigated agricultural development and management. Spatial datasets of the above variables were extracted within the framework of GIS to develop hydrological zones in the MIA. Digital elevation model (DEM) at a 50m resolution (Figure 2a) indicates a potential occurrence of the local groundwater flow. There are five major soil types (Figure 2b): Red Brown Earth (RBE); Self Mulching Clay (SMC) Non Self Mulching Clay (NSMC). Xevi et
al. (2009) provides more detailed description of soil classification and mapping. There is an intensive network of about 864 piezometers across the region. The active piezometers are monitored twice a year in March and September representing pre- and post-rainfall groundwater levels. Datasets of depth to groundwater for September 2002 and September 2006 were selected to represent “dry” and “wet” years, respectively. Two maps showing groundwater tables (the difference between ground surface and piezometric heads) were generated based on about 800 piezometer readings recorded for each year (Figure 2c and 2d). The maps were interpolated from the piezometric data points using an inverse distance weighted (IDW) technique (Child, 2004). Considering that the sampling points were sufficiently dense with regard to the local variation we attempt to simulate, the IDW method is preferred in order to obtain the best representation of the desired surfaces. Hydraulic gradient (HD) was calculated using the observed water level in September 2002 for the Upper Shepparton Formation (Figure 2e). In line with surface flow system, regional hydraulic gradient is from east to west. Figure 2f shows the spatial pattern of vertical hydraulic conductivity \(k_v\) of the Upper Shepparton derived from soil characteristics. Electrical conductivity (EC) of groundwater was observed for two different periods, February 1998 and July 2002. As irrigation is dominant in the summer season (October to March), EC of July 2002 for the Upper Shepparton Formation was interpolated using the IDW technique and the resultant map is given in Figure 2g. Transmissivity (T) is a combination of horizontal hydraulic conductivity \(k_h\) and aquifer formation thickness. An averaged T map (Figure 2h) was derived by multiplying \(k_h\) with the weighted aquifer thickness of each formation (Upper Shepparton, Lower Shepparton and Calivil).

Figure 2. (a) DEM with 50m resolution for MIA. (b) Soils. (c) Depth to watertable (m) - September 2002. (d) Depth to watertable (m) - September 2006. (e) HD of Upper Shepparton - September 2002. (f) \(k_v\) of Upper Shepparton. (g) EC (dS/m) - July 2002. (h) Average T (m²/d) of Upper and Lower Shepparton, and Calivil.
Figure 3 is the available potential groundwater recharge maps for different combinations of landuses and soils in the region under 2002/03 climatic conditions (Xevi, et al., 2010). The maps show recharge/discharge that will occur under each land use, assuming that a particular land use type occupies the entire area of the MIA. Net groundwater recharge was obtained with positive values indicating water entering the saturated zone and negative ones representing water being withdrawn, i.e. evaporation, from the saturated/unsaturated zone.

### 3.2 Zoning Approach

Spatial datasets were processed, integrated and analysed in ArcGIS environment. All data layers were projected into WGS84 UTM Zone 55S coordinate system. Benchmark thresholds for determining risks and suitability for productive and sustainable agriculture were established. An integrative spatial modelling approach was developed for the purpose of delineating hydrological zones and landuse suitability in the MIA (Figure 4). Table 1 lists three critical threshold values of groundwater table, groundwater quality (EC) and recharge used in this process. The threshold value for water table was set taking into account the fact that the root zone of most crops is less than 2m depth. An EC value smaller than 8dS/m is regarded as safe for most crops to maintain a sustainable yield, as well as taking into account the salt tolerance for most native vegetation. The threshold value for recharge was derived statistically. A histogram of recharge values for nine different land uses, or crops, with combinations of five soil types and seven groundwater levels in the period 2002-2003 output from SWAGMAN modelling simulation were plotted. The value of 1ML/ha which occurs most frequently was adopted. In addition, Dwyer Leslie Pty Ltd (1992) also recommended the adoption of a rice industry target recharge figure of 1ML/ha (100mm).

Water depth and EC are two key factors which play significant roles in the zoning process. A general declining trend in the watertable of less than 2m depth was found in the MIA due to drought and decreased area grown to rice (Figure 5a and b). However, the plotted area in 2006 was much smaller than in 2002. Figure 5c indicates changes in watertable between the two years. Figure 5d presents areas with EC values greater than 8dS/m in 2002.

![Figure 3. Recharge maps for various landuse types (2002/2003). Values > 0 represent recharge, and < 0 discharge/evaporation.](image-url)
Groundwater flow and aquifer hydraulic properties are also critical in defining the zones. HD, k_h and k_v, as well as formation thickness in the study area are relatively constant and independent to changes in watertable and the other water balance components (Xevi, et al., 2010). These aquifer characteristics were used to derive a new factor called flux:

\[
\text{Flux} = T \times H = k_h \times \text{Thickness} \times HD = V \times \text{Thickness}
\]

where V is Darcy’s velocity (m/d), and the product of V with thickness is a kind of flux per unit width (m²/d).

Although salinity and water table fluctuation in the Upper Shepparton Formation significantly affect salinisation and crop yield via capillary rise, the Lower Shepparton and Calivil Formations could also have impact on these issues through vertical connectivity, e.g., pumping and return flow. This is especially important to prolonged irrigation fields, such as MIA. In order to incorporate the influence of vertical connection among various formations, an average T of three layers, rather than T of the Upper Shepparton only, was used in equation 1. The flux map was intersected with the k_v. Three classes were interpreted and linked to the drainage capacity (Table 2) with the help of DEM (Figure 6a). Areas where watertable is always high (≥ 2m) in both 2002 and 2006, or watertable is high (≥ 2m) and EC > 8dS/m in 2002, were mapped as severe waterlogging areas (Figure 6b). The map of severe waterlogging areas was then used as a mask to identify a new class in addition to Figure 6a. The hydrological zones were then generated as shown in Figure 6c.

**Table 1.** Definition of threshold values for critical factors.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Threshold</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water depth</td>
<td>2m</td>
<td>≤ 2m is regarded as shallow watertable</td>
</tr>
<tr>
<td>EC</td>
<td>8dS/m</td>
<td>&gt; 8dS/m is regarded as saline watertable</td>
</tr>
<tr>
<td>Recharge</td>
<td>1ML/ha</td>
<td>≤ 1ML/ha is regarded as low recharge, and &gt; 1ML/ha high recharge</td>
</tr>
</tbody>
</table>

**Figure 4.** Flowchart of methodology for hydrological zoning.

**Figure 5.** (a) Areas with depth to groundwater table ≤ 2m (September 2002). (b) Areas with depth to groundwater table ≤ 2m (September 2006). (c) Changes of watertable between 2002 and 2006. (d) Areas with EC > 8dS/m (July 2002).
4. RESULTS AND DISCUSSIONS

Four hydrological zones (Figure 6c) derived from GIS-based spatial modelling provide critical management zones to define hydrologic suitability for sustainable irrigated agriculture in the MIA. It is useful for managing key issues in water resource management, e.g., salinity and waterlogging. Irrigation development may proceed in the good and intermediate zones with reduced risk of waterlogging and salinisation. Irrigation development in the poor zones poses a high risk for waterlogging and salinisation.

Recharge maps shown in Figure 3 were incorporated into the map of hydrological zones (Figure 6c) using the criteria listed in Table 3. The results of the integration of these maps (Figure 7) provide hydrological indicators to assess current hydrologic suitability and land use, as well as potential opportunities for improvement in MIA. The maps in Figure 7 actually represent the hydrologic suitability of each landuse type over the entire MIA. In other words, if there is a choice to develop irrigated land within MIA, which areas are most and least suited to a particular land use type considering the risk of waterlogging and salinisation. The resultant maps also allow managers to look at recharge threshold for different crops, and to optimise crop combinations in each zone, so as to apply different management strategies to achieve sustainable development and ecological benefits of land and water resources.

Table 2. Criteria used to define drainage capacity.

<table>
<thead>
<tr>
<th>Class</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low drainage capacity</td>
<td>flux &lt;= 4 m²/d (or k is very low)</td>
</tr>
<tr>
<td>Medium drainage capacity</td>
<td>4 m²/d &lt; flux &lt; 42 m²/d</td>
</tr>
<tr>
<td>High drainage capacity</td>
<td>Flux &gt;= 42 m²/d</td>
</tr>
</tbody>
</table>

Figure 6. (a) Drainage capacity map showing groundwater flow and aquifer properties. (b) Areas with severe water logging. (c) Hydrologic suitability map.

Table 3. Criteria used to define hydrologic suitability of crops.

<table>
<thead>
<tr>
<th>Zones</th>
<th>High Recharge (≥1 ML/ha)</th>
<th>Low Recharge (&lt;1 ML/ha)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good Drainage</td>
<td>Suitable</td>
<td>Suitable</td>
</tr>
<tr>
<td>Intermediate</td>
<td>Intermediate</td>
<td>Intermediate</td>
</tr>
<tr>
<td>Poor Drainage</td>
<td>Unsuitable</td>
<td>Unsuitable</td>
</tr>
<tr>
<td>Waterlogging</td>
<td>Unsuitable</td>
<td>Unsuitable</td>
</tr>
</tbody>
</table>
Figure 7. Crop suitability maps.

The land use map of 2000/2001 season (Figure 8a) representing the wettest year (for the project period) was overlaid with the hydrological zones (Figure 6c) to generate Figure 8b. This is to assess the hydrologic suitability of the land use in the wettest year and the potential risk for waterlogging and salinisation as a result of irrigation development in the MIA. The analysis revealed that about 45% of irrigated agriculture occurred in “good” and “intermediate” zones; 13% was in the “good” zone, while approximately 32% was in the “intermediate” zone. There was 48% of total area of irrigated crops in the “poor” zone and the rest 7% happened in waterlogging area. Irrigation in the “good” zone should be retained since limitations to irrigated cropping in this zone can be overcome by standard irrigation management practices. These areas can be modernised for irrigation agriculture to improve the water use efficiency through land and water management plans. Limitations to irrigated agriculture in the intermediate zone need to be recognised because a decline in productivity caused by waterlogging and salinisation may occur over time and a range of land use problems may also develop if this land is used and managed inappropriately. There may be a need for irrigated land retirement from the poor zone areas. Irrigation practice in severe water logging zone should be restricted and consider for potential retirement. Finally, we believe a large area of good lands distributed in northwest of the MIA has a great potential to be intensified as irrigated agricultural land. We emphasise that our analysis was conducted without considering all factors needed to produce a comprehensive land suitability map. Such analysis will also require, among others, agronomic factors such as soil fertility and other morphological factors of the landscape.

Figure 8. (a) Irrigated land use for 2000/2001 season. (b) Irrigated crops in hydrological zones (2000/2001).
5. CONCLUDING REMARKS

We defined a hydrological zone as a component of a region with similar hydrological characteristics. This, in turn, determines the suitability for managing land use and irrigation practices to reduce accessions to the water table thus minimising the risk of water logging and salinisation due to rising water table levels. The hydrological zones were developed using spatial distribution of soils, groundwater characteristics and aquifer hydraulic properties. The suitability maps derived are valuable for discussions about the impact of recharge, land use and groundwater salinity on irrigated agriculture. They provide a blueprint for identifying areas that require a different management strategy in terms of land use change, or improvement in irrigation technology, or large scale waterlogging and salinity mitigation schemes. When the approach is implemented for actual use, cares should be taken regarding human intervention and the range of parameters that were applied for the definition of hydrological zones, e.g., unsuitable land could be developed or irrigated for crops or vine with high economic value by installing tile in the poor drainage areas. Results from this study will be used to aid the planning of management guidelines concerning current irrigation practices as well as future irrigation development in the MIA, such as the EnviroWise program which aims to maintain and enhance the sustainability of farming, rural industries and associated communities by funding from both the Federal and State Governments. The incorporation of such information into the program will support the protection and enhancement of the region’s natural resources, and to ensure that some key objectives of the EnviroWise program, that is to maintain or increase irrigated agriculture productivity and to keep drainage water quality within agreed standards, are being met.

ACKNOWLEDGMENTS

Data and funding from Murrumbidgee Irrigation Limited is gratefully acknowledged. The assistance of Heinz Buettikofer with mappings, Jason Carroll and James Foley with data collection and collation are much appreciated. Input from internal reviewers within CSIRO Land and Water helped improve the manuscript.

REFERENCES

Khan, S., and A. Abbas, Upscaling water savings from farm to irrigation system level using gis-based agro-hydrological modelling irrigation and drainage, Irrigation and Drainage, 56, 29-42, 2007.
Modelling Urban Land Use Change Using Geographically Weighted Regression and the Implications for Sustainable Environmental Planning

Noresha Mohd Shariff *; Sanjay Gairola; Anita Talib

School of Distance Education, Universiti Sains Malaysia, 11800 Minden, Penang, Malaysia
Centre For Marine and Coastal Studies, CEMACS, Universiti Sains Malaysia, 11800 Minden, Penang, Malaysia

*noreshah@usm.my, drsanjaygairola@gmail.com, anita@usm.my

Abstract: In this study we applied Geographically Weighted Regression (GWR) approach to model urban land use changes in Penang Island from 1990 to 2005, covering the period during which the Island has experienced tremendous urban growth due to in migration from adjacent areas. Land use change has potential impacts on the physical and social environment. We identified spatial variables describing environment, physical and socio-economic factors which are hypothesized to influence the change in the land use in the study area. An ordinary least squares regression (OLS) model is applied to the variables followed by a GWR model and the results are compared. The results show that the GWR outputs explained considerably more variance in the relationship of the explanatory factors compared to conventional OLS models and provided significantly better results. In addition, GWR also provided important insights on location where changes happen and what distance to the city center they appear. The information generated will give understanding of spatio-temporal dynamics of land use changes resulted from different land use policies and can serve as a basis for developing possible growth scenarios which are essential for sustainable urban planning and development. However, more comprehensive studies are needed to understand long term spatio-temporal patterns and complex inter-related challenges the urban areas of Penang Island facing presently.

Keywords: Urban Land Use Change, Sustainable Environmental Planning, Geographically Weighted Regression Modelling

1. INTRODUCTION

With recognition of the ecological, socio-economic, and cultural significance of urban areas and their sensitivity to rapid urbanization, greater emphasis is now placed on the issue of sustainability in urban development. Presently, urbanization has become one of the main factors of land degradation and resulting losses of non-urban land uses worldwide. Researchers emphasized that these changes of non-urban land uses however certainly provide many social and economic benefits, but have adverse effects on natural environment (Tang et al., 2005) including global carbon cycle, climate, biodiversity, and landscape ecology (Houghton et al., 1999; Luvall, 1997; Sala et al., 2000; Reid et al., 2000; Wickham et al., 2000). The urban areas are recognized as one of the complex and highly dynamic landscapes on earth surface which supports more than half of the global human population, as well as hubs of the worlds manufacturing and service industries (Kaplan et al., 2004). In spite of only 3% coverage of the Earth’s land surface, the urban areas are reported to exert marked effects on environmental conditions at both local and global scales (Grimm et al., 2000; Herold et al., 2003; Liu & Lathrop, 2002). In view of the fact that ecosystems in urban areas are strongly influenced by anthropogenic activities, considerably more attention is currently being directed towards monitoring urban land use changes (Stow & Chen, 2002). The spatially explicit modeling of land use changes is an important technique for describing processes of changes in quantitative terms and for testing our understanding of these processes.
(Serneels and Lambin, 2001). Significance of these modeling techniques to understand the urban development process has been highlighted (Schneider and Gil Pontius Jr. 2001; Walsh and Crawford 2001). Moreover, urban land use change models can generate alternative landscape predictions on the basis of different land use policies and environmental constraints and predictions about future urbanization which are critical to the protection of ecosystems and the sustainability of communities (USGS, 2009). This necessitates the linking of the resulted changes to their driving factors for effective land use planning and sustainable management of resources. The driving factors (e.g. population or development), mediated by the socio-economic setting (e.g. market economy, resource institutions) and influenced by the existing environmental conditions or context, lead to changes in land use through the manipulation of the biophysical conditions of the land (Turner et al., 1995). Presently sustainable urban development became a widely recognized goal for human society and interest has burgeoned among urban planners and researchers to maintain sustainable urban environment and devise robust modeling techniques. The theoretical and mathematical models have for long been created for purposes of urban studies, aiming at clarifying processes of urban and regional change. However, currently a new wave of urban spatial modeling for understanding urban environment has come to the force and becoming an integral topic in current urban research agenda.

The Geographically Weighted Regression (GWR) approach of spatial modeling is an important part of these tools which provide technique to deal with spatial non-stationarity in multivariate regression and estimates regression coefficients locally using spatially dependent weights (Fotheringham et al., 1997). GWR is becoming a more commonly used technique in urban geographical and environmental studies as the important feature of GWR is its ability to generate parameter estimates for every regression point by using observations in a given neighborhood. The parameter estimates are characteristically mapped to highlight spatial variation (Mennis, 2006) and resulting maps are thought to be didactic aids for policymakers, and for summarizing the large amount of data generated by the procedure (Cho et al., 2009). More detail about GWR can be found in Fotheringham et al. (2002) and some other recent articles (e.g. Wang et al., 2005; Chang et al., 2008; Propastin et al., 2008). The GWR has been applied to investigate regional industrialization (Huang and Leung 2002), geographic diversity in urban and regional growth (Partridge et al. 2006), commuting patterns (Lloyd and Shuttleworth 2005), modeling urban spatial structure (Noresah and Rainis, 2009) and forecasts of regional employment (Li et al., 2009). Keeping the wide applications of GWR in background, this study aims at modeling urban land use changes using GWR approach over Penang Island, Malaysia. The spatial variable of relationship between the urban land use change and the proximate causes will give the idea of complexity and interconnections between the land use change and associated factors.

2. STUDY AREA

The Penang state is one of the most rapidly developing and industrializing states in Malaysia. It is located on the north-eastern region of Peninsular Malaysia (Figure 1) and consists of the Island of Penang and a coastal strip on the mainland known as Seberang Perai (or Prince Wellesley). The Penang Island is situated in the northern part of Malaysia and geographically situated between 5°12’ to 5°30’ North latitude and 100°09’ to 100°26’ East longitude. Being the most populated Island in the country; Penang Island has a population of about 745,000 in the year 2009 with 293 Km² coverage areas. The terrain of the Island is mainly represented by coastal plains, hills and mountains with much developed lowland areas. The coastal plains of the Island are narrow, the most extensive of which is in the northeast where the state capital Georgetown is situated. The elevation ranges from zero to 830m and the climate is equatorial humid type. The study area is located in the eastern part of Penang Island. This area is more densely populated, urbanized and industrialized as compared to most part of the Island.

3. DATA AND METHODOLOGY

Land use and population data used in this study are obtained from the Urban Planning Department of Malaysia.
The main data required for the model include location and amount of urban land use change from 1990-2005 (Figure 2), access factors (Figure 3), physical constraints (protected areas: forest, water body) (Figure 4) and social factors (population, education) (Figure 5), which are extracted and processed from primary and secondary sources. Urban land use change data is extracted from urban land use change map comprises of land that has been converted to residential, commercial, industrial, transportation and public institution uses for the period of 15 years. They are classified here as urban. Access factors include road networks, city centers/sub-center, employment centres, shopping centres and airport. Penang International Airport is located in the study area and is the only airport located on the northern region of Malaysia. Roads in Malaysia have been classified into hierarchy of expressway, highways, secondary, primary and residential roads. In this study, in order to reduce the uncertainty in classification, only expressway, highway and other roads are identified. The same classification applies to the city centers/sub-centers where the state capital, Georgetown is the major centre and other second order centres are minor centres. Forest or hilly areas takes major percentage of the Penang Island. Population data is available at sub-district level.

GWR incorporates the spatial structure of the data into the estimation of the regression model’s parameters and shows how those estimates vary across space. It also provides the researcher with an analytical tool to explore changes in the relationship between variables over space. In this present study, relationships between urban land use change as dependent variable and the independent variables are modeled using conventional ordinary least squares (OLS) and geographically weighted regression (GWR). Urban land use change model for the study area are developed using OLS and GWR tools. The amount of land use change between 1990 and 2005 is obtained by subtracting urban land area of 2005 to that of 1990. The urban land area of the study site has increased from 4958 hectares in 1990 to 6428 hectares in 2005, which shows an increase of 30 percent over the 15 year period. A total of 4584 square grids of 250m x 250m are generated covering the study area and a centroid of each grid is used as a reference point for spatial analysis and GWR modeling. The geographic location (i.e. x,y coordinates) of each grid centroid, information on urban change and independent variables are stored in an attribute table. Hypothesizing that proximity of the independent variables to the location of new growth influenced the change, GIS spatial analysis is carried out for each variable and stored as attribute in each grid. The land use change is mapped and intersected with the grid layer to calculate the amount of urban land use in each grid. Each grid is equivalent to 6.25 hectares. The proximity between geographic location of each grid and the spatial factors is calculated using Euclidean distance. GWR coefficients and local R² values are mapped to explore spatial variability of relationships between explanatory variables and the land use change. Finally, an F test is used to determine whether the GWR estimates are a significant improvement on the traditional globally estimated OLS and results are compared based on Akaike Information Criterion (AIC). Lower values of AIC indicate a closer fit to the data.
Figure 2. Variables used in the analysis: (a) Urban Growth from 1990 to 2000, (b) Access Factors, (c) Physical Constraints, and (d) Population Concentration 1990

Table 1. Variables used in present study

<table>
<thead>
<tr>
<th>Variables</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent variable:</td>
<td></td>
</tr>
<tr>
<td>URB</td>
<td>Amount of urban land growth 1990-2000 (in hectare)</td>
</tr>
</tbody>
</table>

Independent variables:

<table>
<thead>
<tr>
<th>Variables</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PExp</td>
<td>Proximity to nearest expressway</td>
</tr>
<tr>
<td>PGtown</td>
<td>Proximity to Georgetown</td>
</tr>
<tr>
<td>PMncity</td>
<td>Proximity to the nearest minor city centres</td>
</tr>
<tr>
<td>PHway</td>
<td>Proximity to nearest highway</td>
</tr>
<tr>
<td>PAirpt</td>
<td>Proximity to nearest airport</td>
</tr>
<tr>
<td>P Educ</td>
<td>Proximity to educational institutions</td>
</tr>
<tr>
<td>PMjrds</td>
<td>Proximity to nearest major road</td>
</tr>
<tr>
<td>P Fores</td>
<td>Proximity to nearest forest reserve</td>
</tr>
<tr>
<td>P Popctr</td>
<td>Proximity to population concentration centres</td>
</tr>
<tr>
<td>P Cont120</td>
<td>Proximity to contour of 120 meters</td>
</tr>
<tr>
<td>PAvln</td>
<td>Proximity to land available for development in 1990 (ha)</td>
</tr>
</tbody>
</table>
4. RESULTS

Results reveal that GWR models exhibited a significant improvement in explained variance as compared to the OLS regression models. The AIC score for GWR model decreased from 13586.1 to 13537.5 which reflect better goodness of fit than the global OLS (Table 1). AIC is a measure of spatial collinearity within the model data. The lower is the value of AIC; the better the fit is the model to observed data. This suggests that GWR model for Penang Island is better than the OLS model based on the AIC.

Table 2. Results of ANOVA test for GWR over the OLS urban land use change models

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>DF</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS Residuals</td>
<td>5169.7</td>
<td>12.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GWR Improvement</td>
<td>285.9</td>
<td>103.09</td>
<td>2.7732</td>
<td>2.5375</td>
</tr>
<tr>
<td>GWR Residuals</td>
<td>4883.9</td>
<td>4468.91</td>
<td>1.0929</td>
<td></td>
</tr>
<tr>
<td>GWR Akaike Information Criterion: 13537.5 (OLS: 13586.1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The summary results of the Global OLS and GWR models are presented (Table 3). The results of Global OLS models suggest that urban land use is positively related to predictor variables but the high amount of variation remains unexplained. Moreover, the low r-square (0.36) suggest that approximately 40% of variance of the urban land use growth in the study area can be explained by the explanatory variables whereas 60% of the variance still remains unexplained. The t statistics of the estimated parameters revel that only 4 variables are statistically significant and explaining the variation in urban land use change. GWR models on the other hand explained about 60% of the variance and all the variables are statistically significant in explaining the change in the urban land change in the Penang Island. The Monte-Carlo test shows that all the predictor variables displayed significant non-stationarity and indicating spatial variation in the relationship between urban land use change and predictors variables. The intercept also showed significant non-stationarity in GWR model. Based on these results, it can be inferred that modelling this relationship with Global OLS regression attains with high amount of uncertainty.

Table 3. Summary results of the Global OLS and GWR model

<table>
<thead>
<tr>
<th>Variables</th>
<th>β</th>
<th>t</th>
<th>p-value ¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.20</td>
<td>0.86</td>
<td>0.00***</td>
</tr>
<tr>
<td>PUrb90</td>
<td>-0.00</td>
<td>-0.36</td>
<td>0.00***</td>
</tr>
<tr>
<td>PForest</td>
<td>0.00</td>
<td>1.42</td>
<td>0.00***</td>
</tr>
<tr>
<td>PItown</td>
<td>0.00</td>
<td>2.24***</td>
<td>0.00***</td>
</tr>
<tr>
<td>PMnrcty</td>
<td>-0.00</td>
<td>-1.83</td>
<td>0.00***</td>
</tr>
<tr>
<td>PExp</td>
<td>-0.00</td>
<td>-1.90</td>
<td>0.00***</td>
</tr>
<tr>
<td>PHway</td>
<td>0.19</td>
<td>-2.23***</td>
<td>0.00***</td>
</tr>
<tr>
<td>PAirpt</td>
<td>-0.00</td>
<td>1.07</td>
<td>0.00***</td>
</tr>
<tr>
<td>PPopct</td>
<td>-0.00</td>
<td>-2.85***</td>
<td>0.00***</td>
</tr>
<tr>
<td>PEduc</td>
<td>-0.00</td>
<td>-0.52</td>
<td>0.00***</td>
</tr>
<tr>
<td>PAvland</td>
<td>-0.00</td>
<td>3.95***</td>
<td>0.00***</td>
</tr>
<tr>
<td>PCnt120</td>
<td>-0.00</td>
<td>-0.30</td>
<td>0.00***</td>
</tr>
<tr>
<td>N=</td>
<td>4584</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted r-squared</td>
<td>0.36</td>
<td>0.68</td>
<td></td>
</tr>
</tbody>
</table>

*** = significant at .1% level

¹ Results of Monte Carlo test for spatial non-stationarity (Fotheringham, 2002)
Figure 3 shows the spatial distribution of the intercept and all the parameters of the GWR model. Intercept term or constant parameter determines the basic level of urban land use change without the effects of all the factors across the study area (Huang and Leung, 2002). Local estimates of intercept coefficient ($\beta_0$) range from a minimum of -27.32 to a maximum of 20.13 with a median of 1.36 instead of a constant (0.20) for the global

Figure 3 (a-k). Spatial variation of the parameter estimates of GWR model.
regression. The result appears to show an apparent spatial variation in the constant parameter. All parameters are significant for GWR model and the parameter estimates are mapped to show the spatial variation. Figure 6 shows the spatial variation of the parameter estimates that are hypothesized to have influenced the change in the urban land use in Penang Island. Higher parameter estimate means that the effect of the variable is higher in that region as compared to other parts of the region (Huang and Leung, 2002). The darker is the shaded area the higher is the parameter estimates.

5. CONCLUSIONS AND RECOMMENDATIONS

The Penang Island has experienced tremendous urban growth since the opening of the Penang Bridge in 1989. This has resulted in-migration from the adjacent areas and the spatial structure of urban areas has changed with respect to diverse populace, morphology and their relationship to the core city. In order to improve our understanding of mechanism of these changes, in this present study we examined relationship between urban land use change and its driving forces. This relationship is tested using regression modelling approach by taking eastern region of Penang Island as a target study area because this area is densely populated, urbanized and industrialized as compared to most part of the Penang Island. The OLR and GWR models are developed to study the relationship between urban land use change and determinant factors. Results reveals that GWR models performed better and provide significant improvement over the global regression models. The global OLS models explained only about 40 percent of the variance in the urban land use change as compared to the GWR models which explained about 60 percent of the variance. This is because the GWR method has the advantage of providing local parameters estimates and reveals interesting pattern of spatial variation or non-stationarity of parameters. The spatial distribution of all parameters shows significant spatial variation with higher parameters in some parts of the region.

From sustainable spatial planning of view the present study is particularly important because the spatial characteristics of land use change are useful for understanding various impacts of human activity on the overall ecological condition of the urban environment (Yeh & Li, 1999). It has been widely accepted that the understanding of urbanization pattern and process at local scale is essential in guiding sustainable urban development. In Penang Island (target area) as per population demographic trends, the urbanization process is rapid and may undergo a high spatial restructuring process; therefore focus should be achieving sustainable urban landscapes by establishing sustainable equilibrium between ecological, social and economic functions of urban ecosystems. In order to accommodate growing urban population in a planned and sustainable manner, investigation of urban structure and morphology for planning proper infrastructure facilities becomes crucial. Using GWR approach there is a need of identification and spatio-temporal analysis of suburban areas as their connections with the core city is considered to be important. The spatial information of urban land use change is largely lacking for study area therefore the results of the present study can be utilized to develop urban growth scenario for forecasting possible future changes which may leads towards sustainable urban land use planning. However, further studies with large number of data would require for more reliable predictions about urban land use changes in the region.

REFERENCES


Informing Regional Planning in Alberta’s Oilsands Region with a Land-use Simulation Model

Matthew Carlson\textsuperscript{a}, Terry Antoniuk\textsuperscript{a}, Dan Farr\textsuperscript{a}, Shawn Francis\textsuperscript{a}, Karen Manuel\textsuperscript{a}, John Nishi\textsuperscript{a}, Brad Stelfox\textsuperscript{a}, Mika Sutherland\textsuperscript{a}, Cornel Yarmoloy\textsuperscript{a}, Craig Aumann\textsuperscript{b}, Daiyuan Pan\textsuperscript{b}

\textsuperscript{a}ALCES Landscape and Land-Use Ltd (mcarlson@alces.ca), \textsuperscript{b}Alberta Innovates-Technology Futures (Craig.Aumann@albertainnovates.ca)

Abstract: Planning for regional sustainability requires strategic understanding of ecological and socioeconomic trade-offs associated with alternative land use options. We discuss a scenario analysis being undertaken to assess trade-offs for a 93,000 km\textsuperscript{2} region in northeastern Alberta containing the world’s second largest oil deposit. Due to its immense economic and ecological value, the region presents both an opportunity and challenge for the objectives of sustainable prosperity and healthy ecosystems put forth by the Alberta government’s Land-Use Framework. ALCES\textsuperscript{\textregistered} simulation and mapping software are being applied to inform government planners and stakeholders about possible future outcomes associated with land-use options. Two characteristics of the modelling process in particular have contributed to its success: the comprehensive and integrative nature of ALCES; and reliance on government and stakeholders for model inputs and scenario development. The scenario analysis provides a case study to discuss the technical aspects of ALCES and the Alberta Land-Use Framework’s approach of facilitating learning through iterative scenario analysis.

Keywords: scenario analysis, cumulative effects, regional planning

1. INTRODUCTION

Environmental degradation has increased in frequency and intensity in recent decades due to expanding development and population growth. Responses to the environmental impacts have typically been reactionary and focused on specific symptoms rather than systemic drivers and effects (Reagan 2006). The result is fragmented government policy that is ill-suited to deal with environmental problems that are typically complex, multidimensional, and broad in spatial and temporal scale (Bellamy et al. 1999). Integrated resource management (IRM) was conceived to address the discordance between the complex problems yet simplistic management regimes. Rather than focusing on single developments or environmental issues, IRM seeks to manage all human activities in a system to balance a broad range of objectives (Cairns and Crawford 1991).

Practitioners of IRM are likely to encounter trade-offs when diverse environmental and socioeconomic goals come into conflict due to finite natural resources. An important component of IRM is scenario analysis to assess trade-offs associated with a range of contrasting but plausible assumptions about a region’s management regime and ecological
processes (Peterson et al. 2003). The scenario analysis should be broad enough in scope to consider the impact of all potentially influential anthropogenic and natural disturbances and to incorporate the broad spatial and temporal scales that define many ecological processes. Computer-based land-use simulation models are well suited for such an analysis due to their capacity to track potentially complex inter-relationships among numerous variables. Although incapable of predicting the future state of an ecosystem due to the uncertain nature of key drivers (e.g., human behaviour and natural disturbances), computer models can foster an understanding of how an ecosystem may respond to human actions by trialing management strategies prior to real-world implementation (Ford 1999).

The province of Alberta in western Canada is a poignant example of the benefits and challenges created by rapid and uncoordinated economic development. Over the past century, Alberta’s population has grown at a rate of 2.1% per year, driven first by the conversion of native prairie to agricultural production and then development of the abundant hydrocarbons of the Western Canadian Sedimentary Basin and harvest of the province’s expansive forests. Alberta has transformed into a jurisdiction whose export economy is of international importance, but rapid development has also generated environmental liabilities (Timoney and Lee 2001). The uncoordinated growth mandates of multiple resource sectors have impacted not only ecosystems but also each other, with land uses such as energy and forestry competing for a limited land base (Ross 2002).

In response to mounting concerns that existing management regimes may be inconsistent with long-term sustainability, the government of Alberta is embarking on an IRM planning exercise to identify land use strategies better suited to balance long-term ecological and socioeconomic objectives (Alberta Land Use Secretariat 2008). The planning process, referred to as the Alberta Land-use Framework (ALUF), has been partly informed by scenario analyses for the first two regional plans undertaken to date. In this paper, we discuss the role of scenario analysis in the ALUF and describe ALCES®, the land-use simulation model being applied to explore regional cumulative effects. As an example of how ALCES is informing the regional planning process, we draw from the ALUF’s scenario analysis that focused on the oilsands region of northeastern Alberta. We conclude by discussing some of the challenges and opportunities of applying simulation models to inform IRM. In particular, we focus on the role of strategic modelling, the importance of stakeholder participation, and the need to apply models to inform society at large about land use issues.

2. ALCES AND THE ALBERTA LAND USE FRAMEWORK

The ALUF, mandated by the Alberta Land Stewardship Act in October 2009, represents a shift from project-specific to regional environmental management. The focus is on strategic level issues and solutions such as the desired balance between ecological and economic performance and how provincial strategies such as economic development, protected areas, and regulations should align at the regional scale. While regional plans require approval from Cabinet, the plans are designed by multi-stakeholder groups through an iterative process informed by scenario analysis. The multi-stakeholder groups, referred to as regional advisory councils (RACs), are populated with a diverse collection of individuals who are collectively knowledgeable of the full range of land-use issues relevant to a region. With ongoing guidance from the RAC, a regional planning team (RPT) consisting of experts from relevant provincial government ministries structures a series of scenarios intended to demonstrate the implications of a range of management options. The process is fluid, with each subsequent scenario being informed partly by learnings from previous scenarios. Ultimately, the RAC applies learnings from the iterative scenario analysis, along with additional information and analysis supplied by the RPT, to recommend a regional land use strategy to Cabinet. If accepted, the regional strategy is then used to guide the development of subregional municipal plans that are
consistent with regional objectives. More information about the ALUF is available at www.landuse.alberta.ca.

ALCES is being applied to inform the ALUF due to its capacity to examine inter-relationships among the full range of relevant land-use sectors and natural disturbances, and explore their environmental and socioeconomic consequences at large temporal and spatial scales (Hudson 2002, Salmo Consulting et al. 2001). ALCES is a stock and flow model built using the Stella modelling platform (www.iseesystems.com). The model was first developed by Dr. Brad Stelfox in the mid 1990’s and has gradually expanded in scope to meet the needs of various regional planning initiatives in western North America. The following description provides an overview of ALCES structure and function. More details can be found on the ALCES Group website (www.alces.ca).

To achieve a synoptic view of regional cumulative effects, a wide-range of land uses and ecological processes are incorporated into the model as drivers. The various land uses and ecological processes can be turned on or off depending on the needs of the scenario analysis. For each land use operating in a region, the user defines development rates, the portion of the landscape available for development, and management practices such as the intensity and lifespan of associated industrial footprints. The influence of natural disturbances (fire and insects) and plant succession on landscape composition are also tracked. Hydrological processes are addressed with surface and groundwater modules, and climate change effects can be incorporated by defining temporal changes in natural disturbances rates, successional trajectories, landcover, meteorology and hydrology.

The first-order effects tracked by ALCES are landscape composition and resource production/supply. Using an annual time-step (although monthly time steps can be used for the meteorology module) the model modifies the area and length of up to 20 landcover and 15 anthropogenic footprint types in response to natural disturbances, succession, landscape conversion, reclamation of footprints, and creation of new footprints associated with simulated land-use trajectories. ALCES is a spatially stratified model, meaning that it tracks the area, length, and quantity of each footprint separately for each landscape type. ALCES does not, however, track the explicit geographic location of these features (e.g., latitude and longitude), a feature that greatly speeds up processing time (less than 1 second per simulation year) relative to a spatially explicit modelling approach. ALCES also tracks resource production and supply using approaches that are typical of sector-specific models such as forestry timber supply models and the Hubbert-Naill life cycle approach for simulating exploitation of hydrocarbon deposits (Naill 1973). By tracking resource supply, ALCES can reduce or stop the expansion of a land use if resource supply becomes inadequate. Changes to water quantity are also tracked by applying water use coefficients associated with each land use.

Land base composition and resource production attributes are translated into indicator variables using coefficients. A wide range of indicators are available so that trade-offs between diverse ecological and socioeconomic objectives can be assessed. Types of indicators that can be tracked by ALCES include wildlife habitat and populations, water quality and quantity, biotic carbon storage, air emissions, employment, gross domestic product, and social indicators such as family income and educational attainment.

By applying ALCES Mapper, ALCES tabular and graphical output can be augmented with maps illustrating the plausible future condition of landscapes and indicators. ALCES Mapper is a companion tool to ALCES developed by Alberta Innovates Technology Futures (formerly Alberta Research Council) as an ArcGIS application (www.esri.com). The tool divides the study area into grid cells of user-defined size, and calculates the initial landscape and footprint composition within each cell. Footprint growth and reclamation, landcover change, natural disturbances, commodity production and other variables as reported by ALCES are then applied to each cell, tracked, and displayed spatially. ALCES Mapper allows users to specify the
general location (i.e., where specified land-use footprints can or cannot occur) and pattern (e.g., dispersed versus contagious) of future development. This feature provides flexibility to map transformations of landscapes through time according to different spatial rules, and is useful for visualizing the implications of different zoning or resource utilization strategies. Maps of future landscape condition can then be analyzed to evaluate the spatial response of indicators such as wildlife habitat to potential future landscapes associated with land-use scenarios.

Figure 1. Overview of the ALCES land use simulation tool.

3. CASE STUDY: THE LOWER ATHABASCA REGIONAL PLAN

The Lower Athabasca Regional Plan (LARP) is the first of the seven regions in Alberta to be assessed as part of the ALUF. We do not attempt to summarize scenario analysis outcomes because LARP will not be completed until later this year. Rather, the scenario analysis approach is described to demonstrate how ALCES and ALCES Mapper can be applied to inform regional planning. ALCES simulations for LARP are being completed by the ALCES Group, under direction of the RPT. The scenario analysis approach used for LARP evolved over the course of the plan development and continues to be evaluated to seek refinements in how it can best inform development of future plans.

The Lower Athabasca Region occupies a 92,000 km² boreal landscape in northeastern Alberta. The region’s forests, fen and bog complexes, lakes, and extensive lotic system support a diverse range of species. Although this region has supported First Nation communities for thousands of years, it is only during recent decades that large-scale industrial development has emerged...
including forestry, energy, and agriculture. The region contains the Athabasca Oilsands, one of the largest bitumen deposits in the world.

Land uses and ecological processes included in the simulations were energy, mining, forestry, agriculture, settlements, transportation, protected areas, fire, succession, and meteorology. To support a transparent modelling process, information to model these processes and define the initial composition of the landscape was either provided or vetted by government ministries, and modelling assumptions are extensively documented for each scenario. Further, the model has been previously validated and calibrated through expert review and comparison with outcomes from more detailed, sector-specific models (Hudson et al. 2002). The scenario analysis reported on approximately 50 indicators related to landscape composition, terrestrial and aquatic habitat and biota, water quality and quantity, air quality, and economic health. Indicator coefficients were defined by subject experts, and vetted by the RPT. Indicator coefficients were empirically derived when data were available from sources such as the province’s biodiversity monitoring program (www.abmi.ca), and were otherwise based on literature review or expert opinion gathered during workshops.

As described previously, scenarios are being structured by the RPT based on input from the multi-stakeholder RAC. Scenario outcomes are presented to the RAC in a workshop setting, and feedback from the RAC is utilized by the RPT to structure the subsequent scenario. Scenario development by the government (RPT) and stakeholders (RAC), rather than the modelers, helps ensure that the scenario analysis considers land-use options that are relevant to the planning process. The first scenario evaluated sensitivity of indicator outcomes to the rate of bitumen development, the most important land use in the region. The next scenario assessed the “best available” management strategies, within technological and economic constraints, for limiting environmental degradation per unit of resource production. The suite of strategies, referred to as best practices, was identified through consultations with experts from government and industry. Included were strategies for minimizing the size and duration of industrial footprint, old forest protection, water conservation, and emissions reduction. A subsequent scenario focused on the potential of mitigating impacts to wildlife species through vehicular access management. The scenario analysis is ongoing, with land-use zoning being one potential scenario yet to be assessed.

A series of fifty 200-year simulations of stochastic fire and meteorology regimes in the absence of land use were completed to estimate the range of natural variability (RNV) for ecological indicators as a comparative benchmark to help interpret scenario results. The likelihood and severity (i.e., risk) of negative impacts to native wildlife and ecosystem services increase as environmental conditions depart from natural conditions. As such, natural conditions are a relevant benchmark for assessing the compatibility of land-use strategies with the ALUF’s outcome of “healthy ecosystems and environment”. For indices of wildlife abundance, departure from natural conditions was categorized into four zones of risk based on species evaluation thresholds used by Alberta’s Endangered Species Conservation Committee (ESCC): stable (within 10% of RNV), low risk (10-50% from RNV), moderate risk (50-70% from RNV), and high risk (>70% from RNV). The quantity of information associated with outcomes for approximately 50 indicators has the potential to obscure key strategic lessons conveyed by land-use simulations. We therefore aggregated indicators into a small set of indices by converting outcomes to a common scale (percent departure from RNV) and averaging across indicators. Although useful for summarizing results, the indices must be carefully interpreted because of the potential to gloss over declines in specific species or ecosystem services.

When creating maps of simulation outcomes with ALCES Mapper, a variety of spatial map themes were applied to direct the location of future land-use features. These themes included bitumen deposits, the area allocated for timber production, towns and cities, the agricultural zone, and protected areas. Land-use expansion followed anchored, contagious, or dispersed growth patterns depending on the type of footprint. Levels of access management differed
across the region depending on sub-regional management priorities. An example of map output is presented in figure 2.

Figure 2. Potential response of a wildlife habitat suitability index at simulation year 60 in the Lower Athabasca Region without (left map) and with (right map) access management. Colours reflect the following levels of risk to moose: stable (green); low risk (yellow); moderate risk (orange), and high risk (red).

4. CONCLUSIONS AND RECOMMENDATIONS

The Alberta government has initiated the ALUF to identify land-use strategies capable of achieving: a healthy economy supported by land and natural resources; healthy ecosystems and environment; and people-friendly communities with ample recreational and cultural opportunities. Doing so requires decisions to address conflicting values and competing interests. Although the LARP process is still incomplete, we believe that the scenario analysis is succeeding in its role as an interactive learning tool to assist the RAC in identifying regional strategies capable of achieving a suitable balance between ecological and socioeconomic objectives. Two characteristics of the modelling process have contributed to this success: its focus on strategic rather than operational issues; and reliance on government and stakeholders for model inputs and scenario development.

To facilitate learning, a model should suit the characteristics of the planning process and the needs of its participants. The complexity of IRM and the need to evaluate strategic trade-offs calls for models that are broad in scope with capacity to integrate the effects of numerous land-uses via indicators that span the diverse outcomes of interest. The integrative capacity of models may be their most important contribution to public policy processes such as land-use planning (Sterk et al. 2009). Accordingly, ALCES has been designed to be comprehensive and reviews of the model have concluded that it is unique with respect to its capacity to integrate the cumulative effects of various land uses existing in western Canada (Hudson 2002, Salmo Consulting Inc. 2001).
To achieve a comprehensive scope, ALCES sacrifices detail relative to many models. ALCES’s spatially stratified approach tracks the cumulative effects of multiple land-uses and natural disturbances on landscape composition but not juxtaposition impacts of these disturbances. The focus on composition rather than configuration is appropriate given that biodiversity is primarily affected by habitat loss rather than fragmentation (Fahrig 2003). Experts engaged in the modelling process to help parameterize ALCES do voice concerns about the coarse nature of ALCES relative to the detailed models that they are familiar with. However, comprehensive strategic modelling is an appropriate focus during land-use planning where the task at hand is visioning potential futures to support political decision processes (Couclelis 2005). Narrower, more detailed models are better suited for operational issues like planning specific developments once a land-use plan is in place.

The regional focus of ALCES is not without its limitations, and opportunities exist to enhance the modelling process with more spatially detailed information. Stakeholders and government staff frequently desire that scenario results be reported subregionally to better understand issues in their “own backyard”. ALCES Mapper results can be summarized at these sub-regional scales to highlight issues that should be the focus of subsequent, more detailed sub-regional planning. In addition, an exciting opportunity exists to integrate ALCES with more tactical modelling tools to assess specific issues (e.g., water use or acid deposition) in greater detail within the strategic regional scenario assessment.

The scenario analysis process has relied on government and stakeholders to provide the information needed to parameterize the model and define scenarios. Active participation of planners and stakeholders in the modelling process has created trust in the validity of simulation outcomes and helped to ensure that the modelling results have played a meaningful role in the RAC’s decision making. A model is unlikely to contribute to multi-stakeholder problem solving unless it is involved in the interactions between participants (Sterk et al. 2009). While effective, participatory modelling can be hard to make operational because planners and stakeholders may be reluctant to engage in modelling due to unfamiliarity or competing time demands (Borowski and Hare 2005). The commitment of the RPT, RAC and modelling team to invest substantial time communicating with each other about modelling assumptions and results has been essential to the success of the LARP modelling process to date.

The modelling process is structured such that substantial effort is applied to develop a small set of scenarios that are supported by government and stakeholders as being representative of a plausible future trajectory. A drawback of this participatory approach is that little opportunity is given to explore a broader spectrum of possible futures. As an example, the implications of rare but potentially influential events such as very large fire years, severe weather events, or catastrophic industrial accidents have not been evaluated. Future ALUF regional analyses would benefit from the inclusion of a sensitivity analysis to assess the potential implications of a wider range of future events, thereby creating greater resilience in the plan.

Even if modelling succeeds in informing planning process, its influence on land use is uncertain. The general public will ultimately determine the acceptibility of land-use strategies and modelling should therefore strive to build societal awareness of the benefits and liabilities of land use. With support from a range of industry, nongovernment, and government organizations, we have developed free, web-based modelling tools including an educational version of ALCES (www.albertatommorow.ca), an urban growth simulator (http://www.abll.ca/aref/), and a library of the historical trajectories of ecological, landscape, and land-use attributes in the province (http://www.abll.ca/library/Library_Home). We believe that such tools will help foster broad public support for tough decisions that must be made during regional planning by informing the broader public about the long-term effects of land use in Alberta.
ACKNOWLEDGEMENTS

The scenario analysis described in this paper was funded by the Alberta Government. We thank the Alberta Land Use Secretariat, Regional Planning Team, the Regional Advisory Council, and numerous subject experts for providing information and guidance during the analysis. We congratulate the Alberta Government for its foresight and courage in taking on the challenges they have accepted with the ALUF.

REFERENCES

Couclelis, H., “Where has the future gone?” Rethinking the role of integrated land-use models in spatial planning, Environment and Planning A 37: 1353-1371.
A new tool for integrated and interactive sustainability impact assessment of urban land use changes: the PLUREL iIAT

Dagmar Haase a, b, Annette Piorr c
Nina Schwarz a, Ingo Zasada c

a Helmholtz-Centre for Environmental Research – UFZ, Department of Computational Landscape Ecology
b Humboldt-University Berlin, Institute of Geography
c Centre for Agricultural Landscape Research – ZALF

dagmar.haase@ufz.de

Abstract: The new integrated Impact Analysis Tool (iIAT) synthesises the modelling results from the impact analysis of land use changes of the EU-project PLUREL on peri-urban land use relationships and the effects of peri-urban land use change (www.plurel.net) into one stand-alone tool. It facilitates the integration of manifold aspects of problems of land use (change) and its functions and services related to urbanisation. It further considers conflicts of interest of different stakeholders such as residents, planners or developers within a planning process. The iIAT covers all dimensions of sustainability, namely the economic, the social and the environmental. Based on land use change maps, economic, social and environmental impacts are evaluated in terms of their sustainability impact using a range of indicators. The iIAT is an internet-accessible tool that displays results in form of sustainability spidergrams, which provide a surface that enables an easy and holistic perception of multilevel information on land use change impacts. The PLUREL iIAT consists of two modules: the iIAT-EU and the iIAT-Region. Hence, it covers two spatial levels: the EU27 and the regional (urban region) level. Being user determined, e.g. by displaying comparative spidergrams for different land use scenarios or for urban regions across Europe, the tool facilitates discussions for heterogeneous user groups ranging from an EU policy assistant to a regional or local planner. The paper presents the prototype version of the iIAT.

Keywords: impact assessment tool; urban region; spidergrams.

1. INTRODUCTION

1.1 Peri-urbanisation in Europe

Urbanisation has arguably been the most significant process of land use change in Europe since the Second World War. Over 70% of Europe’s population now lives in urban areas, which in turn have grown in area by almost 80% over the last fifty years (EEA, 2006). The most obvious signs of this shift towards urbanisation are urban sprawl and the emergence of peri-urban areas, characterised by scattered built-up residential, industrial or commercial areas and dense transport networks, but also by the establishment in some places of green belts, recreational facilities, urban woodlands and golf courses, the conversion of farmstead
complexes into housing and changes from conventional agricultural land uses into hobby farms and rural areas within easy reach of the city.

1.2 The PLUREL approach

The EU-project PLUREL (www.plurel.net) on peri-urban land use relationships and respective land use changes in rural-urban regions aims to achieve a deeper understanding of the changing relationships between urban and rural land use with an emphasis on the most dynamic portion, that of peri-urban areas. It develops methods and tools to assess the environmental, social and economic impacts of land use changes. Potential strategies and good practice examples will be identified in order to promote the sustainable development of land use systems in Rural-Urban Regions, especially the peri-urban. A multi-level approach is essential, both to identify driving forces and pressures, and to explore policy responses and opportunities. Thus the results will be targeted to the pan-EU level as well as for several case studies (Nilsson et al., 2009).

For the pan-European level, PLUREL develops typologies for Rural-Urban Regions, as well as future scenarios for spatial development (Ravetz and Rounsevell, 2008). These scenarios are assessed regarding their effects on land-use change, peri-urban land use relationships, as well as wider sustainability impacts, delivering outputs at NUTS2/3 level across the EU. For the case study level, PLUREL combines detailed collaborative case studies and stakeholder scenarios for peri-urban development pressures, planning and governance systems – with the development of quantitative land use scenarios, for the assessment of peri-urban land use relationships and sustainability impacts, both from regional policies and external driving forces.

What policy questions PLUREL is supposed to answer? It is expected to support end users towards a better and more integrated understanding of urban-rural interlinkages, of trends and processes specifically occurring in peri-urban areas, and of the possibilities to steer them. “End users” refers to policy makers at EU level as well as national policy makers and stakeholders in the case study regions.

1.3 The need for an integrative presentation of modelling results

Within the project, a huge number of modelling results in terms of land use change scenarios and impact analysis has been created. Discussions of scientists and (regional as well as EU-level) stakeholders showed the need for a summarising and comparative tool to provide an overview over all these results. Thus, the integrated Impact Analysis Tool (iIAT) was developed in discussion with stakeholders to synthesise the modelling results from the impact analysis of modelled land use changes into one tool. It is a tool for an integrated result presentation. The tool is multi-purpose and interactive in nature. It allows the integration of manifold aspects of problems of land use (change) and its functions and services related to urbanisation or land consumption. It considers conflicts of interest between different stakeholders within a land development, planning or governance process (Nijkamp et al., 2002; Nijkamp and Vreeker, 2000). The iIAT covers all dimensions of sustainability, namely the economic, the social and the environmental, as required for integrative tools (Schetke and Haase, 2008).

2. THE TOOL

Physically, the PLUREL iIAT is an internet-accessible tool that displays results in form of spidergrams. Those spidergrams provide a surface that enables an easy and holistic perception of multilevel information. These spidergrams allow for a visualisation of changes in indicators, as positive or negative trends according to different scenarios are immediately visible as shifts in the lines of the spidergrams. Different directions of shifts for two or more indicators thus show trade-offs between different dimensions of sustainability.
The interactive nature lies in the possibility for an in-depth view into different thematic scopes and different scales, chosen according to individual user interest. The tool accesses the impact assessment result database of PLUREL and generates the demanded outputs in the graphical user interface (GUI). Via the GUI it is possible to explore the effects of global drivers, national planning policies and local governance on land use change in rural-urban regions and, consequently, their impacts on sustainability. The iIAT is Java based. To use it, no installation etc. is necessary, but a connection to the internet has to be established.

The PLUREL iIAT consists of two modules: the iIAT-EU and the iIAT-Region. Hence, it covers two spatial levels: the EU27 in form of NUTSX regions (size-harmonized NUTS 2/3 regions) and the regional (urban region) level. Being user determined, e.g. by displaying comparative spidergrams, the tool facilitates discussions for heterogeneous end user groups ranging from an EU policy assistant to a local planner. What is more, collaboratively working with the iIAT provokes learning processes about different view on land use development (Haase et al., 2009).

2.1 The iIAT-EU

The iIAT-EU shows how the impacts of urbanisation under future scenario conditions will differ from the current situation (Nielsen et al., 2009). The initial situation 2000 (baseline) can be compared to up to four scenarios of future development (Ravetz and Rounsevell, 2008) for the two time slices 2015 and 2025. The main purpose of the iIAT-EU is to create awareness on how sustainability trends develop at different scales for different types of regions and where policy action might be necessary, thematically and spatially. It also allows distinguishing the impacts of trends in predominantly urban, peri-urban and rural regions. Underlying data are derived from modelling results at the spatial NUTS3 (administrative unit across the EU comparable with districts or counties) or NUTSX scale.

The major indicators the iIAT provides at the EU27-scale address the three pillars of sustainability (Level 1). Level 2 of the iIAT-EU allows gaining a deeper insight on the processes that we expect for different scenario settings. The indicators at level 3 specify the environmental categories of level 2 ‘habitat and biodiversity’, ‘recreational value’ and ‘regulation function’ or the economic indicator ‘agricultural production’. For example, trends show an increase of horticultural areas and of small farms, but a decrease of forests and semi-natural areas (Figure 1).

In order to carry out comparisons, the user can choose different scales of outputs, e.g. the average EU-27 or in the predominantly peri-urban regions of Europe. The user may also choose different typologies for comparisons, so, for example a rural-urban and a mono- and polycentrism typology both developed in PLUREL (Loibl et al., 2008) or a European planning regime typology also developed in PLUREL but also typologies frequently used by the EC such this of the coastal areas. Each NUTSX region has a characteristic profile, resulting from attributes that are derived from the different typologies. In the iIAT functionality they act as filter for the generation of grouped average values of one or more attributes (characteristics). So doing, it allows for comparisons with

- the national average,
- the same or other RUR type regions,
- other coastal areas,
- spatial planning types,
- innovation regions,
- low accessibility regions or
- high natural hazard regions.
By making use of the typologies the user will be enabled to carry out comparisons between a single NUTSX region and an average of other NUTSX groups or between groups (types). The data themselves will be transformed into standardized values in order to unify the scale of output data values between indicators. In the conduction of the standardization, at first we removed the outliers. Therefore all variables with a value above the 97.5 percentile were reduced to that value and all variables with a value below the 2.5 percentile were raised to this level. After that a z-standardization was made. For the z-standardization, the standard deviation was taken into account. We used the OECD Statistics Working Paper in which the method is explained (Nardo et al., 2005). The data input into the iIAT is realised using an open source database format (PostgreSQL) which can be easily and freely updated by the user. Figure 2 shows the preliminary version of the GUI of the iIAT-EU in a stylised form.
2.2 The iIAT-Region

The iIAT-Region approach, technically similar to the iIAT-EU, allows selecting regional land use related impact indicators, the case studies to be compared, and the scenarios that should form the basis of the comparison, as well as thresholds or target values for single indicators. Currently, the six European PLUREL case studies are included: Haaglanden, Koper, Leipzig-Halle, Manchester, Montpellier, and Warsaw.

All values entering the iIAT-region database are change-values compared to the baseline of 2000 and/or a respective target value. They are given as relative changes in %. As output the iIAT computes interactively composed integrated spidergrams for a) different scenarios for one selected urban region or b) a range of indicators comparing different urban regions (cf. Figure 3).

![Figure 3: Sustainability spiders provided by the iIAT Region: A comparing assessment is possible between different scenarios for a selected case study or several case studies with only a single scenario.](image)

Although relative numbers are visible, absolute numbers are available and stored in the database (as reference values and for additional or alternative computations). Compared to the iIAT-EU, the iIAT-Region additionally provides the possibility to enter target values for indicators by the user. Although not a full participatory model, the iIAT-Region facilitates participatory decision processes of practitioners or policy makers.

3. THE INPUT DATA

3.1 Scenarios for land use change

The scenarios used for the iIAT consist of narrative letters related to the AB-scenarios of IPCC SRES and of quantitative demographic (Skirbekk, 2008) and economic trends (NEMESIS model). For the pan-European level, these scenarios were translated into input parameters for the RUG model (see below). For the case study level, planning and governance strategies are additionally incorporated based on regional stakeholder workshops and a joint “regionalization” of the pan-European scenarios to be used as input for the MOLAND model (see below). Thus, planning is translated into land use neighbourhood attraction curves and suitability maps that stand for planners’ decision-making. These scenarios are comparable in terms of their storylines, but not in terms of model specifications as different models are used for different spatial levels and the scenarios for the case studies are locally adapted. Additionally to these scenarios, for each level of analysis a “baseline” in terms of land cover in the year 2000 is available.
3.2 Land use change models

The scenarios of future land use development of rural-urban regions across Europe are computed using the European-wide RUG-model (Rickebusch and Rounsevell, 2009) or the regional MOLAND cellular automaton. The RUG model for all urban regions across Europe uses regression functions, gravity and cost-distance functions to simulate land use change from non-urban to urban and vice versa at a 1km-grid (covering nearly 4.2 million pixels). The main input to RUG is a projection of the quantity of artificial surfaces per NUTS 2 region for 2025. This is derived from projected population and GDP (Gross Domestic Product) per capita, both outputs of the NEMESIS model. To allocate these artificial surfaces within each region, the model also uses data such as travel times to the nearest cities (medium or large), distance from the coast and the presence of flood risk zones (Rickebusch and Rounsevell, 2009). The cellular automaton MOLAND and MOLAND light, respectively, work at the case study scale. MOLAND simulates land use change patterns for 100x100m grid cells stratified into about 15-30 land use classes for the time slices of 2005 and 2025. Growth rates (land use pressures) are determined by expert based regional estimates using targets for the future, e.g. residential land use based on population growth scenarios where additional demands for residential land per capita are based on the new (future) population projections. Particularly MOLAND light is a specific development within the PLUREL project: it is a tool that makes integrated dynamic modelling available to stakeholders (Petrov et al., 2009).

3.3 Indicator analysis

Based on the resulting land use change maps, economic, social and environmental impacts were evaluated for both the pan-European and the case study level. Due to data and model availability, the indicator sets for the two levels are not identical, although both cover the three sustainability dimensions.

The indicators the iIAT provides at the EU27-scale address, as already mentioned the three pillars of sustainability (level 1). It is possible to select more specific impact categories and indicators for each dimension. The level 2 impact categories of the iIAT-EU allow gaining a deeper insight on the processes that we expect for each scenario: economic performance, food production, income, living environment, regulation, demography and housing. At level 3, the respective indicators specify e.g. the environmental impact categories such as habitat and biodiversity, recreational value and ecological regulation function. All indicators are computed in form of land use (change) response functions and represent mathematically bivariate regression functions.

For the case study level, a range of indicators such as share of impervious surface, climate regulation potential, carbon storage, recreation area, biodiversity, employment in the industry or agrarian sector and population density was used (e.g. as proposed in Schetke and Haase, 2008, Haase and Schetke, 2009; Burkhardt et al., 2009). Land use impact models which use the land use change maps as input data are either statistical models (regression functions or machine learning algorithms), empirical models (in form of look-up tables), additive models or physically-based differential equation models (for calculating water supply or filtering functions for example).

4. CONCLUSIONS

To summarize, the relevance and practical use of the PLUREL iIAT is first of all the comparative and summarising presentation of impact modelling results. Additional values are the fostering of discussion processes on various levels of decision making: (1) The iIAT facilitates the exploration of possible future situations under the perspective of sustainable development. (2) The iIAT-EU helps to identify future policy issues and territorial action agendas for European policy-makers. (3) It provides an information basis and encourages European stakeholders within the different policy fields to search for a more sustainable land use development in urban regions with particular focus on the peri-urban space. (4)
The iIAT-Region facilitates the inter-regional comparison and the identification of hot-spot regions with particular need to policy/planning intervention.

The iIAT will be available for public use. This publication raises the question of risks associated with this tool. Generally speaking, those results could be misinterpreted by the public, and this risk cannot be fully eliminated. However, the PLUREL scientists will provide accompanying documentation of all methods in a publicly available tool, the XPLORER (which will actually be the access point for the iIAT). In the XPLORER, descriptions of all methods, including scenario development, land use change modelling, and impact analysis related to all indicators, can be found and are especially targeted at practitioners and not a scientific audience.

The PLUREL iIAT is expected to be online available by the end of 2010. It will be introduced to the public in the International Conference “Managing the Urban Rural Interface”, 19–22 October, 2010, in Copenhagen, Denmark (www.plurel.net/conference).

ACKNOWLEDGMENTS

This work is part of the PLUREL Integrated Project (Peri-urban Land Use RELationships) funded by the European Commission, Directorate-General for Research, under the 6th Framework Programme (project reference: 36921). The authors would like to thank all colleagues working on the iIAT within the PLUREL project.

REFERENCES


Lessons learnt on Requirement Analyses to establish new model systems: Prerequisites to assure model use towards policy outcome

Stefan Sieber*, Karen Tscherning*, Klaus Müller*, Peter Verweij*, Torbjörn Jansson*

---

Abstract: The objective of the paper is to evaluate the applied non-standardised Model Requirement Analysis (MRA). Using the presented methods and results of the MRA, we discuss the suitability for appropriate ways to improve applied methods using the example of the Sustainability Impact Assessment Tools (SIAT). Special focus is given to the prerequisites of project design to assure model use towards model outcome.

The applied methods of the MRA consist of evolutionary prototyping, which provides a way to structure the subsequent group discussions with end-users. The results first summarise the conducted interactions with potential end-users and evaluate the usefulness of conducting MRA for each end-user meeting. A direct outcome of the MRA is the identification of four categories of requirements for the SIAT design: (a) Spatial, time and thematic integration, (b) technical performance and system advancements, (c) quality assurance of data and model systems results and (d) organisational linkages for model system embedding.

We give a number of reasons why the undertaken development process of SIAT was not sufficient for actual operational use towards outcome at the level of policy decision making. A number of recommendations and rules for stakeholder involvement and development methods are suggested in the conclusions.

Keywords: Sustainability Impact Assessment, decision support systems, model use

1 Introduction

Model systems often simulate potential impacts of policy options in order to support decision making (Van Ittersum et al., 2008). There exist still gaps between design and use of model systems (McIntosh et al., 2008). To ensure a high use of model systems among policy decision makers, there is a need for model systems to be aligned to user needs and to overcome the gap between design and use at the science-policy interface (Norse and Tschirley, 2000).

In this paper, we analyse the applied method of a non-standardised Model Requirement Analysis (MRA) using the example of the Sustainability Impact Assessment Tools (SIAT) model system. SIAT was developed in the frame of the Integrated Project SENSOR funded by the 6th Framework Programme of the European Commission, which brought together teams of researchers from 36 institutes in 15 European countries.

SIAT supports integrated ex-ante impact assessments in the context of multifunctional agriculture and sustainable development (Helfing et al., 2008). SIAT was designed to simulate land use policies up to the year 2025 at a regional scale of 570 European regions. Its 83 implemented indicators implicitly synthesize the agriculture sector and the related sectors of forestry, tourism, nature conservation, energy and transport (Sieber et al., 2008). SIAT allows for regionalised trade-off analysis of sustainability.
indicators and conducts evaluations of sustainability decision choice spaces (Helming et al., 2008).

The objective of this paper is to evaluate the applied non-standardised Model Requirement Analysis (MRA) according to its usefulness for developing SIAT with regard to incorporate the results into environmental policy and management actions. Based on the presented methods and results of the MRA, we give recommendations on improving the process, lessons learnt and perquisites to assure transition towards outcome.

2 Material and Methods of MRA
This section clarifies the terminology and discusses first the key methods of (a) evolutionary prototyping and (b) prototype-based group discussions with end-users, which have been applied to conduct the non-standardised MRA. Then we classify these methods into the phases of the project design for model evolvement.

2.1 Defining the MRA
Among the various MRA methods documented (Wiegers, 2003; Ricca et al., 2007; Araujo et al., 2007), ‘user involvement’ and ‘prototyping’ emerge as two key components of MRA (Young, 2001; Sommerville, 2006; Wiegers, 2003).

Software prototyping as a means to guide the model development process should focus the functionalities and model design on the final version of the model, which should be discussed in terms of the specific use of end-user groups (Guida et al., 1999). Typically, certain features mark the functionality of specific domains, which are developed step-wise in collaborative discussions with interdisciplinary researchers (Davis, 1992). Model components require a demand-pull design in their initial orientation (Reeve and Petch, 1999) and may have to use ‘socio-technical’ methods, such as Soft Systems Methodologies, to reflect organisational needs (Winter et al., 1995). The requirements often focus on functional user interface design (Hu et al., 1999). According to studies, non-functional requirements such as quality criteria seem to be less discussed in requirement analyses. Virzi et al. (1993) discussed intensively the usability problem of applying different types of prototypes.

Group discussions with involved potential end-users are often seen as a complementary measure for prototyping. He and King (2008) have widely discussed the importance of user participation with regard to model systems for decision making support. The selection of potential end-users, either as individuals or groups, highly affects decisions regarding model design (Van Daalen et al., 2002; Vetschera, 1997) in terms of both functionality and architecture (Hemmati, 2002).

In establishing mixed end-user groups, different roles, interactions and applied methods should be analysed preferably by a broad spectrum of participants in order to take into account as many requirements as possible (Checkland and Holwell, 1999). Often a broad and good narrative in discussions is more engaging and useful than the best science (Checkland and Holwell, 1999).

For any existing process of decision-making, the organisational structure of the model use plays an important role in design (Fredman et al., 1999; Aggarwal and Rajesh, 1996). More diverse, cross-departmental user groups may lead to an increase in required support. Internal work processes, specific roles of actors or the fact that decision makers often tend to delegate responsibilities to scientific contractors, such as downstream operators, should be taken into account (Funtowicz and Ravetz, 1994). Specific hierarchies and the degree of cross-organisational use of a given system create different requirements with respect to design (Vetschera, 1997).

The key question of the above given processes is, whether the applied processes of user involvements were sufficient to implement the model system for actual use at the level of the European Commission among policy decision makers. To survey model requirements we applied software prototyping, which also allowed us to demonstrate the model system design and implemented functionalities. The model system-prototype was presented in user group discussions.

2.2 Implementing MRA in project design
To conduct the non-standardised MRA, we embedded the key components of ‘evolutionary prototyping (see fig. 1) (c)’ and ‘group discussions with end-users (d)’ in the project design. The project design involves a procedure of the following development phases to develop SIAT. The individual stages of the project design can be outlined as follows.

Fig. 1. The evolution of prototype features in the SIAT development process

(a) Basic requirement evaluation
Based on project design, the basic requirements, capacity settings and deliverables were surveyed. The project specifications allowed for broad decisions in fulfilling contract objectives. Feedback surveys with the contracting body served as an effective measure for priority-setting and maintaining continuous communication, thereby ensuring positive support and minimised risk of late negative judgments and evaluations.

(b) Capacity resource utilisation
Reviewing and benchmarking previous in-house endeavours and expertise helped to increase efficiency and minimise the risk of redundancies. This critical review may lead to the re-use of existing software components. Knowledge resources may also be efficiently used or shared among individuals, groups and organisations. The common development of software across research projects, such as visualisation mapping tools, has been realised for the model evolvement.

(c) Evolutionary prototyping
Our prototypes provide a way of structuring group discussions. The hands-on prototype contains a Graphical User Interface (GUI) with an exemplary implemented model simulation. As one key objective, the prototype should provide a professional ‘look and feel’ on a range of model functionalities.

The group of developers was composed of natural scientists with different backgrounds, including software engineers, landscape planners and agricultural economists. Subcontracted graphical artists delivered on-demand design elements. The result was a SIAT prototype composed of (1) a user interface, (2) a topic-structure for content management of fact sheets, (3) a simplifying model application for one policy simulation, and (4) a visualisation mapping component. Reliability on simulation results was not significant while continuous feedback on design and functionality served as the input for the second and third prototype.

(d) Prototype-based group discussions with end-users
Mixed groups in our study consisted of software engineers, natural scientists and policy experts as potential end-users. Specifically, major clients such as the Research Directorate General, the Institute for Environment and Sustainability (IES) and Institute for Prospective Technological Studies (IPTS) acted as end-users. The IES and IPTS are part of the Joint Research Centre (JRC) of the European Commission. National institutes and various European institutions were involved in the context of scientific conferences as well as through subsidiary informal interviews.

(e) Organisational analysis using semi-structured interviews
Along with group discussions, semi-structured interviews on organisational requirements were also conducted. Within the project consortium, a different group of scientists was responsible for this research field.
Given the above procedure, we considered steps (c) and (d) key elements of applied non-standardised MRA. Steps (d) and (e) allowed active involvements of potential users of the model. They have to be convinced by prototype demonstrations on the usefulness of the model system for discussion support.

3 Results
The results can be structured in (a) the evaluation of end-user interactions to assess the usefulness of the non-standardised MRA, (b) the relevant requirements as the outcome of the group discussions in the frame of the applied non-standardised MRA and (c) the actual description of the design of the SIAT model system components.

3.1 Evaluation of group discussions
The SIAT developer group was involved in 79 internal and external interactions in group discussions. We note that pre-defined compositions of a stable end-user group could not always be realised due to organisational disposition, which was caused by a lack of permanent members. The open discussions were intended to analyse model requirements, which were focused on design potentials.

The evaluation of these interactions was based on available documentation of results and expert assessments. The latter method was implemented by the SIAT developer group using self-assessment to express the usefulness of conducting MRA.

The engagement of the MRA by involved potential end users increased over the project period due to improved prototype demonstrations. The characteristic of the MRA can be described as an outcome of a dynamic group discussion process over the project lifetime. The degree of discussion on model requirements was driven by the given prototype development phases, while the degree of SIAT demonstrations increased due to improved prototypes. Initially lacking discussions on embedding the model system in an organisational structure increased, although it never reached the operational level of actual implementation. While the model design was a stable and central discussion issue, complementary essential issues such as model maintenance beyond the project lifetime and targeted marketing for model promotion were discussed towards the end of the project.

3.2 Results on Model Requirements
The MRA addresses the needs in the context of the evolvement of SIAT and they are specifically tailored to European-wide applied policy for IA analysis (Sieber et al., 2008).

The following described profiles for the SIAT model design are clustered according to most relevant thematic fields of requirement criteria. They are the direct outcome of the applied non-standardised MRA and have been revealed in the group discussions. Either available documentation of the discussions (minutes), and major statements of individuals or general discussion foci have been filtered to summarise requirement profiles.

3.2.1 Integration levels for Sustainability Impact Assessment
Sustainability IA requires a high integration of thematically diverse indicators. End-users expressed the need for thematic integration across social, economic and environmental impacts. Analytical requirements necessary to conduct such a trade-off analysis include spatial scales with a high regional disaggregation. Results should be presented at all available spatial scales from local to regional, and national to EU level. Multiple sector perspectives should be integrated, provided that specific sectors are relevant for the analysis. Cross-sectoral trade-off analysis demands the synthesis of sector analyses and should be considered complementary to in depth sector-specific analysis. The number of indicator variables should be balanced over thematic dimensions. Information should only be explicit if it improves the reliability and plausibility of the results.

3.2.2 System performance
Databases are a key technical requirement, that is the most up-to-date, consistent and consolidated databases. The use of European and official data records was considered the minimum standard. Expert judgment shall be allowed to close eventual data gaps, provided estimation methods are analytically sound and well documented.

The response time for computations of scenarios provided a comparative advantage to existing macro and sector approaches. Thus, it was necessary to have a specific model concept, which allowed for a technical default of scenario calculation time of up to one
minute. The emphasis on multiple mapping and visualisation components required non-standardised technical solutions. A basic compatibility among all implemented model system components was required. Overall, these requirements present technical bottlenecks in the software architecture and model components.

3.2.3 Quality assurance of data and model system results
The reliability criteria emphasise the comprehensibility of scenario results and provided statistics. These quality criteria could be categorised as quality criteria on assumptions, measures to ensure transparency and tracing of calculations.

Criteria on indicator results are expressed in group discussions in the form of available causal chains and process knowledge of indicators, explicitness related to illustrated circumstances of the policy and the availability and visibility of the data records used. Furthermore, it was considered important to use evidence of indicator aggregation rules and input-output relations for all observed scales as well as accuracy of veracity (rather than reproducibility) as average accuracy on a given scales. Slightly less importance was given to the accuracy in predicting the results (or correlations) in the future compared to alternative measures. The complexity of results was seen as an unavoidable reality, but the access to information on context and interrelations in defined systems presents a disclosure condition at all output levels.

Transparency requirements should indicate all defined assumptions that determine results at all levels of scenario resolution, including data sources, exogenous scenario assumptions, indicator definitions, calculation rules and causal chains regarding interdependencies among indicators. Explicitness regarding results is necessary, but if implicit knowledge is applied then access to background information should be guaranteed. Flexibility when choosing the depth of required information is important.

Specific tracing techniques should allow the convergence of simulation results into single calculation steps. The procedure should disclose implemented calculation chains. Not the highest disaggregation of traceable calculation components, but simplicity without loosing the overview was requested and considered as essential.

3.2.4 Organisational linkages for model system embedding
An early permanent link to key contacts within the target organisation was considered a requisite for success regarding the acceptability and dissemination of a model. MRAs regarding user requirements should be as specific as possible, especially if individuals in organizations can be persuaded to collaborate with the model system developer.

It was stated that potential end-users should be involved as early as possible. The continuity of interactions with potential end users during the iterative model development phases is a key action to be fulfilled. Concepts on formal engagements to further develop model systems either through project-funding should be taken into account.

3.3 System design
This section describes the SIAT design, which has been developed based on the outlined model requirements in 3.2. We first illustrate the component-based SIAT framework, and, based on this explanation, we explain the methodology and resulting functionalities of the SIAT model system.

3.3.1 The SIAT framework
Fig. 2 illustrates the basic framework from which SIAT. The external model chain framework generates SIAT input in the form of pre-calculated response protocols. These ‘policy response functions’ describe the effects of exogenous policy instrument changes on (intermediate) endogenous indicator variables, which are implemented throughout as exogenous terms in ‘indicator response functions’. Thematically clustered indicators in the LUF-component are normalised and then aggregated to Land Use Functions (LUF) as indices of compounded results. Based on these policy settings, scenario-based simulations are conducted. Iteratively-solved simulation results can be retrieved from a cache to be displayed and compared by means of visualisation means.

The SIAT contains a server-based SQL-data base that consists of either mathematical functions or knowledge rules (“rules of thumb”) to describe the relation between the policy instrument and the resulting impacts on indicators. The users choose the intensity of the policy instrument (e.g. percentage of direct support of CAP measures) on the Graphical
Use Interface and results will be calculated to be illustrated in visualisation tools (e.g. maps) or to be retrieved in further tools for processing of results (e.g. Excel sheets). New policy cases can be calculated by the established model framework and be uploaded into the SIAT system.

3.3.2 Model chain framework interface

The quantitative results used in SIAT are produced by a system of interlinked models in the model chain framework. It aims to synthesise multiple sectors while maintaining sector-specific details. Each relevant sector for sustainable land use was modelled in detail. The agricultural sector was modelled using the agricultural economic model CAPRI, which runs at regional level within the 27 EU member states (Britz et al., 2008). CAPRI simulates the impact of changing agricultural policies on agricultural production, prices, subsidies and input use (including fertiliser application) in addition to several important environmental indicators, such as nutrient surpluses and greenhouse gas emissions from agriculture. Forest management practices are computed by EFISCEN (Sallnäs, 1990; Schelhaas et al., 2007; Lindner et al., 2002). Based on input data, such as wood demand and forest area, the state of the forest is described using matrices for each forest type in which area is distributed over age and volume classes. SICK, TIM and B&B are models for urban growth, transport infrastructure and tourism respectively. The interdependencies between sectors are handled by linking sector models to an economy-wide economic model called NEMESIS (Zagamé et al., 2002; Brécard et al., 2006) and a land use model called Dyna-CLUE (Verburg et al., 2006). NEMESIS handles competition for production factors, such as land, that are shared among the individual sectors and is also able to simulate the relationship between research and development (R&D) spending and technical progress. NEMESIS also provides results for macro-economic indicators, including employment and economic growth. Dyna-CLUE disaggregates land claims per sector and member state to square kilometres and re-aggregates them into regional scales as required by CAPRI and EFISCEN. Dyna-CLUE also implements spatially-specific policy instruments.

3.3.3 Response function components

The response functions closely replicate the behaviour of the linked system of models and reduce the computation response time. The response functions are static and allow for calculating results within a pre-defined indicator range. The results change depending on the scenario chosen and chosen intensity of the policy instrument (e.g. market support, direct support of CAP measures in percent).

Response functions can be distinguished between indicator response functions and policy response functions. The set of policy response functions for a policy case is generated by performing and analysing simulation experiments with the models, whereas the policy
instruments are systematically varied to cover the entire domain for particular policy cases, with all remaining policy instruments unchanged.

The indicator response function component assures the synthesis of different indicator calculation methods and allows for IA using 83 indicators. Policy response functions provide a linkable component to compute indicators, which are implemented throughout as exogenous function terms in the ‘indicator response functions’. This function type transforms the intermediate policy variables into indicators by means of knowledge rules. The indicator response function component results in different types of indicators either by using model output as indications for gross assumptions (i.e., qualitative indicators) or by transforming quantitative model outputs in indicator results with appropriate units, scales and necessary meta-information. The availability of specific and pan-European data was a driving constraint in selecting the actual indicators. Each indicator function requires exogenous sets of variables on specific indicator-relevant information such as state variables (e.g., soil types) and sets of specific intermediate variables, such as land use-change. Indicators can be expressed by quantitative, semi-quantitative or qualitative functions. The value of each observed variable changes depending on the policy settings.

3.3.4 Graphical User Interface (GUI)

The GUI provides users with two pathways, namely, the simulation of policy scenarios using SIAT or the retrieval of background information. It contains a range of functionalities to assure transparency and traceability at user interaction steps along all navigation bars. The SIAT application solves policy scenarios in five steps.

Step (1) computes the macro-economic reference scenario values of the impact indicators for the target year. The reference scenario includes ‘business as usual’ assumptions, namely, that there will be no changes in terms of anticipated trends of oil prices, expenditures for research and development, the labour force, demographic changes and world economic demand.

Step (2) provides a navigation menu that allows for the selection of policy cases. Each policy case contains sets of policy instruments. Within each case, the user can select and combine different policy instruments. Both the inclusion of one instrument as such and the degree of policy intensity can be chosen in the scenario design.

Step (3) illustrates the computed scenario results of the impact indicators as a consequence of the policy settings. Results are presented using interactive visualisation components such as maps, tables and graphs. Photorealistic visualisations support impressions on changes within landscape views. Map layers using Google data superimpose additional geographical information for specific analyses.

Step (4) evaluates policy impact indicators by means of critical limits, thresholds and targets. These sustainability criteria define sustainability choice spaces, which are based on ex-post scenario comparisons of run simulations.

Step (5) aggregates groups of indicators to Land Use Functions (LUF) that indicate the degree of the relative fulfilment of goods and services at the regional level. Nine LUFs have been implemented, including ‘Provision of work’, ‘Human health and recreation’, ‘Cultural landscape identity’, ‘Residential and non-land-based industries and services’, ‘Land-based production and Infrastructure’, ‘Provision of abiotic resources’, ‘Support and provision of habitat’ and ‘Maintenance of ecosystem processes’.

A range of visualisation components have been integrated into the GUI design. They comprise interactive geographical maps, spider diagrams, numerical tables and textual summaries. The logical structures used to retrieve visualisation components in the chain from (a) policy settings to (b) policy impacts to (c) the valuation of policy effects should be self-explanatory.

4 Conclusions

We have applied a non-standardised MRA using the SIAT model system. The research question was, whether the applied outlined methods were sufficient to actually implement the model system for actual use at the level of the European Commission among policy decision makers. Besides the outlined model development process of the SIAT model system design, the following considerations support points of improvements to ensure actual use by involved potential end users as outcome of cumulated actions:
We would like to state at the beginning that the SIAT model is not currently being used at the level of the European Commission. It was not possible to establish within the project lifetime an entire operational tool ready to utilize for policy advice. The integration of model components is still being processed as post-project endeavour. A workshop to prove the entire implementability with the major aim of demonstrating credibility is being planned for summer 2010.

Expert analysis of the usefulness of the 79 interactions leads to the conclusion that the SIAT demonstrations and MRA discussions were, in general, effective, while the involvement of a stable end-user group over the project lifetime hindered an efficient model system embedding into the organisation structure. The organisational environment to conduct the non-standardised MRA was non-homogeneous and discontinuous among different DIRECTORATE-GENERALS of the European Commission.

The effectiveness for model success through interaction sessions with end users does not depend on the absolute number of interactions, but the quality of sessions (e.g. well prepared questions on possible options of model features) and the presence of important end users, who has decision rights on model use, are essential.

The impact of a stable stakeholder group is ideally a targeted model design that maximises the potential use of the SIAT for present end users and their related organisations. These end users pose as catalyser for organisational linkages and model system embedding.

But in order to transform model development to an outcome of actual using the SIAT model system, following lessons learnt and points of improvement should be taken into account:

- Important seems to be a fully operational model system that convinces potential users with an added value compared to traditional already implemented systems.
- A functioning prototype for demonstration is key strategy for convincing operational performance. Feasibility for new implementable model exercises shall be given.
- Permanent and continuous contact office that is well established and ideally is implemented party in targeted organisation (e.g. contact bureau) assures needed reliability.
- Personal contacts at both front end of the target organisation and the providing research centre that last over a long time increases the probability of adoption of new model system.
- The project design should allocate budget for product life cycle and marketing that we consider as important as the model itself. High Transaction costs were a clearly underestimated factor on effectiveness to meet the defined goals.

Acknowledgements
The authors wish to thank all members of SENSOR (www.sensor-ip.org) for their feedback and input regarding the design of SIAT.

References


User interaction during the development of the Waikato Integrated Scenario Explorer

Hedwig van Delden\textsuperscript{a}, Derek Phyn\textsuperscript{b}, Tony Fenton\textsuperscript{c}, Beat Huser\textsuperscript{b}, Daniel Rutledge\textsuperscript{d} and Liz Wedderburn\textsuperscript{e}

\textsuperscript{a} Research Institute for Knowledge Systems, Witmakersstraat 10, 6211 JB Maastricht, The Netherlands (hvdelden@riks.nl)
\textsuperscript{b} Environment Waikato, Hamilton, New Zealand
\textsuperscript{c} Alchemist, Rotorua, New Zealand
\textsuperscript{d} Manaaki Whenua Landcare Research, Private Bag 3127, Hamilton, New Zealand
\textsuperscript{e} AgResearch, Hamilton, New Zealand

Abstract: The importance of user interaction for the design and development of Decision Support Systems is widely accepted in literature. How this interaction could best take place is however not yet common knowledge and examples of user involvement in the development of Decision Support Systems targeted at environmental policy and sustainable development are scarce. This paper provides a practical example of user involvement during the design, development and implementation of WISE, an Integrated Spatial Decision Support System for planning and policy support for the Waikato region in New Zealand. Focus will be on the process of user involvement during various development phases, the way feedback was collected and handled, and on the changes in the system this resulted in. We conclude that user interaction has had clear implications on the larger design choices as well as the details of the system development and that this interaction had impacts on choices related to the usefulness and the usability. The effort from the user organisation has been substantial and has increased over the duration of the project. The role of champions has been found indispensible in involving users and enhancing the system. The current implementation phase will show if these efforts will also lead to an effective implementation and easy adoption of WISE by the user organisation.

Keywords: Integrated Spatial Decision Support System; Policy Support; Integrated planning; User involvement; Design and development process.

1. INTRODUCTION

The Creating Futures project is developing decision support and process management tools for use in policy development (Huser et al., 2009). These tools are aimed at improving decision making outcomes through more informed consultation and evaluation of future strategy and trade-off choices. One of the tools developed is an Integrated Spatial Decision Support System (ISDSS) named WISE: Waikato Integrated Scenario Explorer. WISE aims to support long-term integrated policy development and planning in the Waikato region in New Zealand by taking into account cultural, social, environmental and economic wellbeing (Rutledge et al., 2008; Huser et al., 2009). WISE is intended to be used within the context of a deliberation that allows for end users to evaluate their policy options against a range of pre determined values and associated indicators (Wedderburn et al, 2009).

WISE consists of a spatially explicit systems model operating at four scales: global, regional, district and local (200 m grid cells). The temporal resolution is one year and its horizon is set at 2050. In the development of WISE there has been a strong emphasis on
the linkage and feedback loops between the different components (e.g. climate, hydrology, water quality, ecological economics, population, land use and biodiversity), rather than on modelling all elements with the highest detail possible. Drivers for the integrated model are climate change scenarios, socio-economic drivers (e.g. fertility, mortality and migration rates, exports and consumption patterns), and policy alternatives (zoning regulations, impact restrictions and construction of infrastructure).

WISE is developed as part of a research project with a duration of four years. The project is led by the system’s intended end-user Environment Waikato, the regional council of the Waikato region and includes frequent interaction with potential users throughout the whole project. The importance of user involvement in the design, development and implementation of innovative products such as WISE is widely mentioned in literature and is assumed to have a positive impact on the system quality, decrease resistance to change and increase their adoption (Lucas, 1974, Keen and Gerson, 1977, Allen and Hauptman 1990). Nevertheless, the financial and cognitive costs associated with user interaction should not be ignored and therefore the appropriate amount of user involvement for a given project should be carefully evaluated (Gales and Mansour-Cole, 1995; Van Delden et al., in press).

This paper describes the user interaction carried out during the design, development and implementation of WISE. After an overview of the various tasks, project phases and ways of user involvement, we describe the results of this interaction. We focus on the way feedback was collected and handled, and on the changes in the system this resulted in. Finally, we conclude with some lessons learnt and recommendations for user interaction in future design and development processes.

2. DESIGN AND DEVELOPMENT PROCESS

WISE is being developed in an iterative and interactive development process, meaning that there is a continuous interaction between all parties involved: the intended users of the system, the scientists and the IT-specialists (Figure 1) (Van Delden et al., in press). Each party brings to the process its own background and understanding and has its own tasks and responsibilities. End-users provide the policy context and define the problems, functions and usage of WISE. Scientists are responsible for the main model processes and choices of scale, resolution and level of detail. IT-specialists design the system architecture and carry out the software implementation of the models and user interface. Several decisions are made in a group processes in which one party takes the lead: Scientists, together with end-users and IT-specialists, are building effective linkages between the individual models and end-users together with IT-specialists and scientists decide on relevant component to include in the user interface.

![Figure 1: Main parties, responsibilities and integration issues during the development of a DSS](image-url)
The interaction between the three groups involved is as important for the quality of the final product as the tasks carried out by each group individually. This interaction enables social learning between all involved, which is crucial for the development of a useful and user-friendly DSS. The architect has an important role in bringing all groups together, creating mutual understanding and respect and keeping the focus on the final product throughout all phases of the project. In various iterations, the system evolves through a scoping phase and a number of prototypes into a final product (Figure 2).

Tasks that can be distinguished during this design and development process are:

1. Scoping
2. Model selection
3. Model integration
4. Bridging the science – policy gap
5. Development of a user friendly interface
6. Implementation
7. Use and maintenance
8. Evaluation

The first five tasks are part of the design and development process; tasks six and seven follow afterwards, but are included because they have major implications on choices made during the design and development process. It should be noted that these seven tasks need not (and often cannot) be executed in a sequential order. After each of the tasks 1 to 7 an evaluation needs to be carried out to assess what adaptations to previous tasks can be made.

The entire process can roughly be divided in a scoping phase, the development of a series of prototypes and the implementation of the system in the user organisation. User interaction was focused on both the usefulness and the usability of the system and took place during all these phases in various forms. User interaction was strengthened by the user organisation (Environment Waikato) leading the project and hence strongly involved in all phases of the process.

Users were also represented on the project team, which consisted of about 10 people (scientists and IT specialists). Two core user representatives or champions were involved: a project manager responsible for innovation projects with a wide network within and outside Environment Waikato and an environmental officer responsible for data management and with a wide experience in geographic information systems. These two people were, as project leader and technical support person, continuously involved in the project and were able to devote between 20-40% of their time on this project. Their input and feedback was provided on an ad-hoc basis, either upon their own initiative or based on feedback requests from developers on design options or draft deliverables.

On an annual basis workshops were organised with a second group of users, a team of policy analysts from Environment Waikato. In-between these workshops ad hoc interaction between the core users and this user group has taken place. For a better understanding of the policy practice and the way WISE could contribute to this, a consultant was hired who carried out in-depth interviews with this user group. The aim of the interaction with this group was to ensure the usefulness of the system for the policy practice of Environment Waikato. To ensure relevance of the system’s approach for other organisations and facilitate the application of WISE outside the Waikato region after the project is finished, interactions have not only taken place with (potential) users within the Waikato region, but also with (potential) users from other geographical areas. At the end
of each workshop session a discussion was organised on the usefulness and usability of the system and users were asked to provide their feedback through a questionnaire. For high level managers and elected councillors from Environment Waikato annual presentations were organised to inform them on progress and to receive their feedback on the overall concept and deliverables.

3. USER FEEDBACK AND SYSTEM ENHANCEMENT

The sections below provide details of the user interaction in each phase together with an overview of the main results.

3.1 Scoping: organization, policy context and WISE functions

The scoping phase of WISE had already started during the development of the proposal of the Creating Futures project. Two research organisations, Landcare Research and AgResearch, started a joint initiative with Environment Waikato to explore end-user needs and scope a potential proposal, before applying for funding to develop decision support tools for long term integrated planning. During this phase conceptual decisions on WISE were already made, mainly those related to the important processes that should be covered by such a system, because this was the starting point to invite additional partners to join the proposal. Selected disciplines for the WISE ISDSS were demographics, ecological-economics, climate, hydrology, water quality, land use and biodiversity.

An important task during the scoping phase is to create a common knowledge base amongst the various parties involved. In the development of WISE this meant that from the early stages of the process user interaction took place by showing similar systems applied to different regions during workshop sessions, allowing users to get an initial understanding of the possibilities, capabilities and limitations of such systems and to enable them to give input into its further development: issues it should provide support to; external factors, policy options and indicators to be included and its overall look and feel.

This common knowledge base is not just related to the system, but also to the policy context. To link WISE to the policy practice, an overview was made of current policy and planning documents as well as internal and external policy processes Environment Waikato is involved in. The processes taken into account included those processes directed by legislation, mainly the Local Government Act 2002 (Long Term Plans) (www.legislation.govt.nz) and the Resource Management Act 1991 (Regional Policy Statement, Regional Plan, Regional Coastal Plan and District Plan) (www.mfe.govt.nz/rma). Applying WISE for non-statutory planning processes was also considered, e.g. the FutureProof growth strategy (www.futureproof.org.nz).

Midway through developing WISE the project team recognised a need to better understand the policy process used by the regional council (Environment Waikato) in order to optimize the use of WISE for council’s current policy practice. Interviews were held with a range of Environment Waikato staff to determine how policy planning has been undertaken within the organization and what role WISE could play in contributing to this process. An overview of these policy processes is described in Fenton (2010). A main conclusion from this analysis was that the way WISE could be used depends on the policy process, the other tools being used and timing and resourcing available. WISE potentially offers an improved method of issue identification, consultation and policy selection and evaluation, which involves a robust process of deliberation around the consequences and trade-offs of different policy options. Consequently the use of these tools may require Environment Waikato staff to work in a way that is different from their current ‘planning paradigm’.

As part of the scoping phase the decision was also made to have a system for strategic thinking and long-term integrated planning that could be used for policy impact assessment and scenario development in workshop sessions. This resulted in a selection of
rather simple process models - to limit the execution time for a simulation run to several minutes - and a focus on dynamic feedback loops between model components. An example of the latter is the link between the economic and the land use model: the economic model is an important driver for land use change in providing land use demand for a range of economic activities, which are subsequently allocated on a grid by the land use model. Only suitable and available locations are taken into account during this allocation. When not all demands can be met the final allocation is fed back to the economic model, which results in a smaller economic growth as would be calculated without resource restrictions. Finally, decisions were made on the user interface. It was decided that the user interface should cater for two types of users: the policy analyst, interested in carrying out scenario and impact assessment studies and the modeller wishing to update data and calibration parameters.

3.2 1st prototype: what can be expected?

The first WISE prototype included the general lay-out of the dual user interface that contained a section for policy users, structured towards a policy impact assessment study and a section for modellers, with access to all underlying data and parameters. In this version only two models were included: the land use change model and the ecological economic model.

In a series of 1-day workshop sessions, the concepts of the first prototype were presented and the system demonstrated to a wide group of (potential) users within and outside of Environment Waikato. Groups outside of Environment Waikato included other regional councils, district councils and national government bodies. The concepts of the entire system were presented in a morning session which was targeted to a wide group of potential users from various hierarchies in the organisations (ranging from technicians to councillors) and a demonstration was offered in an afternoon session which was targeted towards a more technical audience. The system was installed on various laptops and participants were offered a first hands-on experience by following a demonstration that was done on a larger screen. Each session was concluded with a discussion on the usefulness and usability of the system and participants were asked to fill in a questionnaire whenever they left the workshop. This questionnaire included a number of multiple choice questions and a number of open questions. Results of these questionnaires are summarized in Table 1.

Table 1: Results from the questionnaires filled in at the end of the workshops in which the first and second prototype of WISE were presented, discussed and evaluated. After the first series of workshops 32 questionnaires were returned, after the second workshop 24.

<table>
<thead>
<tr>
<th>QUESTION</th>
<th>WORKSHOP WITH FIRST PROTOTYPE</th>
<th>WORKSHOP WITH SECOND PROTOTYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Agree</td>
<td>Disagree</td>
</tr>
<tr>
<td>My organisation would benefit from using Waikato ISDSS</td>
<td>26</td>
<td>3</td>
</tr>
<tr>
<td>Waikato ISDSS enables communication among planners and decision-makers</td>
<td>29</td>
<td>0</td>
</tr>
<tr>
<td>Waikato ISDSS is an easy to use and intuitive tool.</td>
<td>17</td>
<td>5</td>
</tr>
<tr>
<td>I have the impression that in order to use Waikato ISDSS, I need a lot of specific knowledge.</td>
<td>21</td>
<td>11</td>
</tr>
<tr>
<td>I think learning to use Waikato ISDSS is worthwhile, considering the results I can obtain.</td>
<td>27</td>
<td>0</td>
</tr>
<tr>
<td>I would prefer a simpler tool (with less functions), even if that means</td>
<td>6</td>
<td>19</td>
</tr>
</tbody>
</table>
little control on the variables.

I would prefer a more complex tool (with more functions), even if that requires more parameters to deal with. | 9 | 13 | 8 | 15 |

Questionnaire results show that after presenting the first prototype, users were generally positive about the usefulness and usability of WISE and that the chosen complexity seemed to be selected well for the majority of the participants. Most users agreed with the chosen level of complexity, although some would have preferred a more or less complex system. Opinions differed most in the knowledge required to use WISE. The questionnaire results from Table 1 together with the answer on the open questions were put on the project website together with a question and answer document (www.creatingfutures.org.nz/further-resources) to share knowledge between participants from different workshops. Most participants indicated that it was too early to give any suggestions for improvement and that they needed more time with the system before giving specific recommendations.

Besides a few minor adaptations to the user interface not many direct enhancements to the system were made based on this series of workshops. These workshops did however trigger a number of relevant discussions on the inclusion of policy measures in WISE, the operation of the system in an organisation and its relevance at district and/or national level.

3.3 2nd WISE prototype: what needs improving?

The second prototype included the final set of models to be incorporated in WISE as part of the Creating Futures project. Before presenting this system to the wider user groups, one of the core users and a small group of developers spent two weeks together to evaluate the system’s behaviour and to test its user-friendliness. During this interaction data and parameters were updated and model connections and user interface elements improved. This intensive collaborative effort between a user who is familiar with the system and the developers, proved very beneficial in better tuning the system to the actual planning process and information required as part of such a process.

During the development of the second prototype, several case studies were selected that are highly relevant for Environment Waikato and to which WISE could provide an added value: the Regional Policy Statement (statutory), the Future Proof growth strategy (non-statutory) and a set of narrative future scenarios for the Waikato region that were developed as part of the Creating Futures project (www.creatingfutures.org.nz/scenarios). For each of these cases the crucial elements were identified (e.g. urban sprawl, immigration, water quality, protection of natural areas or high quality soils, discussing trade-offs between economy and environment) and case studies and exercises were developed that show how WISE can provide support for exploring these issues.

In a second series of workshops the second prototype was presented and discussed. These workshops were organised in the form of an initial presentation and a set of practical training exercises in a tutorial booklet. Each session was concluded with a discussion on the usefulness and usability of the system and participants were asked to fill in a questionnaire whenever they left the workshop.

Results of these questionnaires are summarized in Table 1. These results are very much in line with the results after the first series of workshops. A main difference was found in the open questions and discussion. After the hands-on training exercises participants felt more familiar with the system and therefore were better able to pinpoint benefits and possibilities for improvement. This resulted among others in a more flexible way on including spatial planning, improved ideas for indicator development and suggestions for additional case studies related to current policy issues at Environment Waikato.
Main expected benefit of WISE that were mentioned by the various user groups, and in particularly those from Environment Waikato, are the ability to:

- explore “what if” scenarios in the development of policy,
- work with district councils in negotiating district and regional policies and rules,
- explore scenarios in public workshops in order for the public to visually “see” options for the region’s future and gain more productive and enthusiastic feedback from them,
- provide support in integrated catchment planning and transport network planning,
- integrate and analyse district policies that vary across adjacent boundaries.

The integrative character of WISE was seen both as a benefit and as a challenge. The complexity created by integrating different sectors includes the interaction between those disciplines and facilitates the assessment of (unwanted) side effects of policies. On the other hand it is crucial that users are able to interpret results and therefore a good understanding of the system is a key to success. It was therefore suggested that a limited number of users would be trained to actually work with the system, while others would learn what type of information the system could provide to them and what type of policy questions it could provide support to.

3.4 Final project version and implementation: are we successful?

At present, work is undertaken to deliver the final version of WISE as part of the Creating Futures project and on implementing the system in Environment Waikato. Foundations for this implementation were placed during the entire development process by carefully deciding for what processes, where in the organisation and by whom the system can best be used and by involving potential future users throughout the development process. Based on feedback from interviews and assessment of the planning processes, a number of barriers to the successful implementation of WISE were identified:

1. Existing planning practice needs to be altered to allow for new tools,
2. Knowledge of WISE needs to be improved with policy staff,
3. Time frames are often under pressure and ‘political expediency’ can drive over good practice,
4. Statutory processes and proposed policies are often legally appealed and hence there is a reluctance to use new and untested tools, approaches or methods,
5. The cost/benefit of a different approach – is there value-add for effort and expenditure?

From this assessment of barriers, it became clear that trialling WISE initially in a non-statutory process rather than in formal legislative processes would be preferable. The non-statutory process would be more focused on a community, an area and its issues. These processes are also more open to idea development and are not burdened by legislative requirements. This should allow for more experiential use of the tools and could therefore improve initial user buy-in.

To limit the time required to train a large group of users in operating the system, it would be beneficial to create an implementation plan and to identify a few more champions, who would – together with the existing champions – apply WISE and:

- improve knowledge about WISE and make WISE useable, credible and supported,
- identify further improvements and modifications to WISE,
- minimise the additional time and resources required to use WISE,
- identifying strategies to change the existing paradigms of approaches to planning and community consultation,
- plan for required technical support for data, training and model set up.
4. CONCLUSIONS AND RECOMMENDATIONS

User involvement during the development of WISE has resulted in a wide-spread knowledge of the tool and an appreciation of its development and enhancements to the system. The latter were especially initiated by those users closest to WISE, the champions. Besides relevant feedback on the behaviour of the system and its user-friendliness, champions also played a crucial role in bridging the gap from the science and developers community to the user organization on a rather detailed system level, thus improving the usefulness of WISE.

Putting innovative spatial concepts in words often proves difficult to both describe and understand. Developing, presenting and discussing various prototypes helps to improve this understanding which results in feedback at rather early stages of the project and builds trust and acceptance for future uptake. Hands-on experience in linking WISE to actual policy issues was valued positively by users and helps to bridge the science – policy gap. Communication channels, both internally and externally, are indispensable in the uptake of a system like WISE, but can easily break down. A project champion who has good face-to-face contact with a wider user group and is dedicated to proactively communicating with them, has proven to be a very good method for overcoming this. Communication and marketing of a project and product requires time, which needs to be budgeted for.

Many potential users do not or cannot place a priority on contributing to projects of innovative tool development due to neither feeling “close” to it, nor having a good understanding of it, or due to heavy workloads and a lack of resources. We suggest that key users need to be identified early on in the project and funding sourced for them so they can contribute in the long term.

ACKNOWLEDGEMENTS

The New Zealand Foundation for Research, Science, and Technology funds the Creating Futures project under contract ENVW0601 to Environment Waikato.

REFERENCES


Empathy: a Unifying Approach to Address the Dilemma of 'Environment versus Economy'

A. Guergachi\textsuperscript{a}, O. Ngenyama\textsuperscript{a}, V. Magness\textsuperscript{a}, J. Hakim\textsuperscript{b}

\textsuperscript{a} Ryerson University, Ted Rogers School of Management, 350 Victoria Street, Toronto, Ontario, Canada, M5B2K3, a2guerga@ryerson.ca

\textsuperscript{b} Hakim Technologies Inc., 1786 Clifton Green, Burlington, Ontario, Canada L7P 2X8

Abstract: When environmentalists warned, back in the 1970s, against pesticides and fertilizers, the then United States Secretary of Agriculture retorted that "before we go back to organic agriculture, somebody is going to have to decide what 50 million people we are going to let starve". This exchange is nothing but an illustration of the dilemmatic choice that societies in general have been asked to make: "develop the economy" or "protect the environment". This paper proposes an integrated framework to resolve the dilemma of 'environment versus economy'. This framework unfolds the essential characteristics of sustainable societies and outlines the mechanics that underlie sustainable development. Referred to as ‘General Theory of Sustainability’ (GTS), it is composed of 10 principles that make statements about a variety of issues including the limitations of systematic thinking and engineering methods, poverty, wealth-generating economic activities, economic theory, civil society and voluntary sector, gender, management of uncertainty and the physical environment. The common thread that weaves through all of these issues in the GTS is a fundamental concept that originated in psychology but is currently popular in other disciplines as well: the concept of empathy.

Keywords: Systematic thinking, economy versus environment, sustainability, empathy.

1. INTRODUCTION

The environment has always been an integral part of the process of building and sustaining civilizations throughout history. It is the source of so many bounties — water, air, food, natural resources, energy, etc. — that are necessary for human civilizations to continue to exist. But the environment is like a coin with two sides; not only is it the source of bounties, but also of threats and challenges.

While some groups (big businesses and others) have downplayed the seriousness of these threats and challenges, there is enough evidence in the literature that environmental damage and dysfunctions represent a significant factor contributing to the collapse of many past civilizations (Diamond, 2006). In our modern world, there are many reasons to be increasingly concerned about the environment. Global warming and climate change, increased frequency of Atlantic hurricanes, huge build-up of toxic chemicals in the environment, depletion of natural resources, energy shortage are all depressing signs that have pushed the issue of environmental damage and sustainability to the top of the agenda of many political parties and governments’ priority lists. Societies have become aware of the risk of a possible collapse and are now eager to act and avoid this risk, but face major difficulties in gathering a consensus among the concerned stakeholders as to what exactly needs to be done to remain sustainable. The Kyoto protocol (1997), Copenhagen Accord (2009), Alberta’s oil sands project (Israelson, 2008), and the cod fisheries in Eastern Canada are but a few examples for which consensus over favoring the economy or protecting the environment has not been easy to obtain.
On one hand, pro-environment groups loudly denounce the destructive nature of many business-oriented activities and demand to reduce these activities to save the environment, but they don’t seem to articulate what needs to be done to compensate for the wealth that will be lost as a result of this reduction. On the other hand, big businesses and their supporters remain relentless in pursuing profit and creating the wealth and jobs that our modern civilization needs to continue to exist in its complex and interdependent form, but they rarely provide convincing and integrated plans to save the environment from collapsing. The result is usually the undesirable situation where either the environment or the business is hurt. The exchange that took place during the 70s between environmentalists and a former US Secretary of Agriculture is a noteworthy illustration of this situation: when environmentalists warned against pesticides and fertilizers, the Secretary of Agriculture retorted that “before we go back to organic agriculture, somebody is going to have to decide what 50 million people we are going to let starve” (Goldstein, 2008).

This paper proposes a new approach to use when considering the dilemmatic choice that societies have been asked to make: “develop the economy” versus “protect the environment”. A basic premise that underlies this new approach is that we need not sacrifice one for the other: not only can organic agriculture feed an entire population, but the principle of developing the economy and protecting the environment at the same time can be applied across the board to all sectors of economy, if the society wants to do so. The approach that we propose in this paper focuses on the smallest entity that makes up a society: the individual. It is argued in this approach that individuals with the appropriate attitude and behaviour towards each other will give rise to a sustainable society that is highly sagacious, yet constantly innovating and discovering new options (at the social, political, economic and technological levels) to effectively manage the dilemma of “economy versus environment”. To explain the mechanics that will allow such a sustainable society to emerge, a ten-principle framework, referred to as the ‘General Theory of Sustainability’ (GTS), is introduced in this paper. GTS implements a fundamental concept that originated in psychology but is currently popular in other disciplines as well: the concept of empathy. By linking the process of resolving the issue of poverty (Principle 5Side Effect I in Principle 5 below) to the process of properly managing/using the physical environment (Principle 5Side Effect II in Principle 5 below) through the nurturing of empathic skills across all classes of the society, GTS captures the essence of how sustainable development can be achieved. Before we present the GTS (Section 4), we first provide a historical background of the concept of empathy (Section 2), and explain why and how empathic skills are needed to achieve sustainability (Section 3).

2. HISTORICAL BACKGROUND OF THE CONCEPT OF EMPATHY

According to the New Dictionary of Cultural Literacy (2002), the concept of empathy refers to the process of “identifying oneself completely with an object or person, sometimes even to the point of responding physically, as when, watching a baseball player swing at a pitch, one feels one’s own muscles flex” (Hirsch et. al., 2002).

The term empathy was coined by American psychologist Edward Titchener in 1909 as a rendering of the German word Einfühlung, which was in turn introduced by German philosopher Robert Vischer in 1873 (Wispe, 1987). While Titchener (1924) defined Einfühlung as a “process of humanizing objects, of reading or feeling ourselves into them”, this German term was originally used by Vischer in aesthetics (the branch of philosophy that deals with the nature and expression of beauty, as in the fine arts) to designate “the projection of human feeling on to the natural [or physical] world” (Pigman, 1995). For instance, to truly appreciate a work of art, one should imaginatively put oneself in the context of that work, time and place of that work. For the last quarter of the 19th century, the term Einfühlung remained focused on the process of perceiving and understanding the non-human, until another German philosopher, Theodor Lipps, extended its application in 1903 to the issue of how we get to know others, and described it as the source of our knowledge about other individuals. Thus, Lipps is the one who is credited with organizing
and developing the theory of Einfühlung for psychology and, as was indicated above, Titchener with introducing the concept in the Anglophone academic world under the term ‘empathy’. During the last hundred years, the research on empathy has spanned several disciplines including not only the traditional ones such as social, developmental and clinical psychology, and philosophy, but also business management, sales and marketing, construction and civil engineering (Valero and Visiland, 2006), and engineering education (Rowland, 1989).

3. EMPATHY TO ACHIEVE SUSTAINABILITY: WHY AND HOW

Addressing the dilemma of “economy versus environment” is equivalent, in essence, to finding the way to sustainable development by taking into account the three bottom-lines (economic, environmental, and social). To deal with the complexities of “sustainable development” and the challenges that are associated with it, several computational approaches have been investigated by researchers, ranging from advanced statistics and knowledge engineering, to artificial intelligence and computer simulations. Based on mathematical models, these are bottom-up methodologies that attempt to look at the basic data sets about the (environmental, economic, business or social) phenomena under study, infer useful information for guiding the process of moving out of un-sustainability, and thus build a body of knowledge for developing plans and policies for sustainable development to be used by professionals and policy-makers. In many cases, these methodologies provide insight into the dynamics of the systems and phenomena under study, and can be of substantial help to policy makers. But they have a number of limitations that cause them to fail miserably in real-world situations, which tend to be dominated by a high degree of non-linearity (i.e., governing variables do not change in a proportional way) and by a large amount of uncertainties. For instance, they failed to predict or help prevent the dramatic collapse of the once plentiful cod fisheries in Easter Canada in 1992 (De Alessi, 2008), as well as the severity of the 1974-1975 and 1981-1982 recessions in industrialized countries (Greenwald, 1994). They also fuel a great deal of debate in the area of climate change, as they don’t seem to provide clear explanations to the numerous uncertainties around the possible impacts of climate change such as, for example, the rise of the sea level (e.g. Oppenheimer and Alley, 2005). More recently, they have been blamed, at least in part, for the subprime financial crisis, as they failed to spot this crisis before it happened (e.g. Rickards, 2008). Some researchers attempted to explain the reasons behind such failures (e.g. Guergachi and Boskovic, 2008), and pointed out that, while extra research on these bottom-up mathematically-based methodologies will definitely contribute toward the enhancement of their effectiveness and expand their applicability, there is no chance for these methodologies to meet, on their own, the challenge of fully understanding and predicting the changes in highly complex systems such as, for example, the climate or the stock market. A Top-down approach must also be developed and investigated to complement the bottom-up mathematically-based ones. This paper intends to propose such a top-down approach in the form of 10 principles that make statements about a variety of issues including the limitations of systematic thinking (this expression is explained below in Section 4) and engineering methods, poverty, wealth-generating economic activities, economic theory, civil society and voluntary sector, gender, management of uncertainty and the physical environment. The common thread that weaves through all of these issues is the fundamental concept of empathy which, as described above, originated in psychology but is currently popular in other disciplines as well.

In this context, the hypothesis that we propose to examine is as follows:

The environmental and economic sustainability of a community as whole is closely dependent on the micro-interactions that take place among the individuals of this community; the more empathy exists in these micro-interactions, the more sustainable the community will be.

This hypothesis attempts to link a macro-phenomenon, which is sustainability, to indicators that are reported at the micro-level in the community. It looks at establishing this link in a way that is consistent with the principle of parsimony, which is a major pillar of science
(Outhwaite and Turner, 2007). Instead of associating sustainability with a long list of individuals’ attitudes and behaviors such as, for instance, being environmentally-friendly and frugal, saving energy, using public transportation, eating healthy, being innovative at work, being honest and socially responsible, etc., we want to identify the smallest set of characteristics which, once they are met by a large proportion of individuals in the community, will lead to sustainability as a direct consequence. In this GTS that we propose in Section 4, we focus on one characteristic: empathy.

Empathizing versus systemizing:

In his book “The Essential Difference: Male and Female Brains and the Truth about Autism” (2004), Simon Baron-Cohen contrasted empathy with systematic thinking, and argued that people can be placed on a mind continuum (MiC) ranging from systematic to empathic. He also advanced his thesis that, on average, females are more likely to be on the ‘empathic’ side of the continuum, males on the ‘systematic’ side, and that autism is nothing but an extreme form of the male condition1. Within the proposed GTS, problem-solving approaches are also considered to form a methodological continuum (MeC) ranging from the purely systematic methods to the purely empathic ones. The GTS attempts to construct bridges between the two extremes of this continuum, by leveraging the MiC (i.e., the respective skills available in the female and male populations and the complementarities that exist among them) to help address the world’s challenge of sustainability.

According to Baron-Cohen, systematic thinking “involves identifying the laws that govern how a system works. Once you know the laws, you can control the system or predict its behaviour”. In essence, this is what the bottom-up mathematically-based methodologies (see our discussion above) are intended to do. To approach the dilemma “environment versus economy” (which is at the heart of the sustainability issue) using systematic thinking, one would formalize this dilemma as a mathematical optimization problem. A fast-growing economy generates a great deal of wealth, but damages the environment, while a slow economic growth will be gentle on the environment, but generates very little wealth; between these two extreme situations, there would be an optimum that we could search for using various mathematical optimization techniques. Identifying such an optimum is useful and can be successful when it focuses on small-scale cases, such as a small community or a specific lake for example. But, it will lead to nowhere if it attempts to tackle a large-scale, multi-dimensional, highly uncertain and nonlinear case, such as the world’s climatic changes and energy/natural resources (it would indeed require a model of the real-world issue, which is very complex and can never be fully accounted for using any mathematical model — without even mentioning the usual obstacles that are posed by the so-called ‘curse of dimensionality’). Thus, we need to start thinking beyond systemizing our problem-solving approaches, and reach out to empathic approaches. In what follows, we explain how we intend to do it, by stating the principles that underlie our GTS.

4. STATEMENTS OF THE PRINCIPLES OF THE GENERAL THEORY OF SUSTAINABILITY (GTS):

Defining the meaning of sustainability:

Before stating the principles of the GTS, it is appropriate to look first at the meaning of ‘sustainability’. A number of definitions have been proposed for the term ‘sustainability’. The most well-known definition is due to Brundtland (1987): Meeting the needs of the present generation without compromising the ability of future generations to meet their needs. This definition is, however, not very practical, because there is no way for us to obtain information about the generations that will live 500 or 1000 years from now, their

---

1 This paper is not concerned with autism, but a few comments on the gender issue as it relates to sustainable development and the internal consistency of our GTS will be made later on in this paper.
technologies, their discoveries, their life styles, etc. This makes it impossible to define what their needs will be and, thus, difficult to operationalize the Brundtland definition. In addition to being impractical, this definition also looks a bit ironic: we did NOT even meet the needs of some people in our current generation\(^2\), and we are thinking about the needs of future generations, whose circumstances are unknown to us.

For the purpose of this paper, the following definition is proposed:

“A system is not sustainable if there is a (high) probability for its current dynamics to lead to a crash (collapse)”

This definition can be applied to any system, whether it is a lake, a watershed, a business enterprise, a water management framework, a society, or a civilization. Also, it allows us to parameterize the task of assessing sustainability, by using different levels of risk and different types of crashes:

- we can define what we mean by a crash and select the severity of such a crash
- we can specify the probability at which the crash becomes a concern

Such a parameterization is useful for assessing sustainability through an interaction with the stakeholders. For instance, one can ask these stakeholders “what level of sustainability are you looking for?” (a question which can be addressed in terms of the severity of a crash and probability of this crash), in the same way as financial advisors regularly ask their clients “what level of risk can tolerate?”

Principles of the general theory of sustainability (GTS):

Now we turn to the principles of GTS. Note that no claim is made herein that the GTS is complete and the statements of its principles are final.

In the statements that follow, the expression “systematic thinking” covers the bottom-up mathematically-based approaches that are traditionally used in engineering and physical sciences, and that consist in developing systematic procedures, mathematical equations and/or computer algorithms to resolve the problem at hand. It also refers to Baron-Cohen’s process of systemizing, as opposed to empathizing (Baron-Cohen, 2004).

**Principle 1:** Systematic thinking will **not** be able to comprehend all the complexities that underlie human nature, the wealth generation and distribution processes in fair economies\(^3\), and many aspects of the physical world including climatic changes, weather patterns, and water cycle dynamics. It will not be possible to resolve the dilemma “environment versus economy” using systematic thinking alone.

**Principle 2:** The concept of empathy can be broadened to become the process of identifying oneself with not only (1) the non-human (i.e., objects such as a painting, a novel, a music, a product, etc.), and (2) the human (other fellow citizens), but also (3) the abstract laws that govern the physical and environmental systems. Under this extended definition, empathic skills, when they are honed correctly and utilized towards these abstract laws, can become a source of knowledge (which may remain in an implicit form) about many of the systems around us. The idea of having humans empathizing with the laws of nature (as opposed to just objects and humans) may seem strange, speculative or metaphysical, but it should not be seen that way, because it is epitomized everywhere

---

\(^2\) Over one billion people in the world lack access to safe water supplies. Roughly two-million people die each year of diarrhea caused by infectious water-borne diseases (Secretariat of the Convention on Biological Diversity, 2010).

\(^3\) Note that we used the expression “fair economies”, not “market”, “planned” or “mixed” economies. It might be possible to create an economy where the wealth generation and distribution processes are in the hand of a small number of people, in which case wealth generation and distribution will be easily captured and predicted using systematic thinking, but this type of economy would not be fair. Depending on the circumstances, a fair economy could be a market, planned or mixed economy.
around us: people learn to swim without having to study Archimedes’ law of buoyancy or take a course on fluid mechanics, babies manage to stand up and walk without knowing anything about mechanics and control engineering, and many basketball players are able to score far away from the basket while they know nothing about Newtonian physics! It is possible to acquire highly complex knowledge about the systems around us without resorting to systematic thinking.

**Principle 3:** Empathic skills will not excel over systematic thinking in the cases where this thinking can be successfully utilized (e.g., most areas of engineering, computing sciences, etc.). However, empathic skills may achieve better results in the cases where systematic thinking fails (e.g., the case of resolving the dilemma “environment versus economy”).

**Principle 4:** While individuals of a society should strive to learn how to empathize with objects [(I) in Principle 2] and with the abstract laws that govern the systems around us [(3) in Principle 2], empathy towards the human [(2) in Principle 2] is the most crucial one for two reasons:

- **Reason I:** Empathic skills towards humans are relatively easy to learn and hone. This is because it is easier to identify ourselves with an individual from our own species than with a painting or with the laws of mechanics. We indeed know how a poor hungry man feels (as we may have experienced hunger and/or poverty before), but we wouldn’t easily feel how an object accelerates when it is hit by another object five times heavier or how fast it will take a tomato to biodegrade in our backyard.

- **Reason II:** Because empathy towards the human is easy to learn, hone and practice, it helps in developing the intellectual, mental and emotional abilities that are necessary for learning the empathic skills towards the non-human and towards the abstract laws of physical and environmental systems.

**Principle 5:** Economies, regardless of their type (market, planned or mixed, fair or not), are likely to produce the following negative side effects:

- **Side Effect I:** some kind of poverty in the society
- **Side Effect II:** some kind of impact on the physical environment

**Principle 6:** Attempting to address the side effects of Principle 5 from within the economic system, such as for example:

a) disallowing layoffs to ensure that poverty will not spread in society;
b) forcing businesses to eliminate the negative impacts on the environment by having them pay the costs of mitigating these impacts;

will end up, in the long-term, damaging the wealth-generating economic activities.

**Principle 7:** To address the two negative side effects of Principle 5 in a sustainable fashion, it is crucial that empathic skills are developed in the society at the individual level in a way that is consistent with Principle 4, i.e., individuals need to start by first learning, honing and practicing empathy towards humans, as it is the most important of all three empathies. Practicing empathy towards the human within the society must lead to the development of a space, outside the wealth-generating economic system, in which charitable giving and volunteering in all forms prevail. The development of such a space will, in turn, lead to solving or, at least, reducing the severity of Side Effect I in Principle 5. Thus, the economy will run in the society as if it was a coin with two faces: one face is the wealth-generating economic system which encourages quality customer service, innovation, competition, efficiency and effectiveness, research and development; the other face is the charity system whose goal is to address Side Effect I in Principle 5. At this point, the reader may recall that, at the very beginning of this document, we also described the environment as a coin with two faces: bounties on one face, and threats on the other one. Therefore, within our GTS, the economy and the environment are both viewed as two metaphorical coins.

**Principle 8:** Because of Reason II in Principle 4, success in solving or, at least, reducing the severity of Side Effect I in Principle 5 will help individuals in the society develop the intellectual, mental and emotional abilities that are necessary for learning the empathic skills towards:
a) the non-human, i.e., objects which would include the physical products that are manufactured by the economy’s businesses
b) the abstract laws of physical and environmental systems, in which the economy’s businesses are necessarily embedded.

These individuals, who would have learnt all three types of empathic skills, will be among the contributors (engineers, accountants, secretaries, technicians, salespeople, nurses, CEOs, etc.) to the wealth-generating economic system. When they get involved in product design, research, sales and marketing, finance and accounting, business development, manufacturing, and decision-making in general, they will act in a way that will not only satisfy the customer [empathic skills (2) in Principle 2] and produce quality products [empathic skills (1) in Principle 2], but also meet the environmental sustainability requirements [empathic skills (3) in Principle 2].

Thus, within our GTS, a society that runs an effective charity system (the second face of the economy’s metaphorical coin) to eliminate or reduce the severity of Side Effect I in Principle 5 is expected to run a successful wealth-generating economic system in an environmentally-friendly fashion. In other words, such a society is expected to have solved the dilemma “economy versus environment”.

Principle 9 — Taking into account the gender issue: In this principle, we start by pointing out the following two facts:

(1) Baron-Cohen (2004) advanced his thesis that most females are on the empathic side of the mind continuum, while males are on the systematic side of this continuum.

(2) Bankers working in the microfinance business have reported that, when women are involved in running sustainable development projects in developing countries, loans are almost always paid back on time, and the projects are successful (Attali, 2006).

Therefore, women’s empathic skills [(1) in Principle 9] seem to contribute beneficially to the complex projects of sustainable development [(2) in Principle 9]. Principle 9’s statement is thus: the lack of involvement of women in the process of resolving the dilemma “economy versus environment” may not lead to successful results.

Principle 10 — Management of uncertainty: Systematic thinking does not have the capability to handle severe or extreme uncertainties. When a society moves away from the path of sustainable development, uncertainties become more and more severe and, therefore, systematic thinking tends to fail. However, solving Side Effect I in Principle 5 through the implementation of Principle 7 and its consequence Principle 8, will lead to a reduction of the severity of uncertainty, thereby making systematic thinking discussed in Principle 1 applicable to a wider range of systems and issues and, thus, fuelling more economic growth — just as it happened at the time of the industrial revolution [when systematic thinking became established], but in a sustainable fashion.

5. CONCLUSION:

This paper proposed an integrated framework to resolve the dilemma of ‘environment versus economy’. This framework unfolds the essential characteristics of sustainable societies and outlines the mechanics that underlie sustainable development. It is composed of 10 principles that make statements about a variety of issues including the limitations of systematic thinking and engineering methods, poverty, wealth-generating economic activities, economic theory, civil society and voluntary sector, gender, management of uncertainty and the physical environment. The common thread that weaves through all of these issues in the GTS is a fundamental concept that originated in psychology but is currently popular in other disciplines as well: the concept of empathy.

As part of future work, the authors intend to test and validate the 10 principles and the logic that binds them together, using socio-economic and environmental data from the OECD countries. They also plan to look at the issue of how one can teach empathy, if
empathic skills prove indeed to be critical for sustainability. Teaching empathy can indeed become a challenging task when there are significant political differences within a society.

ACKNOWLEDGMENTS

The authors would like to acknowledge the financial support of NSERC, CFI, and Ryerson University.

REFERENCES


Israelson, D. “How the oil sands got to the great lakes basin: pipelines, refineries and emissions to air and water”, *Program on Water Issues at Munk Centre for International Studies*, University of Toronto, October 8, 2008.


Formalizing knowledge on international environmental regimes for integrated assessment modeling

M. de Vos¹, P. Janssen², S. Frantzi¹², P. Pattberg², A. Petersen¹, F. Biermann², M. Kok¹
¹Netherlands Environmental Assessment Agency (PBL), P.O. Box 303, 3720 AH Bilthoven, The Netherlands [martine.devos@pbl.nl]
²Department of Environmental Policy Analysis, Institute for Environmental Studies, Vrije Universiteit Amsterdam, De Boelelaan 1087, 1081 HV Amsterdam, The Netherlands

Abstract
International environmental regimes are considered key factors in dealing with global environmental problems. It is important to understand if and how these regimes are effective in solving or improving global environmental problems. In this paper we present a multidisciplinary approach to formalize knowledge on the effectiveness of environmental regimes. We constructed a conceptual framework to enhance systematic analysis of conditions that influence regime effectiveness and implemented this into a computer model using fuzzy logic reasoning. We applied the model in a preliminary analysis of two environmental regimes, the Convention on Biological Diversity and the United Nations Framework Convention on Climate Change. The model can be used to analyze past and future attempts to develop and implement environmental regimes, highlighting the determinants which contribute to success or failure. Results from these analyses can be used to improve scenario storylines in integrated assessment modeling. Although this paper shows that formalizing knowledge on environmental regime theory is not a trivial endeavor, it facilitates and improves the cooperation between scientists from regime theory and integrated assessment.

Keywords: regime; governance; institutions; integrated assessment; fuzzy logic

1. Introduction

The current generation of environmental problems has attributes which distinguish them from earlier environmental problems, and which need to be taken into account when thinking about policy measures. They concern ‘global public goods’, are transboundary and they are characterised by high uncertainty, multiple interests and complexity, requiring transdisciplinary approaches to deal with them. In order to address these environmental problems most effectively and efficiently, society has started to pursue global solutions. This has resulted in a series of systems of rights and obligations and related decision-making procedures in international environmental policy, also known as international environmental regimes [Biermann 2007]. International environmental regimes are considered key factors in dealing with global environmental problems [Biermann 2007]. It is therefore important to understand if and how regimes are effective in tackling these problems.

The creation and performance of international regimes to solve international environmental problems is studied by the field of international relations and more specifically by environmental regime theory. Traditionally, regime theorists have measured the effectiveness of regimes in qualitative terms, especially through the structured, focused comparison of different
case studies. However, to be able to draw lessons for policy making this approach suffers from problems of comparability and generalisability [Biermann et al., 2007]. Integrated assessment is a methodology to analyze global environmental problems by combining knowledge from the social, environmental and economic domains (People - Planet - Profit) relying strongly on quantification and computer simulation, while also recognizing the importance of including more qualitative forms of knowledge [Rotmans and Dowlatabadi, 1997]. For effectively dealing with problems such as climate change and biodiversity loss in a sustainable manner scientists in the field of integrated assessment acknowledge that it is important to include knowledge on environmental regimes in their analyses [Netherlands Environmental Assessment Agency, 2008]. But as attempts to quantify or model political processes have not yet been successful, knowledge on environmental regimes is often of minor importance in integrated assessments of sustainable development [Biermann 2007]. In the multi-disciplinary project Modeling Governance and Institutions for Global Sustainability (ModelGIGS) the challenge of bringing both worlds together this is taken up.

In this paper we present a cutting-edge approach to formalize knowledge on the effectiveness of environmental regimes. Eventual aim of our research is to include this knowledge in integrated assessments of global environmental problems. But as we don’t think that at this stage this type of knowledge can be fully integrated in integrated assessment (IA) models, this is envisioned to be an extension of current IA models. In practice, this means that the analysis of environmental regimes will be used to strengthen scenario storylines for IA simulation models, e.g. by analysing policy options which explicitly account for potential success and fail-factors in establishing effective environmental regimes.

In our research we constructed a conceptual framework to enhance the systematic analysis of conditions that influence regime effectiveness. We expect added value of this framework as a tool to perform quick assessments of regimes and the options to improve their effectiveness. Therefore we implemented the framework in a computer model using fuzzy logic methodology [Zadeh 1965], a simple and straightforward way of linguistic reasoning. Our paper is divided into 5 sections. Section 2 describes the formalization of knowledge from environmental regime theory and the construction of the conceptual framework. Section 3 describes our method to model the framework using fuzzy logic. In Section 4 we apply our model in an illustrative, tentative analysis of two environmental regimes, the Convention on Biological Diversity (CBD) and the United Nations Framework Convention on Climate Change (UNFCCC) In section 5 we discuss our main findings on the usefulness and prospects of our framework and model. Throughout the paper we will reflect on methodological challenges we encountered in our research.

2. Formalizing knowledge and building a conceptual framework

2.1. Regime effectiveness

When measuring the performance of international regimes, scholars by and large focus on the behavioral change of key actors (i.e. states) and not on environmental improvements [Easton 1965; Underdal 2002]. Even though some scholars recognize the need to look at environmental improvement, measurement is difficult because it is almost impossible to disentangle regime impacts from influences that are independent from the regime. In this paper we use the terms ‘likelihood of regime formation’ and the ‘likelihood of regime implementation’ as representatives for the performance of international environmental regimes. Regime formation and implementation are two distinct phases in the development of a regime. While the former includes the negotiations among states, the latter includes the process of putting the regime’s stipulations into practice. We assume that if both stages are completed successfully, a regime
will have a good performance. We also assume that good regime performance will lead to environmental improvement.

In an assessment of literature on environmental regime theory [Frantzi 2010] we identified general findings on regime performance and translated this knowledge into a set of approximately 50 clear rules on the likelihood of regime formation and implementation. As an example the rules on the likelihood of regime formation can be found in appendix B.

2.2. Conceptual framework

From the rules we identified the determinants of ‘likelihood of regime formation’ and ‘likelihood of regime implementation’. The next step was to relate the determinants, i.e. input variables, and the two phases of regime performance, i.e. output variables, together in a conceptual model (figure 1).

We have distinguished the input variables in ‘context’ and ‘design’ variables. There are three categories of ‘context’ variables: ‘problem structure’, which refers to attributes of the environmental problem, '(state) actors’, which refers to the (state)actors who take part in negotiations on the regime, and ‘regime environment’ which concerns the background against which regime formation or implementation takes place. The ‘design’ variables refer to the influences that policy makers can exert to establish regime formation or implementation, and they in fact can mitigate or enhance the impact that the ‘context’ variables have on the likelihood of regime formation or implementation. For example the context variable ‘asymmetry’ describes the asymmetry in actors’ interests, which may negatively affect the likelihood of regime formation (figure 1). Designing policy measures such that ‘differentiated responsibilities’ are allowed for the different (state)actors, can help to mitigate the negative impacts of the asymmetry in their interests, thus enhancing likelihood of regime formation.

![Figure 1 Conceptual framework for the analysis of the likelihood of regime formation.](image)

The conceptual framework serves as a simple representation of determinants of regime performance. The categorization of the identified (input) variables into ‘problem structure’, ‘regime environment’ and ‘(state) actors’ provides some structure and overview. In regime theory, however, these categories are not recognized as separate entities with an associated aggregate indicator referring to their importance for regime performance, nor for a more hierarchical construction of the framework. Besides, although regime formation and implementation are processes in time, we could not find clear information in literature on the
temporal dynamics involved. Therefore time is not explicitly considered in our framework and all input variables are supposed to effect the output variable at the same time. We also expect that there are interactions between the input variables, but again, as we did not find clear information in literature, we so far treated all input variables as independent entities. Including linkages between different variables will then need to be based on expert judgment.

3. Modeling the framework

We modeled the conceptual framework to enable future integration of knowledge on the performance of international environmental regimes in integrated assessment analyses. We had a clear set of rules on the likelihood of regime formation and implementation. However, the corresponding determinants were difficult to define and data were ambiguous, uncertain, lacking or only available as expert knowledge. We therefore chose to model our conceptual framework using the fuzzy logic methodology [Zadeh 1965]. The core of the fuzzy logic methodology is the rule base. We could easily transform our rules on regime performance into a set of fuzzy rules to perform fuzzy logic reasoning. Another basic concept of fuzzy logic is that it uses linguistic variables, thus performing computation with words rather than numbers. It therefore provided a systematic and transparent way of dealing with the determinants of regime performance, for which a straightforward quantification was impossible. Furthermore, fuzzy logic is able to use all types of (quantitative and qualitative) information and yields concrete answers which could eventually be related to quantitative approaches in integrated assessment analyses.

In this section we explain our steps to assess regime performance in a ‘fuzzy logic’ manner. Data on the input variables may be based on indicators or on expert-judgment. We aim for basing the analysis on the more objective indicator-based approach, however due to lack of time in this paper we use expert-judgement for our case study analyses. For the implementation of our model we have used the freely available fuzzy toolbox of Babuška [1995], which runs under Matlab.

3.1. Quantification and fuzzification of input variables

First step is to quantify the input variables and translate them into linguistic categories, a process called fuzzification. Some input variables can only take a binary value, like the presence of ‘differentiated responsibilities’ which may either be true or false. For other input variables we will select quantifiable indicators as representatives, based on relevance and data availability. For example the variable ‘asymmetry’ (of state actors interests) which may be represented by an indicator like Gross Domestic Product of the countries involved, can take real numerical values. We will normalise the numerical values on a uniform scale from 0 to 10, to have a similar basis for all real-valued input variables. Variables with binary values are straightforwardly assigned to two linguistic categories YES – NO (or equivalently, present/absent, true/false). For input variables with numerical values we employ three linguistic categories LOW - MEDIUM – HIGH to expressing them linguistically. The actual translation of these numerical input values into these linguistic categories (i.e. the fuzzification) is established by means of membership functions for these categories (see figure 2 for an example). These membership functions assign to each value x of the input variable a membership grade between 0 and 1 which expresses to what extent this specific input-variable/value belongs to these categories (i.e. is LOW, MEDIUM, HIGH). For example, if the variable ‘asymmetry’ has value 6, it will be categorized as mainly medium (membership grade
0.8) and a little bit high (membership grade 0.2). For the output variables we employ 5 linguistic categories, viz. VERY LOW - LOW - NEUTRAL - HIGH - VERY HIGH, to obtain more differentiation in our statements on the likelihood of regime formation or implementation.

![Membership functions](image)

**Figure 2** Membership functions to translate in and output variables into linguistic categories.

### 3.2. Fuzzy inference model

The heart of the fuzzy logic method is the fuzzy rule base, which in fact reflects our knowledge on the system that we analyse. The fuzzy rules are described in terms of IF-THEN statements, relating input variables with the output variable. The fuzzy rule base for our conceptual framework has been obtained by first translating the rules in appendix B, which reflect the knowledge on regime performance as listed in [Frantzi, 2010], into IF-THEN statements. For example the rule “Great asymmetry of state actors’ interests decreases likelihood of regime formation.” would be translated as

1. **IF asymmetry is high THEN likelihood of regime formation is very low**

This rule involves a single input variable (asymmetry), but there are also rules which describe the combined effect of input variables, as for instance “If a problem is marked with great asymmetry of state actors’ interests, differentiation of responsibilities increases likelihood of regime formation”, in which a context variable (asymmetry) and a design variable (differentiation of rules) is involved. Notice that this rule is in fact a further specification of Rule 1 above, which illustrates the mitigating effect of a specific policy measure (differentiation of rules) on the negative impact of asymmetry on the likelihood of regime formation. To reflect this mitigating effect and to prevent unnecessary rule-conflict in our fuzzy rule base we have therefore replaced Rule 1 by the following set of rules:

1. **IF asymmetry is high AND differentiation is true THEN likelihood of regime formation is low**
2. **IF asymmetry is high AND differentiation is false THEN likelihood of regime formation is very low**

In this way we have constructed our fuzzy rule base directly from the rules in appendix B. Since the rules in appendix B do not cover all possible linguistic values of the input variables (e.g. there are no rules on likelihood of regime formation for situations where asymmetry is low or medium), we have artificially extended the fuzzy rule base by adding rules which conclude on neutral likelihood of regime formation in these situations, e.g.:

3. **IF asymmetry is medium THEN likelihood of regime formation is neutral**
4. **IF asymmetry is low THEN likelihood of regime formation is neutral**

With the fuzzy rule base thus constructed we can for a given set of input variables evaluate the likelihood of regime formation and implementation. This activity is called fuzzy inference and is performed by means of Mamdani’s min-max inference [Jang, 1997]. Each rule is activated by first determining the degree of fulfillment of the rule’s antecedent, which is equal to the membership-grade of the condition in the IF-part. The inferred implication of the rule’s antecedent is established by redefining the membership function of the rule’s conclusion, i.e. the
THEN-part. An aggregation process is performed next to infer the overall conclusion of all activated rules taken together.

3.3. Defuzzification: calculation of likelihood

The fuzzy inference of the previous step has resulted in an overall fuzzy conclusion on the likelihood of regime formation or implementation, represented in terms of an encompassing membership function obtained in the aggregation process. The final step in our fuzzy logic framework involves the back-translation of this fuzzy information into a crisp value for the likelihood. For this defuzzification we use the center of gravity method, which determines at what output value the area under the membership function is divided in two equally sized subareas.

This final output value of our model gives an indication of the likelihood that a regime will be formed or implemented in a given situation. A low likelihood does not necessarily mean that regime formation or implementation do not take place, but rather indicates that formation or implementation is very difficult in the given circumstances. In addition to this overall output value, we have also calculated the results from the rules on the three categories of context variables to determine the contribution of each category to the final output. These results however cannot be seen as intermediate or separate results, but should be used to gain insight in the principal factors influencing the end result.

4. Applying the framework on the international biodiversity and climate regime

As an illustrative example of how our model may be used, we made a tentative calculation of the likelihood of formation of two existing regimes: the Convention on Biodiversity (CBD) and United Nations Framework Convention on Climate Change (UNFCCC) (figure 3). Data were derived by making an educated guess based on expert assessment of the official documents of the regimes (appendix A).

![Figure 3: Tentative calculation of likelihood of regime formation for the Convention on Biodiversity and UNFCCC. Figure shows both overall likelihood results and the contribution of the three categories of context variables.](image)
Looking at the overall results, formation of the biodiversity regime appeared slightly harder than formation of the climate regime. The contribution of the three categories of context variables provides insight in the possible causes for this difference. The characteristics of the biodiversity problem, i.e. the problem structure, were less favorable for regime formation than the characteristics of climate change. Biodiversity is a cumulative problem, which means that it has local causes, but global effects, which decreases the likelihood of regime formation. Whereas climate change is a systemic problem, which means that both causes and effects are global increasing the likelihood of regime formation. Also, society considered the biodiversity problem less urgent and salient, i.e. put it lower on the political agenda, than climate change. On the other hand, factors concerning the (state) actors were more favorable in the biodiversity regime. This is mainly caused by the fact that there was a big coalition of countries in favor of the biodiversity regime, whereas the climate change regime was confronted with a big coalition against opposing it. Both regimes had favorable regime environment because of their scientific advisory bodies, which helped the negotiations with providing necessary scientific information.

The likelihood of regime formation in both regimes could be enhanced by trying to influence the variables that are diagnosed above. Negotiations for both regimes could for example be linked by countries with other issues to enable trade-offs in the negotiation process. Furthermore, negotiations concerning the biodiversity regime may benefit from differentiated responsibilities for the different (state) actors which could mitigate the high asymmetry between the actors’ interests. This analysis of design options to enhance regime formation and implementation will be developed further when we have performed a validation of our conceptual model on various regimes, preferably indicator-based.

5. Discussion an conclusions

The model can be used to analyse the successful or failed attempts to develop regimes in the past, highlighting which determinants contributed to this success or failure. Moreover it can be used to assess future situations where there is no regime formed yet, and to suggest promising options for successful regime formation. Preferably the analysis of regime formation should be linked further to regime implementation, as we would be most interested in those factors from the regime formation phase that are also relevant in the regime implementation phase (as de facto we see a situation of regime formation having taken place for most international environmental problems).

The interpretation of the quantitative output of the model is difficult, as there is neither a universally accepted definition nor metric of the likelihood of regime formation and implementation. However, the quantitative model results enable meaningful comparisons between regimes and policy measures. The contributions of the three different categories to the overall result may help to understand the causes and structures of regime effectiveness. The analysis of the two regimes in the previous section should be considered as illustrative for the type of outcomes we are working towards.

The model enhances the transparency of knowledge on environmental regimes since it explicitly specifies important determinant factors and their potential effects on regime formation and implementation, in terms of the underlying rule base in the fuzzy inference model. In this way it is expected to offer researchers in integrated assessment useful insights in environmental regime theory, whereas it also may challenge regime theorists to structure and increase their knowledge further. In constructing our model from the underlying knowledge base we were namely confronted with questions regarding appropriate chronology, hierarchy and interaction of the identified factors which influence regime performance. Discussing our model and its application with regime theorist could trigger them to address these questions, leading to a further extension and improvement of the knowledge base and our model.
A serious limitation of our approach is the subjectivity in the choice and scoring of indicators and in the definition of membership functions. Both data on indicators and parameters of the membership functions usually cannot be derived entirely from the literature and must be based at least partly on the expert knowledge and judgment of the model developer. Currently we have suggested rather indicative and pragmatic choices on these issues, but in the future more explicit expert elicitation should be invoked to obtain a better underpinning [cf. Cornelissen et al., 2004]

At this stage it is not yet possible to link our model to an integrated assessment model as the scope and variables differ too much. But results from analysis with the model can be used to improve scenario storylines in integrated assessment modelling. The storylines could account for the existence and performance of relevant future environmental regimes as projected by the model and the factors that need to be in place for good regime performance can be identified and included as well (currently that is not happening or even contradicting assumptions may be included). We believe, however, that the integration of the knowledge on environmental regimes in integrated assessment modelling is in its early days and needs further attention.

References


Biermann, F., Earth system governance' as a crosscutting theme of global change research, Global Environmental Change, 17, 326-337, 2007

Biermann, F., Pattberg, P., Heires, M., Governance and institutions for global sustainability, IVM Institute for Environmental Studies, Vrije Universiteit, The Netherlands, 2007

CBD, Convention on Biological Biodiversity, official website: http://unfccc.int


Netherlands Environmental Assessment Agency (PBL), The Netherlands in a Sustainable World. Poverty, Climate and Biodiversity, Second Sustainability Outlook, report nr 500084003, Bilthoven, The Netherlands, 2008


UNFCCC, United Nations Framework Convention on Climate Change, official website: http:// unfcc.int

Appendix A Input variable values (based on expert judgement) for calculation of likelihood of regime formation for the Convention on Biodiversity and UNFCCC

<table>
<thead>
<tr>
<th>Variable</th>
<th>Convention on Biodiversity</th>
<th>UNFCCC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Normalized value</td>
<td>Normalized value</td>
</tr>
<tr>
<td><strong>Problem structure</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urgency</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>Scientific uncertainty</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Regulation costs</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Saliency</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>Systemic problem</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>Collaboration problem</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Initial informal agreement</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Transaction costs</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Initial framework treaty*</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Positive sidepayments*</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Positive issue linkages*</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Negotiation costs*</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td><strong>Regime environment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preceding agreement</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Scientific advisory bodies</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td><strong>(State) actors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assymetry in actors’ interests</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Powerful pushers</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>Powerful laggards</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>Support of important states</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Cumulative cleavages</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>Number of actors</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Homogeneous actors</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Differentiated rules*</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>Positive or negative incentives*</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

* Variables in the category ‘negotiation process’; these can be deliberately influenced by the actors.
Appendix B  Rules on the likelihood of regime formation

1. The existence of a preceding agreement or policy dealing with the same or a similar problem enhances the likelihood of regime formation.
2. The higher the negotiation costs, the less likely is regime formation.
3. The higher the regulation costs, the less likely is regime formation.
4. In case of high transaction costs and scientific uncertainty, initial framework agreements followed by more precise agreements increase likelihood of regime formation.
5. High saliency of the problem, increases likelihood of regime formation.
6. If the environmental problem is urgent, an initial informal agreement increases likelihood of regime formation.
7. If the environmental problem is considered urgent by the majority of actors, then regime formation is more likely.
8. Systemic problems increase the likelihood of regime formation.
9. Cumulative problems decrease the likelihood of regime formation.
10. Scientific uncertainty decreases the likelihood of regime formation.
11. Consensual scientific information by scientific advisory bodies increases the likelihood of regime formation.
12. In case of cumulative cleavages, regime formation is less likely.
13. In case of cumulative cleavages, regime formation is more likely if there are positive or negative incentives.
14. Great asymmetry of state actors’ interests decreases likelihood of regime formation.
15. If a problem is marked with great asymmetry of state actors’ interests, differentiation of responsibilities increases likelihood of regime formation.
16. If the coalition of ‘pushers’ is more powerful than the rest, regime formation is more likely.
17. If the coalition of ‘laggards’ within a regime is more powerful than the rest, regime formation is less likely.
18. If powerful states within the issue area are part of a regime, then regime formation is more likely.
19. The fewer actors are needed to regulate an environmentally harmful activity, the more likely is regime formation.
20. If the actors needed to regulate a harmful activity are homogeneous, then regime formation is more likely.
21. In case of a collaboration problem, regime formation is less likely.
22. In case of a collaboration problem, regime formation is more likely if there are positive side-payments.
23. In case of a collaboration problem, regime formation is more likely if there are positive issue-linkages.
World Sustainability in the Climate Policy Sphere: A scenario assessment of different commitments for curbing CO2 emissions in a Sustainability Framework

Carlo Carraro,a Marinella Davide,b Elisa Lanzic and Ramiro Parrado d

a) University Ca’ Foscari of Venice, Fondazione Eni Enrico Mattei (FEEM) and Centro Euro Mediterraneo per i Cambiamenti Climatici (CMCC), carlo.carraro@feem.it
b) Fondazione Eni Enrico Mattei (FEEM), marinella.davide@feem.it
c) Fondazione Eni Enrico Mattei (FEEM) and Centro Euro Mediterraneo per i Cambiamenti Climatici (CMCC) and Advanced School of Economics (SSE) in Venice, elisa.lanzi@feem.it
d) Fondazione Eni Enrico Mattei (FEEM) and Centro Euro Mediterraneo per i Cambiamenti Climatici (CMCC), ramiro.parrado@feem.it

Abstract: This paper aims at analysing the possible outcomes of the Copenhagen Accord pledges, proposed after the United Nations Framework Convention on Climate Change (UNFCCC) 15th Conference of the Parties (COP-15). We consider different levels of cooperation that may arise and analyse them with a dynamic Computable General Equilibrium (CGE) model of the world economy. In addition, we introduce a novel element in the discussion by also considering the effect of the different climate targets on sustainability. This is done by using a new comprehensive measure, the FEEM Sustainability Index (FEEM SI), which summarises economic, social, and environmental components of sustainability. Being built within a dynamic CGE, the FEEM SI can be projected and compared across different scenarios. We find that whereas the environmental component is greatly improved at regional and world level thanks to the reductions achieved with emissions trading, the economic and social components are generally negatively affected. It is also found that, despite the usual debate on the economic costs of climate policies, the economic pillar is not significantly affected by international emissions trading, while the social one is negatively affected. By constructing a scenario in which revenues from the global emissions trading are recycled in sectors expected to increase economic and social components, we show that climate policies that are implemented along with policies aimed at improving social wealth are likely to lead to a more sustainable future.

Keywords: Climate policy; Computable General Equilibrium (CGE) Models; Sustainability.

1 INTRODUCTION

Dealing with climate change has become a matter of international interest since the adverse effects that come along with it will affect the whole world ecosystem and produce a set of differentiated impacts depending on the characteristics, vulnerability and adaptation capabilities of every country in the world. The difficult task of designing and implementing a feasible climate policy is also controversial due to the economic costs it entails and because it requires efforts that may compromise future growth, in developing countries. In these circumstances there has been a wide set of proposals coming from both developed...
and developing economies. The long path towards a comprehensive international climate agreement is in a decisive phase after the 15th United Nations Framework Convention on Climate Change (UNFCCC) Conference of the Parties (COP-15) held in Copenhagen on December 2009. One of the outcomes of the conference has been the Copenhagen Accord signed by several countries and particularly proposing quantified emission targets by a set of leading countries.

The different pledges of countries regarding emission abatement targets inevitably suggest the task of analyzing the possible outcomes of those efforts in environmental, economic and social terms, as addressed by different studies using also different tools or models. A particularly useful method for the assessment of the effects of mitigation efforts proposals is the scenario analysis. Following this framework, Elzen et.al. (2009) propose three scenarios depicting low and high abatement targets as well as a common effort from Annex I and Non-Annex I countries. Mitigation costs expressed as share of GDP are in the range of 0.01-2.24% with the higher cost when all countries in the world engage in abatement efforts related to the more ambitious target of -30% with respect to 1990 for Annex I and -16% with respect to baseline emissions for Non-Annex I countries.

In another study Mattoo et. al. (2009) perform a scenario analysis related to an abatement effort carried out by all countries in the world by means of a Computable General Equilibrium (CGE) model focusing on different manufacturing exports and output. The different scenarios relate to the adoption of emissions trading systems (ETS) or public transfers to developing countries. Their main findings are that manufacturing output and exports would face a decline especially in countries with higher carbon intensity. For low carbon-intensive countries these effects would be lower, and for some countries they could even be beneficial. Moreover, including trading in emission rights and transfers may intensify the production decline in the manufacturing industry.

This paper proposes a medium term assessment by combining vows for GHG reduction efforts from a set of countries in the brink of a major commitment to address climate change. Moreover, it introduces a new element in the discussion by considering a sustainability framework which takes into account the effect on economic, social and environmental spheres not only separately but also as a whole by means of a composite index - the FEEM Sustainability Index (FEEM SI) - to summarize current and future performance of countries under the different policy choices they may undertake. In this context, we study the effects of the different cooperation and emissions reduction scenarios that may arise from the implementation of the Copenhagen Accord targets proposed by the signing countries. This is done with the purpose of analyzing the costs and effects of those scenarios, not only in terms of economic costs, but also looking at the impact on overall sustainability.

2 The Copenhagen Accord

The adoption of the Kyoto Protocol\(^1\) under the United Nation Framework Convention on Climate Change (UNFCCC), was a first step towards a global emission reduction agreement. However, the Protocol did not reach enough international consensus to achieve the reductions necessary to mitigate the global effects of climate change. Furthermore, recent scientific evidence highlights that the target proposed is not sufficient to reduce the increasing anthropogenic emissions in the atmosphere. If governments take no action to stabilize the greenhouse gases concentration in the atmosphere, the average temperature by the end of this century may increase with a best estimate at the lower end of 1.8°C and at the upper end of 4°C (IPCC, 2007). In addition, a recent study shows that delaying climate policies would lead to higher costs of emissions reductions (Bosetti et. al. 2008). Therefore, there were high expectations on the outcome of the 15th UNFCCC Conference of the Parties (COP-15) held in Copenhagen on December 2009, whose original objective was to

\(^1\) The Kyoto Protocol was adopted in 1997 but entered into force in 2005. It commits most industrialized nations and some economies in transition (Annex I Parties) to reduce their overall emissions of GHGs to 5.2% below 1990 levels, in the period 2008 – 2012 (UNFCCC, 2008).
establish a new international climate agreement to replace the Kyoto Protocol expiring in 2012.

The path towards the future agreement for the post-Kyoto begun in 2007, during the COP-13 that took place in Bali. Focusing on long-term issues, Parties adopted an agreement on a two-year process, the Bali Road Map. This consists of a number of decisions representing the various tracks to reaching a future climate agreement and sets a deadline for concluding the negotiations at COP-15 and COP/MOP-5 in Copenhagen. Since then, conscious that negotiating positions on climate policy are very different, Parties met in numerous sessions of preparatory talks in order to reach a common position by December 2009. At the deadline of January 31st, more than fifty countries submitted their mitigation actions to the UNFCCC secretariat. Most developed countries merely confirmed the emissions reduction targets proposed during the preliminary negotiation talks. Quantitative targets submitted to the UNFCCC Secretariat at the end of January are summarised in Table 1.

<table>
<thead>
<tr>
<th>Country</th>
<th>Autonomous target</th>
<th>Coordinated Effort Target</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Annex I</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>13%</td>
<td>1% to -11%</td>
</tr>
<tr>
<td>Canada</td>
<td>2.52%</td>
<td>2.52%</td>
</tr>
<tr>
<td>European Union</td>
<td>-20%</td>
<td>-30%</td>
</tr>
<tr>
<td>Iceland</td>
<td>-15%</td>
<td>-30%</td>
</tr>
<tr>
<td>Japan</td>
<td>-25%</td>
<td>-25%</td>
</tr>
<tr>
<td>New Zealand</td>
<td>-10%</td>
<td>-20%</td>
</tr>
<tr>
<td>Norway</td>
<td>-10%</td>
<td>-40%</td>
</tr>
<tr>
<td>Russia</td>
<td>-15%</td>
<td>-25%</td>
</tr>
<tr>
<td>USA</td>
<td>-3%</td>
<td>-3%</td>
</tr>
<tr>
<td><strong>Non-Annex I</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brazil</td>
<td>97.4%</td>
<td>88.8%</td>
</tr>
<tr>
<td>Chinaa</td>
<td>156.8%</td>
<td>135.4%</td>
</tr>
<tr>
<td>Indiaa</td>
<td>157.4%</td>
<td>157.4%</td>
</tr>
<tr>
<td>Indonesia</td>
<td>354.0%</td>
<td>262.0%</td>
</tr>
<tr>
<td>Mexico</td>
<td>21.9%</td>
<td>21.9%</td>
</tr>
<tr>
<td>Republic of Korea</td>
<td>35.7%</td>
<td>35.7%</td>
</tr>
<tr>
<td>South Africa</td>
<td>-15.8%</td>
<td>-15.8%</td>
</tr>
</tbody>
</table>

a) Original target in terms of CO2 emissions per unit of GDP and set accordingly in terms of CO2 emissions from the BAU scenario

b) Original target in terms of CO2 emissions per unit of GDP, already achieved in BAU.

For what regards China and India, quantitative targets have been determined adjusting the intensity targets proposed. Analyzing our baseline scenario India already achieved its target so we decide to set its quantitative target as the baseline emissions. In the case of China, although it actually improves its carbon intensity by 27% this is not enough for the proposed goal so we assume a quantitative target that represents a reduction of -39% to -44% of baseline emissions in 2020.

### 3 Evaluating sustainability in a general equilibrium framework

Various computer models are commonly being used for the evaluation of sustainability impacts (Klaassen and Miketa, 2003) and, as argued by Böhringer and Löschel (2004), there is no specific model that would be suitable to assess all the impacts of sustainability. However, the authors suggest that CGE models are flexible in a way that they can incorporate several key sustainability indicators in a single micro-consistent framework. CGE models provide an open framework for linkages to sector-specific models and important relationships to other disciplines adopting an integrated assessment approach and, thus, may be used as efficient tools for sustainability impact assessment (Ibid.).

Although it is expected that climate policies should have a beneficial effect on the environment and climate, the actual effect on sustainability is not straightforward. Sustainability is a complex and multi-faceted concept and in order to study the effect of climate policies it is necessary to take into consideration environmental, as well as
economic and social variables. This can be done by using selected indicators, and then combining them in a unique measure. In particular, we use the FEEM Sustainability Index (FEEM SI) in order to assess the effect of the different anticipated scenarios. FEEM SI in fact is an index built within the dynamic CGE model ICES SI. Constructing the sustainability index within ICES SI allows projecting it in the future under different policy and growth assumptions and for each country and year during a specific time span.

ICES SI is a recursive dynamic model that generates a sequence of static equilibria under myopic expectations linked by capital and international debt accumulation. The static core of the model is based on different additions to the GTAP-E model designed to assess specific climate change impacts (Bigano et al., 2006; Bosello et al., 2006a, 2006b, 2007, 2008; Roson, 2003). The current aggregation used in ICES SI is conformed by 40 regions representing major single countries as well as selected macro regions with 17 economic sectors within every region. The exogenous assumptions from selected sources provide the basis for a mid term scenario based on yearly simulations starting from 2002 until 2020.

Thus, ICES SI is an ideal framework for the construction of a policy-oriented sustainability index. Firstly, the large database the model is based on makes it possible to calculate the index for several regions, and to create indicators using data relative to the different sectors. Secondly, the nature of a CGE model, in which all sectors and regions are interconnected, is ideal to capture the tradeoffs between different indicators. Finally, its dynamic framework produces data relative to a growth path which can be used to calculate the index in the future, and under different policy assumptions. This allows analysing how sustainability may change in the different scenarios conceived and to compare them with a baseline scenario in which there is no climate policy.

4 Emissions abatement scenarios

According to the proposals made by different countries we prepared a set of three mitigation scenarios with the scope of assessing the possible outcomes of autonomous and coordinated efforts to curb GHG emissions within a sustainability framework. For this purpose we set the ICES SI model to simulate an Emissions Trading System (ETS) in which we imposed the proposed reductions for every region with respect to 1990 CO2 emissions only.

1. **Minimum Unilateral Targets:** Resumes the single and autonomous efforts made by a set of leading regions in the climate policy sphere as shown in Table 1. Since these are the minimum efforts proposed without requiring any coordinated action, this represents the least emission reduction scenario that could be expected in 2020 as long as the proposing countries keep their promise.

2. **Concurred Effort Commitment:** The previous scenario may be considered as a starting reference since there are some countries who proposed ulterior reductions if a coordinated commitment is reached within a group of regions. The possibility of a coordinated action with higher abatement targets is depicted in this scenario in which most regions from the first scenario formulate more ambitious targets.

3. **Global Commitment:** Finally, we consider a third scenario where all countries in the world reach a commitment to reduce CO2 emissions. Countries sign up for a global ETS and agree to limit growth of global emissions to 19 % with respect to 1990. For the group of leading regions we set the more ambitious targets from the second scenario while for the rest of the countries, including India, we set the target that emissions in 2020 would be reduced by 30% with respect to Business as Usual.

---

2 The full list of indicators used to construct FEEM SI is detailed in Annex 1, Table A1.1, while Figure A1.1 illustrates its aggregation tree.
3 Detailed information on the model can be found at the ICES web site: http://www.feem-web.it/ices.
4 The regional and sectoral aggregation is in Annex 2 in Tables A2.1 and A2.2 respectively.
5 We consider only CO2 because it is the main GHG and produced mainly due to fossil fuel combustion.
These three different scenarios will allow computing the FEEM SI for the single regions and also for the whole world taking into account that all actions by individual countries will have an effect on future sustainability. For the sake of presenting results in a brief and comprehensible way we decided to show more detail in the leading regions proposing a climate policy target and confronting them with the rest of the world.

All climate policies implemented in the three scenarios take effect starting in 2010 and the proposed targets are reached gradually every year until 2020. The ETS module in the ICES SI model comprises efforts from all sectors in the economy to accomplish the selected target. Moreover, since such mechanism requires the exchange of emissions permits and the associated transfer of financial resources, international capital movements reflect a greater flexibility among regions.

### 4.1 Direct and indirect effects

The efforts made from leading regions and the rest of the world are summarized in Figure 1, where the profile of emissions is depicted in the area behind the graph showing the strong reduction achieved by leading countries in the first two scenarios proposed and also from the rest of the world in the last scenario. Variations in percentage with respect to 1990 levels are represented in the columns for the Business as Usual and the three additional scenarios showing the sets of leading regions, the rest of the world and the World as an aggregate. Growth of World emissions which in BAU are 93% higher than 1990 levels could be reduced to 40% or 34% in the first two scenarios, and to 19% if all countries commit for a coordinated climate policy in the Global Commitment scenario.

Implementing mitigation policies through an ETS system to reduce future emissions produces different effects on Leading regions’ economies and also on the rest of the world. Direct effects are the reduction in total emissions from the group of countries participating in the ETS with the corresponding direct cost reflected in the carbon price as mitigation costs. Indirect effects are interpreted through variations in GDP, summarizing the different interactions in the economic activity; ranging from higher costs due to the carbon price; to the carbon leakage as an increased production of emissions in the rest of the world produced by mitigation efforts enforced by a reduced group of countries. The indirect effect on the leading regions’ economies measured as a variation of GDP in 2020 is shown on Figure 2 along with the GDP in USD trillions (expressed in 2001 USD), confirming that the higher the restriction in future emissions the higher will be the cost. An additional outcome of the exchange of permits is the fact that using and ETS offers the optimal allocation of emission at the lower abatement cost. This explains the different indirect cost faced by the majority of participating regions and also the fact that some regions may

---

6 The details for the abatement targets for the 40 regions present in the ICES-SI model is in Annex 3.
benefit of that scheme such as USA, China and India in all three scenarios and the Annex I leading regions as a group as shown to the right of Figure 2.

![Figure 2: GDP in 2020 and the indirect costs of different abatement efforts](image)

Another interesting insight is that when the rest of the world does not participate in the mitigation policies they are also negatively affected by the final outcome of the policy. This constitutes an additional element of support to the claims from developing economies. Implementing climate policies may reduce growth in their economies, and this is a very important aspect in climate policy design at the global level. Moreover, the indirect cost of the climate policies analyzed in this exercise is between 0.6 to 1% of World GDP, which is close to the findings of Mattoo et al. (2009) and Elzen et al. (2009).

### 4.2 Sustainability effects

The FEEM SI has been calculated for each region separately and for the world as an aggregate. As climate policy is aimed at diminishing the effect of climate change all over the world, it is interesting to check the overall global effect on sustainability. As FEEM SI is based on three main pillars, namely economic, social and environmental sustainability, it is also relevant to check which parts of sustainability are positively and negatively affected by climate policies. If on the one hand, we would expect environmental sustainability to increase, on the other hand we would also expect economic and social sustainability to decrease due to the policy costs.

![Figure 3: Changes in World Sustainability according to the FEEM SI](image)
The overall effect on the world’s sustainability is negative for all policies considered. It is found that the stricter the climate policy is, the lower the sustainability. As it is possible to see in panel a) from Figure 3, the effect of the Minimum Unilateral Targets and of the Concurrent Effort Commitment are similar and show a small but consistent decrease in sustainability compared to the baseline, especially at the end of the period. In the case of the Global Commitment scenario sustainability decreases more, due to the strictness of the climate policy considered.

In order to better understand the causes of this decrease in sustainability, it is interesting to see what happens to each pillar of the FEEM SI. Panel b) of Figure 3 shows that economic sustainability decreases slightly compared to the baseline, with not much difference from the initial values. Social and environmental sustainability are instead much more affected, as shown in panels c) and d) respectively. Social sustainability decrease is greater the more stringent the policy. Environmental sustainability increases substantially and the more so the stricter the policy. Note that the improvement in the Minimum Unilateral Targets scenario and in the Concurred Effort scenario is much lower than the case in which all countries participate to carbon trading.

Calculating percentage changes with respect to the baseline scenario at year 2020 of the FEEM SI and its three main pillars under the three policy scenarios, allows better understanding the impact of the policies. From Figure 4 it is possible to see that the consistent increase in the environmental component does not manage to offset the decrease in the social one, thus leading to an overall decrease in sustainability. This is an interesting result. Usually, the main concerns about the costs of environmental policies regard economic costs, not so social costs. In fact, as highlighted in the previous section, the foreseen GDP losses are what could lead some countries to decide to not sign an international agreement on climate change.

4.3 Improving social sustainability while reducing GHG emissions

Because the scenarios considered above are biased towards the improvement of environmental sustainability, we have also considered a fourth policy defined as a Global Sustainability scenario in which the social component of sustainability is also taken into consideration. In this scenario, the same climate policy as the Global Commitment scenario is considered, but with the addition of a revenue recycling process. Revenues raised from the auctioning of emission trading quotas are redistributed partly within each country and partly from industrialised countries to the rest of the world. In industrialised countries 20% of the revenue raised from emissions trading is collected and redistributed to developing countries where it is invested to increase research and development, education, and health expenditures. This is like an international Carbon Fund where revenues are used to invest in the improvement of Millennium Development Goals. The rest of the revenues

---

7 This additional scenario is proposed in a companion paper with the scope of further analyzing the detailed effects of additional cooperation scenarios at the global level (Lanzi, Parrado, 2010).
8 The countries which give part of their revenues are EU countries, Australia, Canada, Japan, New Zealand, Russia, Switzerland and the USA.
are invested domestically, always in the same selected sectors, and the resources used up to 2.5% of Gross World Product (GWP).

![Figure 5: Changes in World Sustainability with an additional policy](image)

The FEEM SI shows that taking into consideration other indicators besides GDP, sustainability can be accounted for in a more comprehensive way. In the FEEM SI calculations economic sustainability is only mildly decreasing. However, societal aspects seem to be more affected. In fact, when a policy is undertaken in order to limit the adverse effect on social sustainability, the negative effects are limited and the changes in overall sustainability are very low, as shown with the Global Sustainability scenario. The additional policy aimed at giving support to selected sectors not only brings the sustainability path closer to the baseline but also accomplishes the intended environmental goals.

5 Conclusions

In the current debates and negotiations for the achievement of an international agreement on climate change after the UNFCCC 15th Conference of Parties (COP-15), it is interesting to see how the different cooperation scenarios would affect different world regions in terms of emissions reductions, and economic costs. The present work adds an innovative further analysis which studies the effect that the effective implementation of the Copenhagen Accord may have on sustainability. We aim at answering the question of what would be the consequence in terms of sustainability, if it were to be decided to undertake a strict international climate target agreement. This is done by using the FEEM Sustainability Index (FEEM SI), an index taking a novel approach towards sustainability being built within a dynamic Computable General Equilibrium model that allows for scenario studies.

It is found that whereas environmental sustainability is greatly improved at regional and world level thanks to the reductions achieved with emissions trading, economic and social sustainability can be affected to the point of leading to a decrease in overall sustainability. It is also found, despite the usual debate on the economic costs of climate policies, that economic sustainability is not significantly affected by climate policies based on international emissions trading, while social sustainability is worsened.

We believe that this shows the importance of taking into consideration other policies, particularly those aimed at improving societal wealth, simultaneously to climate change ones. To show this we consider a climate policy in which revenues are recycled partly domestically and partly abroad from industrialised to developing countries. All revenues are invested in sectors aimed at increase economic and social sustainability, namely research and development, education and health. This leads to advancements in sustainability, as its social component recovers thereby achieving a more equal distribution between countries and between the three different pillars of sustainability. In conclusion, climate policies that are implemented together with policies aimed at developing other aspects of sustainability are likely to lead to its future improvement and thus the general welfare of society. Emission reductions are not sufficient to ameliorate world sustainability unless an adequate balance between economic, social and environmental outcomes is maintained. Ideas and strategies for the achievement of these objectives are at hand, but the right incentive scheme still needs to be designed and implemented.
ACKNOWLEDGMENTS

This paper is part of the research of the Sustainable Development Programme of the Fondazione Eni Enrico Mattei. The authors would like to thank Alice Favero for collaborating on the climate policy research, and the FEEM SI team for making this work possible.

References

Dimaranan, B. V. (2006), Global Trade, Assistance, and Production: The GTAP 6 Data Base, Center for Global Trade Analysis, Purdue University.
Energy Information Administration (2009), Annual Energy Outlook.
Horridge, M. (2005), SplitCom: Programs to disaggregate a GTAP sector (Preliminary Draft), centre of Policy Studies, Monash University, Melbourne, Australia.
International Institute for Applied System Analysis (IIASA), (2009), GGI Scenario Database Ver 2.0, 2009, Available at: http://www.iiasa.ac.at/Research/GGI/DB/
United Nations Framework Convention on Climate Change (2010a), Information provided by Annex I Parties relating to Appendix I of the Copenhagen Accord, Available at http://unfccc.int/home/items/5262.php
United Nations Framework Convention on Climate Change (2010b), Information provided by Non-Annex I Parties relating to Appendix II of the Copenhagen Accord, Available at http://unfccc.int/home/items/5262.php
Annex 1: FEEM Sustainability Index characteristics

Description of FEEM SI

The FEEM SI is built on a comprehensive methodology divided in different steps\(^9\). Indicators are firstly selected from reliable and internationally-recognised literature, such as the EU SDS indicator list and the indicators from the Commission on Sustainable development of the United Nations. Indicators have been carefully evaluated to choose what we believe are the indicators that best fit in the sustainability framework within the potentials of the ICES SI model.

The chosen indicators are divided into three main components, namely Economic, Social, and Environmental. These main pillars are derived from a tree structure whose leaves are the indicators selected.\(^10\) In the economic components Research and Development (R&D) is combined with Economic Structure, which in turn is derived by combining GDP per capita (GDP p.c.) and Consumption as share of GDP (Cons). For what regards the social component, it is derived from three sub levels, Population Growth Rate (Population), Poverty, and Social Wealth. Poverty is composed by two indicators. The first is Food relevance in primary consumption (Food Relevance), which represents the proportion of national expenditure in primary goods spent on food, and it is a proxy for a poverty indicator according to the Engel law. The second is Energy per capita (Energy p.c.), which is the amount of energy consumed per capita, and it is a measure of the degree of connection to energy systems. Finally, the environmental pillar is composed of the Air, Energy, and Natural Endowment Natural Endow) subcomponents. Air is a composite index of carbon intensity of energy (CO\(_2\) intensity), and Greenhouse Gases per capita (GHG p.c.). Energy is calculated from three indicators, that is Energy Intensity, Imported Energy – a measure of energy security - and the proportion of Clean energy in overall energy consumption (Clean). Finally, Natural Endowment is calculated from an indicator relative to Water, that is water use over total available renewable water resources, and an indicator of Biodiversity, which combines the relative proportions endangered species in plants and animals.

Indicators are calculated from the output of ICES SI, which has been structured to be able to calculate the desired indicators for each region and year. This allows projecting indicators in the future under different policy and growth assumptions. In order to achieve full comparability between indicators, we apply a normalisation procedure. This is done according to a policy target-based benchmarking methodology, where benchmarks are derived from a wide policy review. Finally, indicators are aggregated onto a single sustainability measure according to a non-linear aggregation function. The aggregation methodology for the FEEM SI introduces a very novel approach in this field. In fact, it enhances the importance to consider the interactions between indicators by attributing weights to the single indicators as well as to all possible combinations of indicators belonging to the same theme or sub-theme. Weights are then combined with a special weighted average based on the Choquet integral, which is a tool particularly well-suited to deal with multi-attribute issues like sustainability.\(^11\)

---

\(^9\) Only a brief description of the FEEM SI is given here, as this is enough for the purpose of this paper. However, detailed information is available on the FEEM SI website at [www.feemsi.org](http://www.feemsi.org).

\(^10\) Figure A1.1 illustrates the aggregation tree of the FEEM SI.

\(^11\) A sensitivity analysis can also performed in order to verify the robustness of the aggregation methodology.
### Table A1.1 - List of indicators for the FEEM SI

<table>
<thead>
<tr>
<th>Theme</th>
<th>Sub-theme</th>
<th>Indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECONOMIC</td>
<td>ECONOMIC STRUCTURE</td>
<td>1. GDP per capita</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. Consumption expenditure as GDP</td>
</tr>
<tr>
<td></td>
<td>COMPETITIVENESS</td>
<td>3. Total R&amp;D expenditure as GDP</td>
</tr>
<tr>
<td></td>
<td>POPULATION</td>
<td>4. Population growth</td>
</tr>
<tr>
<td>SOCIAL</td>
<td>POVERTY</td>
<td>5. Share of food in primary goods consumption</td>
</tr>
<tr>
<td></td>
<td>PRIVATE EXPENDITURE ON SOCIAL SERVICES</td>
<td>6. Energy per capita</td>
</tr>
<tr>
<td></td>
<td>EDUCATION</td>
<td>7. Expenditure in insurance and pensions GDP</td>
</tr>
<tr>
<td></td>
<td>HEALTH</td>
<td>8. Public expenditure on education as GDP</td>
</tr>
<tr>
<td></td>
<td>POPULATION</td>
<td>9. Health expenditure by privates as overall health expenditure</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10. Overall health expenditure as GDP</td>
</tr>
<tr>
<td>ENVIRONMENTAL</td>
<td>CLIMATE CHANGE</td>
<td>11. Carbon intensity of energy</td>
</tr>
<tr>
<td></td>
<td>WATER</td>
<td>12. Growth rate GHG emission per capita</td>
</tr>
<tr>
<td></td>
<td>ENERGY</td>
<td>13. Water use as total renewable water resources</td>
</tr>
<tr>
<td></td>
<td>NATURAL RESOURCES</td>
<td>14. Energy intensity (energy/GDP)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15. Imported energy as overall energy use</td>
</tr>
<tr>
<td></td>
<td></td>
<td>16. Share of clean energy in primary energy consumption</td>
</tr>
<tr>
<td></td>
<td></td>
<td>17. Biodiversity index-plants</td>
</tr>
<tr>
<td></td>
<td></td>
<td>18. Biodiversity index-animals</td>
</tr>
</tbody>
</table>

### Figure A1.1 – Tree structure of the FEEM Sustainability Index
Annex 2: The ICES SI model

ICES SI is a recursive dynamic model that generates a sequence of static equilibria under myopic expectations linked by capital and international debt accumulation. The static core of the model is based on different additions to the GTAP-E model designed to assess specific climate change impacts (Bigano et al., 2006; Bosello et. al., 2006a, 2006b, 2007, 2008; Roson, 2003). Industries are modelled through a representative cost-minimizing firm. A representative consumer in each region receives income, defined as the service value of the national primary factors (natural resources, land, labour, capital). Demand for production factors and consumption goods can be satisfied either by domestic or foreign producers which are not perfectly substitutable according to the "Armington" assumption.

The dynamic behaviour of ICES SI has two essential sources. The first is endogenous as it is governed by capital and debt accumulation while the second one is based on exogenous dynamics that allows projecting a baseline in the future using as an input external forecasts of endowments and productivities that are not specifically formulated in the model. Growth is driven by changes in primary resources (capital, labour, land and natural resources) with 2001 as the initial year (GTAP 6 database). Dynamics is endogenous for capital and exogenous for others primary factors. Capital accumulation is the outcome of the interaction of i) investment allocation between regions and ii) debt accumulation.

The general equilibrium structure of the ICES SI - in which all markets are interlinked - is tailored to capture and highlight the production and consumption substitution processes in a social-economic system as a response to external shocks. In doing so, the final economic equilibrium determined takes into account explicitly the “autonomous adaptation” of economic systems.

ICES SI has a wide sectoral and regional disaggregation, which were fundamental for the construction of the sustainability indicators. Compared to ICES, new variables have been included to enrich the model and to calculate the sustainability indicators related to them. In order to do this, changes have been made both to the database and to the model itself. Some specific sectors, namely Research and Development (R&D), Education (Edu) and Health (Hea) were not included in the original GTAP database, though they were part of other more aggregate sectors. Thus, they have been divided with the aid of the SplitCom facility of the GTAP.

Some of the indicators selected for the FEEM SI are related to variables that are not included in the GTAP-6 database, namely use of water, biodiversity and renewable energy. For the construction of these, it has been necessary to include in the model external data relative to these variables. This has been done firstly by adding the variables to the dataset, and then linking them to the model. Linking exogenous variables to variables that are endogenous to the model allows simulating its future behaviour coherently with the endogenous path generated by ICES SI. In doing so it is feasible to obtain level values and percentage changes resulting from projections as well as comparative statics. In the specific case of the FEEM SI, it will be possible to calculate indicators over time.

---

12 Detailed information on the model can be found at the ICES web site: http://www.feem-web.it/ices.
14 SplitCom is a set of programs aimed at facilitating the addition of sectors to a standard GTAP database. With the use of external data, it is possible to split the existing GTAP sectors into a finer aggregation. More information can be found on the SplitCom website: http://www.monash.edu.au/policy/splitcom.htm. The reference manual for SplitCom is Horridge (2005).
15 The process of extending the database and including additional variables is described in Chapter 2 of the FEEM Sustainability Report (2009).
Table A2.1: ICES SI Regional aggregation

<table>
<thead>
<tr>
<th>No.</th>
<th>Code</th>
<th>GTAP 6 Countries</th>
<th>No.</th>
<th>Code</th>
<th>GTAP 6 Countries</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ARG</td>
<td>Argentina</td>
<td>21</td>
<td>MEX</td>
<td>Mexico</td>
</tr>
<tr>
<td>2</td>
<td>AUS</td>
<td>Australia</td>
<td>22</td>
<td>MEast</td>
<td>Bahrain, Iran, Iraq, Israel, Jordan, Kuwait, Lebanon, Palestinian Territory Occupied, Oman, Qatar, Saudi Arabia, Syrian Arab Republic, United Arab Emirates, Yemen</td>
</tr>
<tr>
<td>3</td>
<td>AUT</td>
<td>Austria</td>
<td>23</td>
<td>NorthAfric a</td>
<td>Algeria, Morocco, Tunisia, Libya, Egypt</td>
</tr>
<tr>
<td>4</td>
<td>BNLX</td>
<td>Estonia, Latvia, Lithuania</td>
<td>24</td>
<td>NZL</td>
<td>New Zealand</td>
</tr>
<tr>
<td>5</td>
<td>BRA</td>
<td>Brazil</td>
<td>25</td>
<td>POL</td>
<td>Poland</td>
</tr>
<tr>
<td>6</td>
<td>BUL</td>
<td>Bulgaria</td>
<td>26</td>
<td>PRT</td>
<td>Portugal</td>
</tr>
<tr>
<td>7</td>
<td>CAN</td>
<td>Canada</td>
<td>27</td>
<td>RUS</td>
<td>Russia</td>
</tr>
<tr>
<td>8</td>
<td>CHN</td>
<td>China and Hong Kong</td>
<td>28</td>
<td>ZAF</td>
<td>South-East Asia – Burma, Brunei Darussalam, (East) Timor Leste, Malaysia, Philippines, Singapore, Cambodia, Lao People’s Democratic Republic, Vietnam</td>
</tr>
<tr>
<td>9</td>
<td>DNK</td>
<td>Denmark</td>
<td>30</td>
<td>ESP</td>
<td>Spain</td>
</tr>
<tr>
<td>10</td>
<td>FIN</td>
<td>Finland</td>
<td>31</td>
<td>SWE</td>
<td>Sweden</td>
</tr>
<tr>
<td>11</td>
<td>FRA</td>
<td>France</td>
<td>32</td>
<td>CHE</td>
<td>Switzerland</td>
</tr>
<tr>
<td>12</td>
<td>RoFS</td>
<td>Rest of Former Soviet Union: Armenia, Azerbaijan, Belarus, Georgia, Hungary, Kazakhstan, Kyrgyzstan, Moldova, Republic of Tajikistan, Turkmenistan, Ukraine, Uzbekistan</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>GER</td>
<td>Germany</td>
<td>33</td>
<td>TUR</td>
<td>Turkey</td>
</tr>
<tr>
<td>14</td>
<td>UKI</td>
<td>UK and Ireland</td>
<td>35</td>
<td>AFR</td>
<td>Africa excluding North and South Africa - Angola, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Cape Verde, Central African Republic, Chad, Comoros, Congo (Brazzaville), Cote d'Ivoire, Democratic Republic of Congo (Zaire), Djibouti, Equatorial Guinea, Eritrea, Ethiopia, Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, Kenya, Lesotho, Liberia, Madagascar, Malawi, Mali, Mauritania, Mauritius, Mayotte, Mozambique, Namibia, Niger, Nigeria, Reunion, Rwanda, Saint Helena, Sao Tome and Principe, Senegal, Seychelles, Sierra Leone, Somalia, Sudan, Swaziland, Tanzania, Togo, Uganda, Zambia, Zimbabwe</td>
</tr>
<tr>
<td>15</td>
<td>GCM</td>
<td>Greece, Cyprus, Malta</td>
<td>36</td>
<td>RoAsia</td>
<td>Rest of Asia - Afghanistan, Bangladesh, Bhutan, Korea, Korea, Democratic People’s Republic of, Macau, Maldives, Mongolia, Nepal, Pakistan, Sri Lanka, Taiwan</td>
</tr>
<tr>
<td>16</td>
<td>IND</td>
<td>India</td>
<td>37</td>
<td>RoE</td>
<td>Rest of Europe – Albania, Andorra, Bosnia and Herzegovina, Croatia, Faroe Islands, Gibraltar, Macedonia, Monaco, San Marino, Serbia and Montenegro(Yugoslavia)</td>
</tr>
<tr>
<td>17</td>
<td>IDN</td>
<td>Indonesia</td>
<td>38</td>
<td>RoEU</td>
<td>Rest of the European Union – Czech Republic, Hungary, Romania, Slovakia, Slovenia</td>
</tr>
<tr>
<td>18</td>
<td>ITA</td>
<td>Italy</td>
<td>39</td>
<td>RoLA</td>
<td>Rest of Latin America - Anguilla, Antigua &amp; Barbuda, Aruba, Bahamas, Barbados, Belize, Bolivia, Cayman Islands, Chile, Colombia, Costa Rica, Cuba, Dominica, Dominican Republic, Ecuador, El Salvador, Falkland Islands (Malvinas), French Guiana, Greenland, Guadeloupe, Guatemala, Guyana, Haiti, Honduras, Jamaica, Martinique, Montserrat, Netherlands Antilles, Nicaragua, Panama, Paraguay, Peru, Puerto Rico, Saint Kitts and Nevis, Saint Lucia, Saint Vincent and the Grenadines, Suriname, Trinidad and Tobago, Turks and Caicos, Uruguay, Venezuela, Virgin Islands, British, Virgin Islands, U.S.</td>
</tr>
</tbody>
</table>
### Table A2.2: ICES-SI sectoral aggregation

<table>
<thead>
<tr>
<th>No.</th>
<th>Code</th>
<th>Sector Description</th>
<th>Comprising GTAP Database V6 Sectors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Food</td>
<td>Food-related commodities</td>
<td>Paddy Rice, Wheat, Other Grains, Other Crops, Vegetables and Fruits, Plant Fibres, Cattle, Other Animal Products, Raw Milk, Wool, Cattle Meat, Other Meat</td>
</tr>
<tr>
<td>2</td>
<td>Forestry</td>
<td>Forestry</td>
<td>Forestry</td>
</tr>
<tr>
<td>3</td>
<td>Fishing</td>
<td>Fishing</td>
<td>Fishing</td>
</tr>
<tr>
<td>4</td>
<td>Oth _Ind</td>
<td>Other Industries</td>
<td>Beverages and Tobacco, Textiles, Wearing Appeal, Leather, Lumber, Paper and Paper Products, Fabricated Metal Products, Motor Vehicles, Other Transport Equipment</td>
</tr>
<tr>
<td>5</td>
<td>Coal</td>
<td>Coal</td>
<td>Coal</td>
</tr>
<tr>
<td>6</td>
<td>Oil</td>
<td>Oil</td>
<td>Oil</td>
</tr>
<tr>
<td>7</td>
<td>Gas</td>
<td>Gas, gas manufacture and distribution</td>
<td>Gas, Gas Distribution</td>
</tr>
<tr>
<td>8</td>
<td>Oil _Pcts</td>
<td>Petroleum, Coal Products</td>
<td>Water</td>
</tr>
<tr>
<td>9</td>
<td>Electricity</td>
<td>Electricity</td>
<td>Electricity</td>
</tr>
<tr>
<td>10</td>
<td>MServ</td>
<td>Market Services</td>
<td>Construction, Trade, Other Financial Intermediation, Insurance, Other Business Services, Dwellings, Transport</td>
</tr>
<tr>
<td>11</td>
<td>Ins</td>
<td>Insurance Services</td>
<td>Insurance and Pension Funding</td>
</tr>
<tr>
<td>13</td>
<td>Water</td>
<td>Water</td>
<td>Water</td>
</tr>
<tr>
<td>14</td>
<td>RD*</td>
<td>Research &amp; Development</td>
<td>Research &amp; Development</td>
</tr>
<tr>
<td>15</td>
<td>Edu*</td>
<td>Education</td>
<td>Education</td>
</tr>
<tr>
<td>16</td>
<td>Hea*</td>
<td>Health</td>
<td>Health</td>
</tr>
<tr>
<td>17</td>
<td>NMServ</td>
<td>Non- Market Services</td>
<td>Trade, Retail, Financial Intermediation, Renting</td>
</tr>
</tbody>
</table>

* Additional sectors included to construct the FEEM-SI
## Annex 3: Abatement by scenario

### Table A3.1: CO₂ Emissions growth in 2020 with respect to 1990 for three abatement scenarios

<table>
<thead>
<tr>
<th>Region</th>
<th>Minimum Unilateral Targets</th>
<th>Concurred Effort Commitment</th>
<th>Global Commitment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Target (%)</strong></td>
<td><strong>Effective Growth (%)</strong></td>
<td><strong>Target (%)</strong></td>
</tr>
<tr>
<td><strong>ETS - Leading Regions</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>13</td>
<td>29</td>
<td>1</td>
</tr>
<tr>
<td>Austria</td>
<td>-20</td>
<td>13</td>
<td>-30</td>
</tr>
<tr>
<td>Baltic countries</td>
<td>-20</td>
<td>-49</td>
<td>-30</td>
</tr>
<tr>
<td>BENELUX</td>
<td>-20</td>
<td>53</td>
<td>-30</td>
</tr>
<tr>
<td>Brazil</td>
<td>97</td>
<td>152</td>
<td>89</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>-20</td>
<td>-62</td>
<td>-30</td>
</tr>
<tr>
<td>Canada</td>
<td>2.52</td>
<td>26</td>
<td>2.52</td>
</tr>
<tr>
<td>China</td>
<td>157</td>
<td>96</td>
<td>135</td>
</tr>
<tr>
<td>Denmark</td>
<td>-20</td>
<td>-29</td>
<td>-30</td>
</tr>
<tr>
<td>Finland</td>
<td>-20</td>
<td>15</td>
<td>-30</td>
</tr>
<tr>
<td>France</td>
<td>-20</td>
<td>27</td>
<td>-30</td>
</tr>
<tr>
<td>Germany</td>
<td>-20</td>
<td>-19</td>
<td>-30</td>
</tr>
<tr>
<td>United Kingdom, Ireland</td>
<td>-20</td>
<td>-11</td>
<td>-30</td>
</tr>
<tr>
<td>Greece, Cyprus &amp; Malta</td>
<td>-20</td>
<td>41</td>
<td>-30</td>
</tr>
<tr>
<td>India</td>
<td>157</td>
<td>92</td>
<td>157</td>
</tr>
<tr>
<td>Indonesia</td>
<td>354</td>
<td>380</td>
<td>262</td>
</tr>
<tr>
<td>Italy</td>
<td>-20</td>
<td>29</td>
<td>-30</td>
</tr>
<tr>
<td>Japan</td>
<td>-25</td>
<td>-3</td>
<td>-25</td>
</tr>
<tr>
<td>Mexico</td>
<td>21.9</td>
<td>32</td>
<td>22</td>
</tr>
<tr>
<td>New Zealand</td>
<td>0</td>
<td>75</td>
<td>-20</td>
</tr>
<tr>
<td>Poland</td>
<td>0</td>
<td>-29</td>
<td>-30</td>
</tr>
<tr>
<td>Portugal</td>
<td>-10</td>
<td>75</td>
<td>-30</td>
</tr>
<tr>
<td>Russia</td>
<td>-20</td>
<td>-14</td>
<td>-25</td>
</tr>
<tr>
<td>South Africa</td>
<td>-20</td>
<td>-12</td>
<td>-16</td>
</tr>
<tr>
<td>Spain</td>
<td>-15</td>
<td>71</td>
<td>-30</td>
</tr>
<tr>
<td>Sweden</td>
<td>-15.8</td>
<td>1</td>
<td>-30</td>
</tr>
<tr>
<td>USA</td>
<td>0</td>
<td>8</td>
<td>-3</td>
</tr>
<tr>
<td>Rest of Asia</td>
<td>-20</td>
<td>78</td>
<td>116</td>
</tr>
<tr>
<td>Rest of EU</td>
<td>-20</td>
<td>-18</td>
<td>-30</td>
</tr>
<tr>
<td>Norway, Iceland, Row</td>
<td>0</td>
<td>62</td>
<td>55</td>
</tr>
<tr>
<td><strong>Rest of the World – No ETS</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Argentina</td>
<td>-</td>
<td>41</td>
<td>-</td>
</tr>
<tr>
<td>Rest of FSU</td>
<td>-</td>
<td>15</td>
<td>-</td>
</tr>
<tr>
<td>Middle East</td>
<td>-</td>
<td>247</td>
<td>-</td>
</tr>
<tr>
<td>North Africa</td>
<td>-</td>
<td>153</td>
<td>-</td>
</tr>
<tr>
<td>South East Asia</td>
<td>-</td>
<td>299</td>
<td>-</td>
</tr>
<tr>
<td>Switzerland</td>
<td>-</td>
<td>26</td>
<td>-</td>
</tr>
<tr>
<td>Turkey</td>
<td>-</td>
<td>253</td>
<td>-</td>
</tr>
<tr>
<td>Rest of Africa</td>
<td>-</td>
<td>302</td>
<td>-</td>
</tr>
<tr>
<td>Rest of Europe</td>
<td>-</td>
<td>-8</td>
<td>-</td>
</tr>
<tr>
<td>Rest of Latin America</td>
<td>-</td>
<td>149</td>
<td>-</td>
</tr>
</tbody>
</table>

* in global commitment all countries participate in the ETS

### Table A3.2: Carbon Price for three abatement scenarios USD/TCO₂

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum Unilateral Targets</td>
<td>2.5</td>
<td>5.1</td>
<td>7.8</td>
<td>10.9</td>
<td>14.4</td>
<td>18.4</td>
<td>23.0</td>
<td>28.2</td>
<td>34.3</td>
<td>41.2</td>
<td>49.2</td>
</tr>
<tr>
<td>Concurred Effort Commitment</td>
<td>2.8</td>
<td>5.8</td>
<td>9.0</td>
<td>12.7</td>
<td>17.0</td>
<td>22.0</td>
<td>27.9</td>
<td>34.8</td>
<td>42.9</td>
<td>52.2</td>
<td>63.0</td>
</tr>
<tr>
<td>Global Commitment</td>
<td>3.2</td>
<td>6.4</td>
<td>10.1</td>
<td>14.4</td>
<td>19.6</td>
<td>25.6</td>
<td>32.8</td>
<td>41.1</td>
<td>50.5</td>
<td>61.1</td>
<td>73.0</td>
</tr>
</tbody>
</table>
### Annex 4: Changes in sustainability by scenario

**Table A4.1: Percentage change in Overall Sustainability and its components with respect to baseline in 2020**

<table>
<thead>
<tr>
<th>Region</th>
<th>Minimum Unilateral Targets</th>
<th>Concurred Effort Commitment</th>
<th>Global Commitment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FEEM SI</td>
<td>Eco</td>
<td>Soc</td>
</tr>
<tr>
<td>Argentina</td>
<td>-0.7</td>
<td>-3.1</td>
<td>2.8</td>
</tr>
<tr>
<td>Australia</td>
<td>8.4</td>
<td>-0.7</td>
<td>-5.2</td>
</tr>
<tr>
<td>Austria</td>
<td>-1.4</td>
<td>-3.1</td>
<td>-2.6</td>
</tr>
<tr>
<td>Baltic Countries</td>
<td>-0.4</td>
<td>-6.4</td>
<td>-5.6</td>
</tr>
<tr>
<td>BENELUX</td>
<td>1.7</td>
<td>-0.4</td>
<td>-2.1</td>
</tr>
<tr>
<td>Brazil</td>
<td>0.4</td>
<td>-0.9</td>
<td>-2.8</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>8.2</td>
<td>1.2</td>
<td>-0.5</td>
</tr>
<tr>
<td>Canada</td>
<td>2.2</td>
<td>0.4</td>
<td>-2.5</td>
</tr>
<tr>
<td>Switzerland</td>
<td>0.6</td>
<td>1.3</td>
<td>4.0</td>
</tr>
<tr>
<td>China</td>
<td>17.6</td>
<td>2.2</td>
<td>-11.3</td>
</tr>
<tr>
<td>Denmark</td>
<td>7.4</td>
<td>6.8</td>
<td>-1.0</td>
</tr>
<tr>
<td>France</td>
<td>1.7</td>
<td>-4.0</td>
<td>2.7</td>
</tr>
<tr>
<td>Greece,Cyprus,Malta</td>
<td>-8.7</td>
<td>-16.3</td>
<td>-12.3</td>
</tr>
<tr>
<td>Germany</td>
<td>4.0</td>
<td>-0.4</td>
<td>-6.7</td>
</tr>
<tr>
<td>Indonesia</td>
<td>2.3</td>
<td>1.3</td>
<td>-4.1</td>
</tr>
<tr>
<td>Italy</td>
<td>4.7</td>
<td>0.5</td>
<td>0.9</td>
</tr>
<tr>
<td>Japan</td>
<td>0.4</td>
<td>-1.6</td>
<td>-0.7</td>
</tr>
<tr>
<td>Mexico</td>
<td>-0.5</td>
<td>-1.9</td>
<td>-5.0</td>
</tr>
<tr>
<td>Middle East</td>
<td>-1.7</td>
<td>-3.5</td>
<td>-0.2</td>
</tr>
<tr>
<td>Mexico</td>
<td>-0.5</td>
<td>-1.9</td>
<td>-5.0</td>
</tr>
<tr>
<td>North Africa</td>
<td>-1.5</td>
<td>-2.6</td>
<td>-0.1</td>
</tr>
<tr>
<td>New Zealand</td>
<td>1.3</td>
<td>1.3</td>
<td>-2.1</td>
</tr>
<tr>
<td>Norway, Iceland, RoW</td>
<td>6.4</td>
<td>-3.5</td>
<td>-8.1</td>
</tr>
<tr>
<td>Rest of Former Soviet Union</td>
<td>0.4</td>
<td>-1.6</td>
<td>-0.7</td>
</tr>
<tr>
<td>Rest of Africa</td>
<td>5.4</td>
<td>0.9</td>
<td>-0.7</td>
</tr>
<tr>
<td>Rest of Europe</td>
<td>-1.2</td>
<td>-16.4</td>
<td>-13.0</td>
</tr>
<tr>
<td>Rest of LA</td>
<td>3.6</td>
<td>-2.3</td>
<td>-3.2</td>
</tr>
<tr>
<td>Rest of America</td>
<td>4.6</td>
<td>0.9</td>
<td>3.6</td>
</tr>
<tr>
<td>Rest of LA</td>
<td>-0.8</td>
<td>-2.3</td>
<td>1.7</td>
</tr>
<tr>
<td>Russia</td>
<td>1.1</td>
<td>0.3</td>
<td>1.1</td>
</tr>
<tr>
<td>Rest of LA</td>
<td>4.6</td>
<td>0.9</td>
<td>3.6</td>
</tr>
<tr>
<td>South East Asia</td>
<td>4.1</td>
<td>0.6</td>
<td>8.9</td>
</tr>
<tr>
<td>South East Asia</td>
<td>4.1</td>
<td>0.6</td>
<td>8.9</td>
</tr>
<tr>
<td>South Africa</td>
<td>8.4</td>
<td>-2.7</td>
<td>-6.3</td>
</tr>
</tbody>
</table>
A Taguchi-Based Method for Assessing Data Center Sustainability

Sidney L. Pendelberry
Computing and Information Sciences
Rochester Institute of Technology, Rochester, New York
sidney.pendelberry@rit.edu

Sophie Ying Chen Su
Golisano Institute for Sustainability
Rochester Institute of Technology, Rochester, New York
sophie.chensu@rit.edu

Michael Thurston
Golisano Institute for Sustainability
Rochester Institute of Technology, Rochester, New York
michael.thurston@rit.edu

Abstract: This paper presents a methodology for measuring sustainability performance of data center facilities. Treating data center operation as a continuous production process, a Taguchi framework is used to estimate the sustainability impact due to power utilization from the facility by calculating the loss to society as the facilities power utilization becomes less efficient than the Environmental Protection Agency’s guidelines. By tracking data center sustainability performance with continuous production process metrics, it is easy to identify and study performance characteristics, which if out of specification, have broader environmental impacts. This method evaluates the data center’s excess energy consumption as a performance quality control issue, and estimates the cost of poor performance. Average power usage effectiveness (PUE) is a standard data center performance metric. In this case study PUE is tracked continuously to provide real-time operations and controls feedback. Further, PUE is mapped using the Taguchi framework and the U.S. EPA data center benchmark to the societal cost of non-conformance with the performance benchmark performance deviation based on carbon emissions and associated carbon exchange markets. It is expected that the method used in this paper may provide useful information to policy makers, engineers, and decision makers regarding sustainable IT data centers.

Keywords: Taguchi; data center; sustainability; PUE; carbon emission

1. INTRODUCTION

In 2010, the power consumption of United States data centers was projected to be 95.5 billion kWh (ENERGY, 2007). With data center power consumption at 1.5% of the U.S. total and growing rapidly, “Greener” data centers offer a large opportunity to reduce U.S. power consumption and thus reduce U.S. carbon emissions. This paper asserts that a data center is considered to be “Green” in the United States when the measure of sustainability, Power Usage Effectiveness (PUE) meets or exceeds the specification of the United States Environmental Protection Agency’s (EPA) best practice benchmark (ENERGY, 2007). PUE is defined as the ratio of the total power consumed by a data center to the power
consumed by the IT equipment that populate the facility (Belady, Rawson, Pfleuger, & Cader, 2008):

\[
PUE = \frac{\text{TotalFacilityPower}}{\text{ITEquipmentPower}}
\] (1)

In effect, the PUE represents the energy overhead associated with data center support systems. Total facility power is the sum of the IT equipment power demands and the support system power (security, environmental controls, lighting, etc.)

Treating sustainability as a continuous process of managing change, statistical process control techniques commonly used in manufacturing apply to the sustainability process to ensure that data center process variables such as PUE and carbon emissions stay in control (Di Mascio & Barton, 2001; Montgomery, 2009). A premise of this paper is that a data center facility will maximize its sustainability process as it approaches a PUE characteristic of 1.0, and further that a sustainability loss occurs when the PUE moves away from 1.0. A PUE of 2.0 means that for every watt of energy used to power IT equipment, an additional watt of energy is used to provide cooling, lighting, etc. While a PUE of 1.0 is ideal (given no internal data center renewable energy generation), this cannot be achieved but only asymptotically approached. We suggest a more practical PUE benchmark of 1.5 for the “product” specification (ENERGY, 2007; Google, 2010).

The Taguchi loss function was established to measure the impact of statistical process deviation from a target where a loss or cost function is calculated based on variance around the target value. On target processes incur the least overall loss (Pyzdek, 2003). The intent of this method is to support decision making regarding additional control and maintenance of a process. This paper applies the Taguchi approach (Di Mascio & Barton, 2001; Taguchi, Elsayed, & Hsiang, 1989) to gauge the sustainability process for data centers and the loss to society (as measured by a carbon emission metric) when the actual PUE deviates from the benchmark. The carbon emission related metric is used to quantify and unitize the loss function in sustainability terms, based on societal impact associated with green house gas emissions.

Using the EPA’s benchmark PUE as the ideal value for the performance characteristic (Greenberg, Mills, Tschudi, Rumsey, & Myatt, 2006), the closer to the actual value is to the ideal value, the “Greener” the data center. This is in essence Taguchi’s quality approach. The original purpose of the Taguchi-based measures was to quantify quality control in economic terms. It has been used in many applications, from biotechnology to software reliability.

Data center facility energy demands are a function of both the data center design and the operations and control strategies. While PUE is often used for evaluating data center design, the proposed Taguchi loss metric, calculated continuously, can be used a means of optimizing control in order to minimize overall sustainability impact while meeting service requirements. Once a data center has achieved the benchmark 1.5 average PUE, maintaining that level is a continuous process. The benchmark PUE value may also move (reduce) with technology advances in either data center facilities and IT equipment (Salim, 2009). Therefore, maintaining the same sustainability rating will require continual process or facility improvements. This state of change can be shown as a Deming Cycle (Deming, 2000) for the ongoing process of measuring and implementing data center efficiency. If at some point, the gap between best practice and current practice becomes too great, the process is considered to be out of control.

The PUE rating alone does not tell the whole story of data center energy efficiency. The PUE provides a measure of efficiency relative to the IT equipment consumption, but does not provide a measure of the total energy consumed by a data center or the associated carbon emissions. For example, two data centers of similar capacity and PUE, but with different IT equipment efficiency can have very different carbon emissions.

The next section, Methodology, describes in more detail the general framework for Taguchi’s quality assessment, and the suggested quality measure for data center sustainability. Following the Methodology section, Rochester Institute of Technology’s data center is used as a case study where the Taguchi framework is applied and the result shown. This paper concludes with a discussion of the results and some examples of common energy saving strategies that can aid in achieving the Taguchi product specification for a data center.
2. METHODOLOGY: TAGUCHI FRAMEWORK FOR SUSTAINABILITY ASSESSMENT

Taguchi’s loss function focuses on industrial production, and the need to produce a product within the tolerances as specified in the design. Taguchi argued that quality should start with the understanding of quality costs in various situations. He carried this one step further by considering the cost to society of manufacturing defects. Outside of the costs of rework, any part that is manufactured outside of its design specification and delivered to the customer would result in some loss to the customer through early wear-out; and problems throughout the system a chain effect due to the non-conformance of one part. These external losses, Taguchi argued, would eventually find their way back to the manufacturer. The general form of Taguchi’s loss function is shown in Equation 2 (Di Mascio & Barton, 2001; Taguchi et al., 1989).

\[ L = k(y - m)^2 \] (2)

In equation (2), \( L \) is the loss is in terms of dollars. For sustainability loss, an example would be: if a given data center is not meeting the best practice benchmark in Table 1, the loss in dollars is a function of the extra CO2 produced for the extra power consumed multiplied by the cost of carbon emission credits to offset the additional CO2. Parameter \( m \) is the target value of the measured output. For the data center sustainability application, it is defined as the expected CO2 emissions for the current data center if it achieved the “best practice” PUE. Finally, variable \( y \) is the currently measured (calculated) output value. For our data center sustainability formulation, it is defined as the current CO2 emission based on the actual PUE.

<table>
<thead>
<tr>
<th>PUE Benchmarks Scenario for Mid-Tier Datacenters</th>
<th>Year</th>
<th>2006</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current practice</td>
<td>2</td>
<td>1.9</td>
<td></td>
</tr>
<tr>
<td>Improve operation</td>
<td>2</td>
<td>1.7</td>
<td></td>
</tr>
<tr>
<td>Best practice</td>
<td>2</td>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td>State of art</td>
<td>2</td>
<td>1.5</td>
<td></td>
</tr>
</tbody>
</table>

The variable \( k \) is a scaling constant defined as the cost of the counter-measure or action taken to account for an individual part being outside of the specification, divided by the absolute value of the acceptable variation. This paper defines \( k \) for data center sustainability, as shown in Equation 3. It is the cost of the carbon credit divided by the absolute value of the difference between the CO2 released if the data center were operated at the “State of the Art (SA)” PUE and the CO2 released if the data center were operated at “Best Practice (BP)” PUE. The BP and SA reference values are taken from EPA recommendations and change over time. As technology improves, SA is adjusted to lower PUE values as design and implementation practices improve over time to utilize the new technology, the BP benchmark approaches the new SA. As noted earlier, the PUE benchmark asymptotically approaches an ideal value of 1.0, unless credit is given for local renewable power generation within the data center facility.

\[ k_{dc} = \frac{\text{CarbonCreditCost}}{|SA(CO_2) - BP(CO_2)|} \] (3)

Using the scaling coefficient defined above, our formulation of the Taguchi loss function for data centers is given in equation 4 below

\[ L_{dc} = k_{dc} \left( u(d[Act(CO_2(P_{dc}(t))) - BP(CO_2(P_{dc}(t)))])^2 \right) \] (4)
where Pdc represents the time varying power consumption of the data center, u is a unit (Heaviside) step function, Act(CO2) represents the current actual CO2 emissions of the data center facility at the operating loads and efficiencies, and BP(CO2) is as defined above. The unit step function makes the Loss Function a one-sided loss function, so as not to penalize performance exceeding the target value. The Act(CO2) parameter represents the Taguchi control variable (y) and the BP(CO2) represents the Taguchi target value (m).

2.1 PUE and Carbon Emissions

The data center loss function formulation assumes that Green House Gas (GHG) impact (as measured by Carbon Emission) is the key environmental control metric for data center operations. More comprehensive and sophisticated metrics could be derived, however there is a move internationally to regulate carbon emissions and active trading markets are developing, putting a real value (or cost) on these emissions. Equation 4 requires conversion of electrical power consumption to carbon emissions. While this conversion also varies by region and over time due to the mix of generation sources, this paper uses a conversion based upon the United States national average (ENERGY, 2007).

\[
\text{DC(CO}_2\text{)(tonne)} = \left( \frac{P_{dc}(kW\text{hr}) \times 1329.35(\text{lb}) \times 1(\text{MW}) \times 1(\text{tonne})}{\frac{\text{lb}}{1000\text{kW} \times 2204\text{lb}}} \right)
\]

(5)

2.2 PUE - US Standard

Using PUE as a benchmark standard is the first step for setting process control limits (Salim, 2009). The U.S. Environmental Protection Agency (EPA) published a PUE projection in 2006 for data centers in the U.S. (ENERGY, 2007). Depending on the different scenarios (e.g., server room / server closet, mid-tier, enterprise) the PUE is projected for mid-tier as shown in Table 1. The desired PUE number by 2011 is 1.5 for the best practice; this paper uses that benchmark as the control. The European Union is currently gathering PUE data of data centers from various agencies for monitoring individual data center performance (CENTRE, 2008).

2.3 Carbon Credits

<table>
<thead>
<tr>
<th>Carbon Offset Provider</th>
<th>Price (US$/Metric ton CO2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verus Carbon Neutral</td>
<td>$2.75*</td>
</tr>
<tr>
<td>Carbonfund.org</td>
<td>$10.00</td>
</tr>
<tr>
<td>Terrapass</td>
<td>$13.94</td>
</tr>
<tr>
<td>ClearSky Climate Solutions</td>
<td>$15.00</td>
</tr>
<tr>
<td>Bonneville Environmental Foundation</td>
<td>$29.00</td>
</tr>
</tbody>
</table>

*varies with the Chicago Carbon Exchange Price

Taguchi’s loss function is traditionally formulated such that the loss is represented in economic units. Equation 3 assumes that there is an economic cost (or value) that can be assigned to carbon emissions based on a market in Carbon Credits. Converting CO2 emissions to currency values using carbon credit equivalents are subject to location, market pricing and energy usage. Table 2 illustrates the current variation in carbon credit prices. The large degree of variation observed in the prices is reflective of lack of standardization in the basic definition of the credit, as well as lack of regulation in the markets. As these markets become more regulated, there will still be variation in value over time, as well as between different economies as they try to meet their international commitments to reduce emissions. However, within a particular economy, at a point in time, the variation in pricing should be significantly less. For the purpose of this study, we selected the lowest cost option for achieving data center carbon neutrality ($2.75/ton).

3. ROCHESTER INSTITUTE OF TECHNOLOGY DATA CENTER CASE STUDY
The example system considered here is the data center at the Rochester Institute of Technology (RIT). The most difficult part in calculating RIT’s PUE was collecting the data and determining the time frequency for analysis given the data available.

3.1 Data Collection

PUE is a moving target. The fluctuation in outside air temperature influences cooling demands, which effects energy consumption and introduces variation in the data center PUE. Conversely seasonal (e.g. monthly, quarterly, etc…) computational demands will also introduce variation in PUE. This predicates a need to do real-time data collection and PUE calculations.

The situation of less than ideal power metering locations available at the RIT data center is likely common with older data centers. The RIT data center has been around in one form or another for 30+ years and its growth has been more organic than planned, resulting in periodic adaptations to gather data.

Power consumption data for the RIT data center, measured at Point 1 over a 1-year period, are shown in Figure 1. This figure shows predictable seasonal fluctuations in current draw, with peak consumption during summer months and high demand periods around the end of the school year. The yearly average (2008) of power consumption at point 1 is approx. 250A, subtracting the 20A for outside HVAC service yields an average consumption of approx. 230 (at 480v) for the data center and direct supporting services. The observed seasonal variation results in a shift in consumption of about 60A or 28.8kW.

RIT has an inexpensive and unique data center temperature monitoring system using single wire sensor technology and a php web site that provides real-time temperature data. This data can be used as to continuously monitor performance of the HVAC controls, and also look for hot spots that represents high server consumption or ineffective cooling layout. Figure 2 presents a snap shot of the RIT data center real-time heat map, which can be found at http://data.center.rit.edu

Operating the data center at a higher mean ambient temperature offers opportunities to reduce the energy consumption associated with cooling. However, when increasing the operating temperature, it becomes increasingly important to understand the data center
temperature map and also temperature transients in order to stay within equipment operating specifications. Figure 3 below provides an overlay of the data center temperature with the data center current draw, illustrating the variation in these parameters over a one-day period in February of 2010. The data shows good temperature control (~1 deg C total variation) and a variation in current draw (30 minute moving average) of about 25A, or 10%.

Figure 3. Power Statistics from Point 4 in Real Time

3.2 Results

An engineering study of the RIT data center was commissioned in 2009 (Consultants, 2009) focusing on HVAC and UPS capacity. The data collection systems described above were used for the engineering study. The associated reports provide UPS and generator capacity, total annual power usage, and also break out the HVAC associated power and other auxiliaries. The data summaries from the 2009 report were used for the data center analyses that follow.

Table 3 below shows an analysis of RIT data center power consumption for 2008 and 2009, broken out by sub-system, and also total energy use. The following three years (2010-2012) are projections based on historical growth trends; a 15% annual growth rate for server power and cooling demand is assumed. For purposes of this analysis, it is assumed that the existing UPS system can meet the growing system demand although it is at near to 100% capacity.

Table 3. RIT Data Center Annual Average Power Usage Projections

<table>
<thead>
<tr>
<th>Item</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2008</td>
</tr>
<tr>
<td>Server power (kW) (15% annual growth)</td>
<td>70</td>
</tr>
<tr>
<td>Cooling (kW) (15% annual growth)</td>
<td>73</td>
</tr>
<tr>
<td>Lights (kW) 19 out of 33 lights are on 24 hours</td>
<td>2.8</td>
</tr>
<tr>
<td>UPS power loss (kW)</td>
<td>15</td>
</tr>
<tr>
<td>Total Power (MW-hr)</td>
<td>1410</td>
</tr>
</tbody>
</table>

Using Equation 2, Equation 3 and Table 3, Carbon footprint and PUE of RIT data center is calculated. Table 4 show the carbon emission projection for the RIT data center.

Table 4. RIT Data Center PUE, Carbon Emission

<table>
<thead>
<tr>
<th>Item</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5 shows the calculated Taguchi loss function for the RIT data center, using the EPA values and projections for state of the art and best practice, and using a carbon credit cost of $2.75. Using the actual data center carbon emission for 2008 (851 ton) and the selected carbon cost would suggest that RIT should purchase $2340 of carbon credits to be carbon neutral. The suggested Taguchi loss function formulation calculates a loss of $6360 for 2008. This value is higher than the directly calculated value due to the small spread between SA and BP PUE. The variation of RIT’s data center performance from BP is much higher than the variation between SA and BP. This provides an additional economic penalty for excessive deviation from best practice. Table 5 also gives the calculated Taguchi loss for the RIT data center through 2011, assuming no changes in the existing infrastructure.

Table 5. Taguchi Loss Calculation For RIT Data Center

<table>
<thead>
<tr>
<th>Year</th>
<th>Loss ($$)</th>
<th>k ($/tonCO2)</th>
<th>y (tonCO2)</th>
<th>m (tonCO2)</th>
<th>SA PUE</th>
<th>BP PUE</th>
<th>Actual PUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008</td>
<td>6360</td>
<td>0.185</td>
<td>851</td>
<td>666</td>
<td>1.76</td>
<td>1.8</td>
<td>2.3</td>
</tr>
<tr>
<td>2009</td>
<td>6327</td>
<td>0.107</td>
<td>964</td>
<td>722</td>
<td>1.64</td>
<td>1.7</td>
<td>2.27</td>
</tr>
<tr>
<td>2010</td>
<td>6884</td>
<td>0.070</td>
<td>1095</td>
<td>782</td>
<td>1.52</td>
<td>1.6</td>
<td>2.24</td>
</tr>
<tr>
<td>2011</td>
<td>7811</td>
<td>0.0488</td>
<td>1245</td>
<td>845</td>
<td>1.4</td>
<td>1.5</td>
<td>2.21</td>
</tr>
</tbody>
</table>

3.3 Change Analysis

There are several ways to reduce data center power consumption and associated carbon emissions. Upgrading to higher efficiency IT equipment reduces total energy consumption but may actually result in an increase in PUE. Increasing the overall data center energy effectiveness (reducing non IT related power) results in PUE reductions and associated energy savings. Below are a few examples of changes that can make a positive impact on PUE (examples are relative to the 2008 baseline).

1. Reduction in lamp fixtures from 19 to 6 results in an improvement of approx 0.03 in PUE and a reduction of 10 tonne of CO2 emissions per year.

2. If in addition to turning the reduction in operating fixtures to 6, the remaining fixtures are converted to LED lighting, PUE is reduced by approx 0.04 and CO2 emissions by 15 tonne/year.

3. HVAC cooling reduction – The RIT data center uses Liebert Chillers that are set to utilize free air passive cooling when the outside ambient air temperature drops below 45 deg F. Rochester New York is in a temperate climate and can utilize free air-cooling for approximately 5 months of the year. The 45 deg passive cooling limit is set based upon the targeted data center interior control temperature of 65 deg F. If the data center control temperature is raised to 75 deg F, which would allow passive cooling for approximately 7.5 months of the year, resulting in an overall 25% reduction in cooling load. This improvement results in a PUE improvement of 0.26 and a CO2 reduction of 96 tonnes of CO2.

Increasing equipment operating temperatures raises a concern of reduced equipment reliability or durability. As an estimate of this effect, using the Arrhenius time to failure model indicates that for each 10 °C (18 °F) temperature rise component life is reduced by 50%(Feng & Hsu, 2004; Pinheiro, Weber, & Barroso, 2007). Experience at RIT indicates that the current physical life of IT equipment is longer than the useful life due to technology obsolescence. Additional work research will be needed to insure that increased operating temperatures don’t result in actual reductions in data center reliability and availability.

4. CONCLUSIONS

This paper has described how sustainability metrics can be used for process control using a Taguchi loss function based method. In this case, sustainability performance
(impact on environment and society) is measured using the overall carbon emission. The standard form of the Taguchi loss function penalizes adverse process behavior and variation relative to a performance benchmark. We have suggested that the CO2 emission associated with benchmark data center PUE is an appropriate benchmark.

The employed form of the Taguchi loss function penalizes data center performance deviation that significantly exceeds the variation between best practice and state of the art benchmarks, resulting in increasing cost associated with non-compliance. These results indicate that the RIT data center infrastructure is out of date and requires significant improvements, if not complete overhaul, in order to achieve best practice and zero the loss function. The loss function formulation also results in no economic penalty when the data center is at benchmark performance. The justification for this is that advancement beyond best practice is often not cost effective and there may be better (more cost effective) opportunities to reduce carbon emissions elsewhere.

5. REFERENCES


Consultants, R. I. o. T. 2009. RIT Campus Data Center Power distribution Study.


Evaluation as a Tool to Support Decision Making in Municipal Waste Recycling System for Thailand Sustainable Environment

Vilas Nitivattananon¹ and Yuthtapong Wattanalapa²

¹Asian Institute of Technology (vilasn@ait.asia)
²Thailand Institute of Packaging and Recycling Management for Sustainable Environment (ypw2000@yahoo.com)

Abstract: This paper presents part of the results from a research project aimed at evaluating municipal waste recycling programs, supported by Thailand Institute of Packaging and Recycling Management for Sustainable Environment in collaboration with other recycling program partners in Thailand. The study applied sustainable municipal solid waste (MSW) management approach and program evaluation concepts. The evaluation design includes two main levels. At country level, a survey using a questionnaire was carried out to investigate perceptions on factors influencing success of representative samples including community recyclable banks (CRBs) and school recyclable banks (SRBs). A waste stream analysis at city level was conducted to examine the composition of waste at different selected urban centres. The results show that factors significantly influencing the performance of recycling programs are related to perceptions of source separation, facing inconsistent price problem, lack of skilled operators and financial support. The results of waste stream analysis reflect significant changes in waste proportion of some recyclables. The findings reveal key points, particular trends and status of each recycling program and each type of recyclable materials to the decision makers who are responsible to set sustainable MSW strategies and plans.

Keywords: Local government authorities; municipal solid waste management; private sector participation; program evaluation; sustainability

1. INTRODUCTION

The principles of sustainable municipal solid waste (MSW) management strategies are to: (i) minimize MSW generation, (ii) maximize waste recycling and reuse, and (iii) ensure the safe and environmentally sound disposal of MSW. The sustainable MSW management depends on the overall effectiveness and efficiency of urban managements, and the capacity of responsible municipal authorities (Kaosal, 2009). According to Hardoy et al. (2001), national governments inevitably have the key role in linking local and global sustainability. As Agenda 21 recognizes the need to strengthen and expand national waste reuse and recycling systems, it argues that multinational and national government institution and non-government organization should actively promote and encourage waste reuse and recycling. At the level of city and municipal government, the key policy areas to secure both development and sustainability are to give further incentives to encourage innovative way to reduce pollution and conserve resources.

With rapid growing waste generation rates and high costs of waste disposal, depletion of landfill space and the problem in obtaining new disposal sites (as most of those sites are becoming open dump and nearly exhausted) are unresolved issue in most urban centres. These problems are of critical concern to local authorities that are the centres of MSW management system and play an important role for changing the traditional to more sustainable approaches of the management (Nitivattananon and Gauger, 2004).
In Thailand, the positive sign of recycling promotion has been started in 1997 by Ministry of Science Technology and Environment (MOSTE). In response to the sustainable MSW approach, most on-going efforts regarding waste recycling are focused on encouraging participation of communities such as setting school recyclable bank (SRB), community recycling bank (CRB), and recycling centre at municipal level. Some LGAs have established composting facilities (CFs) and material recovery facilities (MRFs) at a small scale. These models have been developed and promoted since 1999 including: (i) more than 500 SRBs, (ii) more than 300 CRBs, (iii) more than 10 municipal recycling centres, (iv) more than 20 CFs, and (v) 2 MRFs. In addition, itinerant recycling groups have also been promoted. Thailand Institute of Packaging and Recycling Management for Sustainable Environment (TIPMSE), officially established in 2005, aimed to reduce packaging wastes with a safety and sustainable methodology, which focusing to reduce the packaging wastes composition in municipal wastes approximately 19% of total waste within 5 years (TIPMSE, 2007).

Recently, many research studies have been conducted to determine how to make recycling programs more successful and propose various approaches for a sustainable SWM. Thomas (2001) studies how the public understanding’s effect on recycling performance. Williams and Kelly (2003) presents the evaluation of the public perception towards a recycling scheme of a local authority. A strategy planning for drop-off centers has been extensively discussed in the paper by Chang and Wei (1999). Various authors (Ball and Lawson, 1989; Martin et al., 2006; Mcdonald and Oates, 2003) have studied how important factors – economic, political and social conditions- influent the success of the recycling programs and described people attitudes toward recycling programs.

Chula Unisearch conducted a research on packaging industry development for supporting solid waste management and recyclables. The research aimed to study an amount of solid waste which affected to environment, impact of solid waste management and recyclables to packaging industry. The result of waste stream analysis at selected sites was that the biggest amounts of waste are food (54.42%), plastic (24.76%) and paper (10.03%) (as also given in Figure 1). It was found the amount of packaging waste in whole country is 30.80% (Chulalongkorn University, 2004). Although with some data collection and compilation efforts by the government agencies, the overall evaluation of recycling programs and specific recycling models introduced by TIPMSE are yet to be evaluated.

Figure 1. Percent by wet weight of waste composition at the selected sites in 2004 (source: Chulalongkorn University, 2004).

This paper covers evaluation of recycling programs in two levels. At country level, it was carried out using samples of existing small-scale waste management of CRBs and SRBs by investigating their perception on factors influencing the performance of recycling programs. At city level, the investigation was conducted reflect changes in quantity of recyclable waste in municipal waste stream at selected LGAs compared to the previous study similarly conducted 5 years ago.

2. MUNICIPAL SOLID WASTE RECYCLING SYSTEM IN THAILAND

In 2009, a total of 15.03 million tons/year or 41,240 tons/day were generated. Only 3.5 million tons or equivalent to 23 percent of total waste is reported to be recycled. The
current recycling rate is low comparing to the potential recyclable waste of 12.5 million tons (83 percent approximately). The main factors influencing poor performance on recycling rate were limitations of good practices, lack of systematic procedure, public cooperation and budget (PCD, 2009).

LGAs have the direct responsibilities to handle MSW occurred within their governed areas, while central government plays supporting roles to solve the problem. In addition, stakeholders – including NGOs, communities and private sector – function to support the set up policies and implementation in various ways. Private sector takes lead role in forming committee and community network with other similar minded organizations to convey the 3Rs (Reduce, Reuse, Recycle) activities.

Pollution Control Department (PCD), a government agency, has set up the national policy and plan which focused on the sustainable consumption of the natural resources and the application of the ‘cradle to cradle’ concept, including control of waste generation at sources, increase on waste segregation and enhancement of waste utilization efficiency prior to the final disposal. The targets of waste minimization in this plan are to have the waste reduction scheme, to have the waste segregation system for reuse and recycling in every community over the country, and to minimize 30% of total waste generated. In addition, the following strategies have been drafted in order to archive sustainable solid waste management in Thailand: (1) encouragement of cooperation among various stakeholders; (2) promotion of science and technology suitable for 3Rs; (3) capacity building on the 3Rs for local communities; and (4) Initiation of the recycling-oriented society (ONEP, 2008 and PCD, 2006). There are three fundamentally different types of recycling programs, or tracks, implemented in Thailand (as also illustrated in Fig. 2).

![Figure 2. Three types of recycling programs implemented in Thailand (source: adapted from Duston, 1993).](https://example.com/figure2)

### 3. METHODOLOGY

For sustainable development, the appropriate evaluation method must be implemented in accordance with the sustainable SWM which is one of many aspects including in Agenda 21 (Hardoy et al., 2001). The evaluation research in this study involves two approaches - a survey design using a questionnaire (to gather the operational background of CRBs and SRBs and their perceptions on factors influencing a success of the programs), and a waste stream analysis using quartering technique (to explore the composition of waste at selected LGAs representing different political areas). The evaluation results in the research provide key for making decision and enabling government and non-government organizations to promote and encourage waste reuse and recycling in the right direction as well as expand national recycling system for sustainable development as mentioned in the Agenda 21.

The relationship of project evaluation and sustainability has also been adopted in this study. As indicated by Dale (2004), the relationship between sustainability and project evaluation is that sustainability is the maintenance or augmentation of positive achievement induced by the evaluated program or project after the scheme has been terminated. Sustainability may relate to all the levels in means-ends frame-work. The specific example of sustainability in this case is continued ability to plan and manage similar development
work, by organizations that have been in charge of the program or project or any other organizations that are intended to undertake the work.

3.1 Indicators selection

The indicators selection was based on literature review and consultations with key stakeholders as well as applying a set of criteria for performance indicators – validity, reliability, sensitivity, simplicity, utility and affordability. The concept of sustainable SWM would function properly as long as accurate and sensible tools are applied for measuring. The overall indicators used in this research are related to the number of people involved in the recycling program; level of participation, public awareness; degree of members’ perception on recyclables segregation, willingness to participate and quantity of recyclables waste. Those indicators show the results and efficiency of the on-going activities which then also fetch out the gaps of activities. It reflects the important point that still needed to be continuously improved for example - the quantity of recyclable waste imply level and quality of recyclable program, their perception shows their needs and expectation on such program. This would be used as a tool for decision makers to solve and run the program in a sustainable manner.

3.2 Questionnaire Survey

A total of 60 questionnaires were distributed to selected 60 CRBs and 60 SRBs for data collection. A returned rate was 50 percent approximately. Statistical analysis methods using SPSS program including descriptive analysis and multiple regression are used to analyze the data. Regression analysis has been applied to investigate the significant factors influencing the recycling performance both of CRBs and SRBs. These were analyzed among related variables through compositing t-value generated based on the mean value of recycling performance as independent variable.

There are some limitations in the number of targeted programs that are supported by TIPMSE because not many SRB and CRB have been initiated and carried out in Thailand. Therefore, the performance and results of SRB and CRB were evaluated under these constraints. However, the returned questionnaires of around 30 for each should be able to be used for statistical analysis. The key topics of questions are recycling efficiency, participation rate, the report of revenue from selling recyclable materials, expenditure, encouragement methods, problem encountered, the number of cooperative involvement and members’ perception on factor influencing the success of the program.

3.3 Waste Stream Sampling and Analysis Methods

For comparison purposes of quantity and waste composition to the previous study, the followings sampling periods and methods have been applied to this study. The sampling technique used is to collect waste from selected waste drop-off centers. Manual sorting was done by the quartering technique. The collection procedure was seasonally conducted twice that cover two seasons - wet and dry, with the frequency of three days per season. The waste stream analysis was carried out at a total of 10 areas covering three different levels of LGAs scattered throughout Thailand.

4. RESULTS AND DISCUSSIONS

4.1 Recycling Performance of CRBs and SRBs

The recycling performance of CRBs were evaluated through participation rate (number of households involved), diversion rate (amount of recyclables to total amount generated), and
revenue from selling recyclables. While the performance indicators of SRBs were assessed through diversion rate, participation rate (number of students and teachers involved), and benefit-cost (B/C) ratio.

The results, as detailed in Table 1, show that the average diversion rate of CRBs and SRBs are 1.09% and 4.78%. The average participation rates of CRBs and SRBs are 76.54% and 86.80%. Revenue of recyclable waste of CRBs ranges from 300 to 64,026 THB/month, with the average of 6,446 THB/month. The average of benefit-cost ratio of SRBs is 10.15.

**Table 1. The results of performance indicators of CRBs and SRBs.**

<table>
<thead>
<tr>
<th>Performance Indicators</th>
<th>Range</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CRBs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diversion rate (%)</td>
<td>13.02</td>
<td>0.00</td>
<td>13.02</td>
<td>1.09</td>
<td>3.05</td>
</tr>
<tr>
<td>Participation rate (%)</td>
<td>279.40</td>
<td>8.00</td>
<td>287.40</td>
<td>76.54</td>
<td>76.48</td>
</tr>
<tr>
<td>Revenue rate (THB/month)</td>
<td>63,726</td>
<td>300</td>
<td>64,026</td>
<td>6,446.84</td>
<td>14,198.143</td>
</tr>
<tr>
<td><strong>SRBs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diversion rate (%)</td>
<td>40.15</td>
<td>0.00</td>
<td>40.15</td>
<td>4.78</td>
<td>9.92</td>
</tr>
<tr>
<td>B/C ratio</td>
<td>103.50</td>
<td>0.00</td>
<td>103.50</td>
<td>10.15</td>
<td>26.26</td>
</tr>
<tr>
<td>Participation rate (%)</td>
<td>230.98</td>
<td>0.00</td>
<td>230.98</td>
<td>86.80</td>
<td>41.31</td>
</tr>
</tbody>
</table>

Note: 1 Benchmark: Beijing (10%); Thailand (11%); Manila (13%); Hong Kong (36%)
2 Benchmark: Thailand (<10%); USA (64.5%); Ulaanbaatar, Mongolia (<50% Low, >51-100% High)

The overall recycling performance of SRBs and CRBs with mean values representing the central tendencies can be derived from scaled scoring, by assuming that the lower performance score value for a given indicator was set at 0 and the upper performance level was set at 1. They were categorized into five levels: poor (score: 0.00-0.20), fair (0.20-0.40), satisfactory (0.40-0.60), good (0.60-0.80), and very good (0.80-1.0).

The applications of the results show that CRBs’ and SRBs’ recycling performance was found to be satisfactory in terms of participation rate. However, diversion rate of both SRB and CRB are still in poor performance and needed to be improved. In addition the analysis of CRB shows poor performance of the revenue rate and B/C ratio of SRB is in a good performance (see Figure 3).

**Figure 3. Mean values of recycling performance results.**

### 4.2 Factors Influencing the Performance of CRBs and SRBs

The results, as given in Table 2, show that perception of source separation and facing inconsistent price problem significantly affect the CRB’s performance. For SRB’s performance results, the analysis shows that perception of lack of skilled operators significantly affect the SRB’s performance, because in the SRBs, the operators are on voluntary basis.

**Table 2. Results of significant factors influencing the performance of CRBs and SRBs.**
### Table 1: Regression Results for CRBs and SRBs

<table>
<thead>
<tr>
<th>No.</th>
<th>Independent variables</th>
<th>Coefficients</th>
<th>t-value</th>
<th>Significant level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRBs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Perception of source separation</td>
<td>0.071</td>
<td>2.672</td>
<td>0.013**</td>
</tr>
<tr>
<td>2</td>
<td>Facing inconsistent price problem</td>
<td>0.333</td>
<td>1.937</td>
<td>0.065*</td>
</tr>
<tr>
<td>3</td>
<td>Perception of lack of skilled operators</td>
<td>-0.209</td>
<td>-1.163</td>
<td>0.257</td>
</tr>
<tr>
<td>4</td>
<td>Provision of compensatory goods</td>
<td>-0.224</td>
<td>-1.141</td>
<td>0.266</td>
</tr>
<tr>
<td></td>
<td>Adjusted R²</td>
<td>0.197</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>F-test</td>
<td>7.138</td>
<td></td>
<td>0.013</td>
</tr>
<tr>
<td>SRBs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Perception of lack of skilled operators</td>
<td>-0.107</td>
<td>-3.251</td>
<td>0.003***</td>
</tr>
<tr>
<td>2</td>
<td>Financial support by international agencies</td>
<td>0.072</td>
<td>2.441</td>
<td>0.020**</td>
</tr>
<tr>
<td>3</td>
<td>Private sector cooperation</td>
<td>0.273</td>
<td>1.817</td>
<td>0.079*</td>
</tr>
<tr>
<td>4</td>
<td>Provision of interest or money for members</td>
<td>0.233</td>
<td>1.606</td>
<td>0.118</td>
</tr>
<tr>
<td>5</td>
<td>Investment costs less than THB 10,000 (a)</td>
<td>-0.223</td>
<td>-1.535</td>
<td>0.135</td>
</tr>
<tr>
<td></td>
<td>Adjusted R²</td>
<td>0.283</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>F-test</td>
<td>7.715</td>
<td></td>
<td>0.002</td>
</tr>
</tbody>
</table>

*Significant at the 0.1 level or better (p<0.10)
**Significant at the 0.05 level or better (p<0.05)
***Significant at the 0.01 level or better (p<0.01)
(a) USD 1 is approximately THB 34.

### 4.3 Quantity and Composition of Municipal Waste Stream

The results of waste stream analysis, as shown in Figure 4, indicate that the major components of municipal waste at the selected drop-off centres mainly comprise of plastic (31.56%) food (24.01%), and paper (10.57%), which are in range of 17.35-42.22%, 1.33-45.56% and 2.32-19.42, respectively. In comparison to the previous study, it was found that waste proportion has changed – plastic increased by 6.8%, but food has vastly decreased by 30.41%, while paper remains at around 10% of total waste. The overall findings from waste stream assessment found that the plastic waste has been increasing. Instead of food, plastic has moved up to the top components of the waste stream. The possible explanation for the decrease of food proportion is 3R activities that are widely promoted and supported in LGAs.

![Figure 4. The percentage of recyclables in the MSW stream.](image)

There have been many attempts to reduce plastic waste as mentioned in the relevant authorities’ strategies and implemented plan of some agencies. Yet, plastic has been increasing in bigger proportion. In-depth details of the study show that plastic bag is the major type that affected higher percentage in the group of plastic, which equal to 63% of total plastic waste. In comparison to previous study, it is apparently seen that the proportion of three main types of plastic has changed – mixed plastic bottle and foam has greatly decreased, while plastic bag vastly increased. The other types mostly remain unchanged. Based on the result, it reflects the effectiveness of 3Rs programs applied to plastic bottle and foam, while the amount of plastic bag – non-recycling materials has increased according to higher waste generation rate from 2004 to 2009.
The result shows unremarkable change in proportion of metal, glass and aluminum, with a slight increase of 1.42% and 2.87% for aluminum and metal, whereas the glass amount remains constant at 2.62%.

4.4 Evaluation results and sustainability

The strength of implementing recycling activities in the CRB and SRB model is that there are large number of households/students participated in the recycling programs. This shows environmental awareness on source separation and waste minimization. However, the evaluation also investigated their perceptions on the factors that could influence the performance of recycling programs. It is found that facing the recyclable waste price volatility is a major concern of CRB model. The fluctuation of recyclable waste price somewhat affects the number of members attending the recycling programs. For SRB, the skill of operator is considered as the most significant aspect influence SRB recycling performance. In achieving goals of recycling, SRB needs the financial support by agencies and cooperation of private sector in providing grant for establishing recyclable storage building at the initial stage of the program. In this regards, further improvement for sustainable recycling program could be managed by disseminating appropriate management and sustainable strategies for CRB, particularly financial management. In addition, LGA should provide support in encouragement to maintain the current members and expand to more recycling group members in the future. For CRB and SRB, a practical way aimed at sustainability is to hold the community/school workshop which focuses on capacity building on waste separation which would help increase the revenue rate and diversion rate.

The waste stream analysis reflects the quantity of waste at different drop-off centers where all valuable materials have been screened out through various types of activities – exchange between industrial sector, purchase shop, recyclable waste bank etc. The evaluation results indicate that there is a huge decrease in the amount of food proportion, while the amount of recyclable materials tends to be maximized. However, the current recycling rate is still at a poor performance in comparison to the recycling potential. For the sustainable SWM, reducing waste volume prior to the final disposal and shifting waste into raw materials by urging source segregation and source reduction are very important at this stage. Furthermore, recycling performance can be improved by supporting the informal recycling group and private sector with some help from policy makers for the sustainable development.

5. CONCLUSIONS AND RECOMMENDATIONS

The evaluation results of this research links to ‘sustainability’ in three ways: (1) as the results reflect strength and weakness of the current specific recycling system, decision makers could use it as a baseline tool for further improvement such as maintaining its strengths and filling the gaps of the weaknesses; (2) clear and concise conclusions of existing evaluation results will provide a key for decision makers to set up their plans for sustainable SWM; and (3) it shows fact and status of stakeholders involving in the recycling activities as well as their needs which of course significantly affect their participation in the programs. For sustainable SWM, stakeholders must be taken into account and the program benefits should be fairly shared among them.

To further refine the results and outcome of the study, it is recommended that more research should focus on: (1) evaluating recycling activities at program level using case study approach, (2) drawing relationships among different levels of waste recycling programs, and (3) finding effective indicators for evaluating results of municipal waste management to reflect recycling-related programs.
ACKNOWLEDGMENTS

The authors would like to acknowledge Thailand Institute of Packaging and Recycling Management for Sustainable Environment (TIPMSE) and Royal Thai Government (RTG) for their financial support to the research leading to this publication. The assistance for preparation of research results of Ms. Siwaporn Tangwanichagapong and other staff at Asian Institute of Technology as well as TIPMSE in this research project is also acknowledged.

REFERENCES


Chulalongkorn University, Packaging industry development for supporting solid waste management and recyclables project, Chulalongkorn University Press, Bangkok, Thailand, 2004.


Pollution Control Department, Thailand 3Rs Conference Documentation, Waste and Hazardous Substance Management Bureau, Ministry of National Resources and Environment, Bangkok, Thailand, 2006.


Pollution Control Department, PCD Annual Report 2009, Ministry of National Resources and Environment, Bangkok, 2009.


Sustainability, information needs and organisational change in UK water and sewerage companies

Tanner, A. a,b, McIntosh B.S. a, Seth, A. b & Widdowson, D. b

a Centre for Water Science, Cranfield University, UK (a.tanner@cranfield.ac.uk, b.mcintosh@cranfield.ac.uk), b Yorkshire Water, Bradford, UK

Abstract:
Computer-based decision and information support tools (DISTs) have potentially important roles to play in the embedding of sustainability appraisal processes into the planning and operation of water utilities. This paper reports on preliminary outcomes from research employing a particular model of sustainability, the Five Capitals model, to identify and facilitate the exploitation of opportunities for improved incorporation of sustainability appraisal into business process and practice within a major UK water and sewerage company (WaSC). In particular, the aims of this paper are to characterise and critically assess WaSC decision and information support needs by interpreting the findings of having applied the Five Capitals model. Five Capitals sustainability principles were applied as a questioning framework in a series of focus groups within the asset delivery business unit of the WaSC. The approach enabled the researcher to create a shared comprehension of sustainability, whilst mapping the perspectives of the business unit as to the form and efficacy of current sustainability appraisal activities. From the results of the focus group the researcher was able to identify key information support needs and to develop a set of sustainability key performance indicators with WaSC staff to service these needs. The results of the focus groups demonstrated that there was no need for computerised decision support, and that the primary role for information support was twofold – (i) to capture data to provide a basis, in the medium-long term, for improved organisational learning about the sustainability performance of different treatment and distribution assets, and; (ii) to capture data to provide a basis, over the short-medium term, for influencing the decisions made by companies contracted to design and build new treatment and distribution assets for the WaSC. These needs contrast against the standard view of the role of decision support as automating certain aspects of human decision-making.

Keywords: sustainability appraisal; water utilities; decision and information support tools; organisational change

1. INTRODUCTION

The ambitions of sustainability and sustainable development have been argued as being central to the management and delivery of water and sewerage services (Foxon et al. 2002). In England Wales these services are delivered by a set of fully privatised and regulated water companies – water and sewerage companies (WaSCs) and water only companies (WoCs) (see section 2 for more information). The economic regulator for water companies in England Wales, OFWAT (www.ofwat.gov.uk), has been slow to define its position on sustainability, leaving UK WaSCs with the challenge of identifying an appropriate sustainability framework and values, and to develop and embed the corresponding business processes to improve their sustainability performance.
A wide range of computer based decision and information support tools (DISTs) (McIntosh et al. 2008, Diez and McIntosh, 2010) are available to assist in the incorporation of sustainability appraisal as criteria or considerations in organisational strategic planning or operational business processes e.g. life cycle assessment, GIS, integrated assessment models, multi-criteria optimization / decision analysis tools. DISTs may play various roles from the problematic vision of DISTs taking over / automating certain aspects of human decision-making processes, through to playing more limited life-span aides to learning during change processes (McCown 2002). Quite how WaSCs in England and Wales will engage with, and potentially adopt and use DISTs in the context of embedding sustainability is not yet clear, nor is it clear how English and Welsh WaSCs will engage with and embed sustainability systemically.

The main aim of this paper is to identify a set of learning points for the DIST development community generally about the role of DISTs and about the process of developing them. This will be achieved by interpreting the findings of research to identify, pilot and evaluate opportunities for embedding sustainability appraisal into the asset delivery function of a major UK WaSC (see section 2 for a description of UK WaSCs).

Many factors have been identified which inhibit or promote the adoption of new behaviors or technologies, and in doing so enable business change. Models of organisational change identify the requirement for a shared vision or understanding of needs to occur before change processes can begin (Jick, Kanter et al. 1992; Kotter 1996; Lueke 2003), and in turn articulate such visions as being the products of pressure for change (Cooper and Zmud 1990; Jick, Kanter et al. 1992; Van De Ven and Poole 1995; Kotter 1996; Weick and Quinn 1999; Rogers 2003). The Technology Acceptance Model (TAM) shows that perceived complexity and compatibility are the most significant drivers for individual adoption of new technologies (Venkatesh and Bala 2008). With regards the adoption of information systems, Burton and Swanson (1994) have shown that such technologies can come with significant knock-on effects to surrounding business processes, whilst van de Ven (1986) has argued that new innovations may create the additional work through the need to establish new inter-departmental coalitions and resource (re-) allocations.

The combined implications of these findings are that to change WaSC business processes to incorporate sustainability appraisal with appropriate decision or information support tools, (i) a clear, shared vision is needed; (ii) that existing processes and DISTs should be utilized wherever possible, and; (iii) that the scale and scope of process change should be commensurate with the level of buy-in to the vision. The research reported here followed these principles by seeking first to generate a shared vision and to utilise existing DISTs rather than to push complicated or completely novel tools.

2. RESEARCH CONTEXT

This project was undertaken by invitation from a major English and Welsh WaSC with a desire to better embed sustainability appraisal into its asset delivery processes (see below for a description). The nature of the intervention and tools/processes to be identified were not specified in the project brief requiring the researcher to identify the opportunities for incorporating sustainability appraisal. To identify these opportunities the research needed to (i) examine to what extent sustainability was already incorporated in current activities, avoid duplication of process, familiarize the researcher with the existing process, and identify potential gaps in current sustainability appraisal; (ii) appraise the business buy-in for the sustainability project and evaluate the impacts of sustainability appraisal related changes on the business, and; (iii) marry business need and project opportunities to maximize the potential for adoption.

The principle functions of a WaSC in England and Wales are the treatment and distribution of potable water and the safe removal and disposal of sewage (domestic, commercial and municipal). UK WaSC’s are regulated through the activities of three bodies - Ofwat, who
ensure WaSCs do not abuse their natural monopoly positions over service users; the Drinking Water Inspectorate (DWI) who perform potable water compliance testing to ensure distributed water achieves the necessary quality standards, and; the Environment Agency (EA) who regulate both raw water abstractions and treated sewage discharges.

Each WaSC has a large and complicated system of network and treatment assets to replace, improve and maintain under its control (See figure 1). The Asset Delivery Unit (ADU) in the WaSC considered here is responsible for the delivery of solutions to business risks, typically through the replacement of built assets. ADU is divided into five areas referred to as investment Streams. Four of these Streams relate directly to asset infrastructure types - ‘Reservoirs’ are reservoir related assets; ‘Networks’ are typically pipe infrastructure for the transmission of potable water and sewage; ‘Medium Treatment’ is infrastructure used to change the chemical or physical properties of sewage, sludge or water for drinking, and; ‘Other Installations’ are assets such as pumping stations and those related to telemetry. ‘Large Schemes’ differs from the other streams as it is determined by project cost, and can refer to any project that resolves a risk at a cost of greater than ten million pounds. Each ADU Stream has a Stream Manager (SM) who is responsible for the delivery of a stream of solutions to risks. To achieve this, each stream has a team of Project Managers (PMs) who are responsible for managing the resolution of a number of business risks. PMs liaise with partner organizations selected for their skills in design and construction of the solution infrastructure, and commission new assets for delivery from these partner organisations.

3. METHOD OUTLINE

Four steps were undertaken to identify, pilot and evaluate sustainability appraisal changes within ADU as described below. This paper will focus mainly on steps 2 - 4.

**Step 1.** Selection of a relevant sustainability framework to apply to the activities of a WaSC. This step ensured that a coherent, relevant and comprehensive understanding of sustainability was selected to form the conceptual basis for the research.

**Step 2.** Mapping of business processes that influence the management of aspects of sustainability (from the selected sustainability framework). This step enabled the identification of aspects of sustainability which are perceived as less well managed by the business, and where changes to ADU present an opportunity to influence the performance of the business against the aspect identified.

**Step 3.** Identifying opportunities for improvements in sustainability performance as a consequence of changing ADU processes and practices. This step allowed the researcher to marry the adoption of managing an aspect of sustainability with a desired improvement identified by the WaSC.

**Step 4.** Converting research findings into a business change through which the WaSC is better able to appraise and influence sustainability performance. This step was important as the WaSC is early in the adoption process, and as such has identified the need to generate a convincing business case for the incorporation of sustainability appraisal.

4. RESULTS

4.1 **Step 1. Selection of a relevant sustainability framework**

The full results for this step will not be presented in detail here. However, briefly, the framework that was selected was the ‘Five Capitals’ model developed by Forum for the Future (2009). The framework was selected from a literature review of potential frameworks primarily for its breadth of coverage, and for the relevance of the sustainability
dimensions covered to the business of a WaSC. The model was developed for use by organizations wishing to embody sustainability in their practices and processes.

The Five Capitals model defines five capital stocks and describes a number of principles to guide the management of these stocks – see Table 1. The model proposes that a sustainable organisation should seek to maintain and where possible enhance these stocks rather than deplete or degrade them.

Table 1 The Five Capitals model

<table>
<thead>
<tr>
<th>Capital Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural Capital</td>
<td>the natural resources (energy and matter) and processes (direct and indirect) needed by organisations to produce their products and deliver their services.</td>
</tr>
<tr>
<td>Human Capital</td>
<td>incorporates the health, knowledge, skills, intellectual outputs, motivation and capacity for relationships of the individual.</td>
</tr>
<tr>
<td>Social Capital</td>
<td>is any value added to the activities and economic outputs of an organisation by human relationships, partnerships and co-operation.</td>
</tr>
<tr>
<td>Manufactured Capital</td>
<td>is material goods and infrastructure owned, leased or controlled by an organisation that contributes to production or service provision.</td>
</tr>
<tr>
<td>Financial Capital</td>
<td>are assets of an organisation that exist in a form of currency that can be owned or traded, including (but not limited to) shares, bonds and banknotes.</td>
</tr>
</tbody>
</table>

To ensure the identified framework could be easily interpreted when applied discussed with the business, the Project Steering Group agreed to adapt and simplify the language of the ‘principles’ (or rules) where necessary to guarantee the framework would be more easily comprehended across the business. A full description of the adapted ‘principles’ for successful management of each stock are given within Table 2 overleaf.

4.2 Step 2. Mapping of business processes that influence the management of aspects of sustainability

The following activities enabled the identification of sustainability principles that are perceived as less well managed by the business, and where the ADU has opportunity to influence the performance of the business against the principle.

Using the adapted five capitals from ‘Step One’ two focus groups were held with ADU Project Managers. The participants represented geographically separately managed areas of the WaSC covering both clean and waste water service provision. Participants were instructed that the meeting objective was ‘to better understand where and how sustainability is addressed by the business, specifically within investment delivery’.

Participants were presented with the Five Capitals sustainability principles and asked to read through each principle carefully. The author then asked the participants to respond to the following questions for each principle with regards the work of the ADU:

- How does Yorkshire Water as a business manage this sustainability principle?
- Do you believe this principle is effectively managed by the business?
- Can ADU influence the performance of this principle? If so, how?

Participants were asked to answer the questions with regards each of the life-cycle stages in asset delivery (investigation, design, construction, operation and decommissioning), in relation to the business units which have an impact on investment delivery (Human Resources, Program Planning, Supply Chain and Procurement) and in relation to the tools employed by the WaSC during asset delivery (company policy, asset standards, engineering specifications, key performance indicators, and cost models). Responses were categorized as principle perceived as ‘undermanaged’, ‘requiring management’, ‘conditional’ (undermanaged in some situations), ‘effectively managed’, ‘did not know’, or ‘not relevant’ (to the work of the water company). Results were reflected back to the respondents for further comments, to encourage participants to challenge or verify results. The researcher
then used the information gathered to identify those principles perceived as least well managed in investment delivery and under the direct control of ADU (see Table 2).

<table>
<thead>
<tr>
<th>Table 1 Adapted Five Capitals framework and principles</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Principle perceived as undemanged by YW</strong></td>
</tr>
<tr>
<td><strong>%</strong></td>
</tr>
<tr>
<td><strong>Natural Capital</strong></td>
</tr>
<tr>
<td>NC. 1 Protect/improve habitat, biodiversity and ecosystem function.</td>
</tr>
<tr>
<td>NC. 2 Reduce emissions of substances to a concentration that can easily be assimilated by natural systems: a. chemical concentrations and nutrient loads; b. GHG, Ozone depleting substance; c. etc.</td>
</tr>
<tr>
<td>NC. 3 Reduce dependency on materials that are naturally scarce.</td>
</tr>
<tr>
<td>NC. 4 Reduce use of virgin materials and resources</td>
</tr>
<tr>
<td>NC. 5 Reduce dependency on and accumulation of man made substances that may prove harmful to ecosystem or human health substitute all with substances that can be easily assimilated broken down by natural systems.</td>
</tr>
<tr>
<td>NC. 6 Use renewable resources only from well-managed and restorative eco-systems.</td>
</tr>
<tr>
<td>NC. 7 Reduction/elimination of waste</td>
</tr>
<tr>
<td>NC. 8 Increase/full recycling of resources</td>
</tr>
<tr>
<td>NC. 9 Reduce/eliminate dependency in the use of fossil fuels (thereby increasing use of renewable energy resources).</td>
</tr>
<tr>
<td>NC. 10 Reduce energy demand</td>
</tr>
<tr>
<td><strong>Human Capital</strong></td>
</tr>
<tr>
<td>HC. 1 Ensure adequate Health and Safety standards are met</td>
</tr>
<tr>
<td>HC. 2 Respect human rights throughout their operations and geographical regions</td>
</tr>
<tr>
<td>HC. 3 Respect human values and their different cultural contexts</td>
</tr>
<tr>
<td>HC. 4 Give employees (where possible) access to training and education</td>
</tr>
<tr>
<td>HC. 5 Educate and promote for higher standards of health and support mental wellbeing.</td>
</tr>
<tr>
<td>HC. 6 Provide a reasonable living wage and fair remuneration for employees and business partners.</td>
</tr>
<tr>
<td>HC. 7 Allow for and enhance recreation time and support individuals’ active involvement in society.</td>
</tr>
<tr>
<td>HC. 8 Ensure supply chain partners apply the same principles to fulfilling employee needs.</td>
</tr>
<tr>
<td>HC. 9 Create opportunities for varied and satisfying work.</td>
</tr>
<tr>
<td><strong>Social Capital</strong></td>
</tr>
<tr>
<td>SC. 1 Source materials ethically and treat suppliers, customers and citizens fairly</td>
</tr>
<tr>
<td>SC. 2 Reduce emissions of persistent compounds that are harmful to ecosystem or human health.</td>
</tr>
<tr>
<td>SC. 3 Respect and comply with local, national and international law.</td>
</tr>
<tr>
<td>SC. 4 Provide a supportive family friendly labour policy.</td>
</tr>
<tr>
<td>SC. 5 Prompt and full payment of taxes and support of social infrastructure.</td>
</tr>
<tr>
<td>SC. 6 Minimise the negative social impacts of products or services or maximisation of the positive</td>
</tr>
<tr>
<td>SC. 7 Support the development of the community in which the organisation operates, including economic opportunities.</td>
</tr>
<tr>
<td>SC. 8 Assess the wider economic impacts of the organisation activities, products and services on society e.g. in creating wealth in the communities in which the organisation operates</td>
</tr>
<tr>
<td>SC. 9 Encourage and engage in transparent consultation and communication with relevant internal and external stakeholders,</td>
</tr>
<tr>
<td>SC. 10 Fulfil commitments made with suppliers, customers/citizens and regulators.</td>
</tr>
<tr>
<td>SC. 11 Effective Communication throughout the organisation , reflecting shared Values and objectives</td>
</tr>
<tr>
<td><strong>Infrastructure Capital</strong></td>
</tr>
<tr>
<td>IC. 1 Ensure that systems, processes and infrastructure performance is maintained under a robust set of future operating scenarios.</td>
</tr>
<tr>
<td>IC. 2 Seek to maximise the flexibility and adaptability of infrastructure to respond to diverse set of future operating scenarios.</td>
</tr>
<tr>
<td>IC. 3 Develop infrastructure that facilitates ease of maintenance: a. Design for disassembly ; b. Modular designs (to minimize potential negative opex spend)</td>
</tr>
<tr>
<td>IC. 4 Have sought to reduce or eliminate waste and emissions in production systems.</td>
</tr>
<tr>
<td>IC. 5 Where appropriate replace products for service contracts.</td>
</tr>
<tr>
<td>IC. 6 Optimisation of infrastructure/technologies and processes in a way that uses resources most efficiently.</td>
</tr>
<tr>
<td>IC. 7 Optimise the recycling of resources.</td>
</tr>
<tr>
<td>IC. 8 Identifying and utilizing synergistic production systems where one organisation’s waste streams are another’s resources.</td>
</tr>
<tr>
<td>IC. 9 Seek improvements and innovation in the design of product systems (eco-efficiency and eco-innovation).</td>
</tr>
<tr>
<td>IC. 10 Apply sustainable construction techniques when looking at new infrastructure.</td>
</tr>
<tr>
<td><strong>Financial Capital</strong></td>
</tr>
<tr>
<td>FC. 1 Ensure prudent financial management</td>
</tr>
<tr>
<td>FC. 2 Efficient use of financial resources (reducing and minimising costs)</td>
</tr>
<tr>
<td>FC. 3 Management of financial risk (over both short and long term)</td>
</tr>
<tr>
<td>FC. 4 Internalise environmental and social costs and assign an economic value to them.</td>
</tr>
<tr>
<td>FC. 5 Effective total costs under a robust set of future scenarios e.g. a. Unit running costs; b. Unit capital costs; c. Remediation costs of infrastructure; d. Internal manpower costs; e. External services costs ratio; f. Imported (raw and treated) water costs ratio. g. Energy costs ratio h. etc.</td>
</tr>
<tr>
<td>FC. 6 Effective management of financial risk exposure.</td>
</tr>
<tr>
<td>FC. 7 Timely fulfilment of contracts</td>
</tr>
</tbody>
</table>

A. Tanner, B.S. McIntosh, A. Seth and D. Widdowson  
Sustainability, information needs and organisational change in UK water and sewerage companies
4.3 Step 3: Identifying opportunities for improvements in sustainability performance as a consequence of changing ADU processes and practices.

Taking those principles perceived as being less well managed (those with orange highlighted % undermanaged figures in Table 2), a series of interviews was held with the Stream Delivery Managers (SDM). The objective was to identify priority sustainability principles for the ADU business unit, and to reveal the perceived business benefits from improved incorporation of the identified principles into process and practice.

SDM interviews were held on a one to one basis and each interview was allocated 1 hour. Each SDM interviewee was given a description of the process undertaken so far and presented with a list of asset investment classes (investment streams) that corresponded to the asset investment distinctions used by the business. To ensure that the interview captured relevant and informed information on specific investment streams, interviewees identified the stream in which they had most experience and were instructed to proceed with the interview from the perspective of activities carried out within this stream.

The Five Capitals principles were presented to the managers, with principles that had been identified as less well managed by the focus group and under the influence of ADU highlighted in red. The managers were then asked to review all the principles, placing a mark alongside each principle that they believed their stream had a significant impact upon. The interviewees were then asked to identify from the marked principles those which they believe their stream should prioritize (See Figure 1 – stream ‘selected principles’). The researcher then requested the participants to review their selection using two adoption criteria – (1) those principles which would be easiest to make strong performance improvements against, and; (2) those principles which are most likely to result in business benefits and therefore likely to be adopted. Finally, interviewees were asked to select one principle, to improve the sustainability impacts of the asset stream and to describe investment stream improvements they aspired to by adopting the sustainability principle for the stream (See Figure 1 - ‘stream priority indicators’).

4.4 Step 4: Converting research findings into a business change through which the WaSC will be better able to appraise and influence sustainability performance

To enable the research to propose a means of improving the incorporation of sustainability appraisal into investment delivery that was both sensitive to the WaSC’s requirement for change, and which exploited synergies with the internally recognised opportunities for change, the following activities were undertaken. The Project Steering Group requested that the research carried out under steps 1-3 be converted into a set a Key Performance Indicators (KPIs). Consequently potential indicators from the literature review carried out under step 1 were compiled in a spreadsheet, loosely sorted by relevance against capital and principle. The spreadsheet was used to help identify means by which the WaSC could turn stream sustainability objectives into measurable indicators and begin the process of enabling the ADU to manage sustainability performance.

The spreadsheet information was then used in a series of meetings with the WaSC Environmental Strategy Team, the manager responsible for ‘Reporting’ and an employee charged with developing KPMs (Key Performance Measures – contractually binding measures to assess how well delivery partners are performing, as opposed to KPIs which have no contractual status) for ADU for the next 5 years. The meetings provided the researcher with a number of necessary factors to incorporate into the set of proposed sustainability KPIs. In recognition of tightening UK regulatory requirements for reporting and improving WaSC carbon emissions, the Environment Strategy Team felt strongly that the addition of carbon should be a priority data point to evaluate asset investment performance. In order to increase the potential for adoption the Reporting Manager stated it was necessary that proposed indicators should rely solely on data already captured by the business or contractually demanded during asset investment.
The researcher used this information to develop the proposed sustainability KPIs for ADU within the WaSC – see Figure 1.

Figure 1 Research findings and sustainability Key Performance Indicators

Stream investment improvements
Selected Principles
Marked Principles
Networks NC: 4, 8 IC: 1, 2, 8
Reservoirs NC: 1, 3, 4, 8 IC: 1, 2, 3, 6, 7, 8, 9, 10 HC: 8, 9, FC: 4
Other Installations NC: 1, 2, 3, 4, 6, 7, 8, 9, 10 IC: 1, 2, 6, 9, HC: 8, FC: 4
Medium Treatment NC: 2, 3, 9 IC: 1, 6, 9 HC: 8 FC: 4 SC: 2
Large Schemes NC: 7, 8, 9, 10 IC: 2, 3, 6, 7, 8, 9, 10 HC: 1, 3, 4, 8, 9, FC: 1, 2, 3, 5, 6, 7 SC: 1, 2, 3, 6, 7, 9, 10, 11

Sustainability Indicators
indicator data requirement
Embedded Carbon $\% = \frac{N \text{ solution (KgC}_0^2 \text{)} / \text{New Asset (KgC}_0^2 \text{)}}{100}$
Waste Reduction Waste to Landfill (kg) / Construction materials (kg)
Material Intensity $\% = \frac{\text{Construction Materials (kg)} / \text{Recycled Materials (Kg)}}{100}$
Energy on Site On site fuel + energy use (Kwh) / man hours on site (h)
Energy on Transport Transport fuel (l) / man hours on site (h)
Energy Operation Energy use (Kwh) / water flow (m$^3$)
Renewable Energy $\% = \frac{\text{Notional solution (Kwh)} / \text{New Asset (Kwh)}}{100}$
Zero Chemicals $\% = \frac{\text{Notional solution (Chemicals (kg)) / New Treatment H}_2 \text{O (m$^3$)}}{100}$
Safe Chemicals $\% = \frac{\text{Notional solution (Chemicals (kg)) / New Treatment H}_2 \text{O (m$^3$)}}{100}$
Carbon Operation $\% = \frac{\text{N solution (KgC}_0^2 \text{)} / \text{New Asset (KgC}_0^2 \text{)}}{100}$

*eco efficiency by sharing best practice.

Priority indicators dependent on nature of scheme.
5. DISCUSSION

The research reported here is interesting in the context of the development of computer-based DISTs for a number of reasons. The WaSC involved commissioned the research to potentially change and in doing so, improve their performance in a substantive area – appraising, selecting and delivering more sustainable asset investment options. The WaSC involved did not commission research to develop a DIST specifically. Rather, the options for changing the way in which asset investment and delivery activities are conducted and supported were left open and to be identified to best meet the nature of the demand for change. In this respect the research is an example of highly user-centred approach to developing DISTs and approaches in relation to, and in conjunction with, their accompanying business process and practice changes. Finally, having used an explicitly process-based research design the research, once completed, will provide a detailed, chronologically situated case-study about how different factors influence the adoption of sustainability appraisal into water utility asset investment and delivery.

So what can be learned for DIST development and use? Making improvements to an organisation in a contested area such as sustainability requires significant engagement with personnel to ensure (i) that a shared understanding is used to develop organisational changes, and decision or information support systems (processes and tools) from, and; (ii) that the way in which sustainability appraisal changes are made, and support systems developed, correspond to individually and collectively perceived performance deficiencies and opportunities for improvement. Such engagement cannot be done remotely for it requires that the researcher be embedded in the organisation concerned. Here, the researcher will occupy multiple roles including notably being a change agent, being a source of external and expert knowledge on sustainability, and being a sustainability champion.

Improving the way in which sustainability appraisal is undertaken by water utilities in relation to asset investment and delivery requires changing individual and collective (organisational) behaviours. Consequently the process is neither simple, nor necessarily quick. A lesson from the research reported here is that small changes may be required (like the development of KPIs to influence asset delivery partners) to demonstrate an overall positive impact before more widespread changes are considered. The role of making relatively small changes in the first place is an element of organisational learning – the WaSC concerned here has no prior experience of using a set of sustainability KPIs derived from a systemic sustainability framework to influence delivery partners. There is a need for learning about whether, which and how KPIs are effective in influencing partner behaviours through the course of piloting.

Reflecting on two opposing roles for decision support highlighted by McCown (2002) in the context of DSS – that they may be used to automate certain aspects of human decision-making processes, or may be used over the shorter-term to help learn about how to adapt to particular change drivers – the implications of the research reported here are that the information support needs identified are part of a medium term process of learning for the WaSC involved. Contrary to the idea that DSS are suitable for informing decisions in poorly structured problem areas, the incorporation of improved sustainability appraisal is not a process of making decisions about something to be managed, it is a process involving making decisions about how to manage. This is fundamentally a process of learning and quite sensibly, involves the organisation investing in and learning from a limited pilot rather than investing in the development of larger, more complicated means of decision or information support.

5. ACKNOWLEDGEMENTS
The research reported here was jointly funded by the Engineering and Physical Sciences Research Council and Yorkshire Water under award CASE/CNA/07/104.

6. REFERENCES


Diez, E. & McIntosh, B.S., Organisational drivers for, constraints on, and impacts of decision and information support tool use in desertification policy and management, Environmental Modelling and Software, accepted, 2010.


A Low-cost Automatic Water Sampler Equipment

Lúcio Flávio Ferreira Moreira
Rio Grande do Norte Federal University – UFRN (lucio@ct.ufrn.br)

Jean Oliveira de Paiva
Rio Grande do Norte Federal University – UFRN (jpaiwa@gmail.com)

Abstract: This work aims at presenting an automatic low cost water sampler prototype. Data provided by this equipment can be used in water resources management and control. Prototype has an electronic system which controls a peristaltic pump functioning, five solenoid valves and an ultrasonic sensor that provides water level measurement data. An interface with the user provided by a board and LCD display allows sampling parameters input and a report output when sampling is over. Hydraulic module uses a peristaltic 12V pump connected to five solenoid valves, which distribute sample and clean PVC tubes once each sampling is produced. Storage module contains four plastic bottles, 1.95 ℓ. Prototype structure is made of acrylic, 0.72 m height and 0.38 m diameter. Experimental laboratory runs have showed a successful performance.

Keywords: water sampler; equipment; water quality

1. INTRODUCTION

Water management and control measures implementation usually involves systematic water resources monitoring (Coimbra, 1991). Indeed, monitoring systems are conceived to provide fundamental hydrologic and climatologic data to be used for control and management policies definition. Monitoring involves the acquisition of water resources system data by using a variety of measurement devices (Dick, 1996). In general, water resources monitoring involves quantitative and qualitative dimensions. In a basin scale, quantitative monitoring involves a systematic fieldwork which may include stream cross section establishment and flow velocity measurement. In this case, the aim is to estimate water discharge in order to generate a database that can provide a discharge-water level relationship (Porto, 2008). Water quality monitoring usually involves a representative water sample collection by using a well accepted methodology. A punctual-based water sample usually involves the collection of a small amount of water that can be considered to represent instantaneously water body characteristics (CETESB, 1987). In general, water samplers can be classified as manual or automatic devices. During the last decade, automatic devices have had a great and revolutionary use, due to the advance of computation and automation sciences. The main advantage of these kinds of equipments is the possibility of program a previous water sample collection in accordance with the user needs. Furthermore, these equipments provide an interface with the user by using a keyboard or PC, where different sampling options, such as time, frequency sample size can be defined. Once sampling procedure is over, a data measurement report (such as water body level) can be provided.

It can be observed the importance of these kinds of equipments in a context of increasing water demands and pollution in developing countries, such as Brazil. Besides, one aspect of great importance is associated with the high cost.
This paper aims to present a prototype that automatically collects water sample from a water body. The equipment has an automated system which controls a peristaltic pump functioning along with an ultrasonic sensor that measures water body surface level. Water samples are stored in four 2-litres polyethylene bottles. A report is generated when sampling procedure is over, which contains water level and sample number as a function of time.

2. MATERIALS AND METHODS

Prototype was built by using a 6 mm thickness cylindrical acrylic structure 0.72 m (high) and 0.38 m (diameter). Equipment is composed by four internal compartments, which can be seen in Figure 1. The microprocessor and the other electronic components were installed at the automated control system, which was fixed at the upper compartment; into the compartment 2, a peristaltic pump, a 12-V battery, electric cables connection and pump tube. Into the compartment 3 it was installed a PVC tube distribution system, which connects the pump to five solenoid valves; their functioning provides both sample distribution and distribution system cleaning. The bottom compartment was designed to fix four plastic bottles used to store the samples. When sampling procedure is over, the access to the bottles can be made by handling the upper part of the structure (composed by the compartments 1, 2 and 3).

The basic equipment was conceived to provide an automatic water sample collection, previously defined by the user. For this purpose, a control system composed by two main components was developed: a) control system and data acquisition; b) sample collection and distribution system. Control system provides an interface with the user by using a keyboard and a LCD. Once sampling procedure is being made, an ultrasonic sensor provides the sampler water level data as a function of time.

[Image of automatic water sampler]

Figure 1. Front view of the automatic water sampler.

Electric and hydraulic Connectors

The prototype offers serial communication with a PC, input for an ultrasonic sensor electric cable, input for the battery recharge and a power switch. The intake tube is installed to the equipment and uses a PVC strainer attached to the other end. Furthermore, an outlet
tube is attached to the solenoid valve number 5, which is used to clean the distribution system immediately before the sampling procedure. At the bottom compartment, sample bottles are handled by acrylic pieces in order to keep them fixed.

**Sampling Operation**

The equipment was originally conceived in order to offer the option of programming the sampling operation, which must be established by the user. The equipment offers the following programming options: a) manual sampling operation by the user; b) automatic sampling operation, associated to the surface water level variation. Once sensor is installed, water level is measured and made reference. Water level measurement is made at a time interval established by the user. In an urban basin, the occurrence of storm flow associated by a precipitation event will make the sampling program start. At this moment, before the first sample is collected, pumped water stored into the sampler tube system is pumped out and the system is cleaned. The subsequent sampling times must be programmed by the user. c) automatic sampling operation, associated to a time interval program. In this case, the user must establish the date and time of sampling procedure. For every programming option, the equipment turns off immediately after the operation program.

Commonly adopted for this kind of equipment, peristaltic pump operation is provided by a 12-V electric engine. Suction is made possible when a silicon tube is pressed by two dots in a rotational movement.

The equipment was designed to use five solenoid valves attached to a PVC tube system. Four valves provide sample distribution through the system to the bottles. The fifth valve is used to clean the tube before each sampling collection. Water level records supplied by the ultrasonic sensor can be transferred to an automatic control system and be stored in a datalogger as well as sampling time and date.

**Hardware**

The hardware of the equipment is composed by the following components, installed in a plate circuit: (a) microcontroller; (b) storage mechanism; (c) clock; (d) ril; (e) command interface; (f) serial communication interface; (g) sensor-controller communication interface. A basic scheme of the plate circuit can be seen on Figure 3.

**Ultrasonic sensor**

It is composed by a sensor and a circuit installed in a PVC tube (40 mm diameter, 0.2 m length), covered with PVC caps on both ends. The sensor was fixed at one end, while the other end the cable was attached. The sensor must be installed over the water surface at a distance between 1 and 2 m, with its face pointed to the water surface. Sensor is connected to the data acquisition system by using an 8-m length parallel 4-wire cable.

**Software**

Software development took into account a variety of operation options. Software was developed in C++ language, which allows its operation by Windows platform.

**Sample intake mechanism**

Sampling procedure is made by using a rubber tube, ½” diameter, 4.6 m length. At one end it was attached a retention valve and a strainer, which aims to prevent solid matter admission larger than 1 cm diameter. Strainer was built using a PVC tube, ¾” diameter, 0.22 m length, punched.
Figure 3. Basic scheme of the plate circuit equipment hardware.

**Ultrasonic sensor**

It is composed by a sensor and a circuit installed in a PVC tube (40 mm diameter, 0.2 m length), covered with PVC caps on both ends. The sensor was fixed at one end, while the other end the cable was attached. The sensor must be installed over the water surface at a distance (between 1-2 m), with its face pointed to the water surface. Sensor is connected to the data acquisition system by using an 8-m length parallel 4-wire cable.

**Software**

Software development took into account a variety of operation options. Software was developed in C++ language, which allows its operation by Windows platform.

**Sample intake mechanism**

Sampling procedure is made by using a rubber tube, ½” diameter, 4.6 m length. At one end it was attached a retention valve and a strainer, which aims to prevent solid matter admission larger than 1 cm diameter. Strainer was built using a PVC tube, ¾” diameter, 0.22 m length, punched.

**User-equipment communication interface**

Communication interface is an aspect of great importance for automatic equipments. An appropriate interface allows the user a good communication, user visualization and easy operational sampling programming. User-equipment was designed to operate with two mechanisms: (a) membrane keyboard (12 keys, fixed at the top cover); (b) LCD display. Keyboard and LCD display allows accessing the programming menu and defining sampling option. Once sampling operation has been concluded, equipment generates a report.
CONCLUSIONS

The prototype has been subjected to laboratory experimental runs in order to test its hydraulic and sampling operational efficiency. The purpose is testing its capacity on making the programming tasks. Some adjustments and improvements have proved to be of great concern. One of them refers to an easy exchange of the silicon tube by the user. The cost of the materials used on the prototype development was estimated to be 1,620 american dollars.

ACKNOWLEDGEMENTS

This approach was financially supported by CNPQ (Brazilian National Research Council), Ministry of Science and Technology, Brazilian Government.

REFERENCES


B. Geelhoed(a), F.P.J. Lamé(b)

a) Delft University of Technology, Mekelweg 15, 2629 JB Delft, The Netherlands, B.Geelhoed@tudelft.nl
b) Deltares, PO Box 85467, 3508 AL Utrecht, The Netherlands, frank.lame@deltares.nl

Abstract: During the sampling of contaminated soil, sampling errors are unavoidable because of the spatial heterogeneity of the contaminant distribution. The variance is a convenient indicator for the potential magnitude of these errors. Four variance estimators are constructed for use in material sampling, all of which take account of the heterogeneity and the sampling design. Based on large scale three-dimensional computerized models of contaminant heterogeneity in soil stockpiles, these variance estimators are compared using a Monte Carlo simulation of different sampling designs. The Mean Squared Error (MSE) of each variance estimator is used to assess (and compare) the performance of each variance estimator: the lower the mean square error, the better its performance.

Key words: Sampling; soil; variance; estimator; heterogeneity.

1. INTRODUCTION

1.1. Sampling Error

In order to determine if a potentially harmful contaminant in a batch of soil actually poses a threat, a first step in its risk assessment will be to analyze a sample obtained from that batch. The contaminants present will often be expressed as a concentration; e.g. mg/kg. As a series of steps (including sampling, sub-sampling and analyses) is necessary to determine the contaminants concentration, errors occurring during these steps will most likely result in a difference between the “true value” (the real mean concentration in the population) and the estimate (the concentration determined in the sample). This article will only focus on the error caused by primary sampling. Errors occurring during the following steps of sub-sampling and analysis are not considered. The sampling error (es) can be expressed as:

\[ es = c_{\text{sample}} - c_{\text{batch}} \]  

(1)

where:

\( c_{\text{sample}} \) = the concentration of the contaminant in the sample.
\( c_{\text{batch}} \) = the real mean concentration of the contaminant in the batch (unknown).
1.2 Variance

The sampling error is generally expected to differ from sample-to-sample in a random way as $c_{sample}$ of sample i will differ from $c_{sample}$ of sample j, both randomly differing from $c_{batch}$. A useful quantity is therefore the sample-to-sample variance of the sampling error, denoted by the symbol $V(e_s)$. Taking the variance operator on both sides of (1) results in $V(e_s) = V(c_{sample})$, where $V(c_{sample})$ denotes the sample-to-sample variance of $c_{sample}$.

The sample-to-sample variance can (in principle) be estimated by taking multiple samples. Leaving the errors caused by the necessary subsequent steps of sub-sampling and analyses out of consideration, the variance $V(c_{sample})$ can be estimated by:

$$V_{est}(c_{sample}) = \frac{\sum_i (c_{sample,i} - c_{av})^2}{(N-1)}$$

(2)

where:

$V_{est}(c_{sample}) = \text{the estimate of } V(c_{sample})$.

$\sum_i = \text{a summation over index } i$.

$c_{sample,i} = \text{the sample concentration in the } i\text{-th sample}$.

$N = \text{the number of samples}$

$i = \text{an index of each sample } (i=1,\ldots,N)$.

$c_{av} = \text{the average of the } N \text{ values } c_{sample,i}$.

When two or more sampling results are available, (2) can be used to estimate the sample-to-sample variance at the scale of the sample size.

1.3 Scale and variance

Estimating the concentration of the soil batch by the concentration of a sample of a specific scale introduces variability as the sample concentration will differ randomly from the true mean concentration of the batch. Consequently, a relation exists between the scale of the sample and the variance.

Limiting the sample-to-sample variance is of major importance as in practice decisions on the risks posed by the contaminants will often be based on one, or at most a small number of samples. This raises the question how the sample-to-sample variance can be practically determined and what a cost effective sampling strategy will be, considering both the scale of the sample and the number of increments.

1.4 Increments

The new approach of this article is based on information on the scale of the increments of which the sample is composed. This information is generally not available, but it is expected that it can be made available in future at the expense of an increased effort of sampling and increased costs of analyses. In section 2, it will be demonstrated that it is in principle possible to estimate the sample-to-sample variance even when using a single sample, provided that information at the level of the increments is made available (e.g. by analyzing all increments separately).

A batch of soil can be thought of as a population of increments. If each increment in the population is assigned a unique number taken from the set of numbers ranging from 1 to $N_{pop}$, where $N_{pop}$ is the total number of increments in the population, the concentration ($c_{sample}$) can be written as a function of the individual increment properties:

$$c_{sample} = \frac{\sum_i I_i m_i c_i}{(\sum_i I_i m_i)}$$

(3)

where:

$\sum_i = \text{a summation over index } i$. 

$\Sigma_i = \text{a summation over index } i$.
\[ \sum_j = \text{a summation over index } j. \]
\[ m_i = \text{the mass of the } i\text{-th increment in the population.} \]
\[ m_j = \text{the mass of the } j\text{-th increment in the population.} \]
\[ c_i = \text{the concentration in the } i\text{-th increment in the population} \]
\[ c_j = \text{the concentration in the } j\text{-th increment in the population} \]
\[ I_i = \text{the indicator of the } i\text{-th increment in the population.} \]
\[ I_j = \text{the indicator of the } j\text{-th increment in the population.} \]

\( I_i \) (or \( I_j \)) is one when the \( i\)-th (or \( j\)-th) increment is part of the sample and zero otherwise. This ensures that the summations in (3) only sum over the increments in the sample.

2. Estimators for the variance

It was shown by Geelhoed [2004] that (under the conditions of a constant sample mass and constant first-order inclusion probability) the concentration in a sample (defined by (3)) can be seen as a \( \pi \)-expanded estimator (see e.g. Särndal et al. [1992]):

\[ Y_\pi = \sum_i I_i y_i / \pi_i \]

where:
\[ Y_\pi = \text{the } \pi\text{-expanded estimator for the population total of } y_i. \]
\[ \Sigma_i = \text{a summation over index } i. \]
\[ y_i = \text{the variable of interest (here: } y_i = m_i c_i / M_{\text{pop}} \text{ where } M_{\text{pop}} \text{ is the mass of the entire population).} \]
\[ i = \text{the number of an increment in the population.} \]
\[ I_i = \text{the indicator of the } i\text{-th potential increment in the population.} \]
\[ \pi_i = \text{the first-order inclusion probability of the } i\text{-th potential increment in the population (see e.g. Särndal et al. [1992]). Because } \pi_i \text{ appears in the denominator, the estimator } Y_\pi \text{ can only be applied if } \pi_i > 0. \]

Substituting \( y_i = m_i c_i / M_{\text{pop}} \) (where \( M_{\text{pop}} \) is the mass or weight of the entire population) and assuming constant first-order inclusion probabilities \( \pi_i = M_i / M_{\text{pop}} \) indeed results in \( Y_\pi = c_{\text{sample}} \).

The Horvitz-Thompson estimator \( (V_{HT}) \) (see e.g. Särndal et al. [1992]) and the Sen-Yates-Grundy estimator \( (V_{SYG}) \) (Sen [1953], Yates and Grundy [1953]) are defined by the following equations:

\[ V_{HT} = \sum_i \sum_j I_i I_j (1 - \pi_i \pi_j / \pi_{ij}) (y_i / \pi_i) (y_j / \pi_j) \]

and

\[ V_{SYG} = -(1/2) \sum_i \sum_j I_i I_j (1 - \pi_i \pi_j / \pi_{ij}) (y_i / \pi_i - y_j / \pi_j)^2 \]

where:
\[ \pi_{ij} = \text{the joint inclusion probability of the } i\text{-th and } j\text{-th potential increment in the population. Because } \pi_{ij} \text{ appears as the denominator in a fraction, the estimators } V_{HT} \text{ and } V_{SYG} \text{ can only be applied if } \pi_{ij} > 0. \text{ Note the following definition (which is not standard, but which allows the above simplified equations for } V_{HT} \text{ and } V_{SYG}: } \pi_{ij} \text{ exists and equals } \pi_i. \]

Using the “parameter for the dependent selection of particles” [1], \( C_{ij} = 1 - \pi_{ij} / (\pi_i, \pi_j) \), and using \( y_i = m_i c_i / M_{\text{pop}} \) results in:

\[ V_{HT} = -(1/M_s^2) \sum_i \sum_j I_i I_j (C_{ij} / (1 - C_{ij})) m_i c_i m_j c_j \] (4)
V_{SYG} = \left(1/(2M_{s}^{2})\right)\Sigma_{i}\Sigma_{j}I_{i}I_{j} \left(C_{ij}/(1-C_{ij})\right) (m_{i}c_{i} - m_{j}c_{j})^{2} \tag{5}

where:

\begin{align*}
M_{s} &= \text{the mass of a sample} \\
C_{ij} &= \text{the parameter for the dependent selection of particles. Because of the conditions } \pi_{i}>0 \text{ and } \pi_{ij}>0 \text{ for applicability of } V_{SYG} \text{ and } V_{HT}, \text{ it follows that the above equations ((4) and (5)) can only be applied if } C_{ij}<1, \text{ not if } C_{ij}=1. \\
\end{align*}

Hence, the (estimated) variance depends on the parameters \(M_{s}, C_{ij}, m_{i}\) and \(c_{i}\) of the increments in the sample.

It is noted here that the above two variance estimators are entirely general: they are valid for all possible sampling designs (provided \(C_{ij}<1\)). Differences between sampling designs are taken into account by the parameters \(M_{s}\) and \(C_{ij}\); different designs can lead to different values for these two parameters. A condition for the estimator \(V_{SYG}\) to be unbiased is that the number of increments in a sample has a zero variance. \(V_{HT}\) does not have this restriction, but both \(V_{HT}\) and \(V_{SYG}\), as given in (4) and (5), were constructed using the assumption that the sample mass \(M_{s}\) is constant. This is indeed true when both the increment mass and the number of increments are constant.

Given the fact that \(V_{HT}\) is slightly more general than \(V_{SYG}\), it is interesting to investigate possible improvements to \(V_{HT}\). Inspection of (4) shows that \(V_{HT}\) is potentially sensitive to systematic errors made in the determination of \(c_{i}\) and \(c_{j}\). However, the sample-to-sample variance \(V(c_{\text{sample}})\) is not influenced by systematic errors in determination of \(c_{i}, c_{j}\) and \(c_{\text{sample}}\). Therefore it makes sense to adapt \(V_{HT}\) as follows:

\begin{align*}
V_{AD1} &= -(1/M_{s}^{2})\Sigma_{i}\Sigma_{j}I_{i}I_{j} \left(C_{ij}/(1-C_{ij})\right) m_{i}m_{j} (c_{i} - c_{\text{sample}}) (c_{j} - c_{\text{sample}}) \tag{6}
\end{align*}

where:

\begin{align*}
V_{AD1} &= \text{the adapted variance estimator based on } V_{HT}. \\
\end{align*}

A special scenario is Poisson sampling (see e.g. Särndal et al. [1992]). During Poisson sampling, each increment is independently subject to a probabilistic selection process. Therefore \(C_{ij} = 0\) (for unequal \(i\) and \(j\)). Substituting this value for \(C_{ij}\) in the expression for \(V_{SYG}\) leads to \(V_{SYG}=0\) (which underestimates the variance which is generally non-zero). Substitution of \(C_{ij}=0\) and \(C_{ii}=1-M_{\text{pop}}/M_{s}\) in the expression for \(V_{AD1}\) yields:

\begin{align*}
V_{Poisson} &= ((1 - M_{s}/M_{\text{pop}})/M_{s}^{2})\Sigma_{i}I_{i}m_{i}^{2} (c_{i} - c_{\text{sample}})^{2} \tag{7}
\end{align*}

Even if the used sampling design differs substantially from the Poisson sampling design, the estimator \(V_{Poisson}\) may in many cases still provide a reasonable variance estimate if the spatial arrangement of increments in the population can be considered to be completely random. (7) also offers considerable computational simplification with respect to (4), (5) and (6), because (7) contains only a single summation symbol, while (4), (5), and (6) require double summations. Another advantage of \(V_{Poisson}\) is that it does not depend anymore on \(C_{ij}\); the variance can also be estimated when \(C_{ij}=1\), which is not possible using \(V_{HT}, V_{SYG} \text{ or } V_{AD1}\). Moreover, (7) is intuitively easier to understand: the estimate of sample-to-sample variance equals the (weighted) within sample variance \(\left((1/M_{s})\Sigma_{i}I_{i}m_{i}^{2} (c_{i} - c_{\text{sample}})^{2}\right)\) divided by the sample mass \(M_{s}\) in order to convert a within-sample variance to a between sample variance and multiplied by a finite population correction factor \((1-M_{s}/M_{\text{pop}})\) in order to take into account the effect of the finite population size on the variance. However, despite all these advantages of \(V_{Poisson}\), there is no guarantee that the estimate obtained with \(V_{Poisson}\) is accurate, especially because \(V_{Poisson}\) neglects
taking into account the influence of $C_{ij}$ on the variance. Therefore, in this article the performance of $V_{HT}$, $V_{AD1}$, $V_{SYG}$ and $V_{Poisson}$ will be compared.

3. SIMULATION STUDY

A Monte Carlo simulation is presented, which is applied to large scale three-dimensional computerized models of contaminant heterogeneity in soil stockpiles. The procedure is:

1. An independent random sample is drawn without replacement from a virtual population and $c_{\text{sample}}$ and $V_{HT}$, $V_{SYG}$, $V_{AD1}$, and $V_{Poisson}$ are recorded for this sample (calculated using (3), (4), (5), (6) and (7) respectively) wherein the sample is obtained by applying incremental sampling (see section 3.2).
2. The sample “material” is put back in the virtual population.
3. Go back to step 1 until 50,000 repetitions are performed.

3.1 Model of a Population

Thirty virtual models of contaminated soil stockpiles were constructed (Lamé et al. [2005]). These models are based on actual three dimensional concentration data from three soil stockpiles (denoted by “gas”, “dpa” and “rok”) for each of which ten virtual models were constructed. Obviously, only a small fraction of the original soil lots was analyzed. However, the data obtained were used to define, within the same statistical distribution of the observations, all increments in those soil lots, wherein by spatial simulation within the series of ten models, different degrees of large scale heterogeneity were introduced. Hence, the three series of ten models represent soil lots that, with the same statistical distribution, differ considerably in their degree of spatial distribution of highs and lows, thus their degree of large scale heterogeneity.

Here, only a brief description of the geometry of the virtual populations is given. The details of construction are described elsewhere (Lamé et al. [2005]). Each population consists of a rectangular arrangement of cubic “cells” (which represent the increments) wherein for each individual cell the concentration of the contaminant is defined. Consequently, for each of the thirty models the full population in known. The arrangement of cells in each virtual populations is similar to the arrangement displayed in Figure 1, but larger.

Each increment will not only be characterized by its mass ($m_i$) and concentration ($c_i$), but also by its spatial location within the virtual population. Here this location is denoted by the triple of indices $(x,y,z)$. The parameter for the dependent selection of particles ($C_{ij}$) will depend on the spatial locations of increments $i$ and $j$ relative to each other.

**Figure 1.** Model of the population. In the here-depicted model $z$ ranges from 1 to 2; $y$ ranges from 1 to 12; $x$ ranges from 1 to 20; and the population is divided in 10 equally-size blocks. The models that are used here are significantly larger (and therefore more realistic).
Note: for simplicity the assumption $m_i=1$ is made during the simulations, which, under the boundary condition that $m_i$ is constant, is of no effect to the results. The size of these virtual populations is as follows:

- “gas”: height = 15 cells; width = 371 cells; length = 1054 cells. In this article the “gas” populations are cropped to 15x368x1052 because this makes it easier to divide the population in equally-sized blocks.
- “dpa”: height = 16 cells; width = 920 cells; length = 400 cells.
- “rok”: height = 16 cells; width = 500 cells; length = 735 cells. In this article the “rok” populations are cropped to 16x500x732 because this makes it easier to divide the population in equally-sized blocks.

3.2 The Sampling Design

For the purpose sampling, each “block” is subdivided in vertical stacks of increments. Each vertical stack is characterized by its location in the $(x,y)$ plane and consists (for given $x$ and $y$) of all increments with coordinates $(x, y, 1), (x, y, 2), \ldots, (x, y, z_{\text{max}})$ where $z_{\text{max}}$ is the “height” of the virtual model (i.e. the number of horizontal layers, which was 15 or 16 depending on which virtual model was used). Hence each vertical stack contains the increments from all heights at its $(x, y)$ location in the virtual model. If the selection of the sample only takes vertical stacks as a whole, possible large scale heterogeneity in the vertical direction will not result in biased sampling, as all heights will be equally represented in the sample. This will be the case for the here-described sampling design.

From each “block” in the population (see Figure 1) a constant number ($n$) of vertical stacks of increments is selected at random locations in the two-dimensional $(x,y)$-plane. Selection of vertical stacks of increments in a block is by simple random sampling of vertical stacks (see e.g. [3] for a precise definition of “simple random sampling”). The overall sample is formed by combining the simple random samples of vertical stacks from each block (see Figure 2 for a simple illustration).

It is noted that the here-described sampling design can be described as “stratified simple random sampling”. In case the number of blocks (which equals the number of strata) equals one, this mode of sampling reduces to simple random sampling.

![Figure 2. Schematic depiction of the sampling design. This is a top view of the same virtual population as in Figure 1. The grey squares represent the vertical stacks of increments that are part of a random sample. In the here-depicted scenario, the population is divided into 10 equally-sized blocks and from each block two vertical stacks of increments are selected.](image-url)
The sample therefore consists (in this example) of $10 \times 2 = 20$ vertical stacks of increments.

The values of $C_{ij}$ for the here-described sampling design are as follows:

- If increment $i$ and $j$ belong to different blocks: $C_{ij} = 0$. This follows from the fact that increments belonging to different blocks are selected independently.
- If increment $i$ and $j$ belong to the same vertical stack: $C_{ij} = 1 - \frac{1}{q}$ (where $q$ is the first-order inclusion probability of an increment, which is equal to the ratio of the number of increments in the sample and the total number of increments in the population before the sample was taken). This value follows from the fact that for increments belonging to the same vertical stack $\pi_{ij} = \pi_i$ combined with the definition of $C_{ij}$.
- If increment $i$ and $j$ belong to the same block, but to different vertical stacks: $C_{ij} = \frac{1-q}{n-q}$. A derivation is presented in Appendix A.

In a first computer experiment, using the here described sampling design, the thirty models were sampled using three different variations of incremental sampling:

1. (Simple random sampling) $N_{\text{block}} = 1$, $n=10$ (in total the sample consist of 10 stacks of 15 or 16 (depending on which virtual model was used) increments each).
2. $N_{\text{block}} = 4$ (obtained by “quartering” the population in four equal quadrants in the $(x,y)$ plane), $n=8$ (in total the sample consist of 32 stacks of 15 or 16 (depending on which virtual model was used) increments each).
3. $N_{\text{block}} = 16$ (obtained by quartering the population and then quartering each quarter in a similar way, resulting in $4 \times 4 = 16$ blocks), $n=2$ (in total the sample consists of 32 stacks of 15 or 16 (depending on which virtual model was used) increments each).

Hence, a total of $30 \times 3 = 90$ sampling exercises were performed in the first computer experiment, each of which was repeated 50,000 times.

Because the number of increments in each sample was rather low during the first computer experiment (either 150 ($=15 \times 10$), 160 ($=16 \times 10$), 480 ($=15 \times 32$) or 512 ($=16 \times 32$) increments per sample), a second computer experiment was performed to study the behavior of the variance estimators as a function of the sample size (the number of increments per sample). Because of the increase in computational time per sample, only three populations (gas01, dpa03 and rok10) were selected for this second computer experiment. Of these models, gas01 is a model with a strong level of large scale heterogeneity, while the model rok10 does not have a high degree of large scale heterogeneity. The sampling method was again stratified simple random sampling.

For gas01 and rok10, the number of blocks was set to 16. The number of vertical stacks included in a sample per block varied: 2, 4, and 8. As a consequence, the number of increments for dpa01 in the second experiment per sample varied from 480 ($=15 \times 16 \times 2$) to 1,920 ($=15 \times 16 \times 8$). The number of increments for rok10 varied from 512 ($=16 \times 16 \times 2$) to 2,048 ($=16 \times 16 \times 8$).

For dpa03 the number of blocks was set to 1 and the number of vertical stacks per block varied: 10, 20, 40, 80, and 160. As a consequence, the number of increments per sample for dpa03 in the second experiment varied from a minimum of 160 ($=16 \times 10$) to 2,560 ($=16 \times 160$).

### 3.3 Results

The results of the first experiment are presented for all thirty virtual populations. As described in section 3.1, these virtual populations model three actual populations at different levels of spatial segregation. The three actual populations are denoted by “gas”, “dpa” and “rok”. The virtual populations are denoted by “gas01” to “gas10”, “dpa01” to “dpa10”, and “rok01” to “rok10” for “gas”, “dpa” and “rok” respectively. The number that ranges from 1 to ten indicates
the degree of large scale heterogeneity: 1 is the highest degree of large scale heterogeneity and 10 the lowest.

For each scenario (defined by its unique combination of virtual population number, \(n\) and \(N_{\text{block}}\)), a list was obtained of 50,000 independent estimates for \(c_{\text{sample}}\), \(V_{\text{HT}}\), \(V_{\text{AD}}\), \(V_{\text{Poisson}}\) and \(V_{\text{SYG}}\). Using (2) resulted in a numerical value for \(V_{\text{est}}(c_{\text{sample}})\). In view of the high number of replications (\(N=50,000\)), it is expected that \(V_{\text{est}}(c_{\text{sample}})\) is very close in value to \(V(c_{\text{sample}})\).

The Mean Squared Error (MSE) of each estimator was calculated (using the 50,000 values for each estimator in each scenario) as the mean of the squared difference of \(V_X\) (where \(X\) can be HT, SYG, AD1 or Poisson) and the “true” variance \(V(c_{\text{sample}})\) (which was approximated by \(V_{\text{est}}(c_{\text{sample}})\)).

For both experiments, it was observed that for each given sample the three estimators \(V_{\text{SYG}}, V_{\text{HT}}\) and \(V_{\text{AD}}\) resulted in three numerically exactly identical values, but \(V_{\text{Poisson}}\) resulted in a different value (and the difference can vary from sample-to-sample). Using the values of \(C_{ij}\) given in section 3.2 it was then also mathematically proven that \(V_{\text{SYG}} = V_{\text{HT}} = V_{\text{AD}}\) for the sampling design used here (the detailed proof is not given here). The equality \(V_{\text{SYG}} = V_{\text{HT}} = V_{\text{AD}}\) follows from the constant number of increments per block, the sampling of all block, and the specific values of \(C_{ij}\) used here. For non constant number of increments per block, not sampling all blocks, or different \(C_{ij}\) values, \(V_{\text{SYG}}, V_{\text{HT}}\) and \(V_{\text{AD}}\) do not have to be equal.

### 3.3.1 Results for the First Computer Experiment

High Mean Squared Errors were observed during the first computer experiment. The MSE for \(V_{\text{HT}}\) (and also of \(V_{\text{AD1}}\) and \(V_{\text{SYG}}\)) varied from 37% to 2973%. The MSE of \(V_{\text{Poisson}}\) varied from 33% to 2956%. An overview of all results is presented in Figure 3.

The results indicate that \(V_{\text{Poisson}}\) structurally has a slightly lower MSE than \(V_{\text{HT}}\), for the conditions of the first computer experiment. However, when the MSE of a variance estimator is larger than 100% the practical usefulness of the variance estimate has most likely disappeared. Therefore, the result (that the MSE for \(V_{\text{Poisson}}\) is lower than for \(V_{\text{HT}}\) under the conditions of the first computer experiment) should not be used to falsely conclude that \(V_{\text{Poisson}}\) should become the preferred estimator in practice. The MSE of an estimator may also depend on the sample size (where generally lower MSE’s are expected at higher sample sizes).

One possible attempt to deal with the problem of high MSE’s associated with a variance estimator would be to select a larger sample size (i.e. more increments), because selecting a larger sample size can lead to a lower MSE of a variance estimator. This is studied in the second computer experiment. It is noted that selecting larger sample sizes also serves a more immediate interest: the MSE of \(c_{\text{sample}}\) can be decreased by increasing the sample size, because \(V(c_{\text{sample}})\) generally decreases with increasing sample size. It is also noted that the MSE of \(V_{\text{HT}}\) or \(V_{\text{Poisson}}\) may be high, while at the same time the MSE of \(c_{\text{sample}}\) may be low. Vice versa is also possible: low MSE of \(V_{\text{HT}}\) or \(V_{\text{Poisson}}\), while at the same time a high MSE of \(c_{\text{sample}}\).
<table>
<thead>
<tr>
<th>Virt. pop.</th>
<th>Mode</th>
<th>MSE(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rok10</td>
<td>A B C</td>
<td></td>
</tr>
<tr>
<td>rok09</td>
<td>A B C</td>
<td></td>
</tr>
<tr>
<td>rok08</td>
<td>A B C</td>
<td></td>
</tr>
<tr>
<td>rok07</td>
<td>A B C</td>
<td></td>
</tr>
<tr>
<td>rok06</td>
<td>A B C</td>
<td></td>
</tr>
<tr>
<td>rok05</td>
<td>A B C</td>
<td></td>
</tr>
<tr>
<td>rok04</td>
<td>A B C</td>
<td></td>
</tr>
<tr>
<td>rok03</td>
<td>A B C</td>
<td></td>
</tr>
<tr>
<td>rok02</td>
<td>A B C</td>
<td></td>
</tr>
<tr>
<td>rok01</td>
<td>A B C</td>
<td></td>
</tr>
<tr>
<td>dpa10</td>
<td>A B C</td>
<td></td>
</tr>
<tr>
<td>dpa09</td>
<td>A B C</td>
<td></td>
</tr>
<tr>
<td>dpa08</td>
<td>A B C</td>
<td></td>
</tr>
<tr>
<td>dpa07</td>
<td>A B C</td>
<td></td>
</tr>
<tr>
<td>dpa06</td>
<td>A B C</td>
<td></td>
</tr>
<tr>
<td>dpa05</td>
<td>A B C</td>
<td></td>
</tr>
<tr>
<td>dpa04</td>
<td>A B C</td>
<td></td>
</tr>
<tr>
<td>dpa03</td>
<td>A B C</td>
<td></td>
</tr>
<tr>
<td>dpa02</td>
<td>A B C</td>
<td></td>
</tr>
<tr>
<td>dpa01</td>
<td>A B C</td>
<td></td>
</tr>
<tr>
<td>gas10</td>
<td>A B C</td>
<td></td>
</tr>
<tr>
<td>gas09</td>
<td>A B C</td>
<td></td>
</tr>
<tr>
<td>gas08</td>
<td>A B C</td>
<td></td>
</tr>
<tr>
<td>gas07</td>
<td>A B C</td>
<td></td>
</tr>
<tr>
<td>gas06</td>
<td>A B C</td>
<td></td>
</tr>
<tr>
<td>gas05</td>
<td>A B C</td>
<td></td>
</tr>
<tr>
<td>gas04</td>
<td>A B C</td>
<td></td>
</tr>
<tr>
<td>gas03</td>
<td>A B C</td>
<td></td>
</tr>
<tr>
<td>gas02</td>
<td>A B C</td>
<td></td>
</tr>
<tr>
<td>gas01</td>
<td>A B C</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 3.** The Mean Squared Errors (MSE’s) of $V_{HT}$ and $V_{Poisson}$ expressed as a percentage of the sample-to-sample variance of $c_{sample}$ for the 90 scenarios studied during the first computer experiment. The sampling modes are: A = (N\(_{\text{Block}}=1\), n=10), B=(N\(_{\text{Block}}=4\), n=8), B=(N\(_{\text{Block}}=16\), n=2).

### 3.3.2 Results for the Second Computer Experiment

It was observed that the MSE (when expressed as a percentage of $V(c_{sample})$) decreases when the sample size increases for $V_{HT}$, $V_{AD}$ and $V_{SYG}$, while for $V_{Poisson}$ the MSE remained...
approximately constant in one of the three scenarios studied during the second computer experiment. This implies that the MSE’s of \( V_{HT} \), \( V_{AD} \) and \( V_{SYG} \) can be made arbitrarily small by selecting a large enough sample. Despite the high MSE’s observed in the first computer experiment, the second experiment therefore demonstrates the potential practical usefulness of \( V_{HT} \), \( V_{AD} \) and \( V_{SYG} \). Based on the results, it is recommended that \( V_{Poisson} \) is not used in practice. The results are graphically depicted in Figure 4, 5 and 6.

Mathematically, the results can be explained as follows: because \( V_{Poisson} \) neglects the influence of \( C_{ij} \) on the variance it is likely to be biased in most circumstances, while \( V_{HT} \) is unbiased (for the scenarios described in this study). The MSE is composed of a bias component and a variance component. While it can be expected that the variance component reduces with increasing sample size, the bias component does not have to decrease with increasing sample size. This is a possible explanation for the observed trends in Figure 4, 5 and 6.

**Figure 4.** The Mean Squared Errors (MSE’s) of \( V_{HT} \) and \( V_{Poisson} \) expressed as a percentage of the sample-to-sample variance of \( c_{sample} \) for the scenarios studied during the second computer experiment. Virtual population = gas01. The sampling modes are: A = (N\( \text{Block} = 16 \), n=2), B=(N\( \text{Block} = 16 \), n=4), C=( N\( \text{Block} = 16 \), n=8).

**Figure 5.** The Mean Squared Errors (MSE’s) of \( V_{HT} \) and \( V_{Poisson} \) expressed as a percentage of the sample-to-sample variance of \( c_{sample} \) for the scenarios studied during the second computer experiment. Virtual population = rok10. The sampling modes are: A = (N\( \text{Block} = 16 \), n=2), B=(N\( \text{Block} = 16 \), n=4), C=( N\( \text{Block} = 16 \), n=8), and D=( N\( \text{Block} = 16 \), n=16).
Figure 6. The Mean Squared Errors (MSE’s) of $V_{HT}$ and $V_{Poisson}$ expressed as a percentage of the sample-to-sample variance of $c_{sample}$ for the scenarios studied during the second computer experiment ($V_{HT}$ and $V_{Poisson}$ resulted in almost identical values of MSE). Virtual population = dpa03. The sampling modes are: E = (NBlock=1, n=10), F=(NBlock=1, n=20), G=(NBlock=1, n=40), H=(NBlock=1, n=80) and I=(NBlock=1, n=160).

In Figure 7, the error of estimation of $V_{HT}$ and $V_{Poisson}$ (expressed as a percentage of $V(c_{sample})$) of 20 samples are graphically depicted on a “dart board”. The samples where taken from “gas01” using the same sampling modes A, B and C as in Figure 4. Figure 7 clearly illustrates the presence of bias in $V_{Poisson}$ and the reduction of the error (and hence the MSE) of $V_{HT}$ with increasing number of increments per sample.
Figure 7. The errors (expressed as a percentage of $V(c_{\text{sample}})$) of $V_{HT}$ and $V_{Poisson}$ for the same virtual population and sampling modes A, B and C as in Figure 4. The distance of a point to the origin represents the magnitude of the error. Hence the analogy with a dart board: a point at the origin means that no error is made when the corresponding sample and estimator are used to estimate the variance. Three circles are depicted for each “dart board”: a circle indicating the region with error=50%, 100% and 150%. As the number of increments per sample increase going from A to C, the scatter of $V_{HT}$ becomes more centered in a region close to the origin.

4. DISCUSSION

This study focused on the variance estimators and their MSE. The primary question of whether or not a particular sampling mode will yield samples of sufficiently low sampling variance was not directly addressed in this article. But the results of this article are highly relevant for this question, if it leads to the construction of methods for accurate variance estimation.
5. CONCLUSIONS

Virtual populations were used to study the MSE of four variance estimators. In a first computer experiment (at low sample size) the following results with respect to the MSE (expressed as a percentage of $V(c_{sample})$) were obtained:

- The MSE of $V_{HT}$, $V_{SYG}$, and $V_{AD1}$ ranges from 37% to 2973%
- The MSE of $V_{Poisson}$ ranges from 33% to 2956%

Although these results suggest that $V_{Poisson}$ performs slightly, but structurally, better, the MSE’s of all estimators were high. For $V_{HT}$, $V_{SYG}$ and $V_{AD1}$ this was caused by the low sample sizes during the first computer experiment. It is noted that the structurally lower MSE’s of $V_{Poisson}$ during the first experiment do not necessarily indicate that $V_{Poisson}$ itself is structurally lower as well.

A second computer experiment demonstrates that the MSE (expressed as a percentage of $V(c_{sample})$) decreased with increasing sample size for $V_{HT}$, $V_{AD1}$ and $V_{SYG}$. This indicates that these estimators are of interest for future practical application (as opposed to $V_{Poisson}$). For these estimators ($V_{HT}$, $V_{AD1}$ and $V_{SYG}$) the accuracy of the estimate can be improved by increasing the sample size. For $V_{Poisson}$ the MSE does not always decrease as a function of increasing sample size, because of potential bias in $V_{Poisson}$.

ACKNOWLEDGEMENTS

This work was performed as part of a project supported by the Netherlands Technology Foundation STW, under STW grant 7457. NFI, Deltares, Nutreco, Hosokawa and Organon are members of the users’ committee of this project.

REFERENCES


Appendix A

During simple random sampling of vertical stack of particles, every combination of n vertical stacks has an equal probability of being selected. Denoting the number of vertical stacks from which the sample will be taken by $N_{str}$ the probability of becoming part of the sample of each particle is therefore given by $n/N_{str}$, i.e:

$$\pi_i = n/N_{pop}$$

On the condition that the first particle has become part of the sample, the probability that the second particle (belonging to a different vertical stack of particles) becomes part of the sample is given by $(n-1)/(N_{str}-1)$. The joint inclusion probability of the pair is therefore $(n(n-1))/(N_{str}N_{pop})$.
\times (N_u - 1) = \frac{n}{N_u}N_u^2(1-(1-q)/(n-q)), where q is the sampling fraction defined (here) by \( q = \frac{n}{N_u} \).

Expressed as an equation:

\[ \pi_{ij} = \pi_i \pi_j (1-(1-q)/(n-q)) \]

From the latter equality, combined with the definition of \( C_{ij} \), it follows that (for \( i \neq j \)):

\[ C_{ij} = \frac{(1-q)}{(n-q)}. \]
Decision Support for Environmental Monitoring and Restoration: Application of the Partially Observable Markov Decision Process

David Tomberlin
National Marine Fisheries Service
1315 East-West Highway, Silver Spring, MD 20901 USA
david.tomberlin@noaa.gov

Abstract: Environmental monitoring programs can improve management over time, but generally require that correspondingly less time and money be put into direct restoration efforts such as revegetation or dam removal. Thus, budget constraints compel environmental managers to make difficult decisions regarding the allocation of scarce funds and personnel between environmental monitoring and environmental restoration. Among other factors, the best allocation of resources between monitoring and restoration—or, more generally, learning and doing—will depend on the quality of information available from a monitoring program. This paper demonstrates the application of the partially observable Markov decision process (POMDP) as a framework for investigating the optimal intensity of monitoring given stochastic state dynamics and imperfect observations on state variables. Specifically, the paper addresses the problem of choosing among a set of available monitoring protocols that differ in their costs and the type of information they provide. An empirical application of the model to erosion control in California watersheds demonstrates the utility of the resulting decision policy as well as limitations to the approach.

Keywords: monitoring; optimal learning; value of information; adaptive management.

1. INTRODUCTION

Environmental management budgets often seem small relative to the protection and restoration work managers would like to pursue. A natural temptation is to skimp on monitoring because it diverts funds from projects that appear to provide more direct conservation benefit, such as cleaning up polluted sites or staffing nature reserves. While the information derived from environmental monitoring programs has obvious potential to support better long-term management decisions, the opportunity cost of dedicating scarce funds to monitoring is often enough to induce managers to forgo monitoring. And while there is a general appreciation that monitoring can contribute to management, weighing the dynamic and uncertain benefits and costs of monitoring presents a difficult decision problem. This paper develops a framework for analyzing the role of monitoring as part of a more general environmental management program, specifically for assessing the choice of monitoring protocol when managers have access to more than one protocol and when they may also choose to forego monitoring entirely.

While the issue of choosing among monitoring protocols that differ in cost and information quality arises in many areas of environmental management—and the rest of life, for that

1 An earlier version of this paper was given at the HydroPredict International Conference on Predictions for Hydrology, Ecology and Water Resource Management, 15-18 Sept. 2008, Prague, Czech Republic.

2 This paper represents the opinions of the author and not necessarily those of the National Marine Fisheries Service.
matter—the application considered here is the choice among surface water quality monitoring protocols within broader program to control sediment loading. In northern coastal California, excess sediment is considered one of the primary threats to the freshwater habitat of endangered salmonids and is the leading cause for streams to be listed as impaired under the US Clean Water Act. Erosion on roads with stream connectivity is thought to be an especially significant problem. Different approaches to monitoring sediment production on road networks are possible, with the more expensive protocols generally yielding better information (in the sense of providing more precise estimates of the true rate of sediment loading). The question that forest managers face, then, is this: is it better to opt for an expensive monitoring protocol that yields high-quality information, a less expensive protocol that yields poorer quality information, or to skip monitoring entirely and spend the resultant savings on engineering projects to reduce erosion rates?

Below, the challenge of weighing trade-offs between the cost and quality of candidate monitoring protocols is treated within a partially observable Markov decision process (POMDP). The POMDP is an extension of the Markov decision process to handle imperfectly observed state variables (such as erosion rates). That is, in the POMDP not only are the state transitions uncertain, but the state at any given point in time is observed with error. In the case of erosion, it is simply not possible to measure the volume of surface material removed from a site with great exactness, and even rough estimates require staff time and equipment. The modeling approach taken here may be thought of as a particular formalization of adaptive management, in that decisions are made in a dynamic and stochastic environment based on beliefs that change as new information becomes available. Given the pervasiveness of errors in observations on many variables that are key to environmental management (e.g., animal populations), a principled framework for incorporating observation error into decision-making is a valuable tool.

2. BACKGROUND ON ROAD EROSION IN THE REDWOOD REGION

The modeling approach to monitoring design demonstrated below, while widely applicable in environmental management, was inspired by the difficulty of developing erosion control strategies in the Middle Fork Caspar Creek watershed on the Jackson Demonstration State Forest Mendocino County, California, part of the coastal redwood region. This watershed covers approximately 1300 acres, with elevation ranging from 100 to 1000 feet and slopes from 0-100%. The climate is Mediterranean, with average annual precipitation between 40 and 70 inches. Second- and third-growth redwood and Douglas-fir are the dominant vegetative community. The watershed is underlain by the Coastal Belt of the Franciscan Assemblage, consisting of marine sedimentary and volcanic rocks. Much of the watershed has been logged two or more times, and numerous logging roads, mostly dating from the 1960s and 1970s, run through the watershed.

Erosion on forest roads can impair road function, increase transportation costs, reduce access, and necessitate substantial expenditures on repair. It may also cause undesired changes in habitat and hydrologic function, which can in turn lead to higher water treatment costs, increased flood risk, reduced recreational opportunity, and lower aquatic habitat quality. In many northern California coastal streams, the contribution of logging roads to in-stream sediment loads is a particular concern, as excess sediment can adversely affect endangered salmon and steelhead trout.

Beginning in 2005, a field study of erosion rates at ten sites in the watershed was begun, with settling basins used to measure coarse sediment production, laboratory analysis of runoff to estimate fine sediment production, and tipping buckets to estimate runoff (Barrett and Tomberlin 2007 provide preliminary analysis). The cost of instrumentation, staff time, and laboratory analysis confronted agency staff directly with questions about the role of monitoring in management and in particular what level of monitoring best balanced the costs and benefits of the program, given the information being generated. Realizing that significant time and money could be saved by scaling back the monitoring effort to measure only coarse sediment, managers asked for a comparison of the desirability of
continuing the extensive monitoring program vs. pursuing the less intensive monitoring protocol.

In developing a road erosion control strategy, forest managers must consider both natural stochasticity in the erosion processes and uncertainty about the efficacy of various management treatments in reducing erosion risk. Some treatments, such as road removal, may be both quite costly and substantially irreversible, creating an incentive to defer the decision. Maintenance, in contrast, can be continued indefinitely while the landowner retains the option to upgrade or remove the road later. Adding to the challenge of developing an erosion control strategy is the difficulty of knowing which roads or road segments are most in need of treatment. Because budgets are generally nowhere near what would be required to treat entire watersheds, prioritization among possible sites and treatments is key, and monitoring (both before and after treatments) can potentially shed important light on this prioritization. The analysis here is motivated by an important practical question in water quality management, namely, how much time and money should be put into monitoring sediment loading when those same resources could be spent instead on remedial measures to reduce sediment production at its source?

3. A DYNAMIC MODEL OF EROSION CONTROL

This section presents a modeling approach to choosing among monitoring protocols, or more generally among a candidate set of actions that differ in their costs and in the information they yield. The POMDP is a collection of sets \( \{S, P, A, W, \Theta, R\} \) (Cassandra 1994), where \( S \) is the system’s state variables, \( P \) represents state dynamics as transition probabilities, \( A \) is the actions available to an agent, \( W \) is the rewards to taking particular actions in particular states, \( \Theta \) is a set of possible observations on the state variables, and \( R \) is a set of observation probabilities. Observations \( \theta \in \Theta \) are the only information the agent has on the unobservable true state, \( S \). The observation model \( R \) describes the probabilistic relationship between observations \( \theta \) and the true state \( S \). The decision-maker (here, a land manager) uses observations \( \theta \) and the observation model \( R \) to estimate the state \( S \).

The model assumes the manager’s goal is to minimize long-run discounted total costs of sediment control, which includes the cost of monitoring sediment production. The actions that achieve this goal are identified with dynamic programming (Bertsekas 2000) through a recursively defined value function \( V \):

\[
V_i(\pi) = \max_a \left[ \sum_i \pi_i q_i^a + \beta \sum_{i,j} \pi_i p_{ij}^a r_{ij}^a V_j(\pi | a, \theta) \right]
\]

(1)

where

- \( \pi_i \) = subjective probability of being in state \( i \in S \) at time \( t \)
- \( q_i^a \) = immediate reward for taking action \( a \in A \) in state \( i \in S \) at time \( t \)
- \( \beta \) = discount factor
- \( p_{ij}^a \) = probability of moving from state \( i \in S \) at time \( t \) to state \( j \in S \) at time \( t + 1 \)
  after taking action \( a \in A \)
- \( r_{ij}^a \) = probability of observing \( \theta \in \Theta \)
  after taking action \( a \in A \) and moving to state \( j \in S \)
- \( T \) = function updating beliefs based on prior beliefs and observed \( \theta \)

\( V \) is the greatest expected net benefit that the agent can achieve over time, taking into account that as conditions change in the future, different actions may be warranted. The solution of \( V \) yields an optimal policy, which is a mapping from beliefs about the current state, \( \pi \), into the optimal action.
In our setting, the state variable $S$ is a forest road segment’s potential to deliver sediment to the stream system, which for expository purposes takes only two possible values, High Erosion and Low Erosion. The action set $A$ consists of Maintain (i.e., neither monitor sediment loading nor take remedial measures to reduce loading), Monitor Low (i.e., monitor with settling basins only), Monitor High (i.e., monitor with settling basins augmented by laboratory analysis of suspended sediment), and Treat (i.e., take measures to reduce the potential for sediment production). The observation set consists of the same two possible values as $S$, High Erosion and Low Erosion, but an observation of $\theta = \text{High Erosion}$ does not necessarily mean that the true state $S = \text{High Erosion}$. Instead, we define an observation model $R$ as follows:

$$
\begin{bmatrix}
R^1_{\theta} & = & \begin{bmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{bmatrix} \\
R^2_{\theta} & = & \begin{bmatrix} 0.63 & 0.37 \\ 0.25 & 0.75 \end{bmatrix} \\
R^3_{\theta} & = & \begin{bmatrix} 0.83 & 0.17 \\ 0.12 & 0.88 \end{bmatrix} \\
R^4_{\theta} & = & \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}
\end{bmatrix}
$$

Each matrix, with the state $j \in S$ defined by row and each observation $\theta$ defined by column, defines the probabilistic relationship of observation to true state under a different action. $R^1_{\theta}$, for example, tells us that after taking action $a=1$ (Maintain) and moving to the unobservable state $j=\text{Low Erosion}$, we would observe $\theta=\text{Low Erosion}$ with 60% probability and $\theta=\text{High Erosion}$ with 40% probability. That is, maintaining the status quo provides some weak information, presumably through casual observation of the road. $R^4_{\theta}$, in contrast, tells us that implementing a monitoring plan with settling basins ($a=2$, Monitor Low), yields a stronger basis for inference on $S$, and $R^3_{\theta}$ that the more sophisticated scheme Monitor High yields still more information. Finally, $R^4_{\theta}$ indicates that immediately after treating the road, observations tell us nothing about the true state of erosion, a reflection of how treatments often cause transient changes in erosion rates that tell us little about the true state of the road.

The stochastic dynamics of the erosion level $S$ are given by transition probability matrices defined as follows:

$$
\begin{bmatrix}
P^1_{\theta} & = & \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\
P^2_{\theta} & = & \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\
P^3_{\theta} & = & \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\
P^4_{\theta} & = & \begin{bmatrix} 0.95 & 0.05 \\ 0.90 & 0.10 \end{bmatrix}
\end{bmatrix}
$$

The first two matrices embody the assumption that under status quo maintenance or either monitoring program, the state remains unchanged. The final matrix tells us that under $a=4$ (Treat), a Low Erosion road stays in that same state with 95% probability, but there is a 5% chance that the treatment will backfire and create a High Erosion road. Similarly, treating a High Erosion road has an 90% chance of successfully creating a Low Erosion road and a 10% chance of failure (meaning the High Erosion road stays that way). These values are not derived from field data, but are chosen to reflect a plausible scenario for analysis.

Finally, the reward structure (actually, cost structure) is as follows:

$$
W^1_j = \begin{bmatrix} 0 & -15 \end{bmatrix} \quad W^2_j = \begin{bmatrix} -0.5 & -15.5 \end{bmatrix} \quad W^3_j = \begin{bmatrix} -2.5 & -17.5 \end{bmatrix} \quad W^4_j = \begin{bmatrix} -8 & -8 \end{bmatrix}
$$

Here the columns of each vector represent the rewards (in USD x 10^3) of taking a particular action in a particular state. $W^4$ tells us that maintaining the road in Low Erosion state will cost nothing, while the cost of maintaining the road in High Erosion state is $15,000 per year, due to the need for cleanup and repair. $W^2$ and $W^3$, the payoffs to monitoring, are the same as $W^4$ less the periodic cost of the monitoring program ($500 for Monitor Low or $2500 for Monitor High, assuming the equipment is already on hand). $W^4$ tells us that treating the road (by adding rock and mechanically treating likely problems) will cost us the same $8000 regardless of whether the road is in Low Erosion or High Erosion state. Comparing all these costs, it’s obvious that if the manager knew the true state to be Low Erosion, the best choice would be to Maintain ($a=1$), and if the manager knew the true
state to be *High Erosion*, the best thing to do would be to *Treat* ($a=4$). However, the premise of our model, and the reality that managers generally face, is that the true state is unknown.

4. **MODEL SOLUTION AND RESULTS**

Model solution consists of identifying the optimal value $V$ in (1) and the decision rule, or optimal policy, associated with it. This optimal policy is a mapping from the decision-maker’s beliefs about the true state of the system into an optimal action set. That is, it is a state-dependent rule in which the state that is the argument of the rule is the decision-maker’s beliefs. Because these beliefs evolve according to Bayes’ rule and can in principle take on any value on the unit interval, they are uncountably infinite. This is in contrast to most applications in stochastic dynamic programming, in which numerical solution is attained by some form of backward recursion over a discrete set of possible states.

While allowing for uncountably infinite states is a significant complication, it has been shown that the value function associated with (1) can be represented as a convex and finite set of piecewise linear segments (Monohan 1982). This result enables solution of (1) by backward recursion with a nested linear programming routine that identifies optimal candidate segments over the entire belief space (details are available in Monohan 1982 and Cassandra 1994). The backward recursion from the planning horizon $T$ identifies the set of linear segments that comprise the value function $V$, which gives the expected value of the optimal policy for every belief at any given number of periods prior to $T$.

Figure 1 shows the value function, $V$, for the model parameters given above, as it evolves over a 5-period decision horizon (the highest solid line is $V$ at $T-1$, the next down is $V$ at $T-2$, etc.). On the x-axis is $\pi$, the subjective probability that the road is in the Low Erosion state. The solid lines are the segments that constitute the value function. The dashed lines show the division of the state space into policy regions, i.e., the beliefs for which the actions *Treat*, *Monitor High*, *Monitor Low*, or *Maintain* are optimal. The most salient features of the solution are that 1) as the decision horizon lengthens, maintaining the status quo occupies progressively less of the belief space, and 2) *Monitor High* enters the optimal policy at $T-3$ and *Monitor Low* at $T-5$.

Fig. 1. The value function, $V$, and optimal policy as a function of beliefs about sediment production, for five different time horizons. The dotted black lines show the division of the belief space into regions associated with each action for a given decision horizon.
These features derive from the cost of the monitoring strategies and the value of the information they provide: only after the decision horizon has lengthened sufficiently to merit the increased expenditure on information gathering. However, Monitor Low provides relatively little extra information given its cost, hence it is the optimal action for a small set of beliefs. The choice of Treat, by contrast, occupies more than half the belief space: because the cost of maintaining and repairing a high-erosion road is substantial, treating the road without bothering to monitor first is optimal for beliefs (in the 5-period decision horizon) up to \( \pi_1 \) around 0.76. Of course, the results depicted in Figure 1 are specific to the model structure and parameters described above, and should not in any way be construed as general.

Continued iteration on the value function yields an approximately stationary optimal policy in belief space. In this case, the policy showed approximate stationarity by a decision horizon of 10 periods, with a policy in which Treat was optimal on \( \pi_1 \in [0, 0.76) \), Monitor High on \( \pi_1 \in [0.76, 0.93) \), Monitor Low on \( \pi_1 \in [0.93, 0.97) \), and Maintain on \( \pi_1 \in [0.97, 1.0] \).

The performance of a policy can be examined by simulating the policy’s application to the system as it evolves stochastically and provides observations that are used by the decision-maker to estimate the true state. Figure 2 shows the distribution of cumulative performance over a 50-year horizon for 1000 simulations of the approximately stationary optimal policy applied to the model (i.e., 1000 sequences of the policy being applied for 50 years to a sequence of states and observations following the Markov processes given above). The distribution of outcomes is strongly bimodal, due to the particular parameterization adopted: because Treat as an action generates no new information, it does not change beliefs, so for many initial beliefs the optimal policy gets stuck in long series of repeatedly choosing Treat (note that -400 is the cumulative value of choosing Treat every year over a 50-year horizon, regardless of the road’s true state). A similar effect, though not as strong applies to the choice of Maintain; i.e., because little new information is generated when Maintain is the action chosen, it tends to be chosen many times in a row. The relatively few cumulative rewards that lie in the middle of the distribution are associated with initial beliefs such that monitoring is chosen in more periods, i.e., where there is more ambiguity about the current state and where the information signal is strong enough to change beliefs more significantly.

![Figure 2](image_url)

**Figure 2:** Histogram of cumulative reward over a 50-year horizon, based on 1000 stochastic simulations of the optimal policy.
Sensitivity analysis of the assumed signal strengths of monitoring relative to the Treat and Maintain actions, not shown here due to space constraints, revealed that the bimodal distribution of outcomes was still present as more information was gleaned from Treat and Maintain, though in less pronounced form. Similarly, changing the state transitions $P$ to allow for greater randomness in the underlying state variable yielded optimal policies that changed more often over time, leading to a more even distribution of cumulative rewards.

5. CONCLUSIONS

The uncertainty inherent in natural resource management requires that we think carefully about the allocation of scarce resources between monitoring and restoration. The POMDP provides a tool for analyzing the conditions under which more or less monitoring is desirable. This approach accounts for the costs (and benefits) of different actions, including different monitoring schemes, in a formulation that captures stochastic changes in state variables and errors in observations taken on these state variables.

In the application presented here, for a 10-period decision horizon, both of the monitoring schemes considered (more and less intensive) were part of the optimal policy, but only for about 21% of possible beliefs about the true state, while in the other 79% of the belief space the preferred actions were to treat without monitoring or to maintain the status quo (i.e., to do nothing). That is, in this example, for most beliefs the costs of monitoring exceed the expected benefit—monitoring was preferred only when the manager had a fairly strong a priori belief that the road was a low-erosion site, with monitoring serving essentially to rule out the need for more aggressive and expensive treatment.

The optimal policy computed for this problem was then applied to the stochastic environment (i.e., to random sequences of state changes and observations generated from the assumed model), which yielded a strikingly bimodal distribution of cumulative rewards over a 50-year time horizon. The optimal policy of a POMDP is by definition the one that maximizes the expected value of cumulative reward, and is not influenced by the variability of that reward. Many decision-makers may be uncomfortable with a policy that generates a very wide range of possible outcomes. Much work is currently being done in operations research to address risk-sensitive and robust approaches to control problems, but the canonical form of the POMDP as presented here is not risk-sensitive, though changing model parameters, such as increasing the designated costs of certain outcomes, can aid in the search for more conservative policies.

Our example here has been stylized both for ease of presentation and because POMDPs are known to be computationally intractable. Due to the larger (and possibly continuous) state space that will apply in many environmental management settings, model solution will have to draw on heuristic techniques for the POMDP, an active field of research in applied mathematics and computer science. The issues of state space aside and risk sensitivity aside, the POMDP as presented here has several important limitations. The reward structure, transition probabilities, and observation model are all assumed known and fixed. Establishing reasonable parameter sets, especially for the observation model, will be a significant challenge in many applied settings. Applying more general dynamic optimization techniques, such as Bayesian reinforcement learning, may provide a way to address some of these concerns, though allowing more general model formulations will generally come at the cost of requiring more data.

Lastly, a word on the subjective beliefs that form the basis for the model presented here is in order. These beliefs may come from personal experience, field experiments, studies from other areas, or other sources, and evolve over time as new information becomes available. Because the beliefs evolve from some prior, different agents looking at the same sequence of observations may still disagree about the ‘facts’ of a decision setting. Some parties may object that subjective beliefs are not a valid basis for environmental management decisions, but it’s hard to see what alternative we have, given that in most settings relatively little is known with certainty. Indeed, subjective probabilities are the de
facto basis of most management decisions, and we’re probably better off acknowledging this reality rather than ignoring it.

ACKNOWLEDGMENTS

Thanks to Brian Barrett of the California Department of Forestry and Fire Protection and Nikos Vlassis of the Technical University of Crete for helpful discussions, and to Teresa Ish of Kuulakai Consulting for work on an earlier application of the model to erosion control. This work was partially supported by funding from the Southwest Fisheries Science Center of the National Marine Fisheries Service.

REFERENCES


Assessing the Value of Environmental Observations in a Changing World: Nonstationarity, Complexity, and Hierarchical Dependencies

Patrick M. Reed*, Ruchit Shah*, and Joshua B. Kollat*
The Pennsylvania State University (preed@engr.psu.edu)

Abstract: This paper explores three related propositions for designing environmental observation systems: (1) Nonstationarity in environmental data series has the consequent impact of making observation network design itself a nonstationary, stochastic planning problem where the value of alternative observation strategies should be evaluated based on planners' evolving conception of Pareto efficiency given new knowledge, technologies, and policies over long time-scales. (2) Real-world budgetary constraints within observation network design problems yield resource allocation conflicts across space, time, and competing foci that are equivalent in form to the multiobjective d-dimensional knapsack problem (MO-dKP). Consequently, the Pareto efficiency of observation networks can only be determined approximately for non-trivial problem instances. (3) Multiobjective hierarchical Bayesian optimization provides a very promising tool for identifying observation alternatives that are approximately Pareto efficient while simultaneously providing insights into the emergent dependencies of our decisions (both science and management oriented) on critical observations.

Keywords: observation networks, many-objective analysis, Bayesian networks

1. INTRODUCTION

Our detection, prediction and management of critical environmental gradients is fundamentally dependent on our ability to design and manage observation networks. As noted by Reed et al. [2006], environmental change necessitates a shift from myopic, non-adaptive long-term observation strategies towards adaptive design frameworks that link our observations and predictions of evolving human—natural systems. Key to this challenge is properly posing and analyzing the question: what environmental observations are necessary to detect, predict, and manage the risks posed by environmental change? Although a more holistic assessment is justified, at present our national, regional, and local observation strategies are largely ad-hoc, non-adaptive, and generally disconnected from evolving water resources policy and management needs, a condition that has long been recognized [Davis et al., 1979; Langbein, 1979; Moss, 1979; United States Geological Survey, 1999].

Thirty years ago Marshall Moss [1979] eloquently acknowledged these challenges and framed the need for future observation network design strategies to use a “…more integrated measure of information…[that] results from a complex interaction of both the hydrologic knowledge and the procedures that are used to incorporate the knowledge into…decisions” (p. 1673). Moss’s recommendation represents a major departure from the more commonly employed statistical information measures [e.g., Shannon, 1948; Kiefer, 1959] by seeking to understand the value of observables for advancing knowledge while simultaneously characterizing their value to the procedures used to make decisions. This discussion paper draws on our recent research results to highlight three related challenges that must be considered when judging the value of observation systems as well as their gaps through their space, time, and management dimensions:
1. Nonstationarity in hydrologic systems has the consequent impact of making observation network design itself a nonstationary, stochastic planning problem where the value of alternative observation strategies should be evaluated based on planners’ evolving conception of Pareto efficiency given new knowledge, technologies, and policies over long time-scales.

2. Real-world budgetary constraints within observation network design problems yield resource allocation conflicts across space, time, and competing foci that are at least as challenging as the multiobjective d-dimensional knapsack problem (MO-dKP). Consequently, determining the Pareto efficiency of observation networks has a NP-Complete complexity (Nondeterministic Polynomial time-complete), which means that globally optimal tradeoffs cannot be attained with modern computers for non-trivial problem instances.

3. Multiobjective hierarchical Bayesian optimization represents a very promising tool for identifying observation alternatives that are approximately Pareto efficient while simultaneously providing insights into the emergent dependencies of our decisions (both science and management oriented) on critical observations.

The remainder of this paper explores each of these challenges in more detail.

2. TRADEOFFS & NONSTATIONARITY IN AN INTEGRATED MEASURE

It is very difficult to estimate the value of observations in terms of their impacts for improving our understanding of the evolving risks and environmental services of water resources systems. We propose that Pareto efficiency provides a mechanism for discovering and understanding information worth tradeoffs across the broad range of incommensurate objectives that could be of interest for observation network design problems (minimize costs, minimize risk, maximize coverage, minimize uncertainty, etc).

The performance of any potential monitoring alternative \( \mathbf{X}_k \) must be evaluated in a manner that considers performance across the total component objectives that relate to investments, prediction goals, and management needs. Feasible solutions to the problem are evaluated in terms of their nondomination. Assuming minimization of all objectives: an observation alternative \( \mathbf{X}_1 \) dominates \( \mathbf{X}_2 \) if its objectives’ values are less than or equal to those of \( \mathbf{X}_2 \) and there exists at least one objective where \( \mathbf{X}_1 \) attains a lower objective value. This mathematical partitioning rule then serves to identify the feasible space of sampling alternatives that do not have their performance exceeded in all objectives [Pareto, 1896]. Figure 1 provides a two-objective illustration of the concepts of nondomination and Pareto efficient fronts. In the figure, assuming minimization of both cost and error, the goal is to attain the minimum level error for each level of cost. The shaded boxes designate the objective space dominated by solutions 1, 2, and 3 in the figure. The full set of nondominated solutions as plotted represent the Pareto efficient frontier (i.e., the optimal tradeoff between cost and error).

The Cost—Error Pareto front is classified as an a posteriori decision analysis tool, which means that decision makers are provided with an explicit representation (see Figure 1) of their design tradeoffs when seeking compromises between conflicting objectives. Although the water resources literature has largely focused on two-objective formulations, there is a growing body of literature that is advancing a more generalized “many-objective” version of planning and design using problem formulations with 3 or more objectives [Ballinger, 1999; Reed and Minsker, 2004; Bekele and Nicklow, 2005; Fleming et al., 2005; di Pierro, 2006; Kollat and Reed, 2007; Kollat et al., 2008; Zhang et al., 2008; Kasprzyk et al., In-Press; Hadka and Reed, In-Review]. Kollat et al. [2007] use an observation network example to demonstrate how many-objective search and interactive visualization provide a new a posteriori decision aid for discovering tradeoffs, decision interactions, and design consequences across a range of objectives. These issues link strongly to Moss’s [1979] proposal for an integrated information measure that links advances in our scientific knowledge to “…the procedures that are used to incorporate the knowledge…” into
decision”. Assessing Pareto efficiency serves as a unifying decision analysis where decision makers can discover and visually explore tradeoffs in how investments in observations impact our knowledge and management objectives.

Figure 1. Illustration of the concepts of domination and Pareto efficiency assuming minimization of Cost and Error. Solutions 1, 2, and 3 dominate all of the solutions within their respective grey shaded regions in all objectives. The Pareto efficient frontier is the non-dominated solutions that attain the minimum level of Error at each Cost.

Using the concept of Pareto efficiency as an integrated information measure will require innovations in how we define many-objective network design problem formulations, advancements in promising new solution algorithms [e.g., see [Coello Coello et al., 2007]], and new decision analysis technologies incorporating interactive visual analytics [Russell et al., 1993; Keim et al., 2006; Kollat and Reed, 2007; Sanfilippo et al., 2009; Kasprzyk et al., In-Press]. Objectives and decisions are contextually driven by the needs within each management period and would be expected to evolve with new needs and problem insights. Broadly, this issue highlights that the mathematical spaces (or topologies) that define observation systems’ tradeoffs are in fact nonstationary and often highly uncertain. They are nonstationary in the sense that as designers make new discoveries, these discoveries feedback to new hypotheses which then motivate human-guided structural changes in mathematical formulations [e.g., see the de novo planning concepts of Zeleny, 2005]. Given the increasingly severe uncertainties, dependencies and decision tradeoffs for complex environmental systems, the many-objective Pareto efficiency information measure explicitly elucidates the consequences, compromises, and hypotheses that emerge with new information and knowledge.

3. COMPLEXITY OF MANY-OBJECTIVE INFORMATION MEASURE

Beyond the number of objective conflicts, sampling decisions also strongly influence the computational complexity of the environmental observation network design. The n-dimensional binary decisions $X_i \in \{0,1\}^n$ represents a lower bound in the complexity of the problem where yes/no decisions are made across space, time, and different environmental states (flow, water quality species, etc.). A linear increase in sampling decisions yields a $2^n$ exponential growth rate of the number of design alternatives. Other defensible formulations that include real-valued and/or integer decisions could have far more severe growth rates (e.g., factorial or potentially infinite). Our goal in analyzing the lower bound complexity of the problem is to demonstrate the strong computational challenge that environmental observation network design poses. When solving many-
objective formulations of network design problems, their overall problem difficulty will be
governed by maximally difficult sub-problem(s).

Consequently, we can build on known results from computational complexity theory to
show that the environmental observation network design problem is NP-complete
(Nondeterministic Polynomial time-complete). In brief, an NP-complete problem [for a
more formal discussion see Cook, 1971; Garey and Johnson, 1979] represents a severely
difficult problem that cannot be solved exactly using modern computers (i.e., deterministic
Turing machines) for instances that cannot be enumerated. Therefore non-trivial instances
can only be solved approximately and global optimality is not attainable for any algorithm.
The NP-complete computational complexity of network design can be surmised by
considering a simple accounting of cost objectives $j_{\text{cost}}^{f}(A_{k-1}^{f}, X_{k})$ used for the
constrained allocation of sampling investments between D system states. Equation (1)
provides a highly simplified D-objective constrained cost formulation which would be a
subset problem of a full formulation that could include prediction and/or management
objectives. Equation (1) is a simple representation of investment tradeoffs. Moreover,
equation (1) is identical to the multiobjective D-dimensional knapsack allocation problem
[Martello and Toth, 1990; Shah and Reed, In-Review].

$$\text{Min } j_{\text{cost}}^{f} = \left[ j_{\text{cost}}^{1}(A_{k-1}^{f}, X_{k}), j_{\text{cost}}^{2}(A_{k-1}^{f}, X_{k}), \ldots, j_{\text{cost}}^{D}(A_{k-1}^{f}, X_{k}) \right]$$

where

$$j_{\text{cost}}^{i}(A_{k-1}^{f}, X_{k}) = \sum_{j=1}^{n} p_{j} x_{j}, \quad \forall i \in [1, \ldots, D]$$

Subject to:

$$\left[ \sum_{j=1}^{n} p_{j} x_{j} \right]_{k} \leq c_{k}, \forall i \in [1, \ldots, D]$$

$$X_{k} \in \{0,1\}^{n}$$

Although more complex cost equations that incorporate the nonlinearities and complexities
that could be associated with a more economic-oriented formulation that accounts for the
time value of investment would be defensible, equation (1) provides arguably the simplest
meaningful accounting for cost as a simple linear summation of discrete costs $p_{j}$ for the $j^{th}$
sample of the $i^{th}$ state. Taken as a whole, the objectives and constraints of equation (1)
represents the subset sum instance of the knapsack where capacity constraints’ weights
equal items’ respective profit coefficients, see [Martello and Toth, 1990; Pisinger, 2005;
Shah and Reed, In-Review].

Consequently, given that the knapsack problem is a classic NP-complete problem, equation
(1) implies that environmental observation design is an NP-complete problem class.
Returning to our proposal of using many-objective Pareto efficiency as an integrated
information measure, equation (1) implies that robust computational tools are needed to
attain high quality approximations to the optimal tradeoffs. Moreover, alternative instances
of the knapsack problem can be vastly more difficult than others when seeking high quality
approximate solutions. So the immediate concern for environmental observation network
design is answering the question, how hard is our instance of the knapsack?

The historical theoretical work for the knapsack provides insights into the difficulty of the
observation networks problem class. Prior studies [Martello and Toth, 1990; Pisinger,
2005] have clearly shown that a high degree of correlation or interdependence between
the knapsack problem’s binary decisions and/or constraints often dramatically increases the
difficulty of finding high quality approximate solutions. These findings represent a severe
concern for environmental network design because observation decisions across space-and-time
are fundamentally linked and interdependent due to hydrologic systems’ socio-
physical organization. Mathematically the concept of hierarchy provides a useful means of
capturing how a particular decision to observe at the current location and time influences
the impacts of observations at other times and locations. In simple terms, hierarchy may be

1382
viewed as a series of probabilistic if-than-else observation rules across space-and-time. It remains an important challenge to discover and exploit these dependencies observation design frameworks.

4. DISCOVERING HIERARCHICAL DEPENDENCIES

The discovery of hierarchical if-than-else probabilistic rules for complex, adaptive systems has been a focus of the artificial intelligence field since its inception [Simon, 1968]. For environmental observation network problems, we add to this ambition the challenge posed by the NP-complete knapsack problem structure detailed in equation (1). While the socio-physical organization of environmental systems would be expected to engender interdependencies in observation decisions, computationally this knowledge is not present in traditional multiobjective optimization tools [for a detailed review see Coello Coello et al., 2007]. Recently, Pelikan and Goldberg [2003] introduced a new form of evolutionary optimization tool termed the Hierarchical Bayesian Optimization Algorithm (hBOA) to provide the capability to learn and exploit Bayesian network models of decision interdependencies while solving problems. The hBOA innovations are strongly relevant to the environmental observation network design problem’s limiting challenges. Pelikan [2002] demonstrates that the hBOA can attain sub-quadratic computational scaling for severely challenging hierarchically structured problems. In short, this means that the hBOA has reduced computational demands when solving increasingly larger problems with hierarchical dependencies. Moreover, Pelikan and Goldberg [2003] show that more traditional solution tools can have exponentially scaled computational complexities for hierarchically structured problems (i.e., they rapidly fail to attain high quality results unless they resort to enumeration).

At its original inception, the hBOA was a single objective probabilistic model building evolutionary algorithm [Pelikan, 2002; Pelikan et al., 2002]. Expanding the hBOA to address many-objective instances of equation (1), Kollat et al. [2008] introduced the Epsilon Dominance Hierarchical Bayesian Optimization Algorithm (ε-hBOA). The ε-hBOA represents a new type of multiobjective evolutionary algorithm where during evolution, the ε-hBOA selects high performing solutions and builds Bayesian network models of the underlying probabilistic dependencies of their decisions. After learning these dependencies, ε-hBOA uses them within its Bayesian network models to generate probabilistic hypotheses on what decision combinations would yield improved candidate solutions.

An important contribution of the ε-hBOA is that while solving many-objective problems, the algorithm is explicitly building a joint probabilistic density function (PDF) model of what makes observation decisions likely to be non-dominated with respect to many-objectives, simultaneously. The joint PDF provides the potential for discovering and visualizing the hierarchical dependency structure of environmental observation problems. Kollat et al. [2008] used a many-objective groundwater monitoring application to show that the ε-hBOA can dramatically enhance our approximate evaluation of Pareto efficiency for increasingly larger networks. Shah and Reed [In-Review] further showed that the ε-hBOA provides a robust approximation technique for many-objective knapsack instances with strong dependency structures.

Figure 2A provides a network visualization of some of the strongest hierarchical rules proposed by the ε-hBOA that are expected to be satisfied by greater than 98% of all of the non-dominated solutions for the test case. In reality, the algorithm provides a joint PDF model of rules for a full range of probabilities. In the ε-hBOA network illustration in Figure 2A, the numbered circles represent potential decisions for sampling one of the 25 available wells. Green designates wells that are sampled and red represents locations that are not sampled. In Figure 2A, the interior most circles represent the dependent variables in the rules subject to the decisions made in the outer circles. The order of hierarchy for each rule is defined by the number of independent decisions that influence the interior dependent sampling rules. For example, well 22 has a zeroth order hierarchy rule which indicates that it should be sampled by greater than 98-percent of non-dominated solutions independent of all other wells. Alternatively, the ε-hBOA proposes rules of up to the 7th order hierarchy...
for sampling well 2, which is striking given the simplicity of the test case. Note that well 2 has multiple rule instances which represent alternative independent decisions that would motivate that the location be sampled.

Figure 2. (A) The network graphic depicts examples of the hierarchical Bayesian rules proposed by the ε-hBOA when searching for Pareto efficient sampling strategies. The numbering and coloring corresponds to well identifications and sampling decisions (green—sampled, red—not sampled). The interior most green circles are the dependent sampling decisions. Excluding the interior dependent wells, the number of exterior wells in each rule defines its hierarchical order. All rules shown are proposed by the ε-hBOA to occur in greater than 98-percent of the solutions that compose the test case’s Pareto efficient front. (B) A spatial representation of the multi-point sampling wells in the test case’s domain with illustrations of three example rules. Figure 2B provides a spatial visualization of three 1st order hierarchy rules highlighted in Figure 2A. The ε-hBOA’s proposed rules are sources of potential hypotheses on system behavior when you place them within the natural space-time contexts of environmental systems. Figure 2B provides contextual meaning for some of the rules discussed in Figure 2A. For example, wells 1 and 11 jointly sample the defining boundaries of the plume. Similarly, if well 18 is not sampled along the longitudinal axis of the plume, then the ε-hBOA’s rule proposes that well 20 should be, which makes intuitive since it is the next closest well near the mid-line of the PCE plume. Likewise, well 15 is a suggested substitute for well 14. Broadly, the network graphics in Figure 2 provide a classification of how sensitive non-dominated sampling strategies are to a range of interdependent sampling decisions. The rules provided by the ε-hBOA provide decision makers with a means of discerning which sampling decisions have broad impacts over the full plume (e.g., well 1 in the PCE source area) versus those that have more localized effects (well 22 on the edge of the domain). There are two important issues to note in Figure 2. First, although some of the simpler rules may seem intellectually trivial, they represent problem knowledge that traditional solution tools are incapable of capturing, thus enabling the ε-hBOA to deal with severely interdependent problems (a form of severe nonlinearity termed epistasis). Secondly, the more complex high order hierarchical rules in Figure 2A pose interesting hypotheses on the relationships and controls impacting how alternative sampling strategies
can attain near optimal tradeoffs between the test case’s four objectives. Figures 2 supports our contentions that the general environmental observation network design problem class represents difficult knapsack problems with hierarchical dependency structures.

5. CONCLUSION

Environmental change motivates the consideration of the data intensive metrics for the risks, resilience, and adaptability of water resources systems. It is important to recognize that ad hoc and myopic observation management policies for environmental observation systems may pose substantive risks to our ability to manage environmental change over long-time scales. They lack an integrated view of long-term economic costs, impacts on scientific predictive skill, and long-term risk management consequences. Understanding of the value of information within the evolving network-of-networks that characterize national monitoring efforts requires a more direct assessment of their evolving tradeoffs and dependencies, scalable improvements in observation network design methods, and holistic assessments of water resources risks.

ACKNOWLEDGMENTS

The authors of this paper were partially supported by the United States National Science Foundation (NSF) Career grant CBET-0640443. Any opinions, findings, and conclusions or recommendations in this paper are those of the authors and do not reflect the views of the NSF.

REFERENCES

di Pierro, F. (2006), Many-objective evolutionary algorithms and applications to water resources engineering, University of Exeter, Exeter, UK.
Hadka, D., and P. M. Reed (In-Review), Diagnostic Assessment of Search Controls and Failure Modes in Many-Objective Evolutionary Optimization, IEEE Transactions on Evolutionary Computation.
Pareto, V. (1896), Cours D'Economie Politique, Rouge, Lausanne.
Efficiency Criteria for Water Quality Monitoring

Marina G. Erechtchoukova and Peter A. Khaiter

1York University, Toronto, Canada (marina/pkhaiter@yorku.ca)

Abstract: The issues of possible improvements, increased efficiency and/or optimization of a monitoring system in general, and a monitoring design in particular, are urgent. Since monitoring activities are always limited by financial and logistics constraints, algorithms of constrained optimization are deemed more suitable for this purpose. Monitoring designs are developed as solutions of an operation research model. In order to formulate such model the effectiveness function has been introduced. The effectiveness reflects the extent to which a monitoring design meets the objectives of the monitoring program and can be used for comparison of different monitoring designs. The effectiveness function depends on the investigated water quality parameters, selected indicators of water quality and their estimators. The function properties suggest the selection of an optimization algorithm. The proposed approach has been applied to a case study in order to develop temporal monitoring designs. It has been shown that the designs differ significantly only when the levels of the effectiveness are high. With the effectiveness of 80% or less the designs for different water quality parameters and the same indicator can be compromised. Since monitoring data are usually used for various purposes, the preference should be given to simple monitoring designs or to the designs which support efficient reconstruction of chemographs of investigated water quality parameters.

Keywords: Monitoring design; Water quality; Effectiveness; Constrained optimization.

1. INTRODUCTION

Monitoring data reveal important information about the current status of the aquatic environment and the anthropogenic impact water ecosystems are exposed to. Collected periodically at the same locations, these data can be used for projections and evaluation of consequences of management decisions. Monitoring systems have a complex infrastructure supporting all of their sampling and data processing activities. Monitoring systems comprise several components related to different aspects of their functioning. These components include the collection and analysis of physical, chemical, and biological data, as well as quality assurance and control programs which ensure that the data are scientifically valid. A key component of a monitoring system is its network of sampling sites where water quality observations and measurements are conducted. Monitoring activities are always limited by financial and logistics constraints. At the same time, the systems must provide data sufficient for a wide range of scientifically valid conclusions. Therefore, the issues of possible improvements, increased efficiency and/or optimization of a monitoring system in general, and a monitoring design, in particular, are urgent.

Multidisciplinary nature of a monitoring system explains different and sometimes contradicting aspects of the system which must be taken into account during its optimization. It justifies various approaches which are aimed to replace intuitive improvements of monitoring systems in order to increase their efficiency. Thus, the problem of monitoring optimization can be solved as a multi-objective mixed integer...
programming model with constraints [Ning and Chang 2002], or as a constrained optimization using generic algorithms [Icaga 2005; Cieniawski et al. 1995]. Fuzzy optimization approach is also used in optimization of a water quality monitoring network [Ning and Chang 2002].

The current study investigates the approaches to optimization of monitoring networks and, particularly, the development of temporal monitoring designs for water quality sampling. Application of formal optimization algorithms requires formulation of an objective function which reflects the goal of optimization and at the same time allows for quantitative comparison of different monitoring designs. Mathematical properties of the objective function restrict or even determine computational algorithms to be applied to derive monitoring designs which are at least satisficing. Hence, the articulation of an objective function affects the designs deemed to be efficient. Monitoring designs are considered improved when they either are more efficient than other known designs under the same, usually limiting, conditions or the designs have the minimal cost compared to all known alternatives with an acceptable level of efficiency. In both articulations, it is necessary to evaluate the efficiency of a monitoring design. The paper considers different criteria of efficiency of water quality temporal monitoring designs and their implications to suggested sampling programs. Although there are other important aspects in optimization of a monitoring program such as location of sampling site, analytical or sampling methods, they are not taken into account and the study is focused on sampling frequencies.

2. MONITORING FRAMEWORK

Groot and Schilperoort [1983] proposed a framework for developing an efficient monitoring network. The framework consists of five main steps: (1) to identify monitoring objectives; (2) to identify water quality indicators and relevant processes in order to choose approaches to data analysis; (3) to identify sampling programs and determine the effectiveness of a monitoring network; (4) to obtain cost estimates of the monitoring network; and (5) to implement cost-effectiveness analysis. This framework is very generic and can be applied to create an entire monitoring program or its part.

The development of a water quality monitoring system is significantly guided by monitoring objectives formulated for a waterbody. Whitfield [1988] specified the following five general objectives: (1) assessment of trends in variables (i.e. values of water quality parameters) of interest; (2) attainment of water quality standards; (3) estimation of mass discharge; (4) assessment of environmental impact; and (5) general surveillance. In addition to these objectives, monitoring data are also required for specific project and management needs. Each objective implies specific computational procedures employed to derive important information from a set of collected data.

The second step also relies on monitoring objectives since the latter determine important characteristics of selected water quality indicators required for decision making. For example, implementing the total maximum daily load process, it is important to know not only concentrations of water quality constituents and violations of water quality standards, but also the duration and magnitude of the violation [Shabman and Smith 2003]. Monitoring objectives dictate sampling programs. Thus, trend detection requires sampling of selected water quality indicators with a fixed frequency at the same location and at the reference site [Lettenmaier 1978]. Attainment of water quality standards can be investigated by sequential sample collection when the number of observations is determined based on the outcome of observations as they have been made [Whitfield 1988]. Estimations of mass discharge require sampling programs which take into account properties of selected estimators [Robertson & Roerish 1999; Erechtchoukova & Khaiter 2009].

The effectiveness of a monitoring program should quantitatively express the extent to which monitoring results meet the objectives. Although monitoring systems generate series of data, the main outcome of a monitoring program is information which supports decision making in accordance with the monitoring objectives. This information is usually obtained
via processing monitoring data using simple or complex models. Hence, the evaluation of the effectiveness of a monitoring program should include mathematical models transforming observation data into information. It justifies the role of models in developing efficient monitoring designs [Erechtchoukova and Khaiter 2008].

Since monitoring systems operate under financial constraints, cost estimates of a monitoring system are important. They allow for straightforward comparison with available budget and between different monitoring programs. Although Groot and Schilperoort [1983] considered the process of deriving the financial estimates simply as a technical exercise, absolute values of the estimates or their components are not always available. At the same time, the cost of a monitoring program definitely depends on the number of collected samples.

If the effectiveness of a monitoring program could also be expressed in monetary form, the cost-effectiveness analysis can be done based on a direct comparison. However, such estimates are usually not available. In this case, the comparison can be replaced by constrained optimization techniques which do not require expressing a goal function and constraints in the same measuring units, although both the goal function and constraints must have common variables.

3. OPERATION RESEARCH MODEL

The constrained optimization approach can be applied to the development of an efficient monitoring network. For this purpose, it is necessary to define the goal of optimization. With respect to the framework mentioned above there are two possible articulations of the problem of the development of efficient monitoring design: (1) to maximize the effectiveness of the design under limited budget and (2) to minimize the cost of the design within an acceptable level of effectiveness. In both cases it is necessary to provide quantitative estimates of the cost and the effectiveness of a monitoring design.

Although the cost of a monitoring network comprises various components which may not be independent, it is reasonable to assume that the cost of a monitoring network increases monotonically with the number of samples collected at the monitoring sites and the number of sites. This assumption validates the replacements of a cost estimate by the total number of design samples in an operation research model for the development of a monitoring design. This assumption also leads to articulation (2) of the problem of an efficient water quality monitoring design, since articulation (1) requires explicit estimates of the cost of a monitoring design.

Monitoring objectives impose additional requirements on data sets used for data analysis. Thus, trend detection requires data collected with a fixed sampling frequency at the same observation sites for a long period of time. Attainment of water quality standards can be checked using different schemes including fixed frequency, sequential or Markov sampling. For mass transport estimation, simple random sampling or stratified random sampling are preferable. The designs supporting environmental assessment are very project specific. They must be compliant with the type of analysis employed for estimation of the effects of an investigated project. In all the cases, the designs are supposed to provide reliable estimates of selected environmental indicators.

The number of collected samples with additional requirements about sample distribution determines the extent, to which derived data reflect water quality conditions and hence affects the effectiveness of a network. Formalizing the effectiveness as a function of the number of observations ties together cost and effectiveness of the network. Then, the problem can be described by the following operation research model:

\[
\min n \quad \text{subject to} \quad \begin{align*}
E(n) & \geq V \\
(1) \\
(2)
\end{align*}
\]
where \( n \) is the number of required observations in a monitoring design, \( E(n) \) is the effectiveness of the design, and \( V \) is the acceptable level of the design effectiveness.

Following the theory of designs of experiments [Fisher 1971], information derived from the observations is a reciprocal of the variance of an estimator which in its turn depends on the number of observations used in estimation. This implies that information increases infinitely when accurate estimates are obtained. Such representation may restrict the set of automated procedures which can be used to find a solution. Since observations of many water quality parameters cannot be done continuously, daily sample collection is considered as a design providing the best possible estimates with zero error. Under this assumption, the effectiveness can be expressed via variance of an estimator used in the assessment in the following way:

\[
E(n) = (1 - D(I(n)) / \overline{I(n)}) \cdot 100\%,
\]  

(3)

where \( I \) is the selected estimator, \( \overline{I(n)} \) is its estimate on a set of \( n \) observations and \( D(I) \) is its variance. The analytical expression of \( D(I) \) depends on the selected estimator \( I \) which can be a regular statistics or a simulation model describing the selected water quality indicator.

It is worth noting that the effectiveness of the design may depend not only on the total number of observations, but also on the temporal and/or spatial distribution of sampling points. Only temporal monitoring designs for data collection at a given site will be considered further. An efficient distribution of samples over an investigated period depends on an indicator of water quality, seasonal variations in values of water quality parameters and the type of the function chosen to describe the effectiveness of the designs.

While formula (3) is based on the variance of the selected estimator, another approach appraises the effectiveness of a monitoring design by evaluating the extent to which the designs allow to reproduce water quality indicator’s dynamics for a period of observations [Erechtchoukova et al. 2009]. Monitoring data are used to interpolate missing values of water quality parameters and after that an estimate of an investigated water quality indicator is calculated. The relative difference between the estimate and actual value of the indicator evaluates the uncertainty of the estimate. The latter can be used to measure the effectiveness of the design:

\[
E(n) = \left(1 - \frac{I_c - I_a}{I_a}\right) \cdot 100\%,
\]  

(4)

where \( I_c \) is the indicator estimate calculated using interpolation, and \( I_a \) is the actual value of the selected indicator.

According to formulae (3) and (4), the effectiveness can range from -\( \infty \) to 100%. The upper boundary is achieved when the design includes all possible observation points. In order to take into account particular formulae for the indicator \( I \), operation research models (1),(2) and (3) or (1),(2) and (4) must be transformed further. Formulae determine to a great extent computational algorithms used to obtain a solution. Quantification of the efficiency based on formula (3) can result in an analytical expression for calculation of the required number of observations. Then the monitoring design contains sampling points randomly distributed over the segments of the investigated period. However, search techniques may give a set of certain points when the data must be collected in order to achieve the desired level of the effectiveness. Since data collection at specified dates is hardly possible, the formula (4) has been modified to allow for data collection during a certain interval or ‘sliding window’ around proposed dates [Erechtchoukova et al. 2009].
\[ E(n) = \left(1 - \max_{sw} \left( \frac{|I_c - I_a|}{I_a} \right) \right) \cdot 100\% \], \quad (5)

where \(sw\) is the size of the sliding window (i.e., the number of days) around a proposed date when a sample can be collected. Formulae (3) – (5) do not exhaust all possible representations of \(E(n)\). They have been chosen from perhaps the infinity of alternatives due to their simplicity and mathematical properties allowing for application of automated procedures. The criterion of effectiveness (3) has been investigated with respect to the basic statistics of individual water quality parameters: chloride ions, hydrocarbonate ions and total dissolved solids observed at the same sampling site. The results of investigation have been compared with the criterion (5) studied in [Erechtchoukova et al. 2009].

4. CASE STUDY

Given general monitoring objectives, the average annual concentration and the total annual chemical load have been chosen as water quality indicators of interest. Both indicators were evaluated using the population mean and stratified mean estimators. The series of values of instantaneous chemical load were calculated as the multiplication of values of concentrations and corresponding water discharge. In order to verify developed monitoring designs, very detailed monitoring data are required. It was assumed that daily values of concentrations of the selected water quality parameters and water discharges at the selected site would be sufficient to obtain accurate estimates. The data were collected at the cross-section of the Vyatka River near the town of Vyatskiye Polyany. Two years have been selected for calculation: Year_1 with the unimodal type of hydrograph (only spring-summer high flow events) and another one Year_2 with the bimodal type of hydrograph (high flow events took place twice in spring-summer and late fall). Both hydrographs exhibit distinct hydrological seasons with sharp rising and falling limbs for spring-summer high flow events [Erechtchoukova and Khaiter 2007]. The average annual water discharge is estimated about 22.6 km\(^3\). Chemographs of the selected water quality parameters in Year_2 are presented in Figure. 1.

![Figure 1](image_url)

**Figure 1.** Investigated chemographs of the selected water quality parameters.

The observed concentrations in different samples varied from 0.4 to almost 9.0 mg/l for chloride ions, from 47.0 to 231.0 mg/l for hydrocarbonate ions and from 88 to 381 mg/l for total dissolved solids. The last two water quality parameters exhibited similar variability in both years (the coefficients of variation were about 0.25 in Year 1 and 0.34 in Year 2), while the coefficient of variation for chloride ions had higher values (0.34 in Year 1 and 0.51 in Year 2). Time series for all investigated parameters demonstrated correlation with water discharge values indicating that the dilution process affects concentrations of the
investigated constituents to a great extent. The correlation coefficient was higher for total dissolved solids and hydrocarbonate ions (about -0.8) relatively to chloride ions (about -0.6). The correlation coefficients varied from year to year. This means that hydrological conditions change the contributions of individual processes underlying the formation of the constituent concentrations in the water column. The strong correlation between the series of concentrations of total dissolved solids and hydrocarbonate ions showed that the latter represent a significant portion of the total dissolved solids.

Monitoring designs have been developed for evaluation of the annual average concentration and the total annual chemical load of selected water quality parameters with different levels of effectiveness using basic and stratified estimators. The summary of non-stratified designs is presented in Table 1. In this case, samples are randomly distributed within investigated period of time.

### Table 1. Simple random designs for estimation of the annual average concentration (C) and the total annual load (L).

<table>
<thead>
<tr>
<th>Year</th>
<th>Acceptable level of the design effectiveness ((V)), %</th>
<th>The total number of required observation (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Cl</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
</tr>
<tr>
<td>Year_1</td>
<td>95</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>85</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>7</td>
</tr>
<tr>
<td>Year_2</td>
<td>95</td>
<td>169</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>85</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>12</td>
</tr>
</tbody>
</table>

Seasonal variations in water discharges and concentrations of the investigated water quality parameters suggested the application of stratified estimates. As it has been shown previously [Erechtchoukova and Khaiter 2007; Erechtchoukova and Khaiter 2009], stratified estimators require a lesser number of observations to achieve the same effectiveness of a monitoring design. Table 2 presents the summary of stratified designs for the Year_2. The temporal stratification has been implemented according to different criteria in order to reduce the variance of the population means. Thus, for a single water quality parameter and its two indicators, the three factors can be considered: the water discharge, concentrations, and instantaneous loads.

### Table 2. Stratified random designs for evaluation of the annual average concentration (C) and the total annual load (L) in Year_2.

<table>
<thead>
<tr>
<th>Stratification criterion</th>
<th>Acceptable level of the design effectiveness ((V)), %</th>
<th>The total number of required observation (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Cl</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
</tr>
<tr>
<td>Water discharge</td>
<td>95</td>
<td>125</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>85</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>21</td>
</tr>
<tr>
<td>Concentration</td>
<td>95</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>85</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>12</td>
</tr>
<tr>
<td>Instantaneous load</td>
<td>95</td>
<td>245</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>135</td>
</tr>
<tr>
<td></td>
<td>85</td>
<td>84</td>
</tr>
</tbody>
</table>
The temporal stratification was implemented manually and can be considered only as a “good enough” solution. The series of concentrations, instantaneous loads and water discharges have been divided into four strata according to the values of the selected criteria. The exact strata boundaries affect the designs to a great extent. The identified strata are both constituent and indicator dependent. However, they can be used to investigate criteria of effectiveness (3) and (5).

5. DISCUSSION

Monitoring designs presented in Tables 1 and 2 have been developed using model (1), (2), and (3). These designs are sufficient to achieve established levels of effectiveness by providing accurate estimates of the selected water quality indicators. They vary significantly with respect to the total number of observations and required observations per each stratum. From statistical perspective, the deviations can be explained by different variability of the investigated water quality indicators. Such variability depends on water quality parameters and mathematical properties of the selected estimators. As it was expected, the largest numbers of observations are required for estimation of chloride ions indicators, since it is the most varying water quality parameter in the study. Surprisingly, efficient estimation of the total annual load of chloride ion in Year 2 requires a lesser number of observations than those needed for the annual average concentration. It can be explained by the fact that the water flow diminished variations in chloride ion concentrations. In general, stratified estimators are more efficient for this type of hydrological and hydrochemical conditions. However, an efficient stratification must satisfy certain conditions, and strata boundaries are constituent, indicator and site specific [Erechtchoukova and Khaiter 2007]. It is worth noting that temporal stratified designs create additional obstacles in data collection since they require switching the frequencies of sample collection over an investigated period. The stratum boundaries must be determined a priori, whereas dates of typical hydrological and hydrochemical events vary from year to year making the stratification very inaccurate and the corresponding recommendations hard to follow due to a human factor.

For chloride ions, the designs developed from the operation research model (1), (2), and (3) have been compared with the designs derived from the model (1), (2), and (5) with the size of the sliding window of 31 days which means that each observation can be implemented within a month of a suggested date [Erechtchoukova et al. 2009].
criterion (5) generated the designs with significantly lower numbers of required observations from 22 (for the effectiveness of 95%) to 6 (for the effectiveness of 80%) in Year_1 and from 16 (for the effectiveness of 95%) to 6 (for the effectiveness of 80%) in Year_2. The high efficiency of these designs has been achieved by selecting the points in time which support the interpolation of the main sections of the chemograph very well. Thus, the way the designs have been derived makes them site and constituent specific. The comparison of the designs developed using formulae (3) and (5) is shown in Figure 2.

Concentrations of different water quality parameters have been determined from the same set of water samples. That is why suggested monitoring designs must be common for all investigated water quality parameters. For non-stratified estimators, only the highest number of observations from the three identified ones supports the established level of effectiveness for all investigated water constituents. For stratified estimators, although this number can be lower, it is necessary to take into account the differences in temporal strata boundaries for the investigated parameters. Only stratification based on the water discharge values is common for all three water quality parameters. As it shown in Table 2, such stratification is not the best for all investigated water constituents.

The total number of required observations significantly decreases with the lower levels of effectiveness of the designs for all investigated water quality parameters and years. Thus, the numbers of observations suggested by the corresponding designs with the level of effectiveness of 80% are very close.

![Figure 3](image-url)  
**Figure 3.** The total number of the required observations for estimation of chemical load of the investigated water quality parameters versus effectiveness

6. **CONCLUSIONS**

Application of formal procedures to the development of efficient monitoring designs requires quantitative assessment of the design effectiveness. The paper presented an attempt to express the effectiveness as a function of the number of required observations in order to provide necessary information which can be derived from the collected data. This information is obtained via application of statistical procedures or mathematical models which also affect the effectiveness of monitoring programs. The proposed approach implies an articulation of the problem as an operation research model. The efficient designs can be obtained as solutions of this model.
The presented case study demonstrated the dependency of the solutions of the operation research model on selected water quality parameters, indicators, models used for estimates and observation sites. However, the designs differ significantly only when the levels of the effectiveness of monitoring designs are high. With the effectiveness of 80% or less the designs for different water quality parameters and the same indicator can be compromised.

Since monitoring data are usually used for various purposes, some of which are even not known at the time of data collection, the preference should be given to simple monitoring designs or to the designs which support an efficient reconstruction of chemographs of investigated water quality parameters.

ACKNOWLEDGEMENTS

The authors are thankful to anonymous reviewers for their thoughtful suggestions and valuable comments on the manuscript. The study was conducted based on the data sets prepared in Hydrochemical Institute, the Russian Federation.

REFERENCES


Pitimbu River Lowland Portion Water and Sediment Monitoring Data, Natal Brazil

Rafael Takeshi Kobayashi
Federal Institute of Technologic Education - RN
Lúcio Flávio Ferreira Moreira
Rio Grande do Norte Federal University – UFRN (lucio@ct.ufrn.br)
Herbert Tadeu de Almeida Andrade
Rio Grande do Norte Federal University – UFRN (herbet@ufrnet.br)

Abstract: Urbanization in developing countries has had a significant impact on water resources and the environment. Anthropogenic-related factors such as deforestation, uncontrolled urban occupation and the lack of sustainable measures have affected both the quantity and quality of water. The Pitimbu river watershed (126.7 km²) is located within the urban area of Natal, Brazil. The effects of urban occupation have recently been a concern to municipal authorities, since the Pitimbu is the main source of domestic water supply. In this paper authors analyze monitoring data, including the quantitative and qualitative water and sediment factors at the lowland portion of this river system. To this end, two transverse cross sections were established, 5.6 km apart. The monitoring survey included measurement of water discharge and biological sediment analysis over a period of 11 months (Nov 2007 to Oct 2008). Water discharge varied from 0.62 m³s⁻¹ (dry season) to 10.61 m³s⁻¹ (rainy season). Discharge rates were used to fit De puit-Boussinesq baseflow recession equation parameters. The recession constant (0.8) revealed strong river aquifer interaction. Such interaction explains the increase in discharge between the two sections by a factor of 1.98. Increasing sediment input to the river channel has occurred during high magnitude events, resulting in channel sedimentation. Biological sediment analysis revealed both the absence of sensitive benthic species (Corduliidae) and abundance of resistant benthic species (M. Tuberculata), indicating the occurrence of heavy metal contamination.

Keywords: Pitimbu; water quality; monitoring; benthic

1. INTRODUCTION

In Brazil, the growth of urban areas has produced the human settlement of environmentally protected areas, causing impacts on water resources and the environment. In this context, social, commercial and institutional factors interact jointly to aggravate the problem, transforming it into a priority from the sustainability viewpoint. In contrast, the increasing water demands in urban zones highlight the importance of adopting policies in order to control occupation in the basin area.

In the metropolitan region of Natal, the hydrographic basin of the Pitimbu River (BHRP) is strategically important. Water captured directly in the river channel supplies around 30% of the population, with a flow rate 0.722 m³s⁻¹. On the other hand, the pattern of basin area occupancy has increased potential risks of ecosystem and water quality degradation. Since the 1970s the local water company has used an extensive network of deep wells for local supply. In recent years, the increase in nitrate levels in ground water has caused an increase in demand for Pitimbu River surface water, which is used for improving drinking water quality.
Earlier studies associated the process of land occupancy in the BHRP to a series of negative effects in the water and in the ecosystem, caused mainly by the precariousness of the sewage system and the increase in occupancy density [Borges, 2002; Santos et al., 2002; Araújo et al., 2007; Ferreira and Silva, 2009].

The BHRP has a drainage area of 126.76 km² (Figure 1), spanning three municipalities in the metropolitan region of Natal: Macaíba (47%), Parnamirim (43%) and Natal (10%). In recent years, the environmental challenge and signs of degradation have sensitized sectors of society to the urgency and seriousness of a problem that affects the entire population.

This study has the following aims: a) to assess the quality of the aquatic ecosystem using biomonitoring; b) to adjust the parameters of the Deput-Boussinesq equation using water discharge data during the dry season. This equation is used to represent recession water discharge behavior [Hall, 1968]; c) to assess recharge characteristics along a lowland stretch in the urban zone. To this end, two transverse cross sections were established: section 1(BR-101) and section 2(EMPARN), extending 5.69 km. Monitoring took place in both the dry and rainy seasons, over a period of 330 days (November 2007 to October 2008).

2. STUDY AREA

The BHRP lies over sedimentary rocks that form the Barreiras group. The region also contains movable dunes, paleodunes and Quaternary sand deposits. Hydrologic soil properties show their high infiltration capacity. Rainwater infiltrates rapidly into the soil, promoting storage and underwater flow through the non-confined aquifer.

In 2002 the BHRP contained areas of native vegetation, located in preserved military zones. In 2009, most of the basin is covered with anthropized vegetation and to a lesser extent areas of agriculture and urban occupancy. However, the occupied areas occur near the protected zone of the river, causing degradation of both the ecosystem and water quality. Furthermore, the area of the basin not yet influenced by anthropogenic activity (13%) shows the intensity of land occupancy and use [Borges, 2002; Santos et al., 2002].

In the process of BHRP occupancy, degradation is associated to diffuse pollution from different origins: a) domestic and industrial wastes; b) residential septic tanks; c) toxic substances (pesticides) used in agriculture; d) human settlement areas reduce infiltration...

Figure 1. Pitimbu river watershed. Source: LANDSAT on 11-10-2007.
Transverse Cross Sections: • 1 (BR101) • 2 (EMPARN)
and increase surface runoff, generating erosive processes and silting in the river channel. During maximum water level events the transport of organic substances from lateral banks was observed, associated to settlement areas near the main channel. Indeed, the advancing urbanization toward natural floodplains has caused damage to the ecosystem, thereby compromising water quality. In this sense, the failure to implement an effective program to control the occupancy of these areas has put the sustainability of the ecosystem and water quality at risk, which may soon preclude the latter’s use for human consumption.

2.1 Biomonitoring

Biomonitoring can be defined as the systematic use of organism responses aimed at assessing environmental changes, generally caused by anthropic action [Buss et al, 2003]. The results obtained in biomonitoring reflect the wide array of situations that water bodies are subjected to, given that biodiversity and faunal characteristics are the result of pressures exerted on the ecosystem.

Benthic macroinvertebrates, used in the formulation of biotic indices, are considered good indicators of environmental pollution in lotic systems. It is believed that these organisms respond to hydraulic, organic and toxic stresses, leading to a reduction in sensitive species and proliferation of resistant and tolerant species. Their reaction capacity makes them good bioindicators, given that they are differentially sensitive to pollutants, responding gradually to a wide spectrum of stress levels. Moreover, they are abundant and relatively easy to collect and identify, and have a long enough life span to exhibit environmental quality [Silveira, 2004]. Some kinds of benthic macroinvertebrates are shown in Figure 2.

Silveira [2004] underscored the influence of alluvial vegetation on the life of benthic macroinvertebrates. He observed that it interfered in the ecology of aquatic environments, providing shelter, protection, nutrients and organic matter.

Figure 2. Some kinds of benthic macroinvertebrates. a) Tanypodinae; b) Chironominae; c) M. Tuberculata.

3. MATERIALS AND METHODS

Campaigns were conducted to measure water discharge and collect bottom sediment samples in two cross sections located in the lowland section of the Pitimbu River over 330 days, for a total of 11 monthly campaigns. The monitoring period (13/11/07 to 03/10/08) was marked by the occurrence of maximum events, spanned both the dry season of 2007 and the rainy season of 2008. The measurement and sample collection campaigns were followed by sediment analyses, conducted at the entomology laboratory of UFRN.

The measurement of water discharge was carried out using a hydrometric micropropeller at equally spaced vertical elevations in the transverse cross section. Fluvial sediment samples were collected by using a Van Veen dredge (10 kg weight, 2.0 l sample). The methodology used for sediment sample collection considered the wet perimeter of the transverse cross section [Carvalho et al., 2000]. After quartering, the samples were washed in a 250µm sieve and stored in 80% (v/v) alcohol/ethanol solution. In biological analysis, the samples were sorted and the organisms were manually captured with the help of a loupe. The organisms were separated into tubes, identified and counted. Organism identification was based on morphological characteristics and multiple taxonomic keys [Edmunds et al., 1963; Mccafferty & Provonsha, 1994; Niesser & Melo, 1997]. Organism identification and
counting enabled characterization of the environmental quality of organisms as a function of the sensitivity level of the existing biodiversity. To determine the degree of organism sensitivity we used the tolerance values of studies conducted by Alba-Tercedor [1996], Cota et al. [2002] and Figueroa et al. [2003].

4. RESULTS AND DISCUSSION

4.1 Hydrologic Regime

The records of daily rainfall observed in the study period are shown on the graph in Figure 2. The rainy season began on 12/03/2008 and ended on 30/08. Total rainfall for the period was 2495 mm, a value above the historical mean. The events observed on 08/06 (271 mm) and 01/07 (260 mm) were considered exceptional. A decrease in flows was observed in the dry season, despite the occurrence of randomly spaced events with the water level at less than 20 mm. The rainy season accounted for the increase in flow rate to a maximum of 10.6 m³/s, corresponding to 17 times the flow rate at the end of the recession period (Figure 3). This behavior reflects the effect produced by sudden filling, with a significant surface contribution from the impermeable areas of the basin. Depletion behavior of the hydrograph at the end of the rainy season shows the water retention capacity of the soil.

The hydrograph recession phase reflects the behavior of water storage and aquifer transmissivity in the basin. Although recession is a highly variable process in both time and space, it is possible to adjust a regional expression of groundwater flow. Recession reflects the critical state of the water supply. The recession process is influenced by geological factors in addition to land occupancy and use in the basin.

According to Hall [1968], the recession curve allows the adjustment to a linear exponential function using the Deupit-Boussinesq equation,

\[ Q_t = Q_0 e^{-\alpha t} \]  

where \( Q_t \) is the flow at time \( t \), \( Q_0 \) is the flow at the onset of the recession phase, \( e^{\alpha} \) (equal to \( k \)) is a recession constant used as an indicator of the influence of base runoff in the basin. Residence time (three days) corresponds to the mean duration of base runoff. The methodology used in the analysis of the recession hydrograph consisted of a linear adjustment of flows on a logarithmic scale. The values of parameters \( Q_0 \) and \( \alpha \) are presented in Table 1.

The values obtained suggest high recession rates, with a strong effect of base runoff. The residence times obtained in the two sections were 7.7 and 8.0 days, respectively.
Table 1. Pitimbu river recession parameters at the monitoring sections.

<table>
<thead>
<tr>
<th>Section</th>
<th>Qo (m³/s)</th>
<th>α</th>
<th>k</th>
<th>Tres (days)</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.623</td>
<td>0.13</td>
<td>0.879</td>
<td>7.72</td>
<td>0.87</td>
</tr>
<tr>
<td>2</td>
<td>1.24</td>
<td>0.125</td>
<td>0.883</td>
<td>8.01</td>
<td>0.91</td>
</tr>
</tbody>
</table>

Figure 3. Flow behavior during the study period.

Water discharge data revealed an increase between the sections of 1.982. The increased flow reflects the effect of base runoff (river-aquifer) in the stretch of river studied.

Biomonitoring analyzed the presence of benthic macroinvertebrates in the fluvial bed sediment from the transverse cross sections of the Pitimbu River. Sediment samples varied between 0.4-0.45 kg. Organism identification in the laboratory showed a diversity of benthic macroinvertebrates composed of nine taxa.

In section 1, located upstream from the bridge over highway BR-101, a predominance of resistant groups Ortocladiinae and Chironominae was observed, suggesting a high level of ecosystem degradation. The absence of sensitive organisms, along with low biodiversity, confirms compromised quality. Table 2 shows the groups of benthic organisms observed in section 1, grouped by taxonomic family and by degree of sensitivity to organic pollution.

Table 2. Benthic organisms at section 1 at the 2007-2008 dry season.

<table>
<thead>
<tr>
<th>ORGANISMS</th>
<th>TOTAL</th>
<th>Sensitive</th>
<th>Tolerant</th>
<th>Resistant</th>
</tr>
</thead>
<tbody>
<tr>
<td>M. Tuberculata</td>
<td>1</td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orthocladiinae</td>
<td>24</td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chironominae</td>
<td>31</td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polycentropodinae</td>
<td>7</td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tanypodinae</td>
<td>7</td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>0</td>
<td>14</td>
<td>56</td>
<td></td>
</tr>
</tbody>
</table>

In section 2, groups of organisms were observed from the three levels of sensitivity to organic pollution. The dominant group, M. Tuberculata, reflects the damage to ecosystem quality. Table 3 shows the benthic organisms observed in this section, by taxonomic family and sensitivity level.

Table 3. Benthic organisms at section 2 at the 2007-2008 dry season.
The organic sensitivity of each benthic invertebrate group shows that greater damage occurred to ecosystem quality in section 1. In this section, the absence of the sensitive organisms Cordulidae and Helicopsychidae, observed in section 2, signal the degraded quality of the aquatic ecosystem.

5. CONCLUSIONS

The results obtained in the present study allow us to conclude the following:

- Observed water discharge during the dry season enabled the adjustment of Deupiter-Boussinesq parameters. These parameters show relatively high recession rates, residence time of 8 days and strong non-confined river-aquifer interaction. Nevertheless, the BHRP undergoes uncontrolled urban settlement, which tends to impact water quality in the medium-term;
- Biomonitoring results reveal signs of aquatic ecosystem degradation by heavy metals contamination (Copper and Cobalt), evidenced by the lack of sensitive benthic organisms and abundance of resistant organisms;
- The current condition of the aquatic ecosystem underscores the importance of policies to control and protect the basin; otherwise the level of degradation may soon reach levels that will make the water unfit for human consumption.

ACKNOWLEDGEMENTS

This study received financial support from CNPq (Brazilian National Research Council), Ministry of Science and Technology.

REFERENCES

Alba-Tercedor, J., Macroinvertebrados acuáticos y calidad de las aguas de los ríos, paper presented at IV Simposio del Agua en Andalucía (SIAGA), Almería, Spain, [1996].
Cota, L., Goulart, M., Moreno, P., and Callisto, M., Rapid assessment of river water quality using an adapted BMWP index: a practical tool to evaluate ecosystem health, [2002].
Edmunds, G.F., Allen, R.K., and Peters, W.L., An annotated key to the nymphs of the families and subfamilies of mayflies (Ephemeroptera), *University of Utah Biological Series, XIII (1)*, [1963].


Remote Sensing Time Series for Modeling Invasive Species Distribution: A Case Study of *Tamarix* spp. in the US and Mexico

Anna Cord\(^{a,b}\), Doris Klein\(^{b}\), Stefan Dech\(^{a,b}\)

\(^{a}\) German Remote Sensing Data Center, German Aerospace Center (DLR-DFD), Münchner Str. 20, 82234 Wessling, Germany (anna.cord@dlr.de, stefan.dech@dlr.de)

\(^{b}\) Department of Remote Sensing, Institute of Geography, University of Wuerzburg, Am Hubland, 97074 Wuerzburg, Germany (anna.cord@uni-wuerzburg.de, doris.klein@uni-wuerzburg.de, stefan.dech@uni-wuerzburg.de)

Abstract: Detecting invasive species and predicting their potential distribution are crucial to coordinate management responses. Remote sensing data are now available in several spatial and temporal resolutions and can supply environmental models with additional information. This study uses the Maximum Entropy algorithm to model the current distribution of the saltcedar (*Tamarix* spp.) in the US and Mexico and to identify suitable habitats, both already inhabited and not yet occupied. Tamarisk is restricted to specific habitats such as riparian zones, wetlands and agricultural or disturbed areas, which are typically not only characterized by climate. To describe vegetation phenology and thermal seasonality in these habitats, the study uses annual metrics of remotely sensed time series from 2001 to 2008 (Terra-MODIS Enhanced Vegetation Index and Land Surface Temperature) together with WorldClim bioclimatic data. By using occurrence records primarily from the US we were able to model predictive maps of tamarisk distribution correlating very well to the known distribution in the US. For Mexico, where only very few occurrence records exist, we identified potential tamarisk habitats for substantial areas in Baja California, in the states of Sonora and Sinaloa and in the Central Mexican Plateau. These predictive model results can be used to support the early detection and prevention of *Tamarix* spp. invasion.

Keywords: Maximum Entropy; MODIS Time Series; Enhanced Vegetation Index; Land Surface Temperature; Tamarisk.

Remark on the nomenclature of this study: *Tamarix* spp. refers to all species within the genus *Tamarix* L., primarily *T. ramosissima* and *T. chinensis* and hybrids.

1. INTRODUCTION

On a global scale, invasive species belong to the main reasons of biodiversity loss [Ficetola et al. 2007] and cause substantial economic damage. A large number of studies are therefore investigating species invasions with the aim to identify suitable habitats and to coordinate management plans. The conceptual findings and spatio-temporal predictions of these studies contribute to the early detection and elimination of invasive species to reduce their negative impacts. The saltcedar (*Tamarix* spp.) is native to Southern Europe, Asia Minor, Mongolia, Tibet, central China, and North Korea [Carpenter 2003] and has the growth habit of a shrub or small tree. In the 1800s, it was introduced in the United States as ornamental plant and as windbreak for erosion control [Pearce and Smith 2003]. Tamarisk is a facultative phreatophyte, well-adapted to arid environments and – due to its extensive
root system – better able to tolerate drought, water stress, and salinity compared to native mesic species [Glenn and Nagler 2005; Kerns et al. 2009]. Since its introduction, *Tamarix* spp. has become the (sub)dominant woody species along many perennial river systems in the arid south-western United States and north-western Mexico and has increasingly been replacing native vegetation [Glenn and Nagler 2005]. By the 1920s, tamarisk was first recognized as a problem, having several detrimental effects such as decreased stream flow, loss of native biodiversity, changes in food web structure and soil salinization [DiThomaso et al. 1998]. The decline of many native riparian phreatophytes connected with *Tamarix* spp. expansion has been related to altered river flow regimes due to water management programs and the use of ground water for urban, agricultural and industrial purposes [Shafroth et al. 1998]. Today, the tamarisk is considered among the “World's Worst Invasive Alien Species” [Lowe 2000] and one of the most aggressively invasive exotic plants in the US and Mexico. Once the species is dominant at a site, control and restoration efforts are extremely costly in terms of labor, time and money. Zavaleta [2000] calculated the annual costs resulting from tamarisk invasion to be between US$ 280 and 450/ha with almost one million ha already covered in the western United States alone [Pearce and Smith 2003]. Due to its high invasion potential, several previous studies aimed at mapping tamarisk distribution on small scales – often using high spatial resolution [Ge et al. 2006] or hyperspectral data [Hamada et al. 2007]. In addition, previous work [Everitt and DeLoach 1990] showing the specific temporal-spectral signature of *Tamarix* spp. related to vegetation phenology suggests the high potential of remotely sensed time series data in this context. Morisette et al. [2006] used land cover data and three phenological metrics derived from four years of MODIS vegetation index data in a logistic regression approach to create a habitat suitability map for the continental US. Recently, Evangelista et al. [2009] showed that multi-temporal Landsat imagery significantly improved the prediction of tamarisk occurrence compared to single-scene analysis for a small study area in Colorado. The current center of tamarisk distribution includes the states of Arizona, New Mexico and Utah [Glenn and Nagler 2005]. Its distribution in Mexico is not well-known but already documented for the Sonoran coast [Harrison and Matson 2003], the Colorado River delta [Glenn et al. 1996] and Guaymas (Sonora, Mexico) [West and Nabhan 2002]. Building on the findings of previous studies such as Morisette et al. [2006], this study uses multi-year (2001-2008) time series of remotely sensed data to model the distribution of *Tamarix* spp. throughout the southern US and Mexico. We analyzed the Terra-MODIS Vegetation Index and Land Surface Temperature products regarding their usefulness for predicting potential tamarisk distribution together with *Tamarix* spp. occurrence records using the Maximum Entropy (Maxent) algorithm [Phillips et al. 2006]. Apart from Morisette et al. [2006], we are not aware of any other study modeling *Tamarix* spp. habitat covering comparatively large areas with remote sensing time series data in combination with topo-climatic data. Another novel aspect of this study is the approach to reliably predict the as of yet unknown potential range of tamarisk in Mexico.

2. DATA AND METHODS

2.1. *Tamarix* spp. records

We collected *Tamarix* spp. occurrence records from the Global Biodiversity Information Facility (GBIF, http://www.gbif.org) and from the National Institute of Invasive Species Science (NIISS, http://www.niiss.org) covering the current center of tamarisk distribution in the US in California, Arizona, Utah, New Mexico, Colorado and Texas (see Figure 2d). While GBIF is a global data base primarily including herbarium and museum collections, NISSS compiles occurrence records of invasive species for the US alone. NISS records were mainly collected during intensive species-specific field mapping projects such as – for tamarisks – the Southwest Exotic Mapping Program (SWEMP), Arkansas River Watershed Invasive Plant Plan (ARKWIPP) or helicopter surveys, and thus covered a wider range of habitats compared to GBIF samples. Due to a lack of sampling effort, only few samples were available for known populations in northern Mexico. Since model quality and results can be significantly influenced when samples are not representative for the whole climatic niche space of the target species, we used the Mexican *Tamarix* spp. records (10 samples) to run separate models. We are aware that this sample size is at the
lower limit to produce stable model predictions. Another important concern in species records is spatial autocorrelation, which may significantly impact model predictions [Dormann et al. 2007]. Spatial autocorrelation is basically a lack of independence between observations due to the fact that vicinity in space impacts the chance of occurrence. To reduce the effects of inherent spatial autocorrelation (especially between training and test sub-samples), we removed all *Tamarix* spp. records within the same 10 arc-minutes grid cells. The resulting data base comprised 1,726 records (NIISS: 1,449 and GBIF: 277).

### 2.2. Pseudo-absence / Background points

As recent work has shown [Philipps 2009], the influence of spatially biased samples can be reduced by comparing the occurrences with background points which reflect the same spatial bias (rather than using random background points). The idea is that a model based on presence and background data with the same bias will not focus on the sample selection bias but on any differentiation between the distribution of the occurrences and that of the background. Background data should thus be collected using the same methods or equipment as those used for the target species data. We therefore used 1,944 GBIF records of *Salix* spp. (willow), native trees often replaced during tamarisk invasion [Glenn and Nagler 2005]. Both *Tamarix* and *Salix* records showed the same spatial pattern with a concentration along rivers / stream lines and a core area of distribution in the southern US.

### 2.3. Remote sensing data

Two global MODIS-Terra L3 standard products of nominal 1x1 km² spatial resolution were used: (1) *Vegetation Indices* (MOD13A2, 16-day composites, Collection 5) designed for vegetation studies and (2) *Land Surface Temperature* (LST, MOD11A2, 8-day composites, Collection 5). The MODIS data were acquired from January 2001 through December 2007 (MOD13A2) / December 2008 (MOD11A2), as available through the NASA Earth Observing System Data Gateway (https://wist.echo.nasa.gov/api/). Nine MODIS tiles were mosaicked and reprojected to geographic projection (Datum WGS 1984) using freely available MODIS Reprojection Tool software (MRT, Version 4). To identify and replace low quality data, the *Time Series Generator* (TiSeG) software [Colditz et al. 2008] for linear temporal interpolation between valid observations was applied. Annual metrics (Table 1) as measures of vegetation greenness, phenology and thermal seasonality were calculated from the time series and its first derivative and averaged over the seven / eight years of the study period. We decided to use these metrics as the presence or absence of a species in any area is often distinguished not only by absolute levels of vegetation or climate variables, but also by subtle differences in their seasonality.

### 2.4. Bioclimatic and topographic data

The climate data set used (WorldClim, Version 1.4, http://www.worldclim.org) is based on temperature and precipitation values in the period 1950 to 2000 of a global network of 4,000 climate stations [Hijmans et al. 2005]. The meteorological station data had been interpolated to monthly climatic surfaces by using a thin-plate smoothing spline algorithm with latitude, longitude, and elevation data as independent variables [Hijmans et al. 2005]. The full set of 19 bioclimatic variables derived from this WorldClim data set was downloaded at 30 arc-seconds resolution, resampled and gridded to pixel location and cell size of the MODIS data. These bioclimatic parameters express spatial variations in annual means, seasonality and extreme or potentially limiting climatic factors for ecological studies. Since *Tamarix* spp. primarily occurs in riparian habitats [Glenn and Nagler 2005], we also calculated slope values (percent rise) from the SRTM digital elevation model available at an aggregated resolution of 30 arc-seconds through the WorldClim webpage. All environmental data were prepared and converted to ASCII files using DIVA-GIS software (Version 5.4, http://www.diva-gis.org).

### 2.5. Selection of environmental predictors

To improve computing efficiency, we performed a correlation analysis between all environmental variables (19 bioclimatic WorldClim variables, slope, 14 vegetation index
metrics and six land surface temperature metrics). We retained only one of several correlating variables with a Pearson correlation coefficient of $r^2 > 0.75$ from the available input data to act as a proxy for the other co-variables with respect to the eco-physiological requirements of tamarisk as described by Glenn and Nagler [2005].

2.6. Maximum Entropy model

We conducted our analysis using the Maxent modeling software (Version 3.3.1, available from www.cs.princeton.edu/~schapire/maxent/), which has great potential for identifying species distributions and habitat selection patterns [Baldwin 2009] and proved to be very useful in several comparative studies compared to other algorithms such as BIOCLIM, GAM, DOMAIN or GARP [Elith et al. 2006]. Maxent is a general-purpose algorithm for estimating probability distributions based on the principle of Maximum Entropy [Phillips et al. 2006]. The model evaluates the continuous suitability of each grid cell as a function of environmental variables from 0 (unsuitable habitat) to 1 (optimal habitat). Some advantages of Maxent are that it requires presence-only data, appears to be less sensitive than other approaches to the number of presence locations, can incorporate interactions between different variables (by automatically computing “product features” between predictor variables) and has no specific underlying assumptions regarding the statistical properties of the input data [Phillips et al. 2006]. However, one drawback of the Maxent method is that it uses an exponential model for probabilities which is not inherently bounded above. Thus, very high predicted values for environmental conditions outside the range present in the study area may occur [Phillips et al. 2006]. Maxent models were run under the following settings: 10 replicates (bootstrapping), auto features, jackknife test = true, logistic output format (ASCII), random test percentage = 25, regularization multiplier = 1, maximum iterations = 500, convergence threshold = 0.0001, and maximum number of background points = 5,000.

3. RESULTS AND DISCUSSION

3.1. Environmental predictors

Based on the results of the correlation analysis we selected three sets of environmental predictors (see Table 1): (1) TOPOCLIM: 10 topo-climatic variables, (2) RS: 10 remote sensing variables and (3) TOPOCLIM_RS: each top 5 variables of topo-climatic and remote sensing data in sets (1) and (2) measured by variable importance. During each step of the algorithm, Maxent increases the gain of the model by modifying the coefficient for a single feature which is then assigned to the environmental variable(s) that the feature depends on [Phillips et al. 2006]. According to these results, the most important topo-climatic predictor variables were precipitation of the warmest quarter, isothermality and temperature seasonality (Table 1). Out of the remote sensing-derived variables, minimum and maximum surface temperatures together with annual surface temperature range turned out to be the major determinants (Table 1) and were significantly more important than vegetation index features. The climatic parameters observed at tamarisk presence sites are characteristic of continental warm and arid climates. Our results are thus consistent with other studies, e.g. Kerns et al. [2005], who found that Tamarix spp. occurrence is positively correlated to warmer and drier sites with maximum daily surface temperature being the most important variable in an ENFA model. According to Friedman et al. [2005], the frequency of occurrence of T. ramosissima also has a strong positive relation with the mean annual minimum temperature (due to frost sensitivity). This hypothesis was confirmed by the high variable importance of the annual minimum surface temperature in our Maxent model. We assume that the spatial detail of BIO6 (minimum temperature of the coldest month) is not sufficient to reproduce the same relationship within the topo-climatic predictors. For the successful prediction of species invasions between currently occupied and target regions, environmental similarity is the basic requirement [Ficetola et al. 2007]. Several of our predictor variables show a strong north-south gradient, so that the spatial bias of the species occurrence records translates into a bias in environmental predictor space between records in the northern (US) and southern (Mexico) part of the study region
(Figure 1). This effect was taken into account for the distribution model (Sections 3.2. and 3.3.).

Table 1. Environmental predictors of the TOPOCLIM and RS data sets – Data source, ecological / bio-physiological interpretation and variable importance.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Information</th>
<th>Importance (%)</th>
<th>Variable</th>
<th>Information</th>
<th>Importance (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIO18</td>
<td>Precipitation of warmest quarter of the year</td>
<td>32.32</td>
<td>LSTmin</td>
<td>Minimum surface temperature</td>
<td>25.38</td>
</tr>
<tr>
<td>BIO3</td>
<td>Isothermality</td>
<td>19.40</td>
<td>LSTmax</td>
<td>Maximum surface temperature</td>
<td>21.40</td>
</tr>
<tr>
<td>BIO4</td>
<td>Temperature seasonality (standard deviation * 100)</td>
<td>16.69</td>
<td>LSTrange</td>
<td>Annual surface temperature range</td>
<td>13.97</td>
</tr>
<tr>
<td>BIO1</td>
<td>Annual mean temperature</td>
<td>11.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slope</td>
<td>Terrain steepness</td>
<td>8.21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIO6</td>
<td>Min. Temperature of coldest month</td>
<td>4.50</td>
<td>EVImax</td>
<td>Maximum green biomass</td>
<td>7.17</td>
</tr>
<tr>
<td>BIO15</td>
<td>Precipitation seasonality</td>
<td>2.07</td>
<td>EVImin</td>
<td>Minimum green biomass</td>
<td>5.56</td>
</tr>
<tr>
<td>BIO14</td>
<td>Precipitation of driest month</td>
<td>2.04</td>
<td>EVIdatemin</td>
<td>Time of minimum vegetation greenness</td>
<td>5.31</td>
</tr>
<tr>
<td>BIO2</td>
<td>Mean diurnal temperature range</td>
<td>1.90</td>
<td>EVIdatemin</td>
<td>Date of vegetation green-up</td>
<td>4.02</td>
</tr>
<tr>
<td>BIO19</td>
<td>Precipitation of coldest quarter of the year</td>
<td>1.84</td>
<td>EVIdatemin</td>
<td>Date of vegetation senescence</td>
<td>1.64</td>
</tr>
<tr>
<td>SUM</td>
<td></td>
<td>100.00</td>
<td>SUM</td>
<td></td>
<td>100.00</td>
</tr>
</tbody>
</table>

Figure 1. Comparison of mean Land Surface Temperature (LST) values for *Tamarix* spp. records in the US and Mexico compared to 10,000 random background points. The background points from USA and Mexico were significantly different (Mann-Whitney U test, p < 0.001).

3.2. Model performance

ROC (Receiver Operating Characteristic) plots were obtained by plotting all sensitivity values (true positive fraction) on the y-axis against their equivalent values (1-specificity, false positive fraction) on the x-axis. Maxent treats the background pixels as negative instances and species presence pixels as positive instances. The resultant AUC (Area Under Curve) is scaled between 0.0 and 1.0 and provides a measure of model accuracy that is independent of a particular probability cut-off and has therefore become one of the standard methods for assessing performance of SDMs [Anderson et al. 2003]. We obtained: AUCTOPOCLIM = 0.70 (prevalence 0.38), AUC_RS = 0.68 (prevalence 0.41) and AUCTOPOCLIM_RS = 0.71 (prevalence 0.39). According to a common classification [Swets 1988], 0.9 describes ‘very good’, 0.8 ‘good’ and 0.7 ‘useable’ discrimination ability of the respective model. The – according to these standards – ‘poor’ performance of our models is due to the application of *Salix* spp. occurrence records as background data, which leads to *Tamarix* spp. being modeled as more of a generalist than a specialist. Generalist species are not as well discriminated from the background as specialists (whose occurrences are more spatially and environmentally clustered) and thus result in lower AUC values [Veloz 2009].

Analogous Maxent models run with random background points covering the entire study
area revealed AUC values between 0.84 and 0.91 (prevalence 0.22 – 0.28). Nevertheless, we decided rather to correct for the prominent sampling bias present in the species records than to go for the maximum AUC. Using the reduced set of only Mexican occurrence records (10 samples), we obtained significantly different results ($AUC_{\text{TOPOCLIM}} = 0.98$ (prevalence 0.05), $AUC_{\text{RS}} = 0.96$ (prevalence 0.09) and $AUC_{\text{TOPOCLIM RS}} = 0.99$, prevalence 0.05). The combination of TOPOCLIM and RS data resulted in all cases in the highest AUC values. Absence points (even though available through NIIS) were not useful for model validation because tamarisk is currently not at its maximum potential distribution.

3.3. Modeled Tamarix spp. distribution

Applying the different sets of environmental variables resulted in predictive maps (Figure 2) well-related to the known distribution [e.g. Glenn and Nagler 2005, Morisette et al. 2006] of Tamarix spp. when using all occurrence records (Figure 2 (a) – (c)). However, the model predictions suggest that there is more suitable habitat than is currently invaded and indicate that we may still be in the early stage of tamarisk invasion. Morisette et al. [2006] found that especially low- and mid-elevation waterways in the western and central US are still susceptible to tamarisk invasion.

![Figure 2](image-url)  
Figure 2. Modeled potential tamarisk distribution using the TOPOCLIM (a), RS (b) and TOPOCLIM RS (c) sets of environmental variables for all records and (d) – (f) only Mexican records respectively. The distribution of GBIF / NIIS records is shown in (d).

For Mexico, potential tamarisk habitat was mapped for large areas in Baja California, in the states of Sonora and Sinaloa and in the Central Mexican Plateau (Figure 2 (d) – (f)).
According to these model results, we assume a high probability of tamarisk invasion into (semi)arid regions in Mexico in the future. Climate change, hybridization and continued human impact on river flow regimes may even facilitate this trend. Some caveats of the maps should be mentioned: First, pathways or sources for the introduction of *Tamarix* spp. were not available – even though these data are essential prerequisites for the establishment of invasive species. Secondly, the distribution maps are produced at a coarse 1 km² resolution constrained by the amount and individual file size of model input data. Therefore, they show not the actual presence of tamarisk but rather the probability of occurrence or habitat suitability (measured by similarity to vegetation and habitat characteristics of hot spots of tamarisk occurrence). Small-scale habitat patches (such as springs, narrow riparian zones, small tributaries) might thus not be detected by the model. It is therefore appropriate to use our results for large-scale assessments or to identify focus areas to be analyzed with higher resolution environmental data or remote sensing imagery.

4. CONCLUSION

Field surveys, remote sensing data and species distribution models are important tools for identifying priority areas for rapid response and developing efficient, long-term control strategies over large areas for any invasive species. In our opinion, the best interpretation of such model predictions is not as an absolute map of probability of occurrence but rather as a relative ranking of suitable habitat conditions since the true proportion of habitat already occupied by the target species is still difficult to assess by predictive modeling. The Maxent algorithm proved to be very useful in this study and has several features that support the analysis of remote sensing data. The different model scenarios showed that remote sensing time series contributed significantly to discover habitat characteristics even within similar climatic conditions. Even though the environmental data used for this study do not fully cover the spatial extent of currently known *Tamarix* spp. distribution, the models successfully reproduced its known distribution within the southern US. Additionally, we were able to predict the distribution for Mexico, where very little occurrence data exist, but where tamarisk is already recognized as an invasive species. To our knowledge, this distribution map is the first approximation towards modeling the potential range in Mexico. Model predictions produced in this study show that there is a very high risk of *Tamarix* spp. invasion over large areas in Mexico. Comprehensive tamarisk field surveys are thus required to confirm the preliminary results of this study. Future work will also have to include the analysis of higher resolution remotely sensed time series.

REFERENCES


Using Generalized Additive Models to Assess, Explore and Unify Environmental Monitoring Datasets

Russell G. Richards a, Rodger Tomlinson a and Milani Chaloupka b
a Griffith Centre for Coastal Management, Griffith University, Australia
r.richards@griffith.edu.au
a Griffith Centre for Coastal Management, Griffith University, Australia
r.tomlinson@griffith.edu.au
b Ecological Modelling Services Pty Ltd, University of Queensland, Australia
m.chaloupka@uq.edu.au

Abstract:
An on-going challenge for decision makers is the interpretation of temporal trends from monitoring data given that environmental processes often generate complex data that are multivariate and potentially nonlinear. Generalized additive models (GAMs) is a well-suited modelling framework for uncovering such trends and unifying datasets. This approach allows flexible specification of regression splines to represent the functional relationships between a response variable (the parameter of interest) and a suite of temporal and spatial covariates that can be continuous or discrete using a link function and smooth functions of the covariates. We highlight the utility of using GAMs through three case studies. The first highlights the use of a GAM to unify the findings of an established long-term water quality-monitoring program with those of a focused short-term monitoring program. In the second, a GAM is used to evaluate the spatial patterns in a biomonitoring dataset whilst simultaneously accounting for variability in oyster size, which can have a confounding effect on such data. The final case study focuses on a 12 month continuous monitoring program of oceanographic data as part of an evaluation of the environmental conditions for a desalination plant intake pipe. The context for these studies is predominantly water quality in the coastal zone, however the benefits and widespread application to other research areas is clearly evident.

Keywords: GAMs; GAMMs; water quality; nonlinear regression

1. INTRODUCTION

A major challenge to researchers and decision makers alike is interpreting spatial and temporal trends in monitoring data that are multivariate, potentially nonlinear and where spatial structure and dependency might also be important [Bailey et al., 2005]. Generalized additive models (GAMs) and generalized additive mixed models (GAMMs) are well-suited modeling frameworks for uncovering such trends because they allow flexible and non-linear specification of the dependence of some response variable such as the soft-tissue trace metal concentration in an oyster (i.e. biomonitoring data) on a set of temporal and/or spatial covariates without having to specify the model in terms of detailed parametric relationships.

While GAMs have been used extensively in areas such as fisheries [Venables and Dichmont 2004] and species distribution assessments [Guisan et al. 2002], they have been used sparingly on water quality and biomonitoring datasets. Yet these types of monitoring data typically represent a significant proportion of the overall monitoring effort in coastal zone areas. Trend analysis of such data is often assessed through linear models, which is
Richards et al. / Using Generalized Additive Models to Assess, Explore and Unify Environmental Monitoring Datasets

unrealistic for many applications especially if the behaviour of the response variable is poorly represented by a normal, homoscedastic and additive error term [Venables and Dichmont 2004]. There are various techniques utilizing splines to represent nonlinear and multivariate regression analysis e.g. univariate and additive splines, response surface models. However, GAMs provide a more flexible framework for regression analysis allowing the response variable to be drawn from other distributions of the exponential family including the Poisson, Bernoulli and Gamma.

\[ g(y_i) = f_1(x_i) + f_2(z_i) + Z_i b + \varepsilon_i \]  

The form of a GAM follows (1), where \( g \) is some smoothing link function, \( y_i \) is the response variable (i.e. the target contaminant), \( x_i \) and \( z_i \) are some predictor variables, \( f_1 \) and \( f_2 \) are smooth functions that are estimated and \( \varepsilon_i \) are independent error terms with a density function described by \( N(0, \sigma^2) \). Inclusion of a random effects term, \( Z_i b \), extends the GAM to a generalized additive mixed model (GAMM) where \( Z \) is a random effects matrix and \( b \) is a vector of random effects described by \( N(0, \varphi b) \) with \( \varphi b \) representing a covariance matrix.

We present here three case studies located in South East Queensland, Australia (Figure 1) that highlight the utility of GAMs/GAMMs for uncovering trends and processes in coastal monitoring data. The first case study shows how a GAM can be used to compare the findings of a focused short-term monitoring program with an established long-term water quality-monitoring program. Case study 2 provides an example of using a GAMM to evaluate the spatial gradients within an oyster-heavy metal biomonitoring dataset whilst simultaneously accounting for the potentially confounding effects of variability in individual oyster size and spatial structure.

The final case study outlined here provides an insight into the use of a GAMM to explore trends contained within a continuous monitoring program of oceanographic data whilst also accounting for autocorrelation of the error term.

For the three case studies, all continuous variables included in the model(s) were assumed to be potentially non-linear and therefore were fitted with smoother splines. All estimated parameter effects for each model were evaluated by fitting a series of models of decreasing complexity by sequentially removing covariates and comparing the Generalized Cross Validation (GCV) score [Wood 2006]. Diagnostic checks were also carried out to ensure assumptions of independence, normality and constant variance were upheld.

2. CASE STUDY 1: UNIFYING WATER QUALITY DATA FROM TWO MONITORING PROGRAMS

The first case study is based on an intensive short-term water quality monitoring program that was carried out in the Gold Coast Broadwater, a shallow estuary located in the South East corner of Queensland (Figure 1). These types of monitoring programs are regularly carried out in coastal zones for the purposes of baseline and/or impact monitoring. This monitoring was conducted over 3 separate days (Day 1, Day 2, Day 3) in early 2009 as part of an
assessment of extended release times of recycled water into the Broadwater. Extended release is a management strategy being implemented to cope with the higher treated wastewater loads caused by a growing local population. An outcome of this short-term monitoring program was total nitrogen (TN) concentrations (mean: 326 µg-N l⁻¹) that were significantly higher (t-test; p<0.001) than had been measured in the Broadwater (mean: 151 µg-N l⁻¹) during a long-term on-going monitoring program (Ecosystem Health Monitoring Program www.healthywaterways.org). While increased TN concentrations would be expected within the plume of the discharged recycled water, measurements recorded at locations well away from the discharge pipe (ca. 2 km) showed similarly elevated TN concentrations. Assuming that both monitoring programs used identical collection and analytical methods, and even for the sake of this example, the same equipment, personnel and laboratory, comparisons of the results between the two monitoring programs need to be viewed in the context of a range of mechanisms (in addition to the discharge of recycled water) that might be controlling TN concentrations within the Broadwater. Nitrogen concentrations within an estuary typically depend on a range of mechanisms [Eyre and McKee 2002] and the contribution of these will be fluid, both spatially and temporally, and occur at a range of scales [Eyre and McKee 2002]. Consequently, a GAM was used to uncover the functional drivers of TN in the Broadwater based on the water quality data measured during the long-term monitoring (1207 observation sets). The form of the GAM used is shown in (2):

\[
g(TN) = f(\text{conductivity}) + f(\text{year}) + f(\text{month}) + f(\text{lat, long}) + f(\text{TST, by=EBB}) + f(\text{TST, by=FLOOD}) + f(\text{time,wind9am,wind3pm}), \quad \text{family} = \text{quasi(link=log, variance=“mu“)}
\]

Conductivity was included as a proxy measure of catchment runoff and estuarine mixing. The metrical covariates of year and month were included as the data may be confounded by seasonal and longer-term non-linear trends [Cox et al. 2005] while the georeferenced coordinates of each monitoring site were included to account for any underlying spatial
To account for the potential effect of tidal current as a mechanism of TN dynamics, a covariate ($TST$) representing the elapsed time between slackwater, whether high or low, and monitoring was included and was conditioned upon whether the tide was ebb or flood flow (i.e. a varying coefficient). Finally, to account for the possible effects of wind-driven re-suspension, a three-term interaction smoother was specified for monitoring time and the local wind speed recorded at 9am and 3pm.

The model was developed and run on the software platform R (http://cran.r-project.org/) utilising the software package mgcv [Wood 2006] with a quasi family and logarithmic link function specified. The model was checked for over-fitting by randomly selecting 90% of the data to fit the data, which was then used to predict for the other 10% [Wood 2006]. The proportion of deviance explained for the fitted and the predicted model were consistent indicating that over-fitting was not occurring.

All variables, except $TST$ conditioned on the flood flow, were significant ($p<0.05$) and explained 83.3% of the deviance. Of specific interest in this study were the effects of conductivity, because it was used in the GAM as a proxy measure of catchment loading, along with sampling year and the interaction term between monitoring time and wind speed. A strong negative effect of conductivity was obvious (Figure 2a) and indicated that estuary mixing and/or freshwater catchment flows were a prominent driver of TN concentrations. This is not unexpected because the Broadwater catchment is strongly urbanized and is a significant source of nutrient loading [Moss and Cox 1999; Burton et al. 2004]. However, re-running the GAM with catchment rainfall as a covariate instead of conductivity decreased the performance of the GAM even when potential lag effects between rainfall and TN were accounted for by integrating rainfall over a 1-week period. Possible explanations are that rainfall levels measured at a single measuring station do not adequately represent the catchment rainfall and/or the base flow contribution might be significant. Monitoring year (Figure 2b) is characterised by a strong non-linear smooth with a generally positive effect for 2000-2004, a sharply decreasing effect from 2004-2005, a negative effect for 2005-2008 and a sharply increasing effect for 2009. The period of negative effect broadly coincides with El Niño conditions (2005-2006/2007), which would bring drier conditions to the east coast of Australia [Chiew et al. 1998] and presumably an associated reduced catchment loading. The interaction term between monitoring time and recorded local wind speed (9am, 3pm) (Figure 3) indicated that higher wind speeds (> ca. 25 kph) were important effects on TN. In particular, the 3pm wind speed had an increase positive effect on TN, which increased as the monitoring time occurred later in the day. Conversely, the effect of the morning wind speed (as measured at 9am) became more negative as monitoring time occurred later in the day.

![Figure 2](image-url)  
Figure 2. Smoother functions for (a) conductivity and (b) monitoring year. ‘Rug’ marks along the x-axis indicate raw data.
The strong negative relationship between conductivity and TN suggests that freshwater loading into the catchment is an important determinant of observed TN concentrations. Day 3 of the monitoring exhibited the lowest conductivities with a mean of 49 mS cm\(^{-1}\) and individual measurements as low as 43 mS cm\(^{-1}\) while the mean conductivity levels for Day 1 and Day 2 were above 52 mS cm\(^{-1}\). Coincidentally, Day 3 was the only monitoring day that had any significant rainfall (ca. 300 mm) recorded in the preceding 3 weeks. Days 1 and 2 were preceded by relatively low rainfall levels. TN concentrations in the Broadwater appear influenced by the timing of the measurements as the short-term monitoring took place in 2009 when a stronger annual effect is expected (refer Figure 2b). Finally, the effect of wind-driven resuspension might be a significant factor in the TN concentrations observed during the short-term monitoring, especially for measurements obtained in the afternoon. In particular, the interaction term (refer Figure 3) indicated that higher TN concentrations would be expected when there was relatively low 9am wind speed (< ca. 15 kph) and relatively high 3pm wind speeds (> ca. 25 kph). These conditions were observed on Day 1 with 9am and 3pm wind speeds of 13.0 and 26.0 kph respectively suggesting that on this date, wind-driven resuspension was an important factor in the observed TN concentrations.

3 CASE STUDY 2: HEAVY METAL BIOMONITORING DATA

Oysters and other bivalves are commonly-used as biomonitors for water quality assessments because they generally fit the criteria for an effective biomonitor. Specifically, they provide an easily measurable and time-integrated indication of contaminant bioavailability. However, it is often difficult to sample oysters of uniform size both within and across sampling sites [Robinson et al., 2005] and various studies have highlighted a significant influence of oyster size on metal bioaccumulation [Mackay et al., 1975; Hayes et al., 1998]. Consequently, a challenge is to unravel the potentially confounding effect of variable oyster size from biomonitoring data so that meaningful and accurate information can be obtained. Here, this potentially confounding effect was addressed by using a generalized additive mixed modelling (GAMM) approach. This enabled size effects and structured and unstructured (random) spatial effects to be considered simultaneously in the biomonitoring data as shown in (3):

\[
g(\text{Metal conc.}) = f(\text{oyster weight}) + f(\text{latitude,longitude}) + \text{random effects, family = normal} \tag{3}
\]

The GAMM assessment was carried out on native oysters collected from the intertidal zone of Moreton Bay, a large semi-enclosed embayment located in southeast Queensland, Australia (Figure 1) and consisted of 59 observations covering 10 sampling sites. It is important to note that the western catchment of the Bay is dominated by grazing and forestry [Capelin et al. 1998] and the state capital city of Brisbane (population ca. 840,000) while the eastern catchment is dominated by the sparsely populated Moreton and North Stradbroke Islands. This context is supported by the long-term monitoring carried out as part of the Ecosystem Health Monitoring Program (EHMP) in Moreton Bay, which
highlights that the overall water quality consistently exhibits a broad west-east (high-low) gradient in indicators such as chlorophyll-a, nutrients and turbidity. Six adult oysters were collected from the intertidal zone at ten locations around the Bay (Figure 1) during a one-off survey in July 2001 and the soft-tissue of individual oysters were analysed for Al, Cu, Fe, Mg, Mn and Zn [Richards and Chaloupka 2008]. The GAMMs were fitted in a Bayesian inference framework using Markov chain Monte Carlo (MCMC) simulation techniques implemented in the software program Bayes X [Brezger et al., 2005]. Specific details regarding implementation of this modelling including the specification of priors are described in Richards and Chaloupka [2008].

Figure 4. Estimated non-parametric functions of oyster soft-tissue weight (dry weight basis) for Fe and Mg. Posterior mean within the 95% credible region shown.

Figure 5. Correlated spatial effect for Mg (left panel) and Mn (right panel).

The posterior plots of soft-tissue Al, Cu, Fe, Mn and Zn concentrations were characterised by broad 95% credible regions that could each be fitted with a zero-gradient line, indicating no significant relationship between oyster soft-tissue mass and trace metal concentration (see Figure 4 for an example for Fe). Mg (Figure 4) was the only trace metal tested that displayed a clear negative effect of oyster size. The correlated spatial effect for each trace metal highlighted a pronounced west-east (high-low) gradient for Cu, Mn (Figure 5) and Zn, which resembles the spatial pattern previously observed for chlorophyll-a [Dennison et al. 1999]. Opposite east-west (high-low) gradients were observed for Al and Mg (Figure 5) while Fe was characterised by ‘hotspot’ concentrations at Deception Bay to the north and Redland Bay to the south. In this instance, the effect of oyster size as represented by the soft-tissue weight was not a significant effect in the regression modelling for five of the six metals tested. However, there is understandable utility in using a generalized additive modelling approach that explicitly conditions the bioaccumulated metal concentrations on the size of the oysters themselves. This was exemplified for Mg, which was found to have a significant negative size effect and which failure to account for might have resulted in a different spatial gradient and might have resulted in inappropriate management decisions/strategies.

4 CASE STUDY 3: ASSESSMENT OF OCEANOGRAPHIC DATA
The final case study presented here is focused on the results of a continuous monitoring campaign that was carried out to characterise the typical physical oceanographic processes occurring in the vicinity of a desalination inflow and diffuser pipes, South East Queensland, Australia (Figure 1). The implementation of desalination plants in Australia are likely to increase with water demand and a drier environment and therefore understanding the oceanographic processes operating at the intake is an important part in validating the operational design of the plant itself. This monitoring program included the measurement of current speed and direction at 1-meter depth intervals through the water column along with water depth, near-seabed water temperature and turbidity. Current velocity, water depth and temperature were measured using an acoustic Doppler current profiler (ADCP) and turbidity was measured using a YSI sonde. In this case study, a generalized additive mixed model (GAMM) was developed to compliment the assessment of oceanographic processes and help determine the relationship between intake water quality and the ambient environmental conditions. All data was averaged to 30 minute intervals as a means of overcoming the different duty cycles used for the equipment throughout the monitoring program. For example, ADCP velocity measurements were recorded every six, 12 and 20 minutes at various stages of the monitoring.

A subset of the monitoring dataset comprising measurements recorded over a two-month period (January-February 2008) was selected to provide a manageable dataset for the GAMM and this consisted of 1336 observation sets. Seabed turbidity measured by a YSI was selected as the response variable. Turbidity was log-transformed so that it could be sampled from a gaussian distribution, which reduces the computational effort of the model. The predictors initially trialled were the depth-averaged easting (VelocityE) and northing (VelocityN) current velocities and water depth as measured by a seabed ADCP and represented by the Julian day respectively. The two velocity vectors entails that current direction does not have to be specified. Thin-plate tensor product smooths were specified for the four-predictor variables. Autocorrelation was explicitly modelled to address the potential autoregression in the errors due to the small time intervals of monitoring. The model structure is shown in (4):

\[
\log(\text{Turbidity}) \sim f(\text{VelocityE}) + f(\text{VelocityN}) + f(\text{Julian Day}) + f(\text{Depth}), \text{ family = quasi (link = log, variance="mu"), corr=corAR1(})
\]

The best-fit model was found to be the original initial model less the VelocityN covariate. The dominant effect was Julian Day and highlighted a decreasing metrical effect (Figure 6).

**Figure 6.** Smooth functions (a) Julian day, (b) water column depth and (c) depth-averaged current easting current velocity.

Depth was found to have a negative linear effect on TN while there was a negative effect of the magnitude of the easting velocity (VelocityE). A surprising outcome of the analysis was the negative correlation between turbidity and VelocityE, which appears opposite to conventional wisdom of increased current velocity leading to increased turbidity through re-suspension of sediment. However, this observed trend might indicate that there is a turbid layer close to the seabed during ‘calm’ periods that is resuspended into the water
column when current velocity increases with a net decrease in the turbidity levels at or near the seabed, which has implications for the ongoing performance of the desalination plant.

5 CONCLUSION AND RECOMMENDATIONS

We have presented here three case studies that have highlighted the utility and flexibility of GAMs/GAMMs as techniques for evaluating coastal datasets. Their ability to uncover simultaneous nonlinear functional relationships coupled with their relaxation of the assumption of normality in the residual make them particularly powerful for investigating spatio-temporal trends in environmental datasets and tailoring regional- and local-specific water quality guidelines. This latter point is a key philosophy to setting appropriate water quality guidelines in Australia [Cox et al. 2005]. We have found that GAMs/GAMMs often uncover new or emerging knowledge in the study area as exemplified by all three case studies presented here. In Case Study 1, the flexible specification of the GAM allowed the nonlinear effects of rainfall, tidal flow and wind effects to be accounted for when assessing the main drivers of total nitrogen concentrations. This was despite these predictor variables being poorly aligned with the response variable. We applied a combination of interaction terms (wind effect), proxy variables (conductivity for rainfall) and varying coefficient (tidal effect) approach to address this challenge. Furthermore, we simultaneously conditioned the TN data for annual effects, which had important implications for comparing TN concentrations across different years. In Case Study 2, we used a GAMM to condition a biomonitoring dataset for oyster size, correlated spatial effects and uncorrelated spatial effects so that the underlying bioaccumulation effects could be compared. Failure to account for these combined effects can easily lead to incorrect conclusions about such biomonitoring data, which can propagate through to policy decisions. Finally, in Case Study 3, we used a GAMM to assess a continuous monitoring dataset that had been generated for the assessment of ambient environmental conditions at a desalination plant intake. Key concerns in this instance were the relationships between turbidity and current velocity. As measurements were recorded at short time intervals (e.g. 30 minutes) there was a considerable challenge in assessing such continuous data because independence of the error terms is unlikely. We were able to condition the data for autocorrelation by using a GAMM, with the resulting observation of a negative relationship between eacting current velocity and turbidity having potentially important implications for the operation of the desalination plant. Further, the GAMMs allowed for straightforward application of formal diagnostic tests during their development. These tests included the comparison of different models using Generalized Cross Validation scores, as well as tests of normality, constant variance, model overfitting and autoregression; these aspects are of crucial importance in model development but are often overlooked. Finally, the context for the studies presented here, including the biomonitoring study, is water quality in the coastal zone. However, the benefits and widespread application of GAMs to other research areas, whether it is the terrestrial or marine environment, is clearly evident.

ACKNOWLEDGEMENTS

We thank Waterway Management, Gold Coast Water (Case Study 1), the CRC for Coastal Zone Estuary (Case Study 2) and Gold Coast Desalination Alliance (Case Study 3) for providing financial assistance. We are also grateful to the Ecosystem Health Monitoring Program project team for their help in providing access to EHMP water quality data for Case Study 1 and 2. Finally, we would like to thank Peter Bell, Lawrence Hughes, Sally Kirkpatrick, Darrell Strauss and Greg Stuart for their assistance in one or more of the case studies presented here.

REFERENCES


Development of a land-use component for an integrated model of the German biogas system

A. Schaldach¹, D. Lapola¹, U. Jessen², M. Thees², R. Schaldach¹
¹Center for Environmental Systems Research, ²Institute of Production Engineering and Logistics
University of Kassel, Kurt-Wolters-Str. 3, D-34109 Kassel, Germany
(uli.Jessen@uni-kassel.de; schaldach@usf.uni-kassel.de)

Abstract: Germany is one of the biggest producers of biogas in Europe. Biogas can either be combusted in highly efficient combined heat and power units to produce electricity and heat, or upgraded to biomethane for injection into the natural gas grid. As political conditions encourage the installation of new biogas plants, their number and capacity is expected to grow significantly in the next few years. In Germany, energy crops play an important role for biogas production. Due to their relatively low energy content transport distances from the field to the plants are a crucial issue for the efficiency of the biogas system. In order to simulate the interaction of the different processes involved in the biogas system, an integrated model has been developed at the University of Kassel. This article describes the development and testing of one of its central components: a land-use model to simulate the cultivation of crops and grassland management for biogas production under a set of different spatial constraints. First simulation experiments indicate that the model calculates plausible area demands for energy crops used for biogas production.

Keywords: Biogas system, land-use, integrated modelling

1. INTRODUCTION

In Germany, the production of biogas for electricity generation has grown dramatically in the past few years. From 2004 to 2008 the overall capacity of installed biogas plants has increased almost sixfold and reached 1.435 GWel while their number has doubled to more than 4.100 [Thrän et al., 2009]. This development was mainly triggered by the introduction of subsidies for using agricultural commodities as substrate for biogas production. According to Scholwin et al. [2008], 47 % of the substrates (by mass) in use are energy crops, which may account for about 78 % of the electrical energy generated by biogas. The most important energy crop is silage maize (79%), followed by grass silage (8 %), cereal silages (7 %) and grains (6 %). Other important substrates are slurry and organic wastes.

A number of factors put the use of land for production of biogas in a preferable position to biofuels such as ethanol: (1) a wide range of substrates can be used for producing biogas [Amon et al., 2007a], (2) the energy yield for biogas (from maize silage) is far greater (16.600 GJ/km²) than for ethanol from wheat (6.000 GJ/km²) or rapeseed oil (4.500 GJ/km²) [KTBL, 2006], (3) the use of the digestate as a high quality fertilizer [Amon et al., 2007b] supports the closing of the nutrient cycle in the cropping system and (4) the potential of avoiding greenhouse gas (GHG) emissions increases immensely, if slurry from livestock farming is used as coferment. Finally, biogas can either be combusted in highly efficient combined heat and power units to produce electricity and heat, or upgraded to biomethane for injection into the natural gas grid. For both conversion pathways GHG savings in the transportation sector (with optimized engines) are greater than savings using other biofuels such as bioethanol [Campbell et al., 2009].

As current political and legislative conditions in Germany encourage the installation of new biogas plants, their number and capacity is expected to grow significantly in the next years. However, competition to food crops and the use of agricultural land for nature conservation and urbanization will be limiting factors in this development. Concerns about expected
odorous emissions of biogas plants and objections against a landscape dominated by tall-growing maize crops may restrain the expansion of the biogas sector. In 2007 the area for growing renewable resources in Germany reached about 20,000 km², of which an estimated 4,000 [FNR, 2007] to 5,500 km² [Scholwin et al., 2008] produced energy crops for biogas. This accounts for 3.4 - 4.6 % of the arable land, respectively. Different studies have identified a future potential ranging from 30,000 [EEA, 2006] to 56,000 km² [Thrän et al., 2005] for renewable resources, or 25 – 47 % of the arable land, with a growing share of crops for biogas, as the currently dominating area of rapeseed for biodiesel has already reached its peak at about 10,000 km². Due to the generally low energy content of the substrates used for biogas production, transport distances are a major factor determining the profitability of a biogas plant.

The different processes involved in the generation and utilization of biogas and their linkages constitute the biogas system [Lantz et al., 2007]. In order to analyse the effectiveness and sustainability of these systems, often methods from the field of Life Cycle Analysis (LCA) are used [e.g. Börjesson and Berglund, 2007] to model the incorporated energy and material flows. This approach has two disadvantages. First, it does not account for the dynamic nature of processes and, second it only indirectly addresses the spatial distribution of the system’s components. As an alternative proposal the prototype of a dynamic spatially explicit model of the German biogas system has been developed at the University of Kassel. This article describes the development of one central component of this model: a land-use model to simulate the cultivation of crops and pasture as substrate for biogas production under a set of spatial constraints. The following section first gives an overview of the architecture of the integrated model. After that, the structure and a prototypic application of the land-use model is presented and discussed.

2. INTEGRATED MODELLING OF BIOGAS SYSTEMS

The current situation of modelling biogas systems is characterized by a large number of singular models representing different aspects of the process of biogas generation and utilization. Application fields of these models include the cultivation of energy crops, the biochemical procedure of biogas generation and the utilization of the digestate as well as logistic processes, in particular the ensilage as storage process and transports between cultivated areas, storages and the biogas plant. Here a wide range of modelling methods is applied. The modelling of cultivation of energy crops includes a spatial database as well as calculation and simulation models inter alia for land use and energy yield. Continuous simulation is used to model processes such as the biogas generation or the treatment of residual material. Finally, discrete event simulation is the most suitable method for logistic processes where dynamics of discrete objects, e.g. transport and harvest vehicles, has to be modelled. Moreover, discrete event simulation is a valuable tool for the discretized modelling of continuous processes such as stock characteristics of silos.

**Figure 1. Model pipeline of biomass conversion**
In principle, all these models serve well to address the specific problems they were
developed for, and can give answers and predictions with sufficient accuracy. However,
they cannot predict the entire behaviour of a biogas system as they model only details of
the whole system and their system boundaries are based on assumptions. To model the
entire biogas system it is necessary to bring these models together. Basically, there are two
ways to do this. On the one hand, one can model all aspects of the entire system within one
single, new model; on the other hand, it is possible to couple the existing (validated)
models of subsystems by well-defined interfaces. This integration of simulation models
requires both the exchange of data and the time synchronization of the dynamic models.
The modelling approach of the University of Kassel puts the discrete event simulation
model of the entire biogas system in the centre of a data pipeline. This model gets data and
parameters for its configuration from partial models that represent the different subsystems
of the biogas system. Furthermore, it provides result data that can be used for model
evaluation (see Figure 1).

![Simulation model of biomass conversion in Germany](image)

**Figure 2.** Simulation model of biomass conversion in Germany

The dynamic integrated model of the entire system (hereafter “BioSys model”) was built up
utilizing the simulation tool Plant Simulation. The model enables researchers not only to
model single biogas plants and their surroundings, but also, based on a raster map, the
partly automated modelling of any German region (see Figure 2). The regional model
includes areas reserved for cultivation of energy crops, installed biogas and energy
conversion plants and, where applicable, waste treatment plants as well as a logistics
network including transportation and storage (see Figure 2) [Jessen et al., 2009]. For the
configuration, the BioSys model gets data from connected upstream partial models by well-
declared interfaces in form of Microsoft Excel sheets. These include inter alia size of raster
elements, allocation of cultivated areas as well as harvest data from partial model “energy
crops”, data regarding gas flow and consistency from partial model “biogas generation”
and data regarding flow of residual material and its consistency from partial model
“treatment of residual material”. The results of the BioSys model, e.g. dynamic
dependencies of gas yield and content of silos or transports within simulation period, are
lead to a specific module for statistic evaluation. The current BioSys model is the basis of a
holistic examination of biogas systems. In the following the modelling of energy crop
production within the partial model “energy crops” as well as its results, which are used as
input parameters for the BioSys model, are presented.
3. SPATIAL SIMULATION OF ENERGY CROP PRODUCTION

3.1 Input data

The assessment is carried out for Germany on the uniform geographic raster with a cell size of 5 arcmin (~7km x 7km), which is also used by the BioSys model. Each raster cell is characterized by a dominant land-use type, an area fraction that is occupied by settlement structures as well as data on terrain slope, road infrastructure and nature conservation area. Furthermore, it includes information about biomass productivity in terms of crop yields and net primary productivity (NPP) of grasslands. The spatial input data is prepared using the GIS software package ArcGIS 9.2. The initial land-use map for the year 2000 is based on the CORINE CLC2000 land cover database from the European Environment Agency (EEA). The fraction of settlement area is derived from the population density map by Klein-Goldewijk et al. [2005]. Slope data is based on the HYDRO1k dataset from the US Geological Service, while information on road infrastructure is based on the VMAP level 0 dataset. The location of nature conservation areas is taken from the World Database on Protected Areas (http://www.wdpa.org). Biomass productivity is calculated by the LPJmL model [Bondeau et al., 2007]. LPJmL produces output on a 30 arcmin raster, which is then assigned to the 5 arcmin cropland and grassland cells located within each 30 arcmin cell (see section 3.2). According to Weiland [2003] almost 98% of all biogas plants in Germany use combined heat and power plants for the utilization of biogas. Data on the installed electrical power of biogas plants was available on the level of EU NUTS 3 administrative units in the year 2007 [Scholwin et al., 2008], distinguishing four different classes with upper limits of: (i) 1000 kWel, (ii) 10.000 kWel, 20.000 kWel and 40.000 kWel. In order to be able to take into account the location of biogas plants in our later analysis, we have introduced the concept of “virtual biogas plants” (hereafter VBP). For each NUTS 3 unit one VBP is located on the grid cell in its geographic centre. It is assumed that its installed electrical power is equal to the medium value of the respective class, i.e. if unit which falls into class 2, it receives a VBP with 5,500 kWel. The sum of the installed electrical power of all VBPs amounts to 1.150 MWel which is about 5% less than the data provided by Scholwin et al. [2008], who estimate an amount of 1.232 MWel.

3.2 Simulation of crop yields and grassland NPP

LPJmL is a process-based model to simulate global vegetation dynamics and the associated carbon and water fluxes on 30 arcmin raster cells. Agricultural land-use productivity is simulated through the consideration of crop functional types (CFTs), either rainfed or irrigated, representing the world’s most important annual field crops. Moreover, LPJmL’s crop module simulates sowing dates, crop phenology, crop growth and carbon allocation at a daily time step. All four processes respond to climate variables such as precipitation, temperature and insulation. A comprehensive evaluation of LPJmL’s performance for the simulation of crop yields, crop phenology and carbon-fluxes is presented by Bondeau et al. [2007]. For our assessment, the model is applied to calculate rainfed and irrigated crop yields for the most important energy crops maize and rye as well as the NPP of managed grasslands. This means that simulation runs for these CFTs are done for all raster cells within Germany.

Model results for maize and rye are dry mass grain yields. These are converted to fresh mass silage (whole plant) yields using the factors for dry matter (dm) content and ratio of

<table>
<thead>
<tr>
<th>Dry mass content [%]</th>
<th>Conversion harvested part to whole plant</th>
<th>Energy content [kWh/t fresh mass]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maize 35%</td>
<td>1.5</td>
<td>323</td>
</tr>
<tr>
<td>Rye 35%</td>
<td>1.3</td>
<td>274</td>
</tr>
<tr>
<td>Grass 35%</td>
<td>1.0</td>
<td>293</td>
</tr>
</tbody>
</table>

1Kim and Dale [2004]; 2KTBL [2007]
harvested part to whole plant shown in Table 1. For grassland it is assumed that the whole harvested biomass is used for silage production.

### 3.3 Spatial allocation of agricultural land-use

We use the LandSHIFT model [Schaldach and Koch, 2009] to calculate the spatial allocation of crops and grassland used for biogas production. It relies on a “land use systems” approach that describes the interplay between anthropogenic and environmental system components as drivers of land-use change. Central model elements are land-use activities that simulate the land-use decision making of different sectors such as settlement, crop cultivation and grazing on a 5 arcmin raster. Each activity performs two processing steps. First the suitability of each raster cell is calculated with a multi-criteria analysis (eq. 1). In the following step a specified demand for a good (e.g. crop production) or service (e.g. housing) is allocated to the most suitable raster cells.

\[
\psi_k = \sum_{i \in \text{suitability}} w_i \prod_{j=1}^{m} c_{j,k} \quad \text{, with } \sum_i w_i = 1, \quad \text{and } p_{i,k}, c_{j,k} \in [0,1]
\]  

For this study we have implemented a new land-use activity for the spatial allocation of energy crops and grassland (Figure 3). The implemented algorithm iterates over all the 316 VBPs and allocates energy crops to the cells in the geographic neighbourhood (n-order Moore) of the respective VBP to fulfil its energy demand. The maximum neighbourhood searching radius is set following Walla and Schneeberger [2008] in relation to the installed electrical power.

![Processing steps of the model algorithm of the land-use component.](image)

For each VBP the installed electrical power is translated to an energy demand to be provided by biogas. In the first step the suitability of each neighbourhood cell is determined. The multi-criteria analysis considers three factors, which have been identified as important for bioenergy plantations by Hellmann and Verburg [2008], all having the same weight \((w_1 = w_2 = w_3 = 0.333)\): terrain slope \((p_1)\), crop yield \((p_2)\) and available road infrastructure \((p_3)\). We assume that higher slope lowers suitability due to negative implications for the use of agricultural machinery and a higher erosion risk while higher crop yields lead to an increased suitability as well as a better road infrastructure does. Furthermore, suitability decreases linearly with distance from the analysed VBP. This effect is expressed by the constraint \(c_1\). A further constraint \(c_2\) excludes nature conservation areas from being used for the cultivation of bioenergy crops.

In the second step production of energy crops or grassland is allocated to the most suitable cells until the energy demand of the VBP is fulfilled. The energy production of each selected cell is based on the yield and biogas-specific energy content of the allocated crop type (or grassland) (Table 1). In the current version of the algorithm only cells classified as cropland in the initial land-use map are used for crop production. In order to conserve soil carbon stocks and to avoid an additional carbon debt in the sense of Fargione et al. [2008],
grassland and semi-natural vegetation (e.g. forests) cannot be converted. Grassland cells that are located within the neighborhood of a VBP are automatically used to fulfill the energy demand. Agricultural management takes place in a 3-year rotation period, i.e. only 1/3 of a cropland cell area can be used for an energy crop in each year. For each cell a decision is made between silage maize and rye silage production. The cultivation of maize has priority while rye is only used in case of maize yields below a threshold of 38t fresh mass, which is 5% below the yield level classified as low by KTBL [2007].

3.4 Interface to the integrated biogas model

The model output comprises raster maps with the location of energy crops and grassland used as substrate for biogas generation as well as information on the amount of production on each cell in metric tons. This data is further processed before it is handed over to the integrated model. In order to account for the year to year variability of crop and grassland yields, for each cell a 20-year time series of yields and harvest dates is stochastically generated. The statistical distribution function is derived from the census data for the respective crop between 1995 and 2005 and the available sources for harvest dates. The interface to the integrated model is realized as a set of 3 Microsoft Excel tables: (1) A list of raster cells where the VBPs are located, (2) a list of cells where energy crops and grassland are grown including yields and harvested area, and (3) their assigned VBP. Additionally this list includes for each cell the newly calculated time series on crop production and harvest dates. Based on this information the integrated model is able to simulate transport and storage processes between agricultural production sites and a VBP.

4. MODEL EXPERIMENT

We have designed two simulation experiments. The aim of experiment 1 is to provide a first plausibility test of the simulation results generated by the newly developed land-use activity. For the simulation we use data on the installed electrical power of biogas plants for NUTS 3 units in the year 2007 (status quo) provided by Scholwin et al. [2008]. For translating this information into VBP level energy demands we assume that each plant operates 80 % of the time of the year. Furthermore it is assumed that 15 % of the substrate (mass) is provided in form of manure, representing an energy fraction of 1.9%. These values are about half of the threshold value defined in the German law on renewable energies (EEG) for future bonus payments to the plant operator and therefore is a relatively conservative assumption. Other substrates such as biotic waste are not considered in this test run.

![Simulation results for (A) status quo: locations of energy crops and their percentage of total cropland in each Nuts 3 unit and (B) scenario: % energy crops of total cropland in each Nuts 3 unit.](image)

Figure 4. Simulation results for (A) status quo: locations of energy crops and their percentage of total cropland in each Nuts 3 unit and (B) scenario: % energy crops of total cropland in each Nuts 3 unit.

The simulated area of energy crops in Germany amounts to 6.260 km² with 3.810 km², 900 km² and 1.550 km² of silage maize, rye silage and of grassland, respectively. The total simulated area therefore exceeds the estimates of Scholwin et al. [2008] of up to 5.500 km² by 14 %. Figure 4 shows the spatial distribution of cultivation of energy crops.
In experiment 2 (scenario) an eightfold increase of the size of each VBP is assumed to explore possible spatial limitations of the cultivation of energy crops. All other assumptions are the same as in experiment 1. The model result shows an increase of cultivated area to 43,750 km² equaling 26 % of the total agricultural and 37 % of total cropland area in Germany. We further find that the demand of 31 VBPs could not be fulfilled due to the lack of suitable cropland in their neighborhood (Figure 4). Most of these plants are located in Saxony (11) and Lower Saxony (10), followed by Baden-Württemberg (6) and Bavaria (4).

5. DISCUSSION AND CONCLUSION

The article describes a newly developed component of a spatially explicit land-use model to simulate the cultivation of energy crops in Germany. The modified land-use model contributes to the integrated dynamic simulation of processes in the German biogas system and helps to realize the coupling of environmental processes (e.g. crop growth) with technical processes (e.g. transportation of the substrate and electricity generation from biogas). Furthermore, it also takes into account the location of biogas plants into the land-use decisions, comparable to the approach presented by Hellmann et al. [2008].

In the first simulation experiment we could demonstrate that the developed land-use activity produces relatively plausible results. One reason for the overestimation of area needed for the cultivation of energy crops is the lack of data for the use of other substrates for biogas generation. Major source of uncertainty is the regional availability of slurry and organic waste. Here more detailed regional data would be necessary to achieve more realistic results. The second simulation experiment illustrates that there is still a large potential for the expansion of biogas production in Germany. It also becomes obvious that the linear up-scaling approach we used leads into regional problems to fulfill the given demands. In order to produce more reliable results, another model component needs to be developed for the spatial allocation of additional biogas plant capacities. Other uncertainties that have not been taken into account by this study are increasing yield by the introduction of new crop varieties and the possible effect of climate change for plant growth conditions and water availability for irrigation.

The developed model has a relatively coarse spatial resolution and due to the lack of more detailed data it does not consider individual biogas plants. This leads to a highly idealized representation of the real-world system. Nevertheless, the model captures the most important processes of land-use decision making and can be applied to both finer scale raster datasets and more detailed plant location data without major modifications. Beside these spatial constraints, the major simplifications of this study are twofold: (1) Economical competition between biogas crops and other crops (in the way as signalled by Thrän and Kalthoff [2007]) is not considered but instead it is assumed that biogas crops are always given priority on one third of the available cropland area of a raster cell. This limitation could be overcome by introducing economic aspects of agricultural decision making into the suitability analysis. (2) Agricultural management in terms of fertilizer use and ploughing are not considered by the model. Furthermore, the LPJmL runs were summarized for the time period 1991 - 2000. The effect of inter-annual variability of climate variables such as temperature and precipitation on irrigation water requirements and crop yields is taken into account only in the post-processing step. Improvements of the land-use model should therefore also include the application of a more detailed crop/grassland model.

ACKNOWLEDGMENTS

This work has been conducted as part of the project “Development and integration of biogas plants into energy supply structures” funded by the “Fachagentur für Nachwachsende Rohstoffe” (FNR) in Germany. The authors thank all their colleagues at the University of Kassel who were involved in that project. Furthermore the authors thank their colleagues at the Climate Impacts and Vulnerabilities group at the Potsdam Institute for Climate Impact Research (PIK) for providing access to the LPJmL model. Rüdiger
Schaldach acknowledges the support from the EU SCENES project and David Laopla acknowledges the financial support from the MPI IMPRS in Hamburg.

REFERENCES

Amon, T., B. Amon, V. Kryvoruchko, A. Machmuller, K. Hopfner-Sixt et al., Methane production through anaerobic digestion of various energy crops grown in sustainable crop rotations, Bioresource Technology, 98(17), 3204-3212, 2007a.
FNR, Fachagentur für Nachwachsende Rohstoffe, Daten und Fakten zu nachwachsenden Rohstoffen, 2007.
Sustainable forest management for bioenergy

Giulia Fiorese, Giorgio Guariso
Dipartimento di Elettronica e Informazione, Politecnico di Milano,
via Golgi 34/5, 20133 Milano, Italy (fiorese, guariso@elet.polimi.it)

Abstract: Biomass from the forest sector can be an important source of renewable energy and can contribute to climate change mitigation and bioenergy development. However, the removal of biomass from forests can have significant impacts on the forest ecosystems and therefore requires a thorough analysis. The purpose of this work is to compare different alternatives of sustainable forest management with the aim of minimizing greenhouse gases emission. The model used for the analysis, CO2FIX, describes the flows of carbon per unit area of biomass, soil storage and bioenergy products. The model was applied to the forests of the Italian region of Lombardy. We identified four macro-categories: coniferous, deciduous, mixed coniferous and deciduous forests, short rotation forests. For each macro-category, we ran a simulation, with an annual time step for a hundred years horizon, of various management policies: no harvest activities, maintenance of a constant stock, different rotation lengths, maximization of harvested biomass. We identified the most efficient management policy for each macro-category in terms of carbon emissions saved and carbon sequestered. Over the entire region, it emerges that the potential contribution to climate change mitigation amounts to about 1.5 million tons of CO₂eq per year, equal to about 15% of the total reduction needed to meet Kyoto Protocol targets in the region.

Keywords: Bioenergy; Carbon sequestration; Forest management; CO2FIX; Climate change mitigation.

1. INTRODUCTION
To mitigate global climate change our society will have to rely both on low carbon technologies and on maximizing the capacity of the biosphere to sequester carbon from the atmosphere. Even though the sequestration capacity of both soil and vegetation alone will not be enough to compensate the increase of carbon concentration, in the next years their contribution will be crucial. The carbon balance of terrestrial systems has therefore gained more attention because of the connection with global climate change.

Within terrestrial systems, forests play a major role as recognized in Article 3.3 of the Kyoto Protocol, where afforestation, reforestation and deforestation accountings are made mandatory. Countries can also choose to include management activities of existing forests as an addition to their carbon sinks. Furthermore, the use of forest biomass in substitution of more energy-intensive products, such as fossil fuels or other materials, is another major contribution that forests can provide [Brown et al., 1996; Nabuurs et al., 2008].

Globally, forests represent a significant carbon stock. They store 283 Gt of carbon in the biomass, 38 Gt in dead wood and 317 Gt in soils (top 30 cm) and litter. The total content of 638 Gt (for 2005) is more than the amount of carbon in the entire atmosphere. This standing carbon is combined with a gross terrestrial uptake, which was estimated at 2.4 Gt of carbon a year, a good deal of which is sequestration by forests [UNFCCC, 2010].

Management can strongly affect this carbon balance. Forests of new formation sequester carbon and store it in their biomass until an upper limit is reached; at this point, carbon losses due to respiration, mortality, external causes of disturbances and other utilizations may overcome the photosynthetic activity [Odum, 1969]. A recent study that involved boreal and temperate forests [Luyssaert et al., 2008], showed that forests between 15 and 800 years of age accumulate carbon and have a positive net system productivity (including trees and soil), even though there is an age-related decline. This carbon accumulation is
explained with the different rates at which tree mortality and decomposition occur: the first is much faster than the second. Consequently, old-growth forests with tree losses do not necessarily become carbon sources. However, this process strongly depends on the stand structure and the disturbances forests have been subject to.

The biomass extracted and transformed in wood products is itself a limited reservoir of carbon. If a forest has been used to extract biomass or if a forest is lost because of natural events, its pool of carbon will disperse; the same happens when degraded woody products are not replaced by analogous products. On the contrary, the benefits that derive from the replacement of fossil fuels with energy from biomass can be considered irreversible: when energy is produced in substitution of any given fossil fuels, a defined amount of greenhouse gases will be permanently avoided [Tuskan et al., 2001].

Clearly, forests provide a number of other important ecosystem services, such as soil erosion control, wildlife habitat and diversity, as well as relevant economic contribution to sectors like tourism. This paper concentrates only on the carbon biogeochemical cycle to analyse how much the management of forest for the production of substitute of energy-intensive products and the preservation of carbon sinks can contribute to the regional carbon budget.

The approach followed in this study is composed by two main steps. The first is the formulation of a method to compare different management policies and to identify the optimal ones, with the objective of maximizing avoided greenhouse gas emissions and the carbon fixed by the forest system (trees and soil). The second is to assess the environmental benefits that can be derived from the adoption of those policies identified in the first step, over the Region of Lombardy in Northern Italy.

2. STUDY REGION
Forests in the Lombardy region extend over an area of 665,702 hectares, more than one quarter of the overall regional area (24,000 km²). The largest fraction of this area (58%) is covered by deciduous forests, followed by conifer forests (17%) and by mixed forests (13%); the remaining part is not classified into any of these categories [INFC, 2005; ERSAF, 2007]. The most widespread deciduous forests are chestnuts, hornbeams, and ash trees; the most common conifer is the Norway spruce forest. About two thirds of the forest area is public property, while the remaining is privately owned. Only one quarter is classified as a protected or naturalistic area.

About 20% of the total has natural origin, while the vast majority are semi-natural forests derived from sylvicultural and afforestation interventions, which for a small share, was made by replanting of indigenous species. According to the data recently collected by the National Forest Inventory, about 8% of forest can be considered in a juvenile stage, 61% in an adult stage and the remaining 31% in an old stage.

Lombardy forests generally lay in a partially abandoned condition that followed centuries of overexploitation, with consequences of aging and degradation. Removal of woods from the forest sharply decreased after the end of World War II; however a slight resurgence was observed after the 1980s. The removed quantity varies consistently from year to year and ranged between a minimum of 0.8 Mm³ in 2004 to a maximum of 1.8 Mm³ in 1999 [ISTAT, 2006].

3. THE MODELLING APPROACH
Different management strategies directly affect carbon pools and flows, both in trees and soil, and are determined by rotation length, whole-tree or conventional harvesting, thinning intensity, age-class distribution of the forests and many other factors. To compute the carbon budget, we adopted the CO2FIX V 3.1 model [Schelhaas et al., 2004; Masera et al., 2003], which has been widely used for studying, for example: the consequences of forest management policies [Kaipainen et al., 2004], the carbon profile according to forest types [Nabuurs and Schelhaas, 2002], the emissions from silvicultural activities [Markewitz, 2006] and the carbon credit accounting of Italian forest stands [Scarfo and Mercurio, 2009]. This model allowed us to design and compare alternative management policies. We then applied these management policies to the Lombardy forests to estimate how much forest management can contribute to bioenergy production and to climate change mitigation.
3.1 The CO2FIX model

The CO2FIX model describes a forest with a set of modules that represent what happens in the biomass, in the soil and to the wood products (Figure 1). Each of these modules assesses the incoming and outgoing flows of carbon. Finally, there is a fourth module that calculates the overall carbon balance, accounting for all flows, with respect to the atmosphere.

The Biomass module describes the forest biomass growth from the carbon that is absorbed via photosynthesis; the module distinguishes different sections for leaves, branches, logs and roots. Each of these sections is regulated by a set of defined equations that describe growth, mortality, turnover and cutting of the biomass. Mortality, turnover and harvest residues that are left on the soil are the input of the Soil module; this module describes biomass decomposition and the respective carbon flows that depend on climatic conditions and litter quality and composition. The raw material that is harvested is the input of the Products module that describes the various uses of biomass, such as manufacturing of sawn wood, board and panels, pulp and paper, and wood fuels. The biomass used to produce energy is finally described in the Bioenergy module that takes into account the carbon emission flows avoided thanks to the substitution of fossil fuels in electric or thermal energy production. The final module Carbon in the Atmosphere describes the balance of all these flows with respect to the atmosphere. The model has been validated for several climatic regions, including central Europe [Masera et al., 2003]. CO2FIX is a flexible tool that can be applied to several forestry species. For example, the Biomass module, through the cohort model, can describe mono-cultures and mixed forests and can deal with age-structured stands.

All variables are expressed in terms of carbon per hectare (tC/ha) for a single homogeneous stand of forest. The measure of carbon per hectare can be converted, according to the appropriate parameters, into units of weight (dry t/ha) or of volume (m³/ha) of the biomass. The time step used for the simulation of the forest dynamics is one year.

The output of the model is given in the form of two indicators. The first quantifies the annual average amount of greenhouse gases sequestered by the forest (standing biomass and soil). The second quantifies the annual average carbon emissions avoided by using biomass instead of a fossil fuel to produce energy.

![Figure 1. Structure of the CO2FIX model.](image)

3.2 Applying the CO2FIX model to the forests of Lombardy

To study different alternatives for the management of the forests of Lombardy with CO2FIX, we considered four major forest macro-categories: conifer forests, deciduous forests, conifer and deciduous mixed forests, arboriculture tree plantations. These macro-systems are also used in the land use digital cartography of the region [ERSAF, 2007].
Therefore, this correspondence allows to connect the management policies of the macro-categories with their spatial extension and sites by applying GIS software.

Because of their fully different management, arboriculture plantations should be considered separately from the others forests categories. In fact, in this case we assume to grow biomass in short rotation cycles (3 years) that, therefore, have very high yields (up to 33 m³/ha/y) [Fiorese and Guariso, 2010]. Arboriculture has an old tradition in the region and is presently constituted mainly by poplars and some other fast growing species, such as willows and robinia. These areas are all located in the southern flat and fertile part of the region. The figures given by arboriculture are useful to compare its potential with other forestry sectors.

CO2FIX simulations were run for each of the four macro-categories assuming single cohort, even-aged forest stands. For the parameters of the Biomass module (see Table 1), we used values of the stand carrying capacity, annual rate of growth, turnover and mortality derived from a national study by APAT [2002]. For the parameters of the Soil module, the initial soil carbon content and its evolution over time were elaborated from a recent study that covers the region and that estimates the organic content of the soil at various depth [Progetto Kyoto Lombardia, 2008]. The soil carbon dynamics depends also on the local climatic conditions that regulate the moisture and the chemical, physical and biological processes that occur within the soil. In the Product module, we assumed that all the log is used for energy production and that a fraction of leaves and branches (10% for arboriculture and 30% for the other macro-categories) is left on the soil [Masera et al., 2003]. Finally, in the Bioenergy module, it is necessary to define how the biomass will be converted into energy (e.g., electricity or heat and with which conversion efficiency) and what fossil fuel it will substitute. It is assumed here that biomass is used to produce only thermal energy in a plant with 80% efficiency. This thermal energy substitutes that produced by a natural gas plant (assumed to have a 85% efficiency). The avoided emissions are thus estimated with respect to natural gas for all the GHG gases with the appropriate heating value and emission factors for biomass (LHI 16 MJ/kgbiomass, 0.0 gCO₂/kgbiomass, 0.48 gCH₄/kgbiomass, 0.06 gN₂O/kgbiomass) and for natural gas (LHI 42.62 MJ/kggas, 3853 gCO₂/kggas, 0.88 gCH₄/kggas, 0.08 gN₂O/kggas). It is possible to choose different conversion options, such as the use of biomass to generate electricity instead of heat. In this case, the conversion efficiency is more favourable for natural gas (whose conversion efficiency is about 55%) than for biomass (24%). The choice of the energy conversion should, in any case, also depend on local energy demand. Typically in mountainous areas, such as those of the Alps of Lombardy, biomass is mostly used locally in small thermal plants (from hundreds of kWt up to 20 MWt).

Throughout the analysis, we assumed an average carbon content of 0.5 tC per dry ton of biomass and a lower heating value of 16 MJ/kg for all the macro-categories. The dry wood mass density (kg/m³), on the other hand, varies from species to species (Table 1). All other parameters for the simulations have been set as advised in the manual of the model [Schelhaas, et al., 2004]

CO2FIX simulations were run for each of the four macro-categories assuming single cohort, even-aged forest stands. For the parameters of the Biomass module (see Table 1), we used values of the stand carrying capacity, annual rate of growth, turnover and mortality derived from a national study by APAT [2002]. For the parameters of the Soil module, the initial soil carbon content and its evolution over time were elaborated from a recent study that covers the region and that estimates the organic content of the soil at various depth [Progetto Kyoto Lombardia, 2008]. The soil carbon dynamics depends also on the local climatic conditions that regulate the moisture and the chemical, physical and biological processes that occur within the soil. In the Product module, we assumed that all the log is used for energy production and that a fraction of leaves and branches (10% for arboriculture and 30% for the other macro-categories) is left on the soil [Masera et al., 2003]. Finally, in the Bioenergy module, it is necessary to define how the biomass will be converted into energy (e.g., electricity or heat and with which conversion efficiency) and what fossil fuel it will substitute. It is assumed here that biomass is used to produce only thermal energy in a plant with 80% efficiency. This thermal energy substitutes that produced by a natural gas plant (assumed to have a 85% efficiency). The avoided emissions are thus estimated with respect to natural gas for all the GHG gases with the appropriate heating value and emission factors for biomass (LHI 16 MJ/kgbiomass, 0.0 gCO₂/kgbiomass, 0.48 gCH₄/kgbiomass, 0.06 gN₂O/kgbiomass) and for natural gas (LHI 42.62 MJ/kggas, 3853 gCO₂/kggas, 0.88 gCH₄/kggas, 0.08 gN₂O/kggas). It is possible to choose different conversion options, such as the use of biomass to generate electricity instead of heat. In this case, the conversion efficiency is more favourable for natural gas (whose conversion efficiency is about 55%) than for biomass (24%). The choice of the energy conversion should, in any case, also depend on local energy demand. Typically in mountainous areas, such as those of the Alps of Lombardy, biomass is mostly used locally in small thermal plants (from hundreds of kWt up to 20 MWt).

Throughout the analysis, we assumed an average carbon content of 0.5 tC per dry ton of biomass and a lower heating value of 16 MJ/kg for all the macro-categories. The dry wood mass density (kg/m³), on the other hand, varies from species to species (Table 1). All other parameters for the simulations have been set as advised in the manual of the model [Schelhaas, et al., 2004]

### Table 1. Initial values of forest simulation (regional averages).

<table>
<thead>
<tr>
<th>Forest macro-category</th>
<th>Dry wood mass density</th>
<th>Carrying capacity</th>
<th>Standing biomass</th>
<th>Carbon content</th>
<th>Organic C in soil 100 cm deep</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conifer forests</td>
<td>526</td>
<td>339</td>
<td>321</td>
<td>84</td>
<td>154.7</td>
</tr>
<tr>
<td>Deciduous forests</td>
<td>705</td>
<td>183</td>
<td>160</td>
<td>56</td>
<td>126.1</td>
</tr>
<tr>
<td>Mixed forests</td>
<td>616</td>
<td>261</td>
<td>241</td>
<td>74</td>
<td>137.2</td>
</tr>
<tr>
<td>Arboriculture plantations</td>
<td>515</td>
<td>-</td>
<td>115</td>
<td>30</td>
<td>108.0</td>
</tr>
</tbody>
</table>

a INFC, 2005; b APAT, 2002; c elaborated from Progetto Kyoto Lombardia, 2008.

### 3.3 Definition of forest management policies

The management of forests can be defined as sustainable when it – at least – maintains the system biodiversity, productivity, capacity of renewal, vitality, and when it does not compromise the capacity of supplying, now and in the future, ecosystem services [APAT, 2002]. In this study, sustainability is defined only in terms of biomass: the stand biomass at the end of the management horizon is constrained not to be lower than the initial one. However, since this may determine some initial and final transients that are due to the
specific initial conditions (and thus to the management of the past 20-30 years), we will refer in the following only to the average performances, transients excluded. For each forest macro-category, the optimal management problem can be formalized with the objective of maximizing the sum of the average annual CO₂ fixed by the forest ($I_f$) and the average annual CO₂ avoided by the substitution of natural gas with biomass for heat production ($I_a$). These in turn depend on the biomass $B$, whose dynamics is obviously determined by the management policy $u$.

The optimal management policy is thus the solution of the following optimal control problem:

$$\max_u \frac{1}{N} \sum_{t=1}^{N} \left( I_f(t) + I_a(t) \right)$$

(1)

$$B(t+1) = f(B(t), u)$$

(2)

$$I_f(t) = g'(B(t)), \quad I_a(t) = g''(B(t))$$

(3)

$$B(N) \geq B(0)$$

(4)

where constraints (2)-(3) are implemented through CO2FIX, and the time horizon $N$ considered in this study is 100 years.

Additionally, we restrict the set $U$ of possible policies to only six alternatives, which closely resemble those followed in the past. Thus, their social acceptability is guaranteed.

The considered management policies are:

- **Complete protection**: the forest is left evolving according to its natural cycle, without any intervention or biomass removal.

- **Conservation**: each year, biomass is removed from the forest in such an amount that guarantees a constant stock of carbon; the annual net productivity is therefore removed each year.

- **Long, medium and short rotation cycle**: with the first biomass harvest, the forest density is set at a value that allows for maximum growth, therefore guaranteeing a high biomass yield in the following years; after the first cut, biomass is harvested at regular intervals, every 20, 10 or 5 years.

- **Maximum sustainable yield (MSY)**: with the first harvest, the forest density is set at the value that allows the maximum growth from one year to the next; the annual net productivity is then removed each year.

### 4. RESULTS

The values of the two components of the objective for each policy are listed in Table 2 for deciduous forests. Within all the management policies analysed, the optimal policy is the short rotation cycle that, for this macro-category, allows for both the highest carbon sequestration and a high amount of avoided emissions. Table 3 lists the optimal results for all the forest macro-categories. In order to estimate the potential contribution to climate change mitigation of forest biomass in the region, once the optimal management policy has been defined for each macro-category, it is necessary to estimate the extension of such a macro-category. Land use cartography [ERSAF, 2007] has been used for this purpose and figures are listed in Table 4. However, not all the forest area can be managed because of natural or technical constraints (for example, slope limits the accessibility of the forest by machineries and, at the same time, prevents from extracting biomass from parts of the forests where erosion might be more severe).

<table>
<thead>
<tr>
<th></th>
<th>$I_f$ [tCO₂/y]</th>
<th>$I_a$ [tCO₂/y]</th>
<th>$I_f + I_a$ [tCO₂/y]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protection</td>
<td>-1.06</td>
<td>0.00</td>
<td>-1.06</td>
</tr>
<tr>
<td>Conservation</td>
<td>0.51</td>
<td>0.72</td>
<td>1.24</td>
</tr>
<tr>
<td>MSY</td>
<td>0.71</td>
<td>3.38</td>
<td>4.09</td>
</tr>
<tr>
<td>Long cycle</td>
<td>0.16</td>
<td>3.41</td>
<td>3.57</td>
</tr>
<tr>
<td>Medium cycle</td>
<td>0.23</td>
<td>3.31</td>
<td>3.54</td>
</tr>
<tr>
<td>Short cycle</td>
<td>0.93</td>
<td>3.25</td>
<td>4.18</td>
</tr>
</tbody>
</table>
Table 3. Optimal policy and value of the indicators for each forest macro-category.

<table>
<thead>
<tr>
<th>Forest macro-category</th>
<th>Optimal management</th>
<th>Harvested biomass [m³/ha/y]</th>
<th>( I_f )</th>
<th>( I_a )</th>
<th>( I_f + I_a )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conifer forests</td>
<td>Long cycle</td>
<td>5.52</td>
<td>0.25</td>
<td>3.93</td>
<td>4.19</td>
</tr>
<tr>
<td>Deciduous forests</td>
<td>Short cycle</td>
<td>3.40</td>
<td>0.93</td>
<td>2.52</td>
<td>4.18</td>
</tr>
<tr>
<td>Conifer and deciduous mixed forests</td>
<td>MSY</td>
<td>5.32</td>
<td>0.74</td>
<td>4.43</td>
<td>5.17</td>
</tr>
<tr>
<td>Arboriculture plantations</td>
<td>SRF</td>
<td>33.51</td>
<td>-0.41</td>
<td>23.24</td>
<td>22.95</td>
</tr>
</tbody>
</table>

Table 4. Regional extension of each forest macro-category, its manageable part and estimated reduction of greenhouse gases according to the optimal policy.

<table>
<thead>
<tr>
<th>Forest macro-category</th>
<th>Regional forest area [ha]</th>
<th>Manageable area [ha]</th>
<th>Avoided emissions [ktCO₂eq/y]</th>
<th>CO₂ sequestered [ktCO₂/y]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conifer forests</td>
<td>134,352</td>
<td>10,647</td>
<td>41.8</td>
<td>2.7</td>
</tr>
<tr>
<td>Deciduous forests</td>
<td>340,137</td>
<td>97,253</td>
<td>316.1</td>
<td>90.4</td>
</tr>
<tr>
<td>Conifer and deciduous mixed forests</td>
<td>91,555</td>
<td>14,989</td>
<td>66.4</td>
<td>11.1</td>
</tr>
<tr>
<td>Arboriculture plantations</td>
<td>39,323</td>
<td>39,323</td>
<td>918.6</td>
<td>-16.1</td>
</tr>
<tr>
<td>Total</td>
<td>605,367</td>
<td>162,212</td>
<td>1343.0</td>
<td>88.1</td>
</tr>
</tbody>
</table>

We assumed to manage only forest land with a moderate slope (lower than 30%) and close enough to the existing road network (distance less than 200 m). This constraint guarantees that the harvested biomass can be collected and transported to the conversion facility at reasonable costs. These constraints were applied to the forest areas through simple GIS operations on the land use map. Table 4 shows that the extension of the forest that satisfies these two manageability constraints is a small share (27%) of the overall forest area. The management of forests over this area under the proposed policies leads to a decrease of CO₂ of about 1.43 Mt/y from avoided emissions and from sequestration in the forest system (trees and soil). The greatest contribution to the avoided emissions is given by arboriculture plantations; at the same time, however, this forest macro-category is a source of CO₂ from the forest ecosystem (0.016 Mt). This happens because arboriculture plantations are composed by young trees (completely harvested every three years, with a small litter) and thus cannot exploit carbon storage in the soil. On the contrary, for the other forest macro-categories only a part of the biomass is harvested (for example the net primary productivity in the MSY alternative) and a larger fraction of this is left on the soil. The dynamics of carbon in the soil thus plays a major role in determining if the forest can be considered as a source or a sink of GHG.

The amount of biomass that can be harvested could be increased by extending the area of the managed forests. Under the assumptions made on slope and road vicinity, the manageable area could be increased for example by constructing more roads into the forests (which may be positive also for other activities such as fire fighting, but may also...
have negative impacts, such as habitat fragmentation). The maximum potential area that can be managed over the entire region, could be increased in this way from 162 to 262 thousand hectares, resulting in a 30% increase of carbon sequestration and substitution. The map of the province of Como (a part of Lombardy) in Figure 2 shows all the forest area, the part that can be managed and the part that could potentially be managed.

5. DISCUSSION AND CONCLUSIONS

Historical harvests from the forest of Lombardy have covered about 11,000 hectares and have produced an average of little less than one million cubic meter of wood per year, corresponding to about 90 m$^3$ per hectare and year - quite higher than any sustainable policy.

According to the management policies proposed in this paper, harvesting could cover a larger surface of 35,000 hectares, with a total harvest of about half a million cubic meter of wood per year, i.e. an average of about 14 m$^3$ per hectare per year. A substantial contribution to these figures is given by arboriculture that accounts for 74% of the wood production (even if only 24% of the area is presently grown with this macro-category). The suggestion is thus to shift from the overexploitation of only a small area, to a more sustainable harvesting of all the forests, each with its own best policy.

From the figures above, it clearly emerges that the role of forest as bioenergy suppliers is quite more important than their being carbon sinks, but there is no contradiction between these two functions. On the contrary, there might be a positive synergy. In fact, in the absence of harvesting, our forests are not bound to increase their productivity/growth or their carbon sequestration. Indeed, this analysis shows that if forests are let evolving according to their own dynamics, without any intervention (complete protection policy), they might become a source of carbon, instead of being a sink. The difference between the overall sequestration under the optimal solution and the sequestration under such a protection policy, can be considered the “price”, in terms of missing sequestration, that society pays for the lack of proper management of forests.

Moreover, the current abandonment of the Lombardy forests constitutes a form of pressure as well, that may not just impact on the carbon sequestration aspect, but also on the other ecosystem services, such as for instance the spreading of wildfires or the diffusion of parasites. A sustainable set of management operations may contribute to provide healthier forests with a higher capacity of sequestering carbon from the atmosphere and provide wood products.

The management policies proposed in this paper could contribute to the GHG reduction goal set in the Kyoto Protocol. Forests in Lombardy may contribute with a reduction corresponding to about 15% of the total expected reduction in the region [Progetto Kyoto Lombardia, 2008]. Furthermore, this can be achieved in a sustainable way, i.e. without compromising the future biomass production of the forests and even without modifying current land cover.

The model adopted in this study has the single objective of optimizing CO$_2$ reduction. This approach overlooks all the other relevant issues that should be considered to define a sustainable forest management policy such as, for example, forests biodiversity, that depends on standing biomass and on litter quality as well. Our future research will thus focus on investigating policies that consider other ecosystem services as objectives of the optimal forest management problem. Other improvements might regard the emissions caused by the harvest and transport operations and other life cycle emissions of forest biomass with respect to the equivalent life cycle emissions of natural gas. However, previous works [e.g., Fiorese and Guariso, 2010] have shown that the emissions due to logistic operations contribute only to a few percent of the overall balance of the bioenergy systems. Moreover, we plan to evaluate the robustness of the results by adopting other carbon budget models, such as, for example, FORMICA [Böttcher et al., 2008] that has been developed for regional scale studies, or GORCAM [Schlamadinger and Marland, 1996] that accounts not only for the carbon mitigation potential of biomass, but also for possible land use changes.

ACKNOWLEDGMENTS

This work was partially supported by project Consolidamento ECATE – Efficiency and Environmental Compatibility of Energy Technologies, funded by LEAP (Laboratorio
REFERENCES


ERSAF, Destinazione d'Uso dei Suoli Agricoli e Forestali della Lombardia, 2007.


Progetto Kyoto Lombardia, Per vincere la sfida dei cambiamenti climatici e del controllo dei gas serra nella regione più industrializzata d’Italia, ed. Fondazione Lombardia per l’Ambiente, Milano, Italy, 2008.

Scarfo F., and R. Mercurio, Carbon credit accounting: the model CO2FIX v. 3.1 applied to a beech stand under Forest Management in southern Italy, *Forest@ 6*, 215-228, 2009.


An interactive visual decision support tool for sustainable urban river corridor management

Lewis Gill\textsuperscript{1}, Vikas Kumar\textsuperscript{3}, Eckart Lange\textsuperscript{2}, David Lerner\textsuperscript{3}, Ed Morgan\textsuperscript{2}, Daniela Romano\textsuperscript{1}, Ed Shaw\textsuperscript{3}

(acp08lrg, vikas.kumar, e.lange, ed.morgan, d.n.lerner ed.shaw, d.romano)@sheffield.ac.uk

The University of Sheffield

\textsuperscript{1} Department of Computer Science, Regent Court, 211 Portobello, Sheffield, S1 4DP, UK
\textsuperscript{2} Department of Landscape, Crookesmoor Building, Conduit Road, Sheffield, S10 1FL, UK
\textsuperscript{3} Catchment Science Centre, Kroto Research Institute, Broad Lane, Sheffield, S3 7HQ

Abstract: Sustainable integrated catchment management is a complex task that often involves finding compromise between the views of multiple stakeholders that may not easily be brought to consensus. Interactive three dimensional (3D) visualisations of landscapes can facilitate discussion, but these can only partially inform decision makers as many important aspects of interventions are abstract. Such abstract criteria however can be modelled using a Bayesian Network (BN), combining expert opinions, empirical evidence and data derived from existing models. Thus, it is hypothesised that by combining an interactive 3D landscape design software with a BN, a decision support tool is created that brings together the complementary strengths of both techniques. To test this idea, such a tool has been developed for a river management problem in Sheffield, UK. Impounding the River Don that flows through the city are many weirs which form significant landmarks of the urban riverscape and determine the recreational quality for canoeists. Consequently, management decisions must account for the effect of weir modification not only on the visual aesthetic, but also on the abstract recreational quality of the river for canoeists. To support this problem, an interactive tool has been developed that integrates 3D design and visualisation of weir modification options with indicators of the utility of the river from the perspective of the canoeists, modelled using a BN. It is demonstrated that it is possible to feed back the indicator assessment of a design alongside the visualisation or as part the visualisation.

Keywords: Bayesian Network; Decision-support; Integrated catchment management; Weir; Modelling; Interactive 3D landscape; Design process;
1. INTRODUCTION

Sustainable management of urban river corridors is a complex task that, increasingly, involves finding a compromise between the views of multiple stakeholders that may not easily be brought to consensus. One promising method of facilitating discussion between interested parties is to provide interactive three dimensional (3D) visualisations of landscapes, where the viewer is able to move freely around the model (Schroth 2007). When viewing visualisations, people bring their education, experience and interests, each interpreting them from a different viewpoint (Zlatanova 2007). Despite these varied stances, Al-Kodmany (1999) observed stakeholders in collaborative design workshops used digital landscape visualisations as a common language to discuss change. He also found that proposed management interventions can be presented as changes in the virtual landscape, enhancing stakeholder understanding.

Yet, visualisation media alone can only inform decision makers about the spatial and aesthetic nature of management changes. There are many other important non-visual outcomes to a management intervention that the decision maker would like to understand. Logically, the next step is to link visualisation media with modelling techniques that can predict these outcomes, allowing users to more deeply evaluate management interventions. For example, Bishop (2005) proposed an interactive visualisation system for forest management that worked over time and allowed a user to see design changes by moving sliders related to the three pillars of sustainability. Isaacs et al. (2008) showed, using their S-City VT tool, it is possible to show the results of sustainability assessment of built form, modelled with an analytical network process, directly in the 3D visualisation. Gill et al. (2010) suggest constructing interactive 3D landscape models directly connected with assessment models through a defined semantic for areas of landscape using procedural modelling. This allows modelling alterations to directly feed into linked assessments and back to the user in real-time, increasing user understanding of design change.

However, some outcomes of management interventions are more abstract or intricate in nature, proving harder to capture than simple data driven models, and as a consequence new predictive models are now being developing to help address these challenges. Within the uncertain and complex decision problem of integrated catchment management, an increasingly recognised modelling solution is that of the Bayesian Network (BN) (Castelletti & Soncini-Sessa 2007). BNs allow for causal relationships (probabilities) to be specified based on subjective assessments (“expert opinion”), empirical evidence, data derived from existing models, or a combination of all three, making them particularly suitable for integrating predictive information from multiple disciplines (Kumar et al. 2010). Incorporating expert opinion is important because, as is often the case in integrated catchment management, domain expert opinion can be used to create Bayesian networks when there is a paucity of data. However, while increasingly sophisticated modelling techniques are being developed, a barrier to their uptake by water managers is their complexity, and lack of a user friendly interface (Borowski & Hare, 2007). Water managers, who often need to involve stakeholders in the decision making process, do not have time to teach the public to interact with complex models to explore different management options.

Therefore, it is hypothesised that, by combining an interactive 3D landscape design software tool with a BN, a decision support tool can be created that brings together the complementary strengths of both techniques; aiding the understanding of the visual nature of proposed interventions, and making it possible to assess their impact on abstract non-visual criteria. Proposals for changes to river corridors can be fed into the system via user alterations to a 3D landscape model and these alterations can be assessed utilising the knowledge of experts captured in the BN. The overall effect will be that the users of such a system will have an accessible method for increasing their understanding of the consequences of a design decision without the need for extensive training and, thereby, the transfer of knowledge from expert to user will strengthen the process of decision making, as shown in Figure 1.
In this paper, ongoing work to test this hypothesis is introduced. Work to marry visualisation software to a BN, which tackles a real life catchment management problem, is discussed and future research directions expanded upon.

2. CASE STUDY

In Sheffield, UK, many weirs impound the River Don as it runs through the city. While their primary role as infrastructure for the water wheels of early industry is now obsolete, they are significant landmarks of the urban riverscape. Currently, there is much interest from numerous stakeholder groups in modifying these weirs: some anglers would like to install fish passes; potential victims of flooding are interested in removing weirs; proponents of hydro-electricity want to install micro-hydro schemes; enthusiasts of natural history aspire to restore the river to a more natural state while enthusiasts of the river’s heritage want to conserve weirs for their historic value. These modifications also determine the recreational quality of the river for canoeing, a major leisure activity on the river. Weir height, steepness, roughness, plane (orientation relative to the river bank), and presence/absence of attached modifications all determine how canoeists perceive the fun and danger of the weirs. Consequently, decisions on modifying the weirs must account for this abstract impact on the recreational quality of the river for canoeists. To allow this impact to be considered in decision making, an indicator was constructed by building a BN that can predict the influence of weir modification on the quality of the river for canoeing. The process of building this BN and issues regarding its construction are detailed in Shaw et al. (2010).

3. METHOD

To support the testing of the proposed integrated interactive 3D / BN system, a software tool has been developed that presents a graphical view of weir modification options and evaluates the utility of a weir from the perspective of the canoeists. The overall solution comprises of three distinct components: a 3D weir model generator; an assessment system containing the BN and an interactive 3D visualisation system. Miller (1968) states that there can be a detrimental effect on the train of thought of a user when a computer system response to user input is greater than one second. Therefore, to avoid hindering the design process taking place, it is aimed to make the integrated system react at within this time frame, whilst attempting to deliver output in real-time to provide an interactive experience.
3.1 3D weir model generator

The role of this component is to procedurally generate 3D models of weirs, which increases the speed of generating models of the weirs and removes the need to manually construct models. The generator needs several inputs; a 2D polygon region that defines the extent of the weir, two independent edges along the polygon that define the upstream and downstream boundaries, and a distinct set of parameters that allow the user to implement various weir management options. The polygon area of the weir can often be derived from geographic map data, such as the Ordnance Survey Mastermap data. The parameters are simple name/value pairs, such as weir ‘Height’, ‘Bumpiness’, ‘Number of Steps’ in the weir. Given these inputs, the system generates a 3D representation of the weir using the base polygon and parameters to inform the shape of the resultant model. If the weir polygon is altered or the attributes changes, the system will remove the existing model and replace it with one that incorporates the alterations upon a user request for an updated model.

3.2 Assessment system

The canoeist BN, referred to Section 2, has been developed using the Netica software package that allows a BN to be defined using a standard graphical user interface. Each variable (node) in the Netica model has a series of discrete states, e.g. weir height is categorised as ‘High’ (>3m), ‘Medium’ (1-3m) and ‘Low’ (<1m). The nodes are linked in a cause-effect network and the relationships between them are described probabilistically. At the end of the network are indicator nodes; danger and fun of a weir from the viewpoint of canoeists and river quality a combined function of the danger and fun. When the state of an input parameter is set, probabilities will propagate through the BN changing the states of the indicator variables. The compiled Netica model is then linked to the 3D weir model generator through the use of the Netica Application Programming Interface (API). One advantage of using this API is that the system maintains a separation of the BN from the visualisation system and 3D model generator, which allows refinements to be made to the BN outside the visualisation host program, but feed directly into the model assessments.

As a weir model is generated, the various key/value parameters are fed into a weir assessment module in the form of input variables. Additionally, the geometry of the input polygon is analysed to determine the following extra parameters; width, steepness, and weir plane, and these are also fed into the assessment system. Once the input values are assigned to the nodes in the BN, the assessment system can then interrogate other BN nodes to retrieve the altered indicator values, which are then available to be displayed to the user.

The assessment system is designed to incorporate multiple assessment models at once and provide an overall assessment based on all connected models. The BN is integrated alongside a more traditional data and process driven model that examines the potential for hydro-electric power generation. This model, created in Excel, is integrated with the assessment system through COM automation. In a similar fashion to the BN integration, appropriate inputs are fed directly into the Excel model and cell values are available to read back to the assessment system.

Results of the multiple assessments can then be displayed either in a separate window, or within the visualisation, depending on suitability. The system outputs the results of assessments in HTML, which provides the basis for a highly configurable output to the user. So far, the system displays the predicted states for indicator nodes in the Netica model and outputs one result of the micro-hydro model as proof of concept.

The 3D model generator is connected to the assessment system and, when a model is created, it forces an assessment be made. The model generator compiles a list of all the variables needed and passes these, as required, to each analytical model in the assessment system. When inputs alter, it is the assessment system that is responsible for displaying the resultant indicator values to the user. This ensures that when a model is altered an up to date assessment is presented to the user. As the model generator issues the assessment, it
can incorporate the results of the analysis in the resultant model. For instance, the model can be coloured to display the level of weir “fun” derived from the BN to give visual feedback within the visualisation. A threshold value of 40% certainty being reached in either the “high” or “low” states results in the model being coloured green or red respectively, otherwise it is assumed to be in an intermediate state and coloured gray.

3.3 Interactive 3D visualisation system

The weir generator and assessment system have been created in C++ allowing them to be used in a “plug-in” module for the interactive 3D landscape visualisation system, Simmetry3d, which is based on video game technologies. This enables users to make weir modifications within larger landscape visualisations that provide spatial context. Simmetry3d offers interactive views both in birds-eye and eye level “walk-throughs” on a range of computer hardware, from laptops to immersive stereo projection, giving flexibility when supporting stakeholder workshops.

To model a weir in Simmetry3d, all the user needs to do is create a weir polygon in the model using the polyline tool, and designate this as a weir area. The upstream and downstream edges have to be selected and the input parameters dialog is displayed, so the user can begin modelling.

4. RESULTS

The various components of the system detailed in the previous section are functional, linked together within Simmetry3d and respond interactively within the Miller time metric. Users can make weir modifications by changing parameters in a window that appears alongside the visualisation, as shown in Figure 2.

![Figure 2. Weir modification parameter dialog](image-url)

By altering these and pressing an “apply” button, a weir model is generated for the user. In Figure 3, two different weir modification options are visualised, separated for clarity from a larger landscape model. The first option shows a weir with a specified height of 1.4m, and the model shows a green colour to indicate a “high” level of “fun” for canoeists. The assessment window to the right shows the BN output for the overall quality of the stretch of river by the weir for canoeists, and the number of people the electricity generated by a micro-hydro scheme could provide for. The second option shows a weir with a specified height of 0.4m and a stepped profile with the red colour indicating it offers a “low” level of “fun”. The second assessment shows little alteration to the river quality for canoeists (as the BN still needs further refinement), but a marked reduction in potential micro-hydro output.
Figure 3. Two weir modification options for a weir modelled in 3D with their assessments.

The weir modification system has also been placed into a larger landscape model (Figure 4) that allows the landscape context to be understood as the weir is altered. This model is demonstrating the use of a more realistic textured surface rather than representing the “fun” factor of the weir.

Figure 4. Interactive weir model in context within a larger existing landscape model.
5. DISCUSSION & CONCLUSION

Whilst much progress has already been made on the weir modelling tool, at the time of writing there are several key features that still need to be added before user testing is possible. These include modelling the placement and effect of fish / canoe passes in the 3D weir model and allowing changes made in the 3D visualisation to feed back to the associated parameters. The assessment system will also be extended to include a model that determines the effect of weir modifications on the critically endangered European Eel (Anguilla anguilla), providing an ecological indicator of design quality, and a cost model providing an economic indicator. Further work will take place to enhance the amount of information available to the user in the assessment window. With catchment wide processes affected by weir modification (e.g. eel migration), it is anticipated that the modelling tool will be extended to allow views of connected weirs within a catchment giving the user ability to test different weir modifications more strategically.

This research is part of the Urban River corridors and Sustainable Living Agendas (URSULA) project, which is conducting interdisciplinary research into the multifaceted problem of creating sustainable urban riverside landscapes. In the longer term, it is envisioned that the prototype weir design tool will provide the foundations for a more extensive system that will incorporate the future research outputs from URSULA into a more complex BN, allowing the visualisation and interactive modelling of more aspects of the riverscape, whilst providing an integrated assessment of these alterations.

ACKNOWLEDGEMENTS

This paper is based on work undertaken within the URSULA project, funded by the UK Engineering and Physical Sciences Research Council (Grant number: EP/F007388/1). The authors are grateful for this support, but they would like to note that the views presented in the paper are those of the authors and cannot be taken as indicative in any way of the position of the funders or of colleagues and partners. Any remaining errors are similarly those of the authors alone.

REFERENCES


Schroth, O., From Information to Participation - Interactive Landscape Visualization as a Tool for Collaborative Planning. Dissertation No. 17409. ETH Zürich, 2007.

Shaw, E.A., V. Kumar, D.N. Lerner, and E. Lange. Bayesian Belief Networks and Social Objectives: A Canoeing Case Study. iEMSs, Ottawa, 2010.

Bayesian Networks and Social Objectives: A Canoeing Case Study

Ed Shaw¹*, Vikas Kumar¹, David Lerner¹ & Eckart Lange²

¹ Department of Civil and Structural Engineering, Mappin Street, Sheffield, S1 3JD, UK
² Department of Landscape, Crookesmoor Building, Conduit Road, Sheffield, S10 1FL, UK
*email: ed.shaw@sheffield.ac.uk

Abstract: Much of the value of a river lies in the sociocultural values people attach to it. Though such 'social objectives' should form a fundamental part of integrated catchment management, they have typically been neglected due to the qualitative nature of many of the variables that predict them. One such social objective in the Don catchment, UK, is the management goal of maximising the recreational quality of rivers for canoeing within the constraints imposed by other management aims such as reducing flooding. Recreational quality is impacted by the modification of river weirs, an important management intervention in the catchment for which there are multiple potential options. An integrated catchment management decision support tool must predict the impact of the different modification options, raising the question; how to deal with the complex, uncertain, and subjective variables that determine a canoeist’s judgement of river quality? To tackle this issue, we employed a Bayesian Network (BN), which uses probability to describe relationships between variables. As probability can incorporate expert estimates, this enabled us to harness the knowledge of canoeist stakeholders. In this paper we discuss the experiences of building a BN with the collaboration of canoeists to predict how weir modification affects river recreational quality, and comment on implications for the overall utility of the approach for modelling social objectives. We conclude BNs are indeed suitable for modelling social objectives; probabilities capturing the uncertainty and subjectivity of variables such as ‘weir danger’. However the approach also has clear limitations. Though the canoeists found most parts of BN construction engaging, this was not so when eliciting judgements of probability, for which the use of questionnaires made it an abstract process. A further issue was that the number of questions needed in the probability elicitation stage increased dramatically with BN complexity. Interpolating probabilities from a limited number of questions is a partial solution, but even so the high number of questions still necessary was a barrier to completing this step. Ultimately this will constrain the potential complexity of BNs that require expert knowledge to define probabilistic relationships.

Keywords: Integrated catchment management; Bayesian Network; decision support; social objectives; canoeing
1. INTRODUCTION

A large component of the value of a river lies in the sociocultural values people attach to it. These include aesthetic, recreational, cultural and spiritual perceptions of value in the existence of, or in an individual's experience with, elements or states of an ecosystem. Collectively these values are known as non-use or sociocultural ecosystem services [Millennium Ecosystem Assessment, 2005]. Predicting how river interventions affect these values must form a fundamental part of integrated decision support models built to support urban river management. However, such 'social objectives' have tended to be neglected from process based models due to the qualitative nature of many of the variables that need to be included within the model structure [Holzkämper et al., submitted]. In addition water management policy has traditionally focussed on physical objectives such as chemical water quality meaning there has been a lack of a reason why social objectives should be considered in the decision making process, though this is currently changing with the introduction of the Water Framework Directive in the European Union which provides some scope for social objectives to be traded-off against environmental objectives. The integration of human perceptions and values with catchment management models is now considered a key research direction in developing model based tools to aid water management [Borowski & Hare, 2007]. Yet if this integration of social objectives is to be achieved, then the problems that come with it; complexity, uncertainty, and subjectivity of the relationships between variables, must be overcome.

One solution is to take the conceptual modelling approach of a Bayesian Network (BN). In the last decade BNs have increasingly been applied to environmental management problems, and recently also to integrated water management issues [Ames et al., 2005; Barton et al., 2008; Kumar et al., 2008]. BNs use a graphical cause-effect network using probabilities to describe the conditional relationships between the dependent variables (known as child nodes) and their controlling variables (parent nodes). Each variable is described as a range of discrete states so that the model user can specify the state of parent nodes from which the BN predicts the likelihood that each child node state will result. Child nodes may also be parent nodes of further child nodes creating a chain so that the probabilities propagate through the system until a basal child node is reached. By using probabilities, relationships between variables can be derived from expert judgement, capturing uncertainties in expert knowledge.

These qualities make a BN approach suitable for modelling the relationships between weir modification and the recreational quality of rivers for canoeing. The Graphical User Interface (GUI) of a cause-effect network means the model is relatively transparent, with assumed relationships between variables clear to see. By describing relationships between variables probabilistically, mentally held perceptions and judgements of value can be described, allowing the views of stakeholders to be harnessed. Also, probabilities can capture the uncertain nature inherent in subjective variables, where potentially every individual can hold a unique view.

In this paper we discuss the experiences of building a BN that addresses a high profile management problem in the Don Catchment, UK; that of the impact of weir modification on the quality of the River Don for canoeing. We use the experiences of building the BN with the collaboration of canoeing groups to explore issues regarding the utility of the approach for modelling social objectives. What did and didn’t work well in the process of building the model is examined and the suitability of BNs is commented on.

2. CASE STUDY

A current major management problem facing decision makers in the Don Catchment, UK, and in many other catchments around the world is that of the modification of the many weirs that impound river systems to mitigate their negative environmental impacts. The Don Catchment has a particularly high number of weirs as it was a historically important centre of water powered metal working. Despite most now being obsolete, weirs are still a
common sight in the Don Catchment (see Figure 1). Many groups would like to modify the weirs, with some anglers wanting to install fish passes; potential victims of flooding interested in removing weirs; proponents of hydro-electricity wishing to install micro-hydro schemes; enthusiasts of natural history aspiring to restore the river to a more natural state and enthusiasts of the river’s heritage wanting to conserve weirs for their historic value.

All these different modifications have potential to affect the recreational quality of the rivers for another stakeholder group; that of the local canoeists. Canoeing is a popular leisure activity in the city of Sheffield through which the River Don flows. The various weir modification options have both positive and negative effects on the recreational quality of this activity, though the impact is subjective; based on the judgement of the local canoeing groups. If weir modification decisions are to be aided by an integrated decision support tool, then predicting this impact on recreational quality must be included as a social objective.

![Figure 1. The distribution of weirs in the Don Catchment.](image)

### 3. THE CONSTRUCTION OF THE CANOEING BN

In the following sections we describe the four steps to the process of constructing the BN built to predict the impact of weir modification on the recreational quality of rivers. This process of deriving expert judgement to construct the model is known as knowledge elicitation.
### 3.1 Identification of model variables and structure

The first step of building a BN is to identify the variables and structure of the cause-effect network. Two workshops were held to achieve this step, with canoeists from local groups invited to take part. Six local canoeing groups were identified using the internet or through contacts and invited to attend, of which three agreed to send representatives providing a total of five participants. It is not clear why three groups didn’t participate but for these we only had email contact details and it might reflect the impersonality, and frequency of junk mail associated with this communication medium. The process of creating the structure started as an informal group brainstorming exercise, where the canoeists were encouraged to think about how a weir could affect the quality of a river for canoeing, a variable that was to form the basal child node in the network. The canoeists identified that a weir could affect quality in two ways; as a source of danger, and as a source of fun (descending weirs was considered exciting), so these became parent nodes to the child node. The process was repeated, this time by thinking about what factors determined the danger and fun of the weir, and so on until measurable quantitative variables were identified that could form the input nodes to the structure. The next step was to factor in the impact of weir modification options (changing weir height, steepness, orientation, profile of weir face, and installation of microhydro, canoe pass and fish passes), so these were described to the canoeists who then discussed as to whether they would influence any of the model’s variables. Lastly, the nodes least important to the model were examined and dropped if it was thought that they wouldn’t make much difference to the output of their child nodes (see Figure 2 for an overview of the evolution of the network).

**Figure 2.** The evolution of the BN structure in the identification of model variables and structure stage. a) The basal child node of river quality, b) Weir fun and weir danger were identified as the two main ways weirs impact on river quality for canoeing, c) Weir danger was found to be controlled by the weir drawback (cyclical river flow at the base of the weir) and risk of physical injury descending the weir, d) The final canoeing BN structure with all remaining parent nodes and linkages identified.
We consider this stage of model construction to have worked very well. Those involved found the cause-effect network intuitive, and seemed very engaged; enjoying talking about their hobby. The finished cause-effect network could then be directly entered into the BN.

3.2 Categorisation of the variables

One of the characteristics of BNs is that model variables are described as discrete states e.g. weir height can be described as tall, medium and small. This means that the variables have to be defined so that they mean the same thing to all involved in the use and construction of the model, and so that the threshold between states is at a point that results in a meaningful change in the predicted states of the dependent variable. For example, changing weir height from 2m to 3m tall makes a big difference to the risk a canoeist faces descending a weir, but a change in height of 10m to 11m makes little difference.

This process was also conducted at the same workshops and was aided using prepared visual aids for those variables anticipated likely to be important to the canoeists (e.g. see Figure 3). Once again, the canoeists completed this process with relative ease and seemed to enjoy the process.

3.3 Probability elicitation

In contrast to the previous steps, the process of eliciting the probabilities caused problems for the model construction. This stage requires the completion of probability tables for each relationship between a child and its parent nodes (e.g. see Figure 4), where the expert estimates the likelihood of child node being in each state for every permutation of parent node states. However, as the complexity of the model increases, then the number of permutations increases exponentially, so that for example, the sub-network of weir fun with its seven parents has 2916 potential combination of states. This would require the expert to answer 2916 questions to fill out the conditional probability table for this sub-network alone! Clearly this is not acceptable, and for this reason we used an approach developed by Kumar et al. (in prep.) which only requires the expert to provide estimates of probabilities for a limited number of parent node states, from which the remaining probabilities are interpolated. Using this method we reduced the number of questions down to 120.

After the workshops, all participants agreed to complete the questionnaires, which had to be constructed using the outputs of the workshops. Of the five copies of the questionnaire sent...
to participants, and despite repeated reminders and offers of help, not a single copy was returned. As this first attempt at probability elicitation had failed, a change in tactic was tried. Supposing that the number of questions on the questionnaire was still unreasonably long, and the activity was asocial as compared to the workshops the participants had enjoyed, new canoeing contacts were made and were personally supervised when filling out the questionnaires. The process of completing the questionnaires was very time consuming, taking between 2 to 5 hrs. Participants needed to spend time thinking about the questions and often referred to the supervising helper for further explanation, and to check that they had interpreted the questions correctly.

3.4 Model validation

At the point of writing, the model still needs to be validated. There are two planned parts to this stage, both requiring the participation of canoeists. Firstly, the canoeists are to be introduced to the canoeing BN compiled using the data derived from their questionnaires. They will be encouraged to play with the model, testing different permutations of weir states, and to examine if the predicted child node states meet their expectations. New values will be entered into the model if they don’t feel that the predicted outcomes are satisfactory. Once this refinement has occurred, the canoeists will be invited to canoe a length of the River Don in Sheffield. For each weir descended, they will each rate its characteristics e.g. weir fun and danger, and these compared to the weir fun and danger states predicted by the BN. Involvement of canoeists who didn’t participate in model construction will also be sought in order to test how reflective of the model is of the wider canoeing community. We believe this should provide a robust assessment of the canoeing BN’s validity.

4. DISCUSSION AND CONCLUSION

In creating a BN to model the impact of weir modification on the quality of the River Don for canoeing, we believe it has been shown that this technique is appropriate for modelling social objectives. With a working model and clear path to validation and improvement, it is expected that the predictive ability of the final version will be fairly reliable. By using conditional probabilities to describe relationships between variables, the approach successfully captured subjective concepts such as ‘weir danger’ or ‘weir fun’, and the inherent uncertainty in the relationships between these qualitative variables.

As stakeholders must provide the knowledge describing the relationships between variables when modelling social objectives, then it is vital that the process of constructing the model must be stakeholder friendly. In this regard the BN was generally successful. The canoeists found the steps of identifying the variables and structure of the model, and categorising the variables as engaging and relatively intuitive processes. However as in the experience of

---

### Table 1: Conditional Probability Table for Weir Danger and Weir Fun

<table>
<thead>
<tr>
<th>Weir Danger</th>
<th>Weir Fun</th>
<th>River Contact of Weir</th>
<th>Excellent</th>
<th>Good</th>
<th>Medium</th>
<th>Poor</th>
<th>Very Poor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>High</td>
<td>Little</td>
<td>0.400</td>
<td>0.300</td>
<td>0.200</td>
<td>0.100</td>
<td>0.000</td>
</tr>
<tr>
<td>Low</td>
<td>Low</td>
<td>Little</td>
<td>0.100</td>
<td>0.200</td>
<td>0.300</td>
<td>0.400</td>
<td>0.000</td>
</tr>
<tr>
<td>Medium</td>
<td>High</td>
<td>Little</td>
<td>0.300</td>
<td>0.300</td>
<td>0.200</td>
<td>0.100</td>
<td>0.000</td>
</tr>
<tr>
<td>Medium</td>
<td>Low</td>
<td>Little</td>
<td>0.200</td>
<td>0.300</td>
<td>0.300</td>
<td>0.200</td>
<td>0.000</td>
</tr>
<tr>
<td>High</td>
<td>High</td>
<td>Little</td>
<td>0.200</td>
<td>0.300</td>
<td>0.300</td>
<td>0.200</td>
<td>0.000</td>
</tr>
<tr>
<td>High</td>
<td>Low</td>
<td>Little</td>
<td>0.100</td>
<td>0.200</td>
<td>0.300</td>
<td>0.400</td>
<td>0.000</td>
</tr>
</tbody>
</table>

---

**Figure 4.** a) Subnetwork of canoeing BN and b) corresponding conditional probability table
Henriksen et al. [2007] who found aspects of BNs difficult for non-experts to understand, the canoeists also struggled with one part of creating the BN. In our case this was linked to the probability elicitation stage, required to create the conditional probability tables. While we are not certain why no questionnaires were initially returned, it seems reasonable to assume that it was due to the combination of the length and the abstract nature of the questionnaire. Trying to envisage multiple states of a set of parent nodes described textually is difficult, and indeed those supervised to fill out the questionnaire found it challenging. The large number of questions required in the questionnaire, an outcome of dependence on expert knowledge, was also demanding of the stakeholders. As the number of variables in a BN increases linearly, the number of questions needed to fill out the conditional probability tables increases exponentially. Consequently the canoeing BN would not have been practical to build without the method of Kumar et al., [in prep.] to interpolate probabilities from a limited number of questions. Even so the number still required was large and a barrier to completing the knowledge elicitation stage. For some BNs relationships between physical variables can be described using empirical data or data from other pre-existing models, but neither were available for any of the relationships in the canoeing BN. In hindsight more effort should have been taken to simplify the model by removing less important variables, though there would have been an erosion of model usefulness. Ultimately the excessive demands put on experts by questionnaires will constrain the potential complexity of BNs requiring expert knowledge to derive conditional probabilities. Methods need to be developed to increase the ease of completing the probability elicitation process to reduce the burden on those providing the expert knowledge.

Post-construction BNs are also relatively user friendly for stakeholders; important as involvement of stakeholders in the decision making process is considered important if decisions are to be fair and sustainable [Soncini-Sessa, 2007]. With a Graphical User Interface (GUI) of an interactive cause-effect network, BNs are intuitive way to explore how a weir modification options affect river quality for canoeists, and this transparency builds trust. By displaying the likelihood that certain states will be realised due to weir modification, then the inherent uncertainty in the model is communicated to the user (though BNs don’t differentiate between uncertainty in the system and in stakeholder understanding). As the BN structure and variables are defined by stakeholders, then it is automatically constructed at a level appropriate to the stakeholders and with relevant indicators defined. This user friendliness makes it an inclusive tool, meaning BNs can be used to promote understanding between various stakeholder groups, as well as providing information to support decision making. Even so, there is still some scope for the improvement of the GUI. Consider the canoeing BN with its ten input variables. Mentally combining these attributes and visualising actual weir and river they represent is difficult. Given that water managers do not have time to learn or teach the public how to use models that can be difficult to understand [Borowski et al., 2007] then the reliance on abstract mental visualisation is probably a barrier to BN uptake. To improve the user friendliness of the GUI is one reason why Gill et al., [2010] have linked the canoeing BN introduced in this paper to interactive visualisation software. The visualisation communicates what a weir based on user selected input BN states would look like, and allows the user to interact with the weir e.g. changing height, designing different weir modification scenarios, making the BN more accessible to participants in decision making.

However, in addition to the knowledge elicitation difficulties at the probability elicitation stage, there were some other drawbacks in using the BN approach for modelling the canoeing social objective. BNs cannot easily deal with spatial information, meaning that while the canoeing BN could predict how a single stretch of river quality was affected by weir modification, it couldn’t consider the cumulative effect of multiple weirs on longer stretches of rivers. Other approaches must be used in conjunction with BNs if spatial issues are to be dealt with. BNs may also have a limitation in their function as tools for decision makers. Borowski et al., [2007] found there was a need for models that helped the user to develop new solutions to management problems. As discrete management options are predefined in the BN then scope for users to design new management options is restricted. This limitation can be partially overcome by allowing and encouraging users to modify the BN, incorporating new ideas after its initial construction. If the stakeholders don’t know
what the effects of a new modification variable will be, then this means that it cannot be included until this understanding is gained. A further issue is that the categorisation of the variables into discrete states makes the BN unresponsive to small changes in input variables, meaning the tool is not useful for fine tuning designs.

Lastly, some questions still remain on applicability of the BN approach for modelling of social objectives. While the relationships and variables involved in determining the effect of weir modification on the quality of the river for canoeing were clear cut, this may not be the case for other social objectives. Indeed, some cultural values such as perceptions of spiritual or aesthetic value may resist being reduced down to a collection of variables, either because stakeholders are unwilling to do so, or are simply unable. Additionally, for some social objectives there are a wide range of perceptions and judgments of values making it difficult to boil down the model structure into something that is agreed upon by all stakeholders. The elicited probabilities may have so much uncertainty that is too great for a BN to give useful predictions. In order to answer these questions, further research is required on using BNs to predict social objectives.

ACKNOWLEDGEMENTS

This work was jointly funded by the UK Engineering and Physical Sciences Research Council and Economic and Social Research Council. We would like to thank the many canoeists and researchers in the URSULA project who have given time and help making this work possible.

REFERENCES


Gill, L., V. Kumar, E. Lange, D.N. Lerner, E. Morgan, D. Romano and E.A. Shaw, An interactive visual decision support tool for sustainable urban river corridor management. iEMSs, Ottawa, 2010.


Holzkaemper, A., V. Kumar, B. Surridge, A. Paetzold, S. Lautenbach, R. Harris and D.N. Lerner, How to develop a simple tool for a complex task – Decision support for integrated catchment management, (submitted), 2010.


Kumar, V., A. Holzkaemper and D.N. Lerner, Capturing critical state effect in parameterisation of Bayesian network (in prep)


Modelling the ecological impact of discharged urban waters upon receiving aquatic ecosystems. A tropical lowland river case study: city Cali and the Cauca river in Colombia

Holguin Gonzalez J.a, Goethals P.L.M.b

a Laboratory of Environmental Toxicology and Aquatic Ecology, Ghent University, Jozef Plateaustraat 22, B-9000 Ghent, Belgium. jaholgu@hotmail.com
b Laboratory of Environmental Toxicology and Aquatic Ecology, Ghent University, Jozef Plateaustraat 22, B-9000 Ghent, Belgium. peter.goethals@ugent.be

Abstract: The Cauca river is one of most severe cases of pollution for domestic and industrial wastewater discharges in Colombia, principally when it crosses the industrial cities of Cali and Yumbo. The rapid urbanization and major economic development in the Cauca river’s geographical valley has led to dramatic degradation of the environment and increased health risks due to inefficient processing of the increased pollutant load effluents and solid wastes. The city of Cali which is the main urbanization center, with more than two million inhabitants and limitations of the treatment of its wastewaters, discharged in the year 2005, 75 tons of BOD₅ per day. These discharges of wastewater are producing an increasing deterioration of the water quality of the Cauca river. This pollution problem is critical after the river crosses the city of Cali, especially during dry season (low flows), when pollution can reach values of 7.5 mg/l of BOD₅ and concentrations of Dissolved Oxygen (DO) near to zero (0) mg/l. Low DO levels affect the ecosystem equilibrium and the functioning and survival of biological communities. For this reason, the main objective of this research was to contribute to the integrated water quality management of the Cauca river, developing a mathematical model to investigate the ecological quality of this river under actual conditions as well as after different restoration actions. The approach followed was to build habitat suitability models (statistical models) that allow predicting the presence and the abundance of macroinvertebrates in this river under different conditions. An integration of these ecological models with the hydrodynamic and physical-chemical water quality model MIKE11 was performed. The integrated ecological model allows to model and assess the ecological impact of wastewater discharges into the Cauca river and to calculate the needed reductions in discharges of organic matter to meet biological quality criteria in this river.

Keywords: Integrated ecological modelling; habitat suitability models

1. INTRODUCTION

One of the worldwide problems that affect the quality of water resources, has been their use such as receiving aquatic ecosystems of controlled or uncontrolled discharges of wastes from agricultural, urban or industrial activities. These discharges can potentially affect human health and aquatic life, limit water uses, affect riverine ecology and cause loss of amenity. Furthermore, scarcity and misuse of fresh water pose a serious and growing threat to sustainable development and protection of the environment. These problems will
intensify unless effective and concerted actions are taken. Challenges remain widespread and reflect severe problems in the management of water resources in many parts of the world. The optimal balance between the different stakeholder activities needs a very deep insight in the integrated water management. In this context, models can show the limitations of the self-cleaning capacity of water resources. Indeed, water quality modelling is an effective tool to investigate the ecological situation of surface water resources. Nevertheless, until now ecological models have rarely been used to support river management and water policy.

The biotic component of an aquatic ecosystem can be considered as an ‘integrating-information-yielding unit’ for assessment of its quality. Biological communities also integrate the effects of mixed types of stress and in certain cases already respond before analytical detection allows for, [De Pauw and Hawkes, 1993]. Among the biological communities, the macroinvertebrates are by far the most frequently used group for bioindication in standard water management, because they are ubiquitous and abundant throughout the whole river system and they play an essential role in the functioning of the river continuum food web, [Goethals, 2005; Giller and Malmqvist, 1998].

The Cauca river is one of most severe cases of contamination for domestic and industrial wastewater discharges in Colombia. The main urbanization center and source of pollution that affects this river is the city of Cali, with more than two million inhabitants and limitations of the treatment of its wastewaters. This city discharged during the year 2005 around 75 tons of BOD₅ per day, which is around 38% of the total of wastewater discharged load of BOD₅ per day in the Valle del Cauca department. The sewer system of Cali has limitations in the operation and only 40% of the total flow of the wastewaters generated by the city is treated by primary treatment. The rest of the wastewater generated by the city does not receive any type of treatment and it is discharged to the Cauca river. These discharges of wastewater are producing an increasing deterioration of the water quality of the Cauca river. This pollution problem is critical after the river crosses the city of Cali, especially during dry season (low flows), when pollution can reach values of 7.5 mg/l of BOD₅, concentrations of Dissolved Oxygen (DO) near to zero (0) mg/l, values of Faecal and Total Coliforms in the order of 2.4x10⁸ NMP/100ml and critical values for some heavy metals [CVC and Univalle, 2004a; EMCALI and Univalle, 2006].

The mathematical modelling approach has been used to support the generation of policy, plans and projects for the water quality improvement of the Cauca river and the control of wastewater discharges since 1972. During this process, limitations in the knowledge of software and also in the information required in terms of water quantity and quality have been faced. During the last decade (1997-2007) in the framework of the Cauca River Modelling Project (PMC), the water quality model software MIKE11 was used to simulate the hydrodynamic and water quality of the river. This modelling approach allowed a very deep insight of the processes that occur in the river under dynamic conditions, such as temporary variations of flows and polluting loads. However, this approach just considered physical-chemical parameters, therefore biological components of the aquatic ecosystem were not taken into account. For this reason, in order to have a robust, reliable and effective tool to support river management and water policy in the Cauca river, it is necessary to develop and to apply an ecological modelling approach that integrates the hydrodynamic and physical-chemical water quality model MIKE11, with an ecological water quality model developed for this river, which allows to investigate the ecological quality of the Cauca river under actual conditions as well as after different restoration actions.

2. MATERIALS AND METHODS

2.1 Study area

The Cauca river is the second most important river in Colombia and the main hydric resource of the Colombian southwest. The Cauca river geographical valley, situated in the high basin, is especially important for the country’s development and economy. A significant part of the south-western manufacturing industry, the paper production
industry, the sugar cane agricultural industry and part of the coffee zone are located in this area. This study focuses on the stretch of the Cauca river which is situated in the geographical valley, which extends from the station La Balsa until Anacaro (Figure 1) with a total length of 389 km. The most important water quality problems can be found in this zone, especially in the stretch close to the city of Cali.

The Cauca river flows for 445 km in its geographical valley and descends from a height of 1000 meters (Salvajina dam) to 900 meters above sea level. This stretch of the river has an average width of 105 meters and it can fluctuate between 80 meters in the high part of its course (Salvajina dam – La Balsa) to 150 meters in the low part (Anacaro – La Virginia). The depth can vary between 3.5 and 8.0 m. The longitudinal profile of the river shows a concave shape with a hydraulic slope which oscillates between 7x10^{-4} m/m and 1.5x10^{-4} m/m [CVC and Univalle, 2007b].

2.2 Data and information collection to develop and validate models

The database used in this research corresponds to the information collected and analyzed during the years 1997-2007 by the environmental authority in the Cauca river’s geographical valley called CVC and the Cauca River Modeling Project (PMC), [CVC-Univalle, 2004b and CVC-Univalle, 2007a and 2007b]. The database used for the implementation of the ecological models for predicting macroinvertebrates in the Cauca river was selected considering the simultaneousness of biological, physical-chemical and hydraulic measurements in the sample stations. This information corresponds to the years 1997 (high flow conditions), 2001 and 2004 (low flow conditions).
Benthic macroinvertebrates were collected in the Cauca river considering the type of substrate: stone substrate, sand substrate, mud substrate, trunks and floating vegetation substrate. Samples were collected in the middle and at the borders of the transversal section. Thus all type of macroinvertebrate habitats were considered during the monitoring. After samples were collected in each zone and substrate, they were preserved, labelled and transported to CVC’s environmental laboratory where finally an identification and classification of macroinvertebrates, mainly at the taxonomic level of genus (phylum, class, order, family and genus) was carried out [CVC-Univalle, 2004b]. Simultaneously with the biological monitoring, water samples were taken for the analysis of physical-chemical parameters. These water samples were preserved at 4°C and transported to the CVC’s environmental laboratory for the analysis. The hydraulic information reported by the sample stations during the monitoring campaigns was mainly provided by CVC from hydrometric stations. When this information was not available interpolations with the hydrodynamic model (MIKE11) were performed.

2.3 Water quality modelling software used in the catchment of the Cauca river

Hydrodynamic and physical-chemical water quality models: In the framework of the Cauca River Modelling Project (PMC), a physical-chemical water quality and hydrodynamic model (MIKE11) was implemented for the Cauca river. This model was calibrated and verified for dynamic flow conditions and reproduces in acceptable form the values of DO, BOD\textsubscript{s}, temperature, flow, depth and velocity in the monitoring stations of the Cauca river, considering hourly fluctuations. [CVC-Univalle, 2007a and 2007b]. The model is conformed by 62 cross sections, 2 external boundaries (La Balsa and La Virginia), 96 internal boundaries which include 27 rivers and streams, 9 municipal wastewater discharges, 12 industrial wastewater discharges and 37 water extraction sites. Each internal boundary was represented like a lateral extraction or discharge. Water quality modelling was carried out in Level 1 of MIKE11, which includes temperature, BOD\textsubscript{s}, and DO as state variables. In the framework of the PMC project two monitoring campaigns with calibration and verification purposes for the water quality model were carried out during the months of August of 2003 and February of 2005. These campaigns had a duration of respectively five (5) and four (4) days, a monitoring period between 12 and 24 hours per day, with a measuring frequency between 30 and 60 minutes for field parameters (flow, DO, temperature, conductivity and pH) and between six (6) and eight (8) hours for laboratory parameters (BOD\textsubscript{s}, COD, TSS).

Ecological models for predicting macroinvertebrates in the Cauca river: When performing ecological modelling two approaches can be followed which are mechanistic and data driven modelling. In the case of the Cauca river the water quality model MIKE11 allows to calculate water quality variables such as temperature, BOD\textsubscript{s}, and DO. However, there is a lack of information about processes associated with particulate organic matter and nutrients. This lack of information limits the use of ecological mechanistic models (i.e. food-webs) for the Cauca river, for that reason data driven models such as habitat suitability models (e.g. statistical models) for predicting macroinvertebrates were implemented in this research. These models can be used as a first approach for modelling composition of macroinvertebrate communities, and they can be useful for this more detailed type of calculations, where direct relations between a set of predictor variables (physical-chemical and hydraulic) and biological species are calculated, without incorporating feedback loops. The approach followed was to build statistical models called Generalized Linear Models (GLMs) (parametrical method, that provides users with a conventional mathematical function), which are mathematical extensions of linear models for non-linearity and non-constant variance structures in the data. GLMs are better suited for analyzing ecological relationships, which can be poorly represented by classical Gaussian distributions.

Logistic regression (i.e. GLM with logit link function and binomial error distribution) is the most frequently used modelling approach of the GLM techniques, for predicting the probability of species occurrence [Manel et al., 2000; Pearce and Ferrier, 2000; Ahmadi-Nedushan, 2006], because a single record of presence or absence of the target species can
be considered to be a binomial trial with a sample size of 1. This method has become a favourite tool in habitat modelling when the species information is given as presence/absence data, because this information is comparatively easy to collect in the field, even when the zero data set has to be created afterwards by a different sampling strategy. The model estimates the probability of a positive response occurring given a set of explanatory environmental variables (e.g. depth, velocity, DO, substrate, cover). Based on the presence-absence data, a response curve of a species describes the probability of the species being present, \( p \), as a function of environmental variables. The response variable is transformed by the logit link function, which transforms bounded probabilities (between 0 and 1) to unbounded values [Ahmadi-Nedushan, 2006].

In order to select the best multiple logistic regression model (MLRM) for predicting the probability of macroinvertebrate species occurrence in the Cauca river the Akaike’s information criterion (AIC) was used. Logistic models were fitted using the maximum likelihood method [McCullagh & Nelder, 1989] with backwards elimination to select the final predictor variables. The step function, implemented in the statistical software XLSTAT version 2009 used in this research provides a procedure for this purpose using the AIC; this is a penalized version of the likelihood function in which the best model is given by the lowest AIC value. Additionally, the Receiver Operating Characteristics-ROC curve was used to evaluate the performance of the MLRM by means of the area under the curve (AUC) and to compare several models together. The AUC, which ranges form zero (0) to one (1), corresponds to the probability such that a positive event has a higher probability given to it by the model than a negative event [Hosmer and Lemeshow, 2000].

For modelling abundance of macroinvertebrates (numbers of organisms) at new sites and/or future times at the Cauca river, quasi-Poisson regression (other modelling approach based on GLM techniques) was used. Often, data of organisms come in the form of counts, and in ecological modelling the idea is to relate these counts to environmental conditions. Count data in ecology are often “overdispersed” (i.e. for any data set or model the variance exceeds the mean) which is a limitation for Poisson regression. A common way to deal with overdispersion for counts is to use a GLM framework [McCullagh and Nelder, 1989], where the most common approach is the quasi-Poisson regression, because it is widely available in software and it generalizes easily to the regression case [Ver Hoef, and Boveng, 2007]. The regression constants in the QPRM models were estimated by means of the maximum quasi-likelihood method using the statistical software S-PLUS version 6.1. In order to select the explanatory variables in the QPRM, changes in goodness of fit statistics were used to evaluate the contribution of subsets of explanatory variables to a particular model. The deviance (i.e. how much variation is left), defined to be twice the difference between the maximum attainable log likelihood and the log likelihood of the model under consideration, was used as a measure of goodness of fit.

3. RESULTS AND DISCUSSION

Mathematical models are widely applied in science. The application of models in ecology is almost compulsory if we want to understand the function of such a complex system as an ecosystem [Jorgensen and Bendoricchio, 2001]. However, the knowledge of ecological processes in ecosystems and the information available for a very deep insight of these processes have been much less developed and accessible compared with other science fields such as hydrodynamic or hydro-morphologic and physical-chemical processes. This situation is occurring in the Cauca’s river context, where the CVC has many hydrometric stations with hourly database, a few automated measurement stations for continuous water quality monitoring and historical information mainly based on discrete monitoring campaigns, with time intervals of months, however biological information for river quality assessment has been hardly recollected during few years. Therefore, the development of mathematical models for predicting biological communities (the aim of this research) together with the study of biological indicator species are complementary tools for river quality assessment and contribute to the integrated water quality management of this river.
Taken into consideration the aim of this research, which is to build predict models for macroinvertebrates communities present in this river under different conditions by means of an integration with the MIKE11 model, the ecological statistical models proposed in this research included during its calculation a data set of explanatory variables which could be calculated using the model MIKE11 (i.e. temperature, BOD$_5$, DO, flow, depth and velocity). An important consideration for all types of models is which and how many explanatory variables should be included in the model. If there are too few variables, the model will not be able to explain much of the variation. On the other hand, if there are too many variables, then the model will be too specific for the current data set [NIVA, 2007].

A total of three (3) macroinvertebrate predictive models at the taxonomic level of Order, were selected for constructing the ecological models. Ephemeroptera and Trichoptera Orders (pollution sensitive benthos, which belongs to Phylum Arthropoda and Class Insect) as biological indicators for good water quality conditions and Haplotaxida Order (pollution tolerant benthos, which belongs to Phylum Annelida and Class Oligochaeta) as biological indicator for polluted water with high organic matter content. At this taxonomic level these models could be considered too coarse in their predictive ability, however, considering the limitations regarding the database and the financial resources available for biological assessment at the CVC, these models can be thought as a first approach for ecological modelling at the Cauca river, that can be improved afterwards with more detailed data.

The results of the ecological modelling showed that the best MLRM for Ephemeroptera included DO, flow and depth as environmental predictor variables, the best MLRM for Trichoptera included the variables BOD$_5$, velocity, flow and depth and the best MLRM for Haplotaxida included BOD$_5$ and velocity as predictor variables. The assessment of the MLRMs reliability showed that the models for Ephemeroptera (AUC=1), Trichoptera (AUC=1), and Haplotaxida (AUC=0.926) correctly discriminates between occupied (presence) and unoccupied (absence) sites in the dataset. On the other hand, the best QPRM for Ephemeroptera included DO, velocity, flow and temperature as environmental predictor variables, the best QPRM for Trichoptera included the variables DO, depth, flow and temperature and the best QPRM for Haplotaxida included DO, depth and flow as predictor variables. Regarding the predictive validation procedure for QPRMs, it was found that in general the models reproduce with good precision the tendencies and the maximum and minimum values of the abundance data for each macroinvertebrate (i.e. Ephemeroptera, Trichoptera and Haplotaxida) with high coefficient of determination ($R^2$) values ($0.866 < R^2 < 0.998$). An example of the results obtained for the QPRM considering the complete database can be seen in the Figure 2. The same kind of models and graphs were developed for the database divided according to high flow and low flow conditions.

![Figure 2. Results of the QPRM for predicting abundance of Ephemeroptera in the Cauca river](image)

The results obtained in this research showed that flow velocity was the variable most important in four (4) of the six (6) ecological models developed in this research, followed by the depth in two (2) models and DO in one (1) model. Additionally, in five (5) of the six (6) predictive models the DO and the BOD$_5$ as indicators of organic pollution, were the first or the second most important variables. These results are in accordance with the
literature, which reports that other factors than water quality are also important determinants of benthic communities. Of these the related factors of current velocity and nature of the substratum are overriding ones determining the nature of the community, especially in relation to invertebrates [Goethals, 2005]. Since these factors differ along the river in different zones, different communities become established at different sites with the same water quality [Giller and Malmqvist, 1998].

Using the model MIKE11 implemented in the framework of the PMC Project, and the ecological models developed in this research for predicting the (likelihood of) occurrence and the abundance of macroinvertebrates in the Cauca river, some applications were carried out that allowed to study the effects in the water quality of the river generated by the plans and actions for the pollution control proposed by the environmental authorities, the municipalities and the industries in the Cauca river’s geographical valley. The scenarios considered were: reference situation scenario (year 2005); very optimist scenario (year 2015), very pessimistic scenario (year 2015); and the scenario of water quality objectives proposed by the environmental authority (CVC) (year 2015). Profiles of average concentrations of DO, BOD₅, temperature and flow at the Cauca river, were made for each scenario considering the results obtained with the physicochemical and hydrodynamic model MIKE11. In general, those scenarios showed that in spite of the reduction of the total organic matter load discharged into the Cauca river’s basin considered (even in the very optimist scenario), the DO concentrations in the station Paso de La Torre (the most critical zone in terms of pollution, just after the city of Cali) never reach values higher than 2.6 mg/l. Additionally, these DO values are still lower than the minimum standard value established by the Colombian Decree 1594/84 for different uses of the water resource, which means, smaller than 70% of the DO saturation concentration (5.2 mg/l for this river).

The application of the integrated ecological modelling of the Cauca river showed that the MLRMs and QPRMs predicted well the ecological impact of the scenarios for pollution control in the Cauca river’s basin. Thus, in the scenario with the highest pollution reduction an improvement of the water quality of the Cauca river is achieved, which is represented with the presence and/or an increase of the number of pollution sensitive benthos (i.e. Ephemeroptera and Trichoptera) and the absence and/or a decrease of the number of pollution tolerant benthos (i.e. Haplotaxida). On the other hand, if the worst pollution condition scenario is considered a deterioration of the water quality is obtained, which is represented with the absence and/or a decrease of the number of pollution sensitive benthos and the presence and/or an increase of the number of pollution tolerant benthos. Finally, if the scenario of water quality objectives proposed by the CVC is considered, which is an intermediate scenario in terms of pollution reduction, a water quality improvement is achieved in some stations, but there are other stations that still show water quality deterioration (between Puente Hormiguero and Juanchito). This indicates that the pollution reduction proposed by the CVC in this scenario is not enough for increasing significantly the number of pollution sensitive benthos or for decreasing the number of pollution tolerant benthos, and therefore a good biological water quality is not reached in this river.

4. CONCLUSIONS AND RECOMMENDATIONS

The statistical models proposed in this research, allow predicting the occurrence and the abundance of macroinvertebrates for the Cauca river under different hydraulic and physical-chemical water quality conditions. This research demonstrated the high potential of the integration of hydrodynamic and physical-chemical water quality models with ecological models, in helping to get insight in aquatic ecosystems, for finding what is necessary to improve in the integrated water management and policy development. The integrated ecological model proposed in this research is a powerful operational tool, which allows to model and to assess the ecological impact of wastewater discharges into the Cauca river and can help to calculate the needed reductions in wastewater discharges of organic matter to meet biological quality criteria in this river.
Today river quality assessment is mainly based on discrete monitoring campaigns, with time intervals of several hours, weeks, months or even years. For the study of highly dynamical processes such sampling schemes are often insufficient to make a reliable assessment of the river status. In those cases, the application of automated measurement stations for continuous water quality monitoring together with the study of biological indicator species are complementary tools for river quality assessment. Having relatively long life cycles and being confined for most part of their life to one locality on the river bed, aquatic macroinvertebrates act as continuous monitors, integrating water quality over a longer period of time (weeks, months, years). Biological indicator species are unique environmental indicators as they offer a signal of the biological condition in a watershed.

REFERENCES

CVC-Corporación Autónoma Regional del Valle del Cauca, Univalle-Universidad del Valle, Modelling scenarios for defining pollution control plans in the Cauca’s river basin, Streach La Balsa – Anacaro, Volume XII, Cauca River Modelling Project (PMC), Phase III, Cali, Colombia, 2007a.
CVC-Corporación Autónoma Regional del Valle del Cauca, Univalle-Universidad del Valle, Optimization of the water quality simulation model for the Cauca river, Streach La Balsa–Anacaro, Volume XIII. Cauca River Modelling Project (PMC), Phase III, Cali, Colombia, 2007b.
CVC-Corporación Autónoma Regional del Valle del Cauca, Univalle-Universidad del Valle, Identification of critical parameters in the Cauca river and its main tributaries, Volume IX, Cauca River Modelling Project (PMC), Phase II, Cali, Colombia, 2004a.
CVC-Corporación Autónoma Regional del Valle del Cauca, Univalle-Universidad del Valle, Hydrobiology (benthic macroinvertebrates, phytoplankton and zooplankton) and water quality assessment throughout nine (9) monitoring stations in the Cauca river, Volume XII, Cauca River Modelling Project (PMC), Phase II, Cali, Colombia, 2004b.
NIVA-Norwegian Institute for Water Research, Statistical and modelling methods for assessing the relationships between ecological and chemical status in different lake types and different geographical regions, Deliverable 12, Final Conference, REBECCA project (Relationships between ecological and chemical status of surface waters), Edited by S. Jannicke Moe and Robert Ptacnik, Oslo, Norway, 2007.
Modelling the seasonal climate effects on grapevine yield at different spatial and unconventional temporal scales

Subana Shanmuganathan, Philip Sallis and Ajit Narayanan
Geoinformatics Research Centre (GRC), (www.geo-informatics.org)
School of Computing and Mathematical Sciences, Auckland University of Technology,
New Zealand. subana.shanmuganathan@aut.ac.nz

Abstract: The paper briefly outlines recent conventional approaches to modelling/predicting seasonal climate effects on grapevine phenology and wine quality using weather data from national meteorological institutions and yield/vintage ratings provided by sommeliers. The seasonal variability in climatic conditions can cause shifts in grapevine growth stages, phenological events, which in turn affect the formation and ratio of grape berry components, such as sugar, and pro-phenols, that give the colour, aroma and flavour attributes to the vintage relating to its wine style. Although winemaker ability is considered to be the major determinant of the quality of wine, the excellence of any vintage could still be enhanced considerably with grapes ripened under ideal weather conditions; this is evidenced by better vintage ratings and price hikes associated with better weather conditions in the past. Hence, viticulturists and enologists continuously strive to further scientific understanding on climate effects to increase yield and wine quality. Recent studies reveal that conventional rigorous statistical data analysis methodologies used long term data on crop yield/wine quality and weather conditions for studying the associations between the variables and at regional scales. This data requirement impedes any such meaningful study on vineyards established recently, at a micro scale. The Geoinformatics Research Centre (GRC) approaches investigated to overcome this dilemma with data covering only a decade and at a vineyard scale are discussed. Climate data, such as monthly maximum, minimum and average temperature, monthly total rainfall, occurrence of frost days and growing degree days (GDD) (base 10°C) along with yield is analysed using data mining techniques, such as clustering, then with regression and discriminant methods. The results show potential for predicting future yield/wine quality under current weather conditions that could enhance winegrower ability to improve practices for better outcome from the vineyard in terms of yield quality and quantity.

Keywords: Year-to-year climate variability, grapevine phenology and wine vintage

1. BACKGROUND

Modelling the seasonal climate effects on grapevine phenology and wine quality has agricultural as well as economic significance. The seasonal changes in climatic conditions can cause shifts in grapevine growth stages that are observable in terms of phenological events, such as budburst, floraison, veraison, harvest and then in yield as well. In fact, the seasonal changes to a greater extent can influence the formation and ratio (at favourable levels) of sugar and pro-phenols in grapes that in turn contribute towards wine alcohol content, colour, aroma and flavour of the vintage produced from the grapes of that season, of course this is in addition to the grapevine varietal genetics or ‘cultiva’ factors. Despite the fact that the ultimate quality of the vintage is determined by winemaker talent and experience, many studies have proven that ideal climatic conditions lead to better wine quality (with higher wine ratings by sommeliers) and hence price hikes as discussed in Jones [2007]. The literature reviewed for this research reveals that in all such recent studies based on conventional approaches to analysing climate and wine rating data, analysts had
used data at regional scales and over three decades in time span. This data requirement impedes the use of such methods with vineyard yield data covering shorter time span. To overcome this issue, GRC researchers investigated into using novel approaches based on data mining techniques, to analyse data with shorter time span and the initial results of the research showed potential. The results arrived at quantify the monthly average temperature range (maximum, minimum and average temperatures), rainfall and occurrence of frost days, GDD (base 10°C) that are known to be associated with grapevine yield of a vineyard in northern New Zealand. Hence, the approach being investigated is considered to be useful for analysing data sets that are at different spatial and at unconventional temporal scales and this paper details the methodology as well as results arrived at from this research.

2. SEASONAL CLIMATE EFFECTS ON VITICULTURE AND ENOLOGY

2.1 Literature

The literature reviewed for this study in modelling the seasonal climate effects on viticulture and enology under different climate regimes, reveals that the recent trend in this regard has been on analysing the effects of the phenomenon (natural or manmade), on yield / wine quality at macro scales and with date over three decades. The section looks at some of the major modelling approaches that have been successful at this scale.

IPHEN project as described by Mariani, et al., (2007) was initiated with an aim to produce and broadcast phonological maps for Italian wine varieties initially and then to extend the mapping to other fruits with agro-meteorological data. The maps were developed using a base simulation model originally developed to produce fortnightly maps of phonological phases for two widely grown Italian grapevine varieties, namely, the early Chardonnay and the late Cabernet sauvignon. The base simulation model uses summed values of heating normal hours instead of classical GDD and a detailed Digital Elevation Model (DEM) with pixels of 2 x 2 km. Thermal fields of maximum and minimum temperatures created by applying geostatistical procedures to National Agrometeorological Network were then calibrated using bibliographic data obtained from different wine regions of Italy. The model is corrected on a regular basis before the results are broadcast via the Internet (www.ucea.it).

Meanwhile, Tonietto & Carbonneau, (2004) talked about a multicriteria climate classification system, a methodology developed to describe the climate of vineyards on a macroclimate scale for the wine regions in the world using three synthetic climate indices relating to viticulture. The three indices used being 1) dryness index (DI) that in general used as an indicator of the potential water balance of the soil of Rio’s index, but in the study used as a measure of the level of presence-absence of dryness, 2) heliothermal index for Huglin’s heliothermal index, a coefficient of the thermal component that expresses the mean day length in relation to the latitude and 3) cool night index to describe night temperature during berry maturation. The three indices were portrayed to be viticulture descriptors and complementing each other. The authors concluded that the multicriteria climate system called as Geoviticulture, to be a complete one and could be used to represent the variability of the viticulture climates in all the different wine regions in the world, as the system was built with elements that could represent differences in grapevine varieties, vintage quality (sugar, colour and aroma) and styles/ appellations in wine. The system was initially presented for 97 grape-growing regions in 29 counties, as a research tool for classifying grape-growing regions and wine making that could be applied to mostly at larger scales i.e., world’s wine regions or intra-annual variability within a wine region.

Lobell, et al., (2007) studied the yield-climate relationships for forecasting annual crop production and also for projecting the impact of future climate changes on various crops. In this study, authors used 1980-2003 data on annual yield of 12 major Californian crops and climate (minimum, maximum temperature and precipitation) to model the climate effects on the crop yields. The crops studied were; wine grapes, lettuce, almonds, strawberries, table grapes, hay, oranges, cotton, tomatoes, walnuts, avocados and pistachios. Regressions to find the correlations between yield and climate data for each of the crop were performed and from which fairly simple equations were developed that
explained more than two third of the observed yield variance with only 2-3 climate variables, that were selected by explorative data analysis methodologies and described as most important factors.

Ashenfelter, (1995) established the correlations between the price of different vintages and the season’s weather data that produced the vintages with an example set of French red Bordeaux wines (as judged by the prices of mature wines). The logarithm of vintage price was considered as surrogate for yield / quality of vintages. The age of vintage was regressed against some selected weather variables, such as temperature, (during growing season i.e., April-September), rain in September and August, rain in the months preceding the vintage i.e., October-March, average temperature in September R^2 (root mean squared error) for vintages of 1952-1980 (excluding 1954 and 1956, as these wines were rare, the two vintages being considered as the poorest in the decade).

Jones, (2007) discussed of climate and global wine quality factors and elaborated upon a study on a year-to-year comparison over a ten year period. The study included a description of wine quality factors in juxtaposition with prices and vintage ratings. Citing many earlier studies the author of this work pointed out that the analysis of the relationships between climate variables and wine prices to be based on an underlying hypothesis that beneficial climate conditions would invariably improve the wine quality and that in the past these had in turn led to short term price hikes. The paper also reflected the fact that the unavailability of consistent price data for multiple regions and with different styles over many years to be a shortcoming for any complete analysis/ study on long term effects. Furthermore, argued that the vintage ratings to be a strong determinant of the annual economic success of a wine region based on the work in Nemani, (2001) but then went on to say that the ratings could be determinants of wine quality not necessarily a predictor based on Ashenfelter, (2000) where ratings were described to be reflective of wine somewhat in an indirect way i.e., they had the same weather factors documented to be the determinants of the same wine quality.

So far the section looked at the approaches to modelling climate change effects on viticulture and wine quality with data covering over three decades. As far as New Zealand is concerned data available on wine production covers only short periods of time, i.e., less than a decade. The main reason for this being the chequered past of the New Zealand wine industry; it is since the last decade, certain wine appellations, such as Sauvignon Blanc of Marlborough, Pinot Noir of Central Otago, Chardonnay of North Island, became famous as stated by Cooper (2008) hence the lack of long term data impedes modelling climate effects with conventional approaches. The paper details on the novel approaches being investigated to overcome this impediment, as part of the GRC’s overarching project called Eno-Humanas that is aimed at building models to studying the effects of independent factors, such as environmental and climate conditions on dependent factors in viticulture and its products, such as shifts in phenological events, berry ripening process, yield and wine quality. The major issue in the Eno-Humanas project is establishing the links between the independent factors that consist of more precise data and the dependent factors with less precise data, such as wine quality-which is arguably a subjective matter relating to human sensory perception described by sommeliers about wine taste, mainly about wine colour, aroma, mouth feel and aftertaste. Further details on GRS’s projects relating to wine sensory perception are discussed in (Shanmuganathan, et al., 2009 a & b)

2.2 The methodology

The vineyard studied in this research was established in 1996 hence, with this ten year yield data, recent popular conventional methods that require long-term grapevine phenology could not be performed. To overcome this issue, this yield data is analysed initially using self-organising map1 based data clustering (explorative) and then with statistical methods, (regression and discriminant). Weather data used in the study include concurrent monthly climate average values, calculated by National Institute of Water and

1. A self-organising map is an artificial neural network (ANN) based on an unsupervised algorithmic learning. SOM based data clustering and mapping are useful in projecting multidimensional data sets onto low (1 or 2-D) displays enabling the discovery of new knowledge in the form of patterns/relationships in the maps.
Atmosphere (NIWA) using one of its meteorological station’s daily weather recordings, extracted via the institution’s web portal (NIWA, 27 September 2009).

2.3 The data

Yield data being modelled is from a vineyard in northern New Zealand and covers from 1997-2006. The vineyard data was classified into low, moderate and high yield years. Data on monthly maximum, minimum and average temperature, total rainfall, occurrence of frost days, GDD (base 10°C) for this time period, was extracted from NIWA’s web portal. This monthly data was calculated using daily weather data logged at Henderson River Pk, (36.855398, 174.623833) near the vineyard where the yield data comes from.

3. RESULTS AND DISCUSSION

The section elaborates upon the results of data mining and statistical methodologies.

3.1 Data mining

A self-organising map (SOM) was created using a commercial software package called Viscovery2, with the vineyard yield class and monthly climate data to look for relationships between the two sets of variables, (dependent and independent). In the SOM, the year yield class was given higher priority to favour clustering on this factor (low, moderate and high) so that the final mapping enhances the visualisation of any difference/s between the yield year classes and various climate variables being studied. Hence, in the SOM component plane of yield (in the top left corner of fig 1 a), high, moderate (mod) and low yield years with rate codes (3, 2 and 1) are clustered and mapped in the left, centre and right respectively. The corresponding monthly rainfall component planes (MayprainK-MarrainK) show the distribution of each of these variables in relation to the yield year classes. This monthly rainfall distribution is also displayed in a bar chart graph (fig 1b).

Figure 1 a: SOM component planes, each showing the range for that variable in each map unit/cluster. For example, February rainfall 58.1mm, 69mm and 113.8mm relate to low,

---

1 Viscovery SOMine is a commercial software package that is very useful in explorative data mining, visual cluster analysis, statistical profiling, segmentation and classification. The data mapping is based on Kohonen self-organizing maps (SOMs) and classical statistics in an intuitive workflow environment (Kohonen, 2001)
*moderate* and *high* yield year classes respectively. b: graph showing the monthly rainfall distribution in *high*, *moderate* and *low* yield. **regression and ***discriminant predictors.
**Figure 2:** SOM components of (from row 1-4) monthly rainfall (mm), mean temperature (°C), maximum and minimum temperature (°C) each component clustered based on vineyard annual yield year class, left to right high, moderate and low. Monthly rainfall of May, July, September - January show nonlinear (NL) correlations with yield class. August, November and February show linear (L) correlations. Finally, March rainfall shows inverse (IN) correlation with high yield. NL: nonlinear, L: linear and I: inverse correlation. ** regression and *** discriminant predictors.
Figure 3 a: SOM components of monthly total GDD (base 10°C) and total frost days clustered based on vineyard annual yield ratings, left to right high, moderate, low yield. L: linear, NL: nonlinear and I: inverse correlation to yield class. ** regression and *** discriminant predictors.

b: SOM of yield and monthly climate variables as in figures 1 and 2a. c: Results of regression analysis, yield rating against climate variables.
The SOM results displayed in figs 1 a, 2, 3 a - b show the patterns of correlations between yield year classes and the climate variables studied. The SOM findings are verified with regressions and discriminant analyses and the results are discussed here onwards.

3.2 Statistical data analysis results

Based on regression analyses performed on the same data set predictors for the dependent variable yield (tonne per ha) were: (Constant), November GDD, September rain, February maximum temperature, October mean temperature, September maximum temperature, August maximum temperature, November mean temperature, October rain, January mean temperature and June rain (Table 4) for the vineyard being studied in this study.

Table 1: Table showing regression analysis results. Dependent variable vineyard yield (tons per ha) is regressed against monthly total of rain (mm), occurrence of frost days GDD, mean of average, maximum and minimum temperatures (°C) of May to March grapevine growing cycle.

<p>| Model Summary |
|---------------|----------------|-----------------|-----------------|-----------------|</p>
<table>
<thead>
<tr>
<th>Set</th>
<th>R</th>
<th>R Square</th>
<th>Error of the Estimates</th>
<th>F Change</th>
<th>df1</th>
<th>df2</th>
<th>F Change</th>
<th>Watson</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.873a</td>
<td>.762</td>
<td>.738</td>
<td>1.7323</td>
<td>.762</td>
<td>32.062</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>.914b</td>
<td>.835</td>
<td>.784</td>
<td>1.5220</td>
<td>.073</td>
<td>3.954</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>.948c</td>
<td>.894</td>
<td>.855</td>
<td>1.2906</td>
<td>.060</td>
<td>4.517</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>.969d</td>
<td>.938</td>
<td>.903</td>
<td>1.0567</td>
<td>.044</td>
<td>4.933</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>.982e</td>
<td>.965</td>
<td>.935</td>
<td>.8610</td>
<td>.027</td>
<td>4.543</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>.982f</td>
<td>.964</td>
<td>.944</td>
<td>.8006</td>
<td>.000</td>
<td>.052</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>.995g</td>
<td>.990</td>
<td>.982</td>
<td>.4587</td>
<td>.026</td>
<td>15.323</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>.996h</td>
<td>.996</td>
<td>.992</td>
<td>.3006</td>
<td>.006</td>
<td>8.973</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>.999i</td>
<td>.999</td>
<td>.996</td>
<td>.2087</td>
<td>.002</td>
<td>6.371</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>1.000j</td>
<td>1.000</td>
<td>.0982</td>
<td>.001</td>
<td>18.905</td>
<td>1</td>
<td>3</td>
<td>.022</td>
</tr>
<tr>
<td>11</td>
<td>1.000k</td>
<td>1.000</td>
<td>1.000</td>
<td>.0250</td>
<td>.000</td>
<td>36.211</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>1.000l</td>
<td>1.000</td>
<td>1.000</td>
<td>.0021</td>
<td>.000</td>
<td>280.124</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>a-l. Predictors: (Constant), NovpGDD, SepprainK, FebmaxT, OcpmeanTK, SeppmaxT, AugpmaxT, NovpmeanTK, OcpprainK, JammeanTK, JunprainK. m. Dependent Variable: YieldTperHa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Meanwhile, discriminant analysis ran on the same set of data produced November minimum temperature, December rainfall, March and January maximum temperature, and October frost as predictors of yield for this vineyard (Tables 2a and b).

4. CONCLUSIONS

The paper looked at some major recent approaches to modelling seasonal climate effects on grapevine growth stages, phenology and yield. In general the trend is this regard seems to be more towards analysing yield data at larger spatial scales i.e., regional with data covering at least over three decades. This impedes any conventional approaches to modelling at a micro scale i.e., vineyard and with shorter time span. The paper described an approach consisting of data mining and conventional statistical data analysis methods namely, regression and discriminant, investigated for modelling the year-to-year variability...
in climatic conditions and its effects on viticulture using yield data from a vineyard in northern New Zealand and with monthly weather data extracted from NIWA’s web portal recorded at a nearby metrological station. Despite the SOM results that showed many variables as in relation to yield classes (low, moderate and high), regression results produced monthly November GDD, September rainfall, February and August maximum temperature, October, November and January mean temperature, October and June rainfall as predictors of the dependent variable annual vineyard yield (ton/ha). Interestingly, discriminate analysis results produced monthly November minimum temperature, December rainfall, March maximum temperature and October Forest as predictors of the annual yield in the particular vineyard. The former two independent variables influence veraison, March Maximum temperature affects berry ripening and October forest events immensely impact on budburst. Forest events if not handled properly could destroy a whole year’s production. Hence, based on the results of data mining, regression and discriminate analysis approaches it is possible to gain scientific understanding on anecdotal evidence in viticulture in a quantitative manner that could be applied to improve viticulture practices whereby the quality and quantity of the yield could be enhanced in this case the grapes and the ultimate produce the wine.

ACKNOWLEDGMENTS

Michael Brajkovich of Kumeu River Wines is acknowledged for his advice on viticulture and winemaking, and sharing one of their vineyard yield data. The authors wish to thank Peter Sumich of Auckland University of Technology for his efforts in this regard.

REFERENCES


Jones, G. V., Climate change: observations, projections, and general implications for viticulture and wine production 2007. - Vols. Practical Winery & Vineyard: 44-64.


Mariani L [et al.] IPHEN: a model for real time production of grapevine phenological maps (2007). In proceedings of the Climate and Viticulture Congress, Zaragoza, 10-14 April, pp 272-278 See also www.ucea.it


Deceiving Feedbacks: The Challenge of Policy-Design for Inshore Fishery Activities in Complex Ecosystems. The Case of the Ciénaga Grande de Santa Marta

V. Otero*, C. Olaya* and C. Castro Sanguino
*TESO research group, Departamento de Ingeniería Industrial, Universidad de los Andes (v.otero25@uniandes.edu.co)

Abstract: The Ciénaga Grande de Santa Marta is an estuarine lagoon located on the Colombian coast. Given the lagoon's variability, its connections to adjacent ecosystems, and its biodiversity and anthropogenic activity, it has long been a challenge to managers and policy makers to maintain the lagoon for sustainable use. A one-year joint project with the Instituto de Investigaciones Marinas y Costeras (INVEMAR) research institute and the Universidad de los Andes resulted in the development of a system dynamics model to aid in the design of sustainable policies for fishing control. The study presented here underlines the deceiving characteristics of feedback processes like fishing and biological dynamics, which generate unexpected impact on the lagoon. According to environmental variations and the decision-making processes made by local fishermen, salinity changes and fish abundance scenarios were explored in a simulation model. Various technical tests were performed such as structure assessment, behavior analysis and numerical parameter sensitivity. We suggest that the interplay among the reproduction loop, effort loop and catch loop, shows that variations in catches are defined by decision making processes of the fishermen according to the ecosystem and hydrological conditions. Importantly, the simulation model provided useful and novel approximations to support policy-design processes and discussions among local and regional institutions.

Keywords: Ciénaga Grande de Santa Marta, simulation, estuary, system dynamics, policy-design.

1. INTRODUCTION

Appropriate decision and policy making for highly dynamic ecological systems is a challenging task in environmental management. In particular, degraded ecosystems are expected to have different ecological dynamics that might lead to the spread of large-scale changes due to stressors on ecosystems feedbacks and constraints on its biological processes [Benett, et al., 2003]. The Ciénaga Grande de Santa Marta (CGSM) is an impacted estuary located in Colombia, where research has focused on environmental conditions, fishery resources and activity, and mangrove behavior. Management plans and policy implementations are wanted for the sustainable use of this ecosystem [Rueda and Defeo, 2003b; Rueda and Santos-Martínez, 1999; Sánchez et al., 1998]. We developed a simulation model of the CGSM as part of a one-year joint research project with the INVEMAR research institute and Universidad de los Andes. The main purpose of this model was to understand the relevant systemic interactions of the mangrove, fish, and human populations to help understand the complex dynamics that are observed in the lagoon, in order to support decision-making processes to develop sustainable policies for this ecosystem. This document presents the role of important feedback processes that seem to drive the dynamics of the CGSM. The next section describes the ecosystem of the CGSM, while the third section discusses the relevant literature on feedback complexity. The fourth section introduces the primary feedback structures that are proposed to explain
the dynamics of fish and crustacean populations; these structures were mapped in a system dynamics simulation model. Finally, the last section presents the feedback-grounded hypotheses, which might explain the behavior of key variables as a result of the feedback processes interaction, and major results and the lessons that were derived along with implications on future research.

2. A COMPLEX ECOSYSTEM: THE CIÉNAGA GRANDE DE SANTA MARTA

The Ciénaga Grande de Santa Marta is a lagoon estuarine ecosystem located along the northwestern coast of Colombia [Botero and Mancera, 1998]. The estuarine is naturally variable, presenting challenging tasks for policy makers. Several factors, such as its multiple connections with adjacent ecosystems, its biodiversity, the extreme poverty of its human population, the alterations in its hydrological regime, the increase in the mangrove death rate, the reduction of fishery resources, the degradation of water quality, and the lack of commitment by the government and society, have all provided motivation to ensure the sustainable use of this ecosystem via environmental management plans and methods [Rueda, 2001; Rueda and Defeo, 2003b; Rueda and Santos-Martinez, 1999; Botero and Salzwedel, 1999]. However, despite several rehabilitation initiatives put in place since 1981, salinity upsurge, mangrove defoliation, fishery resource variations, intermittent institutional management and government presence, and social problems are still persistent in the lagoon [Vilaridy, 2007]. Above all, past interventions have not taken into account the ecosystem feedback structures. We suggest that knowledge of these structures will likely improve the understanding of the behavior of the ecosystem and thus support decision-making processes for ecosystem management.

3. FEEDBACK COMPLEXITY AND ECOSYSTEMS

Ecological changes and strong feedback processes, like reciprocal interactions between biotic factors and the physical environment, must be considered in environmental management and restoration projects [Benett et al., 2003; Suding et al., 2004]. Because the CGSM is a degraded wetland ecosystem as a consequence of different anthropogenic activities [Botero and Salzwedel, 1999; Rueda and Defeo, 2003a], it is expected to have very different ecological dynamics than less impacted ecosystems [Suding et al., 2004]. For this reason, an understanding of the ecological processes (feedbacks and constraints) in a degraded system is critical. Different stressors on ecological feedbacks lead to the spread of ecological responses that cause large-scale changes [Benett et al., 2003]. For instance, an anthropogenic impact on an ecosystem can lead to soil loss and reductions in water quality and quantity, causing a dramatic change in biodiversity, such as a reduction in the commercial fish stock. In general, self-reinforcing and self-regulative processes driven by positive and negative feedback loops might help to explain the persistence of specific dynamics that usually challenge our understanding of these systems, either by amplifying or stabilizing behaviors [Sterman, 2000]. Hence, ecological feedbacks might become one of the most important factors for ecological decision- and policy-makers [Benett et al., 2003; Cumming et al., 2005].

A relevant tool for understanding these special ecological interactions (feedbacks) is predictive modeling through simulation, which is appropriate for well-understood systems over short time frames [Benett, et al., 2003]. However, most ecological systems are complex, highly dynamic and not well understood [Jorgensen, 1999; Wu and Marceau, 2002]. Simulation models, and software for environmental management, have been constructed for different ecosystems [Woodwell, 1998; Güneralp and Barlas, 2003; Rai, 2008; Gal et al., 2009]. STELLA™ (isee systems) is one of the most versatile tools and has been widely used for ecological modeling in diverse ecosystems around the globe [Costanza and Ruth, 1998; Costanza and Gottlieb, 1998; Ray and Straskraba, 2001; Hulla, et al., 2008]. This tool is based on system dynamics, a methodology suitable for dealing with complex systems driven by accumulation and feedback [Sterman, 2000].
The CGSM has not been studied from the perspective of system dynamics. Research has mainly focused on mangrove ecosystem, economically important ichthysic species, environmental conditions, and some human activities over the past years [Perdomo et al., 1998; Rueda, 2001; Blanco et al., 2006; Rueda, 2007]. Nonetheless, there are no ecological modeling studies that combine all of these variables together, or in pairs. Currently, only Twilley et al. [1998] have studied the CGSM with a modeling approach. In this work, the mangrove ecosystem of the CGSM was evaluated using the FORMAN model, which was developed to simulate the demographic processes of mangroves in a 0.05 ha plot. The model was used to establish trajectories of mangrove recovery according to different restoration criteria at geographically specific conditions and on a decadal time scale, hence contributing to the design and implementation of restoration projects. The simulations were based on differential equations but feedback processes were not taken into account.

Our work introduces a system dynamics model built in iThink™ (isee systems), a software tool similar to STELLA, which allows for the inclusion of diverse variables to capture relevant feedback processes that might help to explain the complex dynamics of the CGSM. The system dynamics method provides a framework to explore and understand the feedback characteristics of a particular system, linking policy decisions to actions, and linking particular behavior to delays in the decision-making process and system response. [Forrester, 1961; Sterman, 2000]. This method can be used for understanding problems in industrial, urban, economic, politic and ecological systems, such as a declining market or industrial productivity, the flux of populations, rising salinity levels and unemployment [Saysel and Barlas, 2001]. An emphasis is made in understanding the system structure, that is, how interactions among physical and decision-making feedback processes produce system behavior. Patterns such as exponential growth or decay, s-shaped behavior, and oscillation and decay can be understood with positive and negative feedback loops [Saysel and Barlas, 2001; Sterman 2000]. Two different tools are available to represent the feedback processes in system dynamics, Causal Loop and Stock-and-Flow diagrams. Causal Loop diagrams are qualitative tools for representing relationships among variables, while Stock-and-Flow diagrams quantitatively describe the system structure using accumulations and rates [Sterman, 2000]. Thus, computer simulation is used to build explanations of structure-behavior relationships with the aim of enhancing the understanding of policy-makers.

4. FEEDBACKS IN THE CGSM

Our model is defined in four sectors (Fig. 1). The environmental sector models the habitat for fish and crustaceans as well as the abiotic (chemical and physical) conditions that are suitable for species survival, such as salinity, and dissolved oxygen. The population sector addresses the most economically and ecologically relevant fish and crustacean species for fishery management at the CGSM (Fig. 1), including reproduction, mortality and migration processes influenced by environmental conditions. The human and fishery sector concentrates on the local human population whose main economic activity is artisanal fishing, which employs many different pieces of fishing equipment, described in Figure 1. The last sector models the mangrove dynamics of the three main species (Rizophora mangle, Laguncularia racemosa and Avicennia germinans) that are locally used to build houses, and for firewood and fishing tackle. The simulation model uses arrays to encapsulate parallel model structures in a single visual arrangement to represent different fish and crustacean species and fishery gear under the same causal structure.
4.1 Conceptualization

The four different sectors generate sixteen feedback causal loops. We will describe only the most pervasive feedbacks that directly affect fish populations by means of generic Causal Loop Diagrams (CLD). CLD use arrows (causal links) to describe the relationship between two variables, and an effect of one variable on another is indicated by the direction of the arrow. Each arrow has an associated polarity sign; a positive sign means that an increase or decrease in the independent variable causes a similar change (further increase or decrease) in the dependent variable, and a negative link means that there is a change in the opposite direction, e.g. an increase produces a decrease in the dependent variable [Sterman, 2000]. Feedback loops are formed when the effect of any variable is fed back as a new cause by passing through any other variables that are connected with causal links. When an increase in the initial variable, for example, fish and crustacean population (Figure 2), results in a further increase in the same variable after going through a loop, or vice versa, it is considered to be a self-reinforcement loop and is denoted with a positive sign surrounded by an arrow (Fig. 2). If an initial decrease in a particular variable results in an increase in the same variable, or vice versa, the loop is labeled as a self-regulating loop denoted by a negative sign (population and carrying capacity interaction in Figure 2). Behavior patterns can be associated with these two types of basic feedback structures; positive feedback loops promote exponential growth while negative loops produce exponential-decay and goal-seeking behavior. Delays and interactions among diverse loops produce more complex patterns such as oscillatory and sigmoidal curves [Sterman, 2000]. Computer simulation helps to build these types of structure-behavior hypotheses.

Figure 2 shows the mutual interactions generated by fish and crustacean populations. A typical self-reinforcement feedback drives the growth or decay of specific species according to its reproductive behavior (Reproduction loop L2). The other feedback process is a negative loop that self-regulates the population and the ecosystem carrying capacity (Fish and crustacean carrying capacity, L1) [Monte-Luna et al., 2004]. Additionally, the reproductive behavior is affected by environmental conditions, and we considered the most relevant factors to be salinity, dissolved oxygen, and the nursing and feeding areas that mangroves provide.
The human population interacts with fish and crustaceans via fishery activities (Fig. 3). The available fishery resources define the total catch, which in turn controls the quantity of those resources due to extraction (Catch loop L3). The total catch is also affected by fishing efforts, in other words, the number of fishery trips is reinforced when the total catch is abundant enough to maintain use of the fishing gear (Effort loop L4). The fishing effort is also affected by fishery resource prices and follows normal supply-demand market dynamics, i.e. price changes modify the effort made by fishermen according to the highest price, which in turn is driven by the resource supply (Price loop L5).

Mangrove dynamics can also be characterized by reproduction and carrying capacity loops (Fig. 4). The growth of mangroves is controlled by available resources (Mangrove carrying capacity loop L6) and their regeneration is defined by seedling production, which in turn also generates further mangrove biomass (Mangrove regeneration loop L7). Growth is also affected by environmental conditions such as the interstitial salinity and water level, depending on the tolerable range for mangrove species. Finally, mangrove biomass is also affected by deforestation. The causal relationships associated with salinity and water level in Figures 2 and 4 do not have causal polarities because they depend on the different range of tolerance for salinity by specific species of fish, crustaceans or mangroves.

4.2 Simulation

Based on the previous conceptualization, a simulation model was built. Technical tests were made such as boundary adequacy, extreme conditions, structure assessment, behavior reproduction, equation formulation, and numerical parameter sensitivity [Forrester, 1961; Sterman, 2000]. Afterwards, we ran simulations to explore various hypotheses to understand structure-behavior relationships. The time horizon was defined as 20 years (2000-2020). Illustrative results are shown in section 5. Simulated and available data were compared. The model showed reliable behavior for aggregated fish and crustacean captures, effort and mangrove density according with seasonal changes in salinity, and trends found in historical data (results are not included in this document).
5. DISCUSSION

Simulations are shown in Figure 5. Four scenarios are proposed to test hypotheses about feedbacks interactions. The first scenario is based on salinity historical data provided by INVEMAR from the years 2000 to 2009, while the data for the future years were selected based on that data which shows salinity periodicity associated with two annual climatic seasons in the ecosystem, one dry and one rainy [Blanco et al., 2006]. The amplitude is estimated according to SOI intensity. The second scenario makes changes to the amplitude in salinity. Seasonal changes in fish abundance are associated with spawning periods and salinity variations, i.e. the ability of fish to cope with stressor conditions. We suggest that the Reproduction loop (L2) reinforces population growth during those spawning seasons, which is simultaneously controlled by the limiting resources in the ecosystem through the Carrying capacity loop (L1) (see Figure 2). The population growth rate is influenced by environmental conditions; for instance, seasons with lower salinity are associated with a higher abundance of fish and vice versa. In Figure 5 (left side), for the first scenario, a variable fish catch is different from the second scenario, where fluctuations are more associated with the reproduction loop than with salinity changes.

We also suggest that the Catch loop (L3) controls the extraction and total catch; this is reflected in changes in the fishing effort when the Effort loop (L4) reinforces the increase or decrease in the number of fishing trips according to the interest of fishermen, i.e. to make the most of fishery. Nevertheless, the interplay among the reproduction loop, effort loop and catch loop shows that the variations in specific catches are defined by decision making of the fishermen and by the hydrological conditions of the ecosystem (see Figure 5 left side).

![Figure 5. Left side: Fish total catch (kilograms/month) behavior with regular salinity, scenario one (1) and modified salinity, scenario two (2). Right side: Changes in fish gear total effort (number of trips per month) with all the species for scenario three (1) and leaving out two fish species (2) for scenario four.](image)

Furthermore, to explore the previous hypotheses, we ran two more scenarios with regular salinity values taken from scenario one. The third considers all of the species shown in Figure 1 and the fourth scenario leaves out the two most captured fish species from 2000 to 2008 (O. niloticus and M. incilis). Figure 5 (right side) shows the simulation results for the fish gear efforts. The interaction between the Catch loop (L3) and Effort loop (L4) for both scenarios influences the total effort of the fish gear in accordance with seasonal variations of the fish population, promoting an increase or decrease in effort, depending on the abundance of fish captured, which is regulated by the Catch loop (L3). We suggest for the fourth scenario that the Effort loop (L4) reinforces the decrease in the total effort of fish gear due to less abundance (L3). The Reproduction loop (L2) still operates, but the dominance of the Catch loop is stronger, as the fourth scenario emphasizes. In addition, the decrease in fish abundance promotes an increase in crustacean fishing efforts, because fishermen are looking for more abundant captures according to the ecosystem supplies. This leads to the hypothesis that fishing trips using certain gear are higher than for others gears based on the resource abundance of the fishery (Fig. 5, right side).
The simulation model is a first approach to explain and understand the interactions and resulting behavior among fish, crustaceans, mangroves and fishermen in the complex ecosystem of the CGSM by means of feedback loops. The model provides a new and available management tool to enhance system thinking and to support policy-decision making among institutions, government and local inhabitants. Additionally, it provides a novel perspective to approach sustainable management in the CGSM, where policy-making can be based on the selection or simulation of certain desirable dominant feedback structures.

However, further work on the understanding of feedback processes in the CGSM is desirable. A detailed discussion of model assumptions and non-linear relationships among the variables is still necessary. The sensitivity of the model in the change-effort function is significant; functions that describe changes in species price, fish, crustaceans and human population migration behavior should also be further explored. Finally, it is important to test alternative measurement units for mangroves.

In summary, we provide an ecological simulation model for enhancing the comprehension of complex interactions of biological, chemical and human factors in the CGSM. Here, computer simulation is used as a tool to examine the consequences of the interaction of feedback structures and to support the decision-making processes regarding the management of the lagoon. An explanation of behavior that is based on feedback processes provides a better and more structured understanding of the ecosystem to enrich decision-making and management approaches used in the CGSM.

ACKNOWLEDGEMENTS

Instituto de Investigaciones Marinas y Costeras INVEMAR and the Research Center of the Engineering Faculty (CIFI) at Universidad de los Andes provided the financial support for the development of this joint project. We would like to especially thank Dr. Mario Rueda, the marine resource use and assessment program coordinator at INVEMAR, for his continuous support, willingness and experience to make this project possible.

REFERENCES

Botero, L. and J.E. Mancera, Síntesis de los cambios de origen antrópico ocurridos en los últimos 40 años en la Ciénaga Grande de Santa Marta (Colombia), Revista de la academia colombiana de ciencias exactas físicas y naturales 78, 465 – 474, 1998.


Dynamics of habitat banking under changing conservation costs and habitat restoration time lags

Martin Drechslera, Florian Hartigb

a Helmholtz-Centre for Environmental Research – UFZ, martin.drechsler@ufz.de
b Helmholtz-Centre for Environmental Research – UFZ, florian.hartig@ufz.de.

Abstract: Tradable permits are a popular instrument for environmental policy. Classical examples are emission control, but tradable permit systems are recently also promoted and applied to the conservation of biodiversity. This raises a number of challenges. One problem is that habitat turnover (the destruction of habitats and restoration elsewhere) is harmful to many species, even if the total amount of habitat is constant. Another problem is that the restoration of habitats often takes time, leading to time lags between the beginning of restoration activities and the time when the restored habitat is available for trading. It is therefore important to understand how restoration time lags and economic conditions affect habitat turnover. Using an agent-based model, we study the dynamics of a tradable permit market. In the model, conservation costs differ among agents and change in time, and habitat restoration takes time. Our results show the existence of trade-offs between conservation costs, the amount of habitat provided and the amount of habitat turnover. We discuss how habitat turnover can be controlled by an additional tax. We also find that restoration time lags can lead to feedback loops and oscillations in key variables and reduce the efficiency of the permit market. We conclude that temporal lags deserve a careful analysis when implementing tradable permit systems for the preservation of natural habitats and biodiversity.

Keywords: agent-based model, environmental policy, habitat turnover, time lag, tradable permits.

1. INTRODUCTION

 Tradable permits have gained increasing popularity as a cost-efficient policy instrument for limiting the overuse of public or common natural resources. Examples of such schemes are the carbon emission trading schemes settled after the Kyoto Protocol, or water permits [Tietenberg 2006]. Recently, permit or credit trading systems have also been discussed and applied to control land use for species protection and biodiversity conservation (e.g., Fox and Nino-Murcia [2005], Latacz-Lohmann and Schilizzi [2005]). Under such a biodiversity credit trading system, the destruction of a natural habitat requires possession of a permit which can only be acquired by restoring a habitat of equal ecological value elsewhere or by buying a permit on the market.

Controlling the amount of natural habitat in a region by means of a tradable permit system raises a number of challenges as opposed, e.g., to the control of carbon emissions. The persistence of biodiversity is highly sensitive to the spatial and temporal allocation of natural habitat. While the spatial allocation has been addressed in various papers such as Drechsler and Wätzold [2009] and Hartig and Drechsler [2009], not much research exists on the temporal allocation. Two factors are particularly important here and special to biodiversity. The first is the presence of long time lags. The restoration of many relevant habitats may take decades or more. If the production of a market good involves a time lag, market participants must decide on their production level before they can sell. The relevance of time lags in markets has been emphasized already about 80 years ago by Hanau [1928] and Kaldor [1934] who showed that markets with time lags may exhibit cyclic fluctuations of prices and the supply of goods. The reason is a feedback loop: at a
current low price, production levels tend to be small, leading to a shortage of the good in the next period, which is associated with a higher price in the next period, which in turn triggers higher production levels in next period, leading to oversupply of the good in the following period, and so on. These cycles are not only undesirable from the point of view of consumers, but may also be inefficient in the sense that the same total amount of the good could be produced at fewer costs if productions levels were held constant in time.

The second problem is a very specific problem in conservation markets: While tradable permit markets lead to more flexibility and to cost savings, the associated activity in these markets is generally harmful for the species. The reason is that each trade implies the destruction of a patch of habitat and its local species populations, while created habitat patches have to be colonized by species before they can contribute to species survival. Even if the total number of habitat remains constant, and even if species are able to disperse between habitats, this spatial reallocation of habitats, called habitat turnover, is often detrimental for the species (e.g., Keymer et al. [2000]).

In this study, we develop a (non-spatial) agent-based model of a market for biodiversity credits. The key variables of this market are habitat restoration time, restoration costs, and the spatial and temporal variation of the opportunity costs of conservation. We ask three questions: How do ecologically and economically important quantities such as habitat turnover and the total costs of the permit scheme depend on these key variables? How can undesired habitat turnover be controlled? And finally, are permit markets for biodiversity likely to develop cyclic fluctuations, similar to other markets with production time lags? Section 2 explains the mathematical formulation of the model. Simulation results are presented in section 3 and discussed with their practical implications in section 4.

2. METHODS

2.1 State variables and model parameters

We assume that the model region consists of \( N = 1000 \) land patches each of which is owned by one agent and managed in discrete time steps. During each period \( t \), a patch can be either managed for agriculture: \( r_i(t) = 0 \), or for conservation: \( r_i(t) = 1 \). Conservation management incurs a patch-specific opportunity cost \( c_i \) per period. Opportunity costs \( c_i(t) \) are modeled as a temporally correlated random walk with mean 1, standard deviation \( \sigma \) and temporal correlation \( \alpha \). A correlation of \( \alpha = 1 \) means perfect temporal correlation so that \( c_i(t+1) = c_i(t) \), and a correlation of \( \alpha = 0 \) means that the costs between consecutive periods are uncorrelated. Costs are further assumed to be uncorrelated among different patches.

The ecological state \( x_i \) of patch \( i \) is given by a value between 0 and 1, with 1 being a habitat of maximum ecological value. The ecological state changes according to the type of management applied to the cell. Management for agriculture immediately changes the ecological state to 0. When managed for conservation, the ecological state is assumed to recover with a speed of \( 1/K \) per time step. Only a patch in state \( x_i = 1 \) is assumed to be of ecological value (“habitat”). Thus, a formerly agriculturally managed patch has recovered to a habitat after a “restoration time” \( K \). We assume that landowners may have to support restoration actively. In this case, they bear an additional cost of magnitude \( d \) per time period during restoration, additionally to the opportunity cost \( c \).

We also assume that any landowner whose patch is not habitat \( (x_i < 1) \) requires a permit. Landowners can, however, hold a permit even if their patch is habitat. Denoting the number of permits held by owner \( i \) as \( z_i \), these conditions are expressed by demanding that each patch \( i \) fulfills

\[
x_i + z_i \geq 1
\]  

(1) (the patch is either habitat, a permit is held, or both). Banking of permits for the purpose of speculation is not considered in our model, so \( z_i \) can be either 0 or 1. This assumption will be justified in the Discussion. The number of land-use permits possessed by all agents in
the model region is denoted as \( Z \). The levels of \( x_i \) and \( z_i \) define the state \( Y_i = (x_i, z_i) \) of patch \( i \). Land-use permits are traded on a perfect market at an endogenous equilibrium price \( p^* \) (see section 2.2). Agents decide on their market and land-use actions under the objective of maximizing their long-term profit (section 2.3). This long-term profit is assumed to be the sum of the discounted profits earned in each period. The discounting factor is denoted as \( q \).

### 2.2 Market dynamics

Each time period is composed of two phases. In the first phase, agents observe the conditions of the current period: the current opportunity costs, \( c_i(t) \), and the state \( Y_i(t) = (x_i(t), z_i(t)) \) of the patch. Based on this information, agents engage in the permit market and adapt their land use which marks the beginning of the second phase of period \( t \). The patch state in this second phase is denoted as \( Y'_i(t) = (x'_i(t), z'_i(t)) \). We assume that the first phase is short compared to the second phase, so that the change into the second phase is practically instantaneous for the purpose of calculating the change of the ecological state \( x_i \).

Assuming that agents maximize their long-term expected profits, we can formulate for each agent \( i \) a demand function \( b_i(Y_i, c_i, p) \) that tells for a given permit price \( p \) whether he is willing to buy a permit on the market \( (b_i = 1) \), sell a permit \( (b_i = -1) \) or do nothing \( (b_i = 0) \). The demand function will be derived in the next subsection. We assume a perfect market where supply and demand are completely settled at an equilibrium price \( p^*(t) \) which is determined by a balance of demand and supply

\[
\sum_{i=1}^{N} b_i(Y_i, c_i, p^*) = 0
\]  

(2)

As a consequence of the market interaction, the number of permits held by agent \( i \) changes from \( z_i(t) \) to

\[
z_i'(t) = \min[\max[z_i(t) + b_i(t),0],1]
\]  

(3)

Next to the market interactions, agents decide on their land use type \( r_i(Y_i, c_i, p) \), depending on the (current) patch state \( Y_i \), opportunity cost \( c_i \) and permit price \( p \). Land may either be management for agriculture: \( r_i = 0 \), or for conservation: \( r_i = 1 \). The choice \( r_i(t) = 0 \) in time period \( t \) implies that the patch degrades to the lowest possible ecological state: \( x_i(t) \to x'_i(t) = 0 \), regardless of its state \( x_i(t) \) that was observed in the first phase of the time period. The decision model of the agents, i.e. the functions \( b_i(Y_i, c_i, p) \) and \( r_i(Y_i, c_i, p) \) will be presented in detail in the following section 2.3.

If \( r_i(t) = 1 \) and \( x_i(t) < 1 \), there are two options: habitat may be restored instantaneously, or it may be restored only after one or more periods. First we consider the second case and assume that patches recover at a rate of \( 1/K \) (with \( K > 0 \) meaning that \( K \) is the time (number of periods) for a patch to recover from an economically used patch \( (x_i = 0) \) to an ecologically valuable habitat \( (x_i = 1) \) if it is managed for conservation \( (r_i = 1) \). Uncertainty in the restoration success (cf. Moilanen et al. [2008]) is ignored for simplicity. Since restoration takes at least until the next period, \( t+1 \), the choice \( r_i(t) = 1 \) implies that the patch remains in its current state \( x_i(t) \) for the rest of the current period \( t \), so the ecological state in period \( t \) is

\[
x_i'(t) = r_i(t)x_i(t)
\]  

(4)

Even though the state of the current period is not affected, the choice of \( r \) changes the ecological state \( x_i'(t+1) \) in the following period: if \( x_i(t) < 1 \), the patch recovers to the next better state \( x_i(t+1)/K \). If the patch is already habitat \( (x_i(t) = 1) \), it remains in that state, so

\[
x_i(t + 1) = \min[r_i(t)[x_i(t) + 1/K],1] = \min[x_i'(t) + r_i(t),1].
\]  

(5)

Restoration activities are assumed to be necessary until the ecologically valuable state, \( x_i = 1 \) has been reached and so we assume that restoration costs occur whenever both \( r_i = 1 \) and \( x_i < 1 \).
The number of permits in the next period, $t+1$, is the number of permits after trading:

$$z_{i}(t+1) = z'_{i}(t)$$

(6)

With the state $Y_{i}(t+1) = (x_{i}(t+1), z_{i}(t+1))$ given by eqs. (5) and (6), we can move to the next period, $t+1$ and proceed in the same way as in period $t$.

Equations (4) and (5) consider the case where habitat restoration is not instantaneous. The case of instantaneous habitat recovery can be considered by replacing eqs. (4) and (5) by

$$x_{i}(t+1) = x'_{i}(t) = r_{i}(t)$$

(7)

obeying the constraints $x_{i} \in \{0,1\}$ and $x_{i} + z_{i} \geq 1$ (eq. (1)).

2.3 The agents’ decision rules

We assume that each agent attempts to maximize his long-term aggregated profit, considering all periods from the present period till infinity and discounting with a discount factor $q = 1/(1+\delta)$ where $\delta$ is the discount rate per time period. For the future periods the agent assumes that the permit price $p$ and the opportunity cost $c_{i}$ will stay the same as in the current period. For the case of non-instantaneous restoration the agents’ optimal decision rules are obtained through dynamic programming (Drechsler and Hartig unpublished):

(a) If the patch is not habitat and a permit is held ($x_{i}<1$ and $z_{i}=1$) the agent can choose between restoration ($r_{i}=1$) or economic use of the patch ($r_{i}=0$). For the decision the agent has to weigh the discounted benefit of selling the permit at price $p$ after restoration is complete against the sum of the discounted restoration costs that accrue from the present period until restoration is complete, and the discounted opportunity costs from the present period until infinity (the latter is henceforth termed the land price of the patch). If the former exceeds the latter $r_{i}=1$ is optimal, otherwise the agent chooses $r_{i}=0$.

(b) If the patch is habitat and no permit is held ($x_{i}=1$ and $z_{i}=0$) the agent can choose between buying a permit ($b_{i}=1$) and converting the patch to economic use ($r_{i}=0$), or buy no permit ($b_{i}=0$) and conserve the patch as habitat ($r_{i}=1$). If the current permit price $p$ exceeds the land price of the patch the latter option is chosen; otherwise the former option is chosen.

(c) If the patch is habitat and a permit is held ($x_{i}=z_{i}=1$) the agent can choose between selling the permit ($b_{i}=-1$) and conserving the patch as habitat ($r_{i}=1$), or keeping the permit ($b_{i}=0$) and converting the patch to economic use ($r_{i}=0$). If the current permit price $p$ exceeds the land price the former option is chosen; otherwise the latter option is chosen.

For the case of instantaneous restoration the results are identical except that in case (a) the permit can be sold immediately after the decision to restore the choice is between restoring and selling on the one side and not restoring on the other.

2.4 Analysis methods

To analyze the stationary dynamics of the market model, we run the model for 1000 “burn in” periods, and then run it for another 1000 periods to generate the data for the analysis. We consider a base parameter combination $\sigma=0.3$ (medium variation of opportunity costs), $\alpha=0.8$ (large level of temporal correlation in the opportunity costs), $q=0.95$ (which corresponds to a discount rate of about five percent per period), $d=10$ (restoration cost equals half the average land price, $1/(1-q)$), and an allocation of $Z/N=0.75$ land use permits in the region. We simulate the market dynamics and record the temporal development of the following aggregated variables: the total amount of habitat prior to decision making ($H_{1}$) (“first phase”: see section 2.2) and after decision making ($H_{2}$) (“second phase”), the amount of habitat turnover between consecutive periods ($T$) the total opportunity cost ($C$) which sums the opportunity costs ($c_{i}$) of all patches under conservation or restoration
(r_i=1), and the total restoration expense (D) that sums the restoration costs (d) of all patches under restoration (r_i=1 and x_i<1).

From the defined base case parameter combination we vary α between 0.1 and 0.5, d between 0 and 1/(1-q), q between 0.92 and 0.98 and Z/N between 0.6 and 0.9. For each varied parameter combination we determine the following statistics of the state variables: the average amount of habitat turnover (ET), the average total opportunity cost (EC), the average total restoration expense (ED). All costs are scaled in units of the mean opportunity cost (which was set to 1 in section 2.1).

These analyses are carried out for instantaneous restoration and for restoration delayed by one time period (K=1). We lastly consider cases of K>1 and investigate the effect of varying the patch number N. Some of the numerical results will be supported by analytical calculations.

3. RESULTS

3.1 The market dynamics

Iterating through the time periods with eqs. (3), (6) and (7) yields the temporal evolution of the state variables for the case of instantaneous restoration (Fig. 1a). One can see that all patches whose owners do not possess a land use permit are habitat: H(t)=N-Z (solid line; recall that N is the number of patches and Z the number of permits). The amount of habitat turnover is about 10% of the total amount of habitat (Fig. 1a, dashed line).

Iterating through the time periods with eqs. (3)–(6) yields the temporal evolution of the state variables for the case of delayed restoration (K=1) as shown in Fig. 1b. Compared to the case of instantaneous restoration the amount of habitat in the first phase of the time periods and the amount of habitat turnover are increased, and most strikingly, oscillate. The reason for this is the time lag in the habitat restoration: The decision to restore a habitat is made in period t on the basis of the current price and opportunity cost, but the habitat and associated land use permit is obtained only a period later where cost and price may have changed.

Figure 1: Amount of habitat in phases 1 and 2 (solid resp. dotted line) and habitat turnover (dashed line) as functions of time for the case of instantaneous habitat restoration (panel a) and for the case of delayed habitat restoration (K=1: panel b). Note that in the case of instantaneous restoration the amount of habitat in phase 1 (cf. section 2.2) of each time period equals that of phase 2 (panel a, solid line). The model parameters are: N=1000, Z=750, σ=0.3, α=0.8, d=0.25/(1-q) and q=0.95.

The time lag leads to a feedback loop: In a period with much excess habitat permit prices are small leading to little restoration activities so the amount of excess habitat declines to the next period, while in a period with little excess habitat permit prices are relatively high, leading to a higher level of restoration activities and more excess habitat in the next period.
3.2 Influence of opportunity cost variation and restoration costs on the market behavior

To analyze the influence of opportunity cost variation and restoration costs on the market behavior, we focus on three variables of the market as introduced in section 2.4: the long-term average amount of habitat turnover, \( ET \), the long term average total opportunity cost, \( EC \), and the long-term average total cost which is the sum of opportunity cost and restoration expenses, \( EC + ED \). For the case of instantaneous restoration, Figure 2a shows that habitat turnover \( ET \) increases with increasing cost variation \( \sigma \) and decreasing restoration cost \( d \), because at large \( \sigma \) cost savings from reallocating habitat are large and at small \( d \) the costs associated with these reallocations are small. The same result is obtained for the case where restoration is delayed by one time period (Fig. 2d).

One would expect that larger restoration costs are generally associated with low habitat turnover and a high total opportunity cost. And indeed, in the case of instantaneous restoration, this is observed (Fig. 2b). However, in the case of delayed restoration we observe that increasing restoration cost \( d \) not only decreases turnover (Fig. 2d) but also the total opportunity cost (Fig. 2e). The reason for this is that in the case of delayed restoration habitat turnover is associated with the restoration of excess habitat, which is associated with excess costs during the phase of restoration (see above). Increasing restoration cost \( d \) therefore reduces both habitat turnover and the associated excess cost.

For small \( \sigma < 0.2 \), the total cost \( (EC + ED) \) first increases, approaches a maximum and then decreases when the restoration cost \( d \) is increased (moving from the bottom to the top of Figs. 2c and 2f). As \( \sigma \) increases the maximum moves towards larger values of \( d \) so that for \( \sigma > 0.2 \) total cost monotonically increases with increasing \( d \) in the interval \( 0 \leq d \leq 10 \).

The effect of the cost variation \( \sigma \) on the total cost depends on the magnitude of the restoration cost. At small \( d < 2 \) the total cost decreases with increasing \( \sigma \) (moving from left to right in Figs. 2c and 2f), because the total cost is dominated by the total opportunity cost \( EC \) which was found to decline with increasing \( \sigma \) (Figs. 2b and 2e). At large \( d \) we find the opposite: that the total cost increases with increasing \( \sigma \). The reason is that at large \( d \) the total cost is dominated by the total restoration expense \( ED \), which increases with increasing habitat turnover \( ET \), which was found to increase with increasing \( \sigma \) (Figs. 2a and 2d).
4. DISCUSSION AND CONCLUSIONS

4.1 Summary of the main results

We investigated how the variation in the opportunity costs among land users (\(\sigma\)), the restoration cost (\(d\)) and the proportion of issued land use permits (\(Z/N\)) influence the amount of habitat turnover (\(ET\)) and the cost of the policy which is composed of the total opportunity cost (\(EC\)) and the total restoration expense (\(ED\)) (respectively summed over all agents).

Habitat turnover generally increases with increasing cost variation (because at large cost variation the cost savings associated with the reallocation of conservation activities and the trade of land-use permits is high) and decreases with increasing restoration cost and habitat recovery time (because large restoration costs and recovery times make restoration unattractive to land users, reducing the supply of permits on the market).

Low habitat turnover, however, leads to higher total opportunity costs, because conservation activities are not allocated to the least costly sites. Total cost (total opportunity cost plus total restoration expense) is maximized at medium levels of the restoration cost \(d\), because at large \(d\) habitat turnover is small and so is the total restoration expense which is the product of habitat turnover and restoration cost. If \(d\) is small increasing cost variation reduces the total cost, because cost-savings from the reallocation of conservation activities are large. If \(d\) is large increasing cost variation increases the total cost, because it generates high habitat turnover which leads to a high total restoration expense.

If restoration is not instantaneous but implies a time lag, the modeled permit market exhibits cyclic fluctuations of habitat and costs (Fig. 1). This behavior has similar causes as, e.g., the well-known pork cycle (e.g., Kaldor [1934]). A feedback loop exists so that in time periods where little of the good (habitat) is present the implied high market prices sets an incentive to produce more goods for the next period, which leads to excess provision of the good in the next period. Conversely a large amount of goods in the present period reduces incentives to produce the good, leading to a shortage in the next period. In our model, any non-zero restoration time increases the total opportunity cost, because it leads to excessive restoration activities (for details, see the discussion of Fig. 1 in section 3.1). This effect however declines if the time and/or cost of restoration become large, because as noted above, large restoration costs and times render habitat restoration unattractive to land users.

4.2 Policy implications

Our results allow drawing some conclusions relevant for the design of permit markets for conservation. In a static setting, one will generally find that a high variation in opportunity costs among land patches offers high efficiency gains through the allocation of conservation to the least expensive patches (e.g., Ando et al. [1998]). In a dynamic analysis, however, it is important to examine how the opportunity costs develop in time, because changes in opportunity costs may lead to ecologically unfavorable habitat turnover.

We find that the cost of habitat restoration is an important factor that determines habitat turnover rates. Therefore, changing this cost by raising a tax on restoration (which within the scope of our model is equivalent to raising a tax on habitat destruction) would reduce habitat turnover. For instantaneous restoration this would be costly, implying a trade-off between total cost and habitat turnover. For delayed restoration, in contrast, the tax would reduce the cyclic fluctuations and on net reduce both turnover and total costs.
4.3 Model assumptions and future research

An important assumption of the model is that landowners base their expectations about future costs and prices solely on their current observations. In contrast to this assumption, it appears also plausible that agents are able to look further back to the past, observe the long-term behavior of costs and prices and include this information into their decision making. Such a memory of agents would probably affect our results in several ways. First, our observation that long restoration times reduces market fluctuations and excessive restoration activities is partly explained by the fact that some agents who engage in restoration in one period may pull out in the next period – a behavior that would become less likely if our model agents were smarter. On the other hand, given that restoration times may be substantial and uncertainty in future costs and prices large, simply extrapolating the current conditions may still be the best option for a real actor. Since the agents face a decision problem under uncertainty, it could also be interesting to allow for errors in the decision process by making the decision rules stochastic or fuzzy.

Another assumption is that spatial issues are neglected. These can be internalized through spatial trading rules which may have consequences on the dynamics of a permit market (Drechsler and Wätzold [2009], Hartig and Drechsler [2009]). Future research should study the interaction between spatial trading rules and rules that target the market dynamics, such as a restoration tax. One can, however, expect that the general conclusions of the present analysis will be retained even if spatial interactions and trading rules are included.

REFERENCES

Hanau A., Die Prognose der Schweinepreise, Vierteljahreshefte zur Konjunkturforschung, Verlag Reimar Hobbing, Berlin, 1928.
Feedbacks in socio-environmental land systems

Joerg A. Priess^a, Nina Schwarz^b, Sven Lautenbach^c
^a, ^b, ^c Helmholtz-Centre for Environmental Research – UFZ, Department of Computational Landscape Ecology
^a joerg.priess@ufz.de, ^b nina.schwarz@ufz.de, ^c sven.lautenbach@ufz.de

Abstract: The dynamics of socio-environmental systems are driven by exogenous forces and by the interaction of endogenous system components, both within the social and environmental realms, as well as between them. In recent years, the number of models and modelling frameworks explicitly representing feedbacks has increased, especially for models of land use systems. Land use changes are on the one hand caused by a complex interaction of human and/or institutional land use demands and the environment, which supports or limits human use in several aspects. On the other hand, land use changes and their effects at least partly influence the respective driving forces and future land-use decisions, e.g. by affecting the productivity of agricultural land, in- or decreasing the quality of life in (residential) urban areas, increasing accessibility and thereby facilitating the economic development of areas and so forth. In this paper we address the complexity of socio-environmental systems via analysing and reviewing the feedbacks implemented in current simulation models, with a focus on feedback loops between the social and the environmental component. We developed an analysis framework distinguishing several categories of information exchange between model components. Results indicate that feedbacks from ‘population’ simulated e.g. as households or average land managers were well represented, whereas institutions or technical changes were rarely addressed. From the environment component mostly the performance of crops or density of population were reported, whereas other environmental changes e.g. concerning soil, weather or water dynamics or structural changes in cities were addressed less frequently or not at all. We conclude that the land-use modelling community started to address system complexity via implementing feedback loops, leaving much room for increasing the realism of information exchange between representations of the social and the environment components.

Keywords: feedback mechanisms; land use modelling; review; urban and rural land systems.

1. INTRODUCTION

Land is a limited resource, especially if we consider land, which is suitable for specific land-use purposes such as agriculture, forestry, livestock-production, housing, recreation or cultural activities. The dynamics of such systems are determined by (1) the initial state of the land system, (2) external (or exogenous) driving forces and (3) internal feedback loops, which may be positive or negative. In system dynamics, the term feedback is used to characterize a bidirectional relation between two or more system components (Morrison 1991), which is exactly the way we use the term in this paper. We clearly distinguish feedbacks from unidirectional relations, which are called drivers, or driving forces (Geist and Lambin 2002), or impacts from the perspective of the affected component. In the land-use literature, instead of the term ‘feedback’ other terms are frequently used, such as link, (complex) interaction, coupled system, connection. For these terms, the authors often not clearly define whether just one component is influencing another one, or whether both influence each other, i.e. whether the relations they address are uni- or bidirectional (e.g. Liu et al. 2007; Parker et al. 2008; Rindfuss et al. 2008; Schaldach and Priess 2008; Verburg 2006; Walsh et al. 2008, Young et al. 2006).
Humans interact with the environment in various ways. Historically, research on these interactions has been split into research on the effects of human actions on the environment and on the effects of environmental changes on human well-being. Coupled socio-environmental systems have been addressed by natural / ecological and social sciences, and the attention for the topic has considerably increased over the last two decades (Science Direct: from 2,800 papers in 1990 to 21,000 papers in 2009). While both scientific realms agree on the importance of understanding the dynamics of socio-environmental systems, they disagree on explanatory approaches (Turner and Robbins 2008), which is also reflected in the range of methodologies of the studies reviewed in this paper. While the complexity of human-environment interactions (here: socio-environmental feedbacks) has been noted early (Marsh et al. 1864), approaches that tackle complexity via addressing the feedbacks between both subsystems have kicked in mainly during the last decade (see Table 1). This can be attributed to the separation of ecological and social sciences (Rosa & Dietz 1998; Liu et al. 2007). The evolving Land-Use Science (also called land change science, e.g. Turner & Robbins 2008) is trying to embrace approaches from either side, contributing to generate new insights into the multiple dimensions of social and environmental subsystems involved in land-use dynamics (GLP 2005). While drivers of land-use dynamics have been analysed in detail (Angelsen and Kaimowitz 1999; Geist and Lambin 2002), to date no systematic approach is available for classifying feedback mechanisms and analysing how they contribute to explain land-use dynamics, although the scientific community seems to agree about the importance to study and simulate the complexity e.g. feedback mechanisms in land systems (GLP 2005; Liu et al. 2007, Parker et al. 2008, Young et al. 2006).

Recent reviews of land use models tackled feedbacks in several ways. Alberti (2008) analysed feedback loops between environmental and human system and distinguished various spatio-temporal scales. In their review of urban models, Haase and Schwarz (2009) differentiated feedbacks (1) of land use and human sphere, (2) of environment and human sphere, and (3) between local and regional scale. Schaldach & Priess (2008) identified socio-environmental feedbacks in land-use models for regional to global scale. Finally, Verburg (2006) distinguished three types of feedbacks (1) between driving factors and the effects of land use change, (2) between local and regional processes, and (3) between agents of land use change and the spatial units of the environment.

This paper seeks to contribute to the maturation of land use science (Rindfuss et al 2008) by analysing important feedbacks in socio-environmental land systems and how they are implemented in the models generated and used by the scientific community. The aims of this paper are thus twofold: first, to provide a framework for analysing feedbacks in socio-environmental land use systems, and second to review existing simulation models of land use systems regarding the feedbacks included. In the review part of the paper, we address the following research questions:

- Which feedbacks are tackled / neglected?
- Which elements of the social and environment components are addressed?
- On which (temporal and spatial) scales do simulated feedbacks occur?

2. METHODS AND DATA

2.1 Analysis Framework

At the global scale, we argue that the only external drivers are the solar activity and the parameters of the orbit, while all other dynamics are endogenous to the system. This per-
spective has been addressed and simulated by Meadows et al. (1972), or with considerably more detail in the GUMBO model (Boumans et al. 2002). At regional or local scales, we expect quite different sets of external drivers (Geist and Lambin 2002) and endogenous feedbacks, which are responsible for the dynamics of land systems, partly originating from human decisions and activities, and partly related to environmental processes and functions and the services they provide for society (e.g. MEA 2005; Rudel et al. 2005; Lambin and Meyfroidt 2010). The differentiation between external drivers and endogenous feedbacks is then a matter of system boundaries. The socio-environmental systems we are addressing in this paper are conceptually divided into two major components: (1) The social component, comprising elements such as persons or households (population), organisations, but also economic sectors or technology, and (2) the environment (be it natural or man-made such as urban or agricultural), including all biophysical properties and processes (see GLP 2005; Schaldach and Priess 2008). At least two types of feedbacks (Figure 1) can be distinguished in these systems, which we here define as:

TYPE 1: feedback between major components
TYPE 2: feedback within major components

TYPE 1 feedbacks between major components of socio-environmental land systems comprise a wide variety of human activities such as where and when to use land for a certain purpose and how to use it (plant potatoes or wheat; build a hut or a castle; irrigate or not; expand land use or abandon a piece of land; protect land for nature conservation or carbon accumulation). The second half of the loop includes all types of environmental changes in functions and services, which are influencing human (land use related) activities and decision making. In either of the social and the environment component cascades of internal (TYPE 2) feedbacks may occur, depending on the social, spatial and temporal scales and resolutions at which the authors study and analyse land systems. It is noteworthy that simple representations of socio-environmental land systems can lack TYPE 1 feedbacks, for example if soils and weather or other environmental conditions are represented as static rather than dynamic elements, resulting e.g. in stable crop yields. Depending on the richness of details included in the models, TYPE 2 feedbacks might be covered implicitly or explicitly. Note that TYPE 1 feedbacks might also pass through TYPE 2 feedbacks before the loop is closed.

Both types of feedbacks as defined above are essential for either stabilising land-use (e.g. a farmer keeps growing potatoes as long as crop yields and prices are within the range of expectations), or triggering land-use transitions, for example if functions or services pass threshold values (Lambin and Meyfroidt 2010).

Figure 1. Feedbacks in socio-environmental systems. Since feedbacks are defined as bidirectional, TYPE 1 feedbacks cover the complete loop between the social and the environment component.

To add more detail to the analysis of feedbacks, the social and the environment component were divided into sub-systems. The social component encompasses Population, Economy, Politics/Planning, Culture, and Technology. The category Population covers the dynamics

---

3 In the remainder of section 2, all categories used in this study are printed in italics.
of population including migrating households as well as quality of life and human health. The category Economy relates to economic activities including farming practices, while Politics / Planning refer to decisions made by policy makers like changes in subsidies, property rights, or protection status of land. The environment component consists of the sub-systems Built environment, Biodiversity, Vegetation/Crops, Soil/Biochemistry, Hydrology and Atmosphere. As far as possible, the feedback analysis is addressing the sub-systems involved. Note that only in cases where the social component was represented without identifiable sub-systems, we used the classification ‘Whole human system’.

Table 1. Case studies for which feedbacks have been analysed.

<table>
<thead>
<tr>
<th>Source</th>
<th>Continent</th>
<th>Simulation method</th>
<th>Thematic focus (one or more)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Berger (2001)</td>
<td>S/C-America</td>
<td>ABM</td>
<td>X X X</td>
</tr>
<tr>
<td>Claessens et al. (2009)</td>
<td>Europe</td>
<td>Regression</td>
<td>X</td>
</tr>
<tr>
<td>Costanza et al. (2002)</td>
<td>N-America</td>
<td>ODE</td>
<td>X X X X X</td>
</tr>
<tr>
<td>Engelen et al. (2007)</td>
<td>Europe</td>
<td>CA</td>
<td>X</td>
</tr>
<tr>
<td>Eppink et al. (2004)</td>
<td>not specified</td>
<td>ODE</td>
<td>X</td>
</tr>
<tr>
<td>Holzkneper and Seppelt (2007)</td>
<td>Europe</td>
<td>GA</td>
<td>X X</td>
</tr>
<tr>
<td>Le et al. (2008)</td>
<td>Asia</td>
<td>ABM</td>
<td>X X</td>
</tr>
<tr>
<td>Lee et al. (2008)</td>
<td>Asia</td>
<td>Regression</td>
<td>X X</td>
</tr>
<tr>
<td>Liu et al. (2008)</td>
<td>N-America</td>
<td>Various</td>
<td>X X X</td>
</tr>
<tr>
<td>Manson (2005) / Parker et al. (2008)</td>
<td>S/C-America</td>
<td>ABM</td>
<td>X X X X X</td>
</tr>
<tr>
<td>Matthews and Pilbeam (2005)</td>
<td>Asia</td>
<td>ABM/ODE</td>
<td>X X</td>
</tr>
<tr>
<td>Oel et al. (2010)</td>
<td>S/C-America</td>
<td>ABM</td>
<td>X X</td>
</tr>
<tr>
<td>Parker et al. (2008)</td>
<td>not specified</td>
<td>ABM</td>
<td>X X X X X</td>
</tr>
<tr>
<td>Priess et al. (2007a,b)</td>
<td>Asia</td>
<td>CA/ODE</td>
<td>X X X X</td>
</tr>
<tr>
<td>Priess et al. (2010)</td>
<td>Asia</td>
<td>CA/ODE</td>
<td>X X X X</td>
</tr>
<tr>
<td>Salvini and Miller (2005)</td>
<td>N-America</td>
<td>ABM</td>
<td>X</td>
</tr>
<tr>
<td>van Delden (2007)</td>
<td>Europe</td>
<td>CA</td>
<td>X X</td>
</tr>
<tr>
<td>Verburg and Overmars (2007)</td>
<td>not specified</td>
<td>Regression</td>
<td>X X X</td>
</tr>
<tr>
<td>Walsh et al. (2006)</td>
<td>Asia</td>
<td>CA</td>
<td>X X X</td>
</tr>
<tr>
<td>Walsh et al. (2008)</td>
<td>S/C-America</td>
<td>CA</td>
<td>X X X</td>
</tr>
</tbody>
</table>

From the social to the environment component, the following aspects of changing land use are distinguished: Intensity, Location, and Extent. Intensity refers to aspects like density, amount of fertiliser used, or irrigation and the like. Location encompasses (re-) location of certain land uses in a spatially explicit way, such as building new houses or allocating crops. Finally, Extent covers changes in the area covered by a land use type.

4 ABM – Agent Based Model; CA – Cellular Automata; GA – Genetic Algorithm; ODE – Ordinary Differential Equations.
From the environment to the social component, the direct influence of land use is addressed, because the presence or absence of a land use in an area might influence social processes like location choices of households depending on existing built-up areas or other factors. Additionally, influences of environmental sub-systems are captured using the ecosystem services (ESSs) concept. The Milenium Ecosystem Assessment (2005) grouped ESSs into the categories: Provisioning, Regulating, Cultural and Supporting. For each of the models under review, we analysed, which of these feedback categories are incorporated.

### 2.2 Selection Process and Case Study Characteristics

For the review presented in this paper, the following criteria were used to select studies on models simulating socio-environmental systems:

- published in peer-reviewed literature.
- focusing on land issues (urban, rural and natural areas; including studies on land - freshwater & land - coastal waters).
- preferably process- or rule-based, to be able to link processes and feedbacks.
- preferably applied at the regional to local scale because studies tend to be less aggregated and to have a more explicit and detailed representation of socioeconomic and biophysical processes and feedback loops.

Based on the criteria above, 28 modelling studies have been reviewed (see table 1). They cover a wide variety of topics, including urban simulation (7), agriculture (21), water management (8), natural vegetation (10) and migration (5). The following modelling techniques are used, with some modelling studies combining two or more methods: cellular automata (6), agent-based models (9), system dynamics models (4), regression models (4) and optimisation methods (2). Geographically, the case studies cover all major regions except Australia and Antarctica: 7 are located in Asia and in North-America, 4 in Europe, 5 in South and Central America, 2 in Africa and 3 model applications could not be related to a specific region.

### 3. RESULTS

![Figure 2](image-url). Feedbacks from the social component via land use (2a, left) to the environment component (2b, right). The colours code the number of times a feedback has been observed in the reviewed case studies (white: no feedback; dark: highest number of feedbacks). The x-axis represents 3 different aspects of land use.
We started tackling the feedbacks of the social and environmental components of the modelling studies by analysing the types of land use changes that are related to various social sub-systems (Figure 2a). The majority of feedbacks originates from the Economy or the Population sub-systems. In the models analysed, the sub-systems of Technology, Planning / Politics, and Culture are much less represented in land use change decisions and information flowing to the environment components. Related to the latter, we noted that changing land use intensity is less often modelled than extent or location of certain land use types. The second step was to analyse the influence of these types of land use changes onto the environmental sub-systems (Figure 2b). Vegetation/Crops and Soil/Biochemistry are the two environmental sub-systems that are influenced the most, hinting at the larger number of agricultural modelling studies in the review. Hydrology and the Built Environment are also directly addressed, whereas issues relating to Biodiversity and the Atmosphere are rarely tackled. However, they might be influenced by TYPE 2 feedbacks occurring within the environmental component. Land use intensity, location, and extent are all relevant for influencing the various environmental sub-systems.

Figure 3. Feedbacks from the environment component via ecosystem services to the social component. The y-axis represents ESSs\(^5\), which trigger the effects in the models. P- provisioning, R – regulating, C – cultural, S – supporting. The colours code the number of times a feedback has been observed in the reviewed case studies (white: no feedback; dark: highest number of feedbacks).

ESSs are an important pathway linking the environment to the social component of land systems (Figure 3). Whenever possible we identified single ESS, of which the most frequently used are food provisioning (P1), water provisioning (P2) wood & fibre provisioning (P3), recreation and tourism (C3) and nutrient cycling (S3). Regulating services are mostly water related (amount & quality), but also including air, climate and soils. Figure 3 also highlights that many potentially important feedbacks via ESSs were not addressed at all, including all major categories – provisioning (inorganic, biochemical) – regulating (pest control, natural hazards) – cultural (heritage, spiritual) – supporting (biodiversity, soil formation). The social sub-systems addressed in the studies mainly refer to population and economy, as these two sub-systems are the most likely to be modelled explicitly, whereas Politics / Planning and Culture are often only implicitly included in scenario configurations.

Figure 4. Feedbacks from the environment component via land use to the social component. The colours code the number of times a feedback has been observed in the reviewed case studies (white: no feedback; dark: highest number of feedbacks).

\(^5\) P1 - food; P2 - fresh water storage and retention; P3 – wood, fiber & fuel; P4 - Inorganic resources; P5 - biochemical & medicinal resources; P6 - genetic materials; P7 - ornamental species; R1 - air quality regulation; R2 - climate regulation; R3 - water quantity; R4 - water quality; R5 - Soil retention & erosion protection; R6 - natural hazard mitigation; R7 - biological regulation & pest control; C1 - cultural heritage; C2 - spiritual & artistic inspiration; C3 - tourism and recreation; C4 – aesthetic; C5 - science & education; S1 - biodiversity & nursery; S2 - soil formation; S3 - Nutrient Cycling
Technologies are mainly present in the form of agricultural practices such as fertiliser use and irrigation, either simulated in process models or via regression coefficients, in either case influenced by crop yields, soil fertility or both. In turn, technologies influence the yield expectations, revenues and land-use decisions of agents or other units of decision-making. Note that the categories e.g. Politics / Planning or Economy do represent a wide variety of different pathways, scales and economic sectors, depending on the major purpose, location and spatio-temporal scale of the study.

Finally, land use can directly link the environment and the social components. The majority of the modelling studies represent those feedbacks by addressing Economy or Population (Figure 4). This is again due to the fact that Economy and Population are the social sub-systems that are most likely to be included dynamically rather than as scenario constraints. Thus, only a few studies consider effects on Policy / Planning, technological or cultural effects. The most important changes in land use considered are the changes of location and of the extent of land use. Land use intensity as well is included in a large fraction of the studies.

4. DISCUSSION AND CONCLUSIONS

Our analysis shows that the land use community started to represent a range of categories of feedback loops in their models. Various representations of social sub-systems such as households, economic sectors (or the economic reasoning of simulated decision-making) are exchanging information with sub-systems of the environment such as crops/vegetation, hydrology or soils. The term feedback (loop) is not at all limited to 1:1 relationships regarding the flow of information between the two conceptual main components, but is also comprising 1:n and m:1 relationships. For example, detailed agent-based models sending various classes of information to their biophysical environment (expansion of farm, plant crop \( x \) in location \( y \), irrigation with technology \( u \); settle in zone \( w \)), “expect” and receive only crop yield levels and distances from the environment component, estimated without any changing weather or soil conditions. The reverse has also been found, i.e. reporting detailed environmental dynamics to “average” decision-makers.

While the majority of studies focus on rural areas and agricultural production, environmental feedbacks from weather, hydrologic and soil conditions are rarely included. As a consequence, decisions simulated for land use transitions like deforestation, which are involving rapid changes in carbon, water and nutrient status of soils, might on the one hand be biased or even spurious. A similar conclusion holds for urban models, in which residents decide upon their location choice with only limited feedbacks of the environment regarding changing living conditions due to human-made restructuring the city. On the other hand, many cellular automata, system dynamics and regression models still provide only “average” land use and management decisions as input for their environment components. Thus, the internal representation of complexity, processes and TYPE 2 feedbacks built into social and environment components is also influencing the amount and categories of TYPE 1 feedbacks between the components (see examples in the previous paragraph). Although many socio-environmental models seem to be developed by multi-disciplinary teams, the old divide between social science and economy vs. natural science and geography and the approaches they preferably use, is also at least partly reflected in the categories of implemented feedbacks. Another example for deficits is including different levels of decision-making in agent-based models or cellular automata. Among others, this encompasses aspects like spatial planning or real-estate developers for urban regions or regional authorities controlling agricultural practices. On one hand it is expected that additional components and feedback loops better explain the complexity of land systems and add more realism to simulations, but on the other hand more feedbacks also require additional data or assumptions and introduce more degrees of freedom and new uncertainties.

Although not always explicitly reported, annual decision-making seems to be a common scheme in most of the 28 studies, triggering cascades and feedbacks both within the components (TYPE 2) and information exchange with the environment components (TYPE 1).
Time lags, occurring for example if intra-annual decisions are required, are treated in different ways. Firstly, predefined land-management strategies are executed between feedback events, even if (process) models of high temporal resolution (monthly, daily) are used to represent environmental components (e.g. the DAYCENT model employed by Priess et al. 2007, 2010). Second, Berger (2001) seems to use a higher frequency of feedbacks between decision-making and the availability of irrigation water (monthly). A third option to bridge temporal gaps makes use of capabilities of environmental models like DAYCENT (Parton et al. 1998; also employed in Priess et al. 2007, 2010) or SWAT (Arnold and Fohrer, 2005) to simulate certain intra-annual land-management decisions like irrigation or fertilisation, triggered by environmental feedbacks, such as soil water status. In the latter case, decision-making and the corresponding feedback loop is partly shifted from the social to the environment component (e.g. in Priess et al. 2010). Water availability and water management/use can also be used to analyse how feedbacks are working across spatial scales. For example Liu et al. (2008) employ a whole suite of models in the SAHRA framework to bridge spatial scales, addressing three levels to transport water management information from the regional scale to the pixel. Others like e.g. Berger (2001), Costanza et al. (2002) or Priess et al. (2010) aggregate water availability from the pixel to the (sub-) catchment scale, at which the information is passed to the simulated water managers/users. Water users decide to irrigate their farm-pixels either based on water rights and crop type (combined agent-based model/cellular automata of Berger 2001), or agricultural pixels based on a multi-criteria suitability assessment and crop type (in the PLM model of Costanza et al. 2002 or in the SITE model of Priess et al. 2010).

To date, only a limited degree of real-world complexity is captured by including feedback loops in current models of land systems. While the model developers’ selection of TYPE 1 (and TYPE 2) feedbacks to capture the complexity and characterise the dynamics of land systems is mostly motivated by the objectives of the study (e.g. research questions directed towards prediction or process understanding; decision support with or without stakeholder involvement), it is less clear why specific feedbacks are neglected. Important feedbacks we expected to be addressed (based on the focus of the study) were not implemented, e.g. agricultural land use has been studied without (direct or indirect) feedback from soils or weather (dry or wet conditions) or urban development has been analysed without considering effects of spatial planning. Furthermore, in some studies it was difficult to identify feedbacks between the different components, because the verbal description of the processes involved left (too) much room for interpretation. The deficits identified may be related either to limited know-ledge or process understanding, or data constraints, or to the scientific concepts / perspectives of the model developers differing from our expectations, or to limitations of the state of the art, or simply to the limited amount of time and resources available for model development. However, ex-post analyses like this review always bare the risk of partly misinterpreting the objectives of modelling studies or model structures, thus over- or underestimating scientific progress or limitations.

As previous work on land use drivers has shown (Angelsen and Kaimowitz 1999; Geist and Lambin 2002), the systematic analysis and classification of elements of land systems has contributed considerably to the progress of the evolving land-use science, and particularly the development of (predictive) models as argued by Loveland et al. (2003). In a similar manner, we expect that the systematic analysis of feedback mechanisms in existing models contributes to identify recent advances and limitations in land-use modelling and ultimately our insights into the dynamics of real-world socio-environmental systems. We consider this review as a first step, as we could only briefly address feedbacks across spatial and temporal scales, while the techniques how different feedbacks have been implemented could not be tackled at all. Additionally, more models need to be analysed to identify successful simulation strategies (Parker et al., 2008). Other types of feedbacks e.g. between different land systems need to be analysed as argued by Liu et al. (2007), contributing to advance the representation and understanding of complex systems.

REFERENCES


Manson, S.M., Agent-modeling and genetic programming for modeling land change in the Southern Yucatan Peninsular region of Mexico, Agriculture, Ecosystems & Environment, 111, 47-62, 2005.


Irrigation Management in Chile: 
Integrated Modeling of Access to Water

T. Arnold\textsuperscript{a}, H. Uribe\textsuperscript{b}, C. Troost\textsuperscript{c}, T. Berger\textsuperscript{c}

\textsuperscript{a}Drinking Water Source Protection, Owen Sound, Canada (t.arnold@waterprotection.ca)  
\textsuperscript{b}INIA Quilamapu, Chillan, Chile  
\textsuperscript{c}University of Hohenheim, Germany

Abstract Fresh and clean water is one of the scarcest and most vital resources to humankind. Agriculture is the largest global water user. Irrigation managers must orchestrate water use at the catchment scale, balancing the management of supply and demand and taking into account the benefits from water use, its distribution among water users and environmental concerns. We present results from the integrated modeling of irrigation water use at catchment scale. An extended hydrological runoff model WaSiM-ETH that depicts inefficient surface irrigation was integrated with and coupled to a parametric model for irrigation water distribution, which is linked to the bio-economic multi-agent model MP-MAS that represents farmers as water users. Models were calibrated empirically, first as standalone models and then with increasing complexity of interactions. The integration of process across such long chain of reasoning resulted in an improved system understanding along disciplinary boundaries. The case study presented is irrigation water use within the Chilean Region of Maule. We analyze how farmers whose endowment with formal water rights is insufficient may depend on spillover water, and specifically the distribution of benefits from improvements in canal conductive efficiency across the farming community.

Keywords: Watershed management; Irrigation; Agent-based model; Integrated Modeling of Feedback

1 INTRODUCTION

1.1 Formal and informal access to water in Chile

Since the 1990s, Chile has become a model for successful free market economies, pairing a strong specialization into few competitive commodities and raw materials, with strong government that proactively supports its producers. With its unique location in the southern hemisphere, export agriculture for Northern markets has experienced rapid and sustained growth. The agricultural sector and its focus on high-quality products generates impressive revenues for producers, processors and traders and contributes significant resources to the government budget. Moreover, it absorbs unqualified and skilled labour and gives the nation a positive image. However, with increasingly modern production systems on one side and the remaining traditional systems, Chile’s income disparity has become one of the most extreme in the world and equitable access to productive resources is a core development objective [Lopez and Anriquez, 2007].

In Chile, water management has always been linked to agricultural development policies, because most of Chile’s agriculture relies on irrigation during the core growing season, December to February. The Water Code, the legal foundation for access to water, has evolved dynamically under different political climates. In 1981, the socialist code of 1973 was reformed and a totally market-oriented policy was adopted. This Water Code defined water as a ‘public property for private use’ and defines water rights as a water equivalent, defined in liters/second. Entitlements to surface water are separated from land ownership and can be traded freely and transferred to other uses, once they are inscribed (‘legalized’) with the Direcccion General de Aguas. The long-term impacts of this Water Code are the subject of national and international debate [Bauer, 2005].

Today, many farmers have not yet fully legalized their water rights with the government. This process is costly and paper works consume time and require good literacy. Nationally, Hearne and Donoso [2005] estimate that 10 to 50 percent of all rights are still not legalized and remain ‘customary’.
Though protected by law [Donoso, 2006], neither is the quantitative value of such customary rights precisely defined, nor if flows may be used continuously or discontinuously [JdV Longavi, 2005]. Due to the volatility of river flows, available water can be less than the amount that right holders are entitled to. In such years, most user organizations interpret water rights ‘traditionally’ as percentages of river flows [Hearne and Donoso, 2005] and every right holder suffers equally from a proportional reduction of water delivery.

The Water rights of all farmers are managed by water user organisations (Juntas de Vigilancia), which ensure that all farmers receive their water. These user-based organisations are also responsible for the maintenance and improvement of the canal conductive system. Mandated by their members, they can also apply for government support for projects that improve the infrastructure of the water conduction system. Such work includes the maintenance, repair and extension of canals, aqueducts, distribution devices and inlets, water gages etc.

1.2 The Model Use Case

The Maule Region is located between Santiago and Concepcion. This area is dominated by the production of apples, pears, berries, vegetables, rice, corn and wheat, and pastures. During the hot and dry summers, temperatures often exceed 30°C and precipitation is as low as 4mm/month, so all summer crops and even pastures require irrigation. Winters are temperate and wet (200 mm/month). Irrigation water is taken from rivers that originate in the Andean mountains, which are fed by precipitation and snow melt that starts with the spring thaw in August and lasts well into January.

To study benefits from water use of heterogeneous farmers, Berger [2001] developed a bioeconomic multi-agent model. This model MP-MAS uses on Mixed Integer Linear Programming, a method established in agricultural economics for farm-level production analysis, and for developing optimal production plans under constrained asset endowments [Hazell and Norton, 1986]. This model estimates incomes and crop yields under water deficit, recursively updating farmers’ asset endowment. With a simulation that uses a statistical representation of every farmer in the study region, Berger used a diffusion of innovation model to demonstrate the impact of Chile’s integration into the Mercosur on different farm strata [Berger, 2001]. For the 2001 model, water rights registries were not available and the essential production input ‘water rights’ was quasi-randomly attributed to farmers. As a calibration benchmark, land use and census data (VI Censo Nacional Agropecuario 1997, INE) were used [Berger, 2001].

Within the project Integrating Governance and Modeling and as part of the CGIAR Challenge Program on Water & Food, Berger et al. [2007] conceptualized how deep integration of economic and hydrological science can generate new insights, extending the crop yield module and a hydrological bucket model that are already implemented in the MP-MAS software. A (semi)empirical model system was built that integrates the watershed-scale distributed hydrological model WASIM-ETH [Schulla and Jasper, 2007] with an intermediate bucket model that parameterizes the distribution of water from rivers, to canal sectors, to individual farmers. Ultimately, the bioeconomic, agent-based farm model MP-MAS is used for an economic analysis of agricultural water use. These models were integrated conceptually and specific components were added that link the cause-effect chains. The value added with a multi-agent model is the simulation of an additional interaction layer, the interactions between a heterogeneous population of farmers. The example elaborated in this paper is spillover water and its relevance for farmer’s access to water.

The intermediate and parametric bucket model pools water associated with the same delivery canals (called irrigation sector), handles canal conductive efficiency, agent-agent interactions such as return flows from inefficient irrigation, spillover water and surplus water that farmers abandon and finally leakage between canals and irrigation sectors (together subsumed as ‘non-attributed water’). Each irrigation sector $j$ is characterized by its total water delivery as a total of all water rights, its canal conductive efficiency $\eta_j$, the irrigation efficiency that is averaged over all fields pertaining to this sector, and the portions of water flows that remain in the pool of spillover water or are lost to other compartments (deep percolation or seepage into other sectors).

Technically, model components were first created as standalone software, to facilitate calibration by disciplinary experts. Interaction variables remained boundary conditions. Then, more complex model setups were created that internalize interactions within a hierarchical coupling scheme [Arnold, 2008].
This paper exemplifies the analysis of a single interaction component in a longer cause-effect chain, as a step toward more integrated system understanding.

Not having room to thoroughly introduce all models, only a few relevant processes are explained here. The multi-agent model represents farmers who transform inputs (land, water, fertilizer, seeds) with the help of investment goods (machinery, irrigation equipment, horses, etc.) into farm produce (crops, dairy products or meat). A range of production technologies are parameterized and may or may not be available to the farmer. As a rational actor, the farmer produces with the objective of maximizing his income, making optimal use of his limited or costly resources. Farmers make an annual production decision that is based on expectations of future markets and hydro-meteorological conditions. On a longer time scale, farmers may purchase investment goods or participate in land markets. Liquidity may be met through short- or longterm credits at the externally determined interest rate.

A wide range of empirical data was collected to parameterize the extended model, including two full agricultural censuses [INE 1997, 2007] with detailed data on land use, farming and irrigation technologies and crops [Troost, 2009], market prices, local crop parameters and hydro-meteorological time series, and complete land- and water right registries that were compiled with local water user organisations [Uribe et al., 2009]. The maximum obtainable yield of each production technology is parameterized. If plant water demands are not met, then yields are reduced using the CropWAT approach [Allen et al., 1998].

Within the larger research project, we were asked to assess the impact of investments into canal efficiency on the population of farmers, and how the existing governmental support program could be improved. The relevant cause-effect chain starts with the annual fluctuation of river flows (the source of irrigation water). River organisations distribute this water into a vast canal system, managed by a second layer of user organisations. Farmers receive irrigation water for cropping and the generation of income. However, the pool of ‘non-attributed’ water is an additional open access resource, as positive externality resulting from spills. This pool also includes a part of canal losses. Finally, farmers can collectively decide to improve the canal conductive system. This collective action feeds back on how much water each farmer receives.

This paper describes two processes: the creation, use and relevance of this pool of ‘non-attributed’ water, as an interaction between farmers that results in a cascade of farm-to-farm interactions. By comparing with-and-without scenarios, we illustrate how an improvement of canal conductive efficiency would impact on the amount of water that each individual farmers receives, in order to understand why user groups would engage into collective action and seek government support for canal improvements.

Data on impacts from canal improvements are not available and the current canal conductive efficiency was estimated through a farm survey. Depending on the irrigation sector, efficiency ranges from 0.5 to 0.9 of the original water delivery and is 0.65 for the largest and most representative sector that is used for illustration purposes.

2 AN INTERMEDIATE MODEL FOR ‘NON-ATTRIBUTED’ WATER AND ITS USE

With improved and more detailed registry data for both land and water [Uribe et al., 2009], we were confronted with a paradox: only 2301 out of the 3594 agents own water rights, and several of these own far less water than required to crop their land. For January of a representative and a dry year, the average water endowment was computed per hectare and farmers are counted for each farm size stratum (Table 1). Assuming a typical crop irrigation requirement for one hectare of 0.5 - 1 liters/second, only about 29% of all farmers have an adequate endowment of water rights in a normal year (23% in a moderately dry year). How do those without adequate water rights operate their farms?

Observing this phenomenon, Donoso [2006] mentions ‘surplus’ water that is taken from rivers and abandoned by their owners. Owners are believed to leave this surplus water in the canals once the irrigation demands of their crops are met. Other farmers, without rights, benefit from this pool of spillover water. Other sources of spillover water are inefficiencies of the canals and on-field irrigation methods. Experts acknowledge the abundance of spillover water in most years and describe the difficulty of enforcing the modest maintenance fees that are attached to legalized water rights in the presence of this free resource. Only in years with stronger droughts does this pool of spillover water

1501
### Table 1: Number of farmers and their level of water endowment, expressed in water right equivalents per hectare, given for normal/moderately dry years. With changing water availability, farmers shift between categories.

<table>
<thead>
<tr>
<th>Water Endowment group (liter/second per hectare)</th>
<th>Specialized small farm [3.5 - 5 ha]</th>
<th>Small farm [5 - 25 ha]</th>
<th>Medium farm [25 - 60 ha]</th>
<th>Large farm [60 - 200 ha]</th>
<th>absolute</th>
<th>percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>184</td>
<td>943</td>
<td>126</td>
<td>40</td>
<td>1293</td>
<td>36%</td>
</tr>
<tr>
<td>&gt; 0 - 0.1</td>
<td>4 / 6</td>
<td>173 / 212</td>
<td>63 / 86</td>
<td>33 / 45</td>
<td>273 / 349</td>
<td>8% / 10%</td>
</tr>
<tr>
<td>&gt; 0.1 - 0.25</td>
<td>13 / 16</td>
<td>238 / 305</td>
<td>108 / 132</td>
<td>33 / 27</td>
<td>392 / 480</td>
<td>11% / 13%</td>
</tr>
<tr>
<td>&gt; 0.25 - 0.5</td>
<td>10 / 12</td>
<td>422 / 463</td>
<td>133 / 142</td>
<td>27 / 28</td>
<td>592 / 645</td>
<td>16% / 18%</td>
</tr>
<tr>
<td>&gt; 0.5 - 1</td>
<td>24 / 24</td>
<td>522 / 434</td>
<td>128 / 92</td>
<td>14 / 10</td>
<td>688 / 560</td>
<td>19% / 16%</td>
</tr>
<tr>
<td>&gt; 1</td>
<td>46 / 39</td>
<td>266 / 207</td>
<td>35 / 15</td>
<td>9 / 6</td>
<td>356 / 267</td>
<td>10% / 7%</td>
</tr>
<tr>
<td>Total</td>
<td>281</td>
<td>2564</td>
<td>593</td>
<td>156</td>
<td>3594</td>
<td>100%</td>
</tr>
</tbody>
</table>

Dry out, leading to aggressive and sometimes violent conflicts over water access. For such an informal resource, empirical data is non-existent. Due to its enormous importance for farmers, any economic production analysis must take it into account.

**Types of non-attributed water.** Parts of the drainage from inefficient irrigation fields returns to the canal system. This return flow may be re-used by other irrigators downstream from the canal. The water that returns from fields into main canals and the river system is called return flows \( RF_j \) to one sector \( j \). The surplus water, \( S_i \), of one agent, \( i \), creates is the water received from legalized rights, \( L_i \), but not used because plant irrigation demand, \( D_i \), is satisfied. The model (Berger 2001) was extended so that surplus water is first accumulated for each sector \( j \) and then a share \( \beta_c \) of it is re-distributed to all farms within that sector (informal arrangements that give preferential access to some are thus ignored):

\[
S_j = \sum_{i \in j} S_i = \beta_c \cdot \sum_{i \in j} L_i - D_i
\]

A third source of non-attributed water stems from canal losses due to low conductive efficiency \( \eta_c \). A part of distribution losses within the canal system are added to the pool of spillover water (e.g. along cracks, weirs and junctions). This use share \( \beta_c \) of these losses remains within the same sector, while the rest is lost to groundwater or downstream sectors. The total amount of water \( T_j = \sum_r \sum_{i \in j} Q_r \cdot WR_i \) that is delivered to an irrigation sector \( j \) aggregates the water rights \( WR_i \) of all its farmers \( i \) and on the river flow \( Q_r \):

\[
C_j = \beta_c \cdot (1 - \eta_c) T_j
\]

**Access to non-attributed water.** Total water delivery \( T_i \) to each agent \( i \) combines access to legalized sources \( L_i = \eta_c \sum_r Q_r \cdot WR_i \) from water rights and access to the pool of non-attributed (or informally managed) water. To describe the usage of canal losses, a flexible function \( f \) was tested with a broad range of parameterizations, the most general one of which is presented here. Two options exist to re-partition non-attributed water to farmers:

1. Each farmer receives water proportional to his share of water rights, as a factor that averages out reuse within each sector. This can be interpreted as institutional allocation: water managers know that canals are inefficient but also know that losses re-appear downstream. Thus, each farmer receives

\[
y = \lambda A^{\frac{\kappa}{A}}
\]

**Figure 1:** Redistribution effect of \( f^{\lambda_c} \), which benefits small farmers increasingly with larger \( \lambda \) (normalized to 1 ha; model only permits farms > 3.5 ha).
more than the canals actually deliver, hoping that with reuse, every right holder can benefit.

Such ‘institutionalized reuse’ increases effective water use efficiency, but does not re-distribute water endowments to those who lack these.

\[ f^{WR} = \frac{\sum_i Q^r \cdot WR^r}{\sum_i \sum_j Q^r \cdot WR^r} \]  

(2)

To test re-distribution based on farm size, a ‘root’-function relates farm area to the benefits from non-attributed water. For large \( \lambda_c \), small farms receive over proportionally

\[ f^{\lambda_c} = \frac{\lambda_c}{\sum_{i \in j} A_i} \cdot (A_i)^{1/\lambda_c} \]  

(3)

For the final model, an equal mix of both modes provided a robust pattern of results. Different parametrization of return flows and institutionalized reuse had only minor re-distributional impact between different farmer groups. For reasons of consistency with the original model, farmers access return flows and surplus water according to their area share within their irrigation sector:

\[ T_i = L_i + f^{WR} + f^{\lambda_c} \cdot C_j + \frac{A_i}{\sum_{i \in j} A_i} \cdot (S_j + RF_j) \]

For model evaluation, results were first analyzed for total flow quantities, such as time series of total legalized water supply \( \sum_i L_i \) and total non-attributed water \( \sum_j (RF_j + S_j + C_j) \). As a reference scenario, canal efficiency was decreased by 10 percent, to estimate the impact of canal improvement ex-post. This reference scenario was tested with ten different canal model parameterizations, to capture the structural uncertainty embedded in model assumptions. The policy scenario then uses the actual canal efficiency data, with the same ten structural model variations. Taking reference and policy scenario of the identical structural model variation, the performance of each individual agent was compared under both scenarios. Indices ranged from the share of expected water supply actually met, over the ratio of expected yields actually harvested, actual revenues and the ratio of planned and actual income. Finally, agents were grouped according to farm size stratum and according to the percentage of plant irrigation water demand that they can satisfy through legalized water rights.

3 Results

To model agricultural land use consistently with empirical data from both land cadastres and water registries, an additional process, e.g. the suggested mechanism with non-attributed water, must be taken into consideration. Only such a redistributive process makes agriculture physically feasible and economically viable for a sufficient number of farm agents. With this mechanism, only 322 agents quit farming over the study period (1996-2006). Without considering this water and while using the best water right registry data, about 2100 agents immediately quit farming.

To demonstrate the interplay of non-attributed and legalized sources of irrigation water under inter-annual variation of water availability, results from the calibrated model are decomposed for two cropping cycles: the hydro-meteorologically normal or typical season 1997/8 and a moderate drought season 1998/9 (Figure 2(a)). During the latter year, extension services report intense and sometimes violent conflicts amongst farmers over access to water. In following years, official records show an increased investment into irrigation equipment [CNR data, internal]. Irrigation water availability is plotted over time, for the total legalized water supply and for the pool of non-attributed water that is available, and also the sum of both as a total (full line). Note that in off-season, when none of the available water is used, much of the legalized water is counted twice because it becomes available again as non-attributed surplus water. Furthermore, flow quantities that farmers expect during their crop planning decision is plotted (dotted line). In the normal year, expected modeled flows and ‘real’ modeled flows coincide closely. In a dry year, agents strongly overestimate supply. Finally, irrigation water demand for both seasons is fairly similar (slashed line), peaking in December. While this demand is easily covered in the normal year, the water shortage during the dry year is significant.

\[ \text{The original model also re-distributed inefficiencies, using a reuse factor that was applied to the legalized water endowments (} T_i = L_i/u_j \text{). Such factor-based approach levels out different inefficiencies within the sector, redistributing from the less efficient to the more efficient. However, this mechanism did not provide water to farmers without rights.} \]
Figure 2: Model output depicts the total delivery of legalized water and the pool of non-attributed water over a normal and a dry cropping season. Furthermore, modeled ‘real’ water delivery is compared with the delivery that farm agents expect during crop planning.

The ratio of total actual flows to farmer’s expectations on total flows drops from 100% in a normal year to a meager 15%, which is even lower than the ratio of irrigation demand that can be satisfied (Figure 2(b)). Specifically, the ratio between non-attributed water and total water almost drops to zero, because farmers use up most of their entitlements and do not replenish the common pool. This highlights the increased vulnerability of those farmers who rely primarily on the spillover.

The impact of canal efficiency improvements is directly linked to the pool of non-attributed water and how users can access it. Canal improvements increase efficiency $\eta_c$ and thus always increase the supply of legalized water $L_i$. Water right owners can either use this additional supply to increase their irrigated area, irrigate more or abandon the additional water for the use of others. However, canal improvements also reduce the share of canal losses that provided spillover water.

For a use portion $\beta_c = 0.3$ and a moderate redistributive parameter $\lambda_c = 2.5$, individual impacts on farmers are evaluated. For illustrative purposes, farmers were assigned to 20 groups, according to their farm size stratum (sub figures in Figure 3) and according to the share of water they meet from water rights (lines). For each of the 3500 farmer agents, differences between the reference and the policy scenario were computed. Then a Gaussian distribution function was fitted for each group and its area was normalized to one for each stratum. This was done for two years, a normal one (solid lines) and a dry year (dotted lines). For most farms with significant water rights endowments, a 10% canal efficiency improvement increases water supply by about 5%, both in normal and in dry years. Farmers with no or a very low share of legalized water rights face more complex impacts: The small stratum that benefits overproportionally from using canal losses may even be impacted negatively by canal improvements, especially in dry years where surplus water generation is lower. Parameter sensitivity experiments show that this effect increases with smaller $\beta_c$ and also with larger $\lambda_c$.

The overall impact on water availability $\Delta T^i$ on an individual agent $i$ can thus be decomposed into three components, each of which fluctuates over time: (1) the increased delivery of river water $\Delta L^i$ because of improved conductive efficiency; (2) the increased overall availability of non-attributed water because those agents that already received excess water now leave even more surplus water $\Delta S_i$ in the canals once these are improved; (3) the decrease of non-attributed water because canal losses that originally produced non-attributed water $\Delta C_i$ are now eliminated.

$$\Delta T^i(t) = \Delta L^i(t) + \Delta S_i(t) - \Delta C_i(t)$$
Arnold et al. /Integrated Modeling of Access to Water

Specialized small farms (3.5-5 ha)
Small farms (5-25 ha)
Medium farms (25-60 ha)
Large farms (>60 ha)

Figure 3: The effect of 10% canal efficiency improvements (x-axis), graphed as distribution functions for a population of farmers normalized to 1 (y-axis). Farmers are grouped by their farm stratum (box) and the share of irrigation water they receive through legalized water rights (lines). Full lines represent a moderately dry year, the dotted lines a hydrologically representative year.

4 DISCUSSION

Perhaps the most important result is that agricultural land use can only be simulated consistently if informal mechanisms such as the allocation of non-attributed water as a source for irrigation is taken into account. However, for such informal water management arrangements and practices, data scarcity exists and it may not be feasible to quantify these through measurements. Thus, all results presented here are merely a qualitative demonstration of how access to non-attributed water, as abandoned or informally shared water and canal losses, may impact farmers. Processes such as water endowments, canal infrastructure, the hydro-meteorological condition, the distribution of water rights across a sector, and the behaviour of others together form the environment in which less endowed farmers must make a living. These processes sometimes function counter-intuitively.

From the policy viewpoint, this model confirms that canal improvements are beneficial to most farmers, especially to those that own water rights and are thus represented in water user associations. However, it also shows that many other farmers may be negatively impacted by improved infrastructure. Furthermore, it offers a direct explanation of why user organizations with many small water rights holders are not eager to improve their infrastructure: it could jeopardize the informal arrangements that benefit many.

The proposed model also reproduces Donoso’s observation of spillover use [2006] quantitatively, using very simple assumptions. From the integrated modeling viewpoint, the existence and relevance of non-attributed water was never a research objective. However, with analysis of both the hydrological water availability and farm-level data, the hydrological model and the economic model could not be reconciled with empirical data. This necessitated the broadening of the scope of analysis and the assessment of every process along the cause-effect chain. With validated water rights registries and other strong empirical data, integration across disciplines pointed to the dominant relevance of this interaction. This interaction may also explain why farmers are willing to engage in canal efficiency projects, which feeds back on this interaction process.

This paper identified another knowledge gap: As production input, non-attributed water fluctuates even stronger than ‘legalized’ water supply. The existing adaptive learning model performed relatively poorly, and many farmers had negative outcomes as a result of severe planning mistakes –
particularly commercial farm agents that invested heavily, based on short-sighted expectations (Figure 2(a)). However it is during drought conditions when conflicts between farmers are severe and local extension workers pointed out to us that informal arrangements become the cause of litigation between neighbors.

5 CONCLUSION

Empirically, this paper underlines that the Chilean model of privatized water rights is complex and must be assessed within its local context, in this case the Maule region. Informal arrangements may be as important as the legally prescribed water rights and management regime, especially for small, traditional farmers as the rural poor. Water access related to these informal arrangements may be an important reason why farmers support the improvement of canal infrastructure, which can benefit those with legalized water more than those without.

Methodologically, the integrated analysis elucidated an agent-to-agent feedback mechanism by extending assessment across a long chain of causes and effects. The interactions ‘between disciplines’ were especially relevant for poorer segments of the population that benefit over proportionally from a niche of informal arrangements. This only became apparent with a disaggregated analysis, combining a multi-agent model with detailed data on water endowments.

ACKNOWLEDGEMENTS

We thank the CGIAR Challenge Program on Water & Food and Alechile (DAAD/CONICYT) for financial support.

REFERENCES


Land use change modelling in an urban region with simultaneous population growth and shrinkage including planning and governance feedbacks

Dagmar Haase

Helmholtz-Centre for Environmental Research – UFZ, Department of Computational Landscape Ecology
Humboldt-University Berlin, Institute of Geography
dagmar.haase@ufz.de

Abstract: In the EU-project PLUREL we develop scenarios for future land use development in European urban regions facing population growth or shrinkage. We combine the spatially explicit modelling work with a participatory approach to involve stakeholder knowledge on drivers and policy instruments to steer land use development. Our scenario technique aims at incorporating feedbacks from planning into land use modelling. For modelling land use futures and options of the urban region of Leipzig-Halle, Germany, the MOLAND cellular automata model was used. To integrate stakeholder knowledge into the functionality of MOLAND and, so doing, to develop locally adapted land use storylines was the major purpose of a respective 1-day scenario workshop with local experts and practitioners. First, MOLAND model results stimulated the discussion about both impacts and steering options of future land development. Second, three predefined storylines were transferred into maps in a planning game. Participants created future land use patterns, assessed their drivers and impacts and reflected the instruments they would use to steer such a land use development. The resulting maps and the new knowledge about steering instruments were again used to improve the MOLAND model setup for Leipzig-Halle. So doing, we can feed stakeholders’ expert knowledge into the modelling work. Overall, quite positive experiences concerning the potentials of such a transdisciplinary approach have been made.

Keywords: land use change model; urban region; stakeholder feedbacks.

1. INTRODUCTION

1.1 The challenge: Land use pressures in growing and shrinking urban regions

Urbanisation is one of the most complex and dynamic processes of landscape change. Although only about 4% of the world’s land area is urbanised and densely populated (Ramankutty et al., 2006), we claim “the millennium of the cities” since more than half of the currently 6.6 billion world population is living in urban areas (UN, 2007; Kasanko et al., 2006). Projections for the future show that urbanisation – in terms of an increasing share of population living in urban and peri-urban areas – is very likely to continue (Batty et al., 2003; EEA, 2006; Lutz, 2001). The ongoing urbanisation process and megacity region growth (Hall and Pain, 2006) affects migration and population allocation between regions in Europe depending on their access to existing growth poles which often ends up in urban growth and peri-urbanisation. This macro-trend may be explained by the fact that economies are increasingly concentrated in urban centres (Hall and Pain, 2006). Moreover, pre-existing settlement patterns are important constraining factors. In some parts of Europe, most notably in Central and Eastern European countries, the fundamental political and economic changes towards capitalism and the market-oriented system lead to a massive
increase in suburbanization and urban sprawl (Tosics, 2004). All in all, this means an increase of urban land, share of impervious surface and of transport infrastructure. However, also urban shrinkage is increasingly observed in the industrialized world and today, shrinkage represents another major challenge for a sustainable urban development in Central Eastern parts of the post-socialist Europe, the former Rust Belt in the US or Russia (Rieniets, 2009; Kabisch and Haase, 2009; Blanco et al., 2009). Shrinkage is understood primarily as the quantitative process of population decline in an area which might be the result of very different processes such as de-industrialisation or demographic change (Haase et al., 2010). Population decline leads to a decrease in population density and to an oversupply/underuse of urban residential and commercial land, infrastructure and services. In consequence, wide-spread housing vacancies and large brownfields come to pass (Haase et al., 2007). In Leipzig, Germany, where in the 1990ies more than 60,000 flats became vacant and nationally financed large-scale demolition actions took place, the term of land use perforation was created (Lütke-Daldrup, 2001). Perforation describes a scattered land use development in the (former very dense) inner parts of the city. It comes clear that an accurate assessment of impacts of urban processes has to relate to both urban growth and shrinkage. Modelling tools are suitable to predict urban land use change and to get a better knowledge of the drivers behind one particular development (Haase and Schwarz, 2009). Regardless good progress in numerics and spatial representation there is still a big gap of integrating social science knowledge into urban land use modelling, particularly for the case of shrinkage (Schwarz et al., in press).

1.2 Land use change modelling in urban regions

As urban regions mostly are very densely populated and their land use components highly interlinked (Liu et al., 2007), developing views about their future is both a major concern in landscape research and a complex task. Modelling urban and peri-urban land use relationships helps to understand underlying drivers of land use change, to create future land use scenarios and assess possible environmental impacts (Lambin et al., 2006). A variety of land use change models, particularly for urban landscapes, already exist, ranging from specific case studies to generic tools for a variety of urban regions. These models differ largely in terms of their structure, their representation of both space and human decisions, and their methodological implementation (Haase and Schwarz, 2009). Compared to land use change models for rural landscapes, urban areas are shaped particularly by human activities, societal processes and human-environmental interactions (Couclelis, 1997).

In the EU-Project PLUREL we develop scenarios for future land use development in European urban regions facing population growth or shrinkage (Haase et al., 2009). So doing, we link the spatially explicit modelling with a participatory approach which involves stakeholders and their knowledge about drivers and policy instruments that are used to steer land use development. Our scenario technique aims at incorporating feedbacks from planning into land use modelling (Haase et al., 2009). For modelling land use options for the urban region of Leipzig-Halle, Germany, the MOLAND cellular automata model was applied. In order to adapt the pan-European scenario framework of PLUREL to the regional MOLAND application, stakeholder knowledge was indispensable to develop locally adapted land use future storylines. This was the major purpose of a respective 1-day scenario workshop with stakeholders, experts and practitioners (Haase et al., 2009).

1.3 Aim and structure of the Paper

The aim of this paper is to present a study where (1) a pan-European scenario framework and (2) the functionality of the MOLAND model were linked to (3) stakeholder knowledge in order to develop regionally adapted land use storylines and, respectively, to modify and to improve the MOLAND simulation by stakeholder input and feedback. So doing, in section 1 an introduction into the challenge of land use modelling in urban regions was given. In the second section, the methods are described, with the empirical case study (2.1), the MOLAND model introduction (2.2), the scenario framework (2.3), and the scenario workshop (2.4). The results and the feedback by the stakeholders are discussed in section 3 before the paper comes to some conclusions in section 4.
2. METHODS

2.1 Case study

For the coupled modelling-stakeholder-feedback study we use the test area of the urban region of Leipzig-Halle, Central Germany, which faces simultaneous population growth and decline. At a size of 4,390 km², the region has 1,073,000 inhabitants. On the whole, the region lost inhabitants since the 1970s, this was spurred by the political change in 1990 and subsequent de-industrialisation. The locally divergent population growth and shrinkage patterns have been accompanied by residential, commercial and infrastructure development. While the post-socialist transformation period with heavy urban sprawl has passed, moderate development in the peri-urban continues to the present-day (Couch et al., 2005). At the same time, considerable parts of the inner city faced a population outflow followed by residential vacancy, large urban brownfields and massive under-utilisation of urban infrastructure (Haase et al., 2007).

2.2 MOLAND model

MOLAND is a complex, computer based land use model, based on the Metronamica simulation software by RIKS, NL (http://www.riks.nl/products/Metronamica). MOLAND simulates land use change patterns for (in our case) 100x100m grid cells stratified into maximum 30 land use classes for time slices until 2025. Growth rates (land use pressures) are determined by statistical or expert based regional estimates using targets for the future, e.g. residential land use based on population growth scenarios where additional demands for residential land per capita are based on the new (future) population projections (Barredo et al., 2003; Petrov et al., 2009). Planning and governance strategies are incorporated based on a regional scenario workshop enabling a ‘regionalization’ of the PLUREL pan-European scenarios which are explained in section 2.3. Thus, planning is translated into land use neighbourhood attraction curves and suitability maps that stand for planners’ decision-making. The model outcomes can be used to assess the effectiveness and suitability of the strategies by the planners. At the same time, planners can feed their ideas and responses on the model results back to the modellers to modify the simulation.

2.3 Scenario framework

Scenarios are ‘stories of the future’, providing a tool for the investigation of possible future conditions and trends, risks and opportunities. They can take different forms such as (a) stories or narratives which can be more fictional, or more realistic, (b) models (technical and quantitative, or, more conceptual and qualitative, (c) visual or narrative images or (d) visions (Ravetz and Rounsevell, 2008). The scenario framework for the land use modelling used in the PLUREL project is twofold (Figure 1): (1) Four scenarios of pan-European trends up to the years 2025 and 2050 with relevance for urban regions provide the framework for the investigation of future trends in land use and its related functions, including risks and opportunities for sustainable development. These scenarios are based on the IPCC SRES scenarios (IPCC, 2007) and can be ordered on two main axes which may be called a governance axis and a value axis. Within PLUREL, the four basic IPCC SRES A1, B1, A2 and B2 scenarios have been extended based on the development of ‘shock’ scenarios and dedicated to particularly urban regions (more information see Ravetz and Rounsevell, 2008). The shocks will allow an analysis of changed drivers that potentially are of great significance to urban-rural land use and represent a novel development of regionally-interpreted SRES scenarios: A1 – rapid development in information and communication technology and rapid counter-urbanisation (Hypertech), A2 – self-reliance (extreme water crisis) due to rapid climate change and water crisis, B1 – sustainability (energy price shock) due to extreme energy prices it comes to a localisation of activity, B2 – fragmentation (social exclusion) due to a pandemic disease followed by a polarisation of cities. Four related narrative storylines are used for the modelling on economic, demographic, and environmental and land use changes (Ravetz and Rounsevell, 2008).
(2) Based on the aforementioned pan-European ‘shock’ scenarios, at the case study level, the storylines were further developed and adapted which is meant by ‘regionalised’. One of the case studies is Leipzig-Halle. In doing so, the population development was adapted in terms of specific in- and outmigration dynamics of specific age groups, localised GDP values were added for each municipality in the region, and transport and peri-urban commercial projects were considered as axial development factors. In addition, the regional planning policy – that is land use planning in form of zoning or regulation but also land preparation plans and financial subventions – was included into the scenarios (which was completely neglected in the pan-European approach due to the scale). So doing, first and foremost flood maps, protection areas and green space networks were incorporated (Petrov et al., 2009). Figure 1 shows the details of the locally adapted scenario storylines which specify the above mentioned assumptions.

### 2.4 Scenario workshop

The scenario workshop’s main objectives were to discuss the MOLAND land use modelling results and maps together with local stakeholders, experts and practitioners. It generally contributes to the dialogue between research and practice on spatial and land use development in the urban region of Leipzig-Halle. The above introduced land use scenarios as well as the MOLAND modelling were presented by scientists in order to stimulate a debate on the implications of future land use development. Policy options for regional planning and governance were discussed. The central element of the workshop, a planning game, was conceptualised to jointly develop future land use patterns for the Leipzig-Halle region under the diverging assumptions (Haase et al., 2009).

The three land use development options for the scenario workshop where developed under consideration of the locally adapted scenario storylines given in Figure 1. Using the same driving forces, the following options were realised by the stakeholders: “Unrestrained growth”, “Managed growth” and “Managed shrinkage”. In practice, according to the scenario, every planning game was supplied with a certain amount of “land use” in the form of coloured paper and a storyline. The stages and the results of the planning game were documented in minutes, photographs and films (Haase et al., 2009). The main idea behind the scenario workshop was to let the practitioners create their own land use maps framed by the land use options. The scientists were interested into the facts where the stakeholders place future housing and transport growth and where they assume land perforation and brownfields will occur and why. In addition, the practitioners listed the instruments they would use to steer the before created land use maps (Haase et al., 2009).

<table>
<thead>
<tr>
<th>Planning Drivers</th>
<th>No restrictions</th>
<th>Current planning system implemented</th>
<th>Strong planning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Growth (population, GDP)</td>
<td>Hypertech Unrestrained growth In-migration new transport axes, commercial investment, gentrification</td>
<td>BAU Managed growth stable population, weakening of protection zoning</td>
<td>Price shock Managed shrinkage Declining population, eco-tourism growth, investment in public transport, urban renewal</td>
</tr>
</tbody>
</table>

Figure 1: The scenario framework (BAU = Business-as-usual): PLUREL shock scenarios in bold, regionalised and locally adapted main ideas of the scenario storyline below.
But how do the land use options used at the scenario workshop relate to the above mentioned pan-European and the regionalised scenarios? They can be understood as an extension of both but rather represent qualitative visions how the region might look like in 2025. Both the regionalised storylines for the MOLAND modelling and the land use options of the stakeholder scenario workshop are compatible so that they can be fed back into the pan-European scenarios (Haase et al., 2009).

2.5 Feedbacks

The reactions of the stakeholders on the simulated MOALND maps, their own manual maps created at the scenario workshop and, last not least, the listing of the drivers and instruments that steer the future land use development (which are indirectly also a reaction on the simulated land use maps so far) are defined as the ‘stakeholder feedback’ in this paper. Respectively, they are presented and evaluated in the results’ section.

3. RESULTS

3.1 Model results

Figure 2 shows the results of the spatially explicit MOLAND land use change modelling for the urban region Leipzig-Halle. Overall, an increase of sealed and built surfaces can be found. Whereas in the ‘Business as usual’ (BAU) scenario the area of discontinuous, low density residential land increases most, in the ‘Hypertech’ scenario most of the growth is due to commercial land development.

![Land use change maps for the time horizon 2025 for two locally adapted scenarios, Business-as-usual and Hypertech, compared to the start year of the simulation 2001.](image)

The ‘Price shock’ scenario shows no further settlement growth but an increase of urban green space instead. As the ‘Hypertech’ scenario assumes a somehow extreme development this goes along with a stronger segregation of the land use and thus the core cities of Leipzig and Halle are faced with no further building activities in their centres, while residential land in the peri-urban area is growing (Figure 3). This inner-urban decline of sealed surface does not occur in the ‘BAU’ scenario where a stronger impact of land use planning is assumed. Most of the built-up and sealed land comes either from agricultural areas, the natural green spaces and floodplains in the cities’ surroundings or the former
sites of mineral extraction which are more or less continuously declining in both scenarios. In the ‘Hypertech’ scenario, the expansion of the Leipzig-Halle airport leads to a big step of land consumption in 2010. In terms of open land, we find a decrease of up to 25%. Natural green spaces, in particular, decline up to 60%. The most effective wetland protection (196%) we find in the ‘Price shock’ scenario where commercial and transport growth is limited (Figure 3).

In the scenarios, the trend of population, residential and commercial growth is prescribed and MOLAND “translates” these trends into numbers of residential and commercial or industrial cells. The way the new cells are allocated in the area is calculated endogenously in the model based on the transition potential driven by suitability, accessibility, zoning, neighbourhood cells and random perturbation (Barredo et al., 2003).

### 3.2 Results of the scenario workshop

Table 1 summarises and compares the results of the stakeholder scenario workshop. Concerning land use, commonalities between the three scenarios were as follows: persistence of existent urban centres, new development mainly within the centres, decline of rural areas, partial abandonment of peripheral settlements. Differences were found mainly in the increase of green spaces and water bodies in the ‘managed’ scenarios, more polarisation of residential and economic land uses, social segregation, urban sprawl in the unrestrained growth story; Table 1):

**Unrestrained growth**
Urban sprawl increases within the Leipzig-Halle beltway. However, the cities of Leipzig and Halle remain the most important centres while rural areas face further decline. At the same time, business locations and residential areas undergo a process of differentiation and polarization. Decay and slums on one hand and growth and gated communities on the other coexist in close proximity (Table 1).

**Managed growth**
The regional development is based on the persistence of the existing structures. The recycling of sites opens up further land potential within existing settlements. Overall, there is a concentration of residential and commercial development within urban core areas. A network of green spaces structures the urban region, enhances the cultural characteristics and the ecological functions of the landscape (Table 1).
Managed shrinkage

Business locations and housing estates concentrate in existing urban cores and along spatial axis. The number of administrative centres and infrastructure sites is reduced. Villages and estates on the periphery of the urban cores disintegrate. Green spaces permeate the region, the access to waterways and lakes is improved (Table 1).

Table 1: Converging and diverging arguments during the game between the three scenarios with respect to land use changes, drivers, land use steering instruments and its valuation by the participants

<table>
<thead>
<tr>
<th>Land use changes envisioned</th>
<th>Unrestrained growth</th>
<th>Managed growth</th>
<th>Managed shrinkage</th>
</tr>
</thead>
<tbody>
<tr>
<td>• urban sprawl, in particular in corridors between the centres</td>
<td>• persistence of overall settlement structure</td>
<td>• persistence of centres</td>
<td></td>
</tr>
<tr>
<td>• persistence of urban centres</td>
<td>• new residential and commercial development within existent centres</td>
<td>• new residential and commercial development within existent centres</td>
<td></td>
</tr>
<tr>
<td>• rural areas face severe decline</td>
<td>• strengthening of green space network</td>
<td>• further suburban development only in exceptional cases</td>
<td></td>
</tr>
<tr>
<td>• polarisation of business locations and residential areas within the settlement structure</td>
<td></td>
<td>• shrinkage of medium-sized settlements, abandonment of peripheral settlements</td>
<td></td>
</tr>
<tr>
<td>• rich and poor residential areas in proximity</td>
<td></td>
<td>• increase of green spaces and water bodies in and around settlements, well connected to centres</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Drivers identified</th>
<th>Unrestrained growth</th>
<th>Managed growth</th>
<th>Managed shrinkage</th>
</tr>
</thead>
<tbody>
<tr>
<td>• positive net migration</td>
<td>• sustainability widely accepted as a guiding principle for public policymaking and planning</td>
<td>• out-migration of population and businesses</td>
<td></td>
</tr>
<tr>
<td>• intense competition of investors over business locations</td>
<td>• cooperation between investors, politicians and citizens</td>
<td>• realisation of housing preferences</td>
<td></td>
</tr>
<tr>
<td>• quickening of land use cycles due to rise and fall of new technologies</td>
<td>• attention to residents’ housing preferences</td>
<td>• limited mobility</td>
<td></td>
</tr>
<tr>
<td>• privatisation of public property and services</td>
<td></td>
<td>• local authorities under financial constraints, decreasing means to influence land use</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Instruments to be used</th>
<th>Unrestrained growth</th>
<th>Managed growth</th>
<th>Managed shrinkage</th>
</tr>
</thead>
<tbody>
<tr>
<td>• privatisation of public services</td>
<td>• federal state, regional and local government reform</td>
<td>• regional parliament</td>
<td></td>
</tr>
<tr>
<td>• decreasing influence of formal planning and recommendations instead of rules</td>
<td>• regional planning for entire region</td>
<td>• federal state, regional and local government reform</td>
<td></td>
</tr>
<tr>
<td>• urban and regional marketing</td>
<td>• transfer of responsibilities in planning from local authorities to regions</td>
<td>• influential role of urban region in politics</td>
<td></td>
</tr>
<tr>
<td>• joint infrastructure planning</td>
<td>• pool of compensation areas</td>
<td>• regional planning for whole region</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• commitment to the protection of green spaces</td>
<td>• joint infrastructure projects</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• continuation of landscape planning</td>
</tr>
</tbody>
</table>

3.3 Feedbacks on drivers and steering instruments by stakeholders

The original assumptions of the MOLAND simulations could be improved by the above listed arguments discussed during the stakeholder process. The following drivers of land use development in the urban region Leipzig-Halle were already incorporated in the model: migration, population growth, GDP and urban regeneration programs. The following drivers were identified by the stakeholders: cooperation between investors, politicians and citizens and concerted actions of urban renewal in form of demolition.

But, whereas in case of population and economic growth investor competition and new technologies were highlighted by the stakeholders, in case of population decline the role of normative planning principles, of private and small-scale investors became more important (Table 1). In the case of economic (and population) growth, commercial development relies on positive netmigration to the region because the local population is diminishing. An intensive competition over business locations increases land consumption for roads and sites, fuelled by short term investment incentives. New technologies and media lead to shorter land use cycles. Financially weak, local authorities seek a further privatisation of
land and services, thereby reducing the public and strategic planning options. In case of a more managed growth the spatial development follows the principle of sustainability. Investors, politicians and citizens have recognized that the long term ecological need and cost benefit of reduced land consumption; they seek cooperation and a better incorporation of housing preferences by the residents. Particularly, the stakeholder feedback concerning the “managed shrinkage” option was interesting: Stakeholders highlighted that the shrinkage of the housing sector – that is predominantly demolition – has to better communicated to service suppliers. In case of further shrinkage economic performance, housing sector and financial transfers are too low to ensure a consistent high level of infrastructure provision in the whole area. The municipal capacities for action are reduced. Within the region there is polarisation. Weak sites cannot be sustained anymore and have to be dismantled (cf. again Table 1). The feedback of the stakeholders here can be regarded as very positive in terms they modified the residential growth and amplify the land abandonment process assumed in the scenarios. Thus, demolition plans were incorporated in the new MOLAND simulations.

4. DISCUSSION AND CONCLUSIONS

Consequently, in a next step, the knowledge/feedback gained in the scenario workshop was incorporated in the MOLAND model to improve the simulation results. This is still work in progress and therefore, in this paper first preliminary results are presented: So, for example, we incorporated residential preferences and demolition plans as suitability and zoning maps into the MOLAND model. Figure 4 proves evidence that we can show inner-urban decline and (empty) brownfield cells – compared to the original ‘Hypertech’ simulation. With respect to the spatial resolution of the model, 100x100 meters, the brownfields in the new simulation represent a realistic picture of a) how demolition already perforated the urban structure in the inner parts of the city and b) where it actually occurs.

![Figure 4: Two runs of the ‘Hypertech’ scenario: the left side map shows the original simulation and the right side the improved simulation including the stakeholder feedback gained at the scenario workshop which clearly shows a loss of continuous urban fabric and land abandonment in the inner parts of the city (dark red and white cells; colours of the land use classes see Figure 2).](image-url)

Further improvements of the MOLAND model will be done using the information gained at the scenario workshop, first and foremost in terms of a) including new zoning features in the model, b) modifying land use suitability of specific areas, c) modifying land use transition rules according to the expected lowering of peri-urban dynamics compared to the inner-urban.
ACKNOWLEDGMENTS

This work is part of the PLUREL Integrated Project (Peri-urban Land Use RELationships) funded by the European Commission, Directorate-General for Research, under the 6th Framework Programme (project reference: 36921). The authors would like to thank all colleagues working on the MOLAND modelling and in the Leipzig-Halle case study within the PLUREL project. Thanks go further to three anonymous reviewers whose comments very much helped to improve the quality of the paper.

REFERENCES


Linking wildfire behaviour and land-use modelling in Northern Mongolia

C. Schweitzer \(^a\) and J.A. Priess \(^a\)

\(^a\) Helmholtz Centre for Environmental Research (UFZ), Department Computational Landscape Ecology (CLE), Permoserstr. 15, D-04318 Leipzig, Germany (christian.schweitzer@ufz.de)

Abstract: Numerous approaches of modelling wildfires have been published covering the regime itself, ignition probabilities, spreading patterns, risks and impacts. However, there is less research linking these validated mature approaches to dynamics of the terrestrial environment. We contribute to filling this gap via integrating a newly developed wildfire module into a land-use model, enabling us to include wildfire-impacts in dynamic simulations of socio-environmental systems. In the forest steppe region of Northern Mongolia, wildfires are a major concern, threatening grassland and forest areas, which are already under pressure of droughts, heavy grazing, (illegal) logging and increasing firewood demand. We employ the generic land-use modelling framework SITE, that includes the ecosystem model DayCent and other components, to develop a wildfire sub-module for simulating wildfire spread and intensity. Outputs are translated into net loss of biomass, changes in carbon and returns of nutrients. Burning and carbon cycling affect biomass and thus fuel load and wildfire risk in subsequent years. Our study presents first results of a coupled land-use-wildfire modelling approach aiming at (i) increasing the model accuracy in land allocation and potential land usability, and (ii) to a more adequate impact analysis based on potentially fire-affected land.

Keywords: Land-use modelling; Wildfire; Mongolia; MODIS burned area

1 INTRODUCTION

Wildfire is a paradox; it kills plants and animals and can cause wide-ranging damages to the ecosystem. On the other hand it can be very beneficial in terms of nutrient recycling and forest regeneration [Rowell and Moore, 1999]. In some areas, natural wildfires have historically adapted with ecologically positive effects. Other ecosystems are susceptible to severe damages, causing a local extinction of species or considerable changes in ecosystem functions (e.g. soil, hydrology). Integrated modelling approaches could provide helpful insights in wildfire-environmental interactions. Globally, the majority of wildfires are caused by human activities in a direct or indirect form. An anthropogenic influenced wildfire regime (frequency, distribution) will potentially affect human acting. This inter-relationship between humans and wildfires has initiated many scientific studies. Millington et al. [2008] mention in their study, presenting an agent based approach simulated land-use management influencing wildfire risk, that only a few models exist who consider human activities and the interactions with vegetation-wildfire dynamics.

In this paper we present preliminary results from a modelling approach which captures the wildfire behaviour in Northern Mongolia. The approach aims at analysing impacts of wildfires on the socio-environment, including feedbacks related to carbon dynamics, biomass availability (in forests and grasslands) and the effects on land use. Therefore we newly developed a wildfire module on the basis of a well established wildfire model and linked it to our dynamic land-use model integrating new model capabilities of simulating...
wildfire spread and intensity. In the following sections we present the general design of our approach including first results from wildfire risk and wildfire behaviour simulations, including impacts on biomass availability and forest use.

2 STUDY REGION

As living in Mongolia is very much dependent on the biophysical environment, the country is extremely vulnerable to natural disasters. Cold winters with heavy snowfall, hot summers marked by droughts and floods frequently set the environment, people and the economy under pressure. Additionally wildfires threaten forests and grasslands which have a high ecological and economic value for the country as they provide firewood or grazing opportunities for the omnipresent nomadic lifestyle.

The presented study was carried out in the Kharaa river basin (105°15’E, 48°41’N) which is located in the forest-steppe region of Northern Mongolia, approximately 30 kilometres north of the capital Ulaanbaatar. The climate is semi-arid, characterized by a mean annual precipitation of 250-300 mm and a mean annual temperature of 0.4°C. The total area is 15,000 km², covered 60 % by grassland, 26 % by forest and 11 % by arable land. Population trend is increasing due to the vicinity of the capital, promising trade and job opportunities. Nearly half of the population (70,000 in 2006) could be characterized as rural, living a modern nomadic lifestyle. For them grazing opportunities and firewood are essential environmental goods. Besides providing wood for cooking, heating and construction, forests have an important hydrological function as they are the source of runoff generation, providing the whole basin with drinking water which is usually extracted from the surface waters by the rural population.

To study the disturbances by wildfires in forest and grasslands, we analysed the historical wildfire regimes using a satellite approach based on the ‘MODIS Collection 5 Burned Area Product - MCD45’ (henceforth, MODIS burned area) [Roy et al., 2008], which detects the approximate day of burning at a spatial resolution of 500 meters (since 2000). Figure 1 shows the total annual area burned (A) and the wildfire seasonality for the observation period 2001 to 2008 (B). Over the time period 2006-2008 the occurrence of wildfires has increased considerably in forested areas (A). Within the observed period, 2.7 % (40,600 ha) of the study area was affected by wildfires, 60 % occurred in forested areas and 40% in the grassland steppe region. The seasonality of wildfires (B) shows a clear pattern of two wildfire seasons, spring and autumn, which is reported as a typical phenomena observed in Mongolia [Goldammer, 2001].

![Figure 1](image-url)

**Figure 1.** Area statistics and wildfire seasonality extracted from MODIS. (A) - Total annual area burned. (B) - Wildfire seasonality for 2001-2008. Note: Due to poor pixel quality and sensor calibration tasks, the year 2000 was not included. The year 2004 could be neglected due to the absence of wildfire scars.
3 DATA & METHODS

The simulation of wildfire behaviour is depending on various, mostly highly uncertain factors such as fuel availability, fuel moisture, weather, and ignition probability. Due to the fact that the land-use model itself offers a high level of complexity, the challenge was to implement an accurate and efficient (in terms of computing time) spatially explicit wildfire spread model, and link its outputs to a process based ecosystem model. The core components are presented in the following paragraph, emphasising the wildfire sub-module and its data requirements. For a detailed description of the data used by the land-model and by the third party model DayCent, please refer to Priess et al. [2010] and Schweitzer and Priess [2010].

3.1 Land-use model

As a platform for model integration and development we use the SITE-Framework (Simulation of Terrestrial Environments), a generic modelling platform for spatially explicit land-use modelling [Mimler and Priess, 2008, Schweitzer and Priess, 2010]. The regional land-use model (SITE-Mongolia) was developed in this framework with the objective to study the dynamics of historical, current and future land-use and land-cover changes including the impacts on water resources [Priess et al., 2010]. Simulations are performed on a 1km x 1km grid, allocating land-use decisions annually following a three step process. First a multi-criteria analysis is carried out for each land-use class and each pixel individually, calculating dynamic suitability maps for each time-step. The resulting normalized weighted values enable a direct comparison and competition between land categories. In the second step, sub-modules are executed (e.g. crop, grassland, settlement, forest) computing land allocation driven by the demand for commodities, space for housing or agricultural products. Finally the linked ecosystem model DayCent [Parton et al., 1998] computes daily plant growth and calculates yield, biomass and carbon feedbacks in cropping systems, grasslands and forests.

3.2 Wildfire sub-module

A new ‘wildfire sub-module’ was developed in addition to the existing sub-modules in SITE-Mongolia to simulate wildfire behaviour and analyse feedbacks (Figure 2). The simulation of wildfire behaviour is based on spreading algorithms described by Rothermel [1972], using a semi-empirical mathematical approach. For our study we use a modified version which is derived from the BEHAVE fire model [Andrews, 1986] and optimized for highly iterative cell-based fire growth simulations [Bevins, 1996]. The new wildfire sub-module consists of several components: (a) a file handler which executes the third party models (e.g. the wind model) including necessary pre- and post-processing steps, (b) the ‘risk analysis’, which performs a multi-criteria analysis to identify the cells which are most ‘suitable’ to burn, (c) a wind model to simulate wind speed and direction for the location of each grid cell, and (d) the wildfire model itself, executing the routines for predicting the spread rate and intensity of free-burning wildfires. Finally outputs of the wildfire sub-module (spread, intensity and flame length) are translated into net change in carbon, which is implemented using a lookup-table to establish the link to the DayCent model. Figure 2 presents a simplified scheme of the modified SITE-Mongolia model. The wildfire sub-module is executed first, calculating wildfire risk maps and wildfire behaviour for the current year. The suitability analysis then excludes wildfire affected cells in the grassland and forest use suitability. Burning intensity is translated to net loss in biomass using the DayCent model. Alteration in biomass availability influence wildfire risk, fuel availability, fuel moisture and affect land-use decisions in subsequent years.
3.2.1 Wildfire risk analysis

In a first step we perform a ‘wildfire risk analysis’ which has two objectives: (i) the identification of areas showing a significant fire risk, (ii) identification of highly suitable cells for potential wildfire ignition. Wildfire risk is computed as a function of three weighted independent categories resulting in a normalized ‘overall fire risk’ (OFR) for each cell (ranging between 0 for no risk and 1 for high risk) comparable to the suitability assessment of the SUIT module (Figure 2). The three OFR categories are: (i) Fuel availability (ii) Weather and (iii) Location. Each category consists of multiple weighted input datasets. A complete calibration and validation process (to be performed in future) can achieve the exact definition of weights for each category, which are currently weighted equally. The estimation of fuel availability is based on biomass. Therefore total forest and grassland biomass are used in combination with respective ratios of live to dead biomass. Weather factors (temperature, relative humidity and wind) generally belong to the most important variables influencing wildfire behaviour. As wind parameters cannot be sufficiently aggregated on an annual scale, we use the relative humidity and the mean annual solar radiation to determine the risk from weather factors. To capture the effect of humidity we count the number of days supporting high flammable conditions (relative humidity <= 45%) for each cell in the current year. Topographic effects influencing the moisture will be reflected in the mean annual solar radiation provided for each cell. Location is used as a representative for spatial ignition patterns. Since most wildfires in Mongolia are related to anthropogenic activities, a distance analysis was performed. Hussin et al. [2008] have reported (from a district located in the study area) a positive relationship between wildfire occurrence and distance to road and distance to rivers and a negative relationship to distance to settlements. In our analysis these factors were used besides an additional settlement buffer excluding areas where wildfires are very unusual.

3.2.2 Wildfire input data

As mentioned above, wildfire modelling is dependent on detailed input to reflect the environmental conditions for starting and propagation of wildfires. In the following subsections we describe the input data required by the wildfire sub-module for the simulation of wildfire behaviour.

**Topography:** An important factor influencing the direction and speed of the wildfire spread is the terrain (slope and aspect). Combinations of wind effects and changes in slope result in the fire propagation exposing potential fuel to additional convective and radiant heat [Rothermel, 1972].
Fuel model: The ‘Fuel Model’ (FM) provides an abstract description of the fuel availability in selected land cover types. Each FM represents a mathematical function that predicts spread and intensity. A numerical value is linked to one of the 13 predefined FMs. Each FM is characterized by the fuel loading for each particle, diameter size class, the surface-area-to-volume ratio, the fuel bed depth, the heat content and the moisture of extinction [Bevins, 1996]. The FMs which have been used in this study are: ‘Short Grass (0.3 m)’, ‘Tall Grass (0.76 m)’ and ‘Timber (grass & understory)’.

Fuel moisture: Fuel moisture is a key factor influencing wildfire propagation and assessing wildfire risk [Chuvieco et al., 2003]. In the model this parameter is described by five input maps consisting of three dead and two live fuel moisture classes. Dead fuel moisture is classified by time-lag, which reflects the time taken by fuels responding to a specified amount to changes in moisture, which is correlated to the burning materials diameter (e.g. 1-hour, 10-hours, 100-hours). The two live categories represent the fuel moisture in herbaceous and woody components. We estimate the fuel moisture of fine dead components (1-hour) by using a simple index developed by Sharples et al. [2009] which includes temperature and humidity for the day of ignition. For the other categories we use a lookup table which reflects the moisture related to the phenological stage of the plant.

Wind direction and speed: Due to the poor availability of high resolution (spatial and temporal) wind data in the region, a wind simulation model to estimate average wind speed and direction for the day of ignition has been applied. We integrated the WindNinja model [Forthofer et al., 2009] which simulates micro-scale winds in mountainous terrains. As input the model requires a digital elevation model and user specified wind speed and direction values. We are glad to be able to use at least one weather station (Baruunkharaa), located in the centre of the catchment, that provides daily wind speed and direction data.

Ignition: To avoid stochastic influences in the model, our approach derives ignition patterns from two factors: (i) frequency and distribution of wildfires occurrences in the last years and (ii) competition of cells corresponding to their OFR. To derive the first factor, single ignition points were extracted from the MODIS data. The algorithm used, identifies all pixels which belong to one fire scar (spatially and temporal) and creates a ‘fire-cluster’. Within this cluster one cell with the earliest date of burning is selected as potential location and time for ignition. For the period 2000 to 2008 it is observed that on an average 16 ignitions occurred annually (60 % in spring and 40 % in autumn).

4 RESULTS

Model runs were performed for risk analysis and wildfire behavior using the above model concept. Here we present two segments of results from SITE-Mongolia simulations:

(1) Wildfire risk and behaviour

(2) Wildfire effects on forest biomass, land allocation and forest use

4.1 Wildfire risk and behaviour

Wildfire risk was simulated (see Section 3.2.1) and risk values calculated for all three categories and OFR (Table 1). For validation of simulated values we use two sets of grid cells: (i) cells corresponding to wildfire detection by MODIS (Set I) and (ii) cells corresponding to no detection of wildfire by MODIS with OFR >0 (Set II). Set I and Set II are compared for risk categories and mean OFR.
### Table 1. Comparison of risk categories and OFR between Set I and Set II.

<table>
<thead>
<tr>
<th></th>
<th>fuel availability</th>
<th>weather</th>
<th>location</th>
<th>overall fire risk (OFR)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Set I</td>
</tr>
<tr>
<td>2001</td>
<td>0.40</td>
<td>0.31</td>
<td>0.59</td>
<td>0.58</td>
</tr>
<tr>
<td>2002</td>
<td>0.38</td>
<td>0.31</td>
<td>0.59</td>
<td>0.57</td>
</tr>
<tr>
<td>2003</td>
<td>0.40</td>
<td>0.31</td>
<td>0.56</td>
<td>0.58</td>
</tr>
<tr>
<td>2005</td>
<td>0.39</td>
<td>0.31</td>
<td>0.53</td>
<td>0.49</td>
</tr>
<tr>
<td>2006</td>
<td>0.37</td>
<td>0.31</td>
<td>0.69</td>
<td>0.58</td>
</tr>
<tr>
<td>2007</td>
<td>0.39</td>
<td>0.31</td>
<td>0.58</td>
<td>0.55</td>
</tr>
<tr>
<td>2008</td>
<td>0.41</td>
<td>0.30</td>
<td>0.54</td>
<td>0.59</td>
</tr>
<tr>
<td>Ø</td>
<td>0.39</td>
<td>0.31</td>
<td>0.58</td>
<td>0.56</td>
</tr>
</tbody>
</table>

Note: The average number of cells used (denominator) for Set I is 117 and Set II is 12449.

From Table 1, it can be observed that despite the large difference in the denominator for calculation of means, all single categories and OFR consistently record a higher value in Set I than Set II. This observation validates the accuracy of risk simulation across the range of categories used. Furthermore, we carried out a spatial validation. Here we present results for the year 2006 as it best represents - (a) presence of both fire seasons (b) area burned (c) distribution of wildfires (spatially and temporally). It is important to note that the spread is dependent on conditions occurred on a specific day within the pixel that is ignited. Hence, not all points of ignition in the wildfire model imply that spreading would occur. Only ‘suitable’ conditions (e.g. low fuel moisture, wind) lead to immediate fire spread failing which, fire is extinguished. Figure 3 shows the simulated area burned and fire behaviour in two areas of interest (A, B) for the year 2006 (left image). Total area burned from model simulations is 17,500 ha. Modelled variables - fire intensity (A2, B2) and flame length (A3, B3) - are important for further calculations of net change in biomass, carbon and nitrogen using the DayCent model. Comparison of simulated burned area (17,500 ha) against MODIS burned area (8,150 ha) show that model simulations to some extent overestimate the burned area in 2006.

#### Figure 3. Simulated results of fire spread (left) and behaviour (right: A2, A3, B2, B3) for the year 2006. For comparison A1 and B1 presents the satellite derived area burned and the corresponding month of burning.

### 4.2 Wildfire effects on forest biomass, land allocation and forest use

In order to highlight the effects of wildfire on the socio-environment, we have chosen the forest sector to demonstrate some impacts. The forest module is driven by the demand for
wood, which is defined as the sum of industrial demand (data from regional statistics) and demand for firewood (linked to rural population dynamics).

We simulated the total aboveground forest biomass including and excluding the wildfire sub-module. Figure 4 presents the comparison of both model runs, showing an increasing trend in diverging biomass availability, which could be related to the amount of wildfire occurrences extracted from MODIS (presented in Figure 1-A). As we know from MODIS observations, wildfire affected areas increase in subsequent years (2007, 2008) which may increase the difference. Due to the lack of daily climate data we are not able to continue the time series for the latter years until now. A considerable reduction in total biomass, due to external disturbances (timber extraction, wildfires) will implicate a land-cover change from ‘Closed Forest’ to an ‘Open Forest’ class in our model. ‘Closed Forest’ is a coniferous type, while ‘Open Forest’ is characterized as mixed, mostly broadleaf (secondary) forest. We observed an additional ‘Open Forest’ allocation of 24% with the new model setup, indicating the indirect effects on land allocation. Furthermore we analysed the distance from settlements to ‘highly suitable’ (> 10% of max. suitability; providing sufficient biomass and adequate re-growth for a sustainable management) forest-use cells to explore changes in utilization activities. We observed (for the period 2001-2006) a 3% (Ø 350 m) farther distance to ‘highly suitable’ cells. We conclude that an increase in wildfire disturbances in Northern Mongolian forests will enlarge the effort related with firewood collection and influence transportation costs in commercial timber production.

5 DISCUSSION & CONCLUSIONS

In this paper we present first results from the process of integrating a wildfire sub-module into a dynamic land-use model, enable us to study feedbacks to the socio-environment. Despite the difficulties (mostly of technical nature) associated with integrating third-party applications into an existing model, we demonstrate that the above concept adds value to the overall model approach in terms of (i) a better estimation of biomass availability, (ii) an improved allocation of land (e.g. ‘Open Forest’ and burned cells) and (iii) addressing changes in the usability of land. Furthermore we expect scientific benefits in: (iv) providing accurate land-use and land-cover information for hydrological modelling purposes, (v) supporting the development of a satellite-based wildfire monitoring concepts (vi) simulation of environmental scenarios. Our results from wildfire simulations indicate good accuracy for fire risk analysis. Simulated burned area in 2006 is higher (53%) than MODIS burned area at first glance. However, several factors may be possible contributors to this observation. Burned area from MODIS uses surface-reflectance dependent algorithms which may contribute to underestimations. Furthermore, we use only ‘High Quality’ pixels (most confidently detected) from MODIS for validation. If burned area is extracted with all detection levels, estimations sum up to 15,375 ha. Hence, it is reasonable to conclude that simulated burned area is in an acceptable range. We also like to mention existing limitations corresponding to the wildfire sub-module. The current state of implementation does not enable handling dynamic input data, (e.g. wind speed and direction) which may affect the spreading pattern simulated compared to real world wildfire scars. Secondly, fuel moisture which is one of the critical variables of fire behaviour modelling needs calibration. The genetic algorithm implemented in the SITE-Framework could be applied to calibrate the fuel moisture index used in the study. Strengthening the human influence in the OFR (since we know that most fires are anthropogenic origin), could be achieved by assigning different weights to risk categories (e.g. location risk).
The increase in wildfire occurrences in Mongolia has initiated a national wildfire satellite-based monitoring system, which is operational since a few years now. Modelling approaches, as the one presented here, could support these efforts providing an appropriate tool to study the impacts and feedbacks to the socio-environment, identify interactions responsible for the increasing wildfire risk and occurrence.

Acknowledgements

The authors thank Subhashree Das for the valuable comments to this manuscript and Jason Forthofer and Kyle Shannon from the USDA Forest Service for providing the latest version of the WindNinja model.

References


Modeling the adaptation of land-use decisions to landscape changes using an agent-based system: a case study in a mountainous catchment in central Vietnam

Quang Bao Le, Roman Seidl, Roland W. Scholz
Natural and Social Science Interface (NSSI), ETH Zurich, Universitaetstrasse 22, CH-8092 Zurich, Switzerland
E-mails: {quang.le, roman.seidl, roland.scholz}@env.ethz.ch

Abstract: A key challenge of land-use modeling for supporting sustainable land management is to understand how environmental feedbacks emerged from land-use actions can reshape land-use decisions in the long-term. To investigate that issue, we use an agent-based land-use change model (LUDAS) developed by Le et al. [2008] based on a case study that was carried out in Hongha watershed (Vietnam). In LUDAS, goal-directed land-use decisions by household agents are explicitly modeled (i.e. agents calculate utilities for all land-use and location alternatives and likely select the alternative with highest utility). The model is run for two mechanisms of adaptation in land-use decisions to environmental changes that emerged from land-use actions. The first mechanism includes only a primary feedback loop learning, in which households adapt to the changing socio-ecological conditions by choosing the best land-use in the best location. The second mechanism builds on the first one but adds a secondary feedback loop learning, in which households can change their behavioural model in response to changing socio-ecological conditions. Patterns of land-use and interrelated community income changes driven from the two feedback mechanisms are compared to evaluate the added value of the inclusion of the secondary feedback loop learning. The results demonstrate that spatio-temporal signatures of the added feedback loops depend on domain type, time scale, and aggregation level of impact variables.

Keywords: human-environment interaction; agent-based modeling; land-use change; adaptive decision-making; feedback loop learning

1. INTRODUCTION

Modeling of land-use changes faces the complexity of coupled human-environmental systems that involves feedback loops between environmental dynamics and human decision-making processes [Scholz, in prep]. Land-use change emerges from the interactions among various components of the coupled human-landscape system, which then feeds back to the subsequent development of those interactions [Le, 2005; Le et al., 2008]. Changes in land allocation occur at the level of land plots, resulting from decisions made by individual actors situated in diverse socio-ecological settings. Accrual of short-term changes over time and aggregation of localized changes over space generate larger-scale emergent patterns of land-use change and economic performance. In turn, changes on the macro level such as climate or policy regimes affect the behavior of the individuals that produce changes at the micro-level [Lambin et al., 2003]. Understanding how such human-nature and cross-scale feedback mechanisms affect the dynamics of environmental and human systems at different spatial and temporal scales is still one of the major challenges in land-use change modeling [Verburg, 2006; Turner et al., 2007].

During the last decade, there has been a rapid growth of agent-based models (ABM) for simulating land-use changes [Matthews et al., 2007]. These models consist of a number of human agents which interact with each other and with their environment. This environment
can also be represented as autonomous land units, i.e. “land agents” (e.g. Le et al. [2008, 2010]). Human agents make decisions influenced by socio-ecological interactions and change their behavior as a result of these interactions, thus offering the opportunity to take into account the adaptation of human decision-making to land-use at different levels of landscape and human organizations.

However, most of ABM for land-use changes have incorporated different feedback mechanisms without carefully testing the new insights added by the inclusions of the feedback mechanisms. Adding unnecessary feedbacks may lead to a dramatic increase in the model’s complexity, in which the model would become too sophisticated (i.e. parsimony principle is not respected). In contrast, if an added feedback mechanism is proven to trigger new insights in the model’s outcomes, it will offer a new quality of the model and an increase in confidence.

In this paper, we test a simple methodology for modeling the adaptation of farmer’s decision-making in coping with long-term changes in socio-ecological conditions in a case study area in central Vietnam. We use an agent-based land-use change model (LUDAS) developed by Le et al. [2008] to examine the effects of a secondary feedback in land-use decision making with respect to different system performance indicators at different levels of aggregation.

2. METHODOLOGY

2.1 Concept of feedback loop learning in coupled human-environmental systems

We use the Human-Environmental System framework (hereafter referred as HES framework), developed by Scholz [in prep.], as the conceptual guide for the detailed investigation of feedback loops in land-use change. An important postulate of the HES framework states that there are different types of environmental feedback loops that represent perception, evaluations, and adaptation of human systems regarding environmental changes. Adaptation of human decision-making to environmental change is the actor’s learning with respect to the adjustment of their decision rules, depending on their static internal model of the human-environmental interactions (i.e. a fixed behavioral program).

In general, adaptive decision-making of human actors involves (i) a primary feedback loop and (ii) a secondary (higher-order) feedback loop learning. With the former, human agents perceive the environment status and react on it, and the human action transforms the environment, with a retroactive effect on the decision-making process of itself and of other agents in a short-term fashion. This first-order feedback learning does not alter the goal-related decision rules of agents. The later type of feedback loop learning is defined by human-driven cumulative changes in social/economic and environmental conditions at larger scales and in the longer term (possibly unintended), leading to the reframing of the actor’s behavioral program. Because of involving multiple interactions and feedbacks among variables and subsystems across different scales, the secondary feedback loop learning can be delayed [Scholz, in prep.], and often causes a legacy impact on the performance of the whole coupled human-environmental system [Liu et al., 2007].

2.2 The LUDAS model

We apply the Land Use Dynamics Simulator (LUDAS) [Le, 2005; Le et al., 2008] to test the effect of the inclusion of a secondary feedback loop learning on land-use and income patterns in the long-term at different aggregation levels. LUDAS is a multi-agent system model for spatio-temporal simulation of a coupled human-landscape system. The model falls into the class of “all agents”, where the human population and the landscape environment are all self-organized interactive agents. The human community is represented by household agents that integrate household, environmental and policy information into land-use decisions. Goal-directed land-use decisions by household agents are explicitly modeled, i.e. the agents calculate utilities for all land-use and location alternatives and
likely select the alternative with the highest utility. The decision model is specific for the livelihood typology of the household. The natural landscape was modeled as landscape agents, i.e., land units that host natural processes and change their nature in response to local conditions exerting influence on each unit of land and its immediate neighborhood. Relevant ecological models (e.g., biomass productivity and vegetation succession models) have been integrated into the structure of the landscape agents (see Figure 1). A detailed specification of the LUDAS model is shown in Le et al. [2008].

**Figure 1.** The conceptual framework of LUDAS model: multi-agent system for the coupled human-landscape system. Source: Le [2005] and Le et al. [2008].

### 2.3 Design of simulation experiments

**Mechanism I:** Household’s behaviour without any secondary feedback loop learning (baseline). In this design, human-environment interrelations are mainly characterised by tenure rules and the primary feedback loop learning. Tenure rules, possibly *de facto* and/or *de jure*, explicitly regulate the household’s access to and the usage of land resources. The primary feedback loop involves direct information/physical flows between household agents and their landscape environments. Household agents perceive the spatial status of the biophysical conditions around them and anticipate benefits that the agent can derive in arriving at decisions. When household agents use land they receive some actual benefits (e.g., agricultural products) that can lead to changes in certain attributes of their profile, and thus the interaction means now become physical. Through land-use activities the household agents modify the structure of spatial organisation in their environments, which then constraint or support his/her decisions in the next few years (via updating variables of the internal decision model).

**Mechanism II:** Household’s behaviour with a secondary feedback loop learning. This adapted decision mechanism builds on the first mechanism but adds a simple secondary feedback loop learning, in which households can change their behavioural model in response to changing socio-ecological conditions. Here, we assume that households can change their land-use behavior model by imitating the strategy of the livelihood group who is most similar to them. In the context of Vietnam uplands, the sustainable livelihood framework concept [Ashley and Carney, 1999] is a relevant basis to specify a land-use strategy, which is a core component of the total livelihood strategy. In LUDAS, there is an automatic classification algorithm, called *AgentCategorizer*, to annually update the livelihood typology of household agents by evaluating the temporal *cumulative changes* in variables of five main household capitals (namely natural, physical, social, human, and financial capitals). These variables – such as land-use structure of household land, agricultural income and so on – are subject of cumulative impacts caused by land-use actions of the considered households and his/her neighbor. *AgentCategorizer* is annually comparing and ranking dissimilarities between the considered household and all livelihood groups in the population, then assigning the household into the
most similar livelihood group. Thus, an imitative learning behavior is assumed. Details of
the algorithm are shown in the Appendix.

The Study Site. The LUDAS model was empirically calibrated for the Hongha watershed
in central Vietnam. The watershed lies about 70 kilometers west of Hue City, at 16°15’04”
– 16°20’17” N latitude, 107°15’01” – 107°23’06” E longitude, and covers an area of
about 90 km². The area is home of three ethnic groups (K’tu, Ta-Oi, and Kinh), fairly
representing the population in the region. By 2003, the population is of about 1200
inhabitants in 240 households. The average population growth is about 4.5%. Agricultural
production and collection of forest products (e.g. firewood, timber, rattan, trapped wild
animals) are the main livelihoods of most villagers.

As in many areas in the uplands along the Central Coast of Vietnam, three types of
agricultural land use most commonly found in the study area are upland crops, paddy rice,
and fruit-based agroforestry. The upland crop system, practiced traditionally by the ethnic
minorities (K’tu and Ta-Oi groups) is a type of shifting cultivation. Main upland crops are
local dry rice, casava and maize. The cultivation of upland crop is rain-fed with almost no
or very low input of chemical fertilizers. Paddy rice is widely practiced by most
households. Most of the paddy rice fields have two crops a year. Chemical fertilizers
(mostly NPK) and pesticides have been increasingly used in the paddy rice system since
1998, along with agricultural extension programs. Fruit-based agroforestry, widely
practiced in Hongha since the 1990s, include bananas, pineapples, jackfruits, lemon, longan
trees and black pepper, which are usually planted in association. NPK fertilizer is
sometimes applied when fruit crops are first planted.

Data Inputs, Impact Indicators and Uncertainty Quantification. Data inputs include
landscape and household attributes. Landscape data were obtained by remote sensing (land
use/cover), soil-landscape (terrain indices), accessibility (proximities to rivers_streams and
roads) analyses, as well as social mapping (holdings, village territory, protection zoning
class). Household data, covering socio-economic attributes (educational status, size, labour,
land endowment, income) and household’s access to policies or developmental programs,
were gathered through surveys using a structured questionnaire. Detailed input data and
calibrated parameters for the model were described and explained in Le [2005].

We assess the effect of the inclusion of the secondary feedback loop learning by estimating
the divergences and bounds of impact indicators between the two adaptive mechanisms
described above. The impact variables/indicators are as follows:

- Global coverage of a land-use/cover type (%) = (area of such a land-use/cover type /
total area of the landscape) × 100%
- Coverage of dense/rich forest within a buffer zone of the main road (%) = (area
covered by dense forest within the buffer zone / total area of the buffer zone) × 100%. By
calculating this localized coverage of dense forest for different extent of the buffer
zone, we expect to measure a spatial pattern of deforestation or forest degradation in
relation to road development.
- Total area of different farm types (ha)
- Average farm size (ha household⁻¹)
- Average yield of different farm types (ton rice ha⁻¹ year⁻¹)
- Average household gross income (1000 VND household⁻¹ year⁻¹) and its partial
components (from different income sources)
- Gini index of household income (varying between 0 and 1; 0 - perfect equality in
income distribution across population, 1 - completely inequality).

Because LUDAS is a stochastic model, it is not recommended to draw any conclusions
from the outputs of a single simulation run. The outputs represent only one realization of a
stochastic process. To quantify the uncertainty in the model outputs induced by the
uncertainties in its inputs, the method of independent replications [Goldsman, 1992;
Nguyen and de Kok, 2007] is used. In its application in this paper, we independently
replicated the simulation 12 times for each mechanism and computed the mean values of
the impact indicators and their confidence intervals at 95% reliability.

3. RESULTS AND DISCUSSIONS
3.1 Landscape responses

Time-series of simulated land-use/cover for the two tested adaptive mechanisms are shown in Fig. 2. The inclusion of the secondary feedback loop learning likely leads to a significant conversion of dense natural forest to open natural forest (degradation of dense/rich forest in the area) after 21-23 years (Figure 2A). Moreover, the simulation result reveals that such an impact likely happens mainly within a buffer zone of 2-4 km distance from the main road, suggesting a spatial scale-dependent impact on forest degradation.

![Figure 2](image)

Figure 2. Time series graphs of simulated land use/cover for feedback mechanisms I and II. (A): area coverage (%) of 5 main land cover types calculated for the whole study area, (B): area coverage (%) of dense/rich forest calculated within different buffer areas of the main road. Note: Vertical bar indicates the confidence interval of the mean values (95%).

The pattern shown in Figure 2 B is in relatively accordance with the observed reality. Apparently, the land trip with 1km from the main road has no more rich forest for logging. Whereas the further land from the road (distance to road > 4km) are covered by dense forests but are not easily accessible due to complex mountain terrain and labor constraint of households. With the added secondary feedback loop, it is likely that there is a temporally progressive shift of household behavior from the strategy of "poor" groups to those of the "better-off". A closer look at empirical data reveals that allocation of a bit more labor to logging and other off-farm activities (e.g. trading and technical work) is characteristics of the livelihood strategy of the “better-off”.

![Figure 3](image)

Figure 3. Time series graphs of simulated cropland area (A), and average farm size (B) for the feedback mechanisms I and II. Note: Vertical bar indicates the confidence interval of the mean values (95%).

The delayed impacts on forest cover in this case study clearly confirms a common awareness that time lags (legacy effects) follow profound non-linear dynamics when considering secondary feedback [Liu et al., 2007; Scholz, in prep.]. The adding of secondary feedback learning likely leads to a significant decrease in the area of upland crop, and the increase of paddy and agro-forestry areas, compared to the baseline (Mechanism I) (see Figure 3 A, B). The overall decline in the average farm size (i.e. total farmland /total household) (Figure 3 B) against the background of increasing cropping area
Figure 5C indicates that the population growth exceeds the expansion rate of farmland, thus likely being an underlying cause for land pressure in the area.

### 3.2 Income responses

The inclusion of secondary feedback loop learning has no significant impacts on overall household income pattern (Figure 4). However, this stable behavior is no surprise. The fact that the “poor” would like to imitate the strategy of the “better-off” does not necessarily include that all poor farmers will be successful regarding their income generation after changing their behavior. The mechanism of changing livelihood typology of the household might not count for all important conditions that support the realization of the new adapted livelihood strategy (i.e. imitative learning can be based on a “wrong” reflection of keys for successful adoption of new strategy). This can be the limitation of the current model algorithm that potentially misses important variables for household’s behavior program adjustment. However, the phenomenon can also reconcile with the genuinely incomplete evaluation of the situation in adoption of new strategies by poor farmers in the real world. For example, poor farmers may not be aware of some “hidden” constraints they face whereas the “similar” ones indeed do not have, or vice versa with some opportunities.

![Figure 4](image_url)

**Figure 4.** Time series graphs of simulated (A) household gross income and (B) income inequality (Gini index) for feedback mechanisms I and II. Note: Vertical bar indicates the confidence interval of the mean values (95%).

### 3.3 Global and local responses

The simulation results shown in Figure 5 A, B show that productivities of main farm types in Hongha commune are relatively non-responsive between the two tested adapted mechanisms. Moreover, agro-forestry farms in general increase over time from small initial values. This agrees with the fact that in 2002 (the initial year) fruit-based agroforestry was still new in Hongha commune. At the beginning of the farm establishment, pineapple and banana crops will be harvested for the first time two or three years after planting. Subsequently, the auto-vegetative propagation of bananas and pineapples increases the density of these crops and subsequently return higher yields. In later years, fruit-trees (e.g. lemon and jackfruit trees) and black peppers will probably increase overall annual yields, while some banana and pineapple crops will be replaced due to declining yields. Thus the annual yield will still increase steadily following a concave up pattern.

At the level of group aggregation, the patterns shown in Figure 5 B indicate that different household livelihood groups have different responses to the inclusion of the secondary feedback loop in terms of the temporal pattern of agro-forestry productivity. Taking into account a second feedback loop (Mechanism II), the productivity of agro-forestry farms under the management of “paddy-rice based and poor” and “upland crop-based and poor” farmers is considerably higher than that of the baseline (Mechanism I). With the “off-farm and better-off” farmers, the phenomenon is the vice versa. It is given that the empirical productivity function for agro-forestry farms used by the LUDAS model is (positively) responsive to only labour inputs and cropping time length [Le, 2005]. Because the setting of cropping time length is the same between the two tested mechanisms, the observed
differences would be only caused by the change in labour allocation of households for agro-forestry farms. Thus, it becomes clear that the adding of the secondary feedback learning triggers poor farmers to invest more time for agro-forestry farm, which can return the benefit the in long run. This is an insightful adaptation of poor farmers to meet their long-term food demand in a difficult context. That is: (i) productivity of farms on hill slope (i.e. upland crop) is already marginal to inputs and facing a high risk of lost yield (Le, 2005), and (ii) the potential access to suitable land for paddy in the narrow mountain valley will be very limited in the future decades.

**Figure 5.** Time series graphs of simulated crop productivity for different levels of farm’s aggregations. (A) whole population, (B) “paddy-based and poor” farmers, (C) “upland crop and poor” farmers, and (D) “off-farm and better-off” farmers. Note: Vertical bar indicates the confidence interval of the mean values (95%).

5. CONCLUSIONS

Understanding how environmental feedbacks emerged from land-use actions can reshape land-use decisions in the long-term is important for integrated system models of land-use change to support sustainable land management. To investigate that issue, we use an agent-based land-use change model (LUDAS) based on a case study that was carried out in Hongha watershed (Vietnam). The model was run for two mechanisms of adaptation in land-use decisions to environmental changes that emerged from land-use actions. The first mechanism includes only a primary feedback loop learning (Mechanism I), in which households adapt to the changing socio-ecological conditions by choosing the best land-use in the best location. The second mechanism builds on the first one but adds a secondary feedback loop learning (Mechanism II), in which households can change their behavioural model in response to changing socio-ecological conditions. Patterns of land-use and interrelated income changes driven from the two feedback mechanisms are compared to evaluate the added value of the inclusion of the secondary feedback loop learning.

The results demonstrate that spatio-temporal signatures of the added feedback loops depend on domain type, time scale, and aggregation level of impact indicators. The inclusion of very simple secondary feedback loop learning can cause long-term delayed effect in forest cover transition, significant change in agricultural area and farm size, and different responses of farming productivity managed by different farmer groups. The interpretation of the change patterns helps to improve our understanding of the co-adaptation between humans and the natural landscape that suggests new insights for alternative land management strategies. However, it also raises a new requirement of validating for the potential added values of the inclusion of more complex feedback loops.

REFERENCES
Appendix. AgentCategorizer algorithm

The algorithm is similar to the K-mean clustering procedure, except that the group centroids here were predefined outside the simulation model by descriptive statistics of household groups, and thus fixed during the simulation runs. The categorising process consists of the following steps:

- A given household $h$ measures dissimilarities in livelihood typology, based on grouping criteria, between himself and all defined household groups in the population:

$$D_{hg} = \sum_{c=1}^{C} w_c \left[ \frac{(H_{hc} - \overline{H}_{gc})^2}{\overline{H}_{hc} + \overline{H}_{gc}} \right]$$

where $D_{hg}$ is the Squared Chi-squared Distance from household $h$ to the centroids of the group $g$ ($g = 1, 2, \ldots, K$), $H_{hc}$ is the instant value of criterion $c$ ($c = 1, 2, \ldots, C$) of household $h$, $\overline{H}_{gc}$ is the mean value of criterion $c$ of the group $g$, $w_c$ is the weight coefficient of the criteria explaining the discrimination of household groups. The default value of $w_c$ is $1/C$.

- Household $h$ assigns himself into the most similar livelihood group ($g^*$):

$$g^* = \arg \min_{g} \{D_{h1}, D_{h2}, \ldots, D_{hk}\}$$

where $g^*$ is the most similar group to household $h$; $D_{h1}, D_{h2}, \ldots, D_{hk}$ are distances from household $h$ to groups 1, 2, ..., $K$, respectively.

- Once the livelihood group of a household has changed, he/she will ask to delete the old land-use decision model and to adopt the decision model of the new group. When adopting a new land-use decision model, there are not only changes in parameter values but possibly also in the behavior structure; some decision variables and production components are added or deleted.
Risk management in an uncertain environment -
A study from semi-arid grazing systems

O. Jakoby a, M. F. Quaas b, B. Müller a and K. Frank a

aHelmholtz Centre for Environmental Research - UFZ, Permoserstr. 15, 04318 Leipzig, Germany (oliver.jakoby@ufz.de)
bChristian-Albrechts-Universität, Kiel, Germany

Abstract: Since livestock grazing is the most important type of land-use in arid and semi-arid rangelands with uncertain and highly variable climatic conditions, well adapted grazing strategies are crucial for effective risk management. However, local environmental characteristics and individual needs and perception lead to a broad range of different management forms used in practice. In this paper we (i) analyse how uncertain climate conditions and the individual farmer’s characteristics like risk aversion, economic constraints and effort cost affect the choice of management, (ii) evaluate the viability of different management options, and in particular (iii) their robustness under climate variability. We use a generic stochastic simulation model that consists of a physiologically well-founded vegetation model combined with a rule-based management model. Thereby, we implement a feedback of grazing management and rangeland state. Changes in the rangeland state are driven by stochastic precipitation and livestock grazing. In turn all management actions taken by a farmer adapt to changes in the rangeland condition. Management is temporally resolved on a weekly scale and spatially on the local scale of an individual farm. With our approach we can characterise strategies that are robust over a wide range of climatic condition and those that are vulnerable to unexpected changes. Hence, we can identify appropriate strategies for an adaptation to climatic risk and therefore support sustainable land-use.

Keywords: semi-arid rangelands, generic simulation model, risk management, environmental uncertainty, climate change

1 INTRODUCTION

Arid and semi-arid rangelands are characterised by highly uncertain and variable climatic conditions. In this uncertain environment, the primary production is low and erratic. This pose the major challenge to extensive livestock farming that is the predominant form of utilisation in these areas and provides livelihood for a large part of the local population. Hence, the need for flexible and viable land-use strategies is especially high. Particularly, on commercial rangelands, where farm borders prevent the formerly widespread large scale moving of early pastoral systems, arises the need for new, well adapted grazing strategies for a risk management on a local scale.

What are the important characteristics of viable grazing strategies in semi-arid rangelands? The four principles of grazing management (i.e. timing, distribution, kind of livestock, and stocking rate) are widely accepted [Heitschmidt and Taylor, 1991]. Based on these principles emerged various kinds of (rotational) grazing systems that are used in practice [e.g. Heady and Child, 1994]. However, the strength of the impact of the different factors is heavily discussed. Whereas, the high importance of an appropriate stocking rate is widely agreed [Holechek, 1988], there are ongoing discussions on the importance of timing, distribution, and in particular on the suitability of rotational grazing systems in general [Briske et al., 2008; Barnes et al., 2008]. Furthermore, the specific strategy and the overall result of rangeland management must always serve the individual goal of the land manager. This goal varies, however, as a function of many social factors of which the economic is usually dominant [Connor, 1991]. Farmers can for instance aim to maximise their income, or might be risk averse and try to minimise the fluctuation in the income to
avoid catastrophically losses. Furthermore, resources like savings, time, or farm structure limit his possibilities. In summary, the variety of different objectives is as manifold as the diversity in management strategies itself. Although the objective and the management goal of a farmer are important variables in the choice of the appropriate strategy, moreover, it is limited by local factors such as climate conditions or available resources.

In this paper we explore several grazing strategies for rangeland management in an uncertain environment, characterised by an uncertain and variable precipitation and observe how they match with different farmers’ objectives. We also analyse the range of robust strategies in order to shifting objectives and climate change.

2 MATERIALS & METHODS

Rangeland management in semi-arid areas depends on multitude of different factors such as precipitation, vegetation growth, livestock grazing as well as farm structure, management strategy, and risk perception of a farmer. To cover these aspects we developed a generic ecological-economic simulation model. The model consists of two submodels: An equation-based ecological submodel describing the dynamics of the grazed vegetation and a spatially explicit, rule-based economic submodel describing the farmer’s management strategy. The ecological part of the system is represented by a physiology-based model of grass biomass production. The biomass is mainly affected by both unpredictably fluctuating precipitation, and the grass consumption of livestock. The way a farmer is managing his livestock has a direct feedback to the ecosystem providing his livelihood. This links the ecological to the economic part of the system. The economic submodel comprises both the management strategy of the farmer and the evaluation of the economic risk at the end of the time horizon. The management strategies considered in this study are characterised by: the rules for selling livestock, and the implemented farming system described by the number of fenced pastures (paddocks), and the rules of herd rotation.

2.1 Ecological Submodel

In the present study, each paddock is described as a grid of small homogeneous cells of rangeland. The ecological submodule describes the dynamics in a single cell that is the result of local interactions of the biotic (vegetation and livestock) and abiotic (precipitation) environment. As major interest of this study is in the effectiveness of different management strategies that are all working with weekly management decisions, the dynamics are described at a weekly time scale. The processes are affected by precipitation and livestock grazing, that are known as the main drivers of arid and semi-arid rangelands dynamics [Walker, 2002].

Precipitation. In accordance with the seasonal rainfall pattern in semi-arid rangelands that is characterised by a low and erratic annual rainfall of 250 - 500mm, we distinguish two different phases per year: (i) the rainy season with interannual fluctuating rainfall, and (ii) the dry season without rainfall. Each year the weekly amount of rain is drawn from a log-normal distribution $LN(\mu, \sigma_r)$ [Sandford, 1982]. The mean $\mu$ and variance $\sigma_r$ of the distribution differs regarding to the different climate scenarios (Tab.1). The rainy season lasts on average four month (18 weeks). It consists of a fixed period of 14 weeks and a flexible beginning and end phase. These are drawn from a uniform distribution of four weeks each.

Vegetation dynamics. The vegetation dynamics within a homogeneous rangeland cell are based on a set of three difference equations that are simulated at a weekly time scale. The plant biomass is subdivided into three components (Fig.1A): The (1) storage biomass (q.v. crown or reserve biomass) represents the vigour of the vegetation [Noy-Meir, 1982]. Further, the (2) green photosynthetically active and (3) brown, dead biomass form the grazeable parts of the plants. The storage biomass is supposed to contribute to the initial growth of green biomass by providing carbohydrates (after the first rainfall or a strong grazing event). If the amount of green biomass increases, the green parts of the plants maintain their growth independently from the storage due to photosynthesis. The growth rate of the green biomass is linearly increasing with the amount of precipitation. Towards the end of the growing season, carbohydrates formed by photosynthesis are transferred to the storage as reserves for regrowth in the next year. After a severe depletion of
green biomass by grazing the return of carbohydrates to storage biomass is diminished resulting in a poorer regrowth in the following year. Brown biomass emerges from dying of green biomass at the end of the rainy season. Both brown and green biomass can be consumed by livestock.

**Livestock grazing at the local scale.** Biomass consumption by large herbivores, i.e. domestic livestock like cattle or sheep has a major effect on the rangeland condition. Furthermore, the impact of different kinds of livestock varies as they differ in dietary preference, nutrient requirements and foraging abilities [Stuth, 1991]. However, in this study we neglect those species specific differences as the main focus lies on the effect of amount and distribution of livestock in general. We presume a general definition of livestock in terms of Large Stock Units (LSU).

Thereby, the biomass loss due to grazing within one cell depends on the amount and the palatability of grazeable biomass. Green biomass has a higher palatability than brown biomass [Stuth, 1991]. Therefore, the biomass discharge is weighted with palatability and quantity of the two types and multiplied by the amount of livestock on the cell.

### 2.2 Economic Submodule

This submodule describes the spatially explicit farm structure and the farmer’s livestock management. It uses the ecological model (see 2.1) for simulating vegetation dynamics in each rangeland cell. Moreover, it describes the livestock dynamics, the spatial grazing behaviour of livestock on one paddock, as well as the set of management strategies considered, and the rules for the economic risk assessment that provides the basis for the farmer’s strategy choice.

**Spatial farm structure.** The modelled rangeland is subdivided into 1024 cells on a rectangular grid (Fig. 1B). The biomass production in each cell is simulated on a weekly basis. The cells are grouped to paddocks that serve as grazing units. Each paddock has access to one water post. As several paddocks use the same water source it is located in one corner of each paddock.

**Spatial grazing behaviour of livestock.** The farmer chooses the paddock that is grazed at a certain time, but he cannot influence the grazing pattern within a grazed paddock. Here, the impact of large herbivores, in particular the degree of herbage defoliation, is largely influenced by the distance from a water source [Andrew, 1988]. Furthermore, livestock has a preference to graze fresh, green biomass [Stuth, 1991]. In the model we assume the relative grazing activity per cell G predicted by a negative exponential decay function [Pringle and Landsberg, 2004]:

\[
G = \exp(-D) \cdot P
\]

where D is the distance of a cell from a water post and P is the preference for grazing in a cell. P is depending on the amount of available biomass weighted with a preference for green biomass.

**Livestock dynamics, stocking and selling rules.** The number of livestock on a farm is determined by intrinsic population growth and the management of the farmer (Eq. (2)). Usually, most decisions regarding the stocking rates are made at the end of the growing season [Holechek, 1988].

---

**Figure 1:** (A) Conceptional description of the vegetation biomass submodel. (B) Spatial layout of the farm in the economic submodel.
Therefore, we assume one calving season per year that is timed towards the end of the growing season. At the same time, the farmer makes the decision to sell livestock to cover his expenses and regulate the stocking density.

In this study we apply an opportunistic stocking rule [Campbell et al., 2006]. This means the farmer varies the maximal number of livestock according to the temporally variable forage supply on the grazed paddocks in the next year, and the stocking rate that correspond to the percentage of available forage that should be used. We assume that the farmer only raises own animals and never purchases livestock. To cover at least some of his expenses he definitely sells 10% of his livestock at the end of each rainy season. Additionally, he sells the surplus of animals that exceed his calculated maximal herd size. In the following, we use the terms sold livestock and income synonymous, as the farmers derives his income only from selling livestock and the livestock prices are defined constant over time.

Moreover, the herd size is determined by its intrinsic population growth. It is based on a linear relationship between growth rate and amount of forage intake during the last year (Eq. (3)):

\[ H_{t+1} = H_t + g_t \cdot H_t - S_t \] (2)

\[ g_t = g_{max} \cdot \left[ 2 \left( \frac{1}{52} \cdot \sum_{k=0}^{51} F_k \right) - 1 \right] \] (3)

\( H_t \) is the number of livestock in the current year, and \( S_t \) is the number of livestock sold. The population growth rate \( g_t \) depends on the maximal population growth rate \( g_{max} \) and the cumulated amount of grazed forage \( F_k \) per week \( k \).

**Rotation rules.** Most livestock ranches in semi-arid areas apply one of various kinds of rotational grazing schemes. That means the farmland is subdivided into multiple paddocks, and grazed in rotational sequence allowing recovery of rested areas. Here, we analyse a non-adaptive rotation management. The herd is shifted from one paddock to the next after a fixed period in time (ST). The selection of the next paddock follows a fixed sequence.

**Evaluation of the management strategies.** In order to evaluate the different management strategies in terms of their appropriateness and to provide insight into the role of the farmer’s management objectives, two different management aims were assumed and compared. In the first case, we consider a risk-averse, utility maximizing farmer whose utility positively depends on the mean number \( S \) of livestock sold and negatively on the coefficient of variation \( CV(S_t) \) (Obj1). In the second case, we assume that the farmer follows a safety-first approach and intends to avoid the economic risk that the yearly number of sold livestock \( S_t \) falls too often below the minimum threshold \( S_{min} \) required for securing his livelihood (Obj2). “Too often” means more frequently than a critical number of years \( Y_{max} \) that can be tolerated at maximum within 10 years. Note that this number \( Y_{max} \) can also be interpreted as indicator of risk aversion: low/large \( Y_{max} \)-values indicate high/low risk aversion.

### 2.3 Scenarios for model simulations

Many different management strategies exist (and are applied) for livestock farming in dry rangelands [e.g. Heady and Child, 1994]. External factors, such as climate conditions, determine the potential management strategies. However, the adequacy of a strategy depends mainly on the preferences and the objectives of a land manager. Therefore, we test different sets of management rules under several climate scenarios and observe how they match with different objectives.

**Management scenarios.** Heitschmidt and Taylor [1991] point stocking rate, number of paddocks and rate of rotation as the important variables in rotational grazing systems. Here, we analyse the impact of paddock number and rotation rate on the success of rangeland management and focus on a medium stocking rate of 50%. Thereto, we define different rotational grazing strategies by combining several paddock numbers (4, 8, 16, 32) with fix standing times (1, 2, 4, 8, 17, 52 weeks). Additionally, we simulate continuous grazing on one paddock.

**Climate scenarios.** The rainfall pattern in drylands differs widely between geographical locations. Moreover, a decrease in precipitation is expected for many parts of the arid and semi-arid
Table 1: Table of climate scenarios. The values were transformed that a medium rainfall is expected to be 1.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Mean rain $\mu$</th>
<th>Variance of rain $\sigma^2_r$</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1-V25</td>
<td>1</td>
<td>0.25</td>
<td>normal rainfall with normal variance</td>
</tr>
<tr>
<td>R1-V40</td>
<td>1</td>
<td>0.4</td>
<td>normal rainfall with increased variance</td>
</tr>
<tr>
<td>R08-V25</td>
<td>0.8</td>
<td>0.25</td>
<td>decreased rainfall with normal variance</td>
</tr>
<tr>
<td>R08-V40</td>
<td>0.8</td>
<td>0.4</td>
<td>decreased rainfall with increased variance</td>
</tr>
</tbody>
</table>

regions over the next century [IPCC, 2008]. Generally, an increased frequency of extreme weather events is predicted. Therefore, it is crucial to know which management strategies perform well under which range of climatic conditions. We evaluate, if the success of different strategies change for different regions or rather under the influences of a locally changing precipitation regimes. Hence, we analysed rainfall scenarios differing in interannual variance $\sigma_r$ and mean precipitation $\mu$ (Tab.1).

**Simulation experiments.** We simulated all management strategies for a period of 100 years. For each climate scenario the simulations were repeated 100 times with differing stochastic precipitation. For the analysis we calculated the mean number of livestock sold ($\bar{S}$) and their coefficient of variation (CV) for each run and averaged the two quantities over of all 100 runs.

3 RESULTS

3.1 Management aiming at maximum stable income

We start with the assumption that the farmer aims at maximizing the mean income from livestock sales by minimizing its variation, i.e. follows an optimisation approach. Therefore, we evaluate the different management strategies from the point of view of the resulting mean $\bar{S}$ and the coefficient of variation $CV(S_t)$ of the livestock sold.

**Influence of paddock number and standing time.** The combination of paddock number (PN) and standing time (ST) shows a strong impact on both $\bar{S}$ and CV (Fig.2). An increasing standing time leads to a typical shape of the resulting curve in the $\bar{S}$-CV-space. Evidently, the CV decreases, except for very long grazing periods on small paddocks, while the mean income $\bar{S}$ shows a unimodal response. This indicates the existence of an optimum ST that balances the trade-off between increasing mean and decreasing variation. We also see that the entire curve and the optimum ST shift with the PN. A simultaneous increase in paddock number and standing time cause both an increase in the mean income $\bar{S}$ and a decrease in CV. Therefore, a higher subdivision

![Figure 2: Relation between mean ($\bar{S}$) and coefficient of variation (CV) of sold livestock for different management strategies. The numbers on the curves indicate the standing time per paddock in weeks. (A) The relation is plotted for different paddock numbers indicated by different colours. The gray dot shows the outcome for a continuous grazing strategy.](image-url)
of the farm combined with an appropriate (and higher) rotation speed improves the success of a farm enterprise in terms of higher income and lower CV. However, in return the farmer’s effort in management increases with more paddocks and faster rotation. Therefore, he is subject to a trade-off between mean income, the degree of fluctuation in his mean income (related to his risk aversion), and the effort cost. Additionally, a continuous grazing strategy performs very poorly (Fig.2). It provides a rather low mean income at an extremely high degree of uncertainty. This means that most rotational grazing strategies perform much better for risk management.

3.2 Management for sustaining a minimum income

We now change the perspective and repeat the analysis for a different specification of the farmer’s objective ($Obj_2$). This means that we consider a farmer who intends to secure a certain minimum income from livestock $S_{min}$ but tolerates a maximum number of years $Y_{max}$ where this minimum is fallen below.

For this analysis, we assess all management strategies regarding their suitability to meet the management objective (Fig.3). This is repeated for different values of $S_{min}$ and $Y_{max}$ to assess the influence of the farmer’s minimum requirements and his capacity to tolerate failing the minimum. Figure 3A reveals that both the income requirements and the tolerance of failure of the farmer markedly limit the range of suitable strategies. If these requirements are low or the tolerance is high, almost every strategy is found to be suitable. In face of increasing requirements or decreasing tolerance, however, the range of opportunities is shrinking. Long ST or small PN become increasingly critical. The suitable strategies are characterised by an appropriate ratio of paddock number and standing time (i.e. the higher the PN the shorter the ST). Altogether, there is only one type of strategy that is suitable for a large range of settings and so most robust: strategies with short standing time and a large number of paddocks. These strategies ensure that the rangeland can regenerate after short-term grazing.

3.3 Influence of climate changes

Additionally, we analysed the influence of the scenarios of climate change listed in Table 1. As far as management objective $Obj_2$ is concerned, we found that alterations in the distribution of precipitation $LN(\tau, \sigma_r)$ cause a shift in the functional response of the strategies in terms of mean

![Figure 3: Possibility space of rotation strategies for farmers with different management aims. Strategies are said to be suitable, if 95% of the simulation runs match the objective. (A) Changes in the possibility space for different minimum incomes ($S_{min}$) and maximal numbers of years that are manageable to fall below ($Y_{max}$). The figure shows an average climate scenario (R1-V25). (B) Suitable management strategies under different climate scenarios (see Tab.1) and a defined minimum income of $S_{min} = 10$.](image-url)
and CV in the income gained from livestock sales, but did not change the shape of this relationship (results not shown). All changes to lower \( \tau \) and higher \( \sigma_r \) result in an increase in the CV of income. Further, in scenarios with lower \( \tau \) the mean income \( \overline{S} \) is diminished as a whole.

Figure 3B shows the performance of the management strategies under the same scenarios of climate change from the perspective of management objective \( Obj_2 \). Evidently, the climatic conditions influence the range of suitable management strategies. A decrease of 20\% in the mean precipitation \( \tau \) causes a dramatic shrinkage of this range. For example, it is not even possible to achieve a fairly low minimum income \( S_{\text{min}} \) on a regular basis, regardless of the standing time or the number of paddocks. Also, an increase in the variability of precipitation results in a smaller set of possible options. Again, the more flexible strategies are characterised by an accurate combination of PN and ST, with dominance of a management on many paddocks with a fast rotation. This shows that the climate conditions limit the chance to manage the income risk by appropriately designed adaptive management strategies.

4 DISCUSSION AND CONCLUSIONS

We used a generic ecological-economic simulation model to incorporate the feedback between the vegetation dynamics in semi-arid rangeland and the management activities taken by a livestock farmer. We evaluated different management strategies from the perspective of different management objectives (maximize stable income \( (Obj_1) \), ensure minimum income \( (Obj_2) \)), and we assessed the robustness of the results against different scenarios of climate change.

We have shown that a higher subdivision of the farm combined with an appropriate (and lower) standing time improve the success of a farm enterprise in terms of higher mean income \( \overline{S} \), lower CV, and the coverage of a certain minimum income requirement. This is consistent with the statement of Hart et al. [1993] that an intensive rotational grazing system has to be coupled with pasture subdivision to produce a grater stocking rate and more uniform grazing. The more homogeneous distribution of grazing in small paddocks [Barnes et al., 2008] is one of the major reasons for the success of a higher number of paddocks in our model. Altogether, our results supports to some extent the findings of Savory and Parsons [1980].

In connection with the management objective \( Obj_2 \) considered, a whole range of suitable management strategies can be determined. Here, the farmer has more opportunities than a single optimum strategy. Nevertheless, the two management objectives \( (Obj_1 \text{ and } Obj_2) \) are interlinked with each other. We have shown that, whenever the minimum income requirements of the farmer are increasing or his economic capacity to tolerate is decreasing, the range of opportunities is shrinking. The most robust strategy that, to a certain extent, stands increasing economic demands or decreasing precipitation is the strategy with short standing time and a large number of paddocks - the strategy that has also been found as promising under management objective \( Obj_1 \).

The range of strategies suitable for ensuring minimum income markedly depends on the farmer’s economic requirements but also on his resources. For example, a full-time farmer with low abilities to stand bad years depends on a continuous and relative high income from livestock farming. He has only a few management options and would be best with a high degree of subdivision and a fast rotation. On the other hand the farm business is just extra income. He can set a lower required income and accept a higher variability as he diversified the risk and can buffer low income years. Here, the economic possibility space would be large, and the selected strategy depends on other resources. Assuming he has limited time due to his main business he cannot put much effort in an advanced management system and would choose a small paddock number, a long standing time, and a low stocking rate.

Further we could show, that a decreasing mean and an increasing variability of precipitation leads to a shrinkage of the range of suitable management strategies that allow an adaptation. To some degree a farmer can follow such changes by adjusting his management system. However, as functional aspects in rangeland management remain constant regardless of social and economic factors [Connor, 1991], after strong changes he has to scale down his requirements, or take other options, such as an increase of the farm size or a diversification of income.

Our study implies feedback loops on different levels of the socio-ecological system. The major feedback in semi-arid rangeland systems couples the availability of natural resources directly to livestock dynamics and management actions. An increasing number of livestock leads to a decrease in green biomass, which in turn causes a decrease in the herd size either by destocking or
by death of livestock. This interaction is connected with two more feedback mechanisms: On the level of the vegetation a positive feedback exists between the vigour of the vegetation and the amount of green biomass. Furthermore, the amount of livestock positively effects the population growth that in turn results in a larger herd size.

Regarding to the huge amount of possible management options, we had to restrict our selection of strategies in this study. So far, we assumed an opportunistic stocking strategy with a fixed selling rule. This management takes into account the negative feedback loop between amount of green biomass and herd size. However, it did not explicitly comprise the positive feedback in livestock production. In this context, an interesting question for future research would be the impact of different stocking and selling rules on the achievability of the management goals. Also, we used the simplest rotation strategy and neglect the feedback between green and storage biomass. To assess the potential of more adaptive rotation strategies that adapt to the recent state of the environment is therefore another task for future research. Overall, this would allow comparing the effectiveness of different approaches of risk management and adaptation. It would strengthen the knowledge base for the design of the institutional framework for sustainable grazing in semi-arid regions.

Conclusively, we can detect cases of management strategies at which the farmer is definitely better of, than in others. Furthermore, we can distinguish areas in the strategy space which are not suitable at all. Altogether the optimal management strategy will vary widely with management objectives and environmental constraints, albeit in our study a relative high paddock number with an adequately high rotation speed appears to be the most robust management strategy. However, it is not suitable to advocate dogmatic a certain kind of management system, as there are plenty different individual needs and goals among livestock farmers.

REFERENCES

Towards the Integration of Economic and Land Use Change Models

Hedwig van Delden* and Garry McDonaldb

*Research Institute for Knowledge Systems, Witmakersstraat 10, 6211 JB Maastricht, The Netherlands (hvelden@riks.nl)

bMarket Economics Ltd., P.O. Box 331 297, Takapuna, Auckland, New Zealand (garry@marketeconomics.co.nz)

Abstract: Although in the past decades several attempts have been made to integrate economic models with land use change (LUC) models, none of them have been fully satisfactory. In this paper we analyse four different integrated models for policy support that include economic and LUC models. We describe several functional forms used to integrate LUC and economic models, highlighting the various strengths and weaknesses of each form, and in turn, suggesting possible pathways for improved integration. Analysing the concepts and underlying assumptions of both types of models show their vast difference. When integrating these models, underlying assumptions and limitations of the existing individual models are passed on to the integrated model. A proper integration therefore requires a thorough understanding of the underlying theories of both types of models and a solution at this theoretical level. We argue that concepts of evolutionary economics and -spatially explicit- agent based modelling, where creation of ideas and learning are embodied into fully integrated LUC and economic models, provide some key mechanisms for bridging the described gap. These approaches are however very data demanding and have -at least at present- major limitations in performing a proper calibration and validation.

Keywords: Ecological economic modelling; Land use change modelling; Model integration, Cellular Automata, Agent Based Modelling, Evolutionary approaches.

1. INTRODUCTION

The first Integrated Spatial Decision Support Systems (ISDSS) capturing both land use and socio-economic inter-linkages appeared in the mid 1990s (Engelen et al., 1995). Since this time, the integration of land use and socio-economic models has involved integrating existing high resolution dynamic LUC models (based on techniques such as cellular automata) with formal sector based macro-economic models that do not have a high level of spatial detail. The notion that economic and land use change processes interact, is common knowledge (Castells, 1977; Harvey, 1985; Lefebvre, 1991). How to simulate this interaction in (integrated) models is not that straightforward. Both disciplines have co-existed for decennia and each has developed its own concepts and (modelling) paradigms. When integrating models from these different disciplines, underlying assumptions and limitations of the existing individual models are passed on to the integrated model. A proper integration therefore requires a thorough understanding of the underlying theories of both types of models.

In this paper we will analyse several ISDSS that integrate economic and spatially explicit LUC models and that aim to support planning and policy-making processes: LUMOCAP (Van Delden et al., accepted), WISE (Rutledge et al., 2008), Eururalis (Verburg et al., 2008) and MedAction (Van Delden et al., 2007). In the first three systems macro-economic models are coupled to LUC models, in the latter economics are incorporated into the LUC model. The ISDSS are selected in such a way that all of them include different types of
economic models. The land use models incorporated are all dynamic simulation models. To provide the reader with a good understanding of the different concepts used by the various models, we will give a brief overview of the theoretical approaches in Section 2. In Section 3 we describe the four ISDSS mentioned above and detail the functional forms of the models used to integrate LUC and economic models as well as the various strengths and weaknesses of each form. Finally we will discuss the current challenges, draw some general conclusions and suggest possible pathways for improved integration.

2. METHODS USED IN ECONOMIC AND LAND USE CHANGE MODELLING

The integration of land use and economic models requires bringing together of two related, but often distinctly different, world views. The methods, concepts and terminology used, for example, by economists and land use modellers can be very different. In section 2.1, we provide a very brief overview of the models (with key concepts highlighted in italics) that regional economists have applied in relation to land use modelling. In section 2.2 we explain common concepts used in LUC modelling. Our attempt is not aimed at being comprehensive, but rather at providing background on key tools used in economic and land use modelling and, perhaps, more importantly, initiating discourse and reflection on the modelling paradigms we are inherently trying to integrate.

2.1 Formal sector based macro-economic models

Econometrics is concerned with developing and applying statistical methods to study economic relationships or principles (Frisch, 1933). It focuses on the statistical properties of key determinants underpinning economic behaviour, and in the case of land use modelling, often extrapolation patterns inherent in time series data to provide future forecasts for key socio-economic variables such as GDP, employment and exports. Because firms within an economy interact with each other, the observational data used in econometrics tends to capture complex equilibrium conditions. This has led econometricians to develop methods for estimating simultaneous equations models (e.g. capturing supply and demand characteristics of an economy). A key problem with the use of econometrics in modelling scenarios is that it often requires that we extrapolate historical trends to determine future states. Such patterns are, however, often peculiar to a series of events of policy interventions or market dynamics, which may not persist through time.

Input-output (IO) models provide a snapshot of the structural interdependencies between industries (e.g. agriculture, manufacturing, and services), primary inputs (e.g. wages and salaries, profit, imports, depreciation) and final demands (household consumption, government consumption and exports) for a given financial year within an economy (Miller and Blair, 2009). Input-output tables are presented in matrix format with row entries representing sales and column entries purchases. Using simple matrix algebra IO tables may be used as an analytical tool to study the short-to-medium implications of comparative static changes in demand (i.e. consumption, exports, and environmental emissions), or supply (e.g. wages and salaries, imports, and environment factors such land, energy, water etc), on an economy. Importantly, IO models capture not only direct, but also indirect (through supply chain purchases) and induced (through consumer spending) impacts associated with economic change. The key drawbacks of IO models is that they are typically linear, ignore important investment (and often employment requirement) dynamic feedbacks, and take no account of price change.

Computable General Equilibrium (CGE) models employ detailed bottom-up micro-economic theories to establish a picture of a given regions macro-economy (Johansson, 1960; Rutherford, 2005). CGE models explicitly model the interconnectedness of agents (households, businesses, investors, government) within any economy, including allowing agents to modify their behaviour in light of economic change. As with IO, a key benefit of this approach is that it allows modellers to comprehensively measure direct, indirect and
induced economic consequences results from these changes. CGE, however, extends IO to include feedbacks associated with investment, labour dynamics and price change. CGE models include non-linear functions (with elasticities determined via econometric analyses) covering both production and consumption. CGE models tend to be recursively dynamic (i.e. there are feedbacks between components within the model), but often neglect transitional dynamics (i.e. how economic agents change through time). Instead CGE models report only on the resulting long run equilibrium or steady state. They are “solved” for a particular economic shock using numerical optimisation. Most CGE models, while mathematically pleasing, are concerned too academic to be useful in applied studies. Key reasons include (1) the data required to determine the elasticities (and also other data inputs) is often unavailable or extremely costly to obtain, and (2) complexities exist with integrating optimisation and LUC models.

2.2 High resolution dynamic land use change models

LUC models that have the objective to simulate spatially explicit dynamics are often based on self-organization or complexity theory (White and Engelen, 1993). They generate an organized but unpredictable behaviour of the land use system. This behaviour is represented by a large set of simple equations or rules that together create a complex behaviour that includes non-linear dynamics and emergent properties. They are simulation models that start with a land use map of the initial year and use a set of drivers (behavioural, institutional and physical) to calculate future developments. These models are exploratory and show what could happen, rather than what should happen. There is no ideal future, nor is there an assumption that the world reaches equilibrium in any point in time.

Cellular automata (CA) are a common means to implement the self-organisation approach. All LUC models in this paper incorporate a CA or use techniques related to CA and hence have the following characteristics: they are grid based applications in which each cell is in a possible state, i.e. occupied by a specific land use. Time progresses in discrete time steps and at each time step all cells update state (land use) simultaneously, based on the state of the previously time step, the neighbourhood of the cell and the transition rules that state under which conditions cell states change.

Most LUC models currently in practice, including the ones described in this paper, make use of a special form of CA, called constrained CA. In this type of model, area demands for each land use are determined exogenously, while these demands are allocated by the model. Furthermore, in most applications land use transitions are not purely based on the cell states in the neighbourhood, but on local characteristics as well, such as accessibility to infrastructure or the inherent suitability of the location for a specific land use. With these additional behavioural components the systems have been named ‘relaxed’ CA (Couclelis, 1985).

3. ANALYSING DIFFERENT APPROACHES TO MODEL INTEGRATION

The four selected integrated models are described below. A brief overview of each model is provided and the link between the economic component and the land use component is detailed. At the end of each section the advantages and disadvantages of the approach are discussed.

3.1 LUMOCAP – Dynamic Land Use change MOdelling for CAP impact assessment on the rural landscape

The LUMOCAP PSS is developed to assess the impact of the Common Agricultural Policy (CAP) on the land use and landscapes of the 27 countries of the European Union (Van Delden et al., accepted). The system incorporates models for agricultural economics, socio-economic regional interaction, land use allocation, crop choice and physical suitability and
uses scenarios for climate change, socio-economic developments and policy alternatives as external drivers. It encompasses four spatial levels: EU-27, national, regional and local (1 km² grid for the entire EU and 4 ha grid for specific case studies). The temporal resolution of all models is a year and the time span of the system is from 2000 to 2030.

LUMOCAP integrates an econometric model, a dynamic multi-product supply model of EU agriculture, with local models simulating land use change and physical suitability. On a yearly basis the econometric model drives the expansion or decline of the agricultural area which is subsequently allocated on the local grid cells using a cellular automaton based LUC model. Based on the competition for space with other land uses, the agricultural land uses will be able to occupy more or less suitable locations. Moreover, suitability of locations is impacted by climate change and in particular changes in temperature and rainfall. After allocating the crop types at the local grids, yield is calculated for each location. Aggregate yield values per region are fed back to the agricultural economic model that uses these for the calculations of the production and area totals per crop type for the next year.

Integration between economic and LUC models is facilitated because both types of models use discrete time steps and operate on a yearly temporal resolution. This makes the feedback of information between components straightforward. Links to other economic sectors at macro-level are only included to the extent they are reflected in the historic data. Although the integration seems to work well, it should be noted that the econometric model heavily relies on historic time series and can therefore not deal very well with long-term scenarios, which is a main application domain of the incorporated LUC model.

3.2 WISE – Waikato Integrated Scenario Explorer

The Waikato Integrated Scenario Explorer (WISE) aims to support long-term integrated policy development and planning in the Waikato region in New Zealand by taking into account cultural, social, environmental and economic well-being (Rutledge et al., 2008; Huser et al., 2009). The system incorporates models at the level of the entire Waikato region (ecological economics), as well as district (demographics), sub-catchment (water quality) and local level (hydrology, land use and terrestrial biodiversity). Drivers for the integrated model are climate change scenarios, socio-economic drivers (e.g. fertility, mortality and migration rates, exports and consumption patterns), and policy alternatives (zoning regulations, impact restrictions and construction of infrastructure). The temporal resolution of all models is a year and the time span of the system is from 2006 to 2050.

WISE incorporates a sector driven economic model based on IO analysis. This model is an important driver for land use change in providing land use demand for a range of economic activities such as industry, commercial activities, dairying, cropping, and beef & sheep farming. The LUC model subsequently tries to allocate these demands at the local level. Only suitable and available locations are taken into account during the allocation. This avoids e.g. allocation of dairying land and industrial locations on steep slopes or urban development in conservation areas. When not all demands can be met, the competition for space between different actors is simulated by the allocation algorithm, and the final allocation is fed back to the economic model. The supply side of the economy is affected by this information and hence economic growth is less than would be expected by a purely demand-driven approach. Because the IO approach captures the interdependencies between industries, the availability of suitable land can restrict growth for different economic sectors.

The key strength of the WISE approach is the integration of available resources in the supply side of the economic model, simulating how physical and institutional restrictions on land resources are limiting the land supply and hence economic growth. Furthermore, this approach has the ability to capture the interdependencies (i.e. supply chain linkages) between industries, and in turn, changes in land use requirements across all industries.
A drawback of the IO model is that this is a linear model and interdependence between industries is assumed to be constant with no technological change. This makes the model less suitable for more creative and long-term scenarios.

When implementing the interactions between the land use and economic components a main difficulty was experienced. For the macro-economic model to operate correctly, the demand and supply side should be in equilibrium for a single year. Because the demand side impacts on the LUC model and the supply side is affected by the LUC model, equilibrium could only be obtained through an iterative procedure between the LUC and the economic component, which would have to be carried out during each time step. Such a procedure would however not match the simulation approach of the LUC model in which action and reaction are modelled over time. After reviewing several alternatives and investigating their results, it was decided to divide the demand and supply calculations over two time steps. This solution is conceptually not ideal (nor is the other solution of iterating between the economic model and the LUC model in the same time step), but was favoured because of its shorter execution time (which was important for the use value of the ISDSS) and its fit with the overall dynamic nature of the integrated model.

3.3 Eururalis

Eururalis provides a tool for a structured discussion between policy makers, stakeholders and scientists about the future of the rural areas of Europe. Results are calculated using a modelling chain including three existing models: an economic model, an integrated assessment model, simulating the impacts of CO2 concentrations and climate change on the agricultural sector and natural biomes, and a LUC model. ‘By combining the economic and integrated assessment model the ecological consequences of changes in agricultural consumption, production and trade can be visualized’ (Verburg et al, 2008). The tool provides information at local (1 km² grid), regional, national and cluster region level for 2000, 2010, 2020 and 2030.

Economic developments are calculated using LEITAP, a CGE model at world level based on the standard GTAP model (https://www.gtap.agecon.purdue.edu/models/current.asp). Changes in LEITAP compared to GTAP are documented in Van Meijl et al. (2006). In an iterative procedure LEITAP and the integrated assessment model IMAGE calculate the agricultural land use changes at the level of individual countries inside Europe and for larger regions outside Europe (Van Meijl et al., 2006). At the same time, these models also calculate changes in other sectors of the economy which are indirectly related to land use. This information is used in a series of simple models. For the industry and services sectors the changes in sector size are translated into land requirements for these sectors. For the natural and residential land use types the claims for land area are based on the exogenous drivers. A spatially explicit LUC model (CLUE-s) finally allocates land use change based on competition between different land uses and the use of spatial allocation rules while including various environmental and spatial policies (Verburg et al., 2008).

Strengths of this approach are the inclusiveness of the economic sectors allowing for an interaction between those sectors. Although there is interaction between the models at global/national level, a waterfall approach is used for the link between the global/national models and the local model: first land use demands are calculated and subsequently these demands are allocated on the grid; there is no feedback from the local level land use to the economic model omitting the feedback from the available land resources to the economy.

3.4 MedAction – Mediterranean Action

The MedAction Policy Support System (PSS) supports regional development and desertification, focusing on sustainable farming, water resources and land degradation in arid and semi-arid regions (Van Delden et al., 2007). MedAction consists of several sub-models that are integrated in a single model that simulates development in the region up to 30 years in the future, using 2000 as initial year. Individual components incorporated in MedAction include: a weather generator and models for hydrology, plant growth,
salinisation, erosion and sedimentation, transitions in natural vegetation groups, crop choice land use and land management. External drivers include climate change and demand for land from international and national economic and demographic growth. The system includes a wide range of policy options, amongst others subsidies, taxes, zoning regulations, reforestation, decisions on water use and extraction, water pricing and construction of infrastructure. Impacts are assessed by means of a range of policy-relevant indicators that are aggregated into three headline indicators: water shortage, environmentally-sensitive areas and long-term agricultural profits.

MedAction makes use of external macro-economic drivers, which are converted into land use demands for economic functions, such as industry, commerce, tourism and agriculture. The LUC model specifies first where agricultural area will be located based on the competition for space with the other land uses. Next, a crop choice component determines which crops will be grown in the agricultural areas. The economic component of MedAction is incorporated in this crop choice model. For each location a utility based function calculates on a yearly basis what crop will be grown. Elements included is this equation are financial (market prices, costs, subsidies and taxes), physical (yield) and social (willingness to change). Financial elements are included as external drivers. The yield and expected yield of substitute crops is calculated by the bio-physical components. When farmers are not able to make a profit, agricultural land becomes abandoned and converts to natural vegetation, thus impacting on the other land use classes.

The main advantage of this approach is that economic drivers are integrated in the LUC models, making the utility approach an integral part of the land use choice. What is lacking in this approach is the link to macro-economic behaviour. Demand is reflected in the market prices, which are exogenous to the model. The regional supply does not have any impact on the agricultural economy of the country or the world, an assumption that is hard to back up for most agricultural regions in the current age of globalization; the interaction of agricultural with other economic sectors is only included in the competition for space at local level, while in reality this interaction has a much wider impact.

4. COMPARING DIFFERENT APPROACHES

All four above-mentioned approaches have integrated land use change and economics. The linkage between macro-economic models and LUC models is typically achieved through a static mapping of sector definition and land use class and is often one-way. LUMOCAP and WISE are exceptions to this by also creating a feedback from the LUC model to the macro-economic model. In MedAction, economic principles are incorporated in the local LUC model.

Although the four ISDSS described in this paper have to some extent succeeded in linking both types of processes, none of the approaches has found a procedure that is fully satisfactory. The problems experienced in this integration do however not stand on itself and are can be found in various other applications that link economic and LUC models (see e.g. Sieber et al., 2008). A key failure of most approaches is that they focus on integration of existing predetermined models as derived from their parent disciplines. The problems are hence closely related to the type of economic model included.

The use of econometric or regression based models, such as LUMOCAP, tends to be favoured in the integration of LUC and socio-economic models because of its –often yearly– temporal resolution which allows for an easy integration with dynamic LUC models, also often operating at the same yearly time step. A link between the economic model in LUMOCAP and any resource model is very difficult, because the lack of resources does not impact on the demand. This is however not a general limitation of econometric models, but a result of the specific choices made in the development of LUMOCAP. The selected economic model only represents the supply side and is therefore not sensitive to limited resources, such as land, water or human resources.
Furthermore, econometric models possess only limited value in assessing emerging behaviour as they attempt to predict future states of key driving economic variables through known historical patterns or trends. More importantly, the underpinning causal mechanisms, as characterized by feedback loops, time lags and non-linearities are typically overlooked or omitted, resulting in an application domain of short term spatiotemporal dynamics. It is worth noting that also comparative static implementations are often driven by econometric or regression analyses and, hence, are susceptible to the same underlying assumptions.

The main benefit of using IO and CGE models is that they capture not only direct, but also indirect (through supply chain purchases) and induced (through consumer spending) impacts associated with economic change. The dynamic link between LUC and economics in WISE thus shows the impact of limited (land) resources not only on the sector for which the demand cannot be fulfilled, but also on all other sectors. A key benefit in using CGE over IO models is that recursive dynamics within an economy (e.g. feedbacks through price changes, labour markets, and household spending) can be captured in the land use dynamics. Nevertheless, several problems exist with this approach: (a) CGE models rely on optimization algorithms to determine long run equilibriums, (b) derivation of the equilibrium typically ignore transitional dynamics, and (c) while the production functions used in CGE are dynamic, economic interdependence occurs through the a nested linear Leontief (IO) matrix – with its own underlying linear assumptions as identified above.

The equilibrium approach of the above-mentioned economic models poses conceptual conflicts with the simulation approach of the dynamic LUC models. A result of these conflicts became apparent in the dynamic interaction between both types of models in WISE. Using a waterfall approach with a one-way interaction seems to bypass this issue, but in reality only neglects to deal with it: making an equilibrium assumption for a future year, deriving land use demands from this and subsequently interpolating these demands and allocating them on a grid is questionable in the least.

Many recent LUC models successfully integrate micro-economic principles into their operation. MedAction is an example of this as are many of the current spatially explicit agent based approaches to LUC modelling. But, we are still left with many exogenous sector-based macro-economic variables unaccounted for and practical applications that include the dynamic interaction between macro-economics and agent-based approaches are not yet available.

5. CONCLUSIONS AND RECOMMENDATIONS

In this paper we have compared four ISDSS in the way they have integrated land use and economics. LUMOCAP, WISE and Eururalis link macro-economics to land use, while MedAction integrates micro-economic principles into its operation. All have in some way succeeded in making the interaction, but none have done so fully satisfactorily. Problems are mainly the results of conflicts in the underlying theories: static equilibrium versus dynamic simulation, and the focus of most macro-economic approaches on short-term prediction based on an extrapolation of historic trends, instead of a dynamic approach that captures cause-effect relationships and transitions.

We believe that for an improved integration the theoretical foundations of economics and land use change processes should receive more attention and that integration between these disciplines should start from a theoretical basis, rather than a software coupling between existing models. Evolutionary economic theories such as those promoted by Nelson and Winter (1982), and the applications of “endogenous growth” theory (Lucas, 1988), embodying the creation of ideas and learning, provide some key mechanisms for bridging the described gap. Linking models incorporating these theories with spatially explicit agent based approaches could improve the representation of human behaviour in the model and facilitate the link between macro-and microeconomics. These models are however very data demanding and have major limitations in performing a proper calibration and validation. We advocate that the trans-disciplinary nature of integrating not only LUC and
economic, but also biophysical models, has a strong fit with the emerging principles of ecological economics (Ayres, 2001). LUMOCAP and MedAction are examples of ISDSS that show the benefit of linked socio-economic and bio-physical models, while WISE shows the potential of including locally defined resource limitations on macro-economic behaviour.

ACKNOWLEDGEMENTS

The New Zealand Foundation for Research, Science, and Technology funds the Creating Futures project (ENVW0601) and the Sustainable Pathways II project (MAUX0906), the European Commission has funded the development of LUMOCAP (FP-6 SSPE-CT-2005-006556), and the DeSurvey IAM (FP-6 IP 003950).

REFERENCES

Frisch, R., Editor's Note, Econometrica, 1, 1-4, 1933.
Verburg, P.H., B. Eickhout and H. van Meijl, A multi-scale, multi-model approach for
analyzing the future dynamics of European land use, *The annals of regional science*,
2008.

White R. and G. Engelen, Cellular automata and fractal urban form: a cellular modelling
approach to the evolution of urban land use patterns, *Environment and Planning A*, 25,
1175-1199, 1993.
A scientific workflow tool for biosphere modelling: Carbon Sink Archives

G.A. Alexandrov and T. Matsunaga
National Institute for Environmental Studies, Onogawa 16-2, Ibaraki 305-8506, Japan
g.alexandrov@nies.go.jp

Abstract: Scientific workflows in the field of biosphere modelling include assessment of the state of the art in modelling global biogeochemical cycles. The process of assessment consists of the following steps: collecting model outputs, performing model inter-comparison study, writing summary for policymakers, formulating an agenda for further research. Here, we present a web-based service that automates the entire process.

Keywords: web-based service, Carbon Sink Archives, biosphere modelling, NPP

1. INTRODUCTION

Carbon Sink Archives (CSA), a web-based service for storing, retrieving and analyzing 2-dimensional data as related to the problem of terrestrial carbon sink, provides a collection of model outputs that can be used for benchmarking newly developed models of the terrestrial carbon sink and its components, and a collection of web-based tools for performing relevant tests [Alexandrov and Matsunaga, 2009]. The current version focuses on the annual productivity of global vegetation, that is, terrestrial Net Primary Production (NPP). Terrestrial NPP is a starting point of carbon sink studies [Alexandrov and Matsunaga, 2008]. It shows limits to further growth of human appropriation of naturally produced organic matter and limits to the usage of human-induced carbon sinks for controlling the atmospheric concentration of CO₂.

2. CONCEPTUAL FRAMEWORK

CSA employs the Web 2.0 conceptual framework [O'Reilly, 2005]. This concept reflects and stimulates changes in the ways software developers and end-users use the Web. A Web 2.0 site allows users to change website content, not only to view the information that is provided to them. Moreover, the content of the website is produced and maintained through collaborative efforts that use peer-production method [Benkler and Nissenbaum, 2006] and are based on the self-correction principle [Suber, 2008].

The CSA data collection and software tools are intended to facilitate development of model independent knowledge on the terrestrial carbon sink that in its turn may serve as the basis for policies on stabilizing carbon dioxide concentration in the atmosphere. (For example, the current version of CSA facilitates development of the model independent knowledge on the biosphere potential to supply primary energy source for all non-autotrophic species including humans.) The knowledge developed to address the needs of policy making is referred to as policy relevant knowledge, and therefore, software tools are considered as services for deriving policy relevant knowledge from the model outputs, which are considered as knowledge components.

The tools for deriving policy relevant knowledge are based on the premises of evolutionary epistemology [Hull, 1988]. Scientific advances are assumed to be unpredictable (like genetic mutations). Their vitality, however, depends on whether they fit-for-purpose. In other words, the evolution of scientific theories is considered as a
Darwinian process of natural selection that determines which theory survives and drifts them toward consensus [Bradie, 1994].

The natural evolution of a particular field of science leads to distinction between frontier and core knowledge [Cole, 1992]. The latter is commonly accepted knowledge that has stood the test of time and is confirmed in a number of independent studies. The core knowledge serves as a normative knowledge -- that is, as the basis for judgement about what ought to be.

Once core knowledge is formed, natural selection prevents acceptance of contradicting concepts, theories, and even facts. This entails risk of coming to an evolutionary deadlock. New collaboration tools presented below are designed to avoid this risk and to accelerate internalization of contradicting facts. The contradicting facts are used to form an alternative knowledge, which in contrast to the normative knowledge serves as the basis for judgement about what might be.

3. DESIGN

The current version of CSA is to facilitate evolution of the normative knowledge on the terrestrial NPP, which is defined as an array of model independent estimates of NPP over a geographic grid of half-degree resolution. CSA consists of four major components: collection of models outputs, tools for benchmarking outputs of a new model, a tool for updating the normative knowledge proceeding from the outputs of a new model, a tool for reporting the updated normative knowledge in a policy relevant form.

![Figure 1. Scientific workflows that Carbon Sink Archives is designed to automate.](image)
3.1 Collection of Models Outputs

The collection of model outputs contains, in present, the outputs of 12 models of terrestrial NPP and the outputs of model inter-comparison studies. The output data make an array of NPP estimates over a geographic grid with half-degree resolution.

3.2 Benchmarking Tools

Every new model of terrestrial NPP is purported to make a visible progress in the state-of-the-art by improving either a conceptual scheme or parameterization of a prior model. CSA includes a toolkit of web applications for comparison of global patterns of NPP produced by different models and for performing benchmark tests.

Perhaps the most fascinating of these tools is the tool for performing “progressivity test”. The purpose of progressivity test is to visualize the progress in our knowledge of the global pattern of NPP. Proceeding from the premises of evolutionary epistemology, we assume that the progress can be measured by the decrease in uncertainty that stems from discrepancies in modelled estimates.

Modelled estimates of NPP for a given biome differ from one another, and so the state of the knowledge is conveyed by the mean value of the modelled estimates. The uncertainty in the mean value is quantified through the width of its confidence interval, which depends on the number of models and the standard deviation of the estimates. Since the width of the interval decreases with the number of the models, every new model is expected to improve the certainty of our knowledge. Hence, the metric of model progressivity may be defined as follows:

$$ z = 100 \frac{x - y}{x} $$

where $x$ is the width of the confidence interval for the mean value of the modelled estimates, excluding the estimate produced by the new model, $y$ is the width of the confidence interval for the mean value of modelled estimates including the estimate produced by the new model.

This metric penalizes the models that produce estimates that deviate largely from the normative mean value. The score depends also on the number of models: the more models, the lower a new model score. It is worth mentioning here that the test measures progressivity with respect to the normative knowledge. A really novel model may have a low score, if it opens new horizons in our vision of the global pattern of NPP.

3.3 Inference Tool

The collection of model outputs is viewed as a knowledge base for deriving normative knowledge on terrestrial productivity -- that is, a consensual estimate of this important characteristic of the Earth system. Every addition to data collection launches a computer program that updates normative knowledge. This program also updates the alternative (or frontier) knowledge.

The normative knowledge on NPP is expressed in the form of an array of numbers. Each number represents normative NPP at a given cell of the geographic grid of half-degree resolution, which is equal to the mean value of the normative estimates for this cell (so called, normative ensemble of estimates). The estimate produced by a new model is included into the normative ensemble if it falls within the boundaries implied by this ensemble and reduces the width of the confidence interval for the mean value. Otherwise, it is included into the alternative ensemble of estimates for this cell.

The details of the algorithm are explained elsewhere [Alexandrov and Matsunaga, 2008; Alexandrov and Matsunaga, 2009]. Here, we emphasize that the inference tool does not categorize models into normative and...
alternative. A model may give a quite usual estimate for one cell and an unusual estimate for another cell. Therefore, the inference tool categorizes the model estimate for a given cell instead of the whole array of model estimates.

3.4 Reporting Tool

The latest version of CSA has a reporting tool for continual updating of ‘summaries for policy makers’. Every addition to the collection of model outputs results in automatic updating of the normative knowledge and generates updated versions of the ‘summaries for policymakers’ (Figure 1).

The normative NPP, expressed in the form of an array of numbers, needs a valid interpretation to be used in policy making. A “summary for policymakers” is to report the numbers in a policy relevant way. The reporting tool summarizes the current version of normative NPP according to given rules and updates corresponding fields (and figures) in the “summary for policymakers” template.

The current template is focused at the human share of terrestrial NPP. Human appropriation of NPP is assumed to be 1-2 tC/person/y. Hence, in the regions where NPP per capita is less than 2.5 tC/person/y, one may strive only for a survivable development, because nothing or little remains for other non-autotrophic species. The “summary for policymakers” reports

- the percentage of land that falls within the regions of high-productive, mid-productive, low-productive and non-productive climate, and the percentage of human population living there;
- the percentage of land that falls within the regions of survivable, barely sustainable, sustainable, and conservable development (defined in terms of NPP per capita), and the percentage of human population living there;
- the percentage of land that falls within the regions where sources of energy are highly technogenic, technogenic, biogenic, and highly biogenic (defined in terms of the ratio of industrial carbon dioxide emissions to NPP), and the percentage of human population living there.

4. DISCUSSION

The CSA software tools can be run under Mathematica [Wolfram, 1999] in an offline mode. The online access to these tools is provided through webMathematica [Wickham-Jones, 2006] that integrates Mathematica with the web server technology. The obvious advantage of moving to software-as-a-service paradigm [Mell and Grance, 2009] is that the users need not have Mathematica to run the tools. The disadvantage is that the burden of maintaining computational infrastructure falls on the developers.

This burden is mainly of an institutional nature: the developers need to find an institution that will endorse and support the developed web-based service. In the case of CSA, this is not a trivial task. Research institutions (as well as funding agencies) are preoccupied with the impact. To be endorsed a research project should promise a research breakthrough. Although model assessment, benchmarking and inter-comparison are essential for understanding what is a real breakthrough and what is not, they cannot bring something that can be named as a research breakthrough by in and of itself.

The lack of a proper institutional framework does not serve, however, as an excuse for postponing community-based efforts to form consensual estimates of biosphere characteristics such as NPP and to form consensual understanding of the directions for further research. These efforts require new scientific workflow tools, because “our capacity to generate unprecedented quantities of new data means that the collaboration tools must enable researchers to quickly understand the information produced by their collaborators” [Frame et al., 2009]. Carbon Sink Archives provides the tools of this sort: for quick evaluation of a new model (or a new parameterization of an old model of terrestrial NPP), for quick updating of normative knowledge, and for quick formulation of an alternative hypothesis regarding the global pattern of NPP.

For example, terrestrial NPP is currently estimated at 60 GtC/yr. The consensus about this value was built in 1970s, although the estimates varied from 40 to 80 GtC/yr at that time, and re-analysis of the data [Alexandrov et al., 1999]
revealed that estimates depend on how the data were classified with respect to the major regions of the world and that terrestrial NPP could be estimated at 50 or at 70 GtC/yr from the same set of observations.

Modelled estimates vary widely [Cramer et al., 1999] and the retrospective analysis of modelling efforts [Alexandrov and Matsunaga, 2008] revealed no convergence in modelled estimates of terrestrial NPP. The range of discrepancy remained roughly the same as it was ten years ago reflecting the structural uncertainty [Manning et al., 2004] in our knowledge of this essential biosphere characteristic and inefficiency of traditional means (such as occasional model inter-comparison projects) for reducing the structural uncertainty. Perhaps the progress in the biosphere science would be spurred by new collaboration tools like those provided by Carbon Sink Archives.

5. CONCLUSION

The policy relevant knowledge (or related research agenda) is traditionally produced by groups of scientists appointed to do this work on behalf of a research community. This is essentially a literary work the purpose of which is to write a review and a summary of the views presented in research articles. The widely known example is Intergovernmental Panel on Climate Change, which is considered as an adequate model for a system for assessing the state of knowledge of all key natural cycles including the carbon cycle [Seitzinger, 2009]. Nevertheless, if we are to substantiate words with numbers, then scientific workflow tools, like Carbon Sink Archives, are crucial for boosting evolution of the policy relevant knowledge. Since current knowledge on biosphere characteristics is expressed in digital form, most of the routine work related to model evaluation, forming research hypotheses, and developing policy relevant knowledge may be transferred to machines.

REFERENCES

Seitzinger, S., Is there an adequate system to assess the state of knowledge of all key natural cycles?, *Global Change*, 74, 3, 2009.
Exposing the Kepler Scientific Workflow System as an OGC Web Processing Service

Andrew Pratt a, Chris Peters b, Siddeswara Guru c, Brad Lee b and Andrew Terhorst b

a Hydro Tasmania, Hobart Tasmania, Australia (andrew.pratt@hydro.com.au)
b CSIRO Tasmanian ICT Centre, GPO Box 1538, Hobart Tasmania, Australia (firstname.lastname@csiro.au)
c CSIRO Marine and Atmospheric Research, GPO Box 1538, Hobart Tasmania, Australia (siddeswara.guru@csiro.au)

Abstract: The Open Geospatial Consortium (OGC) - Web Processing Service (WPS) provides an interface for distributed geoprocessing: from discovery and description of available processes, description of input and outputs, to execution and retrieval of results. While the integration of various geoprocessing libraries into WPS implementations negates the need to redevelop common algorithms, authoring of processes for the WPS is still very much in the hands of the software developer, rather than the scientist. We propose that by integrating a workflow execution engine (in our case, Kepler) as a WPS processing engine, process content authoring becomes more accessible to the scientist. Existing geoprocessing framework integration solutions have a large reliance on additional user-generated metadata to properly define capabilities as required by the WPS specification. There may be many different possibilities for integrating Kepler but we are particularly interested in solutions that reuse and extend the model markup description already available in the Kepler workflow descriptions to make the task of adding new processes less onerous for the user. However, runtime typing in Kepler and mismatch in complex type definition between Kepler and the WPS have made this difficult. As a result, we have had to define our own input and output parameter types in a separate file. The proposed Kepler-WPS integration work was inspired by the need to reuse a process for creating gridded rainfall time-series from point-based measurements for ingestion in a rainfall-runoff model.

Keywords: Kepler workflow, Web processing Service, Scientific workflow

1 INTRODUCTION

In recent times, experiments conducted by researchers have become more collaborative in nature. This is partly due to the willingness of scientists to share data, resources and knowledge. This has also enabled scientists to conduct complex experiments with distributed data and resources. Therefore, the need has emerged for systems to couple data with processing programs and visualisation tools. Scientific workflow systems have become a popular choice to couple disparate environments, executable programs and data from different sources [Altintas et al., 2006]. For example, it is possible to provide data from file systems or databases to programs written in Java and visualise the result in an independent visualisation tool. In scientific workflow, the scenario explained is composed into tasks and they are decomposed into components of the workflow.

Even though scientific workflow tools can handle web services to access data and processes, the workflow tool itself is generally a stand alone desktop application. The common practice is to use the desktop application to design the workflow and later share it with others via a website.
like myExperiment.org. From the WPS perspective, there has been a general lack of geopro-
cessing algorithms implemented [Olaya, 2010], inspiring the coupling of several geoprocessing
libraries like SEXTANTE and Geographic Resources Analysis Support System (GRASS) with the
WPS. Currently, the WPS authoring process is more often served by a one-off broad-brush integra-
tion activity of existing low level geo-algorithms rather than piece-wise high level user-generated
content. The non-existence of high level processes produces a need for additional orchestration
technologies on top of the OGC services stack, adding further technologies and complications.

In this paper, we describe the exposure of Kepler [Kepler, 2004] as a Open Geospatial Consortium
- Web Processing Service (OGC-WPS) and some of the challenges and issues that arise. The aim is
to allow users with little software development experience the means to author algorithms visible
to the wider Service-Oriented community. Kepler thus becomes both an authoring tool for web
service accessible process and the composition and execution engine behind the web service.

The remainder of the paper is organised as follows: Section 2 gives an overview of Kepler scien-
tific workflow system. In Section 3, the WPS is explained. Section 4 describes the Kepler-WPS
Integration design and Section 5 describes progress of the prototype. Section 6 gives some of the
challenges in mapping concepts of Kepler to WPS. The conclusion and potential future work is
outlined in Section 7.

2 Kepler workflow system

Kepler is a popular scientific workflow visual authoring and execution tool widely used in ecology,
bioinformatics and hydrology domains [Kepler, 2004]. Kepler introduces domain polymor-
phism and modal models [Brooks et al., 2008] and supports actor-oriented design. Domain polymor-
phism enables kepler to use same component in different domains and modal models allow the
combination of different models of computation. Actor-oriented modelling separates two mod-
elling concerns: component communication (dataflow) and overall workflow coordination (or-
chestration) [Bowers and Ludäscher, 2005].

Directors, Actors, Ports and Parameters are the fundamental components of Kepler. Directors
implement models of computation and handle workflow orchestration. Actors are configured
using parameters and pass data to each other using ports. Actors implement the algorithms or
processing steps that are chained together. An Actor can have a single port or multiple ports.
Ports in an actor are used to define input and output data and are categorised as input, output or
input/output. Relations are used to branch the data flow, and to send the same data to multiple
actors. Workflows in Kepler are stored in an XML-based format called Model Markup Language
(MoML). Figure 1 shows the kepler workbench environment.
There are several projects in different domains that use Kepler for data acquisition, archiving, and analysis. The Science Environment for Ecological Knowledge (SEEK) is a system designed to facilitate data acquisition and archiving, integration, transformation, analysis and synthesis of ecological data [SEEK, 2005]. Real-time Environment for Analytical Processing (REAP) focuses on developing scientific workflow tools that can be used to access, monitor, analyse and present information from field-deployed sensor networks [REAP, 2007]. Due to its flexible design, pPOD has adopted Kepler for the bioinformatics domain [pPOD, 2010]. Jäger et al. propose using Kepler to orchestrate distributed geospatial processing in its traditional role as a desktop application, but without the focus on interoperable geospatial web services [Jäger et al., 2005]. We choose Kepler for its flexibility of reusing existing components, and the separation of its model of computation (in the form of directors). These two features give us the ability to reuse existing environmental data capture, processing actors and execute workflows with different models of computation without any modification to the workflow.

3 WEB PROCESSING SERVICE

The OGC-WPS is a web services based standard for providing description and execution of processes with a particular focus on geospatial processing of vector and raster data, designed around the publish, find, bind pattern [WPS, 2007]. Previous work in WPS-based research has exposed various Application Programming Interfaces (APIs) and geospatial toolsets to the WPS, with the aim of increasing the number of available low level geoprocesses and improving the runtime performance of resource-intensive models. Woolf and Shaon propose an approach to encapsulate grid infrastructure within the WPS, providing the means for running resource-intensive processing through a standard processing interface [Woolf and Shaon, 2009]. The SEXTANTE geospatial analysis tools are exposed through the WPS, providing access to more than 200 geoprocesses [Schäffer, 2009]. Brauner and Schäffer exposes the GRASS to the WPS, and proposes a method of chaining low-level processes together using Business Process Execution Language (BPEL) [Brauner and Schäffer, 2008].

The WPS provides several operations for discovering, querying and executing processes via various interfaces including Simple Object Access Protocol (SOAP), Hypertext Transfer Protocol (HTTP) GET and POST. The WPS Specification summarises the operations as follows [WPS, 2007]:

![Figure 2: Typical process discovery and execute sequence](image-url)
Pratt et al. / Kepler as OGC Web Processing Service

- **GetCapabilities**: This operation allows a client to request and receive back service metadata (or Capabilities) documents that describe the abilities of the specific server implementation.
- **DescribeProcess**: This operation allows a client to request and receive back detailed information about the processes that can be run on the service instance, including the inputs required, their allowable formats, and the outputs that can be produced.
- **Execute**: This operation allows a client to run a specified process implemented by the WPS, using provided input parameter values and returning the outputs produced.

Figure 2 shows a typical client-server call sequence to discover, query and execute a process on WPS.

Schäffer proposes a transactional profile, defining a DeployProcess and UndeployProcess operation for dynamically registering new processes with the WPS [Schäffer, 2008]. These transactional operations introduce the concept of deployment profiles, which allow technology specific profiles to be implemented on WPS instances. The default profile for the 52°North WPS implementation is BPEL. The solution proposed in this paper includes a deployment profile to allow Kepler Model Markup Language (MoML) to be deployed to the WPS, using the 52°North deployment profile design.

### 4 Kepler-WPS Integration Design

![Diagram of workflow lifecycle](image)

Figure 3 illustrates the full lifecycle of the proposed deployment of kepler workflow on WPS. In step 1, a user designs, experiments and tests a workflow using the traditional Kepler workflow environment. In Step 2, they use the DeployProcess call of the transactional WPS to submit the workflow in the form of a Kepler MoML file (xml-based). In Step 3, the user annotates the workflow through the admin webpage of the WPS. The WPS parses the Kepler MoML file, and prompts the user with the inputs and output ports of the workflow requiring annotation with supported WPS simple or complex types. This creates a WPS configuration file on the server that associates the relevant port descriptions with valid WPS type specifications. Steps 4, 5, 6 and 7 are the typical sequence for interacting with the WPS through process discovery, additional process metadata retrieval, and finally execution and analysis of results. Step 8 may occur when the user decides that the workflow is of no further use or they wish to edit and replace with a newer version of the workflow.

Figures 4, 5 and 6 show the interaction of the components for each of the WPS request types. In Figure 4, the WPS hosts an algorithm repository (a folder on the file system) where Kepler MoML files are stored along with port data type annotations that conform to WPS simple and complex type definitions. Process names are retrieved directly from the top-level Actor in the Kepler workflow file. During the DescribeProcess request in Figure 5, metadata from the Kepler
Figure 4: GetCapabilities request

Figure 5: DescribeResponse request

Figure 6: ExecuteProcess request
workflow file is combined with the corresponding WPS configuration file containing the annotated port types to provide a description of the process along with names and data types of its inputs and outputs. In Figure 6, graphical actors are replaced with non-graphical equivalents in order to allow the workflow to execute without a Graphical User Interface (GUI). This modified workflow is then further modified so that it writes inputs and outputs to a uniquely named job folder for the workflow run, and its ports are dynamically connected to Actors linking it to input parameters provided through the WPS interface. Upon completion of the process execution, the workflow outputs are retrieved from the uniquely named job folder and returned through the WPS interface.

5 IMPLEMENTATION

We have implemented a bare skeleton of the solution, by adapting the 52\°North WPS implementation. The handling of complex types, deployment profiles and type annotation has not yet been implemented.

The significant working parts of the prototype are:

- Implementation of the WPS GetCapabilities interface via linking process names to MoML files in the file system
- Implementation of the WPS DescribeProcess by parsing associated Kepler MoML files for simple types.
- Draft implementation of the WPS ExecuteProcess call through the Kepler runtime engine via non-graphical replacements actors using Hydrant libraries for parameterless processes [King, 2010].

6 ISSUES AND CHALLENGES

The aim of this work was to provide a type marshalling layer between the WPS and Kepler that would allow workflows designed in Kepler to be uploaded to the WPS and run seamlessly without additional process metadata. However, there are some challenges to make this a reality.

There is a mismatch between the Kepler runtime typing and the WPS static typing. In the WPS, the names and types of input and output parameters must be declared during description of the process. In Kepler, while the input and output parameter names can be inferred from port names, the types of those parameter may not necessarily be defined declaratively. Kepler actors can choose to implement runtime type checking allowing them to cope with a variety of different input parameter types. For instance, an actor that takes two numbers as input, sums them and returns a single number as output need not be rewritten for each numeric type available, but instead checks that the type of each of the input parameters conforms to the set of intended types that can be sensibly summed, and asserts the output parameter type as that type at runtime. While this feature is useful in Kepler for reuse of actors across a range of well-defined types, it lacks the sort of declarative definition that the WPS requires. As a result of this mismatch, the number of actors that can be automatically mapped to the WPS is drastically reduced, and artificial constraints are imposed during the workflow design phase in order to cope with the WPS context later in the workflow lifecycle.

Kepler has only fairly rudimentary capability for declaring complex types as the inputs to workflows or actors (workflow components). Kepler ports can be labelled with a type, but for all types that aren’t primitives (int, float, double, string etc.), the base type Object is declared on the port, and it is up to the actor to type cast from Object to the appropriate class. Developers can extend Kepler to define new port token types, but this involves software changes to the Kepler distribution for each new type added. This is in sharp contrast to the WPS that defines complex types with a combination of mime-type, encoding, and schema descriptors. Since the WPS is defined as an interface specification, it expects that inputs and outputs need to be defined in an unambiguous, platform-neutral manner in order to deserialize data streams through the interface in a format that the process can understand. The schema property is used for XML based formats, and is a
URL pointing to the XML Schema definition. Table 1 lists a variety of different triples of mime-type, encoding and schema descriptions for some well known and not-so-well-known formats and compares them against their declarative type in a standard Kepler distribution.

<table>
<thead>
<tr>
<th>WPS datatype</th>
<th>Mime-type</th>
<th>Encoding</th>
<th>Schema</th>
<th>Kepler datatype</th>
</tr>
</thead>
<tbody>
<tr>
<td>JPEG</td>
<td>image/jpeg</td>
<td></td>
<td></td>
<td>Object</td>
</tr>
<tr>
<td>NetCDF</td>
<td>application/x-netcdf</td>
<td></td>
<td></td>
<td>Object</td>
</tr>
<tr>
<td>GML</td>
<td>text/xml</td>
<td>UTF-8</td>
<td>./om/1.0.0/observation.xsd</td>
<td>Object</td>
</tr>
<tr>
<td>O&amp;M</td>
<td>text/xml</td>
<td>UTF-8</td>
<td>./gml/3.2.1/gml.xsd</td>
<td>Object</td>
</tr>
</tbody>
</table>

Table 1: Example Complex Data Type descriptions (note: schemas truncated for brevity)

7 CONCLUSIONS AND RECOMMENDATIONS

We initially set out to provide a simple wrapper over Kepler to allow its workflows to be exposed to the OGC Web Processing Service without annotation to the workflow descriptions or links to external metadata. This was not an easy task due to a mismatch between input and output datatype descriptions in the two systems. Therefore, we resorted to using the existing WPS process description file in use by the 52°North implementation to describe those parts that Kepler can’t be guaranteed to describe through its MoML. This proposed solution provides the ability to author and test workflows in a traditional GUI environment before deploying to the WPS for later discovery and execution by a wider community of Web service client applications. This provides the research community with a means to deploy larger workflows without the need for complex orchestration technologies to bridge numerous web services calls and pass complex data sets. Additionally, it abstracts server-side processing, allowing scientists to be content authors in the Software as a Service (SaaS) paradigm, alongside software developers.

While there are many existing workflow engines, we were particularly interested in reusing Kepler workflows already produced for the project. Future work would be most useful in examining typing systems in other workflow tools and solutions that close the conceptual gap in input/output data type definition in order to make switching from visual workflow authoring to WPS-based process execution more transparent.

The major limitations in coupling Kepler with the WPS are related to the way in which input and output types are declared. The Kepler optional runtime typing doesn’t conceptually fit well with the WPS declarative type system. Additionally, the use of Object for complex types in Kepler, which within a Java-only environment allows actors to runtime type check and provides maximum flexibility and reusability, wasn’t a good conceptual fit for the WPS declarative type system based around mime-type, encoding, and schema for complex types. Therefore, future studies around these aspects and how they are addressed in various workflow environments and the potential for addressing with enhancements to Kepler may be considered. Additionally, the integration of workflow tools with other OGC services may be explored, to examine the benefits of different interface descriptions on the ease of coupling the service interface with the execution engine.

ACKNOWLEDGMENTS

Part of this work was conducted when Andrew Pratt and Siddeswara Guru were working in CSIRO Tasmanian ICT Centre.

This project is jointly funded by the CSIRO Water for a Healthy Country Flagship and the Tasmanian Government. The Tasmanian ICT Centre is jointly funded by the Australian Government through the Intelligent Island Program and CSIRO. The Intelligent Island Program is administered by the Tasmanian Department of Economic Development and Tourism.
REFERENCES


Hydrologists Workbench – a hydrological domain workflow toolkit

Susan M Cuddy and Peter Fitch
CSIRO Land and Water, Canberra, Australia. susan.cuddy@csiro.au

Abstract: Large-scale hydrological modelling exercises are becoming routine to address concerns about future water availability. These involve coupling multiple models to simulate the water cycle; time is of the essence; and the questions to be answered (and associated indicator metrics) are multi-dimensional. As these exercises become more complex and the volume of data increases exponentially, automating management of the flow of data through models, accessing relevant tools, while ensuring auditability and compliance, become essential. The Hydrologists Workbench (HWB) is being developed to meet this need and its shape and content are informed by recent large-scale sustainable (water) yield modelling exercises in Australia. Built on commercial off-the-shelf scientific workflow software which provides the workflow, audit and governance utility, it draws together public domain and proprietary hydrological, statistical and GIS toolkits with tailored workflows to provide an extensible portal for the provision and management of (one-off or routine) modelling exercises. This paper describes the intent, design structure and current state of development of the HWB and uses the example of a reporting workflow that executes a series of data transformations to produce maps, tables and plots for a monthly water situation report. The paper concludes by identifying key challenges that have emerged, and evaluates progress to date against a priori design objectives.

Keywords: scientific workflows; hydrological modelling

1. INTRODUCTION AND BACKGROUND

1.1 Project background

The Hydrologists Workbench (HWB) as a product and a project developed from the authors’ experiences with several large-scale modelling exercises in Australia (from 2007 to present). These studies (the Murray-Darling Basin Sustainable Yields project [CSIRO, 2007-2008] and its follow-on projects in Tasmania, south-west Western Australia and Northern Australia) were commissioned by the Australian Government to predict the likely impact of future climate on Australia’s water resources, and were undertaken by Australia’s national research organisation, CSIRO. In Australia, the availability, quality and management of water are significant national issues. Australia’s water resources are under threat from a number of factors including climate change, over-allocation, and increasing consumptive demands. Knowing how much water is available, together with knowing how much water is needed for the environment and consumptive uses, provides the basis for establishing sustainable diversion limits for water resources across the country.

The suite of Sustainable Yields projects coupled climate models (how much rain), catchment water yield models (how much runoff), operational river system models provided by state water agencies, groundwater models, and water accounting models to quantify basin water availability in a suite of reporting products. Models were stitched together within a framework built on the fly by expert software engineers [Yang 2009]. While adequate for the modelling exercise, such a system cannot be expected to (and did not) provide an extensible platform for undertaking further studies. The tasks involved
were extensive, complex, time-consuming and composed of many manual steps. The generation of a typical reporting product required the collection and storage of large volumes of input files (for model calibration), control of execution of complex integrated sets of models, creation and curation of thousands of modelling results (files, databases), modelling and geo-processing artefacts (map and chart templates), and final archive and metadata tagging (repeated for every sub-basin within the project areas). These tasks could be easily envisaged as large scientific workflows, coupled with business and reporting workflows to manage the approval and publishing activities.

An equally significant initiative of the Australian Government in 2007 was to make the Australian Bureau of Meteorology (the Bureau) responsible for compiling, managing and disseminating Australia’s water resources information; and making that information as comprehensive and accessible as weather and climate information through conducting periodical water resources assessments, producing annual national water accounts, and providing national water availability forecasts. This is an enormous responsibility and workload; and requires large-scale modelling exercises and reporting processes that are commensurate with those undertaken for the Sustainable Yields projects. With these shared experiences and needs, the Bureau and CSIRO formed a Water Information Research and Development Alliance (WIRADA) to address the Bureau’s operational needs through CSIRO’s R&D expertise in water and information sciences.

The Hydrologists Workbench project was established within WIRADA and the objectives of the project were defined in response to the Bureau’s needs (and Sustainable Yields learnings). The HWB is also the name of the key deliverable of the project – a workflow toolkit, customised to meet the needs of hydrological modellers and hydrological reporting, allowing the integration and reuse of key processes. This paper describes HWB progress to date. The rest of this section describes evaluation criteria and functional requirements. Section 2 is a short discussion on the current implementation. Section 3 describes the case study and some of the challenges that it presented. The paper concludes with an assessment of progress against evaluation criteria.

1.2 Case for investment in scientific workflow tools

Current practices in hydrological modelling exhibit a heavy reliance on individuals and manual steps – leading to lack of resilience, poor use of valuable skills, and increased risk of non-reproducible errors. Information is dispersed depending on the culture of the modelling team. There is a heterogeneity of toolsets and capabilities with islands of specialisation (e.g. R, Excel, Perl, Matlab, Python, Fortran, C#) with manual transmission in between. There are divergent and non-conforming data formats and idiosyncratic access arrangements; and basic traceability is reliant on ‘off-line’ documentation (if documentation exists at all). These practices were all exhibited in the Sustainable Yields projects. Insights gained from being part of these projects, together with the similar needs of the Bureau, confirmed the need for a tool that would support the rapid construction and execution of computationally demanding integrated modelling and reporting systems within an environment governed by business rules and protocols. Scientific workflow software, as a technology for composing and executing chains of scientific processes, appeared to meet this need; and the HWB project was initiated between CSIRO and the Bureau to undertake the necessary research and development to apply this technology.

An investment in a particular technology by itself cannot guarantee improvement in practice, as there are so many other influences on modes of working (e.g. disciplinary culture, team cohesion, deadlines, organisational business rules). Significant investment needs to be supported by a convincing case identifying realisable benefits. Up-front costs associated with such an investment (research and development, training, etc.) need to be compared to the ‘hidden’ costs of current work practices, especially the costs of high manual handling associated with data manipulation, versioning and archiving, and the unmitigated risk of weak or no audit trails. The four realisable benefits promoted to justify investment in the HWB were that adoption would:

- facilitate investigative research and foster reproducibility
• result in more robust and efficient modelling exercises
• reduce the manual handling and tedium, resulting in less error and faster turn-over for research workflows
• improve efficiency and reduce duplication.

Progress against these evaluation criteria is summarised in the final section of this paper.

1.3 Hydrologists Workbench (HWB) Objectives

The project brief was to develop an integrated hydrological modelling and reporting platform, drawing upon the new technologies of web services and scientific workflows to provide a solution to integration and workflow composition for automation. Most importantly, HWB was to be built on an existing scientific workflow application. Such a platform should provide:
• the ability to discover and visualise spatio-temporal data from Bureau databases and other sources
• interfaces to external services and toolsets that manipulate and visualise hydrological data
• interfaces to existing hydrological models to allow integrated modelling and reporting tasks to be easily generated and executed
• the ability to save workflows to be rerun at a later date, thus providing the benefits of repeatability, transparency and auditability.

Benefits to the Bureau would include:
• enabling Bureau hydrologists to select and integrate components of other tools, workbenches (e.g. ArcGIS) and modelling applications; thus reducing dependency on a single workbench or modelling application
• providing a means to automate repetitive tasks, providing consistency and rigour in modelling and reporting tasks
• reducing the number of GIS and software engineers required to support integrated hydrological modelling.

It was envisaged that the interfacing of HWB to models and data would be achieved through interface tools, thus exposing the models and data for HWB users. Libraries of tools useful for hydrological modelling, e.g. geo-processing and visualisation, would be built. Importantly, the HWB must provide for the inclusion of scripts (R and Python) written by Bureau staff to provide geo-processing and statistical processing.

A further design imperative was that HWB comply with international and national standards and protocols, such as those being developed within WIRADA, including the Australian Hydrological Geospatial Fabric [Atkinson et al. 2008] and the Water Data Transfer standards [Walker et al. 2009] projects.

HWB would also need to be mindful of governance (of the platform and its components) requirements, both from within the Bureau and across CSIRO and the Bureau, to ensure that HWB would fulfil the objectives of reusability, accountability and auditability. An HWB governance framework is posed in Box [2010].

Other non-functional requirements address deployment considerations (lifecycle management) and usability of the product. The requirement to support the production of Bureau data products raises significant and sufficient architectural and implementation challenges to drive the R&D.
1.4 Linkages with Data Service and Modelling Initiatives

There are several hydrological data service and modelling initiatives underway internationally and in Australia; and the HWB needs to take advantage of these. Figure 1 is an early view of the kind of functionality to be provided through HWB.

AWRIS – Australian Water Resources Information Service – database under construction within the Bureau to ingest and serve water resources data [BoM, 2009]; HIS – Hydrologic Information Service [CUAHSI, 2009]; AWDIP – Australian Water Data Infrastructure Project to develop interoperability standards and protocols [BRS, 2009]; IQQM – River system model widely used in eastern Australia; TIME – a software development framework for creating models [Rahman et al. 2003]

Figure 1. An early view of the functionality to be provided through HWB, showing links to external data service and modelling initiatives

2. IMPLEMENTATION

The HWB project is nearing the end of the 2nd year of a five-year programme. The first year was dedicated to exploring the potential of various workflow platforms and determining their suitability for hydrological modelling; and was largely driven by the researchers within the team (see Guru et al. [2009] for details). The second year has focussed on an application case study (described below), working with Bureau staff to identify opportunities and impediments to individual and organisational adoption of the technology. Return on investment must be demonstrated at this stage to ensure ongoing support for the project.

2.1 Choice of Scientific Workflow Platform

Several platforms were trialled with Kepler [Ilkay et al. 2004] adopted to develop demonstration workflows. Subsequently Trident [Microsoft Corporation, 2009] was adopted. Trident is based on Windows Workflow Foundation and is part of the .NET framework. Trident provides a graphical user interface for building workflows, which are composed of activities (coded as .NET classes). These activities and workflows are managed via a registry which can be local or shared through a central registry. Being a .NET product was advantageous because many of the hydrological tools being developed within WIRADA and CSIRO are .NET applications. The selection was supported by a technical evaluation (e.g. flexibility, robustness, software interoperability) that established that Trident met the requirements for use as the core technology for HWB. More details on Trident from a HWB perspective are discussed by Perraud et al. [2010] and Box [2010].
3. CASE STUDY

Monthly reporting of the water situation in selected catchments across Australia was nominated by the Bureau to prototype use of the HWB. The models and analyses to produce the reports were already encoded in Python and R scripts, and it was hoped that this would facilitate rapid implementation. The workflows are to become part of a suite of scheduled monthly tasks, managed by operators other than those who have developed the workflows.

Meetings were held with Bureau staff responsible for production of the reports (and the writing of the scripts) to clarify the processing steps. Figure 2 shows a flow chart of the monthly water situation reporting, comprising:

- accessing and sub-setting gridded rainfall, soil moisture, streamflow and groundwater data based on spatial and temporal qualifiers
- aggregation at specified spatial (sub-catchment) and temporal (monthly) scales
- calculating statistics for the subsetted data
- producing artefacts (maps and graphs)
- transferring these artefacts to other products that combine with interpretive text for reporting.

![Figure 2. Flow chart of basic processes in the monthly Water Situation Report workflow, showing the role of Python and R scripts (reproduced courtesy of Bureau of Meteorology)](image)

The report has been decomposed into separate workflows for each attribute (being rainfall, soil moisture, streamflow and groundwater), with the rainfall workflow being the first to be composed. This composition exposed many behavioural, technical and organisational challenges, some of which are discussed in the following section.
4. CHALLENGES

A preface to this discussion on challenges is an admission that scientific workflow technologies were new to the project team. Thus, it was a learning year, both in coming to grips with the capabilities (and limitations) of the Trident product, and in approaching the design of process integration/coupling differently from the more traditional approaches to design and build of modelling packages (to which the team was well accustomed). For example, there were no user interfaces to build, and no databases to design; this signalled a reduction in control of the design process that took some getting used to.

4.1 Technical Challenges

The most significant technical challenges encountered were how best to:

- reuse existing scripts (with minimal refactoring)
- interface with ESRI’s ArcGIS
- interface with classes of TIME models
- utilise Trident registry databases for managing development, testing and deployment
- package and distribute workflows
- specify Trident activity requirements and document solution
- compose workflows
- balance flexibility with usability in the design of Trident activities and workflows.

While we made headway on all of these, progress on the first two of these challenges is described in this sub-section (paper length precluding discussion on all).

Reusing existing scripts

Existing Python and R scripts contained the controls to partition and manage the flow of data and processes. Rewriting the scripts was not a preferred option as a key HWB objective is to support interoperability of programming languages. A simple solution would have been to write Trident activities that ‘wrapped’ the existing scripts. However, this would result in Trident activities that were not reusable, and reusability is another key objective. Part of the learning exercise was to pull the scripts apart to understand the processes and refactor as smaller, potentially reusable Trident activities. During this process conventions for writing scripts were developed.

Interface with ESRI’s ArcGIS

Spatial subsetting and zonal statistics are common processes in assessing regional water resources. The water situation report uses existing grids of rainfall, soil moisture, streamflow and groundwater levels, over the preceding month, to report on water availability. A library of Python scripts had been written to perform these tasks, calling ArcGIS functions. Initially these were invoked by writing customised Trident activities that ‘wrapped’ the script. A better solution was to provide full access to the ArcGIS commands from HWB (each command having its own Trident activity). Composite ArcGIS tools which string together commonly used geoprocessing sequences (e.g. Select, Mask, Clip) have been written, wrapped by customised geoprocessing workflow fragments. This has proved to be an excellent integration solution which we believe is applicable to integration with other external workbenches.

4.2 Behavioural Challenges

Two significant behavioural challenges arose. The first was the need to demonstrate to the script developers (who are effectively end-users of HWB) the value of using workflows rather than scripts for run control. The perceived overheads were justified in terms of the workflow approach providing transparency of process (ie the visualisation of the processing sequence was appealing), and provenance tracking of data and run execution.
The second was the separation of the role of developer (in this case the programmers who coded the Trident activities) from that of workflow composer. This was intentional as the workflow composer would typically not be a skilled programmer and a key goal of HWB is to reduce the dependence on programmers, i.e. the underlying Trident activities had to be coded in such a way that they were meaningful to and usable by others – always a challenge.

4.3 Governance Challenges

Governance challenges are considerable, but not insurmountable. Trident’s use of a central registry for managing workflows and activities, and its ability to track the provenance of workflows and the outputs of their execution, are useful. However, within an operational environment such as the Bureau, substantial investment is required to establish a governance framework to manage development, production and deployment of activities, workflows, input data and output products – without compromising being able to use HWB as a ‘sandbox’ for scientific experimentation. A theoretical governance framework for HWB has been proposed by Box [2010], recognising that governance arrangements are ultimately determined by the cultural practices and obligations for quality control/assurance of the organisation. In the case of the Bureau, and the production of water resources assessment reports, every aspect will be the subject of extensive public scrutiny and strong governance is of utmost importance.

5. DISCUSSION AND CONCLUSIONS

Progress against evaluation criteria

Challenges, such as those described above, were expected. The reimplementation of the water situation report demonstrated very clearly that the adoption of scientific workflow technology is not just a technological issue – it also requires behavioural and organisational change. At the start of the project, we posed four criteria that would need to be met to justify the investment in workflow technology – that they would facilitate investigative research and foster reproducibility; result in more robust and efficient workflows; reduce the manual handling and tedium; and improve efficiency and reduce duplication. Have we met these criteria? The answer would be – not yet. While the project team was well equipped to solve many of the technical challenges, the emergent behavioural and organisation challenges require longer term investment to resolve, and must be done in partnership with the end-user, the Bureau. However, the project has demonstrated enough promise that we feel confident in our choice of technology as capable of meeting the Bureau's scientific modelling and reporting needs.

Learnings

The design objective of reuse and adapt (with re-invent a last resort) has had a profound effect on the relationship between the HWB professional programmers and Bureau scientists who write code (and scripts). Script writers rarely have formal programming training, yet their knowledge as encoded in their models and scripts form the core of the activities on which the workflows have been built. Any perceived reduction in coding quality is more than adequately compensated by the benefits of ownership and understanding of the underlying models and scripts by the scientists themselves.

Understanding how best to construct workflows (e.g. granularity and complexity of individual activities) is like writing good code – it comes with experience and sharing. Writing workflows that are understandable, usable and re-usable is definitely an acquired art, in fact Gil et al. [2007] describe it as a ‘black art’.
Current State of HWB and Future Plans

The vision of building libraries of tools (delivered as activities) has been constrained to date by the need to build tools to meet the Bureau’s immediate reporting needs. The workflows to produce these reports are well advanced and are becoming operational within the Bureau. Over the next year, the composition of workflows will transfer to the Bureau, allowing the development team to build the functional components envisaged in Figure 1.

Our ambition for HWB is that it is well aligned with the skills and culture within the Bureau such that it becomes the technology of choice for scientific experimentation and water resources assessment. We believe that the development of the rainfall reporting workflow, with the challenges met and those overcome, has successfully demonstrated the capability of HWB and the potential of the technology; and look forward to building HWB’s repertoire together with the Bureau and the wider scientific modelling community.

ACKNOWLEDGMENTS

WIRADA is managed within the CSIRO Water for a Healthy Country Flagship and the authors thank WIRADA management for its continuing support of the project and its objectives. The authors are indebted to the HWB project team within CSIRO and the Water Division at the Bureau of Meteorology for their contributions of background material and diagram to this paper; and the anonymous reviewers of this paper.

REFERENCES

Box, P., Hydrologists workbench: a governance model for scientific workflow environments, [this conference], 2010.
Perraud, J-M, Q. Bai, and D. Hehir, On the appropriate granularity of activities in a scientific workflow applied to an optimization problem, [This conference], 2010.
Hydrologists workbench: A governance model for scientific workflow environments

Box, Paul  
CSIRO Land and Water, Sydney, Australia  
paul.j.box@csiro.au

Abstract: Scientific workflows (SWF) are an emerging approach that enables scientists to compose and execute complex, distributed scientific processes. The approach is premised on the ability to compose, publish, share and reuse workflows across distributed communities of collaborating scientists. Scientific workflow software (SWFS) provides a technical framework to compose, publish, and reuse SWFs together with data, functionality and computational and other resources upon which they rely. Tools and components are built using service oriented architecture approaches with data and functionality exposed through services. Together, interoperable components from different initiatives form information infrastructure (which as socio-technical endeavours, present specific governance challenges. As workflows, and the resources upon which they rely are distributed and under the ownership of different people and organisations, an enabling governance framework is required. Governance comprises authority structures, roles, policies, processes, and mechanism that enable collective decision-making, and collaborative action to achieve common goals. This paper presents key challenges related to the governance of socio-technical aspects of scientific workflows and workbenches. These include the 'human activity systems' that enable the design, creation and sharing of workflows and the technical governance of workflows, and underlying resources throughout their lifecycles. The paper also presents a conceptual model for scientific workflow governance and discusses its application in the Hydrologists’ Workbench (HWB) project.

Keywords: scientific workflow software; governance; information infrastructure

1. INTRODUCTION

Scientific workflows (SWF) are an emerging approach that enables scientists to set up and run in silico experiments through the composition and execution of a chain of scientific processes without the need for programming. Scientific workflow software (SWFS) provide a technical framework to create, publish, reuse and manage workflows, together with data, functionality, models, and computational and other resources upon which they rely.

Although the promise of scientific workflows is clearly recognised, significant challenges associated with the creation, sharing and execution of scientific workflows remain (Gil et al. [2007]). Deelman and Chervenak [2008] describe challenges associated with the workflow cycle and Goderis el al. [2005] identify a number of bottlenecks to the reuse of workflows, including the discovery and reuse of workflow fragments. Given the collaborative, distributed, service-oriented nature of scientific workflow environments and the fact that workflows and the resources upon which they rely are often under custodianship and operation of different people and organisations, governance of scientific workflows is a critical issue. Shon et al. [2008], identify a number of dimensions of scientific workflows that require governance include reusability, workflow reproducibility and platform extensibility.

This paper describes key governance challenges associated with the application of SWFS in the hydrology domain, together with a conceptual framework used to inform governance solutions. The paper uses as a case study, the Hydrologists Workbench project, a scientific
workflow environment, for hydrological analysis and reporting. The paper describes and characterises some key SWF governance challenges associated with workflow design, and composition and the extension of the SWFS encountered though the HWB project. Finally, the paper presents a conceptual model to address governance in collaborative environments and briefly illustrates its application.

2. THE HYDROLOGISTS WORKBENCH

The Hydrologists Workbench (HWB) is a collaborative project undertaken by CSIRO and the Australian Bureau of Meteorology (the Bureau) (Cuddy and Fitch [2010]). The project aims to develop an integrated hydrological modelling desktop application built around scientific workflows. The HWB is intended to support scientific and reporting activities to meet the Bureau’s expanded role under the Water Act 2007 encompassing the analysis, management and reporting of Australia’s water resources information (DEWHA [2007]).

2.1. Trident

Following a review of candidate technology platforms, reported by Perraud et al. [2009] Trident was selected as SWFS platform to be used for the HWB. Trident is a suite of applications developed for the composition, management and execution of scientific workflows (Microsoft Corporation [2009]). Trident provides much of the core functionality for the HWB. However, HWB has a much broader scope than Trident and can be conceived as an integrated environment for hydrological scientific workflow design, composition, execution, publication and management. Thus the HWB also comprises the human activity system and associated mechanisms and tools that enable the collaboration necessary to create, share, execute and manage workflows, their component parts and the resources upon which they depend.

Trident is based on Windows Workflow Foundation, and is part of the .NET framework. Trident includes two applications: Composer for composition of workflows, and Management Studio for the management of workflows. Composer provides a visual design tool for composition of workflows by dragging configuring and connecting ‘activities’, the atomic building blocks for workflow composition. Both applications leverage a registry which handles the registration, management, discovery and access to workflows, activities, data products and other artefacts that comprise or support workflows.

Upon installation, Trident has a limited set of pre-programmed activities that provide data access, transformation and flow control such as the ‘if else’ activity. Communities extend the basic functionality by developing custom activities to perform processing required for scientific workflows in specific domains. Much of the initial effort of the HWB project has focused on developing core functionality commonly required in the hydrology domain.

2.2. Integration with Geo-processing Workbenches

A key objective of the HWB is to interoperate with other workbenches and environments in use by target communities, to enable users to use the most appropriate tools for a specific task and orchestrate their execution through the HWB. As hydrological modelling, analysis and reporting has an intrinsic spatial dimension, the integration of HWB with geo-processing frameworks was a key research priority. The initial geo-processing framework targeted for integration was the ESRI ArcGIS desktop environment, a widely used geo-processing framework within the Bureau and CSIRO.

Integration with ArcGIS is based on Trident interacting with the geo-processing tools exposed through the ArcGIS geo-processing framework. To extend the default ArcGIS tool set and to create geo-processing workflow fragments custom Python scripts are written. These are exposed as tools in the ArcGIS geo-processing framework. A dedicated .Net activity is created to launch each custom (user defined) and default ArcGIS tool provided as part of the geo-processing framework. Default tool activities are generated by parsing xml
files that describe each tool, supplied as part of ArcGIS product. Activities for custom tools are developed using a hand crafted xml file that describes the tool and its parameters. Using this approach, the ArcGIS geo-processing framework provides registration management and access to the geo-processing tools.

2.3 Water Reporting Workflow Design and Composition

An initial focus area for HWB was the development of workflows to generate information products for monthly water situation reports which are under development by the Bureau. Although the workflows developed can be characterised as production rather than scientific workflows, the approaches to the design of workflows and required functionality are considered to be broadly applicable to the scientific workflow context.

The approach used for the design of workflows was loosely based on the work of Gil [2007] who identifies four stages of workflow design with attendant levels of workflow abstraction:

- Workflow sketches - used for initial workflow requirements specification
- Workflow templates - execution-independent specification of the processing steps, components to be used and the data flow between them
- Workflow instances - execution-independent with specification of input data
- Executable workflows – workflows instances assigned to resources for execution

This distinction provides a useful conceptual framework for approaching design and addressing reuse of workflows. In the context of HWB four levels of abstraction were used as the basis for design process. Firstly, Workflow sketches were compiled as UML activity diagrams. These were used to document existing functional components, identify functionality to the developed and for the refactoring of components to enable reuse across multiple workflows. At the next level of abstraction, workflow templates were created in Trident as generic reusable workflows that were intended for reuse. Workflow instances based on templates were created with parameter values including data sources. Finally, rather than executable workflows the project adopted the term workflow instance runs i.e. an execution of the workflow instance that produces concrete data outputs.

There were typically a number of iterations through each step of this design process to refactor components and workflows based on an improved understanding of required granularity and reusability. Perraud et al. [2010] provide an analysis of the appropriate level of granularity as part of evolving workflow design. At each step of the process a number of artefacts are created that express or are implementations of agreements about how a component should behave. These included, workflow sketches as UML activity diagrams, specification of .NET Trident activities and ArcGIS geoprocessing tools to be developed, development, and production versions of activities, workflow fragments and complete workflows and other artefacts such as configuration files, scripts, local data used as input to workflow activities and sample outputs which comprised part of the specification set.

Working in a collaborative environment across several agencies to extend the core functionality of Trident and to develop custom geo-processing tools in ArcGIS, required an overarching governance framework to ensure that agreements and their implementations were properly managed.

3. GOVERNANCE

To collaborate across organisational boundaries and build effective communities that are able to share and re-use information, processes, and knowledge, socio-technical information infrastructures are required. Aanestad et al. [2007] note that these infrastructures are intrinsically socio-technical endeavours and as such, require governance. Governance provides an overarching and enabling decision-making and accountability framework comprising authority structures, roles, policies, processes, and mechanisms that enable collective decision-making, and collaborative action to achieve common goals (Box and
Rajabifard [2009]). Governance provides oversight and an enabling framework for management activities and can be conceived as three interacting dimensions:
- the what – the scope of governance defined by the aspects of a communities’ endeavour that are under governance
- the who – the key roles and relationships between stakeholders and the collective organisational structures through which governance is exercised
- the how – the mechanisms and processes of the human activity system through which governance operates and technical tools that support governance

3.1. Service Oriented Architecture (SOA) Governance

SWFS, is built around (local and web) services and thus many of the governance challenges and potential solutions associated with service oriented architecture (SOA) are relevant to SWF environment governance. The SOA approach is premised on the development, maintenance, discovery and use of interoperable services. These self-contained functional elements are designed to meet specific purposes and are able to interoperate. The publish, find, bind pattern, shown in Figure 1, provides the mechanism for the publication, discovery and use of services in a SOA.

![Figure 1. The publish, find, bind pattern](image)

Services, although under the control of different owners, are interdependent, necessitating collaboration between owners, developers, operators, and users of the service across departmental and organisational boundaries (Josuttis [2007]). In addition to the technical tools for service management, a governance framework is also required to ensure consistency, and predictability of interdependent services (Stanek [2006]). SOA governance provides the business context for the design, development and operation of services and addresses related aspects of the service lifecycle; design-time and run-time. Design-time governance relates to the environment in which services and other components are designed, developed, tested and approved for publication. Run-time governance addresses the governance of operational aspects of SOA including service discovery, access monitoring, security and management.

In SOA, registers (or lists) of resources and registries (the systems used to manage them) play a vital role in publication, discovery and use of community resources. A range of registers of such things as code, users and permission, standards, and other resources that the community care about, are essential to the sustained operation and growth of an initiative. These artefacts document community agreements and enable the discovery and use of the resources necessary to develop, maintain, operate and grow the infrastructure. Thus registries and the registers they manage play a critical role in supporting governance.

3.2 HWB Governance Challenges

Efforts to extend the HWB through the development of activities in Trident and tools in ArcGIS and compose workflows to meet the specific requirements of the hydrology domain, have informed an understanding of key challenges that require resolution through governance. These challenges which relate to functionality, data, people and computation resources across both design-time and run-time domains are presented in Table 1. Run-time domain in this context includes the composition and execution of workflows in Trident and other workbenches.
Table 1. Governance design-time and run time issues

<table>
<thead>
<tr>
<th>Domain</th>
<th>Design-time</th>
<th>Run-time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Functionality</td>
<td>- identification, documentation and prioritisation of workflows to be developed&lt;br&gt;- identification and specification of functionality required for the hydrological domain&lt;br&gt;- development lifecycle management for Trident activities and ArcGIS tools&lt;br&gt;- specification and management of integration with other workbenches&lt;br&gt;- development lifecycle management for other development environments</td>
<td>- discovery, interpretation and reuse of workflows, workflow fragments and activities&lt;br&gt;- sharing and exchange of workflows and activities&lt;br&gt;- discovery, interpretation and use of functionality in other workbenches (e.g. ArcGIS)</td>
</tr>
<tr>
<td>Data</td>
<td>- agreement on common data types and exchange formats&lt;br&gt;- identifying provenance tracking and persistence requirements for data sources, intermediate data sets&lt;br&gt;- management of final products&lt;br&gt;- managing interactions with data providers and data sources under external governance arrangements</td>
<td>- Discovery, interpretation, and access to data sources</td>
</tr>
<tr>
<td>People</td>
<td>- assignment and managing of decision rights and roles in relation to:&lt;br&gt;- governance processes&lt;br&gt;- extending workbench functionality through the creation and registration of activities and tools&lt;br&gt;- design and composition of workflows&lt;br&gt;- managing interactions with external stakeholders and ‘communities’ to enable cross-community development and re-use of workflows and components</td>
<td>- management of permissions related to workflow and component access composition, reuse and execution</td>
</tr>
<tr>
<td>Computational resources</td>
<td>- assignment of workflows to computational resources for execution</td>
<td>- assignment of workflows to computational resources for execution</td>
</tr>
</tbody>
</table>

3.3. Run-time Governance Capabilities

Trident offers some capabilities that support governance. These capabilities are underpinned by a registry used to register, manage and access workflows, activities, data products and other artefacts that comprise or support workflows. The aspects of scientific workflows governance that are addressed through the Trident registry are:

- **functionality** - registration, discovery, management and use of activities that have been developed and the workflows into which they are composed
- **data** - registration of data sources and data provenance tracking
- **people** - management of user permissions to access registries, workflows, and activities
- **computation resources** - registration and management of computation resources for workflow execution

Likewise, ArcGIS through its geo-processing framework provides a number of governance capabilities that, as with Trident, address run-time governance needs. These tools support the registration, discovery management and use of tools. The ArcGIS ArcCatalog application also provides some tools that support data source registration and management.
Although Trident (and to a lesser extent, ArcGIS) support many required aspects of HWB run-time governance, significant aspects of a governance solution are missing. Firstly, the ‘human activity system’ comprising the overarching institutional and process framework that enable a community to work together is missing. This system sets standards and policies, creates processes, assigns roles and permissions and interacts with external and related governance mechanisms. Secondly, a range of mechanisms and tools required to support governance absent. These include:
- the design-time environment – the governance of the entire development lifecycle for functional components from requirements through publication to retirement
- run-time environment for HWB components not registered in Trident such as ArcGIS tools

Trident registry is only able to support governance of things that are registered in Trident. As many components (workflows, models and blocks on functionality) will be developed and stored outside of Trident there is a need to register and govern these artefacts.

4. A REGISTRY BASED FRAMEWORK FOR GOVERNANCE

To address the key governance challenges and provide missing governance capabilities, a governance framework is required. Key requirements for the framework are that it be:
- based on existing information infrastructure governance approaches
- practical, lightweight and commensurate with the resources under governance
- scalable and evolvable
- focus on the governance requirements of technical components that need to be managed for discovery, reuse and sustained interoperability
- consistent with the inbuilt registration capabilities of Trident and ArcGIS geoprocessing framework, supplemented with additional registries where necessary

It was determined that the conceptual model of governance encapsulated in the ISO 19135 standard Procedures for Registration of Geographic Items [2004] be used as the basis for HWB governance. The standard articulates the use of registers (lists) and registries (systems that manage lists) together with a defined roles and processes related to register creation and management. In this model, the registers define the scope of governance, the processes and roles describe how governance is exercised and by whom. The UML diagram shown in Figure 2, highlights the key roles and relationships related to register and registry ownership management and use.
Figure 2. ISO 19135 - Registration Roles

Using this approach, the things that a community cares about and must manage to ensure the achievement of collective goals, can be conceptualised as a number of registers - lists of things such as agreements and resources (people, data, and technology). Users are able to access registers and find information that enables them to create, access or use common components that together constitute the collective system.

At the core of this approach are two processes; the creation and the assignment of roles related to register management and the registration process. The registration process involves a number of roles that together implement governance. A register owner determines who has authority to make submissions (submitting organisation) to the register, to adjudicate submission requests (control body) and to manage the registers (register manager) and the registry systems used to manage them (registry manager).

The governance model implicit in this standard can be applied to collective endeavours that need to register, manage, discover and reuse common information artefacts that are critical to the coherence of collective effort. Using this approach, a governance regime can be developed through the creation and management of registers and the assignment of roles related to their management and use.

4.1 Model Application

This conceptual model was used to develop a framework to address technical aspects of HWB governance. ‘Technical governance’ deals with ‘technical’ artefacts that the community cares about and which must be governed. These artefacts either specify an agreement about how some aspect of a component will behave or are a component that is based on or implements an agreement. For each set of such artefacts to be governed, a register is created and registration roles identified in ISO19135 assigned with respect to register management and the registration process.

A number of registers were identified as being part of the HWB governance framework including: a data register, to log data used for development and testing; a code register, to manage code (under version control); an ArcGIS tool register to manage geo-processing tools and Trident activity, workflow and users register. Registry capabilities of a number of tools were used to implement the registers. For example the Trident registry was used to register activities, workflows and will be used to register users and execution nodes. ArcGIS was used for registering geo-processing tools. In addition version control software (subversion) and underlying repositories were used to register and manage code used as the basis for Trident activities and ArcGIS and other tools. The activity register evolved from an initial excel-based solution to a software development management tool (JIRA) which enabled detailed tracking of the entire development lifecycle of functionality.

5. CONCLUSIONS

The governance model presented in this paper provides a conceptual framework for a governance solution. The way in which the model is applied will to a certain extent be determined by the implementation environment and organisational and IT governance regimes within which the SWFS is used. The governance model is intended to provide an over-arching framework for the human activity system within which the system operates. Wherever possible, the inherent capabilities of SWFS and other interacting workbenches be used to manage the functionality, data and users to create an integrated governance solution. Where necessary the governance framework can be implemented using registry capabilities provided by a variety of software.

Experience in implementing this governance model within a small collaborative team for the HWB project has shown that establishing and effectively operating governance, entails a significant overhead cost. This investment may in some cases be difficult to justify as the benefits of collaboration are not evident until such times as reuse begins in earnest. In order
to realise the longer-term promise of increased efficiencies based upon the creation, sharing and reuse of activities and workflow fragments leading to faster scientific discovery, change in working practices are also required. Collaborating team members who may be used to working alone or in silos to meet their own needs, must move to more collaborative models in which interoperable pieces of processing, functionality and the workflows into which they are composed, are designed, developed and maintained in a manner that enables discovery and reuse. The complex interwoven socio-technical nature of information infrastructures within which SWF are embedded, and the behavioural aspects of communities are critical aspects of scientific workflow environments. An improved understanding of these phenomena will inform approaches to creating and sustaining successful collaborative SWF environments and thus warrant further investigation.

ACKNOWLEDGEMENTS

This work is part of Hydrologists’ Workbench project in the Water Information Research and Development Alliance between CSIRO’s Water for a Healthy Country Flagship and the Bureau of Meteorology.

REFERENCES


Cuddy, S., and P. Fitch, Hydrologists workbench: a hydrological domain workflow toolkit, [this conference], 2010.


DEWHA (Department of Environment, Water, Heritage and the Arts), the Water Act 2007, 2007


Perraud, J.-M., B. Qifeng, and D. Hehir, On the appropriate granularity of activities in a scientific workflow applied to an optimization problem, [this conference], 2010.


On the appropriate granularity of activities in a scientific workflow applied to an optimization problem

Perraud, Jean-Michel\textsuperscript{a}, Qifeng Bai\textsuperscript{a}, David Hehir\textsuperscript{a,b}
\textsuperscript{a}CSIRO Land and Water, Canberra, Australia
\textsuperscript{b}Australian National University, Canberra, Australia
email: jean-michel.perraud@csiro.au

Abstract: Scientific workflow management is an active area of research and development responding to an increase in the complexity of computational models, data analysis and data size. In the authors’ experience, the distinction between scientific workflow software (SWS) and modelling frameworks or toolsets is at best not clear in the minds of users or developers, at least in the hydrology domain where the interest in concept of scientific workflow appears rather recent. It is understandably tempting for some users to assume that these new software tools aim to replace their existing modelling tools, with varying expectation depending on their prior satisfaction. While it is arguably clear that SWS can play a role in improving practices for high-level orchestration and traceability of scientific workflows, we explore in this paper the granularity at which activities can be usefully defined. We describe a case study, the calibration of a model, wrapping the components of a modelling framework (TIME) from the Trident and Kepler SWS. The problem is decomposed in several workflows comprising activities of differing granularities. We assess each approach against a set of criteria such as runtime performance and flexibility, discuss the feasibility and trade-off. The main findings are the design benefits stemming from having to clearly identify separate activities in the process, and that the difficulty of decomposing the optimization problem into finer-grained activities increases most markedly when needing iterative control flow capabilities.

Keywords: Scientific workflow software; granularity; optimization

1. INTRODUCTION

Modelling environmental systems in a holistic manner leads to the integration of usually disparate software systems and manual tasks. Unsurprisingly, the resulting pipelines of modelling tasks and analyses reach a size and level of complexity that can be daunting compared to that of smaller components of the modelling system, and needing enhanced capabilities for the capture of provenance and metadata. Recent examples of such projects in Australia are the Sustainable Yields projects (http://www.csiro.au/partnerships/SYP.html). These projects usually highlight the need for tools to better handle long-lived workflows. These workflows capture modelling processes whose execution can span anything from hours to months, possibly needing execution on different computing environments over their execution lifespan (Shukla and Schmidt [2006]).

In the authors’ experience, scientists and modellers will tend to approach the task of handling long-lived workflow either by using ad-hoc batch files, shell scripts or other type of glue code as a means for tool integration and automation (McPhillips et al. [2009]), or by trying to somehow wrap heterogeneous software into their modelling software of predilection. Modelling software systems are primarily aimed at helping to represent a real-world system. This model-centric view of the modelling process may not be, in the authors’
opinion, the most conducive to capturing information on data and model provenance as first class entities in the software system. If considered, this tends to be in the form of tools for the management of scenario and configuration information that are model-specific, and not generic enough to be applied to arbitrary workflows.

Conversely, for some scientists and modellers newly introduced to scientific workflow software (SWS), everything can end up looking like a workflow. This seems to be particularly the case if they perceive shortcomings in their main modelling tools. Arguably, many actions referred to as "models" in the terminology of modelling software platforms are indeed pipelines of analysis potentially amenable to be usefully captured as a workflow. However, the value of representing fluxes and states of the physical entities modelled into a scientific workflow rather than more “traditional” modelling software is very debatable. The boundary between scientific workflow tool and modelling platform is somewhat ill-defined, quite possibly gradual and varying depending on the specific SWS used. There are consequently questions on which granularity is appropriate for activities in these SWS. In a modelling and analysis pipeline that needs orchestration at several levels of granularity ranging typically from high-level process batching down to the stepping through space and time of a simulation model, how appropriate are SWS to handle each of these levels of granularity?

1.1. Software

A sample of six fairly different workflow systems and their application to the scientific domain are reviewed in Curcin and Ghanem [2008]. For the present paper the authors consider one SWS listed in Curcin and Ghanem [2008], Kepler (https://kepler-project.org), and another SWS not listed in that publication, Trident (http://connect.microsoft.com/trident).

Kepler (Altintas et al. [2004]) is a Java tool based on the Ptolemy II system, which supports multiple models of computation based on the director/actor paradigm. This gives Kepler the potential to cover a large spectrum of granularities, ranging from the orchestration of high-level tasks to iterating over a temporal environmental simulation model. We consider in this paper the version 1.0 of Kepler. We may use the term ‘activity’ in this paper in lieu of the term ‘actor’ used in the Kepler terminology.

Trident (Microsoft [2009]) is a SWS tool released by Microsoft Research. It is based on Windows Workflow Foundation, currently using version 3 thereof (WF3), part of the .NET framework version 3.5. We consider in this paper version 1.0 of Trident.

We mainly use in this case study the SWS Trident and the TIME environmental modelling framework (Rahman et al. [2003]) to assess the practical feasibility of a calibration task as a workflow. TIME is a software development framework for creating, testing and delivering environmental simulation models. It includes support for the representation, management and visualisation of a variety of data types, as well as support for testing, integrating and calibrating simulation models.

2. CASE STUDY

We explore in this paper several approaches to perform the calibration of a simulation model, i.e. an optimization task. Some relevant contextual and more technically detailed information on this case study can be found in Perraud et al. [2009], while general considerations as to the use of SWS in the hydrology domain are in Guru et al. [2009]. The main activities of the conceptual workflow are depicted in the UML activity diagram in Figure 1. This modelling exercise is of particular interest to assess the granularity of a workflow as it has the following characteristics:

- The core of the system is a temporal simulation model, typically developed upon a modelling framework
- It is possible to decompose the overall workflow in three high-level steps, each in turn comprising several nested steps.
- The step of optimization is in essence a step “looping” over some sub-steps. Here the term ‘loop’ may actually cover serial or parallel evaluations, or a mix thereof.

![Calibration Task Workflow](image)

**Figure 1.** Calibration task conceptual workflow

There are several nested levels of granularities at which the sequencing of the steps in Figure 1 can be expressed usefully via a workflow. Starting from the high level granularity:

1. The entire task can be expressed as a single activity. While this may look at the first consideration as a “degenerate” case, this has value when used as part of an iteration over several differing models, for instance when having to calibrate several water catchments, or comparing the performance of alternate model structures on one or more catchments.

2. The pipeline of defining the model (the system), defining the optimization problem, and performing the optimization. The high level conceptual steps follow those in Talbi [2009], where more information can be found on optimization problems and related software frameworks.

3. Decomposing in the SWS the activity named ‘Optimize’ in Figure 1. There are a couple of possibilities at this level. The first is to offer in the workflow the possibility of composing a sequential pipeline of several optimization algorithms, as is quite common for instance to start by using a global search algorithm followed by a local search that uses more or less the gradient of the response curve. Pictorially in Figure 1 this corresponds to working on the step of “Optimization Strategy”. The second possibility is to express the optimization algorithm itself in the SWS, and corresponds to the aggregate of the “Strategy” and the nested “Optimize” loop. This is arguably the most interesting part to work on as this puts to the test SWS in terms of expressiveness (can the algorithm(s) be expressed by the system), control flow (serial loops and/or opportunities for
parallelization e.g. for population based strategies) and data structure (can the inner information in the optimization process be handled in the SWS).

4. The finest granularity we may consider for explicit handling by the SWS is the “Run Model” activity in Figure 1. This takes over the time stepping, handling of input and output and in essence removes the need for a third party modelling framework.

3. **CRITERIA**

Each approach is primarily assessed against the criteria listed in Table 1. It should be noted that these criteria are chosen for the specific calibration case study, and are not a comprehensive list. McPhillips et al. [2009] for instance give a different list for general workflows.

**Table 1 Main criteria considered for assessing the workflows**

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practical feasibility</td>
<td>Are there technical or conceptual “show stoppers”? Note that an assessment on this criterion is somewhat dependant on the assessment of the others.</td>
</tr>
<tr>
<td>Cost of wrapping</td>
<td>The code required to include the library functionalities should be minimal</td>
</tr>
<tr>
<td>Versatility</td>
<td>The workflow can be configured such that most likely options in the case study are covered (for instance, applicable to several model systems)</td>
</tr>
<tr>
<td>Extensibility</td>
<td>The core workflow can be extended with other activities.</td>
</tr>
<tr>
<td>Clarity</td>
<td>The purpose and expected behaviour of the workflow should be conveyed by the visual appearance</td>
</tr>
<tr>
<td>Performance</td>
<td>The runtime is within a maximum of ~200% of the runtime obtained when using directly the tools and modelling components in TIME.</td>
</tr>
</tbody>
</table>

4. **IMPLEMENTATIONS**

Figure 2. Implementations in Trident of two of the approaches, namely (a) for approach (2) and (b) for approach (3), in the second case only the simplest decomposition possible for the iterative loop.

Figure 2 is a visual screen capture of two of the four solutions when implemented in Trident. The four approaches considered follow. Note that these are not necessarily mutually exclusive approaches. These approaches will be referred to in the rest of this paper as numbered thereafter:
1. Having a single activity for the whole calibration task
2. The high level calibration task is decomposed in three steps (model definition, optimization definition, and optimization)
3. The iterative process of the calibration algorithm is explicitly handled in the workflow
4. The time stepping of the model is handled in the workflow.

5. RESULTS

A summary of the qualitative assessment of each approach is listed in Table 2. Overall the approach (2) appears the best level of decomposition, allowing some degree of visual composition and extension for the user. Approach (3) is found to be feasible but difficult in both Kepler and Trident, more due to the respective fundamental design choices of their workflow engines than simply idiosyncrasies. It remains an appealing approach as it has the prospect to allow users to compose their pipeline of optimization with the greatest level of flexibility, something than in our personal experience many users not keen to program are asking for.

The remainder of this section discusses the salient points of the assessment of each approach that lead to the summary assessment in Table 2, and elaborates on some of the related technical characteristics of the SWS.

Table 2. Qualitative assessment of implementation approaches (+++ ideal, --- not acceptable, 0 neutral, ? not assessed)

<table>
<thead>
<tr>
<th></th>
<th>1 Single activity</th>
<th>2 Model and optimization problem builders</th>
<th>3 Optimization algorithm in the SWS</th>
<th>4 Model iterations in the SWS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practical Feasibility</td>
<td>Yes</td>
<td>Yes</td>
<td>Difficult</td>
<td>No. (marginal in Kepler)</td>
</tr>
<tr>
<td>Cost of wrapping</td>
<td>+++</td>
<td>++</td>
<td>--</td>
<td>---</td>
</tr>
<tr>
<td>Versatility</td>
<td>0</td>
<td>+</td>
<td>+++</td>
<td>?</td>
</tr>
<tr>
<td>Extensibility</td>
<td>0</td>
<td>+</td>
<td>++</td>
<td>?</td>
</tr>
<tr>
<td>Clarity</td>
<td>-</td>
<td>++</td>
<td>+</td>
<td>?</td>
</tr>
<tr>
<td>Performance</td>
<td>+++</td>
<td>++</td>
<td>-</td>
<td>---</td>
</tr>
</tbody>
</table>

The implementation at the finer granularity (4) is assessed as not practically feasible, using either Trident or Kepler. Simple tests show that both platforms exhibit a prohibitive overhead in data exchange between activities. In Trident, the model of execution for activities implies that a new activity instance must be created at each time step. In effect each conceptual activity is stateless by default. It is feasible to maintain and restore the state of an activity (or part thereof) between iterative steps but the overhead is important. As the execution of the model is usually at least the significant majority of the runtime, the solution drastically fails on the performance criteria. This result is hardly a surprise given the performance optimization required even without the overhead of a SWS, as described in Perraud et al. [2009] which showed the significant performance tuning steps and architectural changes required for enabling multi-threading to speed up calibration.

The highest granularity solution (1) is feasible on both platforms with rather little wrapping ‘glue’ code, although a fair bit more for Kepler than for Trident, due to the difference of virtual machine with TIME, but this is a small concern for this paper. The information of the input and outputs of the activity, being high level, can be captured with simple types that both workflow software tools can handle ‘natively’.

The approach (2) brings some interesting design questions and considerations to light when seeking implementation through both Kepler and Trident. This decomposition leads to the
necessity of passing between workflow activities the definition of the model (the system being optimised) and of the optimization task itself (conceptually a superset of the former). The type system of Kepler would require translating this information to simpler types, passing it around as ‘anonymous’ objects, or declaring additional specific types to Kepler. While all appear possible, the first approach is model specific, the second not ideal in terms of clarity, and the third is relatively costly in terms of code. Trident uses the standard .NET type system for the activity inputs/outputs and in this respect is much more accommodating to defining custom types to pass between activities.

The approach (3) is the first one that brings an explicit looping structure in the solution workflow (Figure 2-b). This approach definitely brings to light some behaviours, capabilities and limitations of both Kepler and Trident. Thus, this is the approach we discuss most in the present section.

In Kepler, the models of computation can be regarded as a framework for component-based design, where the framework defines the interaction mechanism between the components. The scheduling of execution is a global decision for a given workflow execution (Brooks et al. [2005]). The simplest and arguably most used director is the synchronous data flow director (SDF). Emulating a “for” loop in Kepler can also be achieved using the “PN” (process networks) director. It is of particular interest as this director is potentially useful for parallel and distributed processing. It is also appropriate for iterations where there are feedback loops between actors, which the SDF director could not handle. One problem we have identified is cases with nested loops using PN directors. When a step in the process fires a ‘Stop’ signal, it is not limited to the current PN domain (current iteration) and is propagated to outer loops, which means it will stop the outer iteration. While simple loops can be implemented with relative ease, the feasibility of nested control flows is uncertain at the time of writing, at least if approaching the task from a point of view of a traditional, imperative programming. Another aspect already noticed in approach (2) implemented in Kepler is the data system. Expressing the optimization process at a finer granularity requires further constructs (parameter sets, population of points, objective function calculators, etc.) to be handled. They need to be re-cast into Kepler’s native data types, passed as anonymous objects, or added as first class data constructs. The undeniable benefits in terms of data polymorphism of the Kepler data types have a clear downside when wanting to handle domain-specific constructs.

If implemented in Trident, the simplest case for decomposing the optimization step (approach 3) can use a ‘for’ loop, as shown in Figure 2-b. The workflow composition and its runtime behaviour are somewhat intuitive, although a large amount of execution logic is still embedded within each activity executed. One particular characteristic of Windows Workflow Foundation version 3 (WF3) quickly becomes apparent when implementing the activity that compares objective scores between iterations. WF3 activities, or activity instances to put it more correctly, follow an activity automaton, typically the three states Initialized, Executing and Closed (Shukla and Schmidt [2006]). An instance cannot get back into an executing state when closed. At every step of the iteration, a new activity instance is created. Retaining the memory of the “best known parameter set so far” in the score comparer in Figure 2-b is not straightforward, as opposed to when iterating using in-memory stateful objects as is the paradigm with TIME components and most object based systems. The good side of this in WF3 is that the interactions between activities are primarily and natively dictated by the definition of bindings between the inputs and outputs of these activities, rather than relying on the state of activities from their last execution. This arguably fosters the clarity of what a workflow is doing, but this is a significant departure from traditional programming or scripting, aside perhaps from stateless transactions such as those found in Web-based systems.

6. DISCUSSION

One outcome of the exercise the authors found interesting is the incentives for certain design felt when using workflow software. Chief among these is the lead towards using something akin to the Strategy design pattern (Gamma et al. [2004]), and decoupling the
steps in the workflow. The main constructs in the process are arguably the model
definition, the optimization problem definition including the objective function, and the
parameter sets trialed during the optimization. Given the execution model of the workflow
software, even the simplest decomposition of the case study leads to using these constructs
to transfer information between workflow activities. While the use of these constructs in an
imperative, object oriented programming context is also beneficial, implementation leakage
and implicit coupling between the different participating objects can occur more easily.
One downside of the data handling and execution model in Trident is the temptation to use
the Singleton pattern to maintain state in an iterative loop in a manner similar to imperative
programming. This would have negative consequences for parallelizing the iteration of the
optimization algorithm.

One enticing prospect of SWS for scientific workflows requiring iteration is the greater
emphasis on declarative programming over imperative programming, the ‘what’ over the
‘how’. The approach aims to be more conducive to write scalable workflows, easier to
parallelize (Chappell [2009]). It is worth noting that it comes at a time where there is also a
renewed interest in functional programming, of a similar philosophy, with recent functional
language additions both on the .NET and Java virtual machines. Declarative programming
does require a change in mindset, perhaps more so for users versed in programming.

Facilities to parallelize tasks are highly desirable for many heuristic optimization problems,
most of them being readily data-parallel with little exchange of messages between sub-
tasks. For both products used for this paper, we find limits to the usability of such facilities.
If and when underlying capabilities are present, they require a level of investigation that in
effect limits the availability to users with a significant level of programming skills.
Currently it appears more feasible to implement the parallelization of computing tasks
within an activity rather than composing it at the workflow level. This is admittedly an
assessment that may be biased by prior work described in Perraud et al [2009]. The design
and implementation of simple, entry-level parallel workflow iterators for the simpler
parallel problems is a task the authors are undertaking as a follow-up to this case study. We
anticipate that the granularity of parallelization that can be expressed usefully through
workflows is larger than that described in Perraud et al. [2009], which is located within the
model. Rather, an evolutionary optimization algorithm would benefit from workflow
iterators parallelizing the evaluation of model configurations i.e. different
parameterizations.

7. CONCLUSION

Optimization tasks are common in many domains, stemming from the generic nature of
most algorithms. The case study in this paper is chosen as it spans several granularities
ranging from the temporal iteration of the model to the high level workflow defining the
model, optimization and executing that optimization. We find that decomposing the
optimization problem as a workflow improves the expressiveness when compared to what
users have to do with scripting engines or programming directly to the application
programming interface of optimization toolboxes and modelling frameworks. Four
approaches distinguished primarily by the level of granularity of the decomposition are
assessed. The finest level of granularity is not practically feasible due mostly to
performance degradation. The intermediate approaches that decompose respectively the
overall task and the optimization process itself have positive incentives. These stem from
the workflow software, and the ability to design sensible constructs defining the temporal
model and the optimization task to decouple steps in the workflow. We find that the nested
control flows necessary to express the optimization algorithms in use in the hydrology
domain are difficult to implement by graphical means in both SWS, due to fundamental
characteristics of the software, albeit different ones in each case. While it has some clear
positive aspects elsewhere, the lack of native maintenance of state values between
iterations in Trident is the main logistical hurdle. In Kepler the main issue is the apparent
complications of using the PN director to emulate iterative control flows, notably nested
control flows. Notwithstanding these limitations both workflow software have the potential
due to design to support iterative control flow structures, notably to better support the
parallelization of portions of workflows. We suggest that both platforms could benefit from better or at least more readily usable iterative control flow facilities to express, among other things, optimization algorithms as workflows.

ACKNOWLEDGMENTS

This work is part of Hydrologists’ Workbench project in the water information research and development alliance between CSIRO’s Water for a Healthy Country Flagship and the Bureau of Meteorology.

We thank Paul Box, Joel M. Rahman, and two anonymous reviewers for their valued contributions.

We gratefully acknowledge the help from the Trident team at Microsoft Research and from contributors to the Kepler users’ mailing list.

REFERENCES


Curcin V. and M. Ghanem, Scientific workflow systems – can one size fit all?, Proceedings of the 2008 IEEE, CIBEC’08, 2008

Gamma, E., R. Helm, R. Johnson, and J. Vlissides, Design Patterns: elements of reusable object oriented software, Addison Wesley, 1994.


Opportunities and limitations of DelftFEWS as a scientific workflow tool for environmental modelling

Peter Gijsbers
Deltares USA Inc., Silver spring, Maryland, USA, Peter.Gijsbers@deltares-usa.us

Abstract: DelftFEWS is a software platform for real-time operational forecasting systems in the hydrological, hydraulic and water quality domain. Operational water management agencies around the world deploy over 40 simulation models in manual and automated time series processing workflows to produce real-time forecasts information with DelftFEWS. DelftFEWS has also been used outside the scope of real-time applications in studies requiring execution of workflows composed of long term simulations and statistical post-processing tasks. DelftFEWS has not specifically been designed for this purpose and does not intend to be a generic scientific workflow tool similar to e.g. Kepler. However, it offers features which make it, under some conditions, a useful production environment for scientific workflows where a chain of time series are being manipulated by generic (spatial) time series handling operations and dynamic simulation models. Data that cannot be represented as a (spatial) time series cannot be properly accommodated by the system. DelftFEWS is not a suitable environment for discovery types of workflows as its architecture, workflow composition process and its graphical user interface do not offer suitable facilities for the user.

Keywords: scientific workflows; DelftFEWS; environmental modelling

1. INTRODUCTION

1.1 DelftFEWS

DelftFEWS is a state-of-the-art, real time software infrastructure for operational water management and forecasting [Werner et al. 2006]. The system is a sophisticated collection of modules, which can be chained in workflows to build an operational water management system customised to the specific requirements of an individual agency. The philosophy of the system is to provide an open shell system for managing the operational management process [Werner et al. 2004]. The shell system offers workflow based data handling capability, which can combine a comprehensive set of general data transformation functions with an open interface to external models. Over the past 7 years, more than 40 different simulation models of different suppliers have been made operational in the field of surface water and groundwater hydrology, water quality and inland and coastal hydraulics. The modular and highly configurable nature of the system allows it to be used effectively both in rudimentary systems and in highly complex systems utilising several simulation models. DelftFEWS can either be deployed in a manually driven stand-alone mode, as a fully automated client-server application, or as a web-service component.

1.2 Applications

DelftFEWS is mostly applied for real-time forecasting purposes. The overall workflow structure for such forecasting application is composed of a set of workflows, each composed of one or more sub-workflows or activities (see Figure 1). The workflows are
typically executed independently, with different time intervals. Fall back mechanisms can be applied to adjust the strategy in case one or more activities fails to provide data.

![Diagram of DelftFEWS workflow]

**Figure 1.** Typical DelftFEWS (forecasting) workflow

In addition to its operational use by forecasting agencies around the world, DelftFEWS is also being used for several research and management studies in the water domain.

One of the most interesting types of DelftFEWS applications from the scientific workflow context is the application in the GRADE project (Generator of Rainfall And Discharge Extremes (GRADE) for the Rhine and Meuse basins). This project [Wit et al. 2007] utilized DelftFEWS to conduct extreme long simulation runs of hydrological runoff and routing to assess flooding conditions using generated rainfall series of 10,000 years length. A workflow was established using the HBV model for runoff computation and hydrological routing [Lindström et al. 1997]. Whenever a peak event was detected in the computed time series, the workflow engine adapts its execution strategy and assigned the flood routing computation activity to a more precise - and computation intensive – hydraulic routing model. Simulations have been conducted with a stand-alone application.

Another interesting scientific workflow application using DelftFEWS is the National Groundwater Modelling System (NGMS), implemented for the Environment Agency of England and Wales. DelftFEWS provides the software platform to conduct impact assessment of groundwater abstractions for policy and planning purposes using numerical groundwater models, such as Modflow, and recharge models [Farrell et al. 2008]. Within this application four default water abstraction scenarios are identified (Historic, Naturalized, Recent Actual, Fully Licensed) to reflect the past and current water management situation. For each scenario, a ‘default’ workflow has been defined to conduct a groundwater simulation run, computed the differences with the other scenarios and calculate statistical aggregates such as mean monthly values. Three additional workflows have been formulated to accommodate what-if scenarios, either derived from the Historic, the Recent Actual or the Fully Licensed situation. For each what-if scenario, the difference with its ‘default’ scenario is calculated, including associated monthly statistics. Within this client-server system, workflow runtimes can vary between 20 minutes and 20 hours depending on the size of the associated groundwater model.
1.3 Scientific Workflows

Scientific workflows combine data and processes into a configurable, structured set of steps that can be implemented in semi-automated computational solutions to address scientific problems. Scientific workflows have become an increasingly important paradigm to accelerate scientific research, as it allows scientific experiments to be conducted through massive computation instead of labour intensive laboratory work. This experimental research - called “in silico” by Woollard et al. [2008] – can be characterized by three phases, a classification that also can be applied to scientific workflow environments:

- discovery: rapid investigation of a scientific principle in which hypotheses are formed, tested, and iterated on rapidly,
- production: the application of a newly formed scientific principle to large data sets for further validation
- distribution: sharing of data results for vetting by the larger scientific community.

While DelftFEWS has been designed for real-time application in the water domain, its functionality may be suitable for some environmental scientific workflows. Using the general (idealized) requirements of Ludäscher et al. [2006] and the categorization of Woollard et al. [2008] as a guidance for discussion, this paper will illustrate that the value of DelftFEWS as a scientific workflow system is primarily in the production phase for time series simulation based research.

2. CAPABILITIES TO MEET SCIENTIFIC WORKFLOW NEEDS

2.1 Access to Data, Resources and Services

Workflows need input data, components that can do the work and resources where the work can be executed. The data ingest into DelftFEWs is a pull-based activity, where an Import workflow reads available files at a directory or FTP-site, or queries a web-service. DelftFEWS supports various common standards and a large number of proprietary formats for the import of scalar and grid time series [Gijsbers et al. 2008]. The supported global standards include Grib, NetCDF (CF-convention) and the DelftFEWS Published Interface. Unfortunately, the hydrological world does not have a global standard for time series data yet. When such standard emerges, it is likely to be supported by DelftFEWS. Services and access to computational resources will be discussed in the next two sections.

2.2 Service Composition and Reuse and Workflow Design

Scientific workflows can be seen as a chain of data handling services that are executed in a particular fashion, e.g. sequentially, parallel, event driven or time step based. In many scientific workflow systems, web-services and Grid computing are the resources that need to be combined and orchestrated. DelftFEWS is more traditional in its architecture, as the services used in its workflows are internal modules, typically executed sequentially inside the workflow engine. Modules themselves may be able to run multiple processing tasks in parallel. The most important DelftFEWS modules are the data transformation module and the general adapter module. Workflows call instances of these modules in a particular sequence, while data flows from one module instance to the next via the input and output time series defined. DelftFEWS workflows are typically configured by hand and/or by scripts which takes the knowledge from another data source. Because of this nature, DelftFEWS is not a suitable candidate workflow environment for discovery type of work.

Data transformations provide the ‘plumbing’ to transform data in the proper format for the models, called the actors by Ludäscher et al. [2006]. The spatial-temporal and data handling transformations offered by this DelftFEWS module are of a high granular level. The available functions are primarily aimed at preparation of a consistent and complete set of model inputs, e.g. time step harmonization, (dis)aggregation, spatial and temporal gap filling. To accommodate post-processing a large set of functions are available within DelftFEWS to conduct time series and ensemble statistics. Where required, these
transformations can be made conditional, either by time (e.g. do transformation m during the summer and transformation n during the winter season) or by time series value (e.g. if parameter value at location A > 3 do transformation n else do transformation m).

The DelftFEWS General Adapter is the module which facilitates the execution of a wrapped external model by way of communication via the DelftFEWS Published Interface (PI) data exchange format [Deltares 2010]. As described in Gijsbers et al. [2008], the General Adapter exports model data sets (schematizations) and model states in native format, while time series are exchanged in PI-XML or NetCDF (grids). A model specific adapter converts this data into the native format for the model. After model execution, the results are converted back by the model adapter into the PI-format for ingest by the General Adapter. Using this concept, all model specific knowledge is embedded in the model adapter, while DelftFEWS only manages time series and model states and data sets.

2.3 Scalability and Detached Execution

Dependent on the kind of workflow, large volumes of data may need to be processed in long running workflows, requiring high-end computational resources, running in the background (so-called detached execution), possibly in distributed fashion. DelftFEWS offers scalability in both stand-alone and client-server applications. Workflows can be configured to run in deterministic (i.e. single trace) or ensemble mode. While most ensemble workflows are based on multiple input traces, some forecasting applications use multi-model ensembles by composing different model chains that execute with the same input. DelftFEWS offers parallelization capabilities to conduct an ensemble workflow on a multi-processor machine by parallel execution of different ensemble member traces.

Detached execution capabilities, i.e. running jobs in the background, are available in the client-server setup of DelftFEWS. Client applications can send a workflow for execution to the Master Controller, which adds the workflow to a job queue and assigns the job to a logical instance of the DelftFEWS engine (a ‘shell server’), when it becomes idle [Gijsbers et al. 2008]. Typically, such system layout involves multiple logical shell servers, allowing different workflows to be conducted in the background at the same time. Ensemble workflows may be deployed server side in parallel fashion using a Condor Grid.

2.4 Reliability and Fault-Tolerance

Reliability is very important for an operational task such as flood forecasting. DelftFEWS therefore offers mechanisms, both at the workflow level and module level, to ensure that results can be generated even when the preferred data sources or modelling services are unavailable. The strategy towards a reliable workflow execution starts with mechanisms to prevent model failure due to missing input data. The data transformation module offers the following gap filling functions for this purpose:

- Data hierarchy: a data merge function selects the preferred data source if available, while offering an alternative source if the preferred source is missing.
- Interpolation: a variety of spatial and temporal interpolation functions are available to fill data gaps.

The second defence mechanism against workflow execution failure is the ability to specify a fall back activity for each activity (nested workflow, internal module or external model) in the workflow. This fall back activity is executed in case the primary activity within the workflow fails. Finally, workflow results may trigger other workflows to be kicked off.

2.5 User Interaction

Different types of workflow applications require different types of user interaction. Discovery workflow systems require a user interface which accommodates composition of workflows, while production and distribution workflows systems require user interfaces that accommodate traceability and insight.
DelftFEWS is very much oriented towards production type of workflows. Initially its focus was on automated scheduling of forecast runs as this was the predominant paradigm of flood forecasting in Europe. In the USA, flood forecasters are working more interactively with the data, requiring new capabilities to accommodate user interaction during workflow execution. The so-called modifiers concept has been implemented and the granularity of the workflow steps has been adapted such that the user has the ability to interact and modify the data before each processing step. The graphical user interface has been adjusted to accommodate adjustment of time series and parameter values and switching between input series. After each modification, the user needs to re-run the sub-workflow in which the change has been conducted.

DelftFEWS offers no graphical user interface capabilities to discover new models or compose new workflows. Insight in workflows is provided via the so-called Workflow Navigator (Figure 2). The tool displays the workflow structure in a tree view, allowing the user to drill down to the (computed) time series, model parameters and configuration of each activity. Configuration mistakes are highlighted by a red cross icon, supported with a message holding more information on the error.

2.6 Smart Reruns

Modification of intermediate data will require rerun of a portion of the workflow. Which portion of a workflow needs to be re-run depends on the topology of the sub-workflows. Only sub-workflows downstream of the modified data will be re-run. An example is given using the workflows of Figure 2, assuming that CHPM4, MICM4, and REPM4 are nodes in the downstream order of the topology. An input time series change to a model in the MICM4_Flow_Forecast workflow will require a rerun of MICM4_Flow_Forecast and REPM4_Flow_Forecast. Workflow CHPM4_Flow_Forecast will not need to be rerun.

In addition, DelftFEWS offers a decision structure in which the workflow can decide to conduct a portion of the run with a more detailed and more computation intensive module.
This facility has been utilized in the GRADE application to run a computation intensive hydraulic model when a pre-defined threshold (a flood stage) was exceeded in the more simple model. The workflow can also decide to conduct a re-run using another module for a specific time window around the maximum value of a given period. For example, within the GRADE project, the workflow scanned for each simulation year the data computed by the simple routing model, detected the annual flood peak and deployed the hydraulic model for a period of a month around the occurrence of the peak. Such smart decision mechanism prevents the need for computation intensive runs if less computer intensive jobs can indicate that no thresholds are passed or peak events occur.

2.7 Smart Semantic Links

Ludäscher et al. [2006] indicates that a scientific workflow system should assist workflow design and data binding phases by suggesting which actor components (i.e. external models) might possibly fit together. DelftFEWS is not helpful for this discovery type of application. It offers no semantic facilities, nor any workflow configuration editing capabilities. All workflow configuration needs to be conducted with external tools (e.g. XML-editors and scripts). The only assistance offered is primary (xsd) and secondary (overall content consistency) validation and error visualization in the workflow navigator.

2.8 Data Provenance (Reproducibility)

“In silico” research experiments with scientific workflow system should be reproducible. Ludäscher et al. [2006] indicate that a scientific workflow system should be able to automatically log the sequence of applied steps, parameter settings and (persistent identifiers of) intermediate data products. With legal liability issues looming at each flood event, reproducibility of forecast runs is of major importance. The DelftFEWS database keeps track of each piece of data. It registers the producing data source (module instance), the producing task run (workflow), and the date-time when the data is written to the database. All information is included when archiving the run. For complete reproduction, the associated configuration and the binaries used can be archived as well.

2.9 Distribution

While data distribution is not a primary purpose of DelftFEWS, its data synchronization capabilities between the Master Controller and client applications allows users to get easy, automated or on-demand, access to workflow results. Better access to models and model results was the main driver behind the National Groundwater Modelling System.

3. LIMITATIONS TO MEET SCIENTIFIC WORKFLOW NEEDS

3.1 Data Types

DelftFEWS is tailored to deal with transient data of coarse granularity, i.e. regular and irregular time series in scalar, longitudinal or grid structure. All time series are location bound, and are either geo-referenced or bound to a dummy location placeholder. Static data can only be handled as part of the configuration. While powerful if time series are the key data concept in the scientific workflow, DelftFEWS should not be used as a scientific workflow application if data connectivity is not achieved through passing time series.

However, even if time series are the major carrier of information exchanged, attention needs to be paid to the semantics and multiple characteristics of time in the forecasting domain. The forecasting domain is characterized by the fact that the clock moves forward in time. Typically each forecast can be assigned to a specific start time of the run and an offset to the actual forecast time (e.g. the forecast of 12Z UTC, may have models starting 6
hours earlier). This concept underlies the different types of time series distinguished within DelftFEWS. The first categorization is by data source:

- External: ingested from an external source
- Simulated: internally generated (computed).

The second categorization is by their relation to time:

- historical time series: continuous in time, i.e. one parameter value per timestamp
- forecasting time series: characterised by the base time (start time) of the forecast, i.e. allowing multiple runs and multiple parameter values per timestamp.

### 3.2 Run Management and Run Comparison

When applying DelftFEWS outside the forecasting domain, careful attention needs to be given to proper semantic understanding of the time series types and valid ways of application by the modules in a specific workflow. The following section explains which time series types can be used under which circumstance to enable a certain type of data comparison.

Data generated for the same time stamp can be stored in historical or forecasting time series types. The historical time series types can be used to create a baseline run, i.e. a run which provides the basis for comparison with other runs. The forecasting time series types can be used in workflows that accommodate what-if scenarios. The time series generated can be compared against historical times series by using the module reference that generated the historical time series.

Forecasting time series from module A can be compared against forecasting time series from module B within the same run. Comparison against a series generated by module B in a previous run is only feasible if the time series from module B is generated in another workflow. A proper semantic understanding of the ‘current’ forecast is important to conduct such comparison.

Within the forecasting domain, each forecast is basically identified by its base time (T0), i.e. the forecast of 12 o’clock. A forecast run needs to be approved to make its results ‘current’ hence making its results instantly available to system components. Data is by default only retrieved for the ‘current’ run, i.e. the last approved run of a specific workflow or the workflow run currently being calculated.

Workflows can be auto-approved at configuration time. Alternatively, a data management dialog is available to make another run ‘current’ or to add the results of other runs to the displays, already showing the current run. What-if scenarios specified for a workflow can be used as a key to identify a run in the data management dialog.

### 3.3 How to Make it Work in Practice?

The National Groundwater Modelling System illustrates how a set of scientific workflows can be composed to accommodate run comparisons against a baseline. Table 1 illustrates the time series types used for the different types of workflows.

<table>
<thead>
<tr>
<th>Workflow Type</th>
<th>Time series type</th>
<th>Compared against</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default Historical</td>
<td>Baseline</td>
<td>simulated historical</td>
</tr>
<tr>
<td>Default Naturalized</td>
<td>Baseline</td>
<td>simulated historical</td>
</tr>
<tr>
<td>Default Recent Actual</td>
<td>Baseline</td>
<td>simulated historical</td>
</tr>
<tr>
<td>Default Fully Licensed</td>
<td>Baseline</td>
<td>simulated historical</td>
</tr>
<tr>
<td>Modified Historical</td>
<td>What if</td>
<td>simulated forecasting</td>
</tr>
<tr>
<td>Modified Recent Actual</td>
<td>What if</td>
<td>simulated forecasting</td>
</tr>
<tr>
<td>Modified Fully Licensed</td>
<td>What if</td>
<td>simulated forecasting</td>
</tr>
</tbody>
</table>
As shown, a What-if scenario from a specific workflow cannot be compared against another what-if scenario conducted with the same workflow. This is caused by the fact that a forecasting workflow can only have one current forecast run. This may be an approved run, but as soon as a new calculation starts for this workflow, the run underway becomes the current one.

4. OPPORTUNITIES FOR SCIENTIFIC WORKFLOW APPLICATION

Keeping the semantic limitations in mind, DelftFEWS can successfully be applied as a scientific workflow environment for production workflows in the field of time series based simulation modelling and forecasting. Within this context, it can also distribute the results to a users who is connected to the client-server system. The workflow structure, modules and data type semantics offer a high level of granularity, which opens the door for generation of DelftFEWS production workflows from discovery workflow applications.

5. CONCLUSIONS

Although DelftFEWS is not designed as a generic scientific workflow system, it can under certain conditions be used for workflows which Woollard et al. [2008] categorize as production workflows. Once the workflows are designed, DelftFEWS can provide the scientific production environment for time series data processing and simulation modelling. However, its primary application purpose, forecasting, has resulted in a data model and database design that requires extra attention when designing the workflow and time series handling concept for application in DelftFEWS. DelftFEWS may offer, in its client-server setting, a suitable distribution environment for workflow results. However, DelftFEWS is not a suitable environment for discovery purposes as its architecture, workflow composition process and its graphical user interface do not offer appropriate user facilities.

REFERENCES

Semantic validation and correction of scientific workflows

David Ratcliffe, Geoffrey Squire, Michael Kearney
CSIRO ICT Centre, Canberra, Australia
firstname.lastname@csiro.au

Abstract: Scientific workflows describe steps for orchestrating the execution of a network of computational operators toward some goal, such as data transformation for analysis or visualization. Typically, these operators consume and emit transformed data, or cause some effect. In most scientific workflow systems, the operators are typed to enable compatibility checks for their composition that make up a workflow. However, type checking performed by most such systems today is still largely confined to syntactic checking with limited, if any, semantic type checking support. In this paper, we present a type system incorporating the W3C OWL ontology language to aid in representing the semantics of data and workflow operators. We show how this type system supports the detection of type incompatibility errors in workflow compositions, and how it facilitates a (semi-)automatic type correction procedure using type transformations. We have incorporated our solution into Kepler, enabling users to statically test the type-consistency of workflows typed using our type language, and demonstrate that inconsistent bindings between expressively typed operators can be automatically corrected via a procedure that seeks to compose adapter/shim functions, such as unit transformations, time series interpolations, or some other arbitrarily complex data transformations.

Keywords: workflow; semantics; type system; owl; adapter; shim.

1. INTRODUCTION

Scientific workflow technology is an emerging method used to create complex data processing procedures without the need to author much, if any, programming code. These systems are finding widespread use in diverse domains such as earth sciences, bioinformatics and astronomy. A wide range of these systems exist and are in use today, amongst the most popular being Kepler, Microsoft’s Trident, Taverna (Hull et al. 2006), Triana (Majithia et al. 2004) and NASA’s SciFlo (Wilson et al. 2005) which provide high-level graphical-based environments for the authorship of workflows.

Irrespective of the scientific workflow system used, fundamental similarities exist in the specification of data processing provided by these products. A scientific workflow describes the steps for orchestrating the execution of a linked network of potentially distributed computational operators toward some goal, such as the transformation of data for analysis or visualization. Such operators are often implemented to perform particular tasks over particular data. For this reason, composing heterogeneous operators in a workflow can be a challenging task if their interfaces are not directly compatible. In such a case, a workflow author is required to insert functions (known as adapters or shims; see Hull et al. 2004) that serve only to manipulate the data as output from one operator into a form appropriate for consumption by another operator with a different interface. The requirement to create adapters/shims in a workflow places a large burden on the workflow author, as they are required to cope with many data management issues at once, such as understanding the particular data formats involved, how to manipulate them appropriately, and how to specify a correct implementation; this is an error-prone process.
In our work, we seek to alleviate as much as possible of the data management issues surrounding the construction of adapters/shims by providing automated support for their generation. In order to achieve this, we are required to increase the level of expressiveness of data and operator types used in scientific workflow to a greater degree than that which is currently implemented by most systems, which is largely based on syntactic typing, with little to no support for semantic data and operator typing. Much work around the Kepler project (e.g., Bowers et al. 2005a) has recognized the limitations of such type systems, and various proposals have been put forward for more expressive type systems based on recognizing the syntax, structure and semantics of data and the operators that process them. In our work, we are attempting to extend the state of the art by generalizing the techniques presented in order to permit mediation directly between many fundamental types of data models, such as relational, nested, and multidimensional, and consider various syntaxes of each. We are also exploring the use of a language for describing abstract functions that can operate over such data, along with methods of describing how type information propagates through such functions by analysis of their specification and use in a workflow context.

We have designed and implemented a prototype system and have incorporated its use within Kepler. The system allows users to annotate the operators (actors) and their input and output port data types using an expressive type language which captures the syntax, structure and semantics of data. Using the expressions in this type language, we are able to perform automated tasks such as type consistency checking (i.e., testing if two types are compatible), which then allows us to detect type errors in bindings made between the output of an operator and the input of another operator. The detection of such erroneous bindings then provides us with a starting point for a search problem that seeks to rectify type inconsistencies by automatically inserting data transformation operators (adapters/shims), which are available to the workflow system and have also been described using our type language.

2. DATA TYPES

Most scientific workflow systems will recognize data types and provide type checking as a basic validation of a composition of operators. For example, the typed output of one function bound to the typed input of another in a workflow (as depicted in Figure 1) can be checked by a type system which compares the compatibility of the source data type (function output) with the target data type (function input). Commonly, the data types being compared are expressed in languages which are tightly bound to the syntax of the data, such as ‘the XML document with schema X’ or ‘a Java array of Integer objects’.

![Figure 1](image.png)

Figure 1. Function $f_1$ outputs data with syntactic type $X_1$, which is directly compatible to type $X_2$ as input of $f_2$.

The semantics of the data is typically not explicitly captured in types – instead, it is only implicit in the syntax, if at all. This can be a problem, as while types $X_1$ and $X_2$ may be compatible at a syntactic level, their meaning (semantics) may be incompatible. For example, $X_1$ and $X_2$ could be floating point numbers, but where $X_1$ represents values of speed in metres per second, and where $X_2$ represents temperature in degrees Celsius. A system that does not explicitly capture the semantics of a type cannot determine such semantic inconsistencies. In our work, we aim to increase the expressiveness of types in order to capture more about the structure and semantics of data, such that we can perform more sophisticated type checking. We consider three aspects to the data type problem – that of the surface structure, the underlying core structure, and semantics.

We refer to the syntax of data as being comprised of two aspects, that of surface structure which describes the way information is presented (e.g., XML), and that of core structure, which corresponds to the underlying arrangement of data (e.g., ordered set, nested
collection, record). We also consider transformational mappings between surface structures and core structures, which describe how to derive a core structure from a surface structure, and how to convert a core structure into a surface structure. We also consider the semantics of data, which we associate closely with the core structure of data, as the latter describes the underlying content of data instances in a more abstract way which may permit a range of surface structural representations.

Our aim is to develop a type system that is capable of representing and comparing many common data types as encountered in a scientific workflow system (such as those captured with XML, CSV, relational data, object data, logic databases, multidimensional data such as NetCDF, etc.) at their core structural and semantic types, in order to recognize type compatibilities at these abstract levels. Detecting that a set of types have compatible core structure and semantics can potentially suggest automated transformations between various surface structures. The following sub-sections explore some concrete examples of these three distinct aspects of type.

2.1 Surface Structural Type Differences

Consider the case where we are attempting to locate a transformation from XML source data to CSV target data, where both encode data with a tabular structure. Moreover, the columns in each record of this tabular data have the same semantics (in that they capture equivalent data).

Example source instance (XML): Required target instance (CSV):

```xml
<timeseries>
  <m ts="2010-01-05T09:00:00Z">0.501</m>
  <m ts="2010-01-05T09:15:03Z">0.442</m>
  ...
</timeseries>
<ts>
  <day date="2010-01-05Z">
    <sample time="9:00:00Z" value="0.501"/>
    <sample time="9:15:03Z" value="0.442"/>
    ...
  </day>
</ts>
```

With an explicit model of the surface structural types (XML with a particular structure defined by an XML schema, and CSV with a particular grammar), core structural types (tabular data as a collection of records) and semantic types (defining the semantics of the collection, and records in tuples in the collection, e.g., a time series of rainfall measurements in inches, where records contain data, time and a magnitude value of rainfall in inches), we should be able to infer that a transformation between the XML and CSV based types is legitimate due to the equivalent core structural and semantic types, and possible if we have transformations from these core structural types to each surface structure (XML and CSV).

A procedure to transform the XML to CSV in this example would simply need to iterate over the underlying core structure in the source instance i.e., an ordered sequence of `m` elements and emit, with relevant alterations and in the order visited, the data in each `m` element as a record in the CSV syntax.

2.2 Core Structural Type Differences

In this example, while the semantic type (e.g., a time series of rainfall measurements made in inches) may be the same between source and target types, the underlying data described in each has a different core and surface structure. For example, consider the following two XML fragments.

Example source instance (XML): Required target instance (XML):

```xml
<timeseries>
  <m ts="2010-01-05T09:00:00Z">0.501</m>
  <m ts="2010-01-05T09:15:03Z">0.442</m>
  ...
</timeseries>
<ts>
  <day date="2010-01-05Z">
    <sample time="9:00:00Z" value="0.501"/>
    <sample time="9:15:03Z" value="0.442"/>
    ...
  </day>
</ts>
```
The source instance has a core structure of an ordered sequence of measurement items as flat records, while the target instance core structure is different, in that it nests measurement items into groups per day. While the latter is still an ordered sequence of items, the underlying core structure is different in that the items are not flat record structures as in the source instance, but instead contain ordered sequences themselves with measurements made at particular times per day. A procedure to transform data in the source type to data in the target type would require the construction of named groups which aggregate records per day. If this transformational operation (to group, or to flatten) is expressible over the language describing core structural types, then a mapping correspondence between the core structural types can be constructed. Since the semantic types are equivalent, the presence of such a mapping correspondence describing a grouping or flattening transformation will imply a transformation between the different surface structural types, if we have transformations from these core structural types to each surface structure.

Note that in this example, the semantic description of the target could have been tightened to reflect the grouped organization of data, in which case we would say there is also a semantic difference between the types. However, we consider this to be a core structural type aspect specifically defined to describe the arrangement of the data.

### 2.3 Semantic Type Differences

Continuing the time series example, consider the case where we are comparing two types that have exactly the same surface and core structure, as follows.

**Example instance 1 (XML):**

```xml
<timeseries>
  <m ts="2010-01-05T09:00:00Z">10.51</m>
  <m ts="2010-01-05T09:15:03Z">11.06</m>
  ...
</timeseries>
```

**Example instance 2 (XML):**

```xml
<timeseries>
  <m ts="2009-05-12T17:13:06Z">6.4</m>
  <m ts="2009-05-12T19:44:44Z">7.7</m>
  ...
</timeseries>
```

At a syntactic level, these two instances can both be valid under exactly the same XML Schema constraints. However, the first instance in this example encodes values of measurements of air temperature in degrees Celsius, whereas the second is encoding values of measurements of humidity in grams per kilogram. From a semantic perspective, these are instances of two completely different (disjoint) types, but this cannot be inferred from the instances themselves or the XML schema under which they are both valid. Without an explicit specification of the semantics of these types, a type system cannot automatically infer that they are incompatible.

### 2.4 Separation of Concerns

These examples show that by separating aspects of type into surface structure, core structure and semantics, we can identify type compatibilities through the inspection of each aspect. This provides a powerful capability to determine, for example, that two types are compatible despite their syntax (e.g., Section 2.1), or that they are incompatible even if they have the same syntax (e.g., Section 2.3).

### 2.5 Type Compatibility and Transformation

Figure 2 illustrates the aspects of type we wish to capture explicitly. Core structural types \( S_1^C \) and \( S_2^C \) have transformational mappings \( T_1 \) and \( T_2 \) to surface structural types \( S_1^S \) and \( S_2^S \), and also logical correspondence mappings \( M_1 \) and \( M_2 \) to semantic type descriptions \( O_1 \) and \( O_2 \).
We say that two types (Type 1 and Type 2) are directly compatible only if each of their respective surface structural, core structural and semantic type aspects are compatible. Furthermore, since we permit subtyping, Type 1 is only directly compatible with Type 2 if all type aspects of Type 1 are subtypes of the respective aspects of Type 2. There are certain conditions under which two types (Type 1 and Type 2) are indirectly compatible:

- **Surface structural differences.** If the core structural and semantic types are compatible, then the transformational mappings ($T_1$ and $T_2$) allow us to interpret surface structural types as being compatible. This case was exemplified in Section (2.1), and is depicted in Figure 3.

- **Core structural differences.** If the semantic types are directly compatible, and there exists a mapping $\mathcal{M}_C$ that defines the transformation between different core structural types, then the transformational mappings ($T_1$ and $T_2$) allow us to interpret surface structural types as being compatible. This case was exemplified in Section 2.2.

- **Semantic, core structural or surface structural differences.** If any or all of the type aspects are not directly compatible, then we can still consider the types to be indirectly compatible if there exists a directed type transformation function $f_T$ which has, as input, a type which is directly compatible with all aspects of Type 1 and has an output type which is directly compatible with all aspects of Type 2. This case includes the example in Section 2.3, where we may have some function available (e.g., a software model) to compute transformations between air temperature values and humidity values in a certain context.

Figure 1 depicts what we consider to be a weak kind of type compatibility checking, in that it does not consider the semantic types which could cause a type incompatibility despite compatible syntax. We also consider this type compatibility checking to be unnecessarily strict, in that when determining if two types are compatible, both the surface structure and core structure must be deemed compatible. This kind of type checking fails to recognize cases where core structure is compatible despite variations in surface syntax, so cannot indicate if a transformation from one surface structure to another could take place, as depicted in Figure 3.
3. TYPE AND MAPPING LANGUAGES

3.1 Data Types

To explicitly capture the aspects of surface structural, core structural and semantic types, we require languages for each type, as well as a language to express mapping correspondences between core structural types and semantic types. Additionally, we need an interpretation that maps expressions of core structural types to surface structural types to define their transformation.

In our system, we use a predicate logic language which is expressive enough to capture the core structure aspect of syntax for many common data structures, such as relational and nested data. We also use another language, specifically the logical language W3C OWL, to define semantic types. We also employ the use of a first-order predicate logic language called iMaPl (Cameron et al. 2005) to define mappings between core structural types to semantic types in OWL, in order to assign semantics to core structural types. Specific details of these languages and our use of them in our type system are out of scope for this paper, as our focus is to describe the capability we have achieved in applying the language to type incompatibility detection and correction problems in scientific workflows.

3.2 Function Types

In addition to modelling data types, we also model data processing functions. We describe aspects of these data processing functions such as their inputs and outputs, together with a description of their category using OWL (e.g., we can describe a function which takes as input time series data and certain parameters such as axis labels, and generates a plot as a graph image on the output). From these abstract specifications, we are able to map to ground instantiations of the specification (e.g., such as the WSDL web service, Java function, etc. that performs the function and which has the inputs and outputs as described in the abstract specification). Additionally, we also have a simple model of function signature specifications over particular types, describing how input data may be propagated to output data, in order to facilitate the propagation of specific subtype information on invocations of the function on typed data.

4. IMPLEMENTATION

4.1 Current Implementation

We have currently implemented a prototype of our type system using Java, SWI-Prolog, OWL-API and the Pellet OWL reasoner. In particular, we have implemented the following:

- Language for specifying core structural types and mappings to semantic types described in OWL, and implementations of transformations between particular core structural types to particular surface structural types, such as relational schemas, tabular data including CSV, and a one way transformation from certain surface structural types capturing nested data (such as XML or object data) onto core structural types (allowing us to query existing instances, but not to generate new instances).
- A semantic type checker based on Description Logic subsumption reasoning implemented using Pellet via the OWL-API.
- A limited core structural type checker based on exact structural matching.
- A limited abstract function specification language for describing the propagation of typed input data to typed output data.
- A limited semantic type propagation capability that pushes certain constructions of semantic types through expressions manipulating core structural types.
- A type transformation composition planner implementing a search algorithm.
We have integrated our solution into the Kepler workflow environment for testing, where we take advantage of Kepler’s ability to annotate workflow operators (Kepler actors) with type annotations, to which we attach specifications of our own type declarations (surface structural, core structural and semantic types).

When composing a workflow of Kepler actors which have been typed using our type language, our implementation of the underlying type system can detect binding errors (type inconsistencies between the output of one actor bound to the input of another with a directly incompatible type). The user then has the ability to send specifications of the workflow that produced errors to our type correction service, which seeks to determine if the two directly incompatible types are indirectly compatible via a type transformation procedure. Our underlying type correction service maintains an internal list of abstract operations and their grounded instantiations as Kepler actors. We have not implemented the function abstraction grounding feature as described in Section 3.2 in the Kepler environment, or the user selection of alternate options located by the type system for abstract grounding where alternate options exist (including alternate type transformation chains).

4.2 Future Work

We have not yet implemented the following:

- A language for specifying and interpreting bi-directional mappings between core structural types such as nested and multidimensional data and surface structures (e.g., XML, object data, NetCDF).
- Subtyping in the language for core structural types. We would like the ability to perform subsumption comparisons between core structural types.
- While we are capable of detecting type inconsistencies and resolving them by composing existing type transformation functions, we cannot yet automatically compute the transformations for converting between different yet indirectly compatible core structural types, such as aggregation (grouping) as described in Section 2.2.

5. RELATED WORK

The limitation of syntactic type checking alone in semantic workflow systems has been observed before. The most relevant and extensively published work on this subject surrounds the work related to the Kepler/Ptolemy project, as summarized in Berkley et al. [2005]. Our work closely reflects that of Bowers et al. [2004] where they describe a framework for semantic mediation, namely, the automated integration of workflow actors that operate over heterogeneous data. They identify core structural types of data and describe how to map such types onto a semantic model described by an OWL ontology (of a particular sub-language). When comparing two structurally incompatible but semantically compatible types, they seek to construct a transformation specification based on mappings between structural types and semantic types, as we also describe in Section 2.2. They describe an instantiation of their framework in terms of XML based structural types and how XQuery can be generated to transform between different XML structures which have been mapped to an ontology. In our work, we are attempting to further this capability by developing a language for core structural types that is abstract (and yet expressive) enough to capture many types of data, from nested to relational (and possibly also multidimensional), and transformational mappings from particular surface structures onto the abstract core structural types, in an effort to expand the applicability of such a system to one that can integrate directly across many combinations of heterogeneous data types.

Bowers et al. [2005a] present a hybrid type system that separates data structure from semantics, and links these two aspects using a first order predicate language (which they refer to as a hybridization constraint). In our work, we employ a similar scheme for mapping between core structural types and semantic types, in that we employ the use of iMaPl as defined by Cameron et al. [2005], which was designed for schema integration, in
a similar way as described by Bowers et al. [2004a] which describes mapping from relational predicates to an ontology schema.

The problem of propagating semantic type information through semantically typed actors in a workflow has been considered by Bowers et al. [2005] and Bowers et al. [2006]. In the latter of these works, algorithms for propagating semantic types over actors are presented, where the behaviour of the actors can described using relational queries (including selection, projection, product, union, etc). In our work, we are attempting to adopt similar strategies in order to propagate semantic types through operations that transform core structural types (e.g., nesting, or aggregation as described in Section 2.2). We recognize the importance of sound and complete propagation procedure which preserves semantic data descriptions when applied over functions with subsuming input and output types.

6. CONCLUSION

In this paper, we have described how an expressive type system that captures the semantic, surface structural and core structural aspects of type can be used to significantly improve the type checking capability of semantic workflow systems. In addition, we have implemented a procedure that takes advantage of the ability of the described type system to detect type inconsistencies, in order to perform type correction by composing existing type transformation functions. Our work extends earlier work in that we are achieving a mediation capability directly between data of different syntaxes, such as XML to CSV, or XML to relational. At present, our implementation currently stops short of generating type transformation functions on the fly (as described by Bowers et al. [2004] over XML types). We have shown how this capability can automate the potentially intensive and error-prone task of manually creating adapter/shim functions to mediate between incompatibly-typed workflow operators.

REFERENCES


Appendix A: Motivating Example

A.1 Workflow to Compute Wind Chill

In this use case, we show how a user may attempt to construct a workflow that helps to visualize a time series of wind chill values over time. The user has two data sources available, each an implementation of an Open Geospatial Consortium (OGC) Sensor Observation Service (SOS), where one produces air temperature measurement time series data, the other spatially co-located wind speed measurement time series data. The function available to compute wind chill does so by executing a function over every pair of air temperature and wind speed measurements made at the same time.

Figure 4 describes this incomplete workflow. The intention of the workflow author is to merge the SOS data appropriately for input to a function to compute a corresponding time series of wind chill values. In this deceptively simple example, there are several interoperability problems. At a semantic level, the wind chill function requires a time series of spatiotemporally co-located measurements of wind speed in kilometres per hour, and air temperature in degrees Celsius. However, the hourly air temperature measurements from SOS1 are published in degrees Fahrenheit, requiring a unit transformation to Celsius. The wind speed measurements from SOS2 have units of kilometres per hour and are made every 45 minutes, so do not align temporally with the air temperature values from SOS1, even though both time series start at the same time (as specified in the queries to each SOS). To compound matters further, each SOS is generating time series data with different syntaxes, and the wind chill function requires an input of yet another different syntax.

In order to reconcile the workflow to one that will correctly achieve the desired intention, the following operations must be described within the workflow:

- Convert the units of air temperature measurements (of SOS1) to degrees Celsius.
- Temporally align both time series before merging into one time series as input to the wind chill function (e.g., if we want hourly values for wind chill, this will require the interpolation of wind speed values over time to and sample corresponding values every hour to pair with the air temperature values).
- Generate a semantically and syntactically valid input document for the wind chill function which combines the aforementioned corrected data.

Typically, the author would proceed to manually reconcile the workflow by inserting appropriate functions and adapters/shims to perform these transformations. We will now show that, by encoding the types involved in the workflow as described in the earlier sections, how the type system can be used to automate tasks in completing the workflow in Figure 3 by assisting in the grounding of an abstract function specification to merge the
time series data, and to perform type incompatibility detection and correction by the semi-automated insertion of adapter/shim functions that perform type transformations.
A.2 Function Abstraction and Grounding

Since the workflow in Figure 4 is incomplete because the input data is not linked to the input of the wind chill function, we seek methods for joining them together to create a complete workflow. However, instead of constructing intermediate components for the workflow manually, we aim to take advantage of the system to infer how we can construct a correct sub-workflow by encoding our intension in the abstract as constraints, which the workflow system can resolve to a ground implementation.

Firstly, we consider the types $T_1$, $T_2$ and $T_3$ that we have in the incomplete workflow in Figure 1:

<table>
<thead>
<tr>
<th>Surface Structural</th>
<th>Core Structural</th>
<th>Semantic Type (OWL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>XML Schema 1</td>
<td>$[(t, v_1)]$</td>
<td>A list of time $(t)$ and value $(v_1)$ tuples (time series), where $v_1$ are measurements of air temperature in °F, and where $(t)$ values start at time X and have a time step interval of 60 minutes.</td>
</tr>
<tr>
<td>XML Schema 2</td>
<td>$[(t, v_2)]$</td>
<td>A list of time $(t)$ and value $(v_2)$ tuples (time series), where $v_2$ are measurements of wind speed in km/h, and where $(t)$ values also start at time X (as in those of $T_1$) and have a time step interval of 45 minutes.</td>
</tr>
<tr>
<td>CSV</td>
<td>$[(t, v_1, v_2)]$</td>
<td>A list of time $(t)$ and value $(v_1, v_2)$ tuples (time series), where $v_1$ are measurements of air temperature in °C, $v_2$ are measurements of wind speed in km/h, and where $t$ values have some time step interval.</td>
</tr>
</tbody>
</table>

Firstly, we are required to describe the way the input time series data is to be merged into a single time series. This sort of operation can be described abstractly over core structural types as an anonymous function $f_M$ which takes two lists of tuples each with a time and value vector, and combines this into a single list where value vectors are concatenated together where the times $(t)$ are equal in tuples across both input lists (for example, the merge function signature may be: $f_M([(t, v_1)], [(t, v_2)]) \rightarrow [(t, v_1 ++ v_2)]$, where $++$ is vector concatenation). Figure 5 demonstrates this specification in the current workflow, as follows:

![Figure 5](image-url)

**Figure 5.** Incomplete workflow describing an abstract merge over time series data. The blue arrows represent bindings between typed operator outputs to inputs.

This partial specification of the workflow describes the necessary requirement to merge the two time series $T_1$ and $T_2$, but does not yet specify exactly how it is to be performed. In order to determine this, the system will need to ground the function abstraction (as described in Section 3.2) to a transformation which achieves the specified effect. Consider that the system does indeed have a transformation function already available, as shown below in Figure 6.
The time series vector merge function has the following input and output types:

<table>
<thead>
<tr>
<th>Surface Structural</th>
<th>Core Structural</th>
<th>Semantic Type (OWL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S₁</td>
<td>Java Array</td>
<td>A list of time ((t_1, v_1; v_1)) tuples (time series), with the same time step interval and start time as described in S₂.</td>
</tr>
<tr>
<td>S₂</td>
<td>Java Array</td>
<td>A list of time ((t_2, v_2; v_2)) tuples (time series), with the same time step interval and start time as described in S₁.</td>
</tr>
<tr>
<td>S₃</td>
<td>Java Array</td>
<td>A (time series) sequence of time ((t_3, v_1; v_1 + v_2; v_2)) tuples, with the same corresponding time step interval, start time and value types as described in S₁ and S₂.</td>
</tr>
</tbody>
</table>

The system matches this transformation function to the merge function abstraction as it satisfies the constraints in the types and function signature. Specifically, this function strictly has input and output types which are subclasses of those described by the abstract merge function specification, in that it requires that all types have the same surface structural type (Java), and that the time step interval is the same in both inputs and output.

Individually, the type system finds that the semantic types \(T_1\) and \(T_2\) in the workflow are more specific subtypes of either \(S_1\) or \(S_2\). However, \(T_1\) and \(T_2\) cannot be considered to be compatible as the inputs to the merge function together at the same time, under the conditions against \(S_1\) and \(S_2\) which equate their time step values, since \(T_1\) and \(T_2\) have different time steps.

Similarly, when considering how values propagate through the merge function, we find that the values being transformed by the function (i.e., air temperature and wind speed) are not modified but are propagated to the output of the function, thus have the same units of measure in the output structure with type \(S_3\). However, there is still a mismatch of units between the air temperature in degrees Fahrenheit values as coming out (and going in) of the merge function, against what is required in \(T_3\) (air temperature in degrees Celsius).

Since the merge function is required in this workflow as shown in the specification in Figure 5, the system will attempt to reconcile the types \(T_1\) and \(T_2\) against \(S_1\) and \(S_2\), and \(S_3\) against \(T_3\), using type transformation functions, in order to use the time series merge operator in Figure 6.

### A.3 Type error detection and correction

As described in the last section, the system has resolved the merge function abstraction to a particular ground function but has detected that the input and output types are not compatible with those as specified in the workflow to which they are to be bound.
In order to correct these type inconsistencies, the system will attempt to infer if types $T_1$ and $T_2$ are indirectly compatible to $S_1$ and $S_2$ (and $S_3$ with $T_3$) through some type transformational steps. Since no core structural transformational steps are necessary as these aspects of the types are already directly consistent, there are only semantic and surface structural inconsistencies left to resolve. Firstly, this section considers the semantic inconsistencies of types $T_1$ and $T_2$ against $S_1$ and $S_2$.

Consider the availability of a time series interpolation function as described in Figure 7. This function takes a list of time value pairs (representative of a time series which can be interpreted as having a continuous interval), and attempts to interpolate between the given points in some specified way, such as using curve-fitting for linear interpolation, polynomial, etc. in order to resample the values using a different time interval.

![Figure 7. Generic, ground time series interpolation operator (Java implementation).](image)

The input and output types of this function are summarised in Table 3.

<table>
<thead>
<tr>
<th>Surface Structural</th>
<th>Core Structural</th>
<th>Semantic Type (OWL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_1$</td>
<td>Java Array</td>
<td>$[(t_1, v_1)]$</td>
</tr>
<tr>
<td>$P_1$</td>
<td>URI</td>
<td>$x$</td>
</tr>
<tr>
<td>$P_2$</td>
<td>(Java Float, URI)</td>
<td>$(t_{\text{period}}, t_{\text{ uom}})$</td>
</tr>
<tr>
<td>$R_2$</td>
<td>Java Array</td>
<td>$[(t_2, v_1')]$</td>
</tr>
</tbody>
</table>

In the presence of such a function, the system will recognize that by transforming one (or both) data with types $T_1$ or $T_2$ through the interpolation function that the resultant data type(s) $R_2$ will be compatible with both $S_1$ and $S_2$ (when considered together) as input to the merge function. As the time interval in the semantic type description for $R_2$ is not constrained to be any particular value (must be defined by $P_2$), and does not propagate through the function from the input. However, input data of types $P_1$ and $P_2$ are as yet unbound, and no other information is available to the system as to how to resolve these. Therefore, while the system can suggest that the application of this interpolation function may be appropriate based on the limited information it has about the types, it will be left to the user to decide on the applicability of the function and if so, to define all unbound input parameters manually.

By binding each of $T_1$ and $T_2$ as input to time series interpolation functions, we transform them to new types, that which we will define as $R_2'$ and $R_2''$ (respectively), which capture the propagated type of values being interpolated and re-sampled over, namely, that of air temperature and wind speed measurements, respectively. Similarly, when considering the output of the merge function, we find that the output is in fact a subtype of $S_3$ since this function also propagates input values with particular semantic types that it merges into one.
time series; we will refer to this type as $S_3'$, where $S_3' \subseteq S_3$. In comparing type $S_3'$ to the required input type $T_3$ of the function to compute wind chill, we find that the types are not directly compatible, as summarised:

<table>
<thead>
<tr>
<th>Surface Structural</th>
<th>Core Structural</th>
<th>Semantic Type (OWL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_3$</td>
<td>Java Array</td>
<td>A (time series) sequence of time ($t$) and value ($v_1$, $v_2$) tuples, with the same time step interval as defined by both $S_1$ and $S_2$.</td>
</tr>
<tr>
<td>$S_3'$</td>
<td>Java Array</td>
<td>A (time series) list of time ($t$) and value ($v_1$, $v_2$) tuples, where $v_1$ are measurements of air temperature ($^\circ$F), $v_2$ are measurements of wind speed (km/h), and where $t$ values have a time step interval as defined by $S_3$.</td>
</tr>
<tr>
<td>$T_3$</td>
<td>CSV</td>
<td>A (time series) list of time ($t$) and value ($v_1$, $v_2$) tuples, where $v_1$ are measurements of air temperature ($^\circ$C), $v_2$ are measurements of wind speed (km/h), and where $t$ values have some time step interval.</td>
</tr>
</tbody>
</table>

$S_3'$ is not directly compatible with $T_3$ as while their core structures match, $S_3'$ defines measurements of air temperature in $^\circ$F ($v_1$), whereas the same air temperature measurements ($v_1$) in $T_3$ are described as having units in $^\circ$C; additionally, the surface structures differ (Java Array as opposed to CSV).

The system admits unit transformation functions as depicted in Figure 8 which are capable of transforming from one type ($V_1$) to another ($V_2$) that preserve surface and core structural types (applicable to a common surface structural type) but which translates semantic types. With such a function, the translation between the semantic types only occur in parts of the type expressions that describe compatible primitive data in the core structural type as having differing but compatible units of measure. In determining which units of measure are compatible, such functions rely on the availability of individual functions such as the one depicted in Figure 9 which implement transformations between primitive values of differing units.

**Figure 8.** Type transformation function ($f_T$) applying unit conversion functions over primitives in any core structure permitted by the surface structure (Java implementation).

<table>
<thead>
<tr>
<th>Surface Structural</th>
<th>Core Structural</th>
<th>Semantic Type (OWL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_1$</td>
<td>Java</td>
<td>$O_1$, which describes primitive values within $T$ as having particular units of measure</td>
</tr>
<tr>
<td>$V_2$</td>
<td>Java</td>
<td>$O_2$, which describes primitive values within $T$ as having a different unit of measure to those defined in $O_1$, where the different units are compatible and there exists type transformations $f_t$ to convert between all the specified different units of measure</td>
</tr>
</tbody>
</table>
**Figure 9.** Type transformation function \((f_T)\) to perform a transform from values in °F to °C (Java implementation).

**Table 6.** Aspects of types \(X_1\) and \(X_2\)

<table>
<thead>
<tr>
<th>Surface Structural</th>
<th>Core Structural</th>
<th>Semantic Type (OWL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(X_1) Java Number</td>
<td>(v_1)</td>
<td>(v_1) is a Temperature (°F)</td>
</tr>
<tr>
<td>(X_2) Java Number</td>
<td>(v_1)</td>
<td>(v_1) is a Temperature (°C)</td>
</tr>
</tbody>
</table>

Since \(S_3'\) is not directly compatible to \(T_3\), in testing for an indirect compatibility relationship, the type system will locate the type transformation function in Figure (8) which can apply the type transformation in Figure (9) on data of type \(S_3'\) in order to construct a new type, \(V_2'\):

**Table 7.** Aspects of type \(V_2'\)

<table>
<thead>
<tr>
<th>Surface Structural</th>
<th>Core Structural</th>
<th>Semantic Type (OWL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(V_2') Java Array ([(t, v_1, v_2)])</td>
<td>A (time series) list of time (t) and value ((v_1, v_2)) tuples, where (v_1) are measurements of air temperature (°C), (v_2) are measurements of wind speed (km/h), and where (t) values have an interval of that defined by (S_3').</td>
<td></td>
</tr>
</tbody>
</table>

The last type \(V_2'\) now only differs in surface structure to \(T_3\), but the system can either generate, or may already have available as shown in Figure 10, a function to render CSV data from core structural types matching lists of flat record structures:

**Figure 10.** CSV generator function (Java implementation).

**Table 8.** Aspects of type \(Y_1\), and \(Y_2\)

<table>
<thead>
<tr>
<th>Surface Structural</th>
<th>Core Structural</th>
<th>Semantic Type (OWL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Y_1) Java Array ([(x_1,\ldots,x_n)])</td>
<td>(O_1) (any type)</td>
<td></td>
</tr>
<tr>
<td>(Y_2) CSV ([(x_1,\ldots,x_n)])</td>
<td>(O_1)</td>
<td></td>
</tr>
</tbody>
</table>

By composing the CSV generator in Figure 10 to data of type \(V_2'\), we achieve data with a type of \(Y_2'\), which is directly compatible with type \(T_3\), thus completing a full (yet still only partially grounded) workflow as shown below in Figure 11:
Figure 11. Partially grounded workflow with undefined parameters $P_1$ (interpolation type) and $P_2$ (time series interval).

At this point, the user is required to ground all unspecified parameters of the functions composed to complete the workflow before it can be executed; in this case, the user is required to specify an interpolation type $P_1$ as input to both time series interpolation functions operating over the time series of types $T_1$ and $T_2$ (e.g., linear interpolation). Also, the user is required to select a common time step interval $P_2$, also as input to both time series interpolation functions, which describe how to resample the time series $T_1$ and $T_2$ such that they are temporally aligned before applying the time series merge operation.

In summary, this example has shown the following:

- Abstract function specifications with mappings to ground implementations, and their semi automatic insertion into a workflow.
- Sophisticated type checking capabilities, with type inconsistency detection and correction via type transformation composition and insertion.
AgroHydroLogos: development and testing of a spatially distributed agro-hydrological model on the basis of ArcGIS.

Soulis Konstantinos¹, Dercas Nicholas²
Agricultural University of Athens, Department of Natural Resources Management and Agricultural Engineering, Sector of Water Resources Management, Athens, Greece, (¹soco@aua.gr, ²ndercas1@aua.gr)

Abstract: In this paper the development and testing of AgroHydroLogos is presented, which is a new spatially distributed, continuous hydrological model. The model was developed as an extension of the ESRI ArcGIS Geographical Information Systems’ software package. It calculates, on a daily basis, the main hydrological balance components, and other important parameters, including plants water stress and irrigation water needs. Its conceptual scheme was based on simplified but well established methods for the simulation of the various hydrological processes. The hydrological model is being developed using object oriented programming techniques, and its Graphical User Interface is very easy to use and it follows the standard ArcGIS software extensions form. The model was tested in a Mediterranean experimental watershed in Greece. The final conclusion is that AgroHydroLogos can produce valuable information concerning agricultural water management; it can also produce important information for the assessment of climate change impacts and for the evaluation of various adaptation and mitigation strategies.

Keywords: hydrological model; climate change; water management; GIS (Geographical Information System); OOP (Object Oriented Programming).

1. INTRODUCTION

Much current research in hydrology is directed at improving the ability to predict the effect of various land-use and climate conditions on the water balance, streamflow variability and water quality of river basins. These and most other practical applications of hydrology require the use of hydrological models [Digman 2002]. Models are extremely useful as test beds for ideas, as data processing and analysis aids, and for exploring scenarios that cannot be tested in the real world. However, it is not a unique opinion that modeling is fine as long as it is not confused with the real world [Silberstein 2006]. A simulation model is a representation of a portion of the natural world, which is simpler than the prototype system and which can reproduce some but not all of the characteristics thereof [Dooge 1986]. Specific scales and specific purposes require the use of different modeling approaches. Advances in hydrological research and the abundance of computer power, facilitate the development of more efficient and more sophisticated hydrological models, as well. As a result, the hydrologic literature abounds with descriptions of models. Hydrological models can be categorized based on their spatial representation to “Lumped” models, which represent a catchment as a single entity, or a small group of entities, and simulate state variables and fluxes into and out of the catchment as a whole, e.g. HEC-1 (US Army Corps of Engineers) and SWAT (USDA, US Department of Agriculture), and to “Distributed” models, which divide the catchment into many entities, each representing small parts of the catchment, and the state variables and fluxes between the entities are determined across the catchment, e.g. MIKE SHE [Refsgaard and Storm 1995], CASC2D [Ogden 1997] and SHETRAN [Ewen et al. 2000]. There are also semi-distributed models, in which variables are not explicitly distributed across the catchment, e.g., TOPMODEL [Beven and Kirkby 1979]. Concerning their simulation basis, models can be also
categorized to “conceptual”, “empirical” or “physical” [Digman 2002]. The vast majority of models in the literature are “conceptual”, using a priori relationships to simulate flows and storages [Alemaw and Chaoka 2003, Moretti and Montanari 2007, Meng et al. 2008]. The recent development of Geographical Information Systems (GIS) and the availability of large volumes of spatial data have raised a high demand for closer integration of GIS and hydrological models. Recent efforts to develop hydrological models on top of GIS are presented by Jain et al., [2004], Shen et al. [2005], Efstratiadis et al. [2008], Li and Zhang [2008], and Teng et al. [2008]. Also, advances in computer programming techniques, such as Object Oriented Programming (OOP), triggered the development of new hydrological models [e.g. Band et al. 2000, Wang et al. 2005, Leone and Chen 2007, and Martinez et al. 2008]. Current efforts are also focused on the development of integrated models involving different scientific disciplines [Pietroniro et al. 2006, Krol et al. 2006, Krysanova et al. 2007, and Krol and Bronstert 2007].

In this paper the development and testing of AgroHydroLogos is presented, which is a new spatially distributed, continuous hydrological model, aiming to produce valuable information concerning agricultural water management, assessment of climate change impacts and evaluation of various adaptation and mitigation strategies. The model was developed as an extension of the ESRI ArcGIS, a GIS software package, in order to facilitate the fully spatially distributed calculation of the hydrological balance components, and at the same time to make use of the advanced capabilities of ArcGIS in managing, editing, analyzing and visualizing geographical data. The model calculates, on a daily basis, the main hydrological balance components, such as, soil moisture, direct runoff, deep infiltration, actual evapotranspiration, and base flow. Furthermore, it can calculate other important parameters, including plants water stress and irrigation water needs in order to constitute a valuable software tool for agro-hydrological analysis. The hydrological model is being developed using OOP techniques, thus facilitating further development. The Graphical User Interface (GUI) of the model is very simple to use and it follows the standard ArcGIS software extensions form. The hydrological model is simple and flexible, as its conceptual scheme was based on simplified but well established methods for the simulation of the various hydrological processes. The model was tested in a Mediterranean experimental watershed in Greece, in order to validate the soundness of the conceptual scheme and to evaluate its efficiency.

2. AGROHYDROLOGOS SOFTWARE DESCRIPTION

AgroHydroLogos software consists of a geographical database, a knowledge base, data pre processing and post processing tools and the core hydrological model. The model was developed on the basis of ESRI’s ArcGIS 9.2 software based on ArcObjects functionality. Arc Objects are built using Microsoft’s Component Object Model (COM), therefore it is possible to extend them by writing COM components using any COM compliant development language. Through Arc Objects, ArcGIS exposes all its functionality and its data structure to the modeler, allowing the development of applications dynamically linked to the geographical database. AgroHydroLogos model components were built making use of the above functionality and then were embedded to ArcGIS as an extension. It was developed based on OOP techniques, using Visual Studio .NET 2005. The model was organized in a collection of discrete objects that facilitate the efficient description of the hydrological cycle. AgroHydroLogos GUI follows the form and the operation of ArcGIS extensions. AgroHydroLogos can be activated or deactivated through the “Extensions” menu. The GUI consists of the main toolbar and the various forms activated through it. Each AgroHydroLogos project is embedded in an ArcMap project and the values of all its options and its parameters are stored in the related .mxd file. A number of tools are also added to the existing GIS functionality in order to assist the dynamic preparation of input data layers and the post processing of results.

2.1 Geographical Database and Knowledge Base

The model works as part of the Geographical Information System in direct interaction with the Geographical database. The inputs come directly from the Geographical database and
the resulting outputs, either directly from the model or from post processing tools, are stored in it too. In this way the user has access to all the GIS functionality for data preparation, storage, visualization and analysis in a familiar and user-friendly graphical user interface. Moreover, all the advantages of a database management system are provided.

The structure of the Geographical database must follow the predefined data model and all spatial input data must be geo-referenced. The spatial outputs are also geo-referenced, allowing easy overlapping and comparing the results with data coming from other sources or sharing the results with other users. The required input data are a depressionless Digital Elevation Model (DEM), soil, land cover, geology and weather stations. Based on the above data, all the requested parameters for running the model are calculated in spatial distributed form such as: the soil water holding capacity, the soil permeability, etc. The above data layers can be calculated dynamically, making use of the knowledge base, facilitating the assessment of various scenarios.

The knowledge base contains information needed for spatial data pre-processing and for some procedures of the hydrological model like the soil group - land cover complex and Curve Number correspondence, soil hydraulic property estimation or vegetation characteristics (rooting depth, planting dates, coefficients needed for actual evapotranspiration calculation, etc). All this information is stored in the form of a separate database to promote adding or editing capabilities.

2.2 Meteorological data

In isolated areas, like mountainous regions with steep relief or small islands, the surface meteorological information available is scarce and confined to plain or coastal locations, considered to be representative at a regional or synoptic scale. Therefore, this information cannot reflect the spatial variation of the climate as influenced by topography and advective processes [Brito et al 1999].

The approach used in the presented model in order to overcome this problem is the use of interpolation techniques accompanied with adjustments depending on topography. In the methodology used, the weight of each selected weather station for each grid cell of the simulated area is calculated at the beginning of each execution of the model, using the Inverse Distance Weight or the Thiessen Polygons method. The values of the meteorological parameters are then calculated on a day by day and cell by cell basis during run time, using the calculated weights and taking also into account the geomorphologic conditions in every grid cell. With this methodology a negligible cost is required with regard to performance or storage. Similar techniques, like spatial interpolation of the meteorological data prior to the running of the model, decrease significantly the model’s performance and necessitate significant storage space.

2.3 Spatial Resolution

A primary goal of developing spatially distributed hydrological models is to minimize the number of parameters which need calibration [Schumann and Geyer 2000]. Lumped models need much less computation effort and, most of the time, have comparable accuracy but always under the condition of extensive calibration. However, the results obtained with spatially distributed hydrological models are considerably dependent on their spatial resolution. In general, a small pixel size gives more accurate representation of spatial parameters such as slope, hydrographical network or water divide [Zhang and Montgomery 1994, Garbrecht and Martz 1994, Thieken 1999, and Wang et al. 2000], but increases geometrically the computing time as well as the volume of the input and output spatial data. Even so, small pixel size is not always related to more accurate results. Zhang and Montgomery [1994] suggested that a DEM spatial resolution finer than the original data accuracy can cause artefacts and subsequent wrong slope and flow direction representation. In this model the spatial resolution is set equal to the spatial resolution of the mask grid that is used to define the modelling area. By this way the pixel size can be easily adjusted to achieve the optimum relation between accuracy and performance.
2.4 Hydrological Model Structure

The hydrological model in the core of AgroHydroLogos Software, calculates in spatial distributed form and on a daily basis, the main hydrological balance components, such as, soil moisture, direct runoff, deep infiltration, actual evapotranspiration, and base flow, and other important parameters for agro-hydrological analysis, including plant water stress and irrigation water. The daily temporal discretization was chosen in order to describe with satisfactory detail the hydrological processes and at the same time to allow the simulation of long time periods and the utilization of readily available meteorological data and data concerning future climate projections. The conceptual scheme of the model was designed in order to achieve the following targets:

- to make use of simplified but well established methods for the simulation of the various hydrological processes
- to require easily available spatial data and parameters
- to adapt to semi-arid and arid regions
- to efficiently describe vegetation-water dynamics

The conceptual scheme and the involved stores and flows are presented in figure 1 and described in detail in the following paragraphs.

Water Balance of the reference soil volume

Soil moisture of the top soil layer is directly involved in the calculation of most water balance components such as direct runoff, infiltration, deep infiltration, and actual evapotranspiration. Thus, the principal equation of the hydrological model is equation (1) that describes the water balance of the reference soil volume:

\[ SWC_{i-1} - SWC_i = P_i - Q_i - aET_i - DI_i \]  

(1)

where \( SWC_{i-1} \) and \( SWC_i \) (mm) are the reference soil volume water contents of the day before and the actual day respectively, \( P_i \) (mm) is the total rainfall depth the actual day, \( Q_i \) (mm) is the direct runoff the actual day, \( aET_i \) (mm) is the actual evapotranspiration the actual day and \( DI_i \) (mm) is the deep infiltration the actual day. As reference soil volume is defined the top soil layer, which in deep soils is limited to the rooting depth. A schematic representation of the interactions determining the water balance of the reference soil volume is shown in figure 1.

In wet periods \( SWC \) can increase up to a maximum, which is equal to the water holding capacity of the reference soil volume (\( SWHC \)), resulting in significant decrement of the infiltration rate, increment of the actual evapotranspiration rate and increment of the deep infiltration rate. In dry periods soil becomes very dry limiting actual evapotranspiration rate and deep infiltration rate and increasing infiltration rate when rainfall occurs. \( SWC \) value is one of required initial conditions for the application of the model. Generally the determination of the initial \( SWC \) is very difficult. However, in some specific periods of the year (after a very wet winter period or a very dry summer period) we can set \( SWC \) equal to the \( SWHC \) or equal to zero, respectively.

Direct Runoff

The Soil Conservation Service Curve Number (SCS-CN) method is used to calculate direct runoff in relation to land use, soil type and antecedent moisture. This method was originally developed by the U.S. Department of Agriculture (Soil Conservation Service) and documented in detail in the National Engineering Handbook, Section 4: Hydrology (NEH-4) [SCS 1956, 1964, 1971, 1985, 1993]. Due to its simplicity, it soon became one of the most popular techniques among the engineers and the practitioners, mainly for small catchment hydrology [Mishra and Singh 2006]. The main reasons for its success is that it accounts for many of the factors affecting runoff generation including soil type, land use...
and treatment, surface condition, and antecedent moisture condition, incorporating them in a single CN parameter. Furthermore, it is the only well established method that features readily grasped and reasonably well-documented environmental inputs [Ponce and Hawkins 1996]. Although the SCS method was originally developed in the United States and mainly for the evaluation of storm runoff in small agricultural watersheds, it soon evolved well beyond its original objective and was adopted for various land uses [Rawls et al. 1981, Mishra and Singh 1999] and it became an integral part of more complex, long-term, simulation models [Choi et al. 2002, Mishra and Singh 2004, Zhan and Hang 2004, Tyagi et al. 2008].

Based on the water balance equation and on the assumptions that the ratio of runoff to effective rainfall is the same as the ratio of actual retention to potential retention, and that the amount of initial abstraction is a fraction of the potential maximum retention \( I_a = \lambda S \), the basic equation of the SCS-CN method is obtained:

\[
Q = (P - \lambda S) \left[ P + (1 - \lambda)S \right],
\]

where \( P \) (mm) is the total rainfall, \( Q \) (mm) is the direct runoff and \( S \) (mm) is the potential maximum retention, which is valid for \( P \geq \lambda S \); otherwise \( Q = 0 \). In Eq. (2), the initial abstraction rate is normally set to a constant value (\( \lambda = 0.2 \)) in order for \( S \) to be the only parameter of the method. Furthermore, the potential retention \( S \) is expressed in terms of the dimensionless curve number (CN) through the relationship

\[
S = 25400/CN - 254
\]

taking values from 0, when \( S \to \infty \), to 100, when \( S = 0 \). In this application the CN values can be dynamically calculated from the soil and land cover data, which are stored in the geographical database and the knowledge base. The resulting CN values apply for normal antecedent moisture conditions (AMC II). For dry (AMC I) or wet (AMC III) conditions, equivalent curve numbers can be computed by:

\[
CNI = \frac{4.2CNII}{10 - 0.058CNII} \quad \text{and} \quad CNIII = \frac{23CNII}{10 + 0.13CNII}
\]

In the model, CN value is justified each day, depending on the actual soil moisture.

**Deep Infiltration**

The determination of the drainage through the soil profile is based on the Brooks and Corey [1964] equation:

\[
K = K_s \left( \frac{\theta_s}{\theta} \right)^n
\]

where \( K \) represents the soil unsaturated hydraulic conductivity when soil moisture equal to \( \theta \), \( \theta_s \) is the saturated soil moisture, \( K_s \) is the saturated hydraulic conductivity, and \( n \) is a shape factor, assuming a free drainage (zero pressure head) boundary condition at the bottom of the reference soil volume.

**Base Flow**

Base flow is simulated based on the equation proposed by Arnold et al. [1993] and Hattermann et al. [2004]:

\[
Q_{b,i} = Q_{b,i-1} e^{-\alpha} + W_{R,i} (1 - e^{-\alpha}) \quad \text{when } aq > aq_e
\]

\[
Q_{b,i} = 0 \quad \text{when } aq \leq aq_e
\]

where \( Q_{b,i} \) (mm) is the base flow in day \( i \), \( Q_{b,i-1} \) (mm) is the base flow in day \( i-1 \), \( \alpha_b \) is the base flow reduction, \( W_{R,i} \) (mm) is the aquifer recharge in day \( i \), \( aq \) (mm) is the water quantity stored in the aquifer in day \( i \), and \( aq_e \) (mm) is the limit of stored water, below which there is no base flow.

**Evapotranspiration**

Actual Evapotranspiration (\( aET \)) rate is depending on weather parameters, land cover and water availability. In this study reference evapotranspiration (\( ET_o \)) is calculated from weather parameters and then \( aET \) is determined taking into account the land cover characteristics and the soil moisture. The only factors affecting \( ET_o \) are climatic parameters. Consequently, \( ET_o \) is a climatic parameter and can be computed from weather data. \( ET_o \)
expresses the evaporating power of the atmosphere at a specific location and time of the
year and does not consider the plant characteristics and soil factors. The FAO Penman-
Monteith [FAO 1998] method is used for determining $ET_o$ in this model. This physically
based method has been selected because it closely approximates $ET_o$ and explicitly
incorporates both physiological and aerodynamic parameters. The required weather
parameters are temperature, humidity, wind speed, and solar radiation. Alternatively, in
case that the only available weather parameter is temperature, the Hargreaves method can
be used [Hargreaves and Samani 1985].

Differences in leaf anatomy, stomatal characteristics, aerodynamic properties, and even
albedo cause the plant evapotranspiration $ET_c$ to differ from $ET_o$ under the same climatic
conditions. $ET_c$ is related to $ET_o$ using experimentally determined ratios of $ET_c/ET_o$, called
crop coefficients ($K_c$).

$$ET = K_c ET_o$$ (8)

Due to variations in the plant characteristics throughout its growing season, $K_c$ for a given
plant also changes during the year. The $aET$ is finally calculated by using a water stress
coefficient $K_w$ expressing the effect of water availability on crop evapotranspiration.

$$aET = K_w ET$$ (9)

The $K_w$ parameter is related to soil moisture but differs for different land cover
characteristics [FAO 1998]. The values of $K_c$ and $K_w$ are determined for each grid cell and
each day, depending on the vegetation characteristics of each place, based on the
knowledge base, which contains information for a wide range of land cover types. Using
the above described methodology, the model efficiently describes vegetation-water
dynamics and is able to calculate plants water stress. The model is also able to simulate
irrigation and to calculate irrigation water needs for selected land cover types.

**Runoff Routing**

Daily accumulated runoff is estimated for each grid cell of the river network, for every time step, by adding the
accumulated surface runoff for this time period to the accumulated base flow. Runoff routing through overland
flow and through the hydrographical network is performed with a simplified travel time approach. Travel
time varies depending on the accumulated runoff in each
grid cell, it is calculated with the kinematic wave
approach and it is relative to the topographical slope, and
to a roughness coefficient based on land cover type. In
every time step, the daily accumulated runoff of each grid
cell is travelling downhill until the total travel time is
equal to the time step interval (Fig. 2). The methodology used does not aim to produce very
detailed hydrographs but it is sufficient for this application, which is mainly focused on
water balance components estimation and uses a relatively large calculation time step. The
parameters involved in the flow routing procedure can be directly estimated using data from
the knowledge base.

### 3. MODEL APPLICATION

#### 3.1 Site Description and Input Data

The model was applied to the small scale experimental watershed of Lykorrema stream
(7.84 km²), situated on the east side of Penteli Mountain, Attica, Greece (Coordinates: UL
23°53’33”E-38°04’13”N; LR 23°56’00”E-38°02’28”N) (Fig. 3a). The region is
characterized by a Mediterranean semi-arid climate with mild, wet winters and hot, dry
summers. Precipitation occurs mostly in the autumn–spring period. The yearly average
precipitation value for the five years studied is 595mm. The reference evapotranspiration
rate varies from about 1mm/day during winter to 7mm/day during summer. The watershed
presents a relatively sharp relief, with elevations ranging between 280 m and 950 m. Its
average elevation is 560 m and its average slope is as high as 36%. Geologically it is
characterized by schists formations covering 96% of its area, while the rest is covered by
marbles. Schist formations in the area are not impervious. They are tectonically intensely fractured and their upper layer is eroded. The aquifers system developed within the intensely fractured bedrock contributes significantly to the base flow of the watershed, which is continuous throughout the year. A soil survey in the area showed that the watershed is dominated by coarse soils with high hydraulic conductivities and a smaller part is covered with medium textured soils presenting relatively high hydraulic conductivities. A detailed land cover classification based on remote sensing techniques, showed that the dominant vegetation type is pasture with a few scattered tufts of trees (Fig. 3a). A small part of the watershed is covered by bare rock [Soulis et al. 2009].

The study area is equipped with a dense hydro-meteorological network, which is fully operational since September 2004. The installed equipment consists of five rain-gauges, one hydrometric station at the outlet of the watershed, one meteorological station and four temperature-relative humidity recorders. The data are recorded with a time step of 10 min. In the current application the period from September 2004 to August 2008 was simulated. The required data layers where produced based on detailed DEM and the above described information. These data layers are: weather stations, DEM, land cover, flow direction, flow accumulation, curve number, soil water holding capacity, saturated hydraulic conductivity, roughness coefficient, base flow reduction constant, base flow initiation limit (Fig. 3).

Figure 3. (a) Map of Lykorrema stream experimental watershed. (b,c) Input data layers, (b) CN spatial distribution. (c) SWHC spatial distribution.

### 3.2 Results and Discussion

Calibration of complex, spatial distributed hydrological models is a difficult task. The involvement of many parameters in spatial distributed form intensifies the problem of the existence of a number of local optima rather than a global optimum. Furthermore, the calibration of hydrological models requires extensive and detailed datasets of measured data. However, these models are used in regions with lack of data or for future projections under different conditions. Even when calibration data exist, those normally concern only runoff recordings at specific locations, while their accuracy is limited due to the great difficulties related to flow rate recordings in natural streams. In order to overcome the above mentioned difficulties the approach included firstly the use of simplified but well established methods having a physical basis for the simulation of the various hydrological processes and the involvement of a limited number of easily available spatial data and parameters. Secondly, calibration efforts target to the most uncertain parameters and exploit the modeller experience on the studied site and the involved processes. In order to explore the use of the developed model in the above described framework, the model was calibrated using only the first year of the available data (Sept. 2004 – Aug. 2005), and by adjusting only the average values of the SWHC and $\alpha_b$ parameters. In figure 4 are

<table>
<thead>
<tr>
<th>Year</th>
<th>Rain (mm)</th>
<th>Runoff Meas. (mm)</th>
<th>Runoff Pred. (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>04-05</td>
<td>660</td>
<td>116</td>
<td>118</td>
</tr>
<tr>
<td>05-06</td>
<td>794</td>
<td>151</td>
<td>147</td>
</tr>
<tr>
<td>06-07</td>
<td>726</td>
<td>61</td>
<td>80</td>
</tr>
<tr>
<td>07-08</td>
<td>490</td>
<td>27</td>
<td>53</td>
</tr>
</tbody>
</table>

Table 1. Total yearly measured rainfall and runoff values and total yearly predicted runoff values.
illustrated the measured daily rainfall and runoff values in comparison with the predicted daily runoff values by the model application for the period from September 2004 to August 2008. In table 1 are shown the total yearly measured rainfall and runoff values and the total yearly runoff values predicted by the model for the same period. It can be observed that the model performance is quite good for the first two years, sufficiently good for the third year, and modest for the last year. It can also be observed that the last year of the modelling period is very dry and that in the last two years the main part of the rainfall occurs in a small number of intense storm events. The main remark is that the model significantly overestimates base flow during dry years, probably because the first year, which was used for the model calibration, is characterized by high base flow and low rainfall intensity values. In figure 5 are presented the predicted spatial distribution of the average yearly runoff, aET and aquifer recharge values. These results demonstrate the great temporal and spatial variability of the water balance components and the importance of a model capable to produce results in such a form as to promote integrated water resources management. It can be also observed that aET and runoff spatial distribution is related to the spatial distribution of the soil and land cover characteristics.

Figure 4. Daily rainfall values and daily measured vs. predicted runoff values.

Figure 5. Predicted spatial distribution of average yearly runoff, aET and aquifer recharge.

4. SUMMARY AND CONCLUSIONS

AgroHydroLogos is a new spatially distributed, continuous hydrological model. The model was developed as an extension of ESRI ArcGIS, in order to facilitate the fully spatially distributed calculation of the hydrological balance components, and to make use of the advanced capabilities of GIS in managing, editing, analyzing and visualizing geographical data. The model calculates, on a daily basis, the main hydrological balance components, such as soil moisture, direct runoff, deep infiltration, actual evapotranspiration, and base flow, and other important parameters, including plant water stress and irrigation water needs in order to constitute a valuable software tool for agro-hydrological analysis. The
hydrological model is being developed using OOP techniques, thus facilitating further development. Its Graphical User Interface is very easy to use and it follows the standard ArcGIS software extensions form. Its conceptual scheme was based on simplified but well established methods, which have a physical basis, for the simulation of the various hydrological processes. It was especially designed in order to enable the efficient description of vegetation water dynamics.

The model was applied to a Mediterranean experimental watershed in Greece. This application demonstrated the ability of the model to produce satisfactory results based on a limited set of input data and parameters and on a modest calibration effort. The final conclusion is that AgroHydroLogos is simple and flexible and it can produce valuable information concerning water resources management and especially agricultural water management. Additionally, it can produce important information for the assessment of climate change impacts on water resources and for the evaluation of various adaptation and mitigation strategies.

REFERENCES


Climate change and long term water availability in Western Australia - An experimental projection

M A Bari1, G E Amirthanathan1 and B Timbal2

1Extended Hydrological Prediction, Water Division, Bureau of Meteorology, Australia
2Centre for Australian Climate and Weather Research, Bureau of Meteorology, Australia

E-mail: m.bari@bom.gov.au

Abstract: The population of the Perth-Bunbury region in Western Australia is predicted to increase to 3.1 million by 2050. Water supply is a key issue, as below-average rainfall since the mid-1970s has led to about 40% decline in the streamflow. General Circulation Models (GCMs) project a further decrease in rainfall leading to diminished water resources in the future and posing a threat to water supply and the environment. In this experimental study we assess the impact of climate change at Serpentine Reservoir using data from eleven GCMs which contributed to the latest Intergovernmental Panel on Climate Change (IPCC) assessment report. Data from two emission scenarios (A2 and B1) were used and downscaled, using a state-of-the-art statistical downscaling model, to a 5 km resolution compatible with catchment modelling. The LUCICAT rainfall-runoff model was calibrated for the Serpentine catchment and then changes in runoff were projected using the downscaled rainfall data. Land use and potential evaporation were not changed for the future rainfall-runoff modelling. Nearly all GCMs projected reductions in rainfall by mid (2046-2065) and late (2081-2100) 21st century compared to 1981-2000 period. There was a significant variation in projected rainfall reductions between different GCMs and emission scenarios. Under the A2 climate scenario, there could be a further 14-24% reduction in rainfall, and this would result in a 49-69% reduction in reservoir inflow by the mid to end of the 21st century. Rainfall reduction under B1 scenario would be around 12% and corresponded to streamflow reduction of about 45-46%.

Keywords: LUCICAT model, climate change, A2 and B1 scenarios, downscaling, Serpentine catchment.

1. INTRODUCTION

The south-west of Western Australia has experienced declining rainfall since the mid 1970s (IOCI, 2002). Several decades of below average rainfall and a recent succession of dry years has focused attention on water resource availability and reliability in south-west Western Australia. The Serpentine Catchment is a major water supply catchment within the Integrated Water Supply System for Perth, the capital city in Western Australia. It is located in the Darling Ranges, where most of the water-supply catchments are located. The Darling Ranges has experienced a decline in winter rainfall of up to 20% over the past 30 years, resulting in a 40% or more reduction in runoff to reservoirs supplying the Perth metropolitan area (IOCI 2002; Bari and Ruprecht, 2003; Water Corporation, 2009). The decline in the water yield of these catchments has resulted in a greater dependence on Perth’s regional groundwater resources and has increased the potential for impact on coastal groundwater dependent ecosystems. In Australia, there have been some studies of the effects of climate change on water resources and catchment hydrology (Ritchie, et al., 2004; Charles, et al., 2009). In the south-west of Western Australia, there have been various studies relating to the impacts of climate change on water resources (Bari et al., 2005; Charles et al., 2007). However, most of these studies were focused on specific climate scenarios.

This paper focuses on the Serpentine water supply catchment in Western Australia and investigates the impacts of two climate change scenarios on streamflow and water yield through the application of LUCICAT (Land Use Change Incorporated CATchment) model. Recently developed 5 km grid rainfall generated by the Australian Bureau of Meteorology (Jones et al., 2009), and daily downscaled rainfall projections for two scenarios from 11 GCMs are used across the Serpentine catchments.

2. CATCHMENT DESCRIPTION

The Serpentine catchment is approximately 55 km south-east of Perth, Western Australia (Figure 1). The main land cover is jarrah (Eucalyptus marginata) forest, with only small parts cleared. The Jack Rocks sub-catchment was mined for bauxite in the 1990s, while parts of the Cameron West and Cameron Central sub-catchments were logged in 1995-96 and there has been selective logging within the Jayrup sub-catchment. Prescribed burning has a role in reducing the incidence of wildfires, regenerating native forests and conserving
biodiversity. The climate is temperate, with hot dry summers and cool wet winters. Average annual rainfall varies from 680 mm in the east to 1150 mm in the west of the catchment. The Serpentine River is the main river and Big Brook is the main southern tributary. The annual pan evaporation ranges from 1600 to 1700 mm.

![Figure 1 Location of the Serpentine water supply catchment](image)

**Figure 1** Location of the Serpentine water supply catchment

**Figure 2 Schematic diagram of methodology**

### 3. METHODOLOGY

Daily downscaled rainfalls were computed using variables from 11 GCMs for the periods 2046-2065 and 2081-2100. Two climate change scenarios were considered – A2 and B1. The four main components to this study are: (a) calibration of the LUCICAT rainfall-runoff model, (b) run LUCICAT model with ‘Historical’ (calibrated GCM) rainfall data and compare model output with the calibration, (c) simulation of the impacts of projected climate change scenario (A2, B1) on streamflow through running the LUCICAT model, and (d) analysis and assessment of the projected streamflow data (Figure 2).

#### 3.1 Climate Scenarios and Downscaling

Statistically downscaled rainfall data were obtained from the Australian Bureau of Meteorology’s Statistical Downscaling Model (SDM) based on an analogue approach (Timbal et al., 2009). The daily GCM data were
extracted from the Coupled Model Inter-comparison Project Number 3 (CMIP3), assembled as part of the Intergovernmental Panel on Climate Change (IPCC) 4th Assessment of Climate Change Science (Solomon et al., 2007). As daily data are required to perform the statistical downscaling, only 11 GCM outputs were used in this study. The models were ranked according to a measure of sensitivity (ΔT, Table 1). The daily climate data were available for three time-slices: 1961-2000; 2046-2065 and 2081-2100. Two emission scenarios were considered: A2, a very heterogeneous world with increasing population and technologically fragmented economic development, and B1, a global solution to economic, social and environmental sustainability (IPCC, 2001). The calibration and validation of the SDM was performed as part of an Australia-wide application (Timbal et al., 2009) using the high quality rainfall dataset assembled by the National Climate Centre (Lavery et al., 1997). Once the SDM has been calibrated it can be applied to any suitable observations within the same climate region. That flexibility was applied to the 5km grid resolution rainfall (Jones et al., 2009) generated as part of the Australian Water Availability Project (AWAP). Initial application was in the south-west of Western Australia which covers the Serpentine catchment.

<table>
<thead>
<tr>
<th>Originating group</th>
<th>Country</th>
<th>Model</th>
<th>Grid size (km)</th>
<th>ΔT (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSIRO</td>
<td>Australia</td>
<td>CSIRO3.5</td>
<td>~200</td>
<td>2.11</td>
</tr>
<tr>
<td>CSIRO</td>
<td>Australia</td>
<td>CSIRO3.0</td>
<td>~400</td>
<td>2.11</td>
</tr>
<tr>
<td>NASA/Goddard Inst for Space Studies</td>
<td>USA</td>
<td>GISSR</td>
<td>~300</td>
<td>2.12</td>
</tr>
<tr>
<td>Canadian Climate Centre</td>
<td>Canada</td>
<td>CCM</td>
<td>~300</td>
<td>2.47</td>
</tr>
<tr>
<td>Meteorological Research Institute</td>
<td>Japan</td>
<td>MRI</td>
<td>~300</td>
<td>2.52</td>
</tr>
<tr>
<td>Geophysical fluid Dynamics Lab</td>
<td>USA</td>
<td>GFDL2</td>
<td>~300</td>
<td>2.53</td>
</tr>
<tr>
<td>Meteo-France</td>
<td>France</td>
<td>CNRM</td>
<td>~200</td>
<td>2.81</td>
</tr>
<tr>
<td>Geophysical fluid Dynamics Lab</td>
<td>USA</td>
<td>GFDL1</td>
<td>~300</td>
<td>2.98</td>
</tr>
<tr>
<td>Institut Pierre Simon Laplace</td>
<td>France</td>
<td>IPSL</td>
<td>~300</td>
<td>3.19</td>
</tr>
<tr>
<td>Centre for Climate Research</td>
<td>Japan</td>
<td>MIROC</td>
<td>~300</td>
<td>3.35</td>
</tr>
<tr>
<td>Max Plank Institute for Meteorology DKRZ</td>
<td>Germany</td>
<td>MPI</td>
<td>~200</td>
<td>3.69</td>
</tr>
</tbody>
</table>

4. THE LUCICAT MODEL

4.1 The Model

The LUCICAT model is a semi-distributed hydrological model. A large catchment is divided into smaller Response Units to take into account the spatial distribution of rainfall, pan evaporation and land use. Each of the Response Units is represented by ‘simplified hill-slope’ and a fundamental ‘building-block’ model is applied. Catchment attributes such as soil depth, rainfall, pan evaporation, land use change, groundwater level and salt storage are incorporated into the building-block model (Bari & Smettem 2006). The building-block model consists of: (i) Dry, Wet and Subsurface stores, (ii) a saturated Groundwater Store, and (iii) a transient Streamzone Store. The transient Streamzone Store represents the groundwater induced saturated areas along the stream zone. The fluxes between the top layer Dry and Wet Stores represent the water movement in the unsaturated zone. The dynamically varying saturated stream zone areas represent surface runoff. The Groundwater Store controls the groundwater and salt fluxes to the stream zone. Generated flow from each of the Response Units is routed downstream by the Muskingum-Cunge routing scheme (Miller & Cunge, 1975). Water and salt balances of the lakes, farm dams and reservoirs in the catchment are also computed. The model runs in the LUCICAT Live framework (Bari et al., 2009).

4.2 Model Set Up

The application of the LUCICAT model for the Serpentine catchment involved data preparation, Response Units delineation, calibration and validation. The catchments were divided into 87 Response Units, based on digital elevation models. Daily rainfall and Morton Wet Surface Potential Evaporation (Morton, 1983) series on a 5 km grid were obtained from the Bureau of Meteorology Australian Water Availability Project (http://www.csiro.au/awap/) and then calculated for each of the Response Units using the reciprocal distance weighting method. Response Unit attributes, stream node network, surface topography, land use history and Leaf Area Indices were developed using ArcGIS and the MAGIC system.

4.3 Calibration and Results

The seven key sensitive physically meaningful parameters, which may vary from one catchment to another, were copied from previous applications and then calibrated by trial and error to obtain a better match between the recorded and the simulated streamflow data. The model performance was evaluated over 1975-2007 through the LUCICAT Live framework (Bari et al., 2009) by analysing graphs and comparing other statistical criteria. The performance criteria includes: (1) joint time series plots, (2) scatter diagram, (3) flow-period Error Index, (4)
Nash-Sutcliffe Efficiency, (5) Explained Variance, (6) Correlation Coefficient, (7) overall water balances, and (8) flow duration curves. Reservoir monthly inflow was computed based on rainfall, pan evaporation, water level, draw and release (Water Balance) supplied by Water Corporation (Jeevaraj, C., Pers. Comm., 2009). LUCICAT predicted inflow for the 1975-2007 period was 1% higher than that of the Water Balance method. Figure 3 illustrates the results of the calibration for the annual inflow into the reservoir and the water level.

4.3.1 Annual Streamflow

The annual simulated streamflow at all the gauging stations generally matched the observed records. The comparison of the simulated and the observed flows indicates that the model represents the flow generation very well. Average annual inflows to the reservoir calculated from the Water Balance method and LUCICAT calibration were 55 and 54.9 mm respectively. The highest mean annual runoff was generated from the Jack Rocks sub-catchment. The observed and predicted mean annual runoffs were 85.3 and 86.8 mm respectively. Predicted mean annual streamflow was within ±5% of observed records of all gauging stations. The rainfall has been declining systematically over the last 30 years in the south-west of Western Australia (IOCI, 2002). Due to a reduction in rainfall, the annual streamflow generated has also declined (Bari & Ruprecht, 2003). Given that the climatic trend is moving towards producing lower flow years, the model should perform adequately and represent the flow generation process for the period of simulation (1960 to 2007). The model predicted the low (10th percentile) and high (90th percentile) flows very well for all gauging stations and $R^2$ for annual streamflow ranged from 0.8 to 0.9.

The model also simulated the spatial distribution of runoff – daily, monthly and annual. Generally higher runoff was associated with the greater amount of annual rainfall received within a catchment. For example in 1988, an high-flow year, annual runoff from different Response Units ranged from 9.5 mm to 368 mm.

4.3.2 Monthly Streamflow

For the simulation period, the relationships between the observed and predicted monthly streamflow for most of the gauging stations were strong. The model occasionally poorly predicted some of the high-flow months and this was consistent at most of the gauging stations. Figure 4 shows the relationship between the monthly observed and the simulated streamflow for River Road and Jayrup sub-catchments. Overall the coefficient of determination ($R^2$) ranged from 0.75 to 0.9 for all the gauging stations within the catchment.

4.3.3 Daily Streamflow

Daily simulated and observed streamflow hydrographs matched very well for all of the gauging stations.
(Figure 5). At the Big Brook catchment, the model predicted very well the duration of the flow, peaks and recessions for 1988, a high-flow year. At the Jack Rock catchment, the predicted streamflow started about a month earlier than observed for a recent very low-flow year (Figure 5b), and then ceased flow slightly earlier as well. Overall, the model predicted the flow duration, peaks and recessions very well for all gauging stations.

Figure 5 Observed and predicted daily streamflow (a) Big Brook and (b) Jack Rock catchments

5. PREDICTIONS UNDER CLIMATE SCENARIOS

The calibrated LUCICAT model was then run to assess the change in catchment yield that could occur following a change in climate under IPCC AR4 A2 and B1 scenarios as projected by eleven GCMs (Table 1). The Morton PE data was extended to 2100 and all climate change scenarios were introduced in 2001. The Leaf Area Indices and land use remained unchanged from 2001 onward. The daily rainfall data, as downscaled from GCMs, for the periods 1971-2000 and 2046-2065 were extended respectively to cover the gaps of 2001-2044 and 2065-2080 and a continuous rainfall series of 1961-2100 period was developed. All LUCICAT model simulations started in 1971 with calibrated initial condition.

Figure 6 Canadian Climate Center Model projected within year rainfall distribution (a) A2 and (b) B1 scenarios

Figure 7 Changes in average annual (a) rainfall and (b) reservoir inflow under the B1 scenario

5.1 Rainfall

In the ‘Historical’ simulation, mean annual rainfall (1981-2000) of all 11 GCMs was 2% higher than observed and ranged between ±5%. However, there were significant variations in monthly rainfall distribution compared to observed data for the 2046-65 and 2081-2100 periods (reported as 2055 and 2090 respectively). Figure 6 shows CCM rainfall as an example but this trend was similar in all other GCMs. In the winter wet months of
June to August predicted rainfall was higher than observed during the ‘Historical’ period and was projected to decline during 2055 and 2090 climate for both A2 and B1 scenarios. Under A2 scenario mean annual rainfall reductions were projected to be 15% and 24% respectively during the 2055 and 2090 climates. Rainfall reduction under B1 scenario would be less than A2, about 12%. But there was a large variation in rainfall projections between different GCMs (Figure 7a).

5.2 Streamflow

Inflow to Serpentine reservoir and all other gauging stations were computed based on rainfall projections for A2 and B1 scenarios. During the ‘Historical’ simulation, projected mean annual inflow to the reservoir was 57 GL compared to the calibrated inflow of 38 GL. This large anomaly could be due to the within and between year distribution of rainfall (Figure 6, Figure 8) and the number of rainy days. The variations in reservoir inflows between different years were much higher than that of rainfall (Figure 8).

The response to projected climate change is evident through the reduction in flow to the reservoir for the future (Figure 8b). Compared to ‘Historical’ simulation, average annual inflow to the reservoir under A2 climate is projected to be 49% and 69% lower by mid (2055) end of the century (2090). If B2 climate prevails, the inflow reductions would be slightly lower, 45% and 46% respectively, by the mid and late 21st century. However, like rainfall, there are significant differences in projected inflow reductions between different GCMs (Figure 7b). If the present demand continues the reservoir is predicted to vary significantly due to inflow reduction under both A2 and B1 climates. The magnitude of these inflow reductions and reservoir water balance highlights serious implications for the abundance and availability of surface water resources in the south-west of Western Australia.

5.3 Water Balance

The LUCICAT outputs obtained from future climate change scenarios (A2 and B1) show that each of the water balance components would decrease further. Interflow is the largest contributor to streamflow followed by baseflow and surface runoff. Interflow was projected to remain the largest contributor in the future climate. Baseflow would have the largest proportional reduction under the future climate and can be linked to a decline in conceptual groundwater levels and soil moisture across the catchment. A reduction in groundwater levels can lead to a reduction in baseflow, stream zone saturated areas and surface runoff. These findings are similar to other studies undertaken in Western Australia (Bari et al, 2005; Charles et al, 2007).

5.4 Flow Durations

Two gauging stations, Big Brook and Jack Rocks were taken as examples to show how the flow durations will change under future climate scenarios. In the ‘Historical’ simulation, only the mid-flow events were predicted to be higher than observed at the Jack Rocks (Figure 9a) while at the Big Brook predicted runoff was higher at the mid to high-flow magnitudes (Figure 9c). Both the A2 and B1 simulations show that daily runoff is predicted to decrease under future climates and indicate that streams would flow less frequently than the ‘Historical’ period. Reduction in flow-duration has implications on stream zone ecology, flora and fauna, and environmental water allocation.

6. DISCUSSION

The use of downscaled rainfall data in modelling is still an emerging science. It provides catchment-scale information about the impact of climate change on water resources (rainfall, temperature and potential
evaporation) potentially suitable for hydrological modelling. Downscaling studies have often found that the predictors and the method of downscaling used to generate rainfall have a strong influence on the magnitude of the projected change. The marked shift and bias in intra-annual peaks (Figure 6) highlights this issue and needs further investigation.

Figure 9 Flow duration curves (a) Jack Rocks A2, (b) Jack Rocks B1, (c) Big Brook A2, and (d) Big Brook B1

The LUCICAT model was run with the calibrated parameters, and driven with projected rainfall series for each of the scenarios. Model parameters related to evapotranspiration, land use and soil properties were assumed to remain unchanged until 2100. Wood et al. (1997) proposed that if the differences between the observed and the current GCM climates are modest, then transferring the calibrated parameter set for projections would be acceptable. In this study, with a distinct difference (‘Historical’ compared to observed) in winter rainfall (Figure 6), it may be argued that the LUCICAT calibrated parameter set may not be ideal for projecting the impacts of climate change on water resources.

Temperature changes are generally included in climate change modelling studies as a change in potential evapotranspiration. While it is usually implied that an increase in temperature would result in an increase in potential evapotranspiration, recent studies have found that in some cases increases in temperature have been accompanied by decreases in pan evaporation (Roderick et al., 2009). Further work is needed to clarify how these findings relate to actual and potential evapotranspiration, and ultimately catchment yield. The response of native vegetation to a change in climate is uncertain. How plant water use changes under a warmer and drier climate needs further investigation.

Decreases in inflow to Serpentine Reservoir due to projected climate change would effectively impose further limitations on the surface water supply systems in Western Australia. Options for future sources of supply will need to consider groundwater use, increased surface water yield through better forest management, demand management, water reuse and desalination. The projected drier climate would also cause altered flow regimes and loss in biodiversity and therefore adaptive responses would be necessary to maintain ecological communities.

7. SUMMARY AND CONCLUSIONS

Streamflow to the Serpentine Reservoir supplying Perth has decreased significantly due to reduced rainfall over the last 33 years. Average annual inflow (1981-2000) was 38 GL and ranged between 20 to 69 GL. A simple distributed conceptual water balance model, LUCICAT, was calibrated and the predicted mean annual inflow to the reservoirs was 1% greater than that of other estimates. The predicted streamflow was within ±5% of the observed data for all gauging stations within the catchment.

Eleven GCMs were selected from those which contributed to the latest IPCC assessment report. Data from emission scenarios A2 and B1 were extracted from CIMP3 and then downscaled through an Analogue Method.

1627
for catchment modelling. All GCMs projected relative reductions in rainfall by mid (2055) and late (2090) 21st century compared to 1981-2000 period. There was a significant variation in projected rainfall reductions between different GCMs and emission scenarios. Under an A2 scenario mean annual rainfall reductions were projected to be 14% and 24% respectively during the 2055 and 2090 climates. Rainfall reduction under B1 scenario would be around 12% for both the time periods.

The LUCICAT model set up was then used to predict the catchment yield for all climate change scenarios. During the ‘Historical’ simulation, projected mean annual inflow to the reservoir was 57 GL compared to the calibrated inflow of 38 GL. This large difference is probably due to the within and between year distribution of downscaled rainfall and number of rainy days. Compared to the ‘Historical’ simulation, average annual inflow to the reservoir under the A2 climate is projected to be 48% and 69% lower respectively by mid (2055) and end of the century (2090) respectively. Under the B1 climate scenario, the inflow reductions would be about 45%.

The magnitude of these inflow reductions and reservoir water balance highlights serious implications for the availability of surface water resources in the south-west of Western Australia.

ACKNOWLEDGMENTS

We are grateful to the reviewers – N Plummer and C Ganter of Bureau of Meteorology, C Jeevaraj of Water Corporation and R Dixon of Department of Water, for useful comments and positive feedback.

REFERENCES

Bari, M.A., M.A. Berti, S. Charles, E. Hauck, and M. Pearcy, Modelling of streamflow reduction due to climate change in Western Australia – A case study, In MODSIM 2005, Melbourne, Australia, December 2005.,
IOCI, Climate variability and change in south-west Western Australia, Indian Ocean Climate Initiative, 34p., 2002.
Timbal, B., E. Fernandez and Z. Li, Generalization of a statistical downscaling model to provide local climate change projections for Australia. Env. Mod. & Software, 24, 341-358, 2009.
Climate change impact on agriculture:
Devils Lake basin
Andrei Kirilenko

University of North Dakota; Grand Forks, ND 58202-9011; andrei.kirilenko@und.edu

Abstract: Northern Great Plains is one of the most important agricultural regions worldwide. This is also a region expected to be heavily impacted by climate change. This and two other papers in this session concentrate on studying climate change impacts on water resources of the region, and on the impacts of these changes on agriculture. The major focus of our interest is Devils Lake watershed in North Dakota. The watershed is located in the Northern Great Plains, the area where intensive agriculture has caused an extreme change in land use and land cover, followed by substantial water pollution. Devils Lake is an endorheic (terminal) lake, which makes it especially sensitive to environmental pollutions, land use and climatic changes. Despite occasional severe droughts that heavily impact agriculture of the region, lake level has been steadily elevating since the 1940s, driven by a wetter climate phase. This paper concentrates on generation of a regional climate change scenario that would take into account the existing variability of climate parameters, on one hand, and data and structural uncertainty, on the other. We also introduce preliminary results of modelling climate change impacts on the production of spring wheat in the region.

Keywords: climate; agriculture; water; uncertainty; terminal lake

1. INTRODUCTION

Climate change, water, and food security are closely connected. Despite a huge progress in improving agricultural practices worldwide, in 2009 the number of undernourished exceeded 1 billion [FAO, 2009] for the first time. The Northern Great Plains (NGP), which include parts of North and South Dakotas, Minnesota, Montana, Nebraska, and Wyoming, is one of the principal world regions of intensive agriculture. Within the area, North Dakota is one of the most important producers of food and feed, ranking first in the nation in 14 commodities. That includes the first rank in production of red spring wheat, and the second rank in producing wheat overall [North Dakota Agriculture, 2009]. The total value of wheat production in North Dakota was almost $2.3 billion in 2008 [NASS, 2009].

It can be fairly said that the agriculture of North Dakota and the entire NGP region largely follows two climatic factors, thermal and moisture regimes, both of them frequently being far from optimal. While the region includes some of the worlds’ most fertile lands, the steep north-south temperature gradient, and east-west precipitation gradient also makes the region one of the most sensitive to climate change [Ramankutty et al., 2002]. In the paper, we study the non-adaptation changes in agriculture production in the region due to several contrasting scenarios of climate change impacts.

The specific region of interest is Devils Lake watershed, located in the NE part of North Dakota. Similar to other endorheic lakes, the water surface area and elevation in Devils Lake are highly variable. From the early 1940s, the area of Devils Lake is continuously increasing. This increase leads to flooding of residential areas, loss of agricultural lands, and deterioration in water quality. The total cost of attempts to alleviate these problems is approaching $0.5 bil. (USGS 2010; for additional information see a discussion by Zhang in this volume). Because of the existing dynamic equilibrium between the lake area, from one
side, and climate, land cover and land use, from the other, variations in the lake water level are indicative of the climate and anthropogenic pressure throughout the entire watershed.

Our main objective is to study the impact of climate change on the hydrology and agriculture of the region and, eventually, to find the correspondence between the impacts of climate change and land conversion for agriculture. This paper primarily deals with developing a set of relevant climate change scenarios. We also show how these results can be applied to study the impacts of climate change on agricultural production in the region. Other parts of the project are discussed in this volume by Zhang and by Lim et al.

2. CLIMATE

A coupled atmosphere-ocean general circulation model (AOGCM or GCM) is able to compute a multi-century climate forecast for a large set of meteorological variables, such as temperature and precipitation, at an hourly or even a minute time step, and at multiple altitudes; however, these projections can be hardly used without modification in an impact study. First, the GCM projections at a temporal scale below one month are not considered reliable (e.g., Kilsby et al., 1999). Then, the GCM simulations cover the entire globe with a grid of a few degrees latitude and longitude size; additionally, the projections are valid only for groups of cells. Finally, future climate projections differ both between GCMs and between single runs of same GCMs. Practical application of GCM projections of future climate in an impact study then requires:

- Temporal downscaling of GCM projections from monthly to daily or a finer temporal resolution;
- Spatial downscaling of GCM projections from hundreds of kilometres down to tens of kilometres or finer;
- Accounting for uncertainty in GCM integrations.

2.1 Downscaling

GCM projections are valid at a scale of multiple GCM cells, which corresponds to several hundred kilometres. All regional details, such as the effect of water bodies, altitude change, land cover and similar are lost at such a coarse resolution; additionally, the subgrid-scale processes such as cloud formation are impossible to reproduce. Multiple methods are being applied to downscale GCM projections both temporarily and spatially. The approaches range from simple interpolation to statistical downscaling to dynamic weather generators to regional GCMs (RGCMs). Simple interpolation does not introduce new details, however even simple distributing GCM projections to a finer grid in a coherent way is not a trivial task by itself. E.g., the IIASA global climate database, widely used in climate impact studies, in its early version projected 32 and more rainy days per month in some areas of the UK (the bug found and reported by the author) – a result of incorrect downscaling.

Statistical downscaling is more elaborate; it is based on finding the correlation between the measured parameters of climate (e.g., precipitation pattern) and GCM projections of current climate. The methods include multiple regression, artificial neural networks (ANN) and others. The major drawback of the method is, however, its assumption that the found correlations will stay same in the future climate. Since the method is, in this regard, based on extrapolation, it is difficult even to estimate the method uncertainty due to downscaling. An example of statistical downscaling is the Statistical Downscaling Model (SDSM, www.sdsm.org.uk), which combines a statistical weather generator for temporal downscaling with regression-based spatial downscaling.

The nested regional GCMs (RCMs) are considered by many to be the downscaling method of choice [Giorghi, 2006]. RCMs are “mini-GCMs” running on a regional scale, using GCM projections or the products of global climate reanalysis to drive their boundary conditions. An RCM then would use the fine-resolution details in forcing (e.g., topography) to model the atmospheric circulation within a small region. Currently, RCMs (e.g., the Weather
Research and Forecasting model WRF - Skamarock et al., 2007) are used frequently to simulate regional climate, and model projections are available for the impact studies on large territories. E.g., the North America Regional Climate Change Assessment program NARCCAP will eventually provide the projections of six RCMs (MCSS, RSM, HadRM3, MM5, RegCM3, and WRF) driven by four GCMs (CCSM, CGCM3, GFDL, and HadCM3) and by NCEP reanalysis. Nevertheless, there are indications that, even if computational complexity is not an issue, the RCM projections do not add skill to GCM projections [Castro et al., 2005; Rockel et al., 2008]. This is explained by strong dominating of GCM projections in RCM integrations. Alternative RCM downcasting methods may improve RCM results, generating more realistic fine-scale pattern of climate components [Lo et al., 2008]. For the purposes of our study, however, the major problem with the RCM approach is its high computational requirements, which precludes the researcher from using more than one or a handful of driving scenarios. E.g., in the (incomplete) NARCAPP simulations all GCM projections are driven by the IPCC Special Report on Emissions Scenarios (SRES) scenario A2. This drastically impacts the ability of both estimating the uncertainty and also including the uncertainty in impact projections.

2.2 Treatment of uncertainty

The uncertainties in projecting future climate are due to a variety of factors:

1. Radiative forcing depends on anthropogenic activity, chiefly the amount of greenhouse gases (GHG) released to the atmosphere from agriculture and energy production and from land use change. SRES [IPCC, 2000] quantifies this forcing by defining different paths of economic and societal development.

2. There are disagreements among the results of general circulation models.

3. The intrinsic uncertainty: a GCM running under the same radiative forcing scenario will return different results due to stochastic nature of climate simulations.

A set of simulations combining a variety of model runs under a variety of scenarios is hence obligatory for comprehensive analysis of possible impacts of climate change.

2.3 Future climate scenarios for the Devils Lake basin

We elected to use a simple statistical downscaling approach. The approach is based on using ANUSPLIN for spatial downscaling and a statistical weather generator [Friend, 1998] for temporal downscaling. ANUSPLIN [Hutchinson, 1995; Hutchinson, 2004] uses a thin-plate smoothing spline to interpolate climate variables in three dimensions. The method was used to generate a widely used very high resolution global climate surfaces at a 30” (1-km) spatial scale [Hijmans et al., 2005] and was demonstrated to produce the results similar to other popular high-resolution climatology products [Stillman et al., 1996], e.g. PRISM.

Figure 1 illustrates the process of climate projection generation. We estimate future climate conditions in the region by combining the historical data on temperature, precipitation, air humidity, and wind speed with an ensemble of GCM projections. For base climate, we use the 1971-2000 measurements. For future climate, we extract monthly projections of seven different GCMs: CCMA_T63, CSIRO, GFDL_CM2, GISS_E-R, MPI, NCAR_PCM, and UKMO, for three pre-set time periods, 2020s, 2050s, and 2080s. Additionally, for each of the GCMs we employ three SRES scenarios (A1B, A2, and B1 - IPCC, 2000). To generate samples of weather at a daily temporal resolution, we employ a statistical weather generator [Friend, 1998] and generate 30 year-long samples of climate parameters, bringing the total number of a year long time series for each climatic parameter to 630 for each time slice. In generation of this ensemble we address all three sources of uncertainty, mentioned in (2.2):

1. Scenario uncertainty is addressed through computations with three different scenarios of driving forces;

2. Structural uncertainty is addressed through computations with seven GCMs;

3. Intrinsic uncertainty of weather predictions is addressed through computations with a statistical weather generator.
2.4 Validation

We compared the simulated GCM historical climate to measured daily temperature and precipitation from the Langdon Experimental Farm station of the US Historical Climate Network (USHCN, http://cdiac.ornl.gov/epubs/ndp/ushcn/ushcn.html), which is the closest to the Devils Lake basin USHCN meteorological station. The USHCN is essentially a subset of 1218 NOAA weather stations with the highest quality of data. Langdon experimental farm station is located approximately 80 km NNE of the city of Devils Lake and 60 km NE of the center of the Devils Lake basin. Due to this shift, we can expect the temperature, measured at the station, to be slightly lower as compared to the annual temperature of the entire basin over the same period.

We compared the distribution of 1971-1990 simulated temperature and precipitation with the data of Langdon station (Table 1). There were 10855 valid points; from these points, mean temperature was 2.6°C according to measurements as compared to 3.5°C according to GCM simulations. This difference is partially explained by the northern shift of the US HCN station. More important, observed data demonstrate higher variability. This is the most pronounced during winter months, with the observed extreme low temperature (-38.3°C) being much lower than the simulated low (-25.5 - -20.8°C). At the same time, warm period temperatures in the simulated and observed data sets are much closer (Table 1). The
frequency distribution of simulated and observed temperatures over the May-October period also demonstrates similarity. Similar to daily projections, distributions of mean monthly temperature and precipitation, observed and simulated, are very similar to the already discussed distribution of daily temperatures. Again, the simulated temperature is slightly higher than the observed data, possibly due to the northern location of the meteorological station. Precipitation distribution is almost identical.

The distribution of precipitation over time is controlled by the number of rainy days. Even though GCM output contains the number of rainy days, it does so at a GCM scale, typically 50,000-100,000 km². The immediate problem is that GCM output (and also gridded historical data) contains much higher number of rain days per month than the observed data at any specific location. To model the number of rainy days per month at a specific location, we used a linear model that connects this value with the total monthly precipitation, and used the result as an input of the statistical weather generator. For warm period precipitation, GCM projections are highly variable, especially when modeling high precipitation events. The heaviest precipitation varies from 46 to 139 mm between the models, while the observed daily precipitation maximum is 105 mm. However, distribution of precipitation shows much more similarity between the GCM and observed data, as demonstrated by comparison of percentiles and Figure 2.

<table>
<thead>
<tr>
<th></th>
<th>Annual T</th>
<th>Warm T</th>
<th>Warm P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obs.</td>
<td>GCM</td>
<td>Obs.</td>
<td>GCM</td>
</tr>
<tr>
<td>Mean</td>
<td>2.6</td>
<td>3.6</td>
<td>13.8</td>
</tr>
<tr>
<td>Std.</td>
<td>14.1</td>
<td>13.1</td>
<td>6.6</td>
</tr>
<tr>
<td>Min</td>
<td>-38.3</td>
<td>-25.4</td>
<td>-13.9</td>
</tr>
<tr>
<td>Max</td>
<td>30.0</td>
<td>29.6</td>
<td>30.0</td>
</tr>
<tr>
<td>Percentiles</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>-7.8</td>
<td>-8.9</td>
<td>9.4</td>
</tr>
<tr>
<td>50</td>
<td>3.9</td>
<td>5.7</td>
<td>15.0</td>
</tr>
<tr>
<td>75</td>
<td>15.0</td>
<td>15.8</td>
<td>18.9</td>
</tr>
</tbody>
</table>

Table 1. Comparison of measured and simulated daily temperatures (T, °C) for annual and warm period (May-October) and warm period daily precipitation (P, mm). A mean of projections of seven GCMs.

Figure 2. Distribution of observed (top left) and simulated (seven GCMs) monthly number of days with precipitation over 0.1 mm, 1971-2000.

2.5 Future climate

GCM integrations (Table 2, Figure 3) agree on a moderate temperature increase in the region under the “balanced” scenario A1B: annual temperature change by 0.9 – 1.6 °C in 2020s, 1.6 – 3.3 °C in 2050s, and 2.2 – 4.6 °C in 2080s. The projections under A2 scenario are similar: temperature increase by 0.6 – 1.5 °C in 2020s, 1.5 – 2.7 °C in 2050s, and 2.6 – 5.2 °C in 2080s. The temperatures under B1 scenario, with smaller carbon emissions, but also with reduced emissions in radiation blocking aerosols, result in larger change in 2020s: temperature increase by 0.8 – 1.7 °C, but the smallest changes after that: 1.2 – 2.5 °C in
2050s, and 1.5 – 3.3 °C in 2080s. Overall, the temperature will continue to increase; but the predicted increases vary among different GCMs and for different scenarios. The actual temperature distribution is likely to fall within the range of the predictions. As opposed to this universal increase in the temperature, there is much more diversity in precipitation projections, which vary from a small decrease by up to 12% to an up to 28% increase. The width of the projected corridor demonstrates a need to use model ensemble rather than results of just one GCM or one scenario.

Table 2. Comparison between the projections of seven GCMs under three SRES scenarios. Temperature (dt) is shown as a difference, and precipitation (dp) is shown as a percentage of change to the “current climate”. No ccma_t63 A2 integrations are available from CMIP3 at this time.

<table>
<thead>
<tr>
<th>GCM</th>
<th>A1B</th>
<th>A2</th>
<th>B1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>dt (°C)</td>
<td>dp (mm)</td>
<td>dt (°C)</td>
</tr>
<tr>
<td>CCCMA_T63</td>
<td>3.1</td>
<td>8.4</td>
<td>2.5</td>
</tr>
<tr>
<td>CSIRO</td>
<td>2.3</td>
<td>1.2</td>
<td>2.7</td>
</tr>
<tr>
<td>GFDL_CM2</td>
<td>3.2</td>
<td>-2.2</td>
<td>2.5</td>
</tr>
<tr>
<td>GISS_E-R</td>
<td>1.6</td>
<td>10.9</td>
<td>1.5</td>
</tr>
<tr>
<td>MPI</td>
<td>3.0</td>
<td>4.4</td>
<td>2.3</td>
</tr>
<tr>
<td>NCAR_PCM</td>
<td>2.2</td>
<td>3.0</td>
<td>2.3</td>
</tr>
<tr>
<td>UKMO</td>
<td>3.3</td>
<td>3.5</td>
<td>2.7</td>
</tr>
<tr>
<td>Mean</td>
<td>2.7</td>
<td>4.2</td>
<td>2.3</td>
</tr>
</tbody>
</table>

Figure 3. Mean annual temperature (°C) for the Devils Lake Basin. The projections are based on two GCM integrations under A1B and B1 SRES scenarios.

3. AGRICULTURE: MODEL VALIDATION AND SENSITIVITY ANALYSIS

Decision Support System for Agrotechnology Transfer (DSSAT) was used to study the impact of climate change on spring wheat production in the North Dakota. DSSAT has been used extensively to simulate crop yields across the U.S. under current climate and climate change scenarios [Tubiello et al. 1999]. The minimum data required for DSSAT includes weather, soil and management, and cultivar-specific data, which were extracted from North Dakota Agricultural Network (NDAWN, http://ndawn.ndsu.nodak.edu), USDA SSURGO (http://soils.usda.gov/survey/geography/ssurgo), and DSSAT databases, respectively.

To validate the model, we compared the results of model simulations with the synthetic daily weather from GCM monthly climate. Several warm season-long Markov chain sequences of random daily temperature and precipitation values were consequently used with the DSSAT; the results of multiple DSSAT runs with 15 different replicas of the synthetic weather for the same growing period were then averaged. We then compared the results of DSSAT simulated yield with the yield generated when using the actual historical NDAWN temperature and precipitation for the 1991-2000 period. We found that the mean
IPCC-climate and NDAWN-climate DSSAT-simulated wheat yields for the same time frame were not statistically different (at 0.05 level). Consequent model sensitivity analysis has demonstrated that the increasing temperatures in the region (with fixed precipitation), despite an extended growing season, negatively impact the yield (Table 3).

We completed preliminary simulations of climate change impacts on wheat production under the 2050s climate, using the 15 samples of temperature and precipitation during the growing season, generated as described above from the CCCMA projections under the SRES A1B scenario. Our analysis of the results shows 15% spring wheat yield decrease under the 2050s climate. This result is consistent with the studies done in other parts of the country, which has projected spring wheat yields to decline 10% to 15% by 2040, and 20% to 26% by 2080 [Stöckle et al. 2008].

Table 3. Simulations of spring wheat yield at changing temperature and precipitation

<table>
<thead>
<tr>
<th>Change of</th>
<th>Original</th>
<th>t+1°C</th>
<th>t+2°C</th>
<th>t+3°C</th>
<th>t+3.5°C</th>
<th>t+4°C</th>
</tr>
</thead>
<tbody>
<tr>
<td>temperature</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>-2.5%</td>
<td>-7.7%</td>
<td>-7.5%</td>
<td>-9.5%</td>
<td>-11.5%</td>
<td></td>
</tr>
<tr>
<td>P-20%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-10%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Original</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P+10%</td>
<td>1.4%</td>
<td>1.8%</td>
<td>2.0%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P+20%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P+30%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. DISCUSSION

Current public discussion of reliability of climate change projections has revived an interesting problem in public perception of climate, which, to the best of our knowledge, was first noticed in the beginning of 1990s, when the first wave of public concern about “global warming” has quickly subsided. Ungar (1992) speculated that the apparent lack of public interest was due to the fact that a person tends not to notice, or even misinterpret the signals from changing climate. To better communicate the information about climate change, a number of “common sense” climate indices have been proposed [Hansen et al., 1998], however the communication evidently needs further improvement. In this respect, the cumulative nature of an endorheic watershed makes an endorheic lake an ideal candidate to serve as a proxy for climatic changes in the region.

The same impact-accumulative nature of an endorheic lake makes the nearby communities, dependent upon its resources, particularly vulnerable to even small modifications in water balance in the watershed. Well-known examples of the impacts of such modifications include the Aral Sea and Lake Chad, where the unsustainable agricultural practices in the basin and, possibly, climate modifications, has led to dramatic lake degradation. Despite an existence of long-term probabilistic forecasts of Devils Lake water level [Vecchia, 2008], the uncertainty in climate change projections remains the major challenge to water managers in the region: the hydrological forecasts based on current climate conditions cannot be relied upon anymore. The major challenge here is producing the water managing plans that are robust to climate change-related uncertainty [Stakhiv, 1998].

In our study, presented by two additional papers in this volume by Zhang and by Lim et al., we are discussing the inter-connected effects of climate change, hydrology, agriculture, and land use change in the region. Here, we present a method to build a database of regional projections of climate change, based on an analysis of a multimodel ensemble of GCM results, which would take into account the scenario, the data, and the structural uncertainty in climate projections. Then, we show how this database can be used to project the consequent changes in performance of agriculture. This preliminary study includes studying the variation of wheat yield under one scenario of climate change; on the next step the methodology will be applied for a variety of the scenarios of socio-economic development, GCMs, and crops.

ACKNOWLEDGMENTS

This study was supported by a grant “The Feasibility of Combining NASA Satellite Data, General Circulation Models, and Hydrologic Models to Inform Decision Making for Flood Mitigation of the Devils Lake Basin of North Dakota” from NASA.
REFERENCES


Development of an Agent Based Coupled Economic and Hydrological Model of the Deep Creek Watershed in British Columbia, Canada

John Janmaat1 and Markandu Anputhas2

1Economics, UBC Okanagan, Kelowna, British Columbia (John.Janmaat@ubc.ca)
2Interdisciplinary Graduate Studies, UBC Okanagan, Kelowna, British Columbia (Markandu.Anputhas@ubc.ca)

Abstract: Deep Creek is an important tributary of the Okanagan River system in Southern British Columbia. The creek drains a watershed with sediments deposited by multiple glacial events, resulting in a complex pattern of surface and groundwater connections throughout the basin. Further, the basin is experiencing changes in agricultural enterprises as well as population growth, and may be heavily impacted by climate change. The MIKE SHE integrated hydrological modeling system is being used to develop a model of the basin hydrology. The watershed is relatively small – with only a couple of hundred agricultural water users – lending itself to a high resolution modeling project. We are at the initial stages of developing an Agent Based model to forecast land use change in the basin. The MIKE SHE model will be used to calibrate the hydrologic features of our model, and the water use predictions of our model will serve as inputs into the MIKE SHE model, as an iterative process to forecast the impacts of climate change in the basin. Conventional climate change impacts on precipitation and temperature will be considered, as well as various changes in input and output prices which may occur. The latter have been little studied, although there is general agreement that agricultural production will shift to more northerly regions, and that this will be driven by price changes as well as new opportunities. After considering various approaches and software options, we decided that REPAST provides the power and flexibility to meet our objectives. We are presently at the preliminary stages. This paper outlines the study location and preliminary conceptual design of the model.

Keyword: Agent based modeling; Climate change impact; REPAST; Simulation; Water demand.

Introduction

Land use patterns are the result of a myriad of individual choices made in response to a diverse range of variables. As these variables change, land use patterns evolve. On top of the usual suspects, climate change is likely to impact on many of the variables that are important to land use choice. This paper documents initial efforts to develop a model that will connect hydrologic process and land used decisions into a tool to understand the impacts of climate change on a watershed in British Columbia. This project is part of the continuation of a long running effort to understand the hydrology of the Deep Creek watershed.

It is doubtful that a relatively simple explanation of why we transform the environment the way we do will be forthcoming (Turner et al, 1990). A theory of land-use change needs to conceptualize the relations among the driving forces of land-use change, their mitigating processes and activities, and human behavior and organization. Further,
agricultural land use changes in particular are influenced by individual choices of farmers. A farmer’s land use choices are responses to personal, social, institutional and market factors, made in the face of the opportunities and constraints resulting from the farmer’s location. Simple models that ignore the idiosyncratic aspects of each farmer’s situation may reproduce the general features of land use change, but are unlikely to capture local details. When local details matter, such as where sensitive habitats exist, such results may be of limited value.

Initial activities within the current project focused on understanding the hydrology of the watershed, with particular focus on developing a detailed model of the groundwater resources and their relationship to surface waters. To forecast climate change impacts, a forecast of water demand is required, and this in turn requires understanding how land use will evolve. Land use will respond to several factors. One is certainly the change in opportunities that results from climate change. However, population growth pressures and changing economic conditions – some of which are also a consequence of climate change – will also play an important role. Our objective in the current research is to develop an appropriate model to forecast how land use will change in response to these forces, and an agent based model is the most suitable model for this task.

**Study Area**

The semi arid Okanagan region lies in the southern interior of British Columbia, in the rain shadow of the Coast Mountains. Deep creek and Fortune creek watersheds are two of many watersheds in the Okanagan valley, respectively covering 230 km² and 160 km² (Figure 1 and 2). Two small urban areas Armstrong and Spallumcheen are located in the watershed. The elevation of the bottom part of the basin ranges between 340 – 520 meters while upper part of the basin ranges from 370 – 1575 meters above sea level (Wei, 2009).

![Figure1. Location of the Okanagan Basin in British Colombia, Canada.](image)
Winters are short and cold, while summers are hot. Annual average precipitation and the potential evaporation are within the range of 475-575 mm and 850-950 mm respectively while the mean annual temperature is within the range of 6-10 °C. Much of the winter precipitation falls as snow, particularly at higher elevations. There has been a measurable increase in precipitation, particularly winter snowfall. (Wei, 2009). Temperatures have also increased, most noticeably daily minimums (Cohen and Kulkarni, 2001). The frost free period increased by nearly 3.1 days per decade during the 20th century, and now ranges from 120 – 150 days (Zbeetnoff, 2006). For more southerly parts of the Okanagan, Nielsen et al (2001) expect climate change to significantly alter the ranges for horticultural crops, moving north and to higher elevations. While perennial horticultural crops are not presently important in the north Okanagan, climate impacts are likely similar. Farmers will modify their practices in response to the opportunities and threats created by the changing climate. Further, many crops in the Okanagan rely on microclimates created by the complex topography (Cohen et.al, 2006). These facts further emphasize that predicting the impacts of climate change on land use in the north Okanagan must consider the unique aspects of the landscape and the importance of individual responses to these changes.

The availability and access to water is the vital factor for agricultural activities in the semi arid climate regions. Deep and Fortune creeks have a limited ability to supply water over much of their range, making groundwater particularly important. The study area has seen multiple glaciations, leading to a complex geology and consequently a complex hydrology. Groundwater head levels vary between 350 – 500 meters above average mean sea level through the valley bottom. Over the last 30 years, groundwater levels have dropped in many of the aquifers from which water is withdrawn. Head levels drop in the summer due to irrigation pumping in all shallow and moderate aquifers. The sensitive range of water depth varies from 50 to 70 meters. Mountain system recharge is a major contributor to the deep regional aquifers, from which there is recharge of some overlaying aquifers and in places artesian flows (Wei, 2009). Given the critical role of
water, coupling a land use and hydrologic model is necessary to forecast the land use impacts of climate change.

Important activities in the study area include forestry, agriculture, manufacturing and tourism, with agriculture the most important. Cattle farming is the most prominent, representing about 21% of all operations. Other common farm types include hay and forage (17.2%), horse and pony (15.2), poultry (6.8), and dairy (6.1%) (Zbeetnoff, 2006). In response to high land prices near Vancouver and a growing population of retired residents, agriculture has been shifting towards more commercial animal agriculture, and smaller hobby farms. Such changes, driven by economic and demographic trends, are factors that will affect how we respond to climate change, and should be incorporated into land use change modeling.

Alternative Land Use Models

Two dominant approaches in land use/cover change modeling are pattern based models and process based models. Pattern based models are mainly GIS oriented models which forecast spatial patterns, and generally have minimal behavioral content. Numerous spatially disaggregated and heterogeneous land use change models exist, taking advantage of the vast amounts of spatially disaggregated land use/cover data that are now available (Pontius and Malanson, 2005; Schotten et al. 2001; Verburg et al. 2004; Pijanowski et al. 2005; Verburg and Veldkamp 2004). Spatially-explicit LUCC models typically begin with a digital map of an initial time and then simulate transitions in order to produce a prediction map for a subsequent time. Such approaches do not model the decisions of individual land owners. They therefore cannot incorporate responses of land owners to economic and social changes that are not already driving pattern changes, nor can they effectively include policies that use economic incentives to change land use decisions.

The alternative process based models explicitly model the behavior of individuals. These models allow the incorporation of the human decision making process in a formal and spatially explicit way, incorporating the influence of social interaction, responses to local conditions, etc. on decisions. This approach can explicitly incorporate a range of social processes – interactions among human ‘agents’ – and environmental processes in a spatially explicitly way. There have been a number of recent implementations (Kirtland et.al. 2000; Lei et.al. 2005; Polhill et.al. 2001), The Mathematical Programming based Multi Agent System – MPMAS (Berger, 2001) is one of a small number that consider changes in agricultural landscapes. Given the spatial heterogeneity of the physical landscape, the sensitive nature of some local environments, the relatively small size of the watersheds, the presence of detailed data on the physical features and in particular the groundwater processes, agent based modeling is an appropriate tool for our study area. However, our intention to incorporate spatially explicit features at a smaller scale than MPMAS, together with the need to integrate and utilize GIS data makes MPMAS an inappropriate tool. We have therefore chosen REPAST, the RECursive Porous Agent Simulation Toolkit (North et.al., 2005a), for our model.

Implementation in REPAST

Even though REPAST has not been used much in agricultural land use simulation, it has the flexibility to handle any kind of simulation requirement. REPAST borrows many concepts from the SWARM agent-based modeling toolkit. However, unlike SWARM, REPAST is implemented in JAVA, proving both a well known language, and access to a vast collection of software that can be incorporated into REPAST models. In particular, REPAST is one of the few simulation / modeling software systems that
supports the integration of geospatial data, especially that of vector-based geometries and ability to represent dynamics (Crooks, 2006). One can obtain further information about REPAST from its web link.

A REPAST simulation is composed of agents and a space, bound together in a model. All three (Agent, Space and Model) are JAVA classes. The space class is flexible enough to implement a cellular automaton hydrologic model, enabling one key aspect of the study area to be explicitly incorporated into the simulation. The agent class itself also enables a rich decision set to be considered, as well as sophisticated learning processes and a range of interactions between agents. Finally, the model class can manage both sequential and simultaneous action time steps, and as a programmed class, can embody interactions that are customized to the specific case under study. Figure 3 illustrates the basic design elements of the model we are developing.

Figure 3: Basic design elements of the Agent Based Model (ABM) simulation.

**Space Class**

The space object will be a cellular automaton implementation of a hydrological model of the study area. Each cell will contain information relevant to determining the productive potential of the surface land represented by that cell, and the water table level in the cell. As the focus is on the impact of climate change, the detailed modeling of each cell will be restricted to those variables that will be impacted, or that relate to other variables that will be impacted by climate change, and/or that otherwise restrict the potential uses of the land, relative to other locations in the study area. Soil type – clay, sand, gravel, etc. – impacts crop growth, soil temperature, and water retention. Slope and aspect of the surface affects radiation received. Elevation determines the temperature regime. Location of the cell within the watershed will also determine the precipitation...
received. Table 1 lists a number of characteristics of each cell in the surface class of the model that we are planning to incorporate.

Table 1: Characteristics and processes for cellular automaton surface class.

<table>
<thead>
<tr>
<th>Static Characteristics</th>
<th>Dynamic Characteristics</th>
<th>Interactive Processes</th>
</tr>
</thead>
<tbody>
<tr>
<td>- slope</td>
<td>- organic matter</td>
<td>- depth to water table</td>
</tr>
<tr>
<td>- aspect</td>
<td>- soil nutrients</td>
<td></td>
</tr>
<tr>
<td>- elevation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- top of aquifer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- bottom of aquifer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- soil type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- coordinates</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The ground and surface water model will implement a Darcy’s Law style of relationship between the cells, focusing on the relationship between the upper aquifers from which water is withdrawn for irrigation, and which interact with the surface water in the watershed. Important connections between the modeled water flows and deeper aquifers, mountain recharge areas, etc. will be incorporated as constant flux and/or head boundaries. Parameter values will be chosen to approximate the results generated by Wei et al (2009), in their MIKE SHE based simulation of the hydrology of the Deep Creek watershed. Wei et al’s simulation will be used to calibrate the hydrologic layer of our model. Water use projections from our ABM model will serve as inputs into the MIKE SHE model for trial scenarios. Iteration of this process will continue until water use decisions by the agent and its impacts are consistent with the MIKE SHE hydrologic model.

Agents Class

Agents make management decisions for a collection of cells. In the figure 3, all cells owned by an agent are contiguous, but this need not be the case. Agents choose the enterprise to engage in (livestock production, crop production, etc.), which crops to grow, and the levels of inputs to use. Of particular interest is water use. Agents decide how much water to apply to the land, in response to the land use decision previously made and the climate conditions. Table 2 lists some of the characteristics, choices, and interactive processes at the agent level that we intend to incorporate.

Table 2: Characteristics, choices and interactive processes for the agent class.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Choices</th>
<th>Interactive Processes</th>
</tr>
</thead>
<tbody>
<tr>
<td>- demographics</td>
<td>- enterprise</td>
<td>- expectations formation</td>
</tr>
<tr>
<td>- assets</td>
<td>- crop</td>
<td>- input markets</td>
</tr>
<tr>
<td></td>
<td>- water use</td>
<td>- output markets</td>
</tr>
</tbody>
</table>

The model will initially be developed with farmer agents only, and even there with the focus first on the choice of crop and pumping level alone. The sophistication of the agent will then be expanded to include multiple enterprises and then permit the trading of land between agents. With a land market, a second type of agent will be introduced, the urban resident or city agent. This will enable population growth to be incorporated into
the simulation. If implemented as individual urban resident agents, then these agents derive a relatively small share of their income from the land use choice. In the spirit of the monocentric city model, these agents will also put a large cost on distance from the center of the relevant urban location. Alternatively, a city agent can be added for each urban center, with must enter the land market to expand and accommodate population growth.

**Farm equation**

Agent decision making falls into two broad categories: optimization or heuristics. While often strongly contrasted, both of these approaches assume goal-oriented behavior, and become more realistic when using detailed production and consumption functions (Schreinemachers and Berger, 2006). Optimization models have certain advantages for land use modeling, such as the inclusion of multiple input and output decisions, solving the decision problem simultaneously, and clearer policy relevance of outcomes. Among the possible functional forms, linear functions and fixed input/output relations are generally seen as unrealistic, while quadratic and Cobb-Douglas functions are more widely accepted (Mundlak 2001). With an input like water, where surplus will harm the crop, the quadratic form is particularly attractive.

Alternative functional specifications are also possible, and we will investigate those used in the MPMAS (Berger, 2001) system, as well as linear programming and other constrained optimization approaches. These various approaches will be assessed for their ability to incorporate crop response to water availability, the choice of water use, etc, and the ease of making them sensitive to physical characteristics of the agent’s location on the surface. The speed of calculation will also be considered, an important component in ensuring that the model remains tractable.

**Model Class**

REPAST’s model class implements the schedule that governs the evolution of the model. It is also the logical level at which to base processes that are not a consequence of the evolution the surface or the interactions between the agents located on this surface. This is where climate change scenarios and scenarios for economic variables impact on the model.

Climate scenarios have been down-scaled for the Okanagan by Nielsen et al (2001). The local predictions embodied in these scenarios will determine the agricultural potential for each land parcel that an agent in the watershed owns, as a consequence of the soil characteristics, aspect, elevation, etc. of that parcel. The commodity price scenarios that correspond to the climate scenarios are a novel contribution of this project. The fact that climate change means a shift in agricultural production from tropical to temperate regions has been recognized for some time (Rosenweig and Parry, 1994; Parry et al, 1999). These shifts in production will only occur if agricultural producers in temperate regions find it profitable to expand production. However, Schmidhuber and Tubiello (2007) find only three analyses that have forecasted the impact of climate change on food prices, and these were limited to international trade. Using works such as those cited by Schmidhuber and Tubiello as guides, we will develop price scenarios to accompany the climate scenarios for important crop options in the watershed.
Table 3: Key elements of model schedule.

<table>
<thead>
<tr>
<th></th>
<th>Winter</th>
<th>Spring</th>
<th>Summer</th>
<th>Fall</th>
</tr>
</thead>
<tbody>
<tr>
<td># steps</td>
<td>1</td>
<td>2-3</td>
<td>6+</td>
<td>2-3</td>
</tr>
<tr>
<td>Agent</td>
<td>- predictions</td>
<td>- crop choice</td>
<td>- irrigation</td>
<td>- harvest crop</td>
</tr>
<tr>
<td></td>
<td>- enterprise choice</td>
<td>- plant crop</td>
<td></td>
<td>- realize revenues</td>
</tr>
<tr>
<td></td>
<td>- land market</td>
<td>- pay costs</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3 illustrates the schedule planned for the model. The resolution of the schedule will be finest in the summer and coarsest in the winter. During the late fall, winter, and early spring there is no irrigation activity, and water movement is mostly unaffected by human actions. In the summer, irrigation is the most intensive, and environmental impacts of irrigation are the most significant. Therefore, to track these impacts, resolution is highest.

For the model agents, winter is the period when expectations are formed and the large enterprise decisions are made. For simplicity, the decision to purchase or sell land will also be made during this season. In the spring, the agent chooses the land use and pays the associated costs. Irrigation and any other relevant input choices are made in the summer. In the fall, crop is harvested and revenues are realized.

Progress

At this point, a basic pumping agent model on top of a simple uniform surface cellular automaton hydrologic model has been implemented in REPAST. The hydrologic system responds to constant head sinks and sources located in the grid, and to pumping activity by the agents. Agents choose pumping level in response to pumping costs, which are a function of the head level in the cell in which they are located. Beyond the proof of concept, two key insights have emerged from this work. First, to smoothly manage the implementation of Darcy’s law, calculations must be done in two steps. In the first step, inflows and outflows from each cell must be calculated in relation to all neighbouring cells. Once calculated for all cells, the head level in each cell is adjusted in response to this water movement. The second insight relates to water pumping. With the simple implementation, where there is only one time step per year, agents cannot be programmed to optimize pumping choice. Rather, agents must be programmed as adjusting pumping level. When agents instantaneously optimize, the system enters an oscillating pattern where agents pump a large amount in one period, and then nothing in the next. In a more sophisticated model, where irrigation is decided after the crop is planted and a number of pumping decisions are made during the growing season, this problem may not exist.

The results of this analysis will form a key input into the ongoing modeling of the Deep Creek watershed. The land use change scenarios will inform the water use throughout the watershed, and serve as input into the MIKE-SHE model. Given that the models will run relatively separately, an iterative process will be required to find a land use and resulting water use scenario that is consistent with the hydrologic model.

Conclusions

The Deep Creek watershed in the North Okanagan has been subject to a detailed hydrological modeling effort in recent years. The area is expected to experience
significant climate change in the decades to come, as well as continuing pressures from population growth, and other economic pressures originating elsewhere in the world. Given the importance of spatial heterogeneity, an agent based approach is the best way to model land use change and the evolution of water demand in the watershed. After reviewing several different ABM tools, we have concluded that REPAST is the best tool for implementing our model. A proof of concept implementation has demonstrated the feasibility of using this tool, and highlighted some of the issues that will need to be accounted for.

References


http://www.obwb.ca/
Effects of small reservoirs on large scale water availability

Maarten S. Krol1, Marjella J. de Vries1, Pieter R. van Oel1, José Carlos de Araújo2.
1: University of Twente, Dpt. of Water Engineering and Management, Enschede, Netherlands, 2: Universidade Federal do Ceará, Fortaleza, Brazil
e-mail corresponding author: m.s.krol@utwente.nl

Abstract: Small reservoirs impact on large scale water availability both by supplying water in a distributed sense, and by subtracting water from large reservoirs with large scale water supply functions. Water availability from small and large reservoirs is subject to climate change, which has the potential to alter the relation between small and large reservoirs. For a case study area in semi-arid Brazil, water supply from small reservoirs is found to be more sensitive to climate change than water supply from large reservoirs. Changes in evapotranspiration are more consistent amongst models and its' effects may dominate effects by more uncertain changes in precipitation. The reduction in water availability from large reservoirs by upstream abstractions by small reservoirs is found to be significant and may increase with unfavourable climate changes, even in an absolute sense.

Keywords: small reservoirs, water availability, semi-arid, climate change

1. INTRODUCTION
In semi-arid river basins, the presence of small-scale reservoirs is often numerous. This paper assesses large scale (river basin scale) effects of small scale reservoirs. Relevant large scale effects of these reservoirs are basically twofold:
- firstly, small reservoirs retain water, enabling local distributed usage, and thereby subtract it from availability to the largest and concentrated uses downstream in the river basin, e.g. facilitated through a large scale reservoir downstream in the basin,
- secondly, small reservoirs enhance distributed water availability, an effect accumulating to a large scale effect at the scale of the river basin, through the large numbers of reservoirs, and through its' dominant role as the source of water availability in locations, relatively upstream in the river basin.

In North-East Brazil, and various other semi-arid regions, small scale reservoirs are so numerous, that their effect is expected to be significant [De Araújo and González Piedra, 2009; Ngigi, 2003; Van Oel et al., 2008]. In a basin, reservoir density is considered high, when reservoir capacity exceeds 40% of average runoff [USBR, 2002], a situation plausibly nearing basin closure [Molle, 2004], as an indicator for unsustainable human interference in the river basin. In the same time, however, the representation of small reservoirs in large-scale hydrological models is, at best, implicit [Günther et al., 2004]. This paper investigates an explicit representation of small reservoirs for evaluation of water management at a large scale, using a case study under current and possible future climate conditions. The case study concerns the Bengue catchment in North-East Brazil [Gaiser et al., 2003]; regional climate change scenarios base on results of modelling studies reported on in the IPCC-TAR [IPCC, 2007].

The goals of reservoirs for water storage are manifold, ranging from guaranteeing water supply, via generating electricity or protecting against flooding up to perennizing river flow [WCD,
These goals are generally pursued within the delineation of a river basin, which often shows a fractal-like structure, where the basin can be divided into sub-basins at many different scales; the goals of reservoirs can be pursued at the various scales, consistently involving reservoirs of various scales [van Oel, 2009].

River basin management (RBM) and integrated water resources management (IWRM) call for an integrated consideration of water related issues, societal stakes as well as aspects of spatio-temporal distribution in designing and implementing water-related policies. Drawing on common pool resources theory [Ostrom et al., 1999], factors enhancing the manageability of reservoirs depend on scale in various, opposing ways [van Oel, 2009], and may have significant externalities towards downstream [Van Oel et al., 2008].

Tools to support the design of such policies include physical water balance models of the river basin, implicitly or explicitly representing hydrological responses, water use issues and operation of water infrastructure. Here, we use a hydrological model to drive the explicit routing of water in a river – reservoir network. Simulations are interpreted to large scale water availability, both as large scale distributed availability, and availability in large scale reservoirs for usage downstream.

The effect of small reservoirs is addressed by comparing simulations for situations with and without small reservoirs. The analysis is performed under a set of climate assumptions, adding potential changes in both precipitation and evapotranspiration.

The analyses are done for a case study basin in semi-arid North-East Brazil. The performance of the hydrological schematizations is evaluated by comparing simulations to monitored storage in the large basin downstream.

Conclusions are sensitivity to developments in small reservoir numbers, and to different dimensions of climate change; the general applicability of the lessons learnt from this exercise is discussed.

2. HYDROLOGICAL MODEL AND INDICATORS FOR WATER AVAILABILITY.

The WASA model [Güntner, 2002; Güntner and Bronstert, 2004], developed mostly for the Brazilian state of Ceará, describes the water cycle in river basins, simulating the water balance of soils and water reservoirs, generation of run-off and routing of flow through the river network, in which the reservoirs are located.

The vertical hydrology is represented on spatial units, defined by properties concerning soil, vegetation cover and topographic position in the terrain. These units are treated as sub-basins. Soil moisture, in a location-dependent number of soil horizons, changes due to infiltration of precipitation, runoff generation, evapotranspiration, deep percolation and moisture redistribution. The model allows for lateral exchanges between units. Runoff results at surface and subsurface levels when soil moisture exceeds the water holding capacity, or when rainfall exceeds infiltration capacity.

Table 1. Spatial data used in the WASA model [Güntner, 2002]

<table>
<thead>
<tr>
<th>Theme</th>
<th>Scale</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topography Model</td>
<td>Digital Elevation</td>
<td>grid spacing 30 arcsec</td>
</tr>
<tr>
<td></td>
<td>Vegetation</td>
<td>1:1 million</td>
</tr>
<tr>
<td></td>
<td>Soil associations, landscape units</td>
<td>1:1 million</td>
</tr>
<tr>
<td></td>
<td>Geomorphology, topography</td>
<td>1:1 million</td>
</tr>
<tr>
<td></td>
<td>Climate</td>
<td>interpolated point data</td>
</tr>
<tr>
<td></td>
<td>Reservoirs</td>
<td>municipality</td>
</tr>
</tbody>
</table>

Runoff is described to accumulate in a stream network at a sub-basin scale and to become available by fractions for storage in smaller reservoirs at that scale. These smaller reservoirs are represented in 5 size classes (from <0.1 Mm3 to 10-50 Mm3) in a lumped way, and each class is assumed to receive 1/6th of the locally generated runoff; the last 1/6th flowing into the main
river network connecting the subbasins. Outflow or overflow of small reservoirs is routed to be equally divided over the larger reservoir classes and the main river network. Eventually, runoff and reservoir overflow / outflow enter the river network. Dams larger than 50 Mm³ are represented to intersect the main river network and may thereby store flow originating from upstream sub-basins as well. Reservoir water balances of both smaller and larger reservoirs account for inflow, direct precipitation, evaporation, infiltration, withdrawal from the reservoir, controlled release and overflow. The WASA model is largely process-based. Regional data availability at the resolution of the model was very limited, making the application of conceptual or empirical models (like HBV [Bergström, 1992] or the SCS-approach [SCS, 1972]) very difficult in this context, reducing their ability to reliably represent the water balance of the region under strongly changing conditions. Process descriptions in the model were chosen on their particular applicability under semi-arid conditions, leading to deviations in descriptions compared to e.g. Topmodel [Beven and Kirkby, 1979] or SWIM [Krysanova et al., 2005].

![Figure 1. Benguê watershed [Creutzfeld, 2006].](image)

The model was applied to the Benguê catchment, a 933 km² area with the 19.6 Mm³ Benguê-reservoir at the outlet [Creutzfeld, 2006]; the reservoir is designed to have a water yield of 200 l/s. Within the catchment, 133 reservoirs are identified, of which 113 have a capacity below 0.1 Mm³, 17 between 0.1 and 1 Mm³, one between 1 and 2.5 Mm³, and two between 5 and 10 Mm³; their total capacity (excluding the Benguê-reservoir) is about 21.5 Mm³. In the current paper, runoff is routed to and between reservoirs in an explicit scheme, based on the locations of the reservoirs and a geohydromorphological analysis of the basin using the Digital Elevation Model. The model used meteorological data from [FUNCEME, 2008; Gerstengarbe and Werner, 2002]. Annual precipitation averages 565 mm (ranging between 210 and 1265 mm from 1979 to 2006), while potential evapotranspiration averages 2200 mm. The combination of low precipitation, high infiltration and large reservoir capacity imply a categorization of the basin hydrology as highly affected by the small reservoirs using the definition in [USBR, 2002] entailing a threshold of 40% of runoff for total reservoir capacity. A comparison of observed and simulated values for storage in the Benguê reservoir for the period 2000-2005 shows, that the geo-explicit model with the input data used yields a
qualitatively reasonable but quantitatively quite poor simulation of the dynamics of reservoir storage (Figure 2), possibly due to biased precipitation/local runoff data.

![Storage volume in Bengue reservoir](image)

**Figure 2.** Volume of water stored in the Bengué reservoir, as observed [ANA, 2005; COGERH, 2002; 2003] and as simulated using hydrologic input from WASA [Güntner, 2002] and explicit routing through the 133 reservoir network. Grey and white bands indicate hydrological years.

Models simulations are interpreted as water availability in two ways, relating to the two aspects of large scale water availability distinguished here.

1. The water availability related to large-scale use downstream, achieved through a strategic reservoir, is often referred to as reservoir yield and quantified as the constant controlled outflow that can be guaranteed in 90% of the years, depending on the stochastic characteristics of inflow. An algorithm to determine this reservoir yield, using statistical properties of reservoir inflow, precipitation and evaporation, and based on the regionally applied method of Campos [Campos, 1987] is described in [Van Oel et al., 2008].

2. The large scale distributed water availability within the basin is assessed using the total water storage in the smaller reservoirs, upstream from the Bengué-reservoir as an indicator. The sensitivity of large scale water availability for climate change was assessed by evaluating both indicators for current climate and under possible future conditions.

### 3. CLIMATE SCENARIOS AND REGIONAL MODEL SKILL

Climate scenarios were drawn from the third (TAR) and fourth (AR4) IPCC assessment reports [IPCC, 2001; 2007]. The skill of global circulation models (GCMs) to represent the regional climate of the semi-arid North-East Brazil is evaluated by comparing the simulations of annual and dry season precipitation to the observed climatology: models contributing to the third assessment report [IPCC, 2001] have been analysed earlier [Krol et al., 2003]. Here evaluations of models contributing to the fourth assessment report were added. The climatology used in the comparison is [New et al., 1999].

The skill of a model, to represent regional climate, is here evaluated by its simulations of annual precipitation and dry season precipitation (and thus its seasonality), and defined as

$$
\text{Model \_ skill} = 1 - \max \left( 4 \frac{\text{abs}(\text{annprec}_{\text{GCM}} - \text{annprec}_{\text{obs}})}{\text{annprec}_{\text{obs}}} , 2 \frac{\text{abs}(\text{dryprec}_{\text{GCM}} - \text{dryprec}_{\text{obs}})}{\text{dryprec}_{\text{obs}}} \right)
$$

where `annprec` denotes annual precipitation, `dryprec` dry season precipitation (June – November) and the values 4 and 2 are weights for the severity of model deviations.

Consistent with the observations in [Krol et al., 2003], Figure 3 shows that, for the GCMs contributing to the TAR, only a limited number of models shows a positive skill, and that these models inhibit strongly deviating climate change signals, ranging from -50% to +20% change in
annual precipitation by the year 2100 for a high emissions scenario. For the AR4-contributing models, again a small number of models show a positive skill for representing the regional climate of Northeast Brazil. The models with a reasonable skill however, are unanimous on a modest climate change effect in annual precipitation, projections ranging from -8% to +6% by the year 2100 for a high emissions scenario. Strong regional precipitation change signals are still found in model simulations, ranging from -35% to +35% by the year 2100 for a high emissions scenario, but these signals result from models with a modest model skill in representing NEB climate.

![Change in NE-Brazil precipitation (2100-high emission) as a function of model skill](image)

**Figure 3.** Changes in precipitation in semi-arid Northeast Brazil, as projected by GCMs, as a function of model skill to represent regional climate. Results were taken of models contributing to IPCC assessments; TAR denotes the third assessment report [IPCC, 2001], AR4 denotes the fourth assessment report [IPCC, 2007].

Next to the precipitation change, changes in other meteorological variables may significantly affect the functioning of surface storage facilities. Particularly, as projected precipitation changes have become modest in skill-full models contributing to AR4, we here address the effect of changes in potential evapotranspiration $PET$ and reference evaporation $ET_0$, for which models consistently project increases of up to 15% by 2100. In the WASA model, increases in evapotranspiration affect the water balance in the spatial units in the area, reducing run-off to the stream and river network and the reservoirs. Moreover, increases in reference evaporation increase the evaporative losses from the reservoir itself, as is accounted for both in the WASA model and in the adapted Campos method to determine reservoir yield. The climate scenarios to assess regional sensitivity are chosen as firstly, a 10% reduction in precipitation, being the worst precipitation trend for reasonably skilled models; secondly, a 15 increase in reference evaporation, and thirdly, the combination of these two changes. Additionally, to assess the influence of the impact of small reservoirs on large reservoirs, simulations were done in the original setting, and for a situation without small reservoirs.

4. **RESULTS**

Simulations of the hydrological model, with routing into and through the reservoir network, yield an appreciable sensitivity of both aspects of large scale water availability to potential climate changes.
The distributed large scale water availability, through storage in small reservoirs, is almost equally sensitive to the plausible reduction in precipitation as to the plausible increase in evaporation, with a combined effect of a loss of over half of the stored water volume (Table 2). The large scale availability (largely towards downstream) in the strategic reservoir is found to be appreciably sensitive too, with the effect of both factors similar as well, combining into a reservoir yield loss of over one third. The distributed availability is found to be more sensitive to climate change than the availability towards downstream is.

<table>
<thead>
<tr>
<th>Table 2. Water storage in small reservoirs and large reservoir yield under climate change.</th>
</tr>
</thead>
<tbody>
<tr>
<td>storage in small reservoirs</td>
</tr>
<tr>
<td>yield strategic reservoir</td>
</tr>
</tbody>
</table>

The effect of the small reservoirs on the strategic reservoir can be observed from the reservoir yield of the strategic reservoir. Table 3 illustrates how small reservoirs in the Benguê catchment reduce the potential reservoir yield by 10% for the current climate, and by up to 17% for a plausible future climate with decreased precipitation and increased evaporation. In absolute terms, the negative impact of the small reservoirs on the yield even increases from 45 l/s to 55 l/s, when the climate change would materialize, even while this climate change would reduce the yield already. The negative effects of climate change and upstream losses thus enhance each other in a way that supersedes linear superposition.

With the current density of small reservoirs, the plausible potential impact of climate change is larger than the effect of the small reservoirs (about twice as large).

<table>
<thead>
<tr>
<th>Table 3. Reservoir yield of the strategic reservoir, depending on climate and small reservoirs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>original climate</td>
</tr>
<tr>
<td>no small reservoirs</td>
</tr>
<tr>
<td>with small reservoirs</td>
</tr>
<tr>
<td>-10%</td>
</tr>
</tbody>
</table>

5. CONCLUSIONS
Climate change projections for North Eastern Brazil using global GCMs contributing to the IPCC-AR4 [IPCC, 2007] still vary from significant decreases to significant increases in precipitation. Contrary to the GCMs contributing to the IPCC-TAR [IPCC, 2001] however, the models contributing to IPCC-AR4 with the best skill in representing the regional semi-arid climate of North-East Brazil, tend to show modest trends in regional precipitation. Impacts of climate on drought and on the large-scale effect of small reservoirs may depend at least equally strongly on the changes in evaporation and evapotranspiration (as affected by temperature, humidity, and radiation) as on the changes in precipitation. Skills of current global climate models to represent the sub-continental climate of Northeast Brazil does not allow for firm conclusions, but tend to imply an increase in dryness and in the large-scale effects of small reservoirs.

Both water availability by storage in small reservoirs and the yield of large reservoirs are strongly sensitive to plausible climate changes.

Large scale distributed water availability, supplied by storage in small reservoirs is more vulnerable than large scale availability towards downstream, supplied by storage in large scale strategic reservoirs.

Distributed water availability may halve, due to climate change, indicating that small reservoirs could only sustainably supply water, if water use intensities drop or reservoir numbers increase. Increasing the number of reservoirs regionally is the historically common way to react to
undersupplies of water; this strategy would be deemed ineffective however, as even for current climate reservoir capacity is of the same order of magnitude as accumulated basin runoff.

Large scale water availability may reduce by over one third, even when assuming that upstream users do not guarantee their water supply by increasing storage capacity. Downstream water use may therefore be sustainably met, if use intensities drop, or additional water supply sources are installed.

Small reservoirs have a significant effect on the water yield of the strategic reservoir, and this effect is simulated to even increase under disadvantageous climate changes. The combined effect of small reservoirs and climate change is larger than the superposition of their individual effects.

Scenario studies of regional development [Döll and Krol, 2002], including both changes in population, land use and climate change, indicate that both concentrations of activities in locations more downstream in the basin, as distributed developments may occur. In all scenarios with disadvantageous climate developments, large dams or inter-basin transfers are suggested to be plausible, and to possibly compensate for the increases in demand.

Under disadvantageous climate development, upstream small scale water users may be expected to compensate for their loss in water availability by increasing the upstream storage capacity in small reservoirs [van Oel, 2009]. This would enhance the climate change effect on strategic reservoirs.

Acknowledgements
The authors thank the SESAM project, in particular Benjamin Creutzfeldt and Eva Müller (University of Potsdam), Andreas Güntner (GFZ German Research Centre for Geosciences), and the Companhia de Gestão dos Recursos Hídricos (COGERH) for their cooperation and data.

References
ANA (2005), Monitoramento dos Reservatórios do Nordeste, Agência Nacioanal de Águas, Brasília, Brazil.


Campos, J.N.B. (1987), A procedure for reservoir sizing on intermittent rivers under high evaporation rate, Colorado State University, Fort Collins, USA.

COGERH (2002), Anuário do Monitoramento Quantitativo dos Principais Açudes do Estado do Ceará, Companhia de Gestão dos Recursos Hídricos, Fortaleza, Brazil.

COGERH (2003), Anuário do Monitoramento Quantitativo dos Principais Açudes do Estado do Ceará, Companhia de Gestão dos Recursos Hídricos, Fortaleza, Brazil.

Creutzfeldt, B.N.A. (2006), Remote sensing based characterisation of land cover and terrain properties for hydrological modelling in the semi-arid Northeast of Brazil, University of Potsdam, Potsdam, Germany.


FUNCEME (2008), Séries históricas de dados pluviométricos, edited, Fundação Cearense de Meteorologia e Recursos Hídricos.


SCS (1972), Soil Conservation Service National Engineering Handbook, Sec. 4, Hydrology, USA.

SRH (1992), Plano estadual dos recursos hídricos. Vol 2, estudo de base II, Fortaleza, Brazil.

USBR (2002), Avaliação de pequenas barragens, Bureau of Reclamation, Brasília, Brazil.

USGS (1999), Global 30 arc second elevation data set, edited, USGS Eros Data Center.


van Oel, P.R. (2009), Water-scarcity patterns : spatiotemporal interdependencies between water use and water availability in a semi-arid river basin, University of Twente, Enschede.

Evaluating Climate Suitability for Agriculture in Switzerland

A. Holzkämpera, P. Calancaa, J. Fuhrera
aAgroscope Reckenholz-Tänikon Research Station ART, Air Pollution and Climate Group, Reckenholzstrasse 191, CH-8046 Zurich, Switzerland, annelie.holzkaemper@art.admin.ch

Abstract: With climate change increased water shortage and extreme weather events during the cropping season may cause more frequent crop loss, yield instability, and make cultivated areas less suitable for traditional crops. In order to develop long-term agricultural policies, planners need to understand the likely impacts of climate change on the climate suitability for different cultivation types. Agroclimatic indices have great potential to communicate the impacts of climate change. However, each metric only represents a specific aspect of climate that may or may not be relevant for the growth of a certain crop type. To guide planners and policy makers, different indices have to be aggregated in a comprehensible manner. In this paper we present a framework for estimating agricultural suitability for major crops in Switzerland. The framework is based on an evaluation of agroclimatic indices for relevant phenological phases of a range of crops. This allows for taking into account that climate change may lead to significant shifts in growth phases and sensitive periods. Suitability functions are defined for each index. A weighted linear combination is used to aggregate the different elements of climate suitability for each crop and cultivation type. Suitability functions and weights are derived from scientific literature and expert knowledge.

Keywords: Climate indices; agriculture; suitability evaluation; climate change

1. INTRODUCTION

Climate plays a fundamental role in agriculture. The quantity and quality of yields can be affected by water stress, heat stress or frost or by pests and diseases (Kassam et al. 1991). European agriculture may be especially susceptible to meteorological hazards because it is based on highly developed farming techniques (Alexandrov et al. 2008). In recent decades shifts in plant phenology have been observed, showing that ecosystems are already responding to global environmental change: earlier flowering and extended periods of active plant growth across much of the northern hemisphere have been interpreted as responses to warming (Studer et al. 2007). However, at the same time plants grow faster, leading to decreases in quality and quantity of yields (Orlandini et al. 2009). Such changes lead to shifts in the geographical distribution of climate suitabilities for different crops. Planners and land managers need to understand these changes for strategic resource and development planning and in order to develop long-term adaptation strategies (Salinger et al. 2000).

Agroclimatic or agrometeorological indices have great potential to quantify and communicate the impacts of climate change on agriculture (e.g. Bootsma et al. 2005, Patra and Sahu 2007, Orlandini et al. 2009, Eitzinger et al. 2009). They can be used to describe the effects of climatic conditions on key agricultural aspects, including production, protection, fertilization, site selection, irrigation, etc. (Alexandrov et al. 2008). Therefore, agroclimatic indices can be very helpful for farmers in their decisions about crop management options and related farm technologies (Eitzinger et al. 2009).
However, each index only represents a specific aspect of climate that may or may not be relevant for the growth of a certain crop type. To guide land managers and planners, different indices have to be aggregated in a comprehensible manner. Thereby, possible interactions between different climate indices need to be taken into account. For example, a certain number of growing degree days may only be suitable for the growth of a specific crop if the precipitation sum is also within a suitable range. Such interactions can not easily be represented using empirical modeling approaches such as in Hundal et al. (2003).

In this paper we present a framework for an aggregated evaluation of agricultural suitability for major crops in Switzerland. The framework is based on agroclimatic indices that are calculated for relevant phenological phases of a range of crops. This allows for taking into account that climate change may lead to significant shifts in growth phases and sensitive periods. Suitability functions are defined for each index. A weighted linear combination is used to aggregate the different elements of climate suitability for each crop and cultivation type. Suitability functions and weights are derived from scientific literature and expert knowledge.

2. **METHOD**

2.1 Evaluation concept

A quantitative approach is developed to facilitate the crop-specific climate suitability evaluation. The evaluation involves six steps, which are explained in the following.

*Step 1: Determination of growing degree days for relevant phenological phases*

Crop phenological development is expressed as a function of growing degree days. To represent the various stages of development, growing degree day thresholds have to be identified for each phase and crop. This enables the dynamic determination of phenophase-specific climate sensitivities. For example, winter wheat is assumed to be more sensitive to water stress during flowering than grain filling. Depending on the climate, the phenological development might differ from year to year and thus also the relevance of precipitation deficits at individual days of the year could differ.

*Step 2: Selection of relevant climatic indices*

To quantify phenophase-specific climatic influences on crops, different climatic indices can be selected. Indices of drought, excess rain, frost and, to a minor degree, heat stress are probably among the most relevant in Europe (Eitzinger et al. 2009). For this classification approach, the interpretation of indices has to be intuitive as the evaluation is based on expert knowledge.

Frost and heat stress can be quantified through relatively simple indices such as number of frost days (days with Tmin < 0°C) or number of heat days (days with Tmax > 35°C). Excess rain can be quantified in relation to precipitation percentiles or as daily rainfall exceeding a crop specific threshold. Drought indices have to quantify the lack of water during plant growth. Thus, they have to take account of the physical and biological properties of the particular crop in order to reflect its sensitivity towards water stress (Eitzinger et al. 2009). A large variety of drought indices is available from the literature (e.g. the Standardized Precipitation Index (SPI), the ratio of actual to potential evapotranspiration (ET/ETP), the Palmer Drought Severity Index (PDSI)). In addition to these climate indices also the length of different phenological phases can be relevant for the quantity and quality of yields, as crops that mature faster accumulate less biomass.

*Step 3: Determination of index-specific suitability ranges and weightings*

Once the relevant climatic indices have been identified for the selected phenophasces, both index-specific suitabilities si and weights wi need to be specified. si-values are assumed to range from 0 to 1, with 0 indicating no suitability and 1 indicating optimum suitability of an index value. Weights wi are assigned to the indices according to their importance for the crop development and so that they add up to 1. In Fig. 1 for instance, water and heat stress indices are equally weighted and weighted higher than the index characterising the rate of
development. Weights and index-specific suitabilities are initially assigned based on a literature review and will be refined in future work based on expert evaluations.

![Graph showing Heat days, Standardized Precipitation Index, and Period length](image)

**Figure 1.** Example of index-specific suitability $s_i$ functions and weights $w_i$ assigned to three different climatic indices.

The expert-based evaluation of weightings for a large range of agroclimatic indices is often too complex to be made off the top of one's head. A structured approach is required to facilitate the weight assignment and allow for an aggregated assessment of climate suitability. The Analytic Hierarchy Process (AHP, Saaty 1980) provides a means for dealing with such complex multi-criteria decision problems. It has also been applied successfully for multi-criteria evaluation of land suitability (e.g. Hood et al. 2006, Perveen et al. 2008, Thapa and Murayama 2008, Rahman and Saha 2008, Cengiz and Akbulak 2009, Tienwong et al. 2009). Within the AHP, the evaluation is broken down into the variables determining suitability, which are then arranged in a hierarchical order (Figure 2). Variable weights are determined based on pair-wise comparisons by experts. Thus, AHP provides a framework that allows hierarchical combination of criteria and incorporates expert participation in the evaluation process.

![Diagram showing hierarchical evaluation of crop-specific climate suitability](image)

**Figure 2.** Example of hierarchical evaluation of crop-specific climate suitability.

**Step 4: Definition of evaluation functions**

To evaluate crop-specific climate suitability $S_c$ based on the phenophase-specific climatic indices, a weighted average can be derived from the index-specific suitability values $s_i$. However, in some cases the linear combination of indices based on weightings as shown in Fig. 1 might not be appropriate due to interactive effects between the influencing variables. For example, Bowen and Hollinger (2004) assumed that precipitation, growing days, and winter minimum temperature follow the “law of the minimum”. This means if a variable is limiting, the species can not be grown, even if all the other variables are not limiting. To take such dependencies into account evaluation rules can be introduced in the evaluation function.
Step 5: Spatial evaluation

The evaluation function defined in step 4 will at first be applied at the local scale, on the basis of routine observations carried out at a number of stations by the Swiss Meteorological Service (Figure 3). Thus, crop-specific climate suitabilities $S_c$ will be derived for every location and year. Based on the local time series of climate suitabilities, averages and variability measures can be derived. Average climate suitabilities would give an indication on the average potential yields, while the variability of climate suitability could give an indication on climate-related production risks. To produce crop-specific maps of average climate suitabilities and their variabilities, the local values will be interpolated.

Figure 3. Locations of climate stations in Switzerland (red = all climate data automatically recorded, blue = only precipitation data recorded).

Step 6: Climate suitability classification

Finally, the averaged continuous climate suitability values will be discretized according to the FAO classification (FAO 1976), which is commonly applied for land evaluation (e.g. Triantafilis et al. 2001). Thereby, three suitability classes are distinguished: $S1 = $ Highly suitable with no or non-significant limitations, $S2 = $ Moderately suitable with intermediate limitations, and $S3 = $ Marginally suitable with severe limitations. Non-suitable classes are subdivided in $N1 = $ currently not suitable, and $N2 = $ permanently not suitable. Suitability subclasses reflect different kinds of limitations (e.g. $c = $ temperature regime, $m = $ moisture availability). Class boundaries will be determined based on expert knowledge. Similarly, the variability values can be classified into different risk categories. The evaluations will be integrated in a GIS to enhance the compatibility with other spatial data and allow for spatial analyses.

3. Expected results

First results of a suitability evaluation for winter crops will be presented. As high values of suitability are assumed to correspond to optimal climatic conditions for crop growth, the results of the expert-based suitability evaluation will be interpreted in the light of simulated yields derived with the process-based crop model CropSyst assuming no nutrient limitations (Stöckle et al. 2003). Obvious inconsistencies between suitability and relative yields will indicate where there is a need to revise the evaluation scheme.

4. Conclusions and outlook

The presented framework allows for a flexible evaluation of crop-specific climate suitability. The evaluation function can easily be modified or updated to integrate new information or to test assumptions. The GIS integration will enhance the user-friendliness.
of the derived climate suitability maps as it allows for the integration with other GIS data and for conducting spatial analyses.

The integration of phenophase-specific climate indices allows for a dynamic evaluation of climate suitability. Thus, also the impacts of climate change can be investigated. Furthermore, the consideration of variabilities in climate suitability allows for assessing production risks.

The approach is planned to be applied for evaluating climate suitabilities for the most important cultivation types in Switzerland (e.g. winter cereals, maize, pasture, vegetables, grapes, fruit). Based on these crop-specific evaluations an overall climate suitability map for agriculture in Switzerland will be derived indicating areas of optimum cultivation type. In the long term, the approach could be extended to incorporate a soil suitability assessment in addition to the climate suitability assessment. This would provide an even more comprehensive basis for land resource planning.

ACKNOWLEDGEMENTS

The presented work contributes to the project “Climate Change and Agricultural Production Risks (AGRISK)” supported by the Swiss National Science Foundation within the framework of the National Centre of Competence in Research on Climate (NCCR Climate).

REFERENCES


Modeling Hydrologic Regime of a Terminal Lake Basin with GCM Down-scaled Scenarios

Yeo Howe Lim\(^1\), Joshua Hassell\(^1\), and William Teng\(^2\)

\(^1\)Civil Engineering Department, University of North Dakota, Grand Forks, ND 58202-8115
howelim@mail.und.edu; Joshua.hassell@und.edu

\(^2\)NASA Goddard Earth Sciences Data and Information Services Center, USA
William.L.Teng@nasa.gov

Abstract: A dramatic phenomenon has been occurring in Devils Lake, a terminal lake in North Dakota: its inundated area has continually expanded since reaching its extreme low in the 1940s. Devils Lake’s continual rise in water level has created immediate hardships for local residents, particularly farmers who have lost large parcels of productive land and properties. The search for solutions to the problems caused by the unique lake flooding has motivated our modeling of the lake basin’s hydrology, the lake processes, and climatic change impacts. A distributed rainfall runoff model, HEC-HMS, has been calibrated for several major sub-basins that drain into the lake. A significant challenge has been the implementation of Arc-Hydro within ArcGIS for a basin with relatively flat terrain (using the best available DEM of the region) which has undergone a lot of land-use changes in the past century. Further complexities in modeling are introduced by a pumped-drainage scheme implemented by the state’s water commission to lower the lake level by pumping and draining into an adjacent stream. The scheme operates whenever the stream’s water quality is favorable. A reservoir simulation model, HEC-ResSim, which is coupled with the HEC-HMS model, is calibrated for the lake for a short period of time. For the model calibration, NASA’s remote sensing data, including the TRMM Multi-Satellite Precipitation Analysis (TMPA) data, are used to supplement the limited ground data. The coupled, calibrated hydrologic-reservoir model allows a series of comprehensive climatic-scenario simulations to be carried out, a key feature of which is the use of future regional climatic conditions derived from GCM down-scaled ensembles. The described modeling is expected to yield useful and usable results for planners and decision makers who set long-term sustainability plans for the Devils Lake region.

Keywords: hydrologic, HEC-HMS; HEC-ResSim, lake flooding; terminal lake

1. PHENOMENON OF DEVILS LAKE

A unique lake-flooding phenomenon is the ongoing situation in Devils Lake of North Dakota, which has engaged many stakeholders, both locally and internationally. Being a terminal lake, its inundated area has continually expanded since reaching its record low in the 1940s. The lake has more than tripled in size since the early 1990s, which can be attributed to a series of wet years. Figure 1 shows an infamous plot of the lake level, with a very clear trend showing a continual rise in water level since the early 1990s. The slow but persistent rise in water level has inundated large parcels of productive agricultural land, properties, and civil infrastructures (including roads and water and public utilities). Unlike a river flooding which typically ends in days, lake inundation like that in Devils Lake can continue for years. It has created immediate hardships for many local residents, particularly farmers. Affected counties have been building higher dikes and re-routing numerous flooded highways and local roads. The cost associated with all these mitigation efforts is approaching half a billion dollars.
The U.S. Northern Great Plains, which includes the Dakotas, is predicted by GCM models to have greater amounts of precipitation in various simulation scenarios (IPCC, 2000). In addition to these model predictions, there are many regional hydrologic factors which have to be fully understood, such as evapotranspiration and seepage of the lake, and basin runoff characteristics. A companion paper in this volume by Zhang (2010) covers the present problem of Devils Lake, while another paper by Kirilenko (2010) presents the climate change impacts on agriculture in the Devils Lake Basin. We postulate that a fully distributed hydrologic model, combined with a reservoir simulation model, and used together with downscaled-GCM ensembles, can predict future lake levels under various climate change scenarios. The search for solutions to the problems caused by the unique Devils Lake flooding has motivated our modeling of the lake basin’s hydrology, the lake processes, and climatic change impacts.

2. HYDROLOGIC MODELING

The hydrologic modeling consists of two components: (1) modeling the generation of runoffs from the sub-basins into Devils Lake at seven locations and (2) modeling the water balance of the lake itself.

DEM and Delineations

Digital delineations of watersheds that flow into Devils Lake comprise the first step in the hydrologic modeling. 150 DEMs from the United States Geological Survey (USGS) were input to ArcGIS (from ESRI with Arc-Hydro extension) to delineate the sub-basins (see Figure 2). The total number of cells used varied according to the basin size. The model for Mauvais Coulee basin, e.g., is made up of half a million cells, each of size 10m x 10m.
The delineation using ArcGIS consisted of (1) merging all the DEMs together; (2) filling in the many small pot holes, to allow water to flow; (3) identifying flow direction once the direction of water flow from each cell is determined; (4) identifying points where water will accumulate (flow accumulation) and, hence, outlining the streams and rivers; and (5) constructing the stream network by setting the number of cells used to define a stream. The number of cells is determined not only by the size of the watersheds but also by the total number of watersheds and streams. ArcGIS’ stream link tool connects the streams by helping to define the nodes, junctions, and reaches of the streams; the catchment delineation tool defines the watershed for each stream section. The implementation of Arc-Hydro extension within ArcGIS for a basin with relatively flat terrain could have caused the formation of tiny holes (groups of cells) in the delineated watersheds. An approximation solution was adopted by adding the relatively small areas to the adjacent basins. The small watersheds were combined to form the six large basins (Figure 2). These large basins were then input as background maps into HEC-HMS (USACE-HEC, 2009) to assist in creating a HEC-HMS model. The dendritic network of channels was modeled mainly by considering its main channel lengths, slopes, and routing parameters. The channel lags were estimated using the SCS Soil Curve Number and the empirical equation of SCS (SCS is now NRCS). 20 sub-basins were used for modeling the Mouvais Coulee, with the details shown in Table 1. It can be seen that the basin topography is very gentle, with slopes (10-80) in the range of 0.00015 to 0.0015 for major portions of the basin. Figure 3 shows the model setup for the Mouvais Coulee, featuring the interconnections of sub-basins, channel reaches, junctions, and a sink.

Table 1. Parameters for the sub-basins of Mouvais Coulee.

<table>
<thead>
<tr>
<th>Sub Basin</th>
<th>Area (mi²)</th>
<th>Length (ft)</th>
<th>80% elevation</th>
<th>20% elevation</th>
<th>ΔH</th>
<th>60% length</th>
<th>Slope (ft/ft)</th>
<th>CN</th>
<th>Time of Conc.(min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>38.61</td>
<td>83000</td>
<td>1625</td>
<td>1550</td>
<td>75</td>
<td>49900</td>
<td>0.00151</td>
<td>77</td>
<td>3075</td>
</tr>
<tr>
<td>2</td>
<td>46.33</td>
<td>140000</td>
<td>1595</td>
<td>1465</td>
<td>130</td>
<td>84000</td>
<td>0.00155</td>
<td>76</td>
<td>4745</td>
</tr>
<tr>
<td>3</td>
<td>30.89</td>
<td>80000</td>
<td>1505</td>
<td>1460</td>
<td>45</td>
<td>48000</td>
<td>0.00094</td>
<td>75</td>
<td>4011</td>
</tr>
<tr>
<td>4</td>
<td>42.47</td>
<td>124000</td>
<td>1485</td>
<td>1465</td>
<td>20</td>
<td>74400</td>
<td>0.00027</td>
<td>77</td>
<td>10033</td>
</tr>
<tr>
<td>5</td>
<td>50.19</td>
<td>85000</td>
<td>1490</td>
<td>1465</td>
<td>25</td>
<td>51000</td>
<td>0.00049</td>
<td>76</td>
<td>5657</td>
</tr>
<tr>
<td>6</td>
<td>57.92</td>
<td>83000</td>
<td>1490</td>
<td>1460</td>
<td>30</td>
<td>49900</td>
<td>0.00060</td>
<td>77</td>
<td>4861</td>
</tr>
<tr>
<td>7</td>
<td>30.89</td>
<td>105000</td>
<td>1500</td>
<td>1465</td>
<td>35</td>
<td>63000</td>
<td>0.00056</td>
<td>75</td>
<td>6477</td>
</tr>
<tr>
<td>8</td>
<td>38.61</td>
<td>76000</td>
<td>1540</td>
<td>1525</td>
<td>15</td>
<td>45600</td>
<td>0.00033</td>
<td>74</td>
<td>6687</td>
</tr>
<tr>
<td>9</td>
<td>50.19</td>
<td>98000</td>
<td>1535</td>
<td>1485</td>
<td>50</td>
<td>58800</td>
<td>0.00085</td>
<td>73</td>
<td>5243</td>
</tr>
<tr>
<td>10</td>
<td>54.05</td>
<td>97500</td>
<td>1510</td>
<td>1475</td>
<td>35</td>
<td>58500</td>
<td>0.00060</td>
<td>76</td>
<td>5714</td>
</tr>
<tr>
<td>11</td>
<td>34.75</td>
<td>113000</td>
<td>1490</td>
<td>1455</td>
<td>35</td>
<td>67800</td>
<td>0.00052</td>
<td>77</td>
<td>6722</td>
</tr>
<tr>
<td>12</td>
<td>30.89</td>
<td>59000</td>
<td>1460</td>
<td>1450</td>
<td>10</td>
<td>35400</td>
<td>0.00028</td>
<td>75</td>
<td>5727</td>
</tr>
<tr>
<td>13</td>
<td>38.61</td>
<td>74000</td>
<td>1475</td>
<td>1460</td>
<td>15</td>
<td>44400</td>
<td>0.00034</td>
<td>76</td>
<td>6099</td>
</tr>
<tr>
<td>14</td>
<td>57.92</td>
<td>118000</td>
<td>1470</td>
<td>1445</td>
<td>25</td>
<td>70800</td>
<td>0.00035</td>
<td>77</td>
<td>8414</td>
</tr>
<tr>
<td>15</td>
<td>65.64</td>
<td>90000</td>
<td>1465</td>
<td>1440</td>
<td>25</td>
<td>54000</td>
<td>0.00046</td>
<td>74</td>
<td>6453</td>
</tr>
<tr>
<td>16</td>
<td>54.05</td>
<td>54000</td>
<td>1445</td>
<td>1440</td>
<td>5</td>
<td>32400</td>
<td>0.00015</td>
<td>73</td>
<td>7640</td>
</tr>
<tr>
<td>17</td>
<td>42.47</td>
<td>105000</td>
<td>1550</td>
<td>1475</td>
<td>75</td>
<td>63000</td>
<td>0.00019</td>
<td>76</td>
<td>4298</td>
</tr>
<tr>
<td>18</td>
<td>34.75</td>
<td>168000</td>
<td>1460</td>
<td>1445</td>
<td>15</td>
<td>100800</td>
<td>0.00015</td>
<td>77</td>
<td>17194</td>
</tr>
<tr>
<td>19</td>
<td>50.19</td>
<td>153000</td>
<td>1460</td>
<td>1445</td>
<td>15</td>
<td>91800</td>
<td>0.00016</td>
<td>73</td>
<td>17080</td>
</tr>
<tr>
<td>20</td>
<td>30.89</td>
<td>126000</td>
<td>1500</td>
<td>1450</td>
<td>50</td>
<td>75600</td>
<td>0.00066</td>
<td>77</td>
<td>6479</td>
</tr>
</tbody>
</table>

3. CALIBRATION OF THE HEC-HMS MODEL

A calibration of the hydrologic model for the Mauvais Coulee basin involved an event from 6/13/2001 to 6/20/2001. The flow gauging station at USGS station 05056100 had been in operation from 06/01/1956 to 12/30/2009. The precipitation data from the North Dakota State Water Commission (NDSWC) station 848 operated by volunteers include rainfall. The results of calibration for that storm event, which include discounting the baseflows, are shown in Figure 4. Various initial loss and continuous loss values were used. Initial loss ranging from 0.04 to 0.2 in. (1 to 5 mm) was found acceptable, while the continuous loss is more sensitive. The model performed the best in the range of 0.037 to 0.039 in/day.

Improving the calibration of the Mauvais Coulee basin is important, because it is the only basin with significant discharge data, which will help make our model more accurate. The completed calibration attributes can subsequently be applied to the other six basins.
Precipitation data derived from NASA Tropical Rainfall Measuring Mission (TRMM) satellite were converted into grid files which were input into the HEC-HMS model. These supplementary data helped the calibration process, because only one ground-based rainfall
A uniform rainfall was assumed, even though, in reality, a storm system moving through a basin is often very unevenly distributed. With only one ground station within the basin, an error estimate for this assumption was not possible. Using grid data in the HEC-HMS model will, in theory, make our model more accurate due to the more accurate rainfall measurements. However, there were significant differences in terms of storm magnitude and temporal distributions.

The use of grid input to HEC-HMS encountered a problem. The approach was to treat each grid cell data as if they were observed by a ground station located at the center of the grid cell. The grid cell data was provided in ASCII format. Initially, it was input into ArcGIS as point data, which were then converted to grid format by adding a cell size. When added to the basin file, the grid did not overlap due to ArcGIS interpreting the longitude and latitude as meters instead of degrees. To overcome this, we input the grid data into a separate file and converted the degrees to meters. The saved file is imported into the basin file of HEC-HMS, with the grid overlapping the delineated basins.

### 4. USE OF NASA DATA IN MODEL CALIBRATION

Sources of NASA satellite and model data used in this study include (1) for precipitation, the Tropical Rainfall Measuring Mission (TRMM) (Huffman et al., 2007); (2) for surface air temperature, the Atmospheric Infrared Sounder (AIRS) (Aumann et al., 2003); (3) for soil moisture, the Land Parameter Retrieval Model (LPRM) applied to brightness temperatures from the Advanced Microwave Scanning Radiometer - Earth Observing System (AMSR-E) and the TRMM Microwave Imager (TMI) (Owe et al., 2008); and (4) for solar radiation and snow water equivalent, the Global Land Data Assimilation System (GLDAS) (Rodell et al., 2004). Brief descriptions of these data follow. The TRMM product, 3B42 (V6), is an “organic” merged product, with new sensor data incorporated as they become available. It uses an optimal combination of microwave-based precipitation estimates to adjust the infrared (IR) estimates from geostationary observations. The resulting global precipitation estimates are then scaled to match the monthly rain gauge analyses. The surface air temperature is part of the AIRS L2 Standard Retrieval Product (AIRX2RET), which includes profiles of retrieved temperature. LPRM data provide global soil moisture for the top few centimeters of the soil column. LRPM is a three-parameter retrieval model for passive microwave data and is based on a microwave radiative transfer model that links surface geophysical variables (i.e., soil moisture, vegetation water content, and soil/canopy temperature) to the observed brightness temperatures. GLDAS drives multiple, offline (not coupled to the atmosphere) land surface models, integrates a large quantity of observation-based data, and executes globally enabled by the Land Information System (Kumar et al., 2006). Time series of the above data are prepared for input to HEC-HMS and GCM simulations. NASA data offer a denser and more uniform spatial coverage, which is required to drive the hydrologic models. An isolated rainstorm event recorded at the NDSWC station 848 in the Mauvais Coulee basin was compared with the TRMM data. The TRMM data indicated less precipitation total by up to 13% over a grid area (25 x 25 km). More elaborate comparisons will be carried out for more ground stations located in proximity of the Devils Lake basin.

### TRMM Data Input Scheme

TRMM data available on the grid pattern were prepared for entry into HEC-HMS. However, that turned out to be inefficient because there are only 13 grid cells overlying the Mauvais Coulee, the largest of the six basins. Due to the grid format of HEC-HMS, input of data in a time series can be cumbersome, given the small grid size with which we are dealing. To be able to use the grid data, we set up a precipitation gauge weight time series component. Each grid cell has a gauge applied to the model to represent the cell. The precipitation gauge weight storm setup was deemed satisfactory.

### Calibration of Multi-year Series

Devils Lake is in a temperate region; hence, it was necessary to calibrate the HEC-HMS model against a continuous flow series covering multiple years, including several winter seasons.
Accumulations of precipitation over the winters and snowmelt runoff processes in the spring were modeled in HEC-HMS. The observed temperature series over the basins was used to determine precipitation type, while a melt-rate function determined the runoff generation from the snowpacks. The melt-rate function allows the HEC-HMS model to interpret how fast the snow will melt, based on the temperature above freezing. This function is described in Dewalle & Rango, 2008 as

\[ M_s = C_m(T_a - T_b) \]  

where \( M_s \) is snowmelt (in/Fo day), \( C_m \) is the melt-rate coefficient which varies between 0.04-0.08, \( T_a \) is the air temperature (F°), and \( T_b \) is the base temperature (F°). Along with the ground gauge runoff series at USGS station 05056100 over a four-year period, a project storm derived from TRMM data was created using multiple gauge stations to represent the grid cells. Figure 5 shows a plot of a model simulation over a four-year period using \( C_m \) of 0.04. Generally, the observed storm sequences and the model runoffs coincide very well, though a discrepancy in the peak flow can be seen at the very first runoff. The total volume of the observed runoff was 15.19 inches, while the model computed runoff was 10.44 inches or 68 percent of the observed value, for \( C_m \) at 0.04. For \( C_m \) of 0.08, the generated total volume was 10.94 inches or approximately 72 percent of the observed value. The difference is large, but it is acceptable for the present feasibility study purpose. Moreover, the dependency on just one ground observation is the limitation of the comparison.

5. HEC-RESSIM MODEL

Devils Lake resembles a reservoir with incoming flows from the surrounding seven major river basins. A reservoir simulation model, HEC-ResSim (obtained from the U.S. Army Corps), was implemented for the lake and surrounding areas. A schematic of the model setup can be seen in Figure 5. It includes four flow regimes: (1) inflows feeding the lake from the major basins; (2) a pumped outlet scheme, constructed by the North Dakota State Water Commission, that drains the lake, via a series of canals and pumps, into the Sheyenne River; (3) the natural overflow channel that connects Devils Lake to Stump Lake; and (4) the eventual overtopping pathway of Stump Lake into the Toulna Coulee and Sheyenne River.
The estimated pumped volume is almost negligible, because the pumped outlet scheme was operational starting in June 2007 with an original pumping capacity of 100 cfs. The outlet operates from April 1 to November 30, if lake level is over 1446.0 ft amsl. One main problem was the high sulfate concentrations in both the water of the lake and the receiving Sheyenne River, which prevented the pumping at full capacity. The limiting factor is the sulfate class 1A stream standard on Sheyenne River (450 mg/l). However, the pumping rate reached almost the capacity of permit in 2009. The current pumping rate is modified to 250 cfs. The additional pumping capacity is scheduled to begin on July 1, 2010. Recently, the North Dakota Department of Health (NDDH) adjusted the rate to a site-specific standard downstream of Baldhill Dam (750 mg/l). With that adjustment, the pumped volume is likely to be increased. The outlet discharge by pumping, since its implementation in 2007, based on reports submitted by the NDSWC to NDDH, shows an approximate drop in water level in Devils Lake by more than 65 mm (2.5 in). The accounting of current outflows includes pumping of water from the lake, losses through lake evaporation, and losses through the lake’s pool seepage. The pool seepage and lake evaporation, together with the inflows generated by the HEC-HMS models of the river basins, are calibrated against the observed water level series of Devils Lake provided by USGS. For continuous simulation purposes, DSS (a file format used in the suite of HEC software for data storage and transfer) files generated by HEC-HMS for each basin in the continuous period studied are imported into the reservoir model. The lake levels observed for the period are also entered as benchmarking for reservoir calibration.

6. GCM DOWNSCALED SCENARIOS

The GCM downscaling for this project is described in the companion paper by Kililenko et al. (2010). The whole process of deriving GCM downscaled scenarios for running the hydrologic model entails significant data handling and processing. The main outputs from GCM downscaling are precipitation and temperature. The gridded data need to be transformed to the appropriate basin-level or as single site time series data. Figure 7 shows the integration of the hydrologic model, reservoir model, and the downscaled GCM ensembles for hydrologic runoff simulations in the Devils Lake basin.

7. CONCLUSIONS

The calibration of hydrologic models (HEC-HMS) for the upland sub-basins of the Devils Lake Basin has proven to be challenging, because of the absence of sufficient ground-based precipitation data. Several storm events derived from the NASA TRMM satellite data have been processed and used to supplement the calibration process. The calibration of HEC-HMS model was satisfactory and a continuous loss of 0.037 to 0.039 in/day will be used in the simulation models. The use of NASA TRMM data in generation of runoff through calibrated HEC-HMS model has shown certain promising success. The runoff timing in the HEC-HMS model matches up consistently with a ground station when melt rate coefficients of 0.04 and 0.08 are used. The initial reservoir model has been established using HEC-ResSim, but further investigations are required for running the full reservoir model under various GCM down-scaled scenarios.
ACKNOWLEDGMENTS
This study was supported by NASA grant NNX09AO06G, “The Feasibility of Combining NASA Satellite Data, General Circulation Models, and Hydrologic Models to Inform Decision Making for Flood Mitigation of the Devils Lake Basin of North Dakota.”

REFERENCES
Kirilenko, A. (2010), Climate change impact on agriculture: Devils Lake basin, 2010 International Congress on Environmental Modelling and Software Modelling for Environment’s Sake, Fifth Biennial Meeting, Ottawa, Canada.
Modelling irrigation demand from two grasslands in Switzerland under contrasting climatic conditions and soil properties

P. C. Smith, P. Calanca and J. Fuhrer
Swiss Federal Research Station Agroscope Reckenholz-Tänikon (ART), Reckenholzstrasse 191, CH-8046 Zürich, Switzerland
pascalle.smith@art.admin.ch

Abstract: We examine irrigation demand in Switzerland under current climatic conditions on the basis of simulations with a process-based grassland model. Two sites are considered: Oensingen, on the Swiss Plateau, benefits from a humid climate and is rarely affected by drought; Sion, located in the Rhone Valley, is highly water-limited, with critical soil moisture conditions every year. Water requirements to sustain productivity therefore vary substantially between the sites and in time, with needs up to 800 mm per year at Sion, and average needs of the order of 300 mm at Oensingen if a shallow soil is assumed. We show that simple relations can be found between water requirements and key variables such as the so-called atmospheric water budget, i.e. the difference between precipitation and potential evapotranspiration. Such relations are not only useful to quantify water requirements at other sites, but also for assessing the benefits of irrigation. At Sion, we find for instance that even while restricting the water requirement fulfilment to 500 mm per year, irrigation can increase productivity by 9 t/ha compared to a rainfed situation.

Keywords: Grassland model; Water deficiency; Yield; Irrigation requirement; Switzerland

1. INTRODUCTION

Drought can significantly affect agricultural production. In Europe this is currently the case mainly in the Mediterranean area and in the South-East [IPCC, 2007]. However, climate scenarios developed under PRUDENCE [Christensen et al., 2007] suggest that the situation could change in the future. In fact, a decrease in summer rainfall and associated increase in summer heat-waves is projected in many areas, including Switzerland and the Alps [Calanca, 2007]. In this respect, the summer heat wave of 2003, with its negative impacts on terrestrial ecosystems, is considered by many as a shape of things to come [Schär, 2004; Ciais et al., 2005; Smith et al., 2010].

Coping with more frequent dry and hot episodes during the growing season is a challenge for farmers. Increasing the share of irrigated land could be one of the possibilities to face more frequent droughts. In Switzerland for instance, several regional governments are currently examining options in this direction [Lehman et al., 2001; Fuhrer & Jasper, 2009; http://www.acqwa.ch]. However this could lead to water use conflicts with other economical sectors, as well as to water waste, higher production costs, erosion and nutrient leaching. Moreover, irrigation could affect the local climate in different ways: moist and dark soils absorb more energy from the sun due to a reduced reflectivity which tends to warm the surface, but available moisture allows the soil-vegetation system to respond to the atmospheric demand by evaporating and transpiring more [Boucher et al., 2004].

Simulation models enable to better understand water demand and consumption in agriculture. They can be used for sensitivity analysis and to explore the implication of projections from climate scenarios. In this study we applied the PROdutive GRASSland Simulator (PROGRASS [Lazzarotto et al., 2009]) to investigate the water balance of
grassland ecosystems in two contrasted areas of Switzerland: The Swiss Plateau and the Rhone Valley. On the Swiss Plateau summer precipitation (total amount during June, July and August) is equivalent to about 400 mm in the long-term, implying drought conditions only every tenth year on average [Calanca, 2004]. In contrast, the Rhone Valley, with only 150 mm of precipitation during summer, is characterized by a dry climate. Water deficit during the growing season is therefore a recurrent phenomenon.

In Lazzarotto et al. [2009, 2010], PROGRASS was applied to study the effect of climate change on grassland dynamics and nutrient cycling. The water balance was not addressed directly. To be able to investigate water demand and consumption under varying climatic conditions and soil properties, improvements to PROGRASS in a number of features were needed. Presenting these improvements and the investigation results for both contrasting sites is the aim of this paper.

2. STUDY SITE CHARACTERISTICS

Location and physiographic characteristics of the two sites considered for the present investigation are given in Figure 1 and Table 1.

Oensingen (Oe) is located on the Swiss Plateau and is characterized by a humid climate, with an annual precipitation of 1150 mm and a mean annual temperature of 9.2 °C. The heavy soil (stagnic cambisol eutric, 43 % of clay) has an estimated maximum water storage capacity of 416 mm for a rooting depth of 800 mm. At this site, several aspects of grassland dynamics related to carbon, nitrogen and water cycling have been investigated experimentally since 2001 [Ammann et al., 2007, 2009].

Sion (Si), located in the Rhone Valley (South-western Switzerland), is characterized by a dry climate, with an annual precipitation of 610 mm and an annual mean temperature of 10.3 °C. The soil is a sandy loam, with an estimated maximum water storage capacity of 53 mm, for a rooting depth of 130 mm. Both values were estimated based on information available from the Swiss soil suitability map [BFS, 2004].

![Figure 1 - Study site locations in Switzerland’s topography](image)

Table 1 – Soil properties at Oensingen and Sion. \( \theta_{\text{sat}} \), \( \theta_{\text{fc}} \) and \( \theta_{\text{pwp}} \) are the volumetric soil moisture content [mm water/mm soil] at saturation (porosity), field capacity and the permanent wilting point, respectively. \( \theta_{\text{crit}} \) is the soil moisture threshold considered for computing water requirements. It is assumed that below this threshold drought stress starts limiting plant growth.

<table>
<thead>
<tr>
<th>Rooting depth [mm]</th>
<th>% sand</th>
<th>% silt</th>
<th>% clay</th>
<th>( \theta_{\text{sat}} )</th>
<th>( \theta_{\text{fc}} )</th>
<th>( \theta_{\text{crit}} )</th>
<th>( \theta_{\text{pwp}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oensingen</td>
<td>800</td>
<td>25</td>
<td>32</td>
<td>0.52</td>
<td>0.46</td>
<td>0.35</td>
<td>0.22</td>
</tr>
<tr>
<td>Sion</td>
<td>130</td>
<td>65</td>
<td>25</td>
<td>0.41</td>
<td>0.36</td>
<td>0.12</td>
<td>0.09</td>
</tr>
</tbody>
</table>
3. METHODS

3.1 PROGRASS and its extensions

Originally, PROGRASS was designed to simulate grassland growth in mixed grass/clover swards in response to climate and management. It uses a bucket approach to estimate the soil water balance, where inputs are from precipitation minus interception and outputs are from evapotranspiration and drainage [Lazzarotto et al., 2009]. The model does not consider lateral water movement or water ponding at the surface. Water in excess of the infiltration capacity is removed as surface runoff.

Our extensions include a better formulation of rainfall interception that entails inhibition of transpiration while the intercepted water store evaporates, and an explicit formulation of capillary rise (Qcap) from external groundwater. The latter prevents soil moisture from crossing the permanent wilting point in very dry conditions. To compute the soil moisture budget, at each hourly time step, losses from evapotranspiration (ET) and drainage (Qdrain) are first subtracted from the available soil water store. Subsequently, non-intercepted precipitation (P) is allowed to infiltrate as long as pore space is available, otherwise surface runoff is generated (Qsurf).

Concerning ET, the model distinguishes between potential ET (pET), which represents the water uptake under optimum growing conditions, and actual evaporation and transpiration (aET), which depends on soil moisture status. For all processes, drought stress is defined with respect to specific soil moisture thresholds [Lazzarotto et al. 2009], which in most cases are close to θcrit as given in Table 1.

To test the new formulation of the water balance, water fluxes simulated at Oensingen were compared to experimental data. We found that daily, seasonal and interannual variations of aET correlate well with observations, in spite of a slight tendency of the model to overestimate day-to-day variability and to underestimate the annual mean. Soil drying and wetting phases in response to precipitation events and losses by drainage, runoff and ET, including the duration and severity of the 2003 drought, were well captured by the model (data not shown).

3.2 Computation of water requirements

In Fuhrer & Jasper [2009], needs for irrigation were examined relatively to critical thresholds in the ET efficiency, defined as the ratio aET / pET. In the present study, we directly computed the amount of water required to fulfil the climatic demand (pET) and sustain potential growth in simulations for non-water-limited (Nwl) conditions. In order to estimate ET efficiency and productivity gains we compared these simulations with that for rainfed (Rf) conditions. This results in four output datasets (OeRf, OeNwl, SiRf, SiNwl). In practice, water requirement (Wreq) is defined as the amount of water needed to maintain soil moisture just above the critical level θcrit. Note that pETRf ≠ pETNwl because keeping soil moisture above the critical level increases the Leaf Area Index (LAI). Moreover, aETNwl = pETNwl. In the following, if not stated otherwise, pET refers to pETNwl and aET to aETRf.

3.3 Simulation setup and initialization

All simulations were run for 28 years using hourly weather data for 1981-2008 as recorded by the Swiss Federal Office for Meteorology and Climate (MeteoSwiss) at Wynau (5 km from Oensingen) and Sion. Because of the large differences in the estimated rooting depth, which may render the interpretation of the results more difficult, simulations were carried out assuming for both sites either 800 mm (subsequently denoted as “deep”) or 130 mm (subsequently denoted as “shallow”).

The same management was prescribed for both sites each year, with 5 cuts between May 5 and October 25, and applications of about 240kg/ha of nitrogen fertilizers (190 kgN/ha in mineral form, 50 kgN/ha in organic form). This reflects the intensive management at Oensingen [Fig. 2b-c in Ammann et al., 2009].
Atmospheric CO₂ concentrations were assumed constant at 370 ppm. Initial conditions for root and shoot mass and soil organic matter were specified according to field data from Oensingen. Additional equilibrium simulations for Sion spanning a total of about 1000 years showed that near-steady-state conditions were reached after 5 years. Therefore the first 5 years of each simulation were omitted from the analysis.

4. Results

First we characterize the water supply and demand and consequently the water deficiency at both sites, partly due to climate and partly due to soil conditions. Then we quantify the minimum water amount required to meet the demand and to sustain potential productivity across pedo-climatic conditions. These water amounts are finally analyzed in relation to key variables.

4.1 Climate, evapotranspiration efficiency and water requirements

Basic features relative to climate, drought stress and water requirements at Oensingen and Sion are presented in Figure 2. Note that all values are computed over the period March 1 to October 31, which is assumed to represent the growing season.

Average growing-season precipitation at Oensingen (800 mm) is twice the amount of that at Sion (400mm), but water demand for potential growth at Oensingen is on average only about 70% of the value estimated for Sion. As a result, ET efficiency is of the order of 0.9 (deep) or 0.7 (shallow) at Oensingen, but only of the order of 0.6 (deep) and 0.5 (shallow) at Sion.

![Figure 2](image)

**Figure 2** – Precipitation (solid lines) and pET (dashed lines) (a-b), actual to potential ET ratio (c-d) and water requirement (e-f) for Oensingen (left) and Sion (right) over 1986-2008. Values were summed over the growing season from March to October. Dotted lines indicate the 0.8 common ET ratio reference and maximum at 1, and year 2003. Black lines indicate the runs using a deep soil (800 mm), and gray lines that using a shallow soil (130 mm).
2003 is the only year where ET efficiency falls below 0.8 (the common ET ratio reference [Doorenbos & Kassam, 1979]) in deep rooting conditions at Oensingen, due to an increase in pET rather than a drop in aET. It is also the only year where pET exceeds precipitation, whereas this is the case every year at Sion, leading at this site to semi-arid growing conditions. Low precipitation, low relative humidity, high wind speed and high radiation typical for the Rhone Valley are responsible for the large gap between the low water availability and high water demand.

The amounts of water required to fill this gap along over the growing season are shown in Figure 2e-f. Average needs are 90 mm for a deep soil at Oensingen, with a peak of 300 mm during the summer of 2003. Average requirements in the case of a shallow soil are about 300 mm at this site. At Sion, average water requirements are of the order of 400 mm (deep) or 600 mm (shallow), with variations of about 50 to 100 mm from year to year and peak requirements close to 800 mm.

4.2 Soil properties and water losses

Shallow soils increase drought stress and water requirements (gray lines in Fig. 2c-f) for unchanged precipitation and pET, in a larger proportion at Oensingen than at Sion. Simulations with PROGRASS suggest that water requirements are not simply given by the difference between growing-season pET and precipitation. Losses from surface runoff and drainage limit the precipitation efficiency in meeting plant demand, in particular in the case of high precipitation amounts and/or shallow soils (i.e. limited holding and infiltration capacity). This is the case at Oensingen (and Sion, in the simulation with a shallow soil only), where losses account for about 50% of the seasonal rainfall.

4.3 Irrigation demand as a function of water deficiency

Because of the considerable variability in climate (20 % in P and 6 % in pET) and water requirements, it is interesting from a practical point of view to find simple relations for expressing water requirements as a function of selected key variables. As shown in Fig. 3a, water requirements can, for instance, be expressed as a decreasing function of the so-called atmospheric water budget (the difference between precipitation and pET). There is an asymptotic behaviour for positive values of P - pET, with limits depending on assumed soil depth. Even simpler is the relation between water requirements and the difference between pET and aET (Fig. 3b), which is, by construction, not surprising at the hourly time step but could have integrated non-linearities on the growing season scale. Note, however, the systematic departure from the 1:1-line, with Wreq in excess of pET - aET, indicative of the fact that aET is also sustained by capillary rise (about 200 mm per growing season at Sion when a shallow soil is assumed).

**Figure 3** – Water requirement expressed as a function of the so-called atmospheric budget (precipitation minus pET) (a) and of the difference between pET and aET (b). Oensingen is represented with circles and Sion with triangles. Full/empty symbols refer to results of simulations for a deep/shallow soil.
4.4 Irrigation demand and productivity

From an agronomic perspective it is interesting to examine drought stress and water requirements in relation to productivity. Yield stability is one of the main purposes of irrigation. The absolute gain in productivity as a function of the water requirements is shown in Figure 4a. Expressed as a function of relative ET (or ET efficiency), relative productivity appears to follow a linear relationship (Fig. 4b). A one-to-one relationship is the expected behaviour for a constant water use efficiency, but the simulations suggest that relative yield is generally below relative ET, except in the case of very humid climates and deep soils. Water use efficiency is found to decrease with increasing soil shallowness and climate aridity, so does the yield gain by unit of irrigation water required (Fig. 4a). In any case, once the potential productivity is reached (of the order of 15-16 t/ha/y at the two study sites) no further benefits can be expected from an increase in irrigation.

5. SYNTHESIS AND DISCUSSION

In this paper we presented an estimation of irrigation water requirements in managed grasslands relying on simulations with a process-based ecosystem model. We studied two sites with contrasting climates and soils, Oensingen being characterized by radiation-limited and Sion by soil moisture-limited growing conditions [Seneviratne et al., 2010].

The results suggest that under current climatic conditions, precipitation is a key variable, both in relation to the average requirements as well as to variability. In humid climates, however, soil properties and rooting depths can also play a role, because in shallow soils with limited storage and infiltration capacity a large fraction of the precipitation is lost as surface runoff and drainage, eventually leading to significant soil water deficits. In shallow soil capillary rise from the water table can nevertheless partially compensate for water deficits. In general this shows the importance of a proper description, in ecosystem models, of water fluxes at the rooting zone lower boundary and of the plant rooting strategy under drought stress [Teuling et al., 2006].

Under a future climate involving a possible increase in the intensity of precipitation events, the partitioning of water fluxes into consumption by the grassland ET and losses by runoff and drainage may change towards generating higher drought stress in the intervals. Plant growth can temporarily suffer from anaerobic conditions due to soil moisture exceeding field capacity, which already occur in winter in Oensingen.

In spite of variability across sites and in time, we found that water requirements can easily be understood and estimated using simple relations. For instance we showed how water requirements can be expressed as a function of the so-called atmospheric water budget. The asymptotic behaviour found in our simulations is reminiscent of the relations of seasonal mean fluxes discussed in Budyko [1974]. In particular, the fact that the limiting value depends on soil depth can be well understood in the context of analysis of the seasonal water balance of terrestrial ecosystems developed by Milly [1993].
We also found a linear relationship between relative productivity (or yield) and relative ET, which appears to support the simple productivity model at the base of the Food and Agriculture Organisation (FAO) methodology [Doorenbos & Kassam, 1979]. We expect that such relations can be generalized to other sites, providing guidance for practitioners. But care is needed in doing so, because important aspects were not yet investigated, e.g. situations where water is lost by lateral fluxes driven by topography, when the rooting zone is not in contact with the water table, or when soil and canopy characteristics are affected by management.

Grassland management includes fertilization (which affects potential productivity and pET), cutting regime (which also has implications for both), grazing (seen as a disturbance but also in relation to soil compaction), rotations and the like. Models that are able to account for some if not all of these aspects, as it is the case with PROGRASS, are therefore needed to study the system sensitivity to drought stress and irrigation. This type of model also provides opportunities to study water requirements for specific growth phases, rather than for the whole growing season as presented here. This could be important, for instance, in view of the necessity to optimize water use in irrigation.

Finally, we would like to point out that the present estimates should in any case be considered as a lower bound, because we implicitly assumed a perfect drought stress monitoring and full efficiency with respect to irrigation. In practice, irrigation efficiency depends on local conditions (including soil permeability) as well as various technological aspects (frequency, amount and position), which need to be considered in practical applications. The target yield and related soil moisture threshold for estimating water requirements (currently aiming at zero stress) should in the end be adjusted to account for economical and environmental factors shifting the optimum irrigation level.

6. CONCLUSION AND OUTLOOK

In this study we were able to highlight some of the key features of the relation between water requirements, climate and soil. We limited our attention to current climatic conditions, and showed that even in humid climates irrigation need could be quite substantial in the presence of shallow soils. The model produces daily estimates and is able to consider the interactive effects of management intensity and water deficiency.

The study is part of the ACQWA European project, which is currently being carried out to estimate climate change impacts on water quantity and quality in vulnerable mountain catchments. Next steps towards reaching the aims of this project include revising PROGRASS to allow simulating grassland productivity and water needs along an altitudinal gradient. This implies extensions of the model towards accounting for snow cover, slope (lateral fluxes), heterogeneous and stone-rich soils. Note that in mountain regions grasslands are often used as pastures, and thus the model needs to be adapted to include the impacts of grazing. For model development and testing, input and verification data are essential. Networks of field experiments make data available for this purpose [Calame et al., 1992; Jeangros et al., 1992; Jonas et al., 2008].

Aiming at producing Canton-scale estimates of future water requirements, the modelling approach will also be made spatially explicit and extended to the other relevant crops (such as, maize, wheat, fruit trees and grapevine), cultivated in the lower elevation belt of the Rhone catchment. Climate and hydrological change scenarios downscaled to the alpine region will be used to drive future conditions for agricultural production. Comparison of results with those obtained with the FAO CROPWAT methodology or Fuhrer & Jasper [2009] for instance will help estimate the uncertainty associated with model predictions.

ACKNOWLEDGMENTS

Work is being supported by the EU-FP7 ACQWA Project and the Swiss NCCR-Climate.

REFERENCES


Statistical picture of climate changes in Central Asia:
Temperature, precipitation, and river flow

Darya Savitskaya
daryasavit@yandex.ru

Abstract: The research presented in this paper is dedicated to the temperature, precipitation and rivers runoff observations in Central Asia. Conclusions about the current situation and the impact of future climate change in the region based on observations from more than 600 meteorological stations during the last hundred years. Special interest was paid to the climatic changes that occurred in the region during the first seven years of the 21 century. The changes in runoff of major Central Asian rivers, such as Naryn, Karadarya, Zerafshan, Amu Darya and its tributary Vakhsh, since the beginning of the last century were studied. Trends in the climatologic and hydrologic data were analysed. The research results provide important insight into climate change and its impacts in Central Asia. The rules of water use and resource-management in the region were developed before the appearance of significant climatic changes in the region. These rules cause growing tension in the relations between countries using water from the same rivers. Adaptive actions based on the research results will be required to reduce water problems and conflicts between neighbouring countries of the Central Asian region.

Keywords: climate change; temperature; precipitation; river runoff;

Introduction

Quite often, from the mass media and from scientific articles, we hear about signs of climate change that occurred on Earth [Boyle, 2001]. But always, only a part of the many processes that take place in the climate changes is demonstrated. Most often average air temperature data are demonstrated. It is difficult to find references about the total amount of precipitation. So, the demonstration of issues related to the yearly distribution of temperature and precipitation is a rare accession. However, this aspect is most important. The study of the intra-annual variability in connection with a change in the average annual temperature, runoff and precipitation is the subject of this article. Only precipitation occurring during the growing season can affect crop production. Average annual temperature precipitation increases have a less important impact for humans than climate change within the year. Conclusions presented in this article, represent the results of numerous statistical calculations using series of historical observations at meteorological and hydrological stations. As expected, there is no direct correlation between the observations of temperature, precipitation and river runoff. The absence of such relationships is caused by multi-year water storage in glaciers at the present time in the region. For example, in a year with low temperatures the precipitation may not cause an increase in river flow because the main amount of precipitation will be accumulated in glaciers. In the next period, if high temperatures occur, this water may appear in the rivers, even if the amount of precipitation is not too great. Finally, for countries in Central Asia the main importance is the water factor and that is why much attention will be payed to the description of existing and possible future changes in water quantity in rivers in correlation to the time of year.
Recently, satellite photos were used for climate change evaluation. For example, photos made by the MODIS satellite have been used for short-term predictions of water flow in rivers. But, there is not enough observation time to evaluate long-term changes in climate. For example, MODIS satellite photos are available only since 2000. There are no doubts that future satellites photos will play a main role in predicting future climate changes. But currently the main source of information about climate change is the series of observations collected at the meteorological and hydrological stations by the usual way.

Observations database

In order to evaluate climatic changes taking place in Central Asia, information about average temperature and precipitation measurements were collected and recorded in an ARCMAP database for more than 1000 meteorological stations during a total period of more than 130 years. Figure 1 shows the GIS project dedicated to mapping the location of the meteorological stations with measurements included in the database.

![Figure 1. GIS project mapping the meteorological stations location with considered measurements of monthly air temperature and monthly precipitation.](image)

Despite the abundance of observation stations, consistent time series datasets are limited as many observation stations were only operated for a limited amount of time. However, there were also stations that continued to work for more than 100 years. From such stations with long records we collected measurements used in statistical calculations.

In addition, changes in air temperature and precipitation over many years and the yearly distribution of air temperature and precipitation caused some changes in river discharges in Central Asia. Such changes are of great importance, as millions of people in the five Central Asian countries rely on the water supply for irrigation and developments. In this paper the information about intra-annual runoff changes in the two major rivers in Central Asia has been investigated on the basis of series of hydrological observations for water content in the rivers at two constantly and continuously operating observation posts:

- Naryn River (the most studied and the main tributary of the Syr Darya) - a tributary for the Toktogul reservoir.
· Vakhsh River (a well-studied and main tributary of the Amu Darya) - a tributary for the Nurek reservoir.

Spatial relationship between measurements of temperature, precipitation and river discharge at stations in Central Asia are analyzed. This is done to identify representative stations that could be used to derive conclusions on regional trends of climate change in Central Asia.

**Precipitation**

More than one hundred stations measure precipitation in Central Asia for many decades. Selection of the representative station for the whole region seems to be an impossible task. Precipitation has a high degree of fluctuation. Maximum precipitation occurs in one region in Central Asia, then in another region. Sometimes even the precipitation at two locations separated by a short distance differs greatly from each other. Fluctuations in the air currents over Central Asia interact with the thermal situation and with another unique atmosphere feature creating a variable precipitation distribution in the region.

Figure 2 shows the typical level of the relationship between the stations located for than 150 km apart.

![Figure 2. Relationship of the annual precipitation quantity at the stations in Samarkand and Jizzakh for the period of 1981-2005 years (R = 0.48).](image)

Also, the relationship between the values of the annual precipitation at more faraway stations doesn’t have a higher correlation coefficient.

- Samarkand - Termez (R = 0.47) for the last 100 years
- Samarkand - Namangan (R = 0.38) for the last 100 years
- Jizzakh - Namangan (R = 0.62) for the last 100 years
Observations at 20 longtime working stations (with data for more than 80 years continuously) in Central Asia were statistically analyzed, and the highest correlation coefficient identified was 0.65 (Tashkent-Dgizak stations).

As precipitation patterns are so diverse, the greatest possible number of operating stations for the last 30-50 years is required to examine the rainfall trends over such a large area as Central Asia. This period is taken as the period in which the features of a changing climate are visible and definable statistical characteristics.

Temperatures

The calculations show that the relationship between measurements of average annual temperature at different positions is considerably closer than precipitation. The most common is the relationship with a correlation coefficient around 0.8-0.9. As an example, consider the relationship between annual temperature values measured at the Tashkent station and the Kungrad station (the distance over a thousand kilometers) where the correlation coefficient is 0.82.

![Figure 3. Relationship between annual temperature values at the Tashkent station and the Kungrad station (R = 0.82) for the last 50 years.](image)

The closeness of the relationship between temperatures at different stations allows us to use the historical data series at the most reliable station to determine the general trend of temperature change throughout the region. Tashkent station was selected as such an appropriate station. This station has the longest number of observations, in which there are no gaps in observations.

River flow

Two river basins were considered - the two major Central Asian rivers - the Syr Darya and Amu Darya. In order to evaluate the relationship of flow in both basins, we define representative observation points. These points must have the following properties:

- Water above them should not be diverted for irrigation;
· A long series of observations should be available;
· Points are located at the main tributaries of major rivers of Central Asia;

We cannot take points on the rivers themselves because their natural discharge flow is distorted with reservoirs work and intakes. So, our choice was the following:

- Syr Darya: We chose a point of observation on its Naryn tributary in the inflow zone above the Toktogul reservoir.
- Amu Darya: We chose a point of observation on the Vaksh tributary in the inflow zone above the Nurek reservoir.

Figure 4 displays the relationship between annual discharges at these points of the observation.

![Relation between average runoff on Narin and Vahsh](image)

**Figure 4.** Relationship between annual water discharge at the cross-sections of the inflow to the Nurek reservoirs and Tohtogul reservoirs \((R = 0.72)\) for the last 50 years.

The high correlation between river discharges at Naryn and Vahsh shows that both of these basins have a similar response to climate conditions (i.e., precipitation as a source of snow accumulation and temperature as a factor determining snow melting).

**Trends in precipitation and temperatures, water content in the rivers of Central Asia.**

To evaluate the trends in precipitation changes in Central Asia we selected observation posts operating continuously for the last 50 years. Statistical analysis showed that the precipitation varies from year to year. For example in Namangan (East Ferghana Valley) rainfall varied from 450 mm to 75mm during two years. Such a strong ripple does not allow us to identify any statistical trends in precipitation in the region. Correlations between precipitation at different locations are very low, which indicates a high variability in processes determining the precipitation pattern in Central Asia [Agaltseva, 2007].

The relationship between air temperatures measured at the various observation posts in Central Asia is quite high. For example the relationship between annual temperature values measured at Tashkent and Samarkand stations (distance is more than 300km) has a correlation coefficient 0.82.
In order to evaluate changes in the amount of precipitation linear trends were counted for stations with records longer than 50 years. Average value of the slope corresponding to the linear trend in the precipitation in Central Asia is 5mm per year (this is correct for the previous 50 years). This amount does not include the increasing evaporation caused by increasing air temperatures. It is possible to conclude that increasing precipitation in Central Asia does not affect the increase of water content in rivers.

The annual average air temperature in the region for the last 30 years has increased from 1.5 (Ferghana Valley) to 2.2 degrees C (Tashkent and Khorezm). These numbers are high and coincide with the values found for the Central Asian region in other researches [Ososkova, 2000]. In the intra-annual context, the situation with air temperature changes looks much worse. Figure 5 shows the course of the average temperature for the winter half of the year and it’s averaging by Tashkent Station.

In figure 5 the temperature increased is clearly visible and almost exponential since 1975, and summer in Central Asia has become a bit cooler. This is the only way to explain the 2.2 degree C annual increase, taking into account results on the figure 5.

Winter is a period of precipitation in Central Asia with snowfield and glacier accumulation in the mountains. In summer, during snowmelt, water flows down from the mountains and appears in the rivers providing irrigation needs. Tashkent is a rapidly growing city and it is quite possible that rapid growth of the winter temperatures is associated with this. However, the growth of winter temperatures is significant for other cities too. For example, the temperature growth in winter at Djizzak city for the last 50 years was more than 3 degrees C. In Termez the average winter temperature increased 17 degrees from 40 years ago. Since 2000, the temperature in the same winter period was 22 degrees C. In Termez the anthropogenic factor in relation to Tashkent is insignificant. These values of increasing winter temperatures are higher than the average increase of temperature in the region and suggest that in most of the region warmer winter seasons have appeared.

We will consider two graphs. Figure 6 shows the chronology of the inflow to the Tohtogul reservoir for the last 60 years. Representative position situated at the largest tributary of the Syr Darya River above, which no water abstraction for irrigation occurs [Shults, 1965]. Figure 7 shows the chronology of the inflow to the Nurek reservoir. This is a representative position at one of the main tributaries of the Amu Darya River. There are no water abstractions discharges for irrigation above this position also [Shults, 1965]. Both locations are consistent in the trend representation of the water content in rivers of Central Asia.
Figure 6. Historical data of the inflow in the Tokhtogul reservoir and the trend for the last 60 years (moving averaging).

Figure 7. Historical data of the inflow in the Nurek reservoir and the trend for the last 50 years.

Analysis of the figures 6 and 7 shows an increase of the water content in the Syr Darya and Amu Darya rivers since 1975 that perfectly coincides with the explosive growth of the temperature (Figure 5). And for the Narin River it is possible to see a sudden rough drop of the water volume in 2005. There was a smoother decreasing in water content of the Vakhsh River for the same period of time. It is important to remember that Narin River mostly fills in with melted snow and melting of seasonal snow mainly forms the water in it. Water of the Vakhsh River in a significant amount is obtained by melting of the perennial ice.

The recent data displayed in Figures 5, 6 and 7 gives a reason to assume that the glaciers and snowfields in Central Asia decreased from 1975 to 2005 year. And after all they have reduced so much that in the future the Amu Darya and Syr Darya Rivers will have only a continuous drought in vegetation periods and water content in these rivers will decrease with the income reducing from the irrigation land. Precipitation will transform to runoff.
without stage of freezing and will be lost for irrigation in Summer time. The simple research was done in another river basins [Beven, 1987].

Probably, the temporary increasing of the water level in rivers for period 1965-1970 coinciding with temperature growth represents the consequences of the glaciers degradation. Situation becomes dramatic because the temperature increasing does not cause the water amount increasing in the rivers as result of the temperature rising since 2000 year. The additional water income from melting will become impossible because there are no any significant glaciers for melting. And remain glaciers will not provide additional water. Moreover, there are evidences that Central Asia glaciers are remains from ancient ice ages. Even without any future temperature growth the glaciers will not appear again after melting in the region.

In the very near future, countries located in the basins of the Syr Darya and Amu Darya Rivers will have to change their water management strategies and implement new water-saving technologies. Otherwise, the region has to expect a series of droughts. Based on the presented analysis we can conclude that Central Asia is strongly affected by climate change, which may increase of conflicts of water usage in the future [Chub, 2000]. A large population is almost entirely occupied in the agricultural production capacity does not have flexibility yet to adjust to a changing climate. Central Asia is one of the groups of the vulnerable regions that will assume the impact of climate change and therefore it needs the assistance of the entire world community with solving the appeared and growing problems associated with the climate change.

**Literature**


The Environmental Systems Modelling Platform (EnSym) to Assess Effects of Land Use Changes on Groundwater Recharge

J. Hå, M. Eigenraam, G. Forbes, W. Lewis and J. Chua

Abstract: This paper describes an application of EnSym (Environmental Systems Modelling Platform) to assess the impacts of climate change on the groundwater levels in Victoria. EnSym is a modular and user-friendly software platform to facilitate the use of environmental modelling tools. It enables easy and rapid evaluation of environmental outcomes due to changes in land management and climatic conditions. It contains a number of toolboxes that deal with different aspects of the environment including land based biophysical process, groundwater dynamics, spatial and contextual connectivity, and finally, a set of tools for systematic spatial and temporal reporting. In this paper, we apply the biophysical modelling (BioSym) toolbox of EnSym to estimate the amount of recharge to the Victorian groundwater system for specified land use scenarios. The groundwater recharge obtained from BioSym forms the transient inflow to the groundwater system. The modular three dimensional finite difference ground water flow model (MODFLOW) is used to simulate the response of the groundwater system to the transient recharge.

We report results of simulating climate change (with a focus on lower rainfall) and its impact on groundwater levels and storage over time. The results can be used as a catchment planning tool, a research tool or to aid cost-effective decision making when planning for future water resource use.

Keywords: Biophysical modelling; BioSym; DFlow; EnSym; groundwater modelling; recharge.

1 INTRODUCTION

It is well recognized that many parts of the world will face significant fresh water shortages in the future, due largely to growing populations and increased agricultural and industrial demands. Fresh water sources could also become impaired through the disposal of wastes, from excessive irrigation and fertilization practices in agriculture, or from simple overproduction and overexploitation. For many communities, the development of new water sources increasingly involves the combined use of surface water and groundwater. The effects of excessive and unsustainable groundwater development may not be immediately evident but ultimately can threaten our natural resources. To sustain groundwater as a long-term reliable resource, factors affecting both the quality and quantity of groundwater must be better understood to inform future decision making. These factors include its abundance, distribution, movement and pollution.

It is recognised that groundwater modelling is the best tool to support management of groundwater resources. In the last two decades, there has been a rapid development of the computational tools for groundwater modelling. In most cases these computer programs are very sophisticated, being able to address a wide range of water-related problems very efficiently, provided the appropriate
data are available to calibrate models in a reliable manner. However, such data are not usually available. The knowledge related to reliable data acquisition, data integration and data extrapolation, particularly regarding spatial data up scaling and spatio-temporal data integration is far less developed than the modelling techniques themselves. In the past few years, the Victorian Government in Australia has been funding a work program to develop reliable groundwater models for all the catchment management regions (watershed) of the State. This work program is still ongoing. Groundwater modelling also requires many source and boundary conditions. These include the location of water bodies, evapotranspiration rates and profile, extraction from and recharge to the groundwater system.

The primary aim of this paper is to report on the modelling system that we have developed to integrate surface and subsurface modelling to assess the effects of land use and climatic changes on groundwater. We call this modelling system EnSym - Environmental Systems Modelling Platform. The second aim is to present some preliminary results from an application of EnSym to case study of the responses of the groundwater system of the Corangamite catchment area due to normal and reduced recharge scenarios. The reduction could be the result of land use change or lower rainfall.

In this paper, we apply the biophysical modelling (BioSym) toolbox of EnSym to estimate the areal distribution of recharge for use as sources of supply to a groundwater flow and transport model. In this paper, we also take evapotranspiration into account in groundwater modelling. Again, we apply BioSym in supplying the information about the evapotranspiration surface and extinction depth. Different amounts of recharge to the groundwater system and a different evapotranspiration regime will result from different specified land use and climatic scenarios. The groundwater recharge obtained from BioSym forms the transient inflow to the groundwater system. In this paper, the modular three dimensional finite difference groundwater flow model (MODFLOW) is used to simulate the response of the groundwater system to the transient recharge. In this way, the impact of environmental changes on groundwater levels and storage over time can be estimated. The results can be used as a catchment planning tool, a research tool or to aid cost-effective decision making when planning for future water resource use.

2 ENVIRONMENTAL SYSTEMS MODELLING PLATFORM (EnSym)

In this section, we will discuss the Environmental Systems Modelling Platform (EnSym) and some of its toolboxes that we have developed for integrated surface and subsurface modelling of the environment at the catchment scale. EnSym is a modular and user-friendly software platform to facilitate the use of environmental modelling tools. It enables easy and rapid evaluation of environmental outcomes due to changes in land management and climatic conditions. It contains a number of toolboxes that deal with different aspects of the environment including land based biophysical process, groundwater dynamics, spatial and contextual connectivity and finally a set of tools for systematic spatial and temporal reporting.

The software provides a stand-alone package that allows user to operate in a “black box” mode, which hides implementation details and usages of the modelling tools. The overlying user interfaces are written in Matlab programming language using a modern design with graphical user interfaces. The environmental modelling tools can be written in any computer programming language. This may, in the long run, contribute to new ways of sharing scientific research. By sharing both data and modelling tools in a consistent framework, the integration and application of new modelling tools into environmental and natural resource management will be straightforward.

The input interface of EnSym will automatically subdivide a catchment and then extract model input data from map layers and the associated relational data bases for each catchment. Soils, land use, weather, management, model and topographic data are collected and transferred to appropriate model input variables. These data sets for modelling the Victorian environment had been collected over a number of years by the Victorian Government. The output interface allows the user to display output maps and numerical and graphical output data by selecting a point from the
map. Users can thus visualise, interpret and test outputs such as sensitive changes in climate, land use and land management practices through a single interface.

EnSym is developed by the Victorian State Government using a version control system to assist in collaborative development, documentation, and feature tracking. While users do not need to study EnSym’s source code, collaborators are welcome to become involved and add new modelling modules, tools and functionalities. Matlab provides gateway wrappers to provide easy access to external modelling programs. One particular design aspect of EnSym is that it can handle dynamic model loading and can easily switch between different tools.

Two of the key toolboxes of EnSym are the biophysical (BioSym) and surface flow (D-Flow) toolboxes. The BioSym toolbox simulates daily soil/water/plant interactions, overland water flow processes, soil loss, carbon sequestration and water contribution to stream flow from both lateral flow and groundwater recharge. The agronomic models can be applied to any combination of soil type, climate, topography and land practice. BioSym can thus be used to evaluate the impacts of climate change, vegetation types (e.g. cropping, grazing, forestry and native vegetation) and land management (e.g. forest thinning and stocking rates) in different parts of the landscape. D-Flow predicts surface water flow directions from digital elevation model (DEM). Flow directions are needed in hydrology to determine the flow paths of water and the movement of sediments, nutrients and contaminants. These two toolboxes of EnSym as well as the groundwater flow model, MODFLOW, are described in the next sections.

### 2.1 Biophysical Modelling (BioSym)

BioSym is a continuous time model that operates on a daily time step. The objective in model development was to predict the impact of management on water, sediment, and agricultural chemical yields in the catchment. To satisfy the objective, the model (a) is physically based (calibration is not possible on catchment scale); (b) uses readily available inputs; (c) is computationally efficient to operate on catchment scale in a reasonable time, and (d) is continuous time and capable of simulating long periods for computing the effects of management changes. The modules in BioSym come from publicly available models. They include CAT (Beverly [2007]), PERFECT (Littleboy et al. [1989]), EPIC (Williams et al. [1989]) and SWAT (Neitsch et al. [2002]). Recently, we upgraded our 3PG+ forest model to its latest version (Feikema et al. [2010]). These models are widely used by the environmental modelling community. The readers are referred to the open literature for references of their developments and model validations.

The physically based models in BioSym provide detailed representations of fundamental processes such as plant growth, infiltration, evapotranspiration, runoff, erosion and sediment transport, nutrient and pollutant transport, stream transport and management practices. By modelling each process separately, the simulation is sensitive to climatic change, land use activities and management changes.

BioSym solves for physical processes conceptually by using simplified analytical solutions and empirical equations. The code for BioSym was written with the objective of simulating all major hydrologic components as simply and realistically as possible, and to use inputs readily available over large spatial scales to enhance the likelihood that the model would become routinely used in planning and water resource decision making.

### 2.2 D-Flow

D-Flow uses the principles of single and multiple flow algorithms, such as Deterministic 8 (D8) and $D_{\infty}$, to direct the flow from each cell to one or more of its 8 neighbouring cells based on the steepest downslope drop. It borrows ideas from image processing to correct the shortcoming of the mentioned flow algorithms in their inability to route flow over flats and sinks as well as to take into account the retention capability of depression drained areas (Chua et al. [2009]). D-Flow follows the flow of water in the catchment, from land areas to streams and rivers, through lakes, to
estuaries and ultimately to the ocean. The use of D-Flow is to move the runoff from one part of the landscape to the next. Water movement is related to erosion, to sediment, nutrient and pollutant transport.

2.3 Groundwater Modelling

Groundwater modelling uses numerical models that approximate the solutions of governing partial differential equations that describe the flow of water in the ground. The ground is typically described as a porous medium with varying densities and water holding capacities. In this paper, the modular three dimensional finite difference groundwater flow model MODFLOW (McDonald and Harbaugh [1988]) is used to simulate the response of the groundwater system to the transient recharge.

3 Simulation Results

In this section, we present the results of our preliminary groundwater modelling of the Corangamite Catchment Management Authority (CCMA) region of Victoria, Australia. The CCMA region covers over 1,335,000 ha or 6% of the State of Victoria. The region is bounded by the Victorian coastline to the south-east, the central highlands (Midlands) to the north, stony rises to the west and sedimentary/volcanic plains to the east (Robinson et al. [2003]). Figure 1 shows the DEM and geomorphology of the Corangamite region. The axes in the DEM figure are in cell units. Each cell is 200 m in length. The DEM is made up of 842 columns and 795 rows. The map of Figure 1(b) taken from the report of Robinson et al. [2003] shows the water bodies and the 3 broad geomorphic divisions of the Corangamite region including the Western Uplands, Western Plains and Southern Uplands.

![Figure 1: (a) DEM and (b) geomorphology of the Corangamite catchment, Victoria.](image)

The model for groundwater flow is made up of 5 layers of 842×795 regular cells of dimension 200×200 m². The height of each layer varies spatially. Not all the cells are active cells for groundwater flow. The cells coloured black in the DEM of Figure 1(a) are inactive cells. The 3-dimensional structure of the 5-layered model is shown in Figure 2. For clarify, only the land
surface and the bounding surfaces of layer 5 of the model are shown for two regions of the catchment. The horizontal coordinates in cell units give the precise locations of these two regions in the Corangamite catchment shown in Figure 1. Figure 2 shows the top surface of layer 5 of our groundwater model reaches the land surface at some parts of the catchment. The regions where the top surface of layer 5 becomes the land surface in Figures 2a and b are in the Western and Southern Uplands shown in Figure 1b respectively.

Figure 2: The layered groundwater model showing the land surface (green), the top (blue) and bottom (magenta) surfaces of layer 5 are shown for two regions of the catchment. For clarity, the other layers of the model are not shown. Figures (a) and (b) show the regions near the top and bottom of the catchment shown in Figure 1 respectively.

Figure 3: Cross-section of the surface of the head on the vertical plane through the line joining A and B of Figure 1a; (a) the temporal fluctuation of the heads of the normal recharge scenario; (b) the temporal fluctuation of the differences between the heads of the normal and reduced recharge scenarios.

The water bodies shown on the map of Figure 1(b) are constant head regions. We assume the initial head is the same as the elevation of the region. In the MODFLOW simulations carried out for this paper, the preconditioned conjugate gradient algorithm is used for solving the simultaneous equations resulting from the finite differencing of the governing equations of the groundwater problem.
Figure 4: Potential head (left) and difference in head between normal and lower recharge scenarios (right) at 3 equally spaced stress periods.
Any water from rain and irrigation that are not consumed by vegetation in the landscape, stored in the soil, lost by evaporation or depleted by lateral flow is assumed to be recharge to the groundwater system. This source of water supply to groundwater is estimated from biophysical simulation of the CCMA region using BioSym. The BioSym simulation also provides the evapotranspiration surface and extinction depth data for the groundwater modelling using MODFLOW. Typically, BioSym simulation over a 50 year period is carried out. However, groundwater modelling demands much greater computer resources and CPU cycles. In the transient MODFLOW simulations we carried out for this paper, the modelling is done over 157 constant stress periods each of which is 7 days long. That is, the CCMA groundwater system is simulated for just over 3 years. In order to compare the effect of lower supply of recharge to the groundwater, the surface and subsurface simulations are repeated for reduced rainfall scenario so that the overall recharge is 50% below the normal scenario. It will be an interesting future study to examine the sensitivities of the groundwater system to different levels of overall recharge changes.

The results of the MODFLOW simulations are presented in Figures 3 and 4. On the left column of both figures, the results of the normal recharge scenario are shown. On the right columns of both figures, the difference between the heads of normal and reduced recharge scenarios are shown. In Figure 3, the vertical axis is in metres and the horizontal axis is in cell units measured from point A of Figure 1(a). Figure 3a shows the cross-section of the surface of the head on the vertical cut plane through A and B of Figure 1(a) for all the stress periods. Each curve in the figure represents the level of the head at a point in time. The vertical spread of these curves give a measure of the amplitude of head fluctuation with time. Figure 3b shows the differences resulting from subtracting the heads of reduced recharge scenario from the heads of normal recharge scenario. Again, the figure shows the cross section of the surface of this difference on the vertical cut plane through A and B of Figure 1(a). The figure shows the magnitudes of the differences vary with time. They can be fairly small at times but can be more than 4 m at other times. Figure 3a shows that the head levels change by small amounts in most places from A to B. Larger variations in the head levels occur near A and B. These large fluctuations in head levels occur in the Southern and Western Uplands of the catchment. In Figure 4, the head levels are colour shaded with low values represented by the blue end and high values represented by the red end of the rainbow spectrum. The results at the end of stress periods 57, 107 and 157 are shown. They show the lateral and vertical movements of the head with time in different parts of the CCMA region. For example, the Southern Uplands generally shows an increase in head levels with time. The southern part of the Western Plains, on the other hand, shows a decrease with time. Animations of the figures will show the dynamics of head movement with time more clearly. On the right columns of Figure 4, the difference between the heads of normal and reduced recharge scenarios are shown. The simulation results show that the major differences between the two scenarios occur in the Southern Uplands. It is not surprising to note that the heads for the reduced recharge scenario are lower for all times than that of the normal recharge scenario.

**CONCLUDING REMARKS**

In Section 2, we present the EnSym (Environmental Systems Modelling Platform) software platform to facilitate the use of environmental modelling tools and some of the toolboxes we have developed for it. Two of the toolboxes we used for obtaining the results for this paper are the BioSym and D-Flow toolboxes. The EnSym software platform can also launch external modelling tool such as MODFLOW that we use for modelling groundwater flow.

In this paper, we show an application example of how EnSym can be used to study the impact of land use and climatic changes on the groundwater system. In particular, we show the integration of surface and subsurface modelling through EnSym to predict environmental outcomes. Using the preliminary groundwater model of the Corangamite region, the BioSym and D-Flow toolboxes were applied through EnSym to obtain the required recharge and evapotranspiration data for groundwater modelling. The groundwater modelling results obtained in the last section show the expected lowering of head for the reduced recharge scenario.
Government policy area is one of the key drivers behind the development of EnSym and the continued improvement and validation of its modelling capabilities. In the area of water resource management, there is the need to develop a protocol for quantifying the State’s groundwater budget. This information will be coupled with projected changes in land use and pumping demand to define the effects of several development scenarios on the community’s water supply. Once developed, this protocol will enable other communities to decide how to best protect vital groundwater recharge areas, where precipitation replenishes local aquifers. This will also help communities examine how changes in groundwater levels will affect local streams, lakes and wetlands. The environmental systems modelling platform EnSym facilitates the incorporation of the various interactive influences into the resource management decisions. The simulation results presented in the last section demonstrate that EnSym can be used as a catchment planning tool, a research tool or to aid cost-effective decision making when planning for future water resource use.

REFERENCES


The Fate of Devils Lake: An Interwoven Aftermath of Agriculture and Climate Change

Xiaodong Zhang
Department of Earth System Science and Policy, University of North Dakota
zhang@aero.und.edu

Abstract: A dry period in the mid-20th century caused the lake area to become greatly reduced, allowing people to build and begin farming on the floodplain. Since 1993, Devils Lake has risen a dramatic 8.5 m in elevation to 442 m, flooding both residential and agricultural zones in the surrounding area. Rising another 3 feet to an elevation of 445 m, the lake would overspill. To date, $1 billion in federal, state and local funds have been spent to help mitigate flooding in the Devils Lake basin. It will cost several hundreds of millions if protection of overspill is needed. Some communities, displaced families and businesses have lost everything, and the future of those who remain is clouded by the very real possibility of continued disaster. A model is currently under development that will incorporate satellite observation to investigate the effect on the lake level of agriculture expansion and future climate change. The results will be of great value in helping local community to solve a real problem and make a wise (and educated) decision.

Keywords: Devils Lake; flooding; remote sensing; land use change; climate change.

1. Introduction

The Devils Lake basin is a 9,868 km² closed basin in the Red River of the North Basin (Fig. 1). At current water levels, the lake itself has no natural outlet and is deemed as an endorheic (terminal) lake with evaporation serving as the only mechanism of water loss.
At the elevation of 441 m above sea level (asl), Devils Lake starts to spill into Stump Lake to the east. And at an elevation of about 444.7 m asl, the combined lakes begin to spill through Telna Coulee into the Sheyenne River, a tributary of the Red River, which flows into Canada. Since the end of glaciations about 10,000 years ago, Devils Lake has been fluctuating from overflowing to dry [Bluemle, 1991]. Research by the North Dakota Geological Survey indicates Devils Lake has overflowed into the Sheyenne River at least twice during the past 4,000 years and has spilled into the Stump Lakes several times [Bluemle, 1991; Murphy et al., 1997].

In 1867, when lake surface elevation was first measured, the lake stood at elevation 438 m asl, covered about 337 km², and contained about 1.85 km³ of water. By 1940, after several years of extreme drought, the lake had diminished even further, dropping to a record-low elevation of 427 m asl and covering only 26 km² [Wiche and Puscheck, 1994]. Since 1940, the lake has exhibited a dramatic resurgence – particularly over the past 16 years – largely in response to a substantially wetter climate (Fig. 2). On April 2, 2010, Devils Lake crept to a new record level, surpassing the previous record of 442.18 m asl, set on June 27, 2009. Today, the lake covered about 622 km² and contained about 4.07 km³ of water.

Unlike riverine floods, which are primarily triggered by short-term precipitation or snowmelt events, Devils Lake flooding primarily depends on long-term climatic conditions in the basin. Water rises slowly, but the higher water levels persist. Therefore, understanding future regional climate under all possible scenarios is critical for understanding the hydrology of the lake.

Retreat of glacier about 10,000 years ago not only left Devils Lake but also created numerous potholes, which had given the North Dakota’s prairies the title “America’s Duck Factory”. Rapid agricultural expansion had drained nearly all of the prairie wetlands in North Dakota by 1930’s. The research conducted by West Consultant Group in 2001 showed that 374 km² of wetland in the Devils Lake basin had been drained [West Consultants, 2001]. United States Fish and Wildlife Service estimated that 769 km² of wetland have been drained, as compared to the estimate of 607 km² given by North Dakota State Water Commission [Noone, 2003]. Some of the conservation groups claim that 1416

![Figure 2: Water level of Devils Lake and monthly precipitation recorded at the U.S. Historical Climatology Network station Langdon, ND (48.7622 N, 98.3447 W). The precipitation data was smoothed by a 6-year running window.](image-url)
km² of wetland have been drained. Drainage of wetlands in the glaciated prairie region can greatly increase discharges [Moore and Larson, 2980]. At present, 70% of land in the Devils Lake basin is used for crop production and 21% is used for ranching or participates in Conservation Resource Program (Fig. 3).

Modern records show that water levels in Devils Lake have increased dramatically three times (Fig. 2), during the years 1940 – 1956, 1983 – 1987 and recently since 1993. However, the recent rise inflicted the most significant damage because much of the area’s development occurred during the previous period of low lake levels. So far, an estimated $1 billion has been spent on flood mitigation. However, because of the nature of encroachment of lake rising, many of the flood mitigation projects were planned for short term benefit. Had it be known that the wet cycle would last up to now, many decision would have been made differently, probably saving a significant amount of money (Belford, personal communication).

If the current wet cycle continues, an improved knowledge of how the lake levels will respond to the future climate is critical for making decisions with long term applicability. Currently, a Devils Lake stochastic simulation model developed by the U.S. Geological Survey (USGS) [Vecchia, 2002; 2008] is being used to predict the potential water level. The analyses, in terms of exceedance probability, however, are all based on ground observations (e.g., lake or rain gauges) and on the assumption that the climate is stationary - regional climate do not change with time. Thus, the impacts of future climate change and climate policy on regional hydrology are not incorporated into the decision-making process.

In this study, the hydrological state of Devils Lake in the future was evaluated by driving a distributed hydrological model with different climate change scenarios using multiple General Circulation Models (GCM) projections. A salient feature of our approach is that NASA’s satellite observations data were used, both providing a more uniform and denser distribution of input parameters required for the hydrologic modelling than ground-based observations and serving as a basis for generation of future climate scenarios.

2. Method and Data

Because the Devils Lake Basin is much larger than the lake area itself, a hydrologic model is essential to characterize the precipitation–runoff process. The lakes and channels are frozen for the majority of the winter months, while snow covers the overland basin areas. Thus, the capability to model water movement within the basin is even more critical. For engineering design purposes, the hydrologic and hydraulic models must be able to provide lake levels at sub-feet accuracy. Notably, an inch in lake level can imply a 10,000-acre-foot difference in the calculated water balance [Vecchia, 2008]. A combination of a distributed hydrologic model and a detailed lake/reservoir hydraulic model is, thus, needed to provide lake level information at sub-feet accuracy. The U.S. Army Corps of Engineers’ Hydrologic Modeling System (HEC-HMS) for the rainfall-runoff model and the Reservoir System Simulation model (HEC-ResSim) for reservoir modeling and flow routing were used for hydrological modelling. For detailed description of the models and the associated results, refer to a companion paper by Lim et al. [2010].

Comprehensive analysis of possible impacts of climate change requires a set of simulations combining a variety of GCM runs under a variety of scenarios because 1) different paths of
economic and societal development will have different radiative forcing; 2) the results of different GCMs diverge from each other; and 3) even a single GCM running under the same scenario will return different results due to stochastic nature of weather generation. We will estimate future climate conditions in the Devils Lake region by combining the historical data on temperature, precipitation, air humidity, and wind speed with an ensemble of GCM projections. For base climate, we will use the 1971-2000 monthly values for temperature and precipitation. For future climate, we extract monthly projections of seven different GCMs: CCMA_T63, CSIRO, GFDL_CM2, GISS_E-R, MPI, NCAR_PCM, and UKMO, for three pre-set time periods, 2020s, 2050s, and 2080s. Additionally, for each of the GCMs we employ three SRES scenarios, A1B, A2, and B1, \cite{IPCC}. For detailed description on climate modelling, refer to a companion paper by Kirilenko \cite{Kirilenko}.

A variety of NASA data products and model outputs, including precipitation, temperature, soil moisture, and other parameters, were used as inputs to the project’s hydrological and climate models. Table 1 summarizes the prospective data.

<table>
<thead>
<tr>
<th>Instrument/Model/Parameter</th>
<th>Spatial Resolution</th>
<th>Spatial Coverage</th>
<th>Temporal Resolution</th>
<th>Temporal Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPRM\textsuperscript{a} Soil Moisture (AMSR-E, TMI, SSM/I, SMMR)</td>
<td>¼ deg</td>
<td>Global</td>
<td>Daily</td>
<td>2002-Present</td>
</tr>
<tr>
<td>TMPA\textsuperscript{b} Precipitation (TMI, SSM/I, AMSR-E, AMSU-B)</td>
<td>¼ deg</td>
<td>Global</td>
<td>3-hourly, Daily</td>
<td>1998-Present</td>
</tr>
<tr>
<td>Aqua AIRS Surface Air Temperature</td>
<td>1.0 deg</td>
<td>Global</td>
<td>Gridded</td>
<td>2002-Present</td>
</tr>
<tr>
<td>GLDAS\textsuperscript{c} Surface Radiation, Snow Water Equivalent, etc.</td>
<td>¼ deg</td>
<td>Global</td>
<td>3-hourly</td>
<td>2000-Present</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Land Parameter Retrieval Model; \textsuperscript{b}TRMM Multi-satellite Precipitation Analysis; \textsuperscript{c}Global Land Data Assimilation System.

Fig. 4 summarizes the process of combining climate model, hydrological model and NASA data in predicting future water level. In addition climate scenarios, the models developed
can also be used to investigate the impact of different land use/land change on the lake hydrology.

3. Results

The extent of expansion of Devils Lake can be clearly seen from a time series of satellite imagery (Fig. 5). From 1991 (Fig. 5-a) to 2009 (Fig. 5-c), the main water body of Devils Lake had expanded 453 km² (Fig. 5-d), with immediate loss of fertile land and commodity valued between $60 - $100 million. In Fig. 5, only the main water body of Devils Lake was shown. Including the smaller lakes in the upper drainage basin (Fig. 1), the rising water has inundated 598 km² of land, with a loss of $130 million in land and commodity.

Satellite data from NASA’s Tropical Rainfall Measuring Mission (TRMM) for precipitation, from Advanced Microwave Scanning Radiometer – Earth Observing System (AMSR-E) on soil moisture, and from Atmospheric Infrared Sounder (AIRS) for air temperature were used to provide input parameters for the hydrologic models (Fig. 6). As shown in Fig. 2, the increase in precipitation during the past 17 years is the major reason for the lake rise and expansion. However, the soil moisture concentration, which currently is not measured regularly on the ground, is also important in regulating the discharge of water into the streams. The measurements from AMSR-E, even though only valid for the top few centimetres of soil, can provide at least an approximate estimate of saturation state of the soil. The air temperature near surface is an important parameter for estimating potential evapotranspiration and snow melting. The latter is of particular importance because spring melting of snow leads to rapid rises in water level.

Preliminary comparison showed that the satellite-derived precipitation measurement agree well with the rain-gauge measured values on the ground (Fig. 7), with a correlation
The coefficient of 0.97. Currently, the satellite data on the other parameters of air temperature and soil moisture is being evaluated against the ground measurements.

Fig. 6. Examples of satellite derived parameters used as inputs to the hydrologic and hydraulic models as well as to the weather generator. (a) TRMM precipitation, (b) AMSR-E soil moisture, (c) and (d) time-series of TRMM precipitation and AIRS air temperature near the surface over the entire basin from 1998 to 2009.

Fig. 7. Daily-cumulative precipitations measured by the rain gauge (maintained by National Climate Data Center) and TRMM in the years of 2001 and 2002.
4. Conclusions

With nearly $1 billion has been spent on mitigating flood impacts caused by rising water of Devils Lake, the future variation of the lake is of great concern to people, the government agencies and all parties involved. Because of non-stationary nature of climate, especially under the anthropogenic influence due to continuously increased release of green house gases, future climate projections need to be investigated for a variety of possible scenarios. A physical hydrologic model combining watershed and reservoir is under development and will be used to evaluate how the changing climate in the future will affect the water level. Satellite observations, continuous and well distributed spatially, will improve the accuracy of the model. A combination of climate model, hydrologic model and remote sensing offers great potentials in helping decision making of real world problems related to water.

ACKNOWLEDGMENTS

This study was supported by NASA grant NNX09AO06G “The Feasibility of Combining NASA Satellite Data, General Circulation Models, and Hydrologic Models to Inform Decision Making for Flood Mitigation of the Devils Lake Basin of North Dakota”

REFERENCES

Kirilenko, A. (2010), Climate change impact on agriculture: Devils Lake basin, in 2010 International Congress on Environmental Modelling and Software Modelling for Environment's Sake, edited, Ottawa, Canada.
The Impact of Climate Change on Agriculture Water Resources for Paddy Rice over Southern Taiwan

Pao-Shan Yu*, Hung-Wei Tseng², Shien-Tsung Chen³

¹ Department of Hydraulic and Ocean Engineering, National Cheng Kung University, Taiwan

* yups@ncku.edu.tw

Abstract: This work aims to investigate the impacts of hydrologic drought on agricultural water resources under climate change scenarios. The study area is Tseng-Wen Reservoir basin in southern Taiwan, which receives temporally uneven precipitation. It is thus a basin prone to suffer from drought during dry season. General circulation models (GCMs) are the main tool to tackle climate change issues through the help of prescribed scenarios. This work used several approaches, including spatial downscaling, temporal downscaling and hydrologic model, to solve the coarse-resolution problem of GCMs and then to analyze the effect of climate change in the study area. The following are important findings: (1) According to future climate projections, droughts may become more frequent (hereafter referred to as scenario droughts), but their duration and magnitude may become more diverse than those of the baseline droughts. (2) The times of start and end of scenario droughts may occur earlier than those of baseline droughts. (3) Scenario low flow during the dry period tends to decrease in January and February, but to increase in March and April. (4) Moderate adjustment of irrigation period to adapt to climate change is suggested.

Keywords: climate change, downscaling methods, hydrologic drought.

1. INTRODUCTION

Numerous investigations have studied rainfall trends in various areas worldwide and have found that rainfall characteristics vary according to regions, and have increasing or decreasing trends. Yu et al. [2006] found that annual rainfall in southern Taiwan has decreased significantly during the past century. This work investigated the hydrologic drought pertaining to streamflow in Tseng-Wen Reservoir basin in Taiwan under climate change scenarios. Droughts can be classified as meteorological drought, hydrologic drought and agricultural drought based on the question considered. The threshold level approach proposed by Yevjevich [1967] has been the most commonly applied approach to drought studies, such as Kjeldsen et al. [2000], Shiau et al. [2001], Hisdal et al. [2001], Shiau [2003], Fleig et al. [2006], and Shukla et al. [2008]. This work also defines a hydrologic drought event by using the threshold level approach and aims to investigate impacts of climate change on agricultural water resources in southern Taiwan, in which the temporal precipitation is very uneven. Around 90% of annual precipitation occurs during the period from May to October (wet period) but only 10% of annual precipitation occurs from December to April (dry period). Although there is less precipitation during the dry period, the temperature is more suitable for paddy rice growth. Recently, the shortage of water resource during the dry period always causes the paddy field to lie fallow. Whether the water resource distribution during this period is influenced by climate change in the future is our concern in this work. Tseng-Wen Reservoir, the major water resource for dry period in southern Taiwan, is chosen as the study area in this work.
2. STUDY AREA AND DATA SET

Tseng-Wen Reservoir, with a storage capacity of about $7.8 \times 10^8$ m$^3$, is the largest reservoir in Taiwan. Tseng-Wen Reservoir was completed in 1973, having multifunction of the water demands for agriculture, domestic use, flood control and hydropower generation. Tseng-Wen Reservoir basin encloses an area of 481 km$^2$ (Figure 1), and is at an elevation of from 157 to 3,514 m above sea level. The mean annual precipitation is about 2,740 mm, of which nearly 90% occurs during the wet season (Figure 1).

Figure 1. Tseng-Wen Reservoir basin (left); precipitation (center) and rainy days (right) over the basin.

Climate and hydrological data used in this work contain local-scale and large-scale data. Local-scale data, including precipitation, streamflow and temperature, are provided by Water Resource Agency, Taiwan. Long-term daily precipitations (1974–2008) are available from eight raingauges, from which areal precipitations on Tseng-Wen Reservoir basin were computed using the Thiessen polygon method. In this work, historical streamflow data are regarded as a datum, baseline, against which change is compared.

Large-scale data, i.e. general circulation models (GCMs) data, were downloaded from the Data Distribution Center of the United Nations Intergovernmental Panel on Climate Change. The climate scenarios describe the emission conditions of greenhouse gas and there are more details in IPCC [2007]. Different scenarios used were for historical climate (20C3M), and future climate (A1B and B1). Generally speaking, 20C3M is the scenario to represent past climate, A1B is regarded as the most likely climate scenario in the future and B1 describes a convergent world that the impact on climate change is slighter than A1B. The time periods of historical scenario and future scenario data are 1975–2000 and 2010–2045 respectively. Table 1 lists the summary of six adopted GCMs.

Table 1. Summary of the GCMs used in this work.

<table>
<thead>
<tr>
<th>Model</th>
<th>Country</th>
<th>Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>CGCM3.1(T63)</td>
<td>Canada</td>
<td>T63,L31</td>
</tr>
<tr>
<td>ECHAM5/MPI-OM</td>
<td>Germany</td>
<td>T63,L31</td>
</tr>
<tr>
<td>CSIRO-Mk3.0</td>
<td>Australia</td>
<td>T63,L18</td>
</tr>
<tr>
<td>GFDL-CM2.0</td>
<td>U.S.A.</td>
<td>2°×2.5°,L24</td>
</tr>
<tr>
<td>GFDL-CM2.1</td>
<td>U.S.A.</td>
<td>2°×2.5°,L24</td>
</tr>
<tr>
<td>MIROC3.2(hires)</td>
<td>Japan</td>
<td>T106,L56</td>
</tr>
</tbody>
</table>

3. METHODOLOGY

3.1 Daily Rainfall Downscaling

GCMs are used as the main tool to project climate changes through the use of emission scenarios. However, due to the coarse resolution, GCMs are not able to represent regional topography and land-sea contrast properly, making local climate projection a big challenge. Thus a two-stage statistical downscaling method was applied to generate future daily precipitation data from climate outputs run by six GCMs. In the first stage, spatial statistical downscaling was applied by using the singular value decomposition (SVD)
scheme (Chu, J.L. et al. [2008]) to downscale the monthly precipitation from six GCMs. In the second stage, the projected changes of monthly precipitation were further used in a weather generator to project the daily precipitation, as shown in Figure 2. Daily data are more practical for hydrological purpose. After the daily precipitation is generated, the hydrological model uses daily rainfall as an input to simulate daily streamflow.

3.1.1 Spatial Statistical Downscaling

Spatial statistical downscaling provides relationships between local and large-scale variables to overcome the drawback of GCMs’ coarse-resolution. The first step of spatial downscaling is data reconstruction by the empirical orthogonal functions (EOFs) to filter off noises. This method is the same as principal components analysis. Then the SVD is applied to extract coupled patterns between local precipitation and large-scale variables. The analysis from SVD-based spatial statistical downscaling scheme shows that the rainfall over the study area is closely tied to the large-scale circulation over the east Asia monsoon region, and that GCMs perform reasonably well in simulating the mean states of sea level pressure (SLP) and meridional wind field at 850hPa (V850) over this region. Consequently, the two large-scale variables, SLP and V850, both of which are taken from six GCMs involving 20C3M, A1B, and B1 scenarios, are used as predictors for downscaling. After spatial downscaling process, the results from these scenarios are used to calculate the future change rates of precipitation. The change rates are the ratio of A1B to 20C3M and the ratio of B1 to 20C3M. Historical monthly rainfall multiplied by change ratios are future projections of monthly rainfall under A1B and B1 scenarios.

3.1.2 Temporal Statistical Downscaling

In the second stage, a stochastic weather model was applied to downscale the monthly precipitation, derived in the first stage, to daily precipitation (Figure 3). The daily precipitation generation is based on procedures proposed by Richardson [1981]. The generator uses a Markov chain to model the occurrence of wet or dry days, and a probability distribution to generate the precipitation amount conditional on a wet day modeled by the Markov chain. A first-order two-state Markov chain was used in this work. The occurrence of a dry or wet day is modeled by a transition probability matrix consisting of conditional probabilities, given a previous dry or wet day.

Many probability distributions were applied to generate daily precipitation amount, such as the exponential distribution (Selker et al. [1990]; Tung et al. [1995]), Weibull distribution (Yu et al. [2002]), two-parameter gamma distribution (Richardson [1981]; Coe et al. [1982]; Woolhiser et al. [1982]; Schubert [1994]; Corte-Real et al. [1999]), and mixed exponential distribution (Woolhiser et al. [1979]; Woolhiser et al. [1982] [1986]). Among the probability distributions, the Weibull distribution approximates daily rainfall in Taiwan the best (Yu et al. [2002]); consequently, it was used to generate daily rainfall.
3.2 Hydrological Model

A continuous hydrologic model was needed to simulate future projected streamflow, after the daily precipitation under the A1B and B1 scenarios were obtained in the previous section by the downscaling. This work used a continuous hydrologic model based on the structure of Hydrologiska Byråns Vattenbalansavdelning (HBV) model (Bergström [1976][1992]), which was initially designed for use in Scandinavian catchments by the Swedish meteorological and hydrological institute. Yu and Yang [2000] adapted the HBV model structure to suit catchments in Taiwan. The modified HBV model uses both an upper and lower tanks to model the rainfall-runoff behavior. Model structure mainly consists of three parts: (1) soil moisture module, (2) runoff response mechanism, and (3) water balance functions. Detail description of the modified HBV model, as well as its calibration and validation in this work, can be found in Yu and Yang [2000] and Yu et al. [2002]. Historical daily rainfall and flow data from 1975 to 1998 were used for model calibration. The calibrated continuous hydrologic model was further verified by historical data from 1999 to 2008. The results from this work found the continuous hydrologic model to be able to simulate the rainfall-runoff behavior over the study area.

3.3 Hydrological Drought Event

The threshold level method is applied to define a drought event. In order to study the features of major hydrological drought, this work defines a hydrologic drought as a low flow event when streamflow series is continuously below $Q_{50}$ (50% probability of exceedance) and the minimum flow during the period is less than $Q_{90}$ (Figure 4). The $Q_{90}$ is required here to avoid minor drought events. Once a drought event is specified, drought characteristic can be quantified. Drought frequency [time/year] is the times of drought occurrence per year. Drought duration [day] is the time period between drought start and end. Drought magnitude [mm] is the total amount of streamflow deficit, expressed in depth. Moreover, to avoid the dependent droughts and slight droughts, a 7-day moving average is used to preprocess the simulated streamflow data.

Figure 4. Definition of drought.
4. RESULTS AND DISCUSSION

In following discussion, projected data are in view of the most likely scenario (A1B) and oncoming time period (2010–2045). By comparing differences between historical data (baseline) and projected data, the impact of climate change can be assessed. Table 2 shows drought characteristics, including drought frequency, duration, magnitude, based on historical data (baseline) and projected data (results from GCMs under A1B scenario). The scenario droughts exhibit a more frequent trend (0.87~1.02 times per year, the minimum and maximum of the GCMs, respectively.), compared with the frequency of baseline droughts (0.77 times per year). The durations of projected droughts are between 132 days and 193 days, and the magnitudes of projected droughts are between 105 mm and 179 mm. The duration and magnitude of scenario droughts maybe increase or decrease, as they depend on GCMs. Overall, scenario hydrological droughts may become more frequent, but their duration and magnitude may become more diverse than the baseline droughts.

Table 2. Impact of climate change on hydrologic drought

<table>
<thead>
<tr>
<th>Model</th>
<th>Frequency (time/year)</th>
<th>Duration (day)</th>
<th>Magnitude (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>0.77</td>
<td>174</td>
<td>148</td>
</tr>
<tr>
<td>CGCM3.1(T63)</td>
<td>0.87</td>
<td>132</td>
<td>105</td>
</tr>
<tr>
<td>ECHAM5/MPI-OM</td>
<td>0.99</td>
<td>159</td>
<td>137</td>
</tr>
<tr>
<td>CSIRO-Mk3.0</td>
<td>1.02</td>
<td>193</td>
<td>179</td>
</tr>
<tr>
<td>GFDL-CM2.0</td>
<td>0.98</td>
<td>182</td>
<td>165</td>
</tr>
<tr>
<td>GFDL-CM2.1</td>
<td>0.89</td>
<td>150</td>
<td>126</td>
</tr>
<tr>
<td>MIROC3.2(hires)</td>
<td>0.90</td>
<td>152</td>
<td>127</td>
</tr>
<tr>
<td>Model average</td>
<td>0.94</td>
<td>161</td>
<td>140</td>
</tr>
</tbody>
</table>

Figures 5 and 6 are the time distributions of start and end of scenario droughts respectively. In the figures, the solid line indicates the baseline drought, and the dash line stands for the scenario droughts, which is the ensemble mean of GCMs. Figure 5 indicates that baseline droughts may occur during the period from mid-October (2nd 10-day period of October) to early November (1st 10-day period of November). However, there is a single-peak form in scenario droughts. Relative to historical data, scenario droughts come early and concentrate in mid-October (2nd 10-day period of October). Figure 6 reveals that scenario droughts may end earlier than baseline droughts ending in early May (1st 10-day period of May). Generally speaking, drought may start and end earlier in the future.
Figure 7 presents the distribution of streamflow during the dry period. Solid lines and dash lines represent baseline drought and scenario drought respectively. For solid lines, the upper is 95th percentile, the lower is 5th percentile and the line in bold is median (50th percentile). The same notation is used for baseline data (dash lines). The scenario flow, which is derived from the ensemble average of GCMs, tends to decrease in January and February (at the beginning of paddy rice growth season) but to increase in March and April (at the end of paddy rice growth season). The 90% confidence interval of scenario flow is wider than baseline flow after February, indicating the future projected flow is more diverse.

5. CONCLUSIONS

This work successfully applied downscaling methods and a hydrological model to assess the effect of climate change on agriculture water resources in southern Taiwan. The results of drought characteristics indicate that scenario droughts may become more frequent, but their duration and magnitude may become more diverse than the baseline droughts. Analyzing time distributions of scenario drought finds that 10-day streamflow patterns are changed during the period of paddy rice growth. Generally speaking, drought event will start and end earlier in the future. The scenario flow tends to decrease in January and February (at the beginning of paddy rice growth season) but to increase in March and April (at the end of paddy rice growth season). The 90% confidence interval of scenario flow is wider than baseline flow after February, because diverse property of future flow is projected. Due to the changed pattern of streamflow, moderate adjustment of irrigation period to adapt to climate change is suggested.

ACKNOWLEDGMENTS

The authors would like to thank the National Science Council of the Republic of China (Taiwan) for financially supporting this research under Contract No. NSC 97-2625-M-006-014.

REFERENCES


Performance Evaluation of Environmental Models

Neil D. Bennett1, Barry F.W. Croke1,2, Anthony J. Jakeman1, Lachlan T.H. Newham1 and John P. Norton1

1 Integrated Catchment Assessment and Management Centre, Fenner School of Environment and Society, The Australian National University, Canberra ACT 2600, Australia

2 Department of Mathematics, The Australian National University, Canberra ACT 2600, Australia (barry.croke@anu.edu.au)

Abstract: For environmental models to be effectively used for management and decision making purposes it is necessary to have confidence in their performance. This paper reviews techniques available for performance evaluation of environmental models. Both quantitative and qualitative performance evaluation are considered and application recommendations for environmental models are provided.

Keywords: Model development; model testing; performance indicators; sensitivity analysis

1 INTRODUCTION

Environmental models are increasingly and extensively used in research, management and decision making. Hence, the importance of assessing confidence in the outputs of such models has increased. The question of model evaluation or how well a model represents the system under study has resulted in many different approaches and much debate on the appropriateness of these techniques. The best method will depend on the type of model, the data and aims of modelling, and multiple methods may be needed for the best understanding and decision support. This paper reviews evaluation methods and criteria available for environmental models. Although the primary applications under consideration are environmental, methods that have been developed in other fields are also included.

Modelling is an essential component of environmental management and science, necessary for understanding and representing environmental systems as well as increasing confidence in management decisions. Currently modelling is used across many environmental fields – hydrology, oceanography, climate change to name a few. In each of these fields many different types of model can be used, which will affect how the performance of a model is evaluated. Jakeman et al (2006) separate models into the following model families:

- **Empirical**: data-based, statistical models where a versatile structure is assumed with minimal assumptions. Examples include cluster analysis, time series models and regression analysis.
- **Stochastic**: general form models that have a standard structure allowing the incorporation of previous knowledge and uncertainty e.g. state space and hidden Markov models.
- **Specific process/theory based models** (usually called deterministic): have a set structure specific to the process and justified by prior theory.
- **Conceptual models**: create a structure based on assumed cause-effect links e.g. Bayesian decision networks and compartmental models.
Bennet, N.D. et al., Performance Evaluation of Environmental Models

- Agent-based models: locally structured models that allow for emergent, unpredicted behaviour.
- Rule-based models: a group of models that use rules to represent and simulate the interactions/behaviour between discrete events and decisions. The model can then map the probability of different outcomes including decision trees and expert systems.
- Models incorporating dynamics: a spectrum of models that can give time-spread responses to an input at any instant. The spectrum includes discrete event/state, lumped dynamical, distributed and delay-differential infinite-state-dimensional models.
- Spatial models: including region based, polygon based and pseudo-continuous spatial models.

Both the intended application of the model and its main features will affect the overall behaviour of the model. Therefore, when selecting a performance evaluation procedure, it is necessary to take these details into account.

Verification and validation of models are core components of the modelling process and significant research has focussed on these topics, including philosophical debates to differentiate verification from validation (Jakeman et al. 2006, Oreskes et al., 1994). This paper is not concerned with such debate; instead it deals with the performance evaluation methods assessing environmental models, whether these methods and criteria are used for verification, validation or calibration. At its highest level, performance evaluation can be split into two categories: quantitative methods that test the model against measured data; and qualitative testing, which does not require independent data to evaluate the model.

2 QUANTITATIVE TESTING

Many different methods have been introduced to evaluate the performance of a model. In hydrology previous work has been completed on general modelling frameworks that consider performance criteria as part of the framework (Jakeman et al. 2006 and Wagener et al. 2001), while studies have also been completed that focus explicitly on performance criteria. Moriasi et al. (2007) produced guidelines for systematic model evaluation, including a list of recommended evaluation techniques and performance metrics. Dawson et al. (2007) has also produced a comprehensive list of metrics that can be applied to hydrological forecasting models as well as a web-based toolbox, HydroTest, that can be used to calculate the metrics.

Different techniques that can be used to test models are:

- Data division methods
  - Cross validation (e.g. Kohavi, 1995, Klemes, 1986)
  - Bootstrapping
- Direct comparison methods
  - e.g. plots of data points and frequency distributions
  - statistical metrics comparing modelled with observed values
  - regression of observed and modelled values, or sum and difference of observed and modelled values (Kleijen et al., 1998)
- Residual methods
  - Graphical methods: e.g. residual and Q-Q plots
  - Numerical methods: e.g. bias, mean square error, mean absolute error, maximum absolute error, error in peak, relative volume error
  - Use of transformations to handle heteroscedasticity in the residuals: e.g. Box-Cox transformation (note care is needed as deficiencies in the model structure may contribute to the heteroscedasticity in the residuals)
  - Impact of variations in uncertainty through the data: e.g. Heteroscedastic Maximum Likelihood Estimator to allow for variation in the divergence of the residuals (Sorooshian and Dracup, 1980). See also Croke (2007 and 2009) for a more general form, including allowing for serial correlation.
  - Application of methods listed here on subsets of the data – selection of subsets based on different aspects of the system response (e.g. Boyle et al. 2000) or using a moving window (e.g. Choi and Beven, 2005) for example.
Bennet, N.D. et al., Performance Evaluation of Environmental Models

- Information criteria: Akaike Information Criterion (Akaike, 1974); Bayesian Information Criterion (Schwartz, 1978); Young Information Criterion (Young, 2001, Taylor et al., 2007)
- Model efficiency: (e.g. coefficient of determination, correlation coefficient)
- Parameter error and identifiability: (e.g. Jakeman and Hornberger, 1993; Young et al., 1980; Cecchini et al., 2006, Wagener et al., 2003, Härdle et al., 2003)
- Transformation methods
  - Fourier and Wavelet transforms to convert residuals to the frequency domain (e.g. Lane, 2007, Chou, 2007)
- Spatial methods
  - Global and local spatial methods: global methods act over the entire spatial domain (ignore any spatial characteristics) while local methods are applied over restricted domains.
  - Grouping spatial methods: e.g. defining homogeneous regions (Wealands, 2005); multi-scale approaches, empirical orthogonal functions (Hannachi et al., 2007).
  - Categorical spatial methods: e.g. confusion matrix (Congalton, 1991); kappa statistic; fuzzy maps (Wealands, 2005).
- Multi-criteria methods: typically involve the concept of a Pareto optimal set (Gupta et al., 1997; Yapo et al., 1998)
- Diagnostic based evaluation methods: consider the information contained in the data (Gupta et al., 2008), exploring impact of model structure (Clark et al. 2008).

3 QUALITATIVE TESTING

In the case when data is unavailable for quantitative testing qualitative testing, which aims to provide a consistent means to compare model performance, is the only form of testing possible. It is, however, highly beneficial for performance testing even when data is available for quantitative testing and should be included within a standard model development routine (Jakeman et al, 2006).

The core component of qualitative testing is a face validation or Turing test, which calls for the analysis of the output and operation of the model to see if it behaves as is expected. Two potential methods to contribute to this analysis are standard questions and sensitivity analysis.

3.1 Standard Questions

Standard questions comprise a list of questions the modeller (and potentially an independent expert) should ask about the construction, operation and output of the model. They help to identify uncertainty in model components, unexpected behaviour and areas where improvement is required. (Parker et al, 2002 and Risbey et al, 1996) A list of standard questions is provided in
Table 1.
Table 1: Qualitative questions for model evaluation.

<table>
<thead>
<tr>
<th>No</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.1</td>
<td>How reliable is the input data? What are the uncertainties in the input data? (e.g. measurement error, sampling rate) How does this affect the model?</td>
</tr>
<tr>
<td>9.2</td>
<td>Is the model behaving as expected? How is the model behaviour affected by any assumptions required for the development of the model?</td>
</tr>
<tr>
<td>9.3</td>
<td>Does the model structure reflect the system i.e. Is the model structure plausible?</td>
</tr>
<tr>
<td>9.4</td>
<td>Is the model over-fitted?</td>
</tr>
<tr>
<td>9.5</td>
<td>Is the model flexible/transparent?</td>
</tr>
<tr>
<td>9.6</td>
<td>Have alternative model structures/types been tested? Why was the current model structure selected?</td>
</tr>
<tr>
<td>9.7</td>
<td>Does the model meet its specified purpose?</td>
</tr>
<tr>
<td>9.8</td>
<td>How realistic and optimal are selected parameter values?</td>
</tr>
</tbody>
</table>

4 SENSITIVITY ANALYSIS

Sensitivity analysis explores how a change in parameter values effect the overall change in the output of the model. This can be completed using simple sensitivity analysis, where only one parameter is changed or more complex arrangements that explore the relationships between multiple parameters. This analysis allows the opportunity for more extensive face validation of a model, where the behaviour of each parameter can be compared to the expected behaviour (Saltelli et al, 2000).

Global/sampling methods sample each parameter over their entire distribution to examine the sensitivity of the model. A simple approach to this could simply be random sampling based on the expected distributions of the parameters (Monte Carlo analysis) and using simple visualisations or regression/correlation analysis to examine the model. It is possible to use more complicated methods which use transformation functions to sample the parameter space. For example FAST which uses a Fourier transform function (Cukier et al, 1978) or Sobol’ which uses a dimensionality decomposition (Sobol’, 1993). More complicated search algorithms that optimize the parameter sample (e.g. genetic algorithms) can also be used. For large complex models this analysis can still be computationally expensive and one option is to utilise transformation functions to represent the model with less complexity. An example of this is high dimensional model representation (e.g. Ziehn and Tomlin, 2008) that represents the mapping between input and output with polynomials of different orders.

Algebraic sensitivity analysis takes a different approach by directly examining the equations of the model. For each operation the sensitivity is calculated and then combined algebraically for the model operations. It is completed by considering finite proportional changes to the input and deriving how it changes the output function, applying simplifications where appropriate. A thorough introduction to the method including derivation of sensitivities for basic operations is provided in Norton (2008). This method has many potential benefits including extra insight into the observed sensitivity behaviour of the model. For larger models calculation and derivation may become more difficult. But complexity could be reduced by simplifications of the analysis method or by performing the analysis on individual components independently.

5 COMBINING QUALITATIVE WITH QUANTITATIVE TESTING

A thorough testing procedure will include both qualitative and quantitative evaluation. When this occurs it is necessary to consider systematically both the qualitative and quantitative components. In some modelling communities this has lead to the development of systematic protocols that allow for the consideration of both factors, The Good
Modelling Practice Handbook (STOWA/RIZA, 1999) for deterministic, numerical models and guidelines for groundwater modelling by the Murray-Darling Basin Commission (2000) are two examples of checklists developed to evaluate models systematically.

Another approach is to assign numerical values to each question allowing the models to be rated either numerically or graphically. One system that uses this approach is the Numerical Unit Spread Assessment Pedigree (NUSAP) system. This system combines derived numerical metrics (including some form of error calculation and spread calculation) with more qualitative approaches used to assess the performance of the model and the process used to generate the model. The results from these multi-criteria tests are combined onto a single kite-diagram allowing easy comparison of various models’ performance.

6 APPLICATION RECOMMENDATIONS

As part of their 10 step modelling procedure Jakeman et al (2006) define a set of minimum standards which models should include (but not be limited to). These standards are repeated below:

1. Clear statement of the objectives and clients of the modelling exercise;
2. Documentation of the nature (identity, provenance, quantity and quality) of the data used to drive, identify and test the model;
3. A strong rationale for the choice of model families and features (encompassing alternatives);
4. Justification of the methods and criteria employed in calibration;
5. As thorough analysis and testing of model performance as resources allow and the application demands;
6. A resultant statement of model utility, assumptions, accuracy, limitations, and the need and potential for improvement; and quite obviously but importantly;
7. Fully adequate reporting of all of the above, sufficient to allow informed criticism.

These standards are all applicable to the performance evaluation process including the selection and application of the performance criteria. Each modelling task completed will likely have unique goals and challenges, which means there is no ideal standard technique applicable for all models. However, despite their differences it is possible to suggest a general procedure that would be beneficial to many models. The procedure suggested is summarised in the following four steps.

Step 1: Identification of the model’s purpose

The most important step of the procedure is the initial step, it is necessary to have a clear idea of the modelling purpose. This means having a clear idea of what events are being modelled and what will constitute a ‘good’ model. Having a clear idea of the model’s purpose allows easy selection of error metrics.

Step 2: Identification of data characteristics

The second step involves an analysis of data which is used to test the model. This involves determining whether there is any data available and how much of this data is required for the development and calibration of the model (generally more calibration data is required for models of greater complexity) At this point it should also be determined whether there is enough data and computing resources/time to consider multiple calibration and testing periods.

The data can then be analysed. For the initial analysis a graphical procedure is suggested to detect the general behaviour of the data to be modelled. For time series data, an auto-correlation procedure will detect any periodicity in the data, while calculating the empirical distribution function will give a better impression of the magnitude of events. It may be necessary to examine a time domain plot of events to detect during what period events and outliers occur. At the completion of these tests there will be a clearer understanding of the data in the system and period/s for calibration and testing can be selected with confidence.
**Step 3: Graphical performance analysis**

The third step entails a graphic analysis to judge the performance of the model. From this step there are two main goals: the detection of likely under- or non-modelled behaviour and gaining an overview of the overall performance of the model. The residual plot, QQ plot and a cross-correlation between the input data and residuals are all capable of indicating when a model is not completely representing a system’s behaviour. These results can be used to judge the model’s performance or to help refine the model before the rest of the evaluation is completed.

**Step 4: Select basic performance criteria**

It is necessary to select performance criteria to evaluate the model. Root mean square error (RMSE) or Nash-Sutcliffe efficiency ($R^2$) are ideal candidates for an initial metric as their widespread usage will benefit in communication of the performance of the model. A thorough understanding of the selected metric is, however, necessary. In particular any weaknesses of a metric for a particular purpose must be addressed. Even in the initial model valuation multiple metrics should be considered. Metrics should be paired which help to overcome the error of individual metrics. For example $R^2$, which can suffer from a significant offset error should be paired with bias. RMSE (or again $R^2$) can be paired with a selected data transformation to reduce the effect large events have on the evaluation.

**Step 5: Consideration of advanced methods**

Once an analysis has been completed using the basic performance criteria it is possible to consider how complete the current evaluation has been. The simple metrics are judged against the knowledge gained from the graphical analysis in step 3, and how well the current evaluation differentiates between multiple models and expert knowledge. Depending on the problems that are identified there are many possible advanced methods that can be considered (Table 2).

<table>
<thead>
<tr>
<th>Problem Identified</th>
<th>Potential Solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Changes in model divergence over time not captured by current metrics</td>
<td>Need a windowed metric, more advanced wavelet analysis, or a metric that is able to allow for the changes in uncertainty (assuming uncertainty is driving the divergence)</td>
</tr>
<tr>
<td>Objective functions not effectively differentiating between models</td>
<td>Modify objective function based on sensitivity analysis, or use multi-criteria methods, consider application of Pareto methods if methods are actually the same</td>
</tr>
<tr>
<td>Significant difference between calibration and testing model performance</td>
<td>Period of calibration may not be well chosen, perform sensitivity analysis to determine which parameters are causing trouble, DYNIA for periods parameters are active. Try different/multiple calibration periods.</td>
</tr>
<tr>
<td>Significant divergence in low/high magnitude events not captured by metrics</td>
<td>Use data transformations to highlight the differences, metrics that allow for the divergence (e.g., HMLE). Consider multi-resolution methods</td>
</tr>
</tbody>
</table>

**REFERENCES**


Murray-Darling Basin Commission, Groundwater Flow Modelling Guideline. Murray-
Norton, J., August Algebraic sensitivity analysis of environmental models, Environmental
Modelling & Software 23 (8), 963-972, 2008.
Oreskes, N., K. Shrader-Frechette, and K. Belitz, Verification, validation, and confirmation
Parker, P., R. Letcher, A. Jakeman, M. Beck, G. Harris, R. Argent, M. Hare, C. Pahl-Wostl,
A. Voinov, and M. Janssen, Progress in integrated assessment and modelling,
Pontius, R., Useful techniques of validation for spatially explicit land-change models,
Risbey, J., M. Kandlikar, and A. Patwardhan, November Assessing integrated assessments,
Saltelli, A., K. Chan, and E.M. Scott, Sensitivity Analysis, Wiley Series in Probability and
1978.
Sorooshian, S., and D.A. Dracup, Stochastic Parameter Estimation Procedures for
Hydrologic Rainfall-Runoff Models: Correlated and Heteroscedastic Error Cases,
Water Resources Research, 16 (2), 430-442, 1980.
STOWA/RIZA, 1999. Smooth Modelling in Water Management, Good Modelling Practice
Handbook. STOWA Report 99-05. Dutch Department of Public Works, Institute for
99.036.
Taylor, C.J., D.J. Pedregal, P.C. Young, and W. Tych, Environmental time series analysis
and forecasting with the captain toolbox, Environmental Modelling and Software, 22
(6), 797-814, 2007.
van der Sluijs, J.P., M. Craye, S. Funtowicz, P. Kloprogge, J. Ravetz, and J. Risbey,
Combining Quantitative and Qualitative Measures of Uncertainty in Model-Based
Wagener, T., D. Boyle, M. Lees, H. Wheater, H. Gupta, and S. Sorooshian, A framework
for development and application of hydrological models. Hydrology and Earth System
Wagener, T., N. McIntyre, M.J. Lees, H.S. Wheater, and H.V. Gupta, Towards reduced
uncertainty in conceptual rainfall-runoff modelling: dynamic identifiability analysis,
Wealands, S., R. Grayson, and J. Walker, Quantitative comparison of spatial fields for
hydrological model assessment - some promising approaches, Advances in Water
Yapo, P., H. Gupta, and S. Sorooshian, Multi-objective global optimization for hydrologic
Young, P.C. Data-based mechanistic modelling and validation of rainfall-flow processes.
In M. G. Anderson and P.D. Bates (Eds.), Model Validation: Perspectives in
Young, P.C., A.J. Jakeman, and R.E. McMurtrie, An instrumental variable method for
Ziehn, T., and A.S. Tomlin, Gui-hdmr – a software tool for global sensitivity analysis of
Set-membership approach for identification of the uncertainty in power-law relationships: the case of sediment yield

Karel J. Keesman\(^a\), Lachlan Newham\(^b\) and Tony Jakeman\(^b\)

\(^a\)Systems and Control Group, Wageningen University, Bornse Weilanden 9, 6708 WG Wageningen, The Netherlands (karel.keesman@wur.nl)

\(^b\)CAM, The Australian National University, Canberra ACT 0200, Australia

Abstract: Power laws are used to describe a large variety of natural and man-made phenomena. Consequently, they are used in a wide range of scientific research and management applications. In this paper, we focus on the identification of uncertainty bounds on a power law relationship from experimental data, using a bounded-error characterization. These bounds can subsequently be used as constraints in e.g. optimization and scenario studies. The basic so-called set-membership approach involves outlier identification and removal, feasible parameter set estimation, evaluation of the feasible model output set and tuning of the error bounds. As an example we examine scattered sediment yield versus catchment (or watershed) area data of Wasson, (1994). The key result of this is an appropriate unfalsified relationship between sediment yield and catchment area with uncertainty bounds.

Keywords: identification, set-membership, power laws, sediment

1 INTRODUCTION

In a wide range of science and in many applications power laws are used to describe natural and man-made phenomena in a quantitative way (eg Deng and Jung, 2009; Millington et al., 2009). In particular, power law distributions are widely applied in earth sciences, linguistics, biology, economics and social sciences, when relating sizes to the frequency of occurrence. This distribution describes the phenomenon that large is rare and small is common. In power law distributions, the exponent in the power law function is negative. But it is certainly not limited to this. For instance, an exponent of 0.5 gives a square root function describing the free outflow from a tank, as is commonly derived from Bernoulli’s law. Moreover, an exponent greater than 1 gives rise to a kind of exponential growth, as is frequently seen in biology. In fact, many well-known laws in physics are expressed in terms of a power law function, for instance, Stefan-Boltzmann law, Inverse-square laws of Newtonian gravity and Electrostatics, van der Waals force model, Kepler’s third law, Square-cube law (ratio of surface area to volume), but also Pareto’s principle follows a power law function. In the pre-computer era, scientists plotted all kind of phenomena on a log-log scale to arrive at a linear relationship between the variables, which they most often found. Hence, the power law is a fundamental way of describing a wide range of relationships.

Apart from some fundamental relationships in physics, most frequently power laws are derived from experimental data. Experimental data always contain measurement and sampling errors, which are usually characterized in statistical terms. Consequently, the estimates of the power law parameter are of a stochastic nature. However, for instance, in
case of limited data or after some non-linear transformation of the data, the presumed stochastic characterization is not always valid. Hence, as an alternative to a stochastic characterization a so-called bounded-error characterization, also called set-membership approach, has been proposed in the last decades.

The objective of this paper is to present a set-membership approach to the identification of error bounds on a power law and to provide unfalsified bounds on the data of Wasson (1994) using a power law relationship and a bounded-error characterization. The Wasson (1994) data provides a convenient example but the technique could be applied to any number of similar data sets.

2 BACKGROUND

Power-law functions are polynomials in a single variable, that is

\[ f(x) = ax^k + o(x^k) \]  

where \( a, k \) are real constants and \( o(x^k) \) is an asymptotically small function. The parameter \( k \) is called the scaling exponent, since a typical property of a power law is the scaling invariance. To show the scaling invariance of power laws, let \( x \) be multiplied with a constant \( c \) then \( f(cx) = a(cx)^k = c^k f(x) \propto f(x) \). In other words, multiplication with a constant does not change the shape of the function. Hence, power laws are explicitly used to describe the scaling behavior of natural processes. Allometric scaling laws, for instance, are frequently used to describe the relation between biological variables and thus are some of the best known power-law functions in nature.

When dealing with an experimental data set \((y, x)\), the term \( o(x^k) \) is replaced by a deviation or error term \( e \), so that

\[ y = ax^k + e \]  

where \( y \) is the measured (dependent) variable in (2). In what follows, and from the viewpoint of parameter estimation, Eqn. (2) is also called a nonlinear regression, with unknown parameters \( a \) and \( k \).

Notice from (2) that when \( k \) is given, \( a \) can be simply estimated from the resulting linear regression using ordinary least-squares estimation. On the contrary, the estimation of the exponent \( k \) is not so easy. There are many ways of estimating the scaling exponent in a power law from data. However, not all of them yield unbiased and consistent estimates. A commonly applied technique is to apply a (natural) logarithm transformation to the deterministic part of (2), which results in the linear regression

\[ \ln y = \ln a + k \ln x \]  

It is well-known that logarithmic transformation leads to distortion of the error \( e \) (see e.g. Barlett, 1947; Box and Cox, 1962). If the original error \( e \) is normally distributed, after logarithmical transformation of the data it becomes log-normally distributed. However, in many cases with limited data this assumption about normality of the data is questionable and cannot be thoroughly tested. Considering a stochastic nature of \( e \), and more particular assuming a Gaussian distribution, the most reliable estimation techniques are often based on maximum likelihood methods.

In this paper, given a limited data set - as is quite common in practice - we will follow an alternative route that is based on so-called set-membership estimation (Walter, 1990; Norton, 1994, 1995; Milanese et al., 1996). Let us shortly summarize this approach.

Consider hereto the following non-linear regression type of model in vector form,

\[ y = \mathbf{F}(\mathbf{\theta}) + \mathbf{e} \]  

where \( y \in \mathbb{R}^N \) contains the observed output data, \( \mathbf{F}(\mathbf{\theta}) \) is a non-linear vector function mapping the unknown parameter vector \( \mathbf{\theta} \in \mathbb{R}^m \) into a noise-free model output \( \hat{y} \). The
error or information uncertainty vector $\epsilon$ is assumed to be bounded in a given norm. In what follows, we assume that

$$\|\epsilon\|_\infty \leq \epsilon$$

(5)

where $\epsilon$ is a fixed positive number. Hence, a measurement uncertainty set (MUS), containing all possible output measurement vectors consistent with the observed output data and uncertainty characterization, is defined as

$$\Omega_y := \{ \tilde{y} \in \mathbb{R}^N : \| y - \tilde{y} \|_\infty \leq \epsilon \}$$

(6)

This set is a hypercube in $\mathbb{R}^N$. Let the set

$$\Omega_\theta := \{ \theta \in \mathbb{R}^m : \| y - F(\theta) \|_\infty \leq \epsilon \}$$

(7)

define the feasible parameter set. Then, the set-membership estimation problem is to characterize this feasible parameter set (FPS), which is consistent with the model (4), the data ($y$) and uncertainty characterization (5)-(6).

Hence, instead of trying to find the optimal parameter vector as in an ordinary least-squares approach, our goal now is to find the set with feasible parameter vectors that are consistent with the model and the data with related error bounds. Hence, we will not consider the measurements as such but define intervals for each measurement. This approach avoids the distortion of the original probability density function after some non-linear transformation, because only bounds are considered. Furthermore, a symmetric bound in the log-log space introduces automatically skewed error bounding in the original space, which seems to be natural when considering data which most likely can be described by a power law.

From the set-membership literature, it is well-known that for the linear regression case, the FPS is a polytope found from the intersection of $N$ (number of data points) strips in the parameter space. This will also be demonstrated in our example case, when working in the log-log space. It can even be shown that the well-known weighted least-squares techniques can be used to solve the bounded linear regression problem (Milanese, 1995; Keesman, 1997). But, in general, the FPS can be a complex, and even unconnected, set (see Keesman, 2003, for details and possible solutions).

3 APPLICATION

In this study we examine a data set collated by Wasson (1994) of sediment yields versus catchment area. The Wasson (1994) data, was collated from numerous studies of long-term sediment yields in south east Australia. The data shows a high level of scatter – a function of (i) high inherent spatial and temporal variability of sediment yield across south east Australia; and (ii) the use of a range of different underlying methods to estimate sediment yields.

Our particular interest in the analysis of this data was to identify likely upper and lower bounds of plausibility for sediment yield estimates. Such information is invaluable in informing the development and testing of dynamic, semi-distributed (spatially) models of sediment generation e.g. Newham et al. (2004).

The following figures show the Wasson (1994) data in original and log-log space.
Presume that the catchment area-sediment yield data of Figure 1a can be described by the power law,

$$Y = aA^k$$  \hspace{1cm} (8)

where $a$ and $k$ are unknown parameters. Given these data (Figure 1a,b) and the model (8), our objective is to find an appropriate uncertainty description of the sediment yield $Y$ for a given catchment area $A$, preferably in a log-log space.

Let us start by applying a natural logarithmic transformation of the power law (8). Consequently,

$$\ln Y = \ln a + k \ln A$$  \hspace{1cm} (9)

from which we define $\alpha_1 := \ln a$ and $\alpha_2 := k$. The ordinary least-squares (OLS) estimates (indicated by a hat) and corresponding covariance matrix of the estimation error ($\Sigma$) are given by
\[ \hat{\alpha}_1 = 3.3125, \quad \hat{\alpha}_2 = 0.9177; \quad \Sigma = \begin{pmatrix} 0.0353 & 0.0012 \\ 0.0012 & 0.0020 \end{pmatrix} \] (10)

However, as mentioned before, it is well-known that a logarithmic transformation leads to distortion of the error. Hence, the assumption about normality of the log transformed error is questionable and, because of the limited size of the data set (see Figure 1), cannot even be thoroughly tested. Consequently, the covariance matrix in (3) cannot be directly interpreted, which limits the possibilities for a direct uncertainty analysis. Moreover, due to the error distortion the estimates become biased. Hence, bias correction must be applied to correctly estimate the unknown parameters.

The set-membership approach, as presented in Section 2, avoids these obstacles, since we focus only on the calculation of the bounds and not at all on the probability distributions. However, given the linear regression (9) and the data in Figure 1b, the key question here is how to choose the error bound \( \epsilon \) (see (5)). Notice that outliers, with inappropriate bounds, can easily lead to an empty feasible parameter set (FPS). Hence, the first step is to remove possible outliers. Keesman and van Straten (1989a) suggested a re-iterative min-max estimation, where the maximum error is plotted against the iteration number. After each min-max estimation step in a specific iteration, the data point at which the maximum error occurs is removed and the procedure is repeated. The result of such a procedure for the given data set is presented in Figure 2.

**Figure 2. Results of reiterative min-max estimation.**

Figures 1 and 2 suggest that the data set basically contains two possible outliers (as indicated in Fig. 1b in red) and that a sufficient error bound would be 3 t/a. The min-max estimate from the third iteration, thus after removal of two possible outliers, is given by: \( \hat{\alpha}_1 = 3.5237 \) and \( \hat{\alpha}_2 = 0.9072 \), with maximum error of 2.85 t/a. Hence, choosing the error bounds on the basis of this maximum error, would degenerate the FPS to a singleton. The estimation results related to a constant error bound of 3 t/a and using exact (see e.g. Walter and Piet-Lahanier, 1989; Mo and Norton, 1990) and approximate (Monte Carlo based) bounding techniques (Keesman and van Straten, 1989b; 1990) indicate that, for this error bound, the feasible model output set (FMOS) does not fully reflect the uncertainty in the measurements (not shown here). This would be acceptable when the data points not covered by the FMOS could be considered as outliers. However, in this case there is no evidence to do so. Hence, in the next step, we will increase the error bounds such that the FMOS contains (most of) the measurements.

The variation in the data set, in particular for \( \ln(A) = -4.6 \) and on the intervals \([-2.12, -1.96], [-0.67, -0.56]\) (see Figure 1b), can be estimated from the standard deviations in \( \ln(Y) \). For each of these regions the standard deviations have been estimated as 1.40, 1.49 and 2.33, respectively. Consequently, an error bound of 5 t/a has been chosen to reflect the
3σ-bound for the individual measurements in this data set. The set-membership estimation results are presented in Figures 3 and 4. Notice from Figure 3 that, given the uncertainty in the data, the scaling exponent can possibly be smaller and larger than 1. As expected, the FPS indicated by blue dots and constrained by lower (green) and upper (red) bounds contains the min-max estimate ($\alpha_1 = 3.5237$ and $\alpha_2 = 0.9072$). Increasing the error bound will thus lead to a larger FPS. Hence, it reflects the larger uncertainty considered in the data. Notice that the FMOS in Figure 4 does contain almost all of the measurements and thus we may consider these results as appropriate for further evaluation. For instance, the (interpolated) bounds could be used in the identification of an erosion model, using bounded information, i.e. basically taking into account constraints instead of point measurements.

**Figure 3.** Feasible parameter set related to error bound of 5 t/a.

**Figure 4.** Feasible model output bounds (upper bound: green $+$; lower bound: blue $+$) related to error bound of 5 t/a.
4  DISCUSSION
Considerable challenges exist in the identification of feasible parameter sets for modelling environmental data. This is particularly the case for power law relationships such as those applied to water quality data.

It is interesting to see that the power law, \( Y = a A^k \), is a solution to the differential equation,
\[
\frac{dY}{dA} = k \frac{Y}{A}, \quad Y(0) = 0
\]  (11)

In other words, since \( \frac{A}{Y} \frac{dY}{dA} = \frac{dY}{Y} = \frac{dA}{A} = k \), the relative or normalized slope of the relation between catchment area and sediment yield is constant and equal to the scaling exponent.

Notice that power law functions, as in (1), describe the static, non-linear relationship between variables. In this paper it has been shown that, given bounded-error data, a Monte Carlo-based bounding technique, also known as the Monte Carlo Set-membership Method (MCSM), can solve the parameter estimation problem. However, approximate (Monte Carlo based) bounding techniques are also applicable to the identification of dynamic, non-linear simulation models (see e.g. Keesman and van Straten, 1990). As for all other non-linear deterministic or stochastic estimation methods, the Monte Carlo-based bounding technique is practically constrained to cases with a limited number of unknown parameters. This curse of dimensionality is, in fact, an issue in many estimation and optimization problems. Hence, reduction of the problem via e.g. time scale decomposition or parameter space decomposition is crucial (Keesman, 2002).

5  CONCLUSIONS
A bounded-error characterization leads either to an empty set, a singleton or a (non-convex, not even connected) set of parameter vectors, using deterministic algorithms. As such, it directly reflects the uncertainty in the model and in the data without statistical computations. In particular for data sets of limited size, for which statistical properties are difficult to verify, the set-membership approach provides a good alternative. Given the Wasson (1994) catchment area-sediment yield data and a power law relationship with unknown coefficients, we were able to derive unfalsified model-based bounds on the data for use in constrained optimization and scenario studies.

REFERENCES
Keesman, K.J., State and parameter estimation in biotechnical batch reactors, Control Engineering Practice, 10(2), 219-225, 2002.


Uncertainty propagation throughout an integrated water-quality model

G. Freni and G. Mannina, G. Viviani

a Dipartimento di Ingegneria Idraulica ed Applicazioni Ambientali, Università di Palermo, Viale delle Scienze, 90146 Palermo, ITALY (mannina@idra.unipa.it)

Abstract: In integrated urban drainage water quality models, due to the fact that integrated approaches are basically a cascade of sub-models (simulating sewer system, wastewater treatment plant and receiving water body), uncertainty produced in one sub-model propagates to the following ones depending on the model structure, the estimation of parameters and the availability and uncertainty of measurements in the different parts of the system. Uncertainty basically propagates throughout a chain of models in which simulation output from upstream models is transferred to the downstream ones as input. The overall uncertainty can differ from the simple sum of uncertainties generated in each sub-model, depending on well-known uncertainty accumulation problems. The present paper aims to study the uncertainty propagation throughout an integrated urban water-quality model. At this scope, a parsimonious bespoke integrated model has been used allowing for analysing the combinative effect between different sub-models. Particularly, the different parts of the quantifiable uncertainty have been assessed and compared by means of the variance decomposition concept. The integrated model and the methodology for the uncertainty decomposition have been applied to a complex integrated catchment: the Nocella basin (Italy). The results show that uncertainty contribution, due to the model structure, is higher with respect to the other sources of uncertainty.

Keywords: Environmental modelling, Integrated urban drainage systems, Uncertainty analysis, Receiving water body, Wastewater treatment plant.

1. INTRODUCTION

Since the Water Framework Directive (WFD) enactment, several researchers have been working to develop integrated approaches for water basin management (among others, Rauch et al., 2002; Mannina et al., 2006; Willems, 2008). Indeed, WFD implicitly requires a holistic approach of the whole system, namely: sewer system (SS), wastewater treatment plant (WWTP) and receiving water body (RWB), in order to properly design as well as manage water resources and protect the environment from pollution. A modelling framework of the overall integrated system may allows all domains of the catchment to be modelled and is essential for describing water states in both the temporal and spatial dimensions. An integrated approach generally requires the management of a complex model where several sub-models are linked together and processes and interactions within and between them take place. However, high complexity in general should be avoided and it is not desirable (Chapra, 2003; Rechow, 2003; Beven, 2006; Willems, 2008). Complex models usually require large databases for calibration and they can deliver relevant amounts of uncertainty to the modelling outputs; the study of the interconnection between model complexity and data availability is of paramount interest. In order to reduce model complexity, the system is generally separated into small systems. As a matter of fact, Schmitt and Huber (2006) suggest to divide the whole system into smaller subsystems and to define appropriate system boundaries. With this respect uncertainty assessment may constitute an optimal solution. As a matter of fact, uncertainty quantification of a mathematical model enables one to gain insights about the significance level of the results.
provided by the model. Further, quantitative uncertainty analysis can provide an illuminating role for problems where data are limited and where simplifying assumptions have been used in order to help identify the robustness of the conclusions and to help target data-gathering efforts (Frey, 1992). In the last years, a large debate rose in literature about the most appropriate methodologies for analysing model uncertainty and about the compromise between initial hypotheses of methods and, on the other hand, reliability of uncertainty analysis results. An example is given by the debate about the use of “formal” Bayesian methods and “non-formal” ones like Generalised Likelihood Uncertainty Estimation: the first providing more rigorous and reliable framework but requiring the definition of a formal error model; the latter generally computing larger uncertainties but relaxing many of the Bayesian hypotheses (Beven and Byrnely, 1992; Stedinger et al., 2008; Vrougt et al. 2009).

In urban drainage, especially for water quality modelling, the state-of-the-art is far less advanced compared to other fields. Deletic et al. (2009) pointed out that the state of knowledge regarding uncertainties in urban drainage models is poor, in part due to a lack of clarity and/or consensus on the way in which the results of model uncertainty analyses are obtained, presented and used. Among the possible reasons for the current lack of quantification of model uncertainty in the urban drainage modelling field, the high computational effort (e.g., for a Monte Carlo simulation) required by mathematical models has been cited as the main impediment (Muschalla et al., 2009). Indeed, urban integrated models are basically a cascade of sub-models (simulating SS, WWTP and RWB), and the computational time required for the entire model may constitute a limitation, especially in cases where several Monte Carlo simulations are run (Mannina and Viviani, 2010).

In this context the paper presents the uncertainty assessment and the uncertainty propagation throughout a bespoke integrated model developed in previous studies (Mannina, 2005). Particularly, the variance decomposition concept has been applied for the uncertainty quantification as well as propagation. The methodology and the model has been applied to a real case study, the Nocella catchment (IT), for which an extensive field gathering campaign was carried out and quantity/quality data were therefore available.

2. MATERIALS AND METHODS

2.1 Integrated urban drainage model

For the simulation of the whole system, a previously developed, bespoke model was adopted (Mannina, 2005). The model is able to estimate both the interactions between the three components of the system (SS, WWTP and RWB) and the modifications, in terms of quality, that urban stormwater causes inside the RWB. The general structure of the integrated model consists of three sub-models; each sub-model is divided into quantity and quality modules for the simulations of the hydrographs and of the pollutographs, respectively. The modelling structure can be adapted to the specific application by removing or duplicating sub-models or parts of them, such as the stormwater tank or the Combined Sewer Overflow (CSO). The quantity module of the SS sub-model is described by a cascade of a linear reservoir and a channel, representing the catchment, and a linear reservoir, representing the sewer network. Initially, the net rainfall is computed by subtracting both continuous and initial losses; the latter are modelled assuming constant initial depression storage and a constant runoff coefficient. For the quality module of the SS sub-model, several processes were considered, both on the catchment and in the sewer, as well as during dry and wet weather. In particular, the build-up and wash-off processes for pollutants were considered according to the classical approaches proposed by Alley and Smith (1981) and Jewell and Adrian (1978). Solids deposition in the sewer during dry weather was evaluated by adopting an exponential law depending basically on the duration of the antecedent dry weather and on sewer network characteristics (Bertrand-Krajewski, 1992). Regarding the erosion of sewer sediments (Mannina and Viviani, 2010 Crabtree, 1989), their cohesive behaviour was considered by assuming the bed sediment structures hypothesised by Skipworth et al. (1999). The pollutographs at the outlet of the sewer system were evaluated by hypothesising the complex catchment sewer network as a reservoir and by considering the transport capacity of the flow. Finally, the WWTP inflow
was computed by taking into account the presence of a CSO device, representing its efficiency by the introduction of two dilution coefficients.

The WWTP sub-model simulates the most sensitive units that can be affected by an increase of pollutant load inflow; more specifically, the activated sludge tank and the settler. In particular, the flow substrate and microbial density in the activated sludge tank were calculated with mass balances based on Monod's theory. Conversely, the sedimentation tank performance was simulated using the solid-flux theory according to the methodology proposed by Takács et al. (1991). In particular, the solids concentration profile was obtained by dividing the settler into horizontal layers of constant thickness. Within each layer, the concentration was assumed to be constant, and the dynamic update was performed by imposing a mass balance for each layer. The settling velocity function proposed by Takács et al. (1991) was employed. Regarding the RWB sub-model, the exemplified form of the Saint-Venant equation (kinematic wave) for the quantity module and the dispersion advection equation for the quality module were adopted (Brown and Barnwell, 1987).

2.2 The Case study

The analysis was applied to a complex integrated catchment: the Nocella catchment, which is a semi-urbanised catchment located nearby Palermo in the northwestern part of Sicily (Italy). The entire natural basin is characterised by a surface area of 99.7 km² and has two main branches that flow primarily east to west.

The two main branches join together at 3 km upstream from the river estuary. The southern branch is characterised by a smaller elongated basin and receives water from a large urban area characterised by relevant industrial activities partially served by a WWTP and partially connected directly to the RWB. The northern branch was monitored in the present study. The basin closure is located 9 km upstream of the river mouth; the catchment area is 66.6 km². The cross-section closing the catchment is equipped with a hydro-meteorological station (Nocella a Zucco).

The river reach receives wastewater and stormwater from two urban areas (Montelepre, with a catchment surface area equal to 70 ha, and Giardinello, with a surface area of 45 ha) drained by combined sewers. Both urban areas are characterised by concrete sewer pipes with steep slopes. The Montelepre sewer is characterised by circular and oval pipes with maximum dimensions of 100 cm x 150 cm. The sewer system serves 7,000 inhabitants, and it is characterised by an average dry weather flow of 12.5 L/s (the water supply is 195 L/capita/d), and an average dry weather BOD concentration of 223 mg/L. The Giardinello sewer is characterised by circular pipes with a maximum diameter of 800 mm. The served population is 2,000 inhabitants, and it has an average dry weather flow of 2.5 L/s (here, the water supply is 135 L/capita/d) and an average dry weather BOD concentration of 420 mg/L. The calculated BOD unit loading factors for the two urban catchments are 35 and 45 g/capita/d for Montelepre and Giardinello, respectively. These values are lower than those typically observed in Italy (60 g/capita/d), likely due to the industrial activities present in the urban catchments; the lower concentration of BOD in Montelepre’s urban catchment is also due to the presence of an infiltration flow into the sewer system.

Each sewer system is connected to a WWTP protected by CSO devices. The WWTPs are characterised by simplified, activated sludge processes with preliminary mechanical treatment units, an activated sludge tank and a final circular settler. According to the modelling scheme, particular attention in data acquisition was given to the activated sludge tank and the sedimentation tank. Moreover, such units are the most sensitive to flow and concentration variations during wet weather periods; wet weather loads greatly affect activated sludge settling tanks and can significantly affect the effluent quality.

Rainfall was monitored by four rain gauges distributed over the basin: the Montelepre rain gauge is operated by Palermo University and is characterised by a 0.1-mm tipping bucket and a temporal resolution of 1 minute; the other three rain gauges are operated by the Regional Hydrological Service and they are characterised by a 0.2-mm tipping bucket and a temporal resolution of 15 minutes. The hydro-meteorological station (Nocella a Zucco) located at the catchment end is characterised by an ultrasonic level gage operated by the
Regional Hydrological Service and has a temporal resolution of 15 minutes. Rainfall data for yearly maximum intensity events are available for all the rain gauges from 1955 to the present without a gap. The instruments were integrated by Palermo University by installing an area–velocity submerged probe that provides water level and velocity data with a 1-minute temporal resolution. An ultrasonic external probe was used to obtain a second water level measurement for validation and as a backup in case the submerged probe failed; an automatic 24-bottle water quality sampler was used for water quality data collection. The water quality parameters monitored were Total Suspended Solids (TSS), Biological Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Ammonia-Nitrogen (NH₄-N), Total Kjeldahl Nitrogen (TKN) and Phosphorus (P); the Dissolved Oxygen (DO) was only monitored for the river. All analyses were carried out according to Standard Methods (APHA, 1995). The monitoring campaign was used for model calibration under the present conditions; details of the calibration process can be found in Freni et al. (2010). The model parameters were calibrated for each of the sub-models by means of Monte Carlo Analysis, randomly varying parameters in user-defined ranges and minimising the variance of the model output errors based on the available water quantity and water quality measurements (Freni et al., 2009).

2.3 The variance decomposition concept and quantification of different uncertainty sources

The uncertainties for each sub-model can be decomposed into model input and model-related uncertainties. Model input uncertainties are due to errors in the data used as boundary and initial conditions in the model. Model uncertainties are due to the structure of the model, which includes the equations and algorithms used for the simulations and the coupling of the models and the parameters used to control the equations. As pointed out by Willems (2008), model structure uncertainties can be seen as the remaining uncertainties in the model output after use of error-free input in the model and after the most optimal calibration of the model parameters to the available measurements (e.g., for a given model structure, by optimising the selected goodness-of-fit statistics). Usually, when the comparison of different model structures is not within the scope of the study, model structure and model parameter uncertainties are jointly analysed. In such cases, parameters are assumed to be the only source of uncertainty and structural uncertainty is implicitly distributed among the parameters (Freni et al., 2010-2009). Such a hypothesis was maintained in the present study, which was based on the analysis of uncertainties related to input and calibration data and model parameters.

The variance of the total uncertainty in the model output variables was calculated as the variance of the errors in the model results after comparison with observational data; the variance of the observational errors was subtracted from this variance. The variance of the other model structure-related uncertainties was then quantified as the variance due to the variation of only the model parameters (a lumped approach) and also as a rest term in the description of the total variance, making use of the concept of variance decomposition (a distributed approach):

\[
\sigma_{Y,\text{tot}}^2 = \sum_{i=1}^{n} \sigma_{Y,\text{inp}(X_i)}^2 + \sigma_{Y,\text{str}}^2
\]

where \(\sigma_{Y,\text{tot}}^2\) is the variance of the total uncertainty in the model output variable \(Y\) after subtracting the variance of the observational errors, \(\sigma_{Y,\text{inp}(X_i)}^2\) is the source variance of the uncertainty contribution by model input variable \(X_i\) (\(i=1, \ldots, n\); \(n\) is the total number of input variables) and \(\sigma_{Y,\text{str}}^2\) is the source variance of the contribution of the model structure-related uncertainty.

Once quantified the error for each submodels, these are thereafter propagated throughout the different submodels that constitute the integrated model. To accomplish such a goal, different techniques may be employed. Among such techniques, the Monte Carlo method may be employed.

Similarly to Willems (2008), a Box–Cox (BC) transformation (Box and Cox, 1964) was applied to the all sub-model outputs \(Y\) for which a total variance was calculated:

\[
BC(Y) = (Y^{1+\lambda} - 1)/\lambda
\]
where the parameter $\lambda$ ($0 < \lambda \leq 1$) is calibrated to reach homoscedasticity in the errors (variance of errors nearly independent on the model output magnitude). The parameter was calibrated in this study by trial-and-error after a visual inspection of the homoscedasticity of the model errors. The BC transformation was applied to the model input and the output variables before calculating the variances $\sigma_{Y_{\text{eq}}}^2$ and $\sigma_{Y_{\text{str}}}^2$ of Eq. (1).

After this transformation, standard deviations or confidence interval widths were obtained that were nearly uniform for all time steps. The overall uncertainty was then represented by the average of the variances for all time steps (after BC transformation). Using this methodology and the above-mentioned measurements, total uncertainty was quantified for the output variables for the different sub-models.

Model input uncertainties were quantified based on detailed investigations of the input data. For rainfall, which is the main model input and the driving force of the temporal variability of the system processes, uncertainties in the calibration curves for the rain gauges were investigated at the Hydraulics Laboratory of Palermo, and the following error formulation was found:

$$H = H_{\text{real}} \cdot (1 + \text{err} \cdot H_{\text{real}}^a)$$

where $H$ is the rainfall depth taking into account the error, err is the error randomly generated considering a normal distribution with mean equal to zero and standard deviation equal to 0.035, $H_{\text{real}}$ is the real rainfall depth (unknown) and $a$ is a shape coefficient equal to 0.2332. The error curve parameters were obtained by calibration over several rain gauges similar to those installed in the analysed case study.

In order to conduct the random error simulation, the time series (the input series in this case) were separated into “independent” storm events, i.e., events leading to separate or independent sewer runoff events (events were separated by a dry weather flow period equal to or larger than the concentration time of the sewer network).

Errors in calibration data were analysed by means of a normal statistical distribution with a null mean. The uncertainty in this flow monitoring was assessed after testing the depth–velocity devices in the Hydraulics Laboratory of Palermo. The standard deviation of the flow per monitor ranged from 20% of the flow value for water levels lower than 5 cm to 6% for water levels between 5 cm and 50 cm. The standard deviation of the water quality measurement errors was assessed based on values found in the literature (Ahyerre et al., 1998; Bertrand-Krajewski et al., 2001; Kanso et al., 2003; Willems, 2008) to be as high as 30–40% for BOD and 15–20% for the other variables considered.

In the present study, structural uncertainties for each sub-model were related to parameters, assigning to each an uncertainty quota connected to model algorithms and equations. Parameter error distributions were assumed uniform, and thus the parameters had the same probability of taking any value within a specified range. The ranges were determined in a previous study by means of a model calibration based on several monitored events (Freni et al., 2009). These values were also employed for the variation of the model parameters in the simultaneous assessment of the structural uncertainty by means of the Monte Carlo runs. The propagation of the random input errors to the model output variables was performed by Monte Carlo simulation. Random simulations (1,000 runs) were carried out with the stochastic model input error, and the propagated errors on the model output variables were calculated for each time step. With this procedure, distributions of random errors were obtained, reflecting the uncertainty in the model output variables caused by the total model input uncertainty. These distributions and corresponding error variances or confidence intervals could be obtained for each time step or averaged over all time steps in the simulation period to obtain a “mean overall uncertainty” estimate.

A similar procedure was followed for the random simulation and propagation of the other uncertainty sources considered. Uncertainty sources were either analysed separately to obtain the partial contributions of each uncertainty source or jointly to obtain the total uncertainty considered for a specific modelling output.

Whenever the different types of model uncertainties are assessed separately it is possible to compare and quantify the contributions of the different sources of uncertainty to the total uncertainty in the model output (Willems, 2008). More specifically, it is of paramount
It is advisable to have a balance between data and structure uncertainty. Indeed, whenever the data uncertainty is higher it is recommended to provide more attention on data collection than to research in an attempt to improve the model results. By comparing the contributions of the different uncertainty sources, efficient measures thus can be determined to reduce the total uncertainty in the model results (Willems, 2008).

3. ANALYSIS OF RESULTS

The variance decomposition outlined above was applied to the bespoke integrated urban drainage model for RWB quality modelling of a case study in Italy. The analysis was performed as a step-by-step process starting from the most upstream water quantity sub-model and then propagating the uncertainties to the downstream ones. Initially, the homoscedasticity was verified and corrected by B-C transformation. For each modelling output for which measures were available, the total variance of the errors was computed by jointly accounting for all sources of error defined in the previous paragraph (input, calibration data and model structure, which was limited to model parameters). Partial contributions were singled out by analysing one uncertainty source at a time; their sum, according to the variance decomposition equation, gave the total variance of errors for the analysed modelling output. Any differences between the total variances computed by the lumped and the distributed approaches may have been due to the presence of a correlation between uncertainty sources. Such a correlation was surely absent if measurement and model errors were considered, but could have been present if uncertainties in different sub-models were considered. In the present application, the analysis has been focused on RWB discharge and quality state. The uncertainty in the estimation of these variables has been used as the object of the study.

In Figures 1-3 the results in terms of uncertainty bounds for the different sources of uncertainty are reported. Particularly, Figure 1 shows the uncertainty bounds in the river cross-section taking into account only the rainfall monitoring uncertainty. Such uncertainty was taken into account by applying the error model presented in the previous paragraph and calibrated in Laboratory to measured rainfall. Conversely, Figure 2 shows the uncertainty bounds considering a variation of the model parameters. By comparing the two uncertainty bounds it emerges that the model parameter uncertainty bounds are wider both for the quantity (Figure 2a) and for the quality (Figures 2b-2c). Therefore, for the present application, the uncertainty due to the rainfall could be negligible.

![Figure 1. Confidence limits due to rainfall uncertainty.](image1)

![Figure 2. Confidence limits due to model parameter uncertainty.](image2)
As a confirmation of the previous statement, Figure 3 reports the uncertainty bounds for the case of both model parameter uncertainty and rainfall uncertainty. Although the uncertainty bounds are wider with respect to the ones of the model parameter uncertainty, the differences are not relevant. The uncertainty bounds are in general wider for the quantity rather than for the quality aspects confirming the higher uncertainty that relies on quality aspects. Thus the results suggest to provide much more efforts both on the modelling as well as gathering data for the quality processes.

Figure 3. Confidence limits due to both model parameter uncertainty and rainfall uncertainty

4. CONCLUSIONS

The uncertainty assessment and its propagation throughout an integrated bespoke urban drainage model has been performed. Model structure, input and parameter uncertainty has been assessed and compared at the final cross section of the modelled system: the RWB outlet. Particularly, the analysis was carried out analyzing both quantity and quality phenomena: discharges and BOD and DO concentrations, respectively. The study enabled to draw some interesting considerations:

- Model results revealed that when analysing water quality variables, water quantity sub-models always provide smaller contributions to uncertainty than do water quality ones; this may be due to the higher complexity of water quality processes and their related models.
- The uncertainty contribution of water quantity modules to water quality ones is not negligible, and specific efforts should be provided by the modeller in order to adopt robust water quantity models, as their contribution to uncertainty affects both water quantity and water quality variables.
- The comparison of total variance computation by means of a lumped approach and by variance decomposition demonstrated the relevance of modelling error correlation when moving from upstream to downstream sub-models.

REFERENCES


Using Statistical Bias Description for Multi-objective Calibration of a Lake Water Quality Model

Anne Dietzel and Peter Reichert
Eawag, Swiss Federal Institute of Aquatic Science and Technology, 8600 Dübendorf, Switzerland, and ETH Zurich, Institute of Biogeochemistry and Pollutant Dynamics, 8092 Zurich, Switzerland
Anne.Dietzel@eawag.ch, Peter.Reichert@eawag.ch

Abstract: Models of environmental systems are simplified representations of reality. For this reason, their results are affected by systematic errors. This bias makes it difficult to get reliable uncertainty estimates of model parameters and predictions. We were faced with that difficulty when applying the lake model BELAMO to data from three Swiss lakes. This model combines the description of biogeochemical and ecological processes in lakes. Considering bias in the model output, lead to a description of system observations as the sum of the output of the deterministic model, bias and observation error. The identifiability problem between model output and bias was addressed by specifying informative priors for the standard deviations of the observation errors and choosing means of zero for the observation error and the bias. The resulting multi-objective calibration problem was solved by using the prior of the bias to specify how much model error we are willing to accept for which output variable. To avoid the very high computational demand of conventional Bayesian numerical techniques, the maximum of the posterior was calculated and a local Gaussian approximation was used to estimate parameter and model prediction uncertainty. Parameter estimations for 9 to 20 years until 1995 were conducted. The remaining 10 years of data were used for model validation and to compare with estimated prediction uncertainty. The results show the large influence of the bias for the model output. The results for the validation period indicate the large uncertainty in model prediction, but also the ability to estimate the role of model bias with the suggested technique.

Keywords: Lake water quality modelling; Multi-objective calibration; Bias; Uncertainty

1. INTRODUCTION

Models substantially contribute to formalizing and summarizing knowledge, analyzing observations and testing hypotheses about environmental systems. As all models are simplified representations of reality, results of environmental models are, in addition to input and parametric uncertainty, affected by systematic errors.

When trying to identify model parameters from data by using statistical inference, model bias leads to a violation of typical statistical assumptions and thus makes it difficult to get reliable uncertainty estimates. One option to derive uncertainty bounds of model predictions is to use a statistical description of bias in model output (Craig et al. [1996], Craig et al. [2001], Kennedy and O’Hagan [2001], Higdon et al. [2004], Bayarri et al. [2007]).
Another problem resulting from systematic deviations between model results and data is that frequently used model parameter estimation techniques often cause results that do not fulfil calibration objectives of the user. To address this problem, manual calibration criteria were included in multi-criteria optimization techniques (Yapo et al. [1996], Gupta et al. [1996], Madsen [2000], Madsen et al. [2002], Gupta et al. [2003], Boyle et al. [2003]). This procedure can lead to better model calibration, but does not provide probabilistic information required for estimating prediction uncertainty.

To assess the parameter, structural and prediction uncertainty under the problems mentioned above, a combination of statistical description of model bias and the ideas underlying multi-objective model calibration was used as described by Reichert and Schuwirth [2010]: The prior of a statistical description of bias was used to weigh between different calibration objectives. This multi-objective calibration technique was applied to the lake model BELAMO. Thereby, a calibration period of 20 years (9 years for Greifensee) was chosen (ending in 1995). Validation of the model for the following 10 years was done and uncertainty bounds of our knowledge about the true values of the model output variables were derived for the calibration as well as the validation period.

2. METHODS

We briefly summarize the technique suggested by Reichert and Schuwirth [2010], the used simplified approach to assess parameter and prediction uncertainty bounds and the numerical implementation of the proposed technique.

2.1 Review of inference in the presence of bias

When applying a model to a specific study area, one has to cope with the problem of the “right” parameter choice. As for most parameters direct measurements are missing, the user is faced with the difficulty of the choice of parameter values that are in a realistic order of magnitude and allow for model results that are in agreement with measurements for all state variables for the whole simulation period. This can be done by combining prior knowledge of parameter values with information gained from the observations of model output by Bayesian inference. However, in order not to violate the statistical assumptions of the model, it has to explicitly account for bias in model output. Following the literature on statistical bias description (Craig et al. [1996], Craig et al. [2001], Kennedy and O’Hagan [2001], Higdon et al. [2004], Bayarri et al. [2007]), the observations are described as the sum of the output of the deterministic model, \( y^L(x, \theta) \), bias, \( B^L(x, \xi) \), and observation error, \( E^L(\psi) \):

\[
Y^L_M(x, \theta, \psi, \xi) = y^L_M(x, \theta) + B^L_M(x, \xi) + E^L(\psi),
\]

where \( Y^L_M \) is the vector of random variables representing the observations at the observation layout \( L \). The observation layout, \( L \), defines which variables to observe or evaluate at which points in time and space. \( Y^L_M \) depends on external influence factors, \( x \), unknown model parameters, \( \theta \), and additional parameters, \( \psi \) and \( \xi \), of the error terms.

Equation (1) leads to a hierarchical model with the bias, \( B^L_M \), as an intermediate variable. Integrating out this intermediate variable, a likelihood function can be obtained as a function of the parameters, \( \Gamma = (\Theta, \Psi, \Xi) \), only:

\[
f_{Y^L_M|\Theta, \Psi, \Xi} (y^L_M|\theta, \psi, \xi, x) = \int f_{E^L|\psi}(y^L_M - y^L_M(x, \theta) - B^L_M(x, \xi)|\psi) \cdot f_{B^L_M|\xi}(b^L_M|\xi, x) \, db^L_M.
\]

In this equation, \( f_{E^L|\psi} \) is the probability density of the observation errors and \( f_{B^L_M|\xi} \) is the prior density of the bias given the parameters \( \xi \) and external inputs \( x \).
We assume a normally distributed observation error with mean zero and covariance matrix $\Sigma_{e}(\psi)$

$$f_{e^{T}|\psi}(e^{T}|\psi) = \frac{1}{\sqrt{2\pi}^{d} \sqrt{\det(\Sigma_{e}(\psi))}} \exp\left( -\frac{1}{2} (e^{T})^{T} \Sigma_{e}(\psi)^{-1} e^{T} \right)$$  \hspace{1cm} (3)$$

and a Gaussian stochastic process with a correlation structure that decays continuously with the distance of the corresponding independent input variables, $\mathbf{x}$, to describe our prior knowledge of the bias

$$f_{b_{i}^{T}|\xi,\mathbf{x}}(b^{T}|\xi,\mathbf{x}) = \frac{1}{\sqrt{2\pi}^{d} \sqrt{\det(\Sigma_{b_{i}^{T}}(\xi,\mathbf{x}))}} \exp\left( -\frac{1}{2} (b^{T})^{T} \Sigma_{b_{i}^{T}}(\xi,\mathbf{x})^{-1} b^{T} \right).$$  \hspace{1cm} (4)$$

A simple form of the covariance matrix with matrix elements is used

$$\Sigma_{b_{i}^{T},b_{j}^{T}}(\xi,\mathbf{x}) = \begin{cases} 0 & \text{if } y_{i}^{T} \text{ and } y_{j}^{T} \text{ are of different type} \\ \sigma_{b_{i}^{T},b_{j}^{T}}(\xi)\sigma_{b_{i}^{T},b_{j}^{T}}(\xi) \exp\left( -\sum_{k} \beta_{k}(\xi)(x_{y,i}^{T} - x_{j,k}^{T})^{2} \right) & \text{else} \end{cases}$$  \hspace{1cm} (5)$$

With these assumptions the integration in equation (2) can be done analytically. This leads to the likelihood function

$$f_{y^{T}|\theta,\psi,\xi,\mathbf{x}}(y^{T}|\theta,\psi,\xi,\mathbf{x}) = \frac{1}{\sqrt{2\pi}^{d} \sqrt{\det(\Sigma_{e}(\psi))} \sqrt{\det(\Sigma_{b_{i}^{T}}(\xi,\mathbf{x}))}} \cdot \exp\left( -\frac{1}{2} [y^{T} - y_{M}^{T}(\mathbf{x},\theta)]^{T} (\Sigma_{e}^{-1} + \Sigma_{b_{i}^{T}}(\xi,\mathbf{x}))^{-1} [y^{T} - y_{M}^{T}(\mathbf{x},\theta)] \right)$$  \hspace{1cm} (6)$$

(see Reichert and Schuwirth [2010] for more details). The identifiability problem between model and bias is solved by using means of zero for the bias and observation errors, an informative prior for the parameter(s) of the error model, $\mathbf{E}$, and by using the parameters of the bias, $\mathbf{B}$, for specifying how much bias one is willing to accept for each model variable.

2.2 Numerical implementation

Inference

An observation layout $L$ is used for inference. It represents points in time and space where observations are available for all considered variables. The mode of the posterior of the model parameters given the observations can be calculated by

$$\gamma^{0}(y_{i}^{T}) = \arg\max_{\gamma} \left( \log(f_{y^{T}|\theta,\psi,\xi,\mathbf{x}}(y^{T}|\theta,\psi,\xi,\mathbf{x})) \right)$$  \hspace{1cm} (7)$$

We then approximate the posterior by a normal distribution with the same mode and curvature at the mode, $N(\gamma^{0}, \Sigma)$. Here, the variance-covariance matrix $\Sigma$ is given as

$$\Sigma_{\gamma} \approx \left( \frac{\partial^{2} \log(f_{y^{T}|\theta,\psi,\xi,\mathbf{x}}(y^{T}|\theta,\psi,\xi,\mathbf{x}))}{\partial \gamma^{T} \partial \gamma} \right)_{\gamma^{0}}^{-1} = \left( \frac{\partial^{2} \log(f_{y^{T}|\theta,\psi,\xi,\mathbf{x}}(y^{T}|\theta,\psi,\xi,\mathbf{x}))}{\partial \gamma^{T} \partial \gamma} \right)_{\gamma^{0}}^{-1}$$  \hspace{1cm} (8)$$
where the second derivative of the log posterior can be approximated numerically as

\[
\frac{\partial^2 \log(f_{\text{post}}(\mathbf{y}, \mathbf{y}^0, \mathbf{x}))}{\partial \mathbf{y}_i \partial \mathbf{y}_j} \bigg|_{\mathbf{y} = \mathbf{y}^0}
\]

\[
\log(f_{\text{post}}(\mathbf{y}^i, \mathbf{y}^0 + \delta \mathbf{e}_i + \delta \mathbf{e}_j, \mathbf{x})f_{\text{post}}(\mathbf{y}^0 + \delta \mathbf{e}_i + \delta \mathbf{e}_j))
\]

\[
- \log(f_{\text{post}}(\mathbf{y}^i, \mathbf{y}^0 - \delta \mathbf{e}_i + \delta \mathbf{e}_j, \mathbf{x})f_{\text{post}}(\mathbf{y}^0 - \delta \mathbf{e}_i + \delta \mathbf{e}_j))
\]

\[
+ \log(f_{\text{post}}(\mathbf{y}^i, \mathbf{y}^0 - \delta \mathbf{e}_i - \delta \mathbf{e}_j, \mathbf{x})f_{\text{post}}(\mathbf{y}^0 - \delta \mathbf{e}_i - \delta \mathbf{e}_j))
\]

\[
= \frac{4\delta_i \delta_j}{\dot{\mathbf{y}}^T \mathbf{V} \dot{\mathbf{y}}} (9)
\]

(Gelman et al. [1995]). The approximate posterior distribution of the parameters is then given as \( \Gamma|\mathbf{Y} \sim \mathcal{N}(\mathbf{y}^0, \mathbf{\Sigma}_T) \).

**Prediction**

For the prediction, in contrast to Reichert and Schuwirth [2010], linearised error propagation is used to reduce computation time. The simplified prediction for layout \( L_2 \) is given by

\[
\mathbf{y}^{L_2} + \mathbf{B}^{L_2}_M | \mathbf{y}^{L_2} \sim \mathcal{N}(\mathbf{y}^{L_2}_M | \mathbf{y}^0) + \mathbf{E}[\mathbf{B}^{L_2}_M | \mathbf{y}^{L_2}_M, \Gamma](\mathbf{y}^{L_2}_M | \mathbf{y}^0)
\]

\[
\mathbf{V}^{L_2} \mathbf{\Sigma}_{T} (\mathbf{V}^{L_2})^T + \text{Var}[\mathbf{B}^{L_2}_M | \mathbf{y}^{L_2}_M, \Gamma](\mathbf{y}^{L_2}_M, \mathbf{y}^0))
\]

(11)

with

\[
\mathbf{V}^{L_2} = \frac{\partial(\mathbf{y}^{L_2}_M + \mathbf{E}[\mathbf{B}^{L_2}_M | \mathbf{y}^{L_2}_M, \Gamma]})}{\partial \mathbf{y}^T}
\]

(12)

The conditional distribution of the bias becomes a multivariate normal distribution under our assumptions. \( \mathbf{E}[\mathbf{B}^{L_2}_M | \mathbf{y}^{L_2}_M, \Gamma] \) and \( \text{Var}[\mathbf{B}^{L_2}_M | \mathbf{y}^{L_2}_M, \Gamma] \) are the mean and covariance matrix of the bias. In our case, layout \( L_2 \) represents the validation period of 10 years.

Similarly, we can calculate the prediction for layout \( L_1 \) analogue to \( L_2 \) with mean

\[
\mathbf{E}[\mathbf{B}^{L_1}_M | \mathbf{y}^{L_1}_M, \Gamma] = \text{Var}[\mathbf{B}^{L_1}_M | \mathbf{y}^{L_1}_M, \Gamma] \cdot \Sigma^{-1}_{E_1} \cdot (\mathbf{y}^{L_1}_M - \mathbf{y}^{L_1}_M (\mathbf{x}, \mathbf{\theta}))
\]

(13)

and covariance matrix

\[
\text{Var}[\mathbf{B}^{L_1}_M | \mathbf{y}^{L_1}_M, \Gamma] = (\Sigma^{-1}_{E_1} + \Sigma^{-1}_{B_M})^{-1}.
\]

(14)

The marginals of the predictions (11) and (13) can be combined to marginal predictions of the complete layout \( L_1 \cup L_2 \).

### 3. Didactical Example

As a simple example, we use a continuous-time, linear model with two output variables
as was used in Reichert and Schuwirth [2010]. The observation layout consists of observing both variables, \( g \) and \( h \), at the time points \( \{t_1, t_2, \ldots, t_n\} \). This leads to the deterministic model function

\[
y_{1/t}^l(\theta) = (g(t_1, \theta), \ldots, g(t_n, \theta), h(t_1, \theta), \ldots, h(t_n, \theta))^T
\]

with the parameter vector \( \theta = (a, b, c)^T \). The error model consists of independent normal distributions with standard deviations \( \sigma_{E_g} \) and \( \sigma_{E_h} \) for the variables \( g \) and \( h \) at all points in time. This leads to the parameters \( \psi = (\sigma_{E_g}, \sigma_{E_h})^T \) of the error model. Finally, we assume a Gaussian stochastic process in time with standard deviations \( \sigma_{B_g} \) and \( \sigma_{B_h} \) for each of the model variables \( g \) and \( h \), for describing the bias. Using the correlation time \( t_{\text{corr}} = \beta^{-0.5} \) instead of \( \beta \) to parameterize equation (5), leads to the parameter vector \( \xi = (t_{\text{corr}}, \sigma_{B_g}, \sigma_{B_h})^T \) for the bias.

The model given by the equations (15) and (16) represents in a very simple way coupling of two output variables in a multivariate model (by the multiplicative factor \( c \)). The intercepts and slopes of the two model output variables are not independent. We will produce synthetic data for this model with two types of bias: The first type consists in choosing the intercepts and slopes of the two model output variables independently. Second, we add a “zig-zag” line of bias to the model outputs. As the model cannot fit both linear components equally well, the quality of fit of each component will depend on the choice of the prior of the bias. By specifying the prior of the bias, the user can choose how much bias is acceptable in each of the two variables. As this can hardly be done in absolute terms a priori, we choose an exponential prior for the standard deviations of bias in both output variables. We assume independent normal priors for \( a \) and \( b \) (with means of 1 and 0.2 and standard deviations of 0.5 and 0.2, respectively), lognormal priors for \( c \), \( t_{\text{corr}} \), \( \sigma_{E_g} \) and \( \sigma_{E_h} \) (means 2, 3, 0.2 and 0.2 and standard deviations 2, 0.3, 0.02 and 0.02, respectively), and exponential priors for \( \sigma_{B_g} \) and \( \sigma_{B_h} \) (with means of 0.2 and 0.5).

![Figure 1](image)

**Figure 1.** Data points (markers), median (solid) and 95% credibility interval (shaded area with dashed boundaries) of \( y_{1/2}^{1/2} + B_{1/2}^{1/2} \mid Y_{1/2}^{1/2} \) and median of \( y_{1/2}^{1/2} \).

Figure 1 shows the model predictions for \( y_{1/2}^{1/2} + B_{1/2}^{1/2} \mid Y_{1/2}^{1/2} \) and \( y_{1/2}^{1/2} + B_{1/2}^{1/2} \mid Y_{1/2}^{1/2} \) for one case of prior choice of bias. The results demonstrate that our posterior knowledge is much more
precise in time domains with data than in the extrapolation range. Further results of the residuals between model output and observations show that the suggested technique is able to divide the residuals probabilistically into bias and observation error. A test of a different prior choice of bias indicates that in case of conflicting objectives (good fit of variable $g$ versus good fit of variable $h$) the prior standard deviation of the bias is a crucial value for trading-off one objective versus the other. The results of the linearized error propagation technique are shown in Fig. 1, and show slightly smaller uncertainty bounds than the results without the linearization shown in Reichert and Schuwirth [2010].

4. APPLICATION TO BIOGEOCHEMICAL AND ECOLOGICAL LAKE MODEL

In the following sections we go through the steps described in section 2 for the application case of a joint calibration of the lake model BELAMO applied to three lakes.

4.1 BELAMO: Model description

The calibration technique described in section 2 was applied to the Biogeochemical and Ecological LAke MOdel (BELAMO). It aims at a joint calculation of mass balances of nutrients, oxygen, organic particles, phytoplankton and zooplankton. Its box version describes the lake as four boxes: epilimnion, hypolimnion and two sediment boxes. In these boxes, concentrations of ammonium, nitrate, phosphate, oxygen, degradable and inert dead organic particles and (in the aggregated version used for this paper) one group of phytoplankton and one group of zooplankton are modelled.

BELAMO was implemented in AQUASIM (version 2.1f), a computer program for the identification and simulation of aquatic systems (Reichert [1994], Reichert [1998]). For a detailed description of the box version of the model see Mieleitner and Reichert [2008]. For a detailed description of the changes made compared to the model used in Mieleitner and Reichert [2008] and illustrative figures of the model structure and processes accounted for in each of the model compartments, see Dietzel et al. [2010]. Manual calibrations showed the difficulty of calibrating the model evenly well for all output variables and all lakes.

4.2 Study area

The model was applied to long-term observations of the three Swiss lakes Greifensee, Lake Zurich and Walensee. As measured by prevailing phosphorus concentrations Greifensee is still eutrophic, Lake Zurich rather mesotrophic and Walensee is an oligotrophic lake. For a detailed description of the main lake attributes see Mieleitner et al. [2006].

4.3 Data

Monthly measured profiles of physical, chemical and biological variables for Lake Zurich and Walensee were obtained from 1976 to 2005. For Greifensee, monthly to daily measurements of physical, chemical and biological variables were obtained from 1987 – 2004. Information on inflows into the lakes (physical and chemical parameters) and meteorological data were received from federal and cantonal agencies.

4.4 Prior distributions

The parameters included in the parameter estimation procedure comprise a selection of influential and most uncertain model parameters. To be able to estimate the importance of bias in the calibration, fixed standard deviations were assumed for the observation error representing the parameter set $\Psi$. Finally, exponential priors were chosen for the standard deviations of the bias of all considered output variables, for which the probability density
increases with decreasing value of the standard deviation. This reflects the desire to avoid bias if possible. In general, the bias is assumed to be larger than the observation error.

For the parameter estimation a newer version of UNCSIM (Reichert [2005]), a program package for statistical inference, identifiability analysis and uncertainty analysis, was used by coupling with AQUASIM. Linear approximation of the posterior distribution and estimation of the prediction uncertainty was done with R.

4.5 Results

The calibration technique described above was applied to a calibration period of 9 (for Greifensee) and 20 (for Lake Zurich and Walensee) years. The parameter vector $\theta$ for the inference of this application case consists of maximum specific phyto- and zooplankton growth and death rates, the fraction of inert material in lake inputs as well as maximum aerobic and anaerobic mineralization rates. The prediction uncertainty was estimated for the validation period of the following 10 years. Due to the large computation time, for a first rough estimate of the results, the prediction uncertainty was only estimated with the maximum specific phytoplankton growth and dates rates and the fraction of inert material in lake inputs. Here the deterministic model function $y^L_M(\theta)$ consists of discrete-time model outputs of the different variables oxygen, phosphate, nitrate, phytoplankton and zooplankton in the epilimnion and hypolimnion of all three lakes.

Fig. 2 and 3 show examples of simulation results of BELAMO for both the calibration and validation time. They represent the median of the distribution of model results ($y^L_M$) due to the posterior distribution of model parameters $\theta$ conditional on the model results of layout 1 without contribution of the bias. Furthermore, the 68% credibility intervals are shown. The results depict our knowledge about the true state of the system (without measurement errors). In the first part, observations were used to enhance our knowledge.

![Figure 2](image-url)

**Figure 2.** Phosphate concentration in the epilimnion of Greifensee. Data points (markers), median (solid) and 68% credibility interval (shaded area with dotted boundaries) of $y^L_M + Z^L_M$ | $y^L_M$ (model results plus bias conditional on observations of layout 1) for both layouts and median of $y^L_M$ (results without bias) for the whole simulation time (dashed).

In the second part, observations were neglected, which can be seen as prediction of the “future” state. Hence, plotted data points were used for the calibration of the model only for the observation layout (layout 1). Observations within layout 2 are presented for validation of the predictive power of the model.

Depending on the correlation time (in our case around 2 months), the gained information about the bias for layout 1 gives information also for the beginning of the validation time.
After the correlation between the bias of layout 1 and 2 decreased to 0, the median of \( y_{m}^{1,2} + B_{m}^{1,2} | y_{m}^{1} \) (solid line in Fig. 2 and 3) equals the median of \( y_{m}^{1,2} \) (dotted line in both figures).

Figure 3. Phytoplankton concentration in the hypolimnion of Lake Zurich. See caption of Fig. 2 for description of the different symbols.

5. DISCUSSION

For both the didactical example and the application case, it becomes obvious that simulation results plus bias are mostly able to reproduce the data during the calibration time. This results from sufficient knowledge about the system in times where measurements are available. Furthermore, the observation error was assumed to be small compared to the contribution of the bias. Both aspects lead to small uncertainties during the calibration time, which is especially true for the shown oxygen results in Greifensee.

The observations of the calibration time help to get information about the model parameters, which improve the prediction. However, due to the deviations between model results and data, the uncertainty is estimated to be much larger for the validation period. As the model results are in less good agreement with the data for the application to BELAMO than for the didactical example, the difference in the uncertainties is larger in this case. This difference is even more extreme for the phytoplankton than for the oxygen results. Hence the estimated uncertainty for the validation period is very large. Those results are meaningful, as poor models should not be used for prediction purposes. In case they are, the user should be aware of the high uncertainty.

A comparison of the remaining results not shown in this paper indicates that annual patterns of both phyto- and zooplankton seem to be most difficult to be represented by the model. The deviations most likely result from simplifications in the spatial and functional aggregation in the model. Those simplifications are less realistic for biological than for chemical variables.

Furthermore, the good representation of data by the model results plus bias compared to the relatively poor pure model results demonstrates the large contribution of bias in model results. This is especially true for the rather complex model BELAMO. A more detailed insight into the importance of bias for the different output variables could be given by the comparison of prior and posterior marginal distributions of the model parameters and the standard deviations of the bias not presented in this paper. Those results indicate how much can be learned about the parameters due to the available data.

In general, the main aim should be to decrease bias by improving the model. But especially in complex models bias will be present nevertheless and a transparent way has to be found to deal with it. The proposed technique makes the estimation of uncertainties in model predictions possible while still fulfilling the statistical assumptions of the error model.
Although the technique requires subjective choice, i.e. choice of prior distributions of standard deviations of bias for each of the output variables, it is a transparent way of uncertainty analysis and makes explicit what often is done implicitly. This subjective choice influences the results of uncertainty estimates, but it is unavoidable for multi-objective calibrations, as there are no objective criteria to distribute the bias among different output variables. However, the decision is made by the best prior knowledge the modeller has about the system. In specific, the user gets information about the importance of the bias in model simulation. The results show the importance of using such techniques, on the one hand, and, on the other hand, the need to carefully use deterministic models as BELAMO for prediction purposes. The deviations between data and model results demonstrate the difficulty of a joint calibration of three different application objects over a long-term period.

REFERENCES

Mieleitner, J. and P. Reichert, Analysis of the transferability of a biogeochemical lake model to lakes of different trophic state, Ecological Modelling, 194, 49-61, 2006.
Parameter optimization and Bayesian inference in an estuarine eutrophication model of intermediate complexity

P. Ding, C. E. S. Hoyle, M. E. Borsuk

Thayer School of Engineering, Dartmouth College
Hanover, NH 03755 USA (Peng.Ding@Dartmouth.edu)

Abstract: The Neuse River Estuary, North Carolina, has been experiencing severe consequences of eutrophication in recent years including excessive algal blooms, low levels of dissolved oxygen, declining shellfish populations, large fishkills, and outbreaks of toxic microorganisms. As in many other marine systems, nitrogen has been identified as the pollutant of concern in the estuary because it is believed to stimulate the excessive algal growth that is at the root of other ecological problems. A model incorporating the mechanisms of algal growth and nutrient consumption in the Neuse River Estuary is formulated mathematically and implemented in the computer software AQUASIM. Key model parameters of water quality interest are calibrated to observations of system variables by minimizing the sum of the squares of the weighted difference between actual measurements and simulated results. The calibrated model reproduces the observed seasonal patterns of key system variables and thus demonstrates a predictive capability that is of use to policy makers when they are making decisions for sustainable environmental management. As future work we will implement Bayesian parameter estimation, which would improve the robustness of decision support by accounting for parameter uncertainty using probability distributions. Eventually, our model will be linked with a Bayesian version of the SPARROW watershed model as a Bayesian Network to be used for developing an adaptive implementation modeling and monitoring strategy (AIMMS) for the Neuse River basin.

Keywords: eutrophication; process dynamics; mechanistic models; bayesian inference; predictive uncertainty; environmental management

1. INTRODUCTION

Calibrated mechanistic models, together with a characterization of uncertainty, are indispensable tools for understanding environmental system dynamics and for supporting environmental management decisions. In this paper, we will examine both optimization-based and Bayesian parameter estimation methods using a simulation model of eutrophication in the Neuse River Estuary (North Carolina, USA). The methodological framework is applicable to a wide variety of disciplines (e.g., hydrology, ecotoxicology and air pollution) [Arhonditsis, 2007].

The Neuse River Estuary (NRE, Figure 1) has been experiencing characteristic symptoms of eutrophication since the late 1970s. Investigations conducted by the Division of Water Quality (DWQ) of North Carolina indicated that algal blooms, stimulated by excessive nutrients, especially nitrogen, cause low dissolved oxygen levels contributing to the extensive fish kills. These fish kills, algal blooms, and correspondingly high levels of chlorophyll a prompted DWQ to place the NRE on the 303(d) list of impaired waters in 1994.
A Total Maximum Daily Load (TMDL) for nitrogen was then developed based on the Nutrient Sensitive Waters (NSW) Management Strategy and additional environmental modelling. This TMDL, calling for a 30% reduction in nitrogen loading, was approved by EPA in 2002, and rules to support the NSW Management Strategy were fully implemented by 2003.

However, at this point we have not yet observed any significant decrease in actual nutrient loading to the estuary, although nitrogen loads from point sources have been reduced by 65%. Thus, the goal of a 30% reduction in total nitrogen loading and the anticipated reduction of chlorophyll a standard violations have not yet been achieved. This phenomenon might be due to the accumulation and recycling of nutrients in riverine and estuarine sediments. Therefore, development of a more detailed model to better understand the complex nutrient dynamics in the NRE is necessary.

In this paper, we develop an NRE simulation model of intermediate complexity. On the one hand, our model is more complex than the zero-dimensional Bayesian probability network model (Neu-BERN) [Borsuk et al. 2003] since detailed process mechanism is incorporated. On the other hand, it is less complex than the two-dimensional Neuse Estuary Eutrophication Model (NEEM) (Bowen, 2003) in which daily values of response variables are predicted on both a longitudinal and vertical grid. The model we describe here includes two vertical sediment layers intended to capture the mechanisms of nutrient accumulation and release. In addition to providing model results calibrated by conventional optimization-based parameter estimation, we highlight the use of Bayes’ theorem to describe parameter uncertainty and update our knowledge as new data become available.

2. MODEL DESCRIPTION

The development of our model is based on the vertical 4-box lake model developed by Mieleitner and Reichert [2008]. The estuary model is essentially a combination of multiple lake models in a row. For our modeling purpose, the Neuse River Estuary is divided longitudinally into six sections (Figure 1). These divisions are based on the hydrological and water quality characteristics in different regions of the estuary as well as the available sampling points. The nutrient concentration decreases while the salinity concentration increases as we move downstream in the NRE. The middle four sections (Upper, Middle, Bend, Lower) are ones of interest since they are where most water quality violations take place. Each of the middle regions (Upper, Middle, Bend, Lower) are divided vertically into four “boxes” including two water layers and two sediment layers (Figure 2). The water body is divided into two parts, the epilimnion and hypolimnion, because the NRE has been observed to stratify into two layers between the frequent mixing events. For shallow water such as the NRE, sediments play a critical role in providing nutrients for algae because accumulated nutrients in sediments may support algal growth for a long period of time after external nutrient loading reductions. By including two sediment layers (Figure 2), the mechanisms of nutrient interactions are modelled more precisely, thus providing a better simulation and forecast in the long run.
The model also includes the River and Sound sections. However, the river is modeled as having just one water layer since the movement and mixing in the Neuse River does not allow it to stratify. The Sound does not have any sediment layers as we are not concerned with the sediment dynamics in the sound. The state variables we are tracking throughout the estuary and over time are listed in Table 1.

Table 1: Description and units of state variables as well as the time interval their corresponding measurement.

<table>
<thead>
<tr>
<th>State Variable</th>
<th>Units</th>
<th>Description</th>
<th>Data Span</th>
<th>Approximate Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>X_ALG</td>
<td>gDM/m³</td>
<td>Algae (dry mass)</td>
<td>1994-2000*</td>
<td>Weekly</td>
</tr>
<tr>
<td>X_ZOO</td>
<td>gDM/m³</td>
<td>Zooplankton (dry mass)</td>
<td>1994-2000**</td>
<td>Twice monthly</td>
</tr>
<tr>
<td>X_S</td>
<td>gDM/m³</td>
<td>Degradeable particulate</td>
<td>1994-2000**</td>
<td>Twice monthly</td>
</tr>
<tr>
<td>X_I</td>
<td>gDM/m³</td>
<td>Inert organic particulate</td>
<td>1994-2000**</td>
<td>Twice monthly</td>
</tr>
<tr>
<td>X_inorganic</td>
<td>gDM/m³</td>
<td>Inorganic particulate</td>
<td>1994-2000**</td>
<td>Twice monthly</td>
</tr>
<tr>
<td>S_HPO4</td>
<td>gP/m³</td>
<td>Phosphate-phosphorus - dissolved</td>
<td>1995-2000</td>
<td>Twice monthly</td>
</tr>
<tr>
<td>S_NH4</td>
<td>gN/m³</td>
<td>Ammonia-nitrogen - dissolved</td>
<td>1995-2000</td>
<td>Twice monthly</td>
</tr>
<tr>
<td>S_NO3</td>
<td>mgN/l</td>
<td>Nitrate-nitrogen - dissolved</td>
<td>1995-2000</td>
<td>Twice monthly</td>
</tr>
<tr>
<td>S_O2</td>
<td>mg/l</td>
<td>Oxygen - dissolved</td>
<td>1995-2000</td>
<td>Weekly</td>
</tr>
<tr>
<td>S_S</td>
<td>mg/m³</td>
<td>Salinity</td>
<td>1996-2000</td>
<td>Weekly</td>
</tr>
</tbody>
</table>

*We do not have direct measurement for algae. Instead, chlorophyll data are converted to algal concentration.

** We do not have direct measurement for zooplankton, degradable particulate and inert organic particulate. Instead, TSS measurement data is used as a aggregate measure.
The dynamics of algae (X_ALG) and zooplankton (X_ZOO) comprise the biological part of our model. The growth, death and respiration processes are modeled in both of the water compartments (Figure 2). However, unlike the lake model, we treat algae as one group, not four functional groups. This is reasonable for our model as we are interested in the effect nutrients have on the algal community rather than the different forms of algae. Phosphate, ammonium and nitrate are the most relevant nutrients and, together with dissolved oxygen and salinity, represent the dissolved state variables of our model. In the NRE we are particularly interested in modeling nitrogen, which is considered to be the limiting nutrient and is therefore subject to loading limitations. Biodegradable (X_S) and inert organic matter (X_I) summarizes organic particles resulting from death of algae and zooplankton and from zooplankton excretion as fecal pellets. The biological dynamics, chemical reactions and physical transportation are modeled within various compartments and by the links between the compartments (Figure 2). Table 2 summarizes these processes.

Table 2: Biological, chemical and physical processes simulated in NREM

<table>
<thead>
<tr>
<th>Biological dynamics</th>
<th>Chemical reactions</th>
<th>Physical processes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Growth of Algae and Zooplankton</td>
<td>Aerobic/ anaerobic mineralization</td>
<td>Exchange of oxygen between water and air</td>
</tr>
<tr>
<td>Death of Algae and Zooplankton</td>
<td>Nitrification</td>
<td>Light absorption</td>
</tr>
<tr>
<td>Respiration of Algae and Zooplankton</td>
<td></td>
<td>Sedimentation and diffusion</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Vertical mixing</td>
</tr>
</tbody>
</table>

Data sets representing measured values of the state variables are used as a basis for comparison to check how closely the model results reproduce measured results over the time period modeled. Measurements of chlorophyll a, phosphate, nitrate, ammonia and salinity are available in all compartments (River, Upper, Middle, Bend, Lower, Sound) in both water layers. Unfortunately, we do not have regularly measured data for the sediment layers, only snapshots from detailed investigations.

3. MODEL IMPLEMENTATION

Our model is implemented in AQUASIM [Reichert, 1998], a simulation and data analysis software for aquatic systems. Users are allowed to define state variables and processes within a configuration consisting of compartments and links of the available types (Figure 3). The partial differential equations describing the physical and biological dynamics of our estuary system are solved with the DASSL [Petzold, 1983] implementation of the backward-difference GEAR integration technique [Gear, 1971].

Figure 3. The four functional units in AQUASIM and their relationships.
4. PARAMETER CALIBRATION

4.1 Optimization-Based Parameter Estimation

To perform parameter calibration is essentially to solve a constrained optimization problem. The optimization objective function is to minimize the sum of the squares of the weighted difference between actual measurements and simulated results within the constraints of parameter ranges. Two numerical techniques implemented in AQUASIM are available for this parameter estimation task. The secant method [Ralston and Jennrich, 1978] converts our nonlinear least square fitting problem to a linear one by approximating the objective manifold with a secant plane through previous step objective function values. The downhill simplex method [Nelder and Mead, 1965] finds iteratively an improved searching direction in terms of decreasing the sum of squares of the weighted difference between actual measurements and simulated results using local information of function values only. Both secant and downhill simplex methods are derivative-free algorithms which are much less expensive computationally than derivative-based methods, such as gradient descent. The secant method is much more efficient in terms of convergence rate since it rapidly jumps to the position found by parabolic extrapolation, whereas the downhill simplex method slowly moves down the “gradient” of the objective function. Therefore, in a practical sense, parameter estimation usually starts with a secant algorithm. Also, we usually run multiple parameter estimation processes with different initial guesses of target parameters to check the convergence of the algorithm.

4.2 Bayesian Learning of Parameter Uncertainty

In water quality assessment and management, mechanistic simulation models are powerful in terms of understanding physical and biochemical processes, predicting aquatic ecosystem response to external nutrient loading changes and supporting the environmental policy making process [Reckhow and Chapra, 1999]. But these models are never “perfect” since models are always a simplified version of reality [Stow et al., 2003]. The imperfectness is described by the term uncertainty, usually characterized quantitatively by probability distributions. Parameter uncertainty is the type of uncertainty we are often interested in investigating with an established mechanistic simulation model. However, conventional optimization-based parameter calibration techniques fail to support this analysis in two aspects. One problem is that conventional parameter calibration providing a best fit of the model parameters to the dataset only computes a set of fixed parameter values. This procedure is formally referred as maximum likelihood estimation by frequentist statisticians. Equation (1) illustrates this mathematically, where likelihood represents the data likelihood conditional on the parameter vector $\theta$ of our model.

$$\theta_{\text{optimal}} = \arg \max_{\theta} \left[ \text{likelihood}(\text{data} | \theta) \right]$$  \hspace{1cm} (1)

The other problem is that conventional parameter calibration makes the model data-specific by fitting it to a given dataset at the moment. As new datasets become available, the model has to be recalibrated without a possibility of considering the previous parameter results. In other words, we do not update existing knowledge about model parameters, but rather we make our model “learn” from the very beginning whenever new information is available.

Fortunately, Bayes theorem provides us a means for addressing these two problems. Existing knowledge about parameters, or a prior distribution, is updated according to Bayes updating rule (equation 2) when new information becomes available, resulting in updated parameter knowledge, or a posterior distribution, of parameters.

$$f_{\text{post}}(\theta | \text{data}) = \frac{f_{\text{like}}(\text{data} | \theta) f_{\text{pri}}(\theta)}{\int_{\theta} f_{\text{like}}(\text{data} | \theta) f_{\text{pri}}(\theta)}$$  \hspace{1cm} (2)
In this equation, \( f_{post}, f_{like}, f_{pri} \) denote the posterior distribution, model likelihood and prior distribution, respectively. To reduce the uncertainty about parameters, we can use the Bayesian updating rule iteratively, using the previous step’s posterior distribution as the prior distribution for the next step and obtaining a new posterior distribution with the new data provided. Computationally, the posterior parameter distributions are often approximated by their samples simulated based on Markov Chain Monte Carlo (MCMC) method. Bayesian inference changes our perspective of seeking a single “optimal” value for each model parameter, to finding a joint distribution of parameter sets. These probability distributions provide a straightforward way to quantify parameter uncertainty that can be easily used by decision makers/policy planners [Reckhow, 1994; Arhonditsis et al., 2006].

As future work, we will implement the Bayesian parameter estimation using UNCSIM [Reichert, 2004], a program package for statistical inference and uncertainty analysis. UNCSIM not only provides routines for frequentist (maximum likelihood) but also for Bayesian (Markov Chain Monte Carlo and importance sampling techniques) parameter estimation. Therefore, we can easily compare these two methods.

5. RESULTS

Optimization-based parameter estimation in AQUASIM requires two inputs: parameters of interest and measured data used as the fitting target. To reduce the computational time as well as to fit parameters more precisely, parameters of interest were divided into five groups and they are being calibrated sequentially based on relevant observed data (Table 3).

<table>
<thead>
<tr>
<th>Group Order</th>
<th>Parameter</th>
<th>Calibration Data</th>
<th>Additional Assumptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Dispersion constants</td>
<td>salinity and oxygen (epi* and hypo**)</td>
<td>none</td>
</tr>
<tr>
<td>2</td>
<td>Algal parameters (physical)</td>
<td>oxygen (epi), chlorophyll a (epi and hypo)</td>
<td>Fix phosphate, ammonia and nitrate concentrations at mean values</td>
</tr>
<tr>
<td>3</td>
<td>Zooplankton parameters</td>
<td>oxygen (epi), chlorophyll a (epi and hypo)</td>
<td>Fix phosphate, ammonia and nitrate concentrations at mean values</td>
</tr>
<tr>
<td>4</td>
<td>Sediment/water bacteria parameters</td>
<td>oxygen (epi and hypo), phosphate and nitrate (epi and hypo)</td>
<td>none</td>
</tr>
<tr>
<td>5</td>
<td>Algal parameters (nutrient)</td>
<td>all data</td>
<td>none</td>
</tr>
</tbody>
</table>

epi* represents epilimnion and hypo** represents hypolimnion

Dispersion parameters are important as they characterize the physical properties of the NRE. Thus, we calibrated these parameters according to salinity and oxygen measurements from both of the water layers as the first step. A comparison (Figure 4) of model results and measurements indicates that the simulations generally reproduce the seasonal and intermediate time patterns of salinity and oxygen levels and match the downstream trends. The simulated oxygen concentrations in the hypolimnion are currently too low relative to observed values, but we expect that to improve as we calibrate sediment and water bacterial parameters (Group 4). We are currently working on the remaining four groups of the parameter calibration.
Figure 4. Simulated vs. measured salinity (left) and oxygen (right) concentrations of the two water layers across the Upper, Middle, Bend and Lower sections with time range from 6/1/1996 to 12/27/2000. Concentration units of salinity: mg/m$^3$. Concentration units of oxygen: mg/l.
Black lines: simulated model results in epilimnion, Yellow: measurements in epilimnion
Blue: simulated model results in hypolimnion, Magenta: measurements in hypolimnion

6. OUTLOOK
The next step will be to conduct Bayesian inference on the most significant parameters so that we have full parameter probability distributions for implementation in the Bayesian Network model. This will be linked with a Bayesian implementation of the SPARROW watershed nutrient delivery model [Smith et al., 1997] as part of a larger project. It is intended that these two linked sub models will facilitate a process of Bayesian adaptive learning and management on the basis of new evidence, including remotely sensed land use and water quality data (Figure 5).

ACKNOWLEDGEMENTS
This work was supported by a grant from the EPA Office of Research and Development’s Advanced Monitoring Initiative (AMI) Pilot Projects Focused on GEOSS (Global Earth Observation System of Systems) (PI: K.H. Reckhow, Duke University). The authors thank Martin Frey, Peter Reichert, Johanna Mieleitner, and Anne Dietzel of Eawag for model support and Ibrahim Alameddine of Duke University for assistance with data.
Figure 5. Integrated system of Neuse Estuary Model and SPARROW watershed nutrient delivery model.

REFERENCES


Uncertainty of a biological nitrogen and phosphorus removal model

A. Cosenza, G. Mannina, G. Viviani

Dipartimento di Ingegneria Idraulica ed Applicazioni Ambientali, Università di Palermo, Viale delle Scienze, 90128 Palermo, ITALY. Corresponding Author (tel. +39 0916657756; fax: +39 0916657749; E-mail: mannina@idra.unipa.it)

Abstract: In the last few years, the use of mathematical models in wastewater treatment plant (WWTP) processes has become a common way to predict WWTP behaviour. However, mathematical models generally demand advanced input for their implementation that must be evaluated by an extensive data gathering campaign, which cannot always be carried out. This fact, together with the intrinsic complexity of the model structure, leads to model results that may be very uncertain. Quantification of the uncertainty is imperative. However, despite the importance of uncertainty quantification, only a few studies have been carried out in the wastewater treatment field, and those studies only included a few of the sources of model uncertainty. This paper presents an uncertainty assessment of a mathematical model simulating biological nitrogen and phosphorus removal. The uncertainty assessment was conducted according to the Generalised Likelihood Uncertainty Estimation (GLUE) methodology. The model was based on Activated-Sludge Models 1 (ASM) and 2 (ASM2). Different approaches can be used for uncertainty analysis. In the present study, the GLUE procedure was employed. The GLUE methodology requires a large number of Monte Carlo simulations in which a random sampling of individual parameters drawn from probability distributions is used to determine a set of parameter values. Using this approach, model reliability was evaluated based on its capacity to globally limit the uncertainty. The method was applied to a full-scale WWTP for which quantity-quality data were gathered.

Keywords: activated sludge models; calibration; nitrogen phosphorus removal; uncertainty analysis; wastewater modelling

1 INTRODUCTION

During the last decades, interest in mathematical modelling of wastewater treatment plant (WWTP) processes has increased. The mathematical models have contributed to increasing knowledge in this field. The activated-sludge models (ASMs) (Henze et al., 2000) proposed by the working group of the International Water Association (IWA) have been applied several times in order to best understand how to improve plant design, how to optimise the processes and which control strategies to prefer (Jeppsson et al., 2007; Salem et al., 2002; Flores et al., 2005). The application of WWTP models makes it possible to improve designs: an overall optimization of the involved processes increases efficiency and enables better compliance with increasingly stringent regulations (Belia et al., 2009). Nevertheless, modelling activated-sludge systems is not easy because biological systems, as well as each natural environmental system, are intrinsically complex and are subject to many natural variations. The activated-sludge process cannot be considered a well-characterised process, and some activated-sludge model parameters are uncertain (Flores et al., 2008). Consequently, the application of ASMs requires a great number of assumptions concerning influent wastewater composition and model parameters. Traditionally, WWTP
process simulators assume constant rather than variable model parameters and are thus not capable of taking into account the inherent randomness of these parameters (Flores et al., 2008). Such assumptions have a significant influence on the model predictions and could lead engineers to make erroneous decisions during their design or optimization of a project. Therefore, an accurate analysis and quantification of model uncertainty is imperative. The assessment and presentation of uncertainty are widely recognised as important parts of the analysis of complex water systems (Beck, 1987). They allow modellers to identify the sources of error in the modelling process and to learn how the errors spread to the model outputs.

During the last few years, scientific research in the wastewater modelling field has focused on uncertainty issues, and some publications have appeared in the literature (among others, Neumann and Gujer, 2008; Benedetti et al., 2008; Flores et al., 2008; Sin et al., 2009; Bixio et al., 2002). Different approaches have been proposed for the assessment of uncertainty. These studies have all demonstrated that taking uncertainty into account can affect the decision-making process for a design project or in the prediction of plant behaviour. For instance, Bixio et al. (2002) suggested a methodology for quantification of the uncertainty of a WWTP model using a Monte Carlo simulation. The methodology takes into account the input and parameter uncertainties in order to evaluate how the uncertainty can improve the likelihood of meeting effluent standards without requiring above-average capital investments. Bixio et al. (2002) demonstrated that considering uncertainty can even reduce the capital investment. However, few studies yet deal with uncertainty in wastewater quality modelling. Indeed, as pointed out by Sin et al. (2009), the field of uncertainty analysis of WWTP models is still in its infancy. This conclusion was also one of the main outcomes at the recent WWTPMod2008 workshop on uncertainty (Belia et al., 2010).

Manifold sources of uncertainty in the model predictions have been identified and, as suggested by the literature, can be classified as follows (Belia et al., 2009): 1) uncertainty in external influence factors (e.g., the measurement errors that affect the observed input data), which can have significant effects on model predictions; 2) uncertainty in the model structure, which is attributable to an inappropriate model that is too simple compared with the complexity of the real system that it tries to represent (e.g., inadequate selection of processes algorithms); 3) uncertainty in model parameter values (e.g., wrong estimation of parameter values); and 4) uncertainty in the numerical calculations used to solve the model algorithms (e.g., programming errors).

Bearing these considerations in mind, this study presents an uncertainty analysis of a mathematical model for the simulation of biological nitrogen and phosphorus removal processes. The uncertainty analysis is assessed by means of the Generalised Likelihood Uncertainty Estimation (GLUE) methodology (Beven and Binley, 1992). This methodology is one of the most widely used methods for investigating uncertainties in hydrology, and is now spreading into other research fields. The goal of the study was to test the suitability of such a methodology for WWTP modelling in order to provide an easy and useful tool for uncertainty assessment, a topic that is still only rarely addressed compared with other research fields.

2 MATERIALS AND METHODS

2.1 Uncertainty assessment

As outlined in the introduction, in the present study, the GLUE methodology was used for uncertainty quantification (Beven and Binley, 1992). The GLUE methodology is a non-formal Bayesian methodology that facilitates an easy assessment of uncertainty. On the other hand, in formal Bayesian methodology, a formal description of the likelihood function is always required. This is extensively discussed in the literature. For example, Mantovan and Todini (2006) reported incoherencies of the GLUE methodology with Bayesian inference. Beven et al. (2007) replied that formal Bayesian inference is a special case of GLUE when a formal likelihood description is used. Regardless of this discussion, Freni et al. (2009b) also demonstrated that both methods perform similarly when the GLUE methodology is based on the same assumptions as the Bayesian approach. To apply the GLUE methodology, the model was run using a uniform randomly sampled parameter sets.
By means of a likelihood measure, $E$, parameter sets can be classified; sets with poor likelihood weights with respect to a user-defined acceptability threshold ($T_r$) are discarded as “non-behavioural”. All parameter sets coming from the behavioural simulation runs are retained, and their likelihood weights are rescaled so that their cumulative total sum is equal to one. The likelihood measure $E$ represents the ability of the model to fit real data. On the other hand, the acceptability threshold $T_r$ represents a user-defined critical value indicating the minimum value of $E$ that each modelling simulation should have in order to be representative of the model behaviour with respect to the analysis aim. $T_r$ is usually set equal to zero.

In the present study, the following equation was employed as a likelihood measure (Freni et al., 2009a):

$$L(\theta_i|Y) = \exp\left(\frac{-\sigma^2_{M_j-O_j}}{\sigma^2_{O_j}}\right)$$

where $\theta_i$ represents the $i^{th}$ set of model parameters (randomly generated), $\sigma^2_{M_j-O_j}$ is the variance of the residuals between model and observations of the $j^{th}$ simulated model output and $\sigma^2_{O_j}$ is the variance of observations for the period under consideration.

Treating the distribution of likelihood values as a probabilistic weighting function for the predicted variables, it is possible to assess the uncertainty associated with the predictions (conditioned on the definition of the likelihood function) of the input data and model structure. A method of deriving predictive uncertainty bands from the behavioural simulations using the likelihood weights has been shown by Beven and Binley (1992). The uncertainty bands are calculated using the 5th and 95th percentiles of the predicted output likelihood weighted distribution. Wider bands mean higher uncertainty in the estimation of the modelling output and thus lower confidence in the model results.

The GLUE methodology can also be used to analyse the impact of each parameter on modelling outputs. Plotting the cumulative likelihood distributions for the set of behavioural simulations ($E \geq T_r$) and the set of unconditioned cumulative distributions, respectively, it is possible, by comparing the deviation between the two, to determine if the model output in question is sensitive to changes in the parameter values. If little difference between the two cumulative distribution functions (CDFs) is found, the parameter is considered insensitive with regard to the model output. Conversely, if a great difference is found, the parameter is considered to be sensitive. Applying the nonparametric Kolmogorov–Smirnov d-statistic (maximum distance between the two CDFs), a measure of sensitivity is introduced, i.e., $d = 1$ is the most sensitive and $d = 0$ is non-sensitive (Hornberger and Spear, 1981; Beven et al., 2008; Freni et al., 2009a). This sensitivity analysis is used to determine the relative importance of each parameter in the model structure. It is evident that the GLUE results can be affected by the definition of parameter variation ranges. This definition can influence the analysis because it defines the domain where the model uncertainty is evaluated. The selection of the parameter variation ranges can be accomplished by considering the physical meaning of the parameters, but this approach cannot be used for conceptual parameters that have a weak link to the physical system. In addition, this approach can produce variation intervals that are too wide, thereby leading to the problems described above.

### 2.2 The case study

The municipal activated-sludge WWTP under study was located in Sicily, Italy, and had an outflow to the Mediterranean Sea. The plant was designed for a design capacity of 40,000 inhabitant equivalents (IE). The influent of the WWTP, with average and maximum values of 400 m$^3$/h and 600 m$^3$/h respectively, consisted of domestic and non-industrial wastewater produced by a nearby refinery. After the pretreatment step (coarse grit removal, fine grit removal, filtration with a rotating panel, sand and grease removal), the influent was introduced into an equalisation tank with a volume of 1,700 m$^3$ in which the wastewater was discontinuously aerated (3 h/d). The effluent of the equalisation tank flowed to the biological activated-sludge treatment area, which consisted of an activated-
sludge reactor designed according to a Bardenpho scheme and a secondary clarifier (with a volume of 2,885 m³) where the COD, N and P removal was accomplished. More specifically, the activated-sludge reactor was composed of three completely mixed compartments of different sizes. The first compartment operated as an anaerobic zone, the second as an anoxic zone and the third as an aerobic zone, with volumes (V) of 900, 1,140 and 5,800 m³, respectively. Returned activated sludge (RAS) from the bottom of the secondary clarifier and internal mixed-liquor recirculation (MLR) from the end of the aerobic zone were pumped to the anaerobic and anoxic zones, respectively. Aeration was supplied by 900 fine bubble diffusers positioned on the bottom of the aeration zone.

The influent flow rate (QINF) under normal operating conditions was approximately 400 m³/h; the MLR flow rate (QMLR) and the RAS recirculation (QRAS) were generally set to 3 and 1.5 QINF, respectively. The waste activated sludge (WAS) was simply dewatered by a belt-press filter.

2.3 Model description

The model adopted in this study was able to reproduce the nitrification-denitrification/enhanced biological phosphorus removal processes occurring in a full-scale WWTP characterised by a Bardenpho scheme. The model was built on ASM concepts; in particular, the complexity of the ASM2 model was reduced by omitting processes that did not play a significant role and components that did not have a dominant effect upon the kinetics of the processes (Henze et al., 2000).

The model describes the following variables: ammonia (NH₄-N), nitrate (NO₃-N), total soluble phosphate (Pₜₕₜ), total COD (CODₜₜ), particulate material (Xₜₕₜ) and total soluble COD (CODsol). The model did not take into account the solid-liquid separation in the secondary clarifier. Accordingly, the concentrations of the soluble components in the returned activated sludge were assumed to be equal to the effluent concentrations from the aerobic reactor.

The model was calibrated using the design and operational data of a real WWTP and chemical-physical data collected during an ad hoc field data-gathering campaign carried out during the period from 01 March 2006 to 12 April 2006 in the same plant. In particular, total suspended solids (TSS), total and soluble COD (CODₜₜ and CODsol), orthophosphate (P-PO₄), total soluble phosphorus (Pₜₕₜ), NH₄-N, NO₃-N, dissolved oxygen, temperature, pH and air flow rate were monitored in different sections of the plant. The samples were withdrawn from the effluent of the anaerobic, the anoxic and the aerobic tank (sections 1, 2, 3 and 4) and from the RAS channel (section 5). For further details on the model calibration and the gathering campaign, the reader is referred to Cosenza et al. (2008). In the following section, the model parameters are reported according to the ASM notation (Henze et al., 2000).

3 RESULTS AND DISCUSSION

3.1 Methodology application

In order to reduce the number of parameters, a preliminary local sensitivity analysis was carried out before the model uncertainty analysis (details are discussed in Cosenza et al., 2008). Following this preliminary model parameter analysis, the number of sensitive model parameters was 29. These 29 parameters were allowed to vary during Monte Carlo simulations, while 12 were held constant. In this way, the impact of such a reduction on the reported uncertainty outputs was quantified. In particular, for each sensitive model parameter, a uniform distribution was considered, and the broadest variation range–drawn from the relevant literature (Henze et al., 2000; Weijer and Vanrolleghem, 1997; Petro Alfonso and Maria da Conceição Cunha, 2002; Iacopozzi et al., 2007; Sin et al., 2009; Flores Alsina et al., 2008) was selected in order to explore the overall confidence region. It is important to emphasise that the parameter variation ranges considered during the uncertainty analysis were equal to the ranges that were used during the sensitivity analysis and the model calibration steps.
To apply the GLUE methodology, the defined parameter space for each sensitive parameter was randomly sampled with the Monte Carlo technique. In particular, 1 000 behavioural simulations (approximately 440 000 simulations) were run on randomly sampled parameter sets. This number of simulations has been found to be consistent with the objectives of the present study. Specifically, a sample dimension was selected, verifying that the uncertainty analysis was not affected by any bias linked to the number of Monte Carlo simulations. This study was carried out by analyzing the statistics and variations and changing the sample dimensions between 100 and 1 000 behavioural simulations (Bertrand-Krajewski et al., 2002). For each parameter set, the uncertainty was spread by running the model simulation, and a likelihood measure was computed for each model variable in order to evaluate the ability of the model to fit real data. At the end of this step, we had 1 000 likelihoods and 1 000 respective dynamic profiles for each model variable. According to Equation (1), the likelihood measure varies between 0 and 1, with a likelihood of 1 corresponding to a perfect fit. For large errors, the likelihood becomes 0 as the ratio goes to infinity. In order to evaluate how the parametric uncertainty was spread in the model output variables (owing to the sensitive model parameters), the nonparametric Kolmogorov–Smirnov d-statistic (d K-S) was assessed (Smirnov, 1948).

As outlined in the previous paragraph, in order to evaluate how the uncertainty of the sensitive model parameters was spread in each model output variable and in the global model response, the d K-S was calculated for each model output. Therefore, the cumulative likelihood distribution of the 1 000 likelihoods was computed and compared with the cumulative density function of the uniform distribution (CDFu) for each model output variable and sensitive parameter. Regarding the global model response, the cumulative likelihood of the model (E MOD) was computed by considering the weighted sum of the efficiencies of the n model outputs for each model run and then comparing it with the CDFu. In particular, the d K-S represents the maximum absolute value of the distance between the cumulative likelihood distributions and the CDFu, and it is generally used as a measure of parameter sensitivity. Values of d close to 1 indicate a very high sensitivity, whereas a d-value close to 0 indicates low sensitivity (Thorndahl et al., 2008).

### 3.2 Model parameter uncertainty results and model uncertainty bands

In Figure 1, the d K-S values of some model variables are shown. The results show a different response in terms of the spread level of parametric uncertainty, depending on the model output variable considered. The COD showed the highest values of d K-S for the different parameters (Figure 1a, b and c). Indeed, the d K-S of the rate constant for the lysis of heterotrophic biomass (bH) had the maximum absolute value, equal to 0.9 for COD TOT,2. The CODsol,3 (Fig. 1c) is the model variable for which the parametric uncertainty was more evident. In this case, 42% of the model parameters had a d K-S value higher than 0.2.

In general, $\mu_{AUT}$ and $K_{NH}$ showed the highest sensitivity. The calibrated value of the nitrifying growth rate, $\mu_{AUT}=1.08$ d$^{-1}$, was in agreement with literature values (referred to a temperature of 20 °C): 1.2 d$^{-1}$ (Makinia et al., 2005), 1.8 d$^{-1}$ (Rieger et al., 2001), 1 d$^{-1}$ (Henze et al., 2000) and 0.55 d$^{-1}$ (Ferrer et al., 2004). Conversely, referring to $K_{NH}=1.41$ gN/m$^3$, there was a substantial difference with the value presented by Makinia et al. (2005), where the value is 0.2 gN/m$^3$. However, lower values of this parameter are commonly encountered in pilot plants because of a lower diffusion limitation related to the higher turbulence and smaller flocs in comparison with full-scale plants (Henze et al., 2000). Regarding the calibrated value of $Y_{PO4}=0.11$ gP/gCOD, this value was not close to the default value (Henze et al., 2000). Similar results were obtained by Machado et al. (2009), which the authors explained by the presence of glycogen-accumulating organisms (GAOs) not considered by ASM2. Indeed, under anaerobic conditions, PAOs and GAOs can alternatively store fermentation products, SA. While PAOs utilise the energy obtained from the hydrolysis of polyphosphate and from glycogen degradation, GAOs use only the energy from glycogen degradation. This fact justifies the decreases of the $Y_{PO4}$ value with the increase in the GAO population (Ferrer et al., 2004).

Nevertheless, despite agreement with the aforementioned parameter values, it has to be stressed that with respect to their level of sensitivity, the parameter significance levels may differ from one plant to another because of changes in the process scheme and available data (among others, Ruano et al., 2007).
In terms of global model response (Figure 1d), it is worthwhile to observe that the parametric uncertainty is flatter because it is computed on the weighted sum of the efficiencies of the n model outputs. However, the global model response, like the single model variable, was still the most sensitive model output and was sensitive to the parameters $b_H$, $\mu_H$ and $\mu_{AUT}$. In Figure 2, the cumulative likelihood of $b_H$, $\mu_H$ and $\mu_{AUT}$ are reported. The results show that in terms of the global model response, despite the higher compensation effect among the parameters, the model outputs were strongly influenced by such parameters.

It is worth mentioning that quantitative prioritization of the model parameter was really useful. The finding that almost half the parameters had little effect on the performance is an implicit rebuke to the architects of these models, implying the models are too complex.

The 1000 likelihoods and the 1000 respective dynamic profiles for each model variable were used to compute the cumulative likelihood of each variable at each simulation time. According to the GLUE procedure, the 5th and 95th percentiles of the cumulative likelihood distributions for each simulation time step and for each model output were then used for calculating uncertainty bands. In Figure 3, the uncertainty bands of some model outputs considered during the uncertainty analysis are reported.
Analyzing the graphs reported in Figure 3, it is evident that the results of the uncertainty analysis performed in this study show different response with respect to the model output considered. Indeed, the uncertainty band widths for COD (Fig. 3b, d and f) are generally wider than the nitrogen components (Fig. 3a and e). Such a result is likely due to the different amplitude of the model parameter ranges employed. Indeed, as aforementioned, in this study in order to explore the parametric space without considering different classes of uncertainty, the broadest parameter variation range drawn from literature was employed. Such a fact affects the uncertainty band widths especially for those model outputs influenced by several sensitive parameters. Indeed, the uncertainty of $S_{\text{PO}_4}$ (Fig. 3c) is smaller than the others because among the parameters for which the uncertainty has been studied (sensitive parameters), only three are directly connected to the phosphorus removal processes. Such results are consistent with previous studies carried out on ASMs (Sin et al., 2009). Indeed, due to the fact that ASMs are overparametrized, such models provide different responses in terms of uncertainty band widths. The uncertainty bands of $X_{\text{TSS}}$ model outputs have not been reported because it has almost no uncertainty according to Sin et al., (2009). Indeed, in the model under study, the settling parameters are not subject to uncertainty. It is important to point out that the model uncertainty response is certainly influenced by the subjective hypotheses that have been made applying the GLUE.
methodology such as the choice of the efficiency-measure. Indeed the method has many limitations as to make the results almost useless. So it would be interesting to study how the ASM model uncertainty changes changing the efficiency measure in the GLUE methodology as well addressed by Freni et al., (2009a) in the field of urban-drainage modelling. However, despite these drawbacks, the results demonstrate, according to other authors (Flores et al., 2008; Benedetti et al., 2008; Melcer et al., 2003), that when uncertainty in the ASM model inputs is considered, the results of a well structured and calibrated model might be questioned; so an accurate uncertainty analysis is important depending on the objective of the study. As a matter of the fact, although the model calibration provided acceptable results giving efficiency ranging between 0.42 and 0.75 (Cosenza et al., 2008), in terms of uncertainty a significant proportion of the measured data fall near or on the extremes of the uncertainty bands. Such a fact confirms even more the importance in the quantification of the model uncertainty. Indeed, the quantification of the uncertainty pointed out that the model structure has to be improved in order to provide a better reproduction of the simulated phenomena. The GLUE is confirmed to be a good tool for uncertainty assessment also for WWTP modelling. Such a methodology, although can be affected by subjective hypothesis, it is a valuable and easy to use tool for uncertainty. With regards to the computational time needed for the implementation, in particular with regards to the Monte Carlo simulations, the Latin hypercube sampling could be an optimum choice especially for computational demand models.

4 CONCLUSIONS

The uncertainty analysis of a mathematical model simulating biological nitrogen and phosphorus removal processes was performed using the GLUE methodology (Beven and Binley, 1992). In order to evaluate how the model parameter uncertainty can influence the response of the model, a uniform distribution of the broadest model parameter space was considered, and several Monte Carlo simulations were conducted in order to investigate this space. The output of each simulation was compared with measured data and a measure of the likelihood was created. The 5th and 95th percentiles of the cumulative likelihood distributions were then used for calculating uncertainty bands for each model output variable. From the analyses, the following conclusions can be drawn:

- The uncertainty analysis performed in this study gave different responses for several of the model outputs considered. The results were strongly dependent on the width of the parameter range and on the parameters selected during the sensitivity analysis. In addition, the correlation between the sensitive and non-sensitive parameters was ignored.
- The uncertainty assessment showed that despite the fact that the best-fit model response between the measured and simulated values was acceptable, the model approach needs to be improved in order to correctly simulate the system. Indeed, the model showed a wide band of uncertainty, with a significant proportion of the measured data results (far more than 5%) falling near or on the extremes of the uncertainty bands.
- The study demonstrated the suitability of the GLUE methodology for wastewater treatment plant modelling, although the methodology is based on some subjective choices that can affect the results. Nevertheless, as a first screening study (i.e., studies for evaluating the magnitude of the polluting emission, the classification of pollutant impacts, etc.) it could be a feasible and good solution.

REFERENCES

A comparison of Fuzzy, Bayesian and Weighted Average formulations of an in-stream habitat suitability model.

C.J. Brookes, Vikas Kumar, S.N. Lane

Institute of Hazard, Risk and Resilience and Department of Geography, University of Durham, Durham, UK

University of Sheffield, Sheffield, UK

ChristopherBrookes@btinternet.com

Abstract: Three variations of a simple in-stream habitat suitability model were implemented and the effect on output for one organism at 22 sites on one short section of a river was examined. The model uses only two factors, depth and velocity, to calculate quality and a third factor, width to quantify utility. The implementations were based on 3 multi-criteria evaluation approaches: Weighted Average, Fuzzy Sets and Bayesian probability. There was broad agreement between the formulations but important differences in detail. The model outputs indicate that uncertainty arising from model formulation is significant and can have a bearing on planning decisions. There is a complex interaction between the formulations and the characteristics of the sites. Suitability models should be used thoughtfully and implemented in ways that: facilitate exploratory analysis; present ranges of possible outputs; present indicators of uncertainty; and facilitate back tracking to explain the outputs.

Keywords: Uncertainty; Multi-criteria evaluation; Habitat suitability; Fuzzy; Bayes

1 INTRODUCTION

Environmental management decisions have to be taken in the context of uncertainty. Nowhere is this more the case than in relation to river habitat. Although gauging river flows is itself a challenging science, gauging river habitat is more difficult because: (1) it can vary continuously over spatial scales varying from a few m to the whole catchment (Lammert and Allen, 1999; Folt et al., 1998); and (2) the impacts of changes in drivers (e.g. river flow) are continuous in time and space. Although the relationship between habitat and use of habitat by organisms is a complex issue, there is still an academic focus upon determining the habitat template and the effects of interventions (e.g. flow regulation) on that template for rivers that are, effectively, ‘ungauged’ because of the difficulty of measuring ecological elements at scales that match the spatial and temporal variability in the drivers of those elements. Thus, habitat template modelling is an important element of river management (e.g. Leclerc, 2005).

This paper looks at one particular source of uncertainty in habitat modelling, model formulation. Three alternative implementations of a habitat suitability model: (1) Fuzzy sets; (2) Bayesian probability; and (3) Weighted Average approaches, are compared. The driving question is: could different model formulations lead to different decisions?
2 THEORY

2.1 Habitat suitability analysis

Habitat suitability analysis (HSA) is a methodology for informing environmental conservation and restoration decisions. The viability of species and organisms depends fundamentally on the availability of suitable habitat to support them. By analysing habitat in a model framework, in a way that recognises the interactions of factors that influence habitat, it is possible to establish a base-line, to estimate the predicted effects of management interventions and to monitor habitat change. Here, the focus is on instream river habitat, driven by two of the most common factors used to determine habitat suitability: flow depth and flow velocity (Lane et al., 2006).

This kind of analysis has a very long and established history. It is based upon the assumption that the ecologically-useable habitat in a river depends on several key parameters, notably flow velocity and depth, wetted perimeter, substrate, water temperature and pH (e.g. Elso and Giller, 2001, Maddock et al., 2001, Leclerc, 2005). Traditionally, emphasis has been placed upon hydraulic variables, notably velocity, depth and wetted perimeter because: (1) they are important controls upon organism metabolism, both directly and indirectly, controlling the balance between access to food and expenditure on swimming; and (2) the spatial and temporal changes in higher order parameters (such as pH and temperature) should track changes in velocity and depth to some extent. Width, depth and velocity are commonly incorporated into some form of habitat score such as PHABSIM (e.g. Milhous et al., 1984), which can be used to determine habitat suitability in situations where flow is uniform, or approximately uniform (Milhous et al., 1989). Concerns with the simple hydraulic basis of PHABSIM have resulted in a much wider research field concerned with developing hydraulic models of habitat, including development of more sophisticated one-dimensional hydraulic treatments as well as two-dimensional habitat modelling (e.g. Leclerc et al., 1995, 1996; Tiffan et al., 2002). The latter is important because the habitat that can be used by an organism varies continually in space (cross-stream and downstream) but may be prohibited by the data and computational demands when habitat assessment is needed at the scale of entire river basins. This is the subject of much debate in the habitat modelling literature (see Leclerc, 2005). However, there is an important additional issue that is the focus of this paper. It is now recognised that the choice of habitat suitability analysis is about more than the choice of appropriate hydraulic representation. Rather, it must capture the severe uncertainty associated with only poor or even no measured ecological data. Likewise, different suitability analyses use models with very different kinds of assumptions over how to handle this uncertainty as well as how to combine the different criteria (e.g. depth, velocity) to determine habitat.

2.2 Multi criteria evaluation

Habitat suitability analysis is a form of multi-criteria evaluation (MCE). MCE maps physical attributes such as depth to a value. Value is an abstraction; it cannot be measured directly but is inferred from measurements. Value can be expressed in two main ways: as a number such as an index or score (e.g. 80 out of 100), or as a class (e.g. "Good"). The transformation from attribute to value is by a value function. The second step is to combine multiple values into a single value using a combination process. Before the values can be combined they must be normalised to a single scale to avoid implicit weighting. If criteria are not equally important they should be explicitly weighted. There are two interpretations of value: quality and utility. Quality is calculated by the value and combination functions. Habitat utility depends on intrinsic quality and how much there is, so there is a third step - the utility function.

MCE is a huge field with a large literature and many formal models (see Jankowski 1995). Here three approaches are considered. The first approach is pragmatic. Attributes are scored independently, on a common scale of value, and then combined by taking the mean of each attribute score (e.g. Jiang and Eastman, 2000). In a two-step process the value functions map
attributes to scores and the combination function averages the scores. Because criteria can be weighted this method is known as weighted average (WA).

The next two approaches, Fuzzy and Bayes, address uncertainty by using soft classes. In conventional Boolean classes membership is binary and exclusive. Soft classes allow partial membership of classes and multiple memberships of alternative classes. Partial membership expresses how definite the classification is. An entity can simultaneously be classed as, say, both "good" and "medium" with different degrees of definiteness. For example, if there are three classes, poor, medium and good, an entity can be classified using a membership vector of the form \( \langle M_{(\text{poor})}, M_{(\text{medium})}, M_{(\text{good})} \rangle \). A vector \( \{0,100,0\} \) indicates certainty that the class is medium, \( \{10,80,10\} \) indicates less certainty while \( \{33,33,33\} \) says all classifications are equally likely. The value function maps attributes to class membership functions for each criterion. The next step is to combine criteria classes to a resultant classification using cross-memberships. With 2 criteria and 3 classes for each there are 9 class combinations. These can be mapped onto fewer resultant classes as illustrated in Table 1 of section 3 where good-good maps to excellent and both good-medium and medium-good map to very good.

Fuzzy and Bayes differ in the interpretation of membership and, consequently, the calculation of cross-membership. In Fuzzy, partial membership represents vagueness about the meaning of the classes while in Bayes partial membership represents the probability of membership (Fisher 2000). Fuzzy relates to conceptual uncertainty and Bayes relates to factual uncertainty. The Bayesian approach is grounded in probability theory and is consistent with statistical error modelling (Aspinall and Veitch 1993). It is the basis for Bayesian Belief Networks, which integrate quantitative and qualitative uncertainties in a single rigorous framework (Henriksen et al., 2006). The Fuzzy approach is based on Fuzzy set theory, which was developed to address the problem of vagueness (Fisher 2000, Legleiter and Goodchild 2005). Implementation of Fuzzy is less rigorous than Bayes (Fisher 2000) and here only one typical implementation is considered. The Bayesian class combination function is the joint probability. With reference to Table 1, the probability of habitat being excellent is the joint probability that depth is good and velocity is good. For Fuzzy it is a set operation to calculate the intersection of two classes. Typically, it is implemented as a fuzzy_and (see Box 1) operator (Fisher 2000). The Fuzzy and Bayes cross membership functions are shown in Box 1.

**Box 1 Fuzzy and Bayes Cross membership functions**

<table>
<thead>
<tr>
<th>Given two membership vectors ( {X_1, X_2, \ldots X_i} ) and ( {Y_1, Y_2, \ldots Y_i} ) for criteria X and Y, there are ( i \times j ) cross-memberships ( M_{ij} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayesian joint probability ( M_{ij} = X_i \times Y_j ) for all ( i, j )</td>
</tr>
<tr>
<td>Fuzzy_and ( M_{ij} = \min(X_i, Y_j) ) for all ( i, j )</td>
</tr>
</tbody>
</table>

### 3 THE IN-STREAM HABITAT MODEL

The in-stream habitat model (ISH) uses two criteria, flow depth and flow velocity, to evaluate habitat quality. Separate value functions were created for each criterion using presence/absence data from the literature. Problems in this approach are discussed elsewhere (Lane et al., 2006). Each numerical value of depth or velocity maps to a quality value. In the original model (Lane et al., 2006) quality is expressed as fuzzy membership of 3 classes: good, medium and poor. Combining depth and velocity criteria gives 9 cross membership possibilities but these are mapped to 6 final classifications of habitat quality: nil, very poor, poor, good, very good and excellent. Table 1 shows the correspondence between the cross memberships and the final habitat quality class. To calculate utility, quality is expressed as a score, from 0 for Nil up to 6 for Excellent (Table 1) and multiplied by the width.
Table 1. Cross-classifications and associated scores

<table>
<thead>
<tr>
<th>Depth</th>
<th>Velocity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Poor</td>
</tr>
<tr>
<td>Poor</td>
<td>Nil-0</td>
</tr>
<tr>
<td>Medium</td>
<td>Very Poor-1</td>
</tr>
<tr>
<td>Good</td>
<td>Poor-2</td>
</tr>
</tbody>
</table>

The Bayes and Fuzzy methods use different combination functions and generate different cross membership values, as explained above. Utility is also calculated differently. For Bayes utility is based on expectation, where expectation is the sum of (probability * score) for each class. So for example, if excellent habitat scores 6 and the membership is 80%, and good habitat scores 4 and the membership is 20%, the overall score is (4.8 plus 0.8) which gives 5.6. Thus Bayes quality is a continuous function. Fuzzy uses the most likely outcome. The class with the highest membership is selected to give a single, Boolean, class. This means that the fuzzy model generates discrete quality scores \( \{0,1,2,3,4,6\} \). The WA method maps attribute values map to a criterion score in the range 0 to 3. The combined quality score is thus between 0 and 6 giving the same range as the Fuzzy and Bayes methods.

Each site is a cross-section of a river channel and comprises a number of elements each with a modelled velocity, depth and width so there is a final step to generate a summary value for each site. Quality and utility are modelled per element. Quantity is expressed by scaling to width. Utility of a site is the sum of the utility of each element. Aggregation means that at the site level Fuzzy quality is not restricted to integers and can be any value from 0 to 6. With factors \( d(\text{depth}), v(\text{velocity}), \text{classes } M_{ij} \text{ and scores } S_{ij} \), quality per element, \( Q \) is:

3.1.1  WA: \( Q = \frac{Q_d + Q_v}{2} \)
3.1.2  Fuzzy: \( Q = S_{ij} \) where \( i,j \) are given by \( M_{ij} = \max(M_{ij}) \) for all \( i,j \)
3.1.3  Bayes: \( Q = \sum(M_{ij} * S_{ij}) \) for all \( i,j \)

4  PROCEDURE AND RESULTS

The models were run for a single organism, adult brown trout, on 22 sites on a section of the River Don in Sheffield. The same habitat requirements were used in each run. Requirements were input as vectors specifying class, class boundaries and precision \( ((\text{poor, min, max}, P)) \) per class, criterion and species. The precision parameter \( P \) makes the classes soft. A value has 100% membership if it lies between \( (\text{Min} + P) \) and \( (\text{Max} - P) \), and 0% membership if it is outside \( (\text{Min} - P) \) to \( (\text{Max} + P) \). Intermediate membership values are interpolated by a linear function. High values of \( P \) represent large uncertainty. The Bayes and Fuzzy models used the same membership function. The WA value function was derived from the same input data, essentially by multiplying the membership by the relevant class quality score (3 for high, 1 for medium, 0 for poor) for each criterion (see Figure 1). \( P \) is not used in WA.
Before running with field data the models were run with depth and velocity values in the range 0 to 1.5. Figure 2 plots quality against depth and velocity and shows some differences between the models. The models produce different ranges of outputs: Fuzzy (0,1,2,4), Bayes (0-3.56) and WA (0-6). This is a consequence of the high uncertainty in the habitat requirements for brown trout. No depth or velocity results in un-ambiguously good classification but many values are unambiguously poor. Also, off-diagonal classifications (Table 1) are more likely than the diagonals. For example, very good results from a combination of good-medium or medium-good while only one possibility, good-good, leads to excellent. There is thus a bias towards lower scores in the Fuzzy and Bayes models but not in WA where there is no uncertainty treatment of the input. Another difference in the models is the trade-off between factors. As Table 1 shows, medium-medium is better than good-poor in the classification models whereas in WA (1+1) is less than (3+0). The integer values for Fuzzy are the result of hardening the fuzzy classes into a single Boolean class for output. The plots indicate uncertainty in the outputs; steep gradients indicate high sensitivity to the inputs.

With the field data, there was enough general agreement between the models in terms of quality and utility to indicate that all three are consistent with each other. Flow velocities at most sites are too high for good habitat and most sites would be classed as poor by the velocity criterion. In contrast the depth is often good. Thus the expected upper bound for quality with the given sites is nearer 3 than the theoretical maximum of 6. Table 2 shows the minimum and maximum values for quality and utility produced by each model.

<table>
<thead>
<tr>
<th>Model</th>
<th>Average quality</th>
<th>Utility</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>min</td>
<td>max</td>
</tr>
<tr>
<td>Fuzzy</td>
<td>0.99</td>
<td>2.09</td>
</tr>
<tr>
<td>Bayes</td>
<td>0.94</td>
<td>2.37</td>
</tr>
<tr>
<td>WA</td>
<td>1.19</td>
<td>3</td>
</tr>
</tbody>
</table>
The WA model has a similar range of values to Fuzzy and Bayes but has higher minimum and maximum values. This can be explained by the way that WA trades-off depth and velocity scores. A poor velocity (0) can still generate an aggregated score of 3 if the depth is good, whereas for Bayes and Fuzzy the best would be 2 (see Table 1). For Bayes and Fuzzy the velocity has to be classed as Medium before the overall score can be 3 or more. With regard to utility, Fuzzy and WA are more similar to each other while Bayes has a larger range and more extreme values. The following section compares the results on a site-by-site basis.

Figure 3 shows quality per site for each model, sorted by Fuzzy quality. In general the differences between the models appear systematic; WA generates the highest scores, then Fuzzy with Bayes least. However, two sites, 2 and 13, have higher scores for Bayes, and in fact reverse the normal ordering of scores from WA-Fuzzy-Bayes to Bayes-Fuzzy-WA. These two sites are the highest ranked by the Bayes method but not by the other models. WA ranks 10 sites higher than site 13 and Fuzzy ranks 5 sites higher than site 2. Similarly, the models do not agree on the worst sites, which are: 7 and 12 (fuzzy), 4 and 14 (Bayes) and 2 and 12 (WA). Perhaps the most interesting result is that site 2 is ranked 2nd highest by Bayes and 2nd last by WA. The anomalous results for site 2 indicate uncertainty in the output. This is supported by looking at the membership vector used to generate the summary score. For site 2 it is {13, 27, 19, 14, 20, 7} compared to the more typical site 1 {25, 36, 38, 0, 0, 0}.

A more detailed look at site 2 and site 21, which has a similar quality score by the Fuzzy model, reveals how the models differ. Figure 4 shows the quality scores along the cross-sections. Site 2 is wider and shallower than site 21 and the velocity is lower. At site 2, velocity is medium and depth is poor/medium. At site 21 velocity is poor while depth is good. The improvement in velocity more than compensates for the poorer depth in the Bayes model but not the WA.

Figure 5 shows utility per site for each model, sorted by Fuzzy. Considering utility, the models agree more on the best and worst sites. Site 13 is ranked highest by all 3 models. The worst are 14 and 12 (Fuzzy), 14 and 18 (Bayes) and 14 and 12 (WA). At sites 12, 13 and 14 there is a big change in width that damps the quality differences and that explains the change in utility values and accounts for the agreement between the models (Table 3).

<table>
<thead>
<tr>
<th>Site</th>
<th>Width</th>
<th>Fuzzy Quality</th>
<th>Fuzzy Utility</th>
<th>Bayes Quality</th>
<th>Bayes Utility</th>
<th>WA Quality</th>
<th>WA Utility</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>6.51</td>
<td>1.14</td>
<td>7.41</td>
<td>1.13</td>
<td>7.33</td>
<td>1.20</td>
<td>7.79</td>
</tr>
<tr>
<td>13</td>
<td>15.60</td>
<td>2.1</td>
<td>32.73</td>
<td>2.37</td>
<td>37.03</td>
<td>2.02</td>
<td>32.09</td>
</tr>
<tr>
<td>14</td>
<td>6.03</td>
<td>1.28</td>
<td>7.71</td>
<td>1.01</td>
<td>6.1</td>
<td>1.93</td>
<td>11.65</td>
</tr>
</tbody>
</table>
CONCLUSION AND RECOMMENDATIONS

Looked at broadly, the models appear to agree, but in detail there are some big differences. Notably two models generated almost completely opposite quality results for one site. The models are very sensitive to the physical characteristics of the sites, as implied by site 2. Differences between the models are greater with respect to quality than utility because of the masking effect of width. In each model the sites with highest and lowest utility are consecutive sites on the river so although the different models do not agree in detail for each site, they do direct attention to the same part of the river.

Model implementation is a source of uncertainty and different formulations could lead to different decisions. This uncertainty should be conveyed to the end user. Standard statistical error modelling expresses both a range of values and confidence that the true value lies within that range. While the Bayes and Fuzzy models represent and propagate uncertainty in habitat requirements their output classifications and the WA scores do not convey uncertainty in the output. Taken together, the range of outputs produced by the three implementations indicates a range of possible outputs. This approach should be extended, using Monte-Carlo or similar methods, to generate a range of outputs in response to uncertainties in data, habitat requirements and model formulation. Models should output as much information as possible so that users can examine what leads to particular results. This would help users to assess the confidence in the outputs.
REFERENCES


Limitations in Interpretable Top-Down Effective Rainfall-Runoff Modelling

J.P. Norton

Integrated Catchment Assessment & Management Centre
The Fenner School of Environment & Society
The Australian National University, Canberra 0200, Australia
john.norton@anu.edu.au

Abstract: The effective rainfall-flow module of a "top-down" hydrological model has linear, fixed dynamics, giving a two-exponential instantaneous unit hydrograph (IUH) often interpreted as summing slow- and quick-flow components. Such components suffer uncertainty due to ignorance of rainfall variation between samples and ignorance of delay in the IUH. If they are obtained from linear-in-parameters models, ill conditioning in the conversion amplifies the uncertainty. The extent of these uncertainties is analysed, with examples. The paper also considers alternatives to performance assessment of such models by Nash-Sutcliffe efficiency. Rather than employing the sample mean as an output-prediction benchmark, they use almost equally simple predictors taking the flow correlation structure into account. Again numerical examples are given.

Keywords: Hydrology; uncertainty; model calibration; model performance measurement.

1 INTRODUCTION

This paper has two aims. The first is to analyse the possible extent of various errors arising in the parameter estimation of models relating effective (excess) rainfall and river flow (Evans and Jakeman [1998]). Attention is confined to linear, time-invariant (LTI), discrete-time models, calibrated on rainfall and flow records. Section 2 examines errors arising from the nature of the modelling process rather than from observational errors or from the model-structure-selection or parameter-estimation technique. The influence of non-linearity in catchment dynamics on uncertainty is not considered, being less amenable to analysis.

The second aim is to seek alternatives to Nash-Sutcliffe efficiency to measure how well a model fits observed flow. Section 3 examines measures based on simple flow predictors. These topics are not novel; rainfall-flow models are recognised as subject to many sources of error and Nash-Sutcliffe efficiency has been widely criticised ([Croke [2007]; Schaefli and Gupta [2007]). The contributions here are in showing how the ranges of certain errors depend on model characteristics and the observation interval, and in offering less flattering model-performance measures.

2 UNCERTAINTIES IN EFFECTIVE RAINFALL-FLOW MODELS

2.1 Errors in model due to unmodelled features of rainfall and IUH

The next two subsections discuss modelling errors, and hence uncertainties, due to inability of the hydrological model specification to capture features of the underlying physical
behaviour. These errors are fundamental and do not depend on details of the model or peculiarities of particular records. Straightforward analysis can establish bounds on the resulting uncertainties in unit-hydrograph components. To establish their distributions or to turn them into uncertainties in a modelled flow time series, distributional properties of the records have to be known, so that is not attempted.

2.2 Error due to ignorance of rainfall distribution over observation interval

Error arises first from the discrete-time nature of the model. The instantaneous unit hydrograph (IUH) is conceptually the continuous-time flow response to unit rainfall arriving all at time zero, i.e. the unit-impulse response \( h(t), t \geq 0 \). The model actually relates rainfall and flow observed synchronously at time intervals \( T \), say. Furthermore, unit input at time zero is rainfall registered as 1 at that instant, i.e. unit total rainfall over \(-T < t \leq 0\). The model produces the unit-pulse response \( \text{UPR} \) \( h_k, k = 0, 1, \ldots, \infty \), of flow values at successive observation instants \( t = kT \) resulting from unit rainfall registered at time zero. The response to rainfall registered by an observation depends on how the rainfall varied in the preceding time step. A calibrated model averages such responses over the record. Use of the model incurs errors through this average differing from the response appropriate to each time step of rain. The following analysis examines such errors, for the common interpretation of the UPR of an effective rainfall-flow model (Jakeman and Hornberger [1993]) as representing an IUH with quick- (\( q \)) and slow- (\( s \)) flow components:

\[
h(t) = g_q \exp\left(-\frac{t}{\tau_q}\right) + g_s \exp\left(-\frac{t}{\tau_s}\right), t \geq 0
\]

with time constant \( \tau_q > \tau_s \). The IUH gives the flow response due to rainfall \( u(t) \) in the period \(-T < t \leq 0\) as

\[
q(t) = \int_{-T}^{0} u(\tau) h(t - \tau) d\tau, t \geq 0
\]

The effects of not knowing the time course of rainfall over an observation interval are readily seen for a single IUH component (component \( i \) in general)

\[
h_i(t) = g_q \exp\left(-\frac{t}{\tau_i}\right), t \geq 0.
\]

The corresponding actual flow component due to unit rainfall in the observation interval up to time zero is found by substituting this IUH component and the actual rainfall history into (2). Comparing the extremes and an intermediate case, consider rain falling as (i) a very short pulse \( u(t) = \delta(T+5) \) at the start of the period, (ii) a very short pulse \( u(t) = \delta(0) \) at the end, or (iii) steady rain \( u(t) = \frac{1}{T}, \) all for \(-T < t \leq 0\), where \( \delta(t) \) is the Dirac \( \delta \) function, a unit-area impulse at \( t = 0 \). The corresponding responses are

\[
q_i(t) = \begin{cases} 
(i) & g_q \exp\left(-\frac{(t+T)}{\tau_i}\right) = g_q \exp\left(-\frac{T}{\tau_i}\right) \exp\left(-\frac{t}{\tau_i}\right), t \geq 0 \\
(ii) & g_q \exp\left(-\frac{t}{\tau_i}\right) \\
(iii) & \frac{g_s \tau_i}{T} \left(1 - \exp\left(-\frac{T}{\tau_i}\right)\right) \exp\left(-\frac{t}{\tau_i}\right)
\end{cases}
\]

They differ only in amplitude, by factors depending only on \( \tau/T \). The ratios of responses (ii) and (iii) to (i) are plotted in Figure 1 for a range of \( \tau/T \). The large ratios for time constants comparable with \( T \) imply high uncertainty in the quick-flow response to rainfall in any interval, due solely to ignorance of its time course. Although calibrating a model averages the responses over the record, the model’s performance on each large flow peak, as well as overall, may well be of concern, and Figure 1 implies high uncertainty for individual peaks. The effect of averaging in calibration can be analysed by assuming that, in finding \( h_k \) for given lag \( kT \), the rain
intensity in the time step ending \( kT \) ago is uniform on average over the record. This just says that no particular point between \((k+1)T\) and \(kT\) ago has different average rain intensity from any other; it says nothing about the shape or duration of rainfall events. The mean response to rain in the time step ending \( t \) ago is the response at time \( t \) to the mean rain, since expectation and integration are linear operations:

\[
E[q(t)] = E \left[ 0 \int_{-T}^{T} w(\tau) h(t-\tau) d\tau \right] \\
= \int_{-T}^{T} E[w(\tau)] h(t-\tau) d\tau
\]

so (iii) of (4) also gives the mean response to rainfall not favouring any particular point in the time step.

Figure 2 replots (4) with this mean response as reference, to show the variability of unit-rainfall response about the mean. The variability is large. For example, if \( \tau_i = T \) the response to rainfall in any one time step may approach 58\% higher or 42\% lower than the mean. Even with \( \tau_i = 2T \) it may approach 27\% higher or 23\% lower.

Figure 2. Ratios of responses to concentrated rainfall and mean response

2.3 Error due to uncertainty in delay of start of response

Flow response also incurs uncertainty from uncertainty in the dead time (pure delay) between rainfall and the start of the response to it. Rainfall-flow models usually take this delay as constant, yet transport delay through catchments varies with flow. Indeed, event-by-event variation of dead time with rainfall intensity is often visible in rainfall-flow records. The delay registered also varies with the pattern and movement of rainfall with respect to the rain gauge(s). A discrete-time model can be made responsive to varying dead time by allowing parameters to time-vary (Norton [1975]), but modelling of all influences on dead time would be complicated and difficult even with enough rainfall observations. The extent of error due to ignoring variation of dead time is readily analysed, however.

As before, consider an exponential IUH component, now subject to dead time \( t_d \), so that

\[
h_i(t) = g_i \exp\left(-\left(t - t_d / \tau_i\right)\right) = g_i \exp\left(t_d / \tau_i\right) \exp\left(-t / \tau_i\right), \quad t \geq t_d
\]
Thus dead time just alters the amplitude of the exponential at any instant during it. Underestimating the dead time by \( \delta t_d \) causes the amplitude to be overestimated by a factor \( \exp(\delta t_d / \tau_i) \), plotted for positive \( \delta t_d \) as a function of \( \delta t_d / T \) for various \( \tau_i / T \) in Figure 3.

The errors due to modest errors in dead time can be large. They depend strongly on the time constant of the response component, so are most serious for the quick-flow component, but Figure 4 shows that they are still substantial even for slow-flow components with time constants considerably greater than the observation interval.

2.4 Errors in quick-flow slow-flow model due to transformation from linear model

An effective-rainfall flow model is often obtained initially as an input-output equation linear in its parameters and variables, e.g. an ARMAX model (Jakeman et al. [1990]; Ljung [1995]). Linearity greatly simplifies parameter estimation and uncertainty analysis (Norton [1986]), but for physical interpretability the model must be transformed to another form, such as a unit-pulse response with additive quick-flow and slow-flow components. The transformation is easy for a linear second-order model but typically amplifies uncertainties. Denoting by \( q_k, u_k \) and \( e_k \) the flow, effective rainfall and equation error at time \( t = kT \), the pseudo-regression

\[
q_k = -a_1 q_{k-1} - a_2 q_{k-2} + b_0 u_k + b_1 u_{k-1} + e_k
\]

(7)

can be rewritten in \( z \)-transform form to give the input-output relation

\[
Q(z^{-1}) = \frac{b_0 + b_1 z^{-1}}{1 + a_1 z^{-1} + a_2 z^{-2}} U(z^{-1}) = \left( \frac{g_1}{1 - p_1 z^{-1}} + \frac{g_2}{1 - p_2 z^{-1}} \right) U(z^{-1})
\]

(8)

Here the transfer function linking the effective rainfall and flow transforms \( U(z^{-1}) \), \( Q(z^{-1}) \) has been split into partial fractions whose inverse \( z \)-transforms give sampled exponential components \( g_1 p_1^k \) and \( g_2 p_2^k \), \( k \geq 0 \), of the unit-pulse response, with time constants \( \tau_1 = -T / \ln p_1, \tau_2 = -T / \ln p_2 \). The transformation is algebraically simple:

\[
p_1 = \left(-a_1 + \sqrt{a_1^2 - 4a_2}\right), p_2 = \left(-a_1 - \sqrt{a_1^2 - 4a_2}\right),
\]

\[
g_1 = \frac{b_0}{p_1 - p_2}, g_2 = \frac{p_2 b_0 + b_1}{p_2 - p_1}
\]

(9)

but its effects on the uncertainties in the parameters of (7) must be examined. The local normalized sensitivities (proportional change in factor)/(proportional change in result) of the \( p \)'s and \( g \)'s to the \( a \)'s and \( b \)'s are
J. P. Norton / Limitations in interpretable top-down effective rainfall-runoff modelling

\[ S_{a1}^1 = \frac{a_1}{p_1} \left( p_1 + p_2 \right) \left( p_1 - p_2 \right), \quad S_{a1}^2 = \frac{p_1 + p_2}{p_2 - p_1}, \]
\[ S_{a2}^1 = \frac{p_1^2}{p_1 - p_2}, \quad S_{a2}^2 = \frac{p_1^2}{p_1 - p_2}, \quad S_{a2}^3 = \frac{p_1^2}{p_1 - p_2}. \]

\[ \left( p_1 + p_2 \right)^2 \left( p_1 g_2 - p_2 g_1 \right) \]
\[ \left( p_1 - p_2 \right)^2 g_1 \]
\[ \left( p_1 + p_2 \right)^2 \left( p_1 g_2 - p_2 g_1 \right) \]
\[ \left( p_1 - p_2 \right)^2 g_2. \]

A numerical example will show that some of these sensitivities are likely to be high. For \( p_1 = 0.5, p_2 = 0.9, g_1 = 3 g_2 \), i.e., time constants of 1.44 and 9.49 days and a quick-flow component three times the size of the slow-flow component, (10) gives

\[ S_{a1}^1 = -3.5, S_{a1}^2 = 3.5, S_{a1}^3 = -6.42, S_{a1}^4 = 19.25, \quad S_{a2}^1 = 2.25, S_{a2}^2 = -1.25, S_{a2}^3 = 1.875, S_{a2}^4 = -5.625. \]

\[ S_{b1}^1 = -1.667, S_{b1}^2 = 9, S_{b1}^3 = 2.667, S_{b1}^4 = -8. \] A few percent uncertainty in \( a_1 \) or about 10% uncertainty in \( a_2 \), \( b_0 \) or \( b_1 \) is enough to make the estimate of slow-flow amplitude unusable. The quick-flow amplitude also is very sensitive to uncertainty in \( a_1 \). Indeed, every normalized sensitivity is greater than unity in size, so uncertainty is amplified in every relation in computing the partial fractions. The numbers in this example are not contrived; the high sensitivities found here are typical.

It is also easy to show that the normalized sensitivity of \( \tau \) to \( p \) is \( \tau / T \), so the estimated slow-flow time constant, particularly, is highly sensitive to uncertainty in the corresponding pole. However, this is less serious than it seems, as the slow pole is normally well defined by the flow record in catchments displaying clear quick- and slow-flow components.

3 MEASURING MODEL PERFORMANCE

3.1 Alternative measures of model fit

The fit of the output of a hydrological model to the observed output is most often measured by the Nash-Sutcliffe efficiency (coefficient of determination)

\[ R^2 = 1 - \frac{\sum (y - \hat{y})^2}{\sum (y - \bar{y})^2} \] (11)

where \( t \) indicates time, \( y \) the observed output with sample mean \( \bar{y} \), and \( \hat{y} \), the model output. It is commonly interpreted as the proportion by which the output mean-square error (MSE) of the model reduces the sample variance of \( y \). Equally it can be viewed as comparing the model output MSE with that of the crude estimator \( \hat{y} = \bar{y} \). Usually \( R^2 \) measures the performance of a model driven by observed input alone, with any past output required supplied by earlier model-output values rather than observations, i.e., with the model run in simulation mode. The model could be compared with any other input-independent output estimator. If some such estimator performs better than \( \hat{y} = \bar{y} \), its performance is a more stringent benchmark. Specifically, consider an input-independent predictor of the output \( k \) time steps ahead. For fair comparison, \( k \) must conform with the aim of the model. If the model is to predict over \( l \) steps, \( k \) should equal \( l \). On the other hand,
if the model aims to explain the output on the sole basis of past observed input (as in simulation mode), one can ask how much better the output is explained by the input, via the model, than by the earlier output via a \( k \)-step predictor. The comparison most favouring the model is with \( k \) large enough for past output to be uninformative. The least favourable is with a 1-step predictor.

A simple, \( k \)-step, output-to-output predictor, linear in its parameters, can be found by minimizing the mean squared error

\[
s = E[(y_t - \hat{y}_t)^2]
\]

by choice of the parameter vector \( a \) of the predictor

\[
\hat{y}_t = a^T f_{t-k}
\]

where \( \hat{y}_t \) predicts \( y_t \) from time \( t-k \), \( a^T = [a_0, a_1, a_2, \ldots, a_m] = [a_0, a^T] \), \( f_{t-k} \) comprises any collection of \( m \) functions of observed output values no later than time \( t-k \) and \( f_{t-k} = [1, f_{t-k}^T] \). The predictor includes a constant term \( a_0 \times 1 \) to cover any long-term component not dependent on recent output behaviour. For a stationary point of \( s \),

\[
\frac{\partial s}{\partial a} = 2E[f_{t-k} (a^T f_{t-k} - y_t)] = 0
\]

so the minimum-mean-square-error (MMSE) estimator is

\[
a = E[f_{t-k} f_{t-k}^T]^{-1} E[f_{t-k} y_t]
\]

providing the inverse exists. It does if \( E[f_{t-k} f_{t-k}^T] \) is positive-definite, which is so unless \( a^T E[f_{t-k} f_{t-k}^T] a = E[(a^T f_{t-k})^2] = 0 \) for some non-zero \( a \), implying that \( a^T f_{t-k} = 0 \) in every realisation, clearly not true. As the Hessian of \( s \) with respect to \( a \) is \( 2E[f_{t-k} f_{t-k}^T] \), the positive-definiteness confirms that (15) minimizes \( s \). Parameter vector \( a \) is constant if the output series is stationary, which seems dubious for river flow but is justified by the expectation in (12) being over an indefinitely long flow sequence, averaging out seasonal and other variations. This conforms with \( R^2 \) measuring fit over the entire record.

Given the definition (12) of \( s \), it is no surprise that the MMSE estimator (15) is the probabilistic counterpart of the ordinary least-squares estimate of \( a \). Now

\[
E[f_{t-k} y_t] = E\left( f_{t-k} \left[ -\frac{1}{\sigma} \frac{1}{\sigma} \right] (y_t - \bar{y} + \bar{r}_k + \bar{y}) \right) = \left[ \frac{\bar{y}}{\sigma^2} - \frac{2}{\sigma^2} r_{t-k} + \bar{y} \right]
\]

\[
E[f_{t-k} f_{t-k}^T] = E\left( f_{t-k} \left[ -\frac{1}{\sigma} \frac{1}{\sigma} \right] (f_{t-k} - \bar{r}_k + \bar{f}) \right) = \frac{1}{\sigma^2} R_{t-k} + \bar{f}^T
\]

with \( \bar{y} \) the mean of the output, \( \sigma \) its standard deviation, \( \bar{f} \) the mean of \( f_{t-k} \), \( r_{t-k} \) an \( m \)-vector with \( E[f_{t-k, i}] / (\sigma, \sigma) \) as element \( i \), \( \sigma \) the standard deviation of \( f_{t-k, j} \), the \( i \)th of the functions making up \( f_{t-k} \), \( R_{t-k} \) has \( E[f_{t-k, i} f_{t-k, j}] / (\sigma, \sigma) \) as element \((i,j)\). Substituting (16) and (17) into (14) gives

\[
a_0 = \bar{y} - \bar{r}^T a' \left\{ \bar{f}(y - \bar{r}^T a') + (\sigma^2 R_{t-k} + \bar{f}^T) a' = \sigma^2 r_{t-k} + \bar{f} \right\}
\]

From the second equation of (18),
\[ a' = R_{t-k}^{-1} r_{t-k} \]  \hspace{1cm} (19)

In practice, \( \bar{y}, R_{t-k}, \) and \( r_{t-k} \) are replaced by unbiased estimates based on the output record. The simplest special case of this predictor is when \( f_{t-k} \) consists merely of \( y_{t-k} \), giving \( R_{t-k} = I, r_{t-k} = r_k, a_1 = k, a_0 = \bar{y}(1-r_k) \) with \( r_k \) the output autocorrelation at lag \( k \), so using sample mean and lag-\( k \) autocorrelation of the output,

\[ \hat{y}_t = \hat{\bar{y}} + \hat{r}_k (y_{t-k} - \hat{\bar{y}}) \]  \hspace{1cm} (20)

This predictor takes an initial observation-free estimate \( \hat{\bar{y}} \) and corrects it by a term proportional to observed error \( y_{t-k} - \hat{\bar{y}} \) and weighted by the correlation between \( y_{t-k} \) and \( y_t \). By contrast, the estimator \( \hat{y}_t = \bar{y} \) implicit in \( R^2 \) makes no correction and effectively takes the autocorrelation as zero. The output MSE corresponding to (20) is

\[ E[(y_t - \hat{y}_t)^2] = E[(y_t - \bar{y} - r_k(y_{t-k} - \bar{y}))^2] = (1 - r_k^2)\sigma^2 \]  \hspace{1cm} (21)

As \( E[(y_t - \bar{y})^2] = \sigma^2 \) by definition, if \( R^2_k \) denotes the statistic comparing the model MSE with that of the \( k \)-step predictor (20), its theoretical value is

\[ R^2_k = 1 - \frac{E[(y_t - \hat{y}_t)^2]}{E[(y_t - \bar{y})^2]} = 1 - \frac{1 - R^2}{1 - r_k^2} = R^2 - \frac{r_k^2(1 - R^2)}{1 - r_k^2} < R^2 \]  \hspace{1cm} (22)

### 3.2 Practical examples

The differences between \( R^2 \) and the measures of model fit based on MMSE \( k \)-step output-to-output predictors will be illustrated on roughly 30-year records of daily flow in contrasting catchments: at Tinderry below 506 km\(^2\) of the Googong catchment and at Gingera below 148 km\(^2\) of the Cotter catchment, both near Canberra, ACT, Australia. The former catchment is largely grazing or low-intensity forestry with some farm dams, and the latter is higher, steeper and near-natural.

Figures 5 and 6 show the sample autocorrelation functions of these flow records. The autocorrelation is strong (\( > 0.4 \)) at lags up to 2 days for Tinderry and 15 or so days for Gingera. The more prominent slow flow component of the latter, due to its greater natural

**Figure 5.** Sample autocorrelation function of daily flow records at Tinderry

**Figure 6.** Sample autocorrelation function of daily flow records at Gingera
storage capacity, accounts for the difference. Figures 7 and 8 display the resulting measure of fit $R^2_k$ given by (22) as a function of prediction interval $k$, for a range of values of $R^2$.

$R^2_k$ tests how much of the output variation left by a very simple output-to-output predictor is removed by the model. The discrepancy between $R^2_k$ and $R^2$ is quite small for the Tinderry record except for $k = 1$ or 2, but for the Gingera record it is large at $k = 5$ and notable even at $k = 15$. Thus for some catchments crude flow predictors of the form (20) explain enough of the flow variation to show a given model much less favourably than does $R^2$.

4 CONCLUSIONS

Uncertainties in the instantaneous unit hydrograph due to ignorance of detailed rainfall history and rainfall-flow dead time, and conversion from a linear-in-parameters to a quick-flow slow-flow model, have been examined. They may be large unless the flow sampling interval is very short compared with the response time constants.

A family of model-performance measures analogous to Nash-Sutcliffe efficiency but based on simple $k$-step predictors of flow has been presented. They offer an alternative to $R^2$ whenever comparison with a predictor using earlier flow, but not rainfall, is permitted.

REFERENCES


Robust estimation of the total unit hydrograph

F.T. Andrews\textsuperscript{a}, B.F.W. Croke\textsuperscript{a}, and K.W. Jeane\textsuperscript{a}

\textsuperscript{a}Fenner School of Environment and Society, and Department of Mathematics, The Australian National University (felix.andrews@anu.edu.au, barry.croke@anu.edu.au, kevin.jeanes@anu.edu.au)

Abstract: This paper investigates three techniques for estimating the parameter values for the unit hydrograph: full model optimisation (the CWI and CMD versions of IHACRES were used), simple rainfall scaling approaches for estimating effective rainfall and an inverse filtering approach. The approaches are tested using synthetic streamflow data generated by the catchment moisture deficit (CMD) version of the IHACRES model, based on observed rainfall data for Cotter catchment (near Canberra, Australia) and the Singkarak catchment (Sumatra, Indonesia). The impact of rainfall and streamflow errors are considered by reducing the raingauge density and introducing errors in the rating curve. The inverse filtering approach performed well for the synthetic streamflow derived from the rainfall data for the Cotter catchment. For the Singkarak catchment data, the differences between the approaches was small, due possibly to either the higher frequency of rainfall in that area, or to the lower correlation between the rainfall gauge locations.

Keywords: effective rainfall; unit hydrograph; transfer functions; calibration

1 Introduction

Spatially-lumped, conceptual hydrological models are often based on two components: a Soil Moisture Accounting (SMA) component which generates “effective rainfall” and a routing component (based in the unit hydrograph) which uses this to generate outflow. In its original form, unit hydrograph (UH) theory addressed the impulse response of runoff, requiring the subtraction of the baseflow component from the observed streamflow. With the application of the transfer function approach; for example, the IHACRES [Jakeman et al., 1990] and Data-Based Mechanistic (DBM) models [Young, 2003], the UH was expanded to include the total impulse response of the catchment (i.e. including the baseflow component). Investigation of the unit hydrograph is of interest for studies looking at changes in catchment response (e.g. due to climate or land use change, or variations in the rainfall intensity between events).

In calibrating and assessing models using observed streamflow, the SMA and UH components are typically treated as a single unit. While calibrating all the parameters simultaneously will give a good fit to the observed flow, the unobserved state variables are not necessarily well estimated. Thus, in some contexts it can be useful to attempt to decouple the two components. This is an example of the use of theory-based hydrological signatures suggested by Gupta et al. [2008]. Decoupling is achieved by estimating the UH from observed streamflow, using approximations and transfer function methods. This can improve the efficiency of model calibration, and can be used to generate detailed diagnostics to guide model development. Of course, one must be aware of the approximations and sensitivities inherent in these methods.
Optimal estimation of the parameters of the UH requires minimal uncertainty in the effective rainfall estimates. Typically, rainfall runoff models use an assumed functional form for estimating the effective rainfall based on estimates of climate inputs (areal rainfall and a measure of potential evaporation) as well as an indicator of the antecedent catchment soil moisture. This results in considerable uncertainty in the estimated effective rainfall due to uncertainty in the model inputs and error in the model structure. This, in turn, leads to considerable uncertainty in the UH parameter values and difficulty in assessing the deficiencies in the UH model structure.

One approach to bypassing structural errors in the SMA model is to use non-parametric estimation of effective rainfall [e.g. Young, 2003]. Indeed, Kirchner [2009] has demonstrated non-parametric estimation of a complete hydrological model, although that approach is quite different in that it does not distinguish effective rainfall, or the unit hydrograph, but rather considers a storage-discharge relationship. An alternative approach has been developed using an iterative technique where the effective rainfall is estimated from the observed streamflow using an assumed UH, with the constraint that the effective rainfall is less than the observed rainfall (an inverse filtering approach). The effective rainfall sequence is then used to update the UH parameters. The process is repeated until the UH parameter values stabilise. This approach uses the minimum amount of information from the precipitation time series and doesn’t require the estimation of a measure of catchment moisture state. Thus the inverse filtering approach described in this paper limits the impact of uncertainty in the precipitation estimates. The result is an improved estimate of the UH and ability to evaluate the adequacy of the UH model structure.

This study explores three methods for estimating the unit hydrograph: full model calibration; simple rainfall scaling methods; and an inverse filtering approach. The synthetic data was generated using the CMD version of IHACRES, and noise added to the rainfall and streamflow data by varying the rain gauge density and introducing a random perturbation to the streamflow rating curve. Observed rainfall data for 2 catchments with very different climates (Cotter catchment near Canberra, Australia, and the Singkarak catchment on Sumatra, Indonesia) were used to generate the syntethic streamflow data.

2 The general Unit Hydrograph model

The Unit Hydrograph routing function translates a time series of effective rainfall into a corresponding time series of streamflow. As such it influences the size of flow peaks and the duration of flow after a peak. A Unit Hydrograph recession curve shows the streamflow response resulting from an impulse of 1 unit of input effective rainfall.

The transfer function used here is an ARMAX-like (auto-regressive, moving average, with exogenous inputs) model, where the input series is denoted $U$ and the output $Q$:

$$Q[t] = a_1 Q[t-1] + \ldots + a_n Q[t-n] + b_0 U[t-\delta] + \ldots + b_m U[t-m-\delta]$$ (1)

The order is denoted $(n,m)$, with delay $\delta$. The number of parameters is $n + m + 1$.

In hydrology it is natural to decompose the transfer function into a system of exponentially receding components, which may be in a parallel and/or series configuration. Each component is defined by a recession rate and peak response, or equivalently, a time constant $\tau$ and fractional volume $v$. In this form the parameters are physically interpretable. For further discussion, see for example Young [2003].

3 Unit Hydrograph estimation methods

We tested three types of methods for estimating the Unit Hydrograph from rainfall and streamflow data: full model parameter optimisation, simple rainfall scaling methods, and the inverse filtering approach. Only the first type are normal hydrological models; the
other methods make use of observed streamflow data and so cannot be used for prediction. Rather, they are methods for estimating the unit hydrograph (response characteristics of the catchment). The direct estimation technique developed by Croke [2006] has not been explored here as that approach is best suited to ephemeral streamflow regimes where flow events are well-separated. The direct estimation technique doesn’t require rainfall data, and so higher resolution streamflow data that is typically available can be used. As daily rainfall data has been used to generate synthetic streamflow data, higher resolution streamflow data is not available for this study.

In data checking and model diagnostic contexts, it is often adequate to estimate the response characteristics alone. This can reveal changes in catchment response over time, or allow evaluation of a SMA model performance independently from streamflow routing. Alternatively, if a full hydrological model is required, one could fit a SMA model while keeping the UH fixed.

### 3.1 Full model parameter optimisation

This approach involves joint estimation of the Soil Moisture Accounting model parameters and the Unit Hydrograph model parameters (in the hydrologically-meaningful exponential components form). This is a complex, but general, non-linear optimisation problem.

In this study we used the Shuffled Complex Evolution (SCE) algorithm [Duan et al., 1992], a popular calibration method in the hydrology literature. The objective function was a linear combination of the $R^2$ coefficient of efficiency [Nash and Sutcliffe, 1970] of square-root transformed data, and the relative bias.

Two models were used with this approach: the “true” model, which is the IHaCRES CMD (Catchment Moisture Deficit) SMA, defined below; and an alternative model with a very different structure, the IHaCRES CWI (Catchment Wetness Index) model of Jakeman and Hornberger [1993].

### 3.2 Simple rainfall scaling methods

Simple rainfall scaling methods involve a simple transformation of the rainfall data to estimate effective rainfall, which is then used to fit the Unit Hydrograph with specialised transfer function methods – in this case the Simple Refined Instrumental Variable (SRIV) algorithm [Young, 2008]. Such approaches have been recommended for estimating the complexity and parameter values of the UH, applicable to simple models such as IHaCRES-type models [Jakeman et al., 1990]. This approach is also typical of the early stages of Data-Based Mechanistic modelling [Young, 2003].

We tested two methods: firstly a runoff ratio model, where rainfall was scaled by the runoff coefficient (flow as a fraction of rainfall) as calculated in a moving (triangular-weighted) window of 60 days. For added flexibility, a threshold parameter was included, and no effective rainfall was generated when the runoff ratio fell below the threshold. Secondly we tested a typical model structure used in Data-Based Mechanistic (DBM) modelling of catchment hydrology [Young, 2003]. In this case rainfall is scaled by the observed streamflow on the same time step, raised to a power.

The single parameter in each case was selected (from 20 regular samples) according to the objective function described above. For the runoff ratio model, the threshold parameter was varied between 0 and 0.3. For the DBM model, the power parameter was varied between 0 and 1. In both of these methods, a constant mass balance term was also included in order to set total effective rainfall equal to total observed streamflow.
3.3 Inverse filtering approach

Inverse filtering refers to a method of deconvolution – estimating effective rainfall time series $U$ from an observed time series $Q$ of streamflow volumes. The inverse filter for the standard model can be easily derived from the simulation equation and written as:

$$U[t] = (Q[t] - \sum_{i=1}^{n} a_i Q[t - i] - \sum_{j=1}^{m} b_j U[t - j])/b_0 \quad (2)$$

Importantly, the effective rainfall $U$ is also constrained to be less than the observed rainfall $P$ on the corresponding time step (actually a 10% excess is allowed).

The estimation method is an iterative algorithm. Effective rainfall $U$ is estimated from $Q$ by inverse filtering, then this is passed on to the Simple Refined Instrumental Variable (SRIV) algorithm [Young, 2008] to re-fit the unit hydrograph parameters. This process is repeated until the sum of total differences between $U$ in successive iterations is less than 1/1000th of total $Q$.

Initialisation could be done with either an initial parameter set or an initial estimate of $U$. We chose to start with $U$ estimated as the observed rainfall $P$, scaled by the runoff ratio in a moving window, i.e. starting with the runoff ratio model defined above.

4 Test Simulations

For this study, we chose a simple model as the “true” model, i.e. the model used to simulate the test datasets. For the Soil Moisture Accounting, we use the IHACRES-type CMD (Catchment Moisture Deficit) model of Croke and Jakeman [2004]. For the unit hydrograph we used a transfer function with two exponential stores in parallel ($n = 2, m = 1$ in the notation above, with three parameters). The chosen parameter values were: $d = 200, f = 0.7, e = 0.2, \tau_s = 30, \tau_q = 2, v_s = 0.5$.

For simulation inputs, we use rain gauge and temperature data from two contrasting catchments. One is in the wet tropics and the other is a relatively dry temperate site. Note, however, that the details of these catchments are not important because we use only simulated outputs, and make no use of observed streamflow.

The Cotter catchment is an inland catchment in the Australian Capital Territory, Australia. Rainfall data from six rain gauges in and around the catchment were used in this study (070310, 070316, 570915, 570946, 570948, 570958). The “true” areal rainfall was constructed as a simple average of these.

The Singkarak lake catchment is an inland, mountainous catchment in West Sumatra, Indonesia. Rainfall data from six rain gauges in and around the catchment were used in this study (100001, 100012, 100013, 100027, 100033, 100035). This is an extremely wet series, with rainfall on about 75% of days, and a mean daily rainfall of 6 mm.

5 Error simulation

The UH estimation methods described above were tested with simulated data to assess their robustness to various types of error.

The accuracy of areal rainfall estimation is fundamentally limited by the spatial coverage of rain gauges relative to the spatial variability of rainfall. This type of error was simulated by taking a sub-sample of the available rain gauges. Samples of $n = 5$ and $n = 3$ gauges were taken (out of a total of 6), and 6 replicates were run for each of these levels. The results are shown in Figure 1.
Figure 1: Simulated areal rainfall series with error, plotted against the “true” areal rainfall. Each group of panels refers to the number of rain gauges in the sub-sample (out of a total of 6 gauges). Regression lines (dashed) are shown along with the 1:1 reference line.

Streamflow gauging is subject to error, particularly in the rating curve, the transformation from stage height to flow volume. This is likely to be a systematic error, and to be more severe at high flow levels Croke [2009]. This type of error was simulated by a non-linear transformation applied to the true streamflow $Q$. Specifically, we define four random scaling factors along the range of $Q$, interpolated by a cubic spline, and use this as a multiplicative transformation function. The scaling factors were chosen to be within ±5% ($d = 0.05$), ±15% ($d = 0.15$) or ±30% ($d = 0.3$). 5 replicates were run in each case. This approach is similar to the error-in-rating-curve simulation of Croke [2009], but is less structured.

The effect of mis-specification of the UH transfer function is represented by the use of different model orders as defined in Equation 1. The “true” UH is a second-order transfer function, and we tried each of the calibration methods with the correct form as well as a simpler first-order transfer function.

6 Results and Discussion

Figure 2: Example of fitted models based on one variant of the Cotter dataset: the streamflow series with error type $d = 0.05$, and rainfall series with error type $n = 5$; taking the first replicate in each case.

The aim of this study was to investigate the performance of methods for fitting the unit hydrograph. As the unit hydrograph is convolved with the modelled effective rainfall time series to produce streamflow, effective rainfall is also of interest.

To aid comprehension, results for a single dataset variant are shown in Figure 2. It shows
Figure 3: Results for the Cotter dataset, showing accuracy of the modelled unit hydrograph recession curve, and the modelled effective rainfall time series, when fitting by each of the methods shown. In both cases, low values indicate a good fit. The labels $n = 6$, $n = 5$, $n = 3$ indicate the number of rain gauges used, with fewer gauges implying a larger error. The right-most column gives results from fitting a first-order unit hydrograph rather than the true second-order form. Results include 6 replicates of each level of error in rainfall, and for each of these, 5 replicates of each level of error in streamflow.

Figure 4: Results for the Singkarak dataset, showing accuracy of the modelled unit hydrograph recession curve, and the modelled effective rainfall time series, when fitting by each of the methods shown. Interpretation is the same as Figure 3.
a section of the fitted streamflow time series, and the fitted unit hydrograph recession curve, from 3 out of the 5 methods. It is clear that, while large streamflow peaks are fitted reasonably well, the recession curve is far too rapid in the runoff ratio method, and a little too rapid in the slow component for the DBM method. The CMD result, which is the “true” model fitted by full optimisation, does reproduce the unit hydrograph very well in this case.

We chose to assess the results from the simulation study in terms of the absolute error in the modelled UH recession curve, and the R Squared [Nash and Sutcliffe, 1970] of square-root transformed effective rainfall. The latter was also calculated from fitting with a reduced-order (first order) unit hydrograph.

The results from fitting with each of the 5 methods, as the level of error in rainfall input data increases, are shown in Figures 3 and 4. These figures show the distribution of results over the three levels of streamflow error (5%, 15%, 30%) and the replications of both rainfall and streamflow error simulation. It was found that the level of error applied to streamflow data made little difference to the results, and that the results were dominated by the level of error in rainfall; for that reason the individual effect of streamflow error is not presented. The effect was noticeable only when the rainfall error was low, and particularly for the inverse method.

The results from the Cotter dataset in Figure 3 show that the inverse filtering method generally performed better than the other 4 methods tested. Full optimisation of the true model (CMD) produced good results in terms of effective rainfall, although it was susceptible to errors in the rainfall data. However, the true model performed as poorly as the alternative (CWI) model in reproducing the UH. When a first-order UH transfer function was specified (i.e. one exponential component rather than the true two components), the inverse method gave much worse results, as expected. However, it still gave a better fit to effective rainfall than all methods except the optimised true model. These results confirm that the inverse filtering approach is minimally affected by errors in the rainfall data, and independent of the choice of SMA model.

The results are very different on the Singkarak dataset, shown in Figure 4. All methods were strongly constrained by the level of error in rainfall. For the low error cases (n = 6, n = 5), the best results were generally from the CMD (true) model and the DBM method. Inverse filtering method did less well. This may be due to either the higher frequency of rainfall in that area, or to the lower correlation between the rainfall gauge locations.

As expected, the performance of full model optimisation was strongly dependent on the choice of SMA model: the true SMA model (CMD) gave much better results than optimisation of an incorrect SMA model (CWI). Of the simple rainfall scaling methods tested, the DBM method consistently gave better results than the runoff ratio method.

7 Conclusions and further work

The inverse filtering method performed well at reconstructing the true unit hydrograph and effective rainfall in the simulations in this paper, except in the case of high frequency rainfall inputs. However, these results are based on reconstructing simulations with the correct model structure of the unit hydrograph transfer function, which in this study was a second-order (two stores) form. The results when fitting a first-order transfer function were much less good. This highlights the need to try different model structures.

Further work should investigate the performance and robustness of methods with different choices for the true unit hydrograph: possibly three or four stores in parallel and/or series, or perhaps with variable partitioning between quick and slow stores, depending on rainfall intensity. Another option is the simulation of groundwater linkages.
Furthermore the effect of the true SMA model should also be investigated. The CMD model used in this study is relatively simple, and other process representations may influence the results to some extent.

**Computational details**

The results in this paper were obtained using R 2.11.0 [R Development Core Team, 2010] with the packages **hydromad** 0.8–3, **lattice** 0.18–5 [Sarkar, 2008] and **zoo** 1.7–0 [Zeileis and Grothendieck, 2005]. R itself and all packages used are (or will be) available from CRAN at [http://CRAN.R-project.org/](http://CRAN.R-project.org/). The R code used to produce the results in this paper is available from [http://www.nfrac.org/felix/papers/2010/iemss/](http://www.nfrac.org/felix/papers/2010/iemss/).

**References**


The Use of Entropy as a Model Diagnostic in Rainfall-Runoff Modelling

I.G. Pechlivanidis\textsuperscript{a}, B. Jackson\textsuperscript{a}, and H. McMillan\textsuperscript{b}

\textsuperscript{a}School of Geology, Environment and Earth Sciences, Victoria University of Wellington, Kelburn, Wellington, New Zealand (ilias.pechlivanidis/bethanna.jackson@vuw.ac.nz)

\textsuperscript{b}National Institute for Water and Atmospheric Research (NIWA), Riccarton, Christchurch, New Zealand (h.mcmillan@niwa.co.nz)

Abstract: Recent papers have called for the development of robust model diagnostics (in addition to traditional “measures of fit”) that provide insights on where model structural components and/or data may be insufficient. The potential of entropy measures to provide these in hydrology has not been adequately explored. Further, flow duration (FD) curves provide a useful visual diagnostic of catchment response, but attempts to quantify the fit of modelled versus observed FD curves to date have relied on using time series measures of fit. We note that Shannon entropy of flow is strongly related to the FD relationship, so suggest it provides a more appropriate quantitative measure of fit. This paper presents initial results from a study calibrating two rainfall-runoff models to 4 years of hourly data from the Mahurangi catchment, NZ. Kling-Gupta efficiency (KGE), Nash-Sutcliffe efficiency (NSE) and two entropy measures were considered. When assessed using a range of model diagnostics, KGE was overall the single best measure, outperforming NSE at all times. Entropy outperformed KGE over particular hydrograph sections, and we show performance may improve further with careful choice of discretisation. We demonstrate entropy’s strong relationship to FD and interrogate the performance of entropy measures in the presence of timing and bias errors. As entropy is insensitive to timing errors but very sensitive to most other errors (in sharp contrast to, e.g., the NSE measure) it potentially provides a useful diagnostic of the types of error present in combination with other OFs.

Keywords: Shannon entropy; Model diagnostics; Model uncertainty; Rainfall-runoff modelling.

1. INTRODUCTION

Traditional “measures of fit” are commonly used in hydrological modelling to provide an objective assessment of the “closeness” between simulated and observed hydrological observations (e.g. streamflow). Different measures emphasise different systematic and/or dynamic behaviours within the hydrological system; hence a robust assessment of model performance using single measures is difficult [Krause et al., 2005; Schaefli and Gupta, 2007]. Recent papers have highlighted the importance of moving from model calibration to diagnostic model evaluation, which aims to: 1) determine the information contained in the data and in the model, 2) examine the extent to which a model can be reconciled with observations, and 3) point towards the aspects of the model (or data limitations) that need improvement [Gupta et al., 2008].

Information theoretic entropy-based measures provide a promising avenue to allow us to better identify where information is present and/or conflicting [Singh, 2000]. If placed within a hydrologically relevant context, these measures may assist in the generation of more robust modelling frameworks allowing us to better diagnose model/data/hypotheses inconsistencies [Jackson et al., 2010]. Therefore, the main objective of the current study is
to investigate the potential of entropy-based measures as objective functions and as model
diagnostics in hydrological modelling. Entropy-based statistics are introduced in Section 2,
where we also highlight barriers to robust implementation in hydrological contexts. Section
4 describes the two rainfall-runoff model structures and the identification method followed
using observed records from the Mahurangi River catchment, New Zealand. Section 5
presents initial results consisting of statistical analysis based on observed data and
modelling results which use entropy as an objective function. The paper concludes with a
brief discussion on possible ways forward.

2. USE OF ENTROPY AS A MODEL DIAGNOSTIC

According to Schreiber [2000], information is equivalent to the removal of uncertainty;
hence uncertainty and informational entropy are in some senses identical. In recent years,
entropy-based statistics have been applied to several hydrological problems [e.g. Maruyama
et al., 2005; Hejaki et al., 2008; Ruddell and Kumar, 2009]; however their potential to be
used as objective functions (OFs), or better as model diagnostics, is largely unexplored
(exceptions are the papers of Amorocho and Espildora [1973] and Chapman [1986]). This
paper is not concerned with repeating what is already discussed in previous reviews (see
Singh [2000], Jackson et al. [2010]), but rather on presenting preliminary results and
insights on the potential use of entropy as an objective function and model diagnostic test.

2.1 Shannon entropy

Entropy is variably described; examples include “a measure of the amount of chaos” or “of
the lack of information about the system” [Koutsoyiannis, 2005]. The Shannon entropy
[Shannon, 1948] of a discrete random variable \( X \) with \( N \) possible outcomes is given by:

\[
H_X = - \sum_{i=1}^{N} p(x_i) \cdot \log_{\text{base}} p(x_i)
\]  

(1)

where \( p(x_i) \) is the probability of occurrence of outcome \( x_i \) and \( \text{base} \) is the base of
the logarithm used (entropy is a unit of binary digits, bits, when \( \text{base}=2 \)). Shannon [1948]
defined entropy as the average number of bits (assuming \( \text{base}=2 \)) needed to optimally
encode independent draws of \( X \) following a probability distribution \( p(x) \). A low value of
entropy indicates a high degree of structure and a low uncertainty. It can be easily shown
that with complete information entropy equals 0, otherwise it is greater than 0. If no
information is available then entropy will reach its maximum equal to \( \log_{\text{base}}(N) \). When
using entropy as an OF or diagnostic, we suggest normalizing Equation 1 by dividing
through by \( \log_{\text{base}}(N) \). This normalised entropy remains 0 with complete information
/ maximum order and takes a maximum value of 1 with no structure/maximum disorder.
Note also that, if applied to flow data, this probabilistic measure is closely related to the FD
curve; Shannon entropy becomes a quantification of the information (and hence shape) of
the histogram or discretised distribution of flow, i.e. flow duration.

Although a continuous analog to the Shannon entropy is available, we rarely possess the
analytical form of our variable \( X \)’s probability distribution so generally work with the
discrete form given in Equation 1. Unless \( X \) is ordinal, a number of discrete bins must be
specified, with accompanying ranges. In this case, estimation of the probability density and
its associated entropy is influenced by the resolution of this data, the number of bins, and
the locations of divisions between these bins. The introduction of arbitrary partitions could
result in “edge effects”. According to Ruddell and Kumar [2009], with too few/many
partitions, “edge effects” become severe and entropy estimates are positively biased,
resulting in underestimated mutual information and transfer entropy.

Different approaches can be used to discretise the data set to probability “bins”. These
include function fitting [Knuth, 2005], kernel estimation [Nichols, 2006], and binning with
fixed mass or fixed interval partitions [Ruddell and Kumar, 2009]. Fixed interval partitions
are usually applied since the approach is simple and computationally efficient [Ruddell and Kumar, 2009]. These fixed interval partitions are used in this preliminary study.

### 2.2 Limitations of Shannon entropy when used as an OF

Although Shannon entropy is a quantification of the amount of information within a dataset, its static probabilistic nature cannot capture the temporal variability of information. It therefore shows no sensitivity in time. In addition, this probability-based measure does not depend on the range of the data, so mass balance error could be introduced. The following example illustrates the lack of sensitivity of entropy in time and mass balance. Three streamflow hydrographs ($Q_t$, $Q_t/2$, and $Q_{t+DT}$) are generated and presented in Figure 1. $Q_t/2$ is the streamflow data of $Q_t$ divided by 2 at each time step, while $Q_{t+DT}$ is the $Q_t$ data lagged by $DT$ hours. Entropy considers the probability that a value (or range of values) occurs within the data series, and does not take into account their location within the time series. Therefore, the three hydrographs in Figure 1 have identical Shannon entropy.

The lack of sensitivity of entropy to mass balance was demonstrated above. However, mass conservation between observed and simulated streamflow is important in many hydrological applications. Therefore, time series analysis based on the Shannon entropy may need to introduce methodologies that can (explicitly or implicitly) account for mass balance. In this preliminary study we aimed to create a performance measure sensitive to mass balance by linearly fixing the bins between the maximum observed or modelled data range (maximum (simulated, observed) - minimum (simulated, observed)). To distinguish the binning methods from this point forward, we call **scaled**, $H_s^F$, the Shannon entropy using fixed bins for both simulated and observed data (this attempts to conserve mass), and **unscaled**, $H_u^F$, the Shannon entropy using different bin ranges for simulated and observed data based on their individual specific maximum range (this measure ignores mass conservation).

### 3. CATCHMENT DESCRIPTION

The analysis is based on observed data from the experimental Mahurangi River (Figure 2) in northern New Zealand, which drains 46.6 km$^2$ of steep hills and gently rolling lowlands. A network of 28 flowgauges and 13 raingauges has been installed, collecting records at 15 minutes intervals as part of the MARVEX project [Woods, 2004]. The catchment experiences a warm humid climate (frosts are rare and snow and ice are unknown), with mean annual rainfall and evaporation of 1,600, and 1,310 mm respectively. The catchment elevation ranges from sea level to 300 m. Most of the soils in the catchment are clay loams, no more than a metre deep, while much of the lowland area is used for grazing. Plantation forestry occupies most of the hills in the south, and a mixture of native forest, scrub and grazing occurs on the hills in the north. Further details are given in Woods [2004].

Historical hourly rainfall, streamflow and potential evapotranspiration data were provided by the National Institute of Water and Atmospheric Research, New Zealand, for the period...
1998-2001. The arithmetic average of the 13 raingauge records was used as the mean areal precipitation; this was homogeneously distributed over the catchment. Only the flowgauge at the outlet of the catchment was considered in the present study.

4. RAINFALL-RUNOFF MODELLING

4.1 Model description

Two rainfall-runoff model structures within the Rainfall-Runoff Modelling Toolkit [Wagener et al., 2004] were used to describe the hydrological behaviour of the catchment. The first model is the Probability Distributed Moisture (PDM) model [Moore, 2007]. This allows for a varying distribution of storage capacity over the catchment (Figure 3), described by a Pareto distribution according to Eq. 2.

\[ F(C) = 1 - \left( 1 - \frac{C}{C_{\text{max}}} \right)^b \]  

\( C \) is the storage capacity in the catchment, \( C_{\text{max}} \) is the maximum capacity at any point in the catchment, and the parameter \( b \) (\( \cdot \)) controls the spatial variability of storage capacity over the catchment. Within each time step, the soil moisture storage is depleted by evaporation as a linear function of the potential rate and the volume in storage, and augmented by rainfall. Effective rainfall is then equal to the soil moisture excess.

![Figure 3. Structure of the Probability Distributed Moisture model.](image)

The second model is a simple bucket model which could be described by setting parameter \( b \) in the PDM equal to 0. Both models were combined with a routing component consisting of two linear reservoirs in parallel, representing the quick and slow response of the system. This model component has three parameters: a residence time for each reservoir, \( K_q \) and \( K_s \) (hours) and \( q \), the proportion of total effective rainfall going to the fast response reservoir. Streamflow is finally delayed by a parameter \( T \) (hours) to adjust time to peak response.

4.2 Model identification method

A Monte Carlo uniform random search procedure was used to explore the feasible parameter space and to investigate parameter identifiability (30,000 samples). The first year (1998) was used as a warm-up period, the next two years for calibration (1999-2000) and the final one year for independent evaluation (2001). Both models were calibrated using streamflow data at the catchment outlet using four objective functions (OFs): the Nash-
The Sutcliffe Efficiency, NSE (Eq. 3) [Nash and Sutcliffe, 1970], the recently proposed Kling and Gupta Efficiency, KGE (Eq. 4) [Gupta et al., 2009], the absolute difference in unscaled entropy between simulated and observed series, US-Ent (Eq. 5), and our new entropy OF; the maximum of the scaled and unscaled Shannon entropy difference, SUS-Ent (Eq. 6).

\[
NS_E = 1 - \frac{\sum_{i=1}^{n} (Q_{obs,i} - Q_{sim,i})^2}{\sum_{i=1}^{n} (Q_{obs,i} - Q_{obs})^2}
\]

\[
KGE = 1 - \sqrt{cc - 1) + (\alpha - 1)^2 + (\beta - 1)^2}
\]

\[
US - Ent = \text{abs} (H_{U, sim} - H_{U, obs})
\]

\[
SUS - Ent = \max [\text{abs} (H_{U, sim} - H_{U, obs}), \text{abs} (H_{S, sim} - H_{S, obs})]
\]

\(Q_{sim}\) is the calculated flow using the parameter set \(\theta\), \(Q_{obs}\) is the observed flow, \(n\) is the length of the time series, \(cc\) is the linear correlation coefficient between \(Q_{obs}\) and \(Q_{sim}\), \(\alpha\) is a measure of variability in the data values (equal to the standard deviation of \(Q_{sim}\) over the standard deviation of \(Q_{obs}\)), and \(\beta\) is equal to the mean of \(Q_{sim}\) over the mean of \(Q_{obs}\). See Gupta et al. [2009] for further details of the KGE and its components. As explained earlier, US-Ent is sensitive to the shape of the time series, but not scale (given the example in Section 2.2). SUS-Ent has been introduced as a trade-off between shape (and hence information) and scale conservation, in a first attempt to address the possible mass balance issue highlighted in Section 2.2. The simulated runoff using the two entropy measures is insensitive to timing errors and hence is completely insensitive to the final routing delay parameter. To overcome this, after the US-Ent and SUS-Ent calibrations were performed, this routing parameter \(T\) was individually adjusted through manual calibration. This decoupling of time sensitive versus time insensitive parameters can be seen as both an advantage (reducing problem dimension) and a potential disadvantage (it may cause issues in more complicated models where time sensitive and time insensitive parameters are strongly inter-dependent; we leave this question for future work).

### 5. RESULTS

Table 1 summarises model performance over both the calibration and validation period, considering NSE, KGE, US-Ent and SUS-Ent. Overall, there was no significant difference in performance of the bucket versus the PDM model when considering NSE and KGE; however, the PDM can represent the information in the data distribution (as described by the entropy OFs) better than the bucket model. As expected, information in the flow is well conserved when entropy is used as an objective function. The optimum parameter set based on SUS-Ent derives only slightly lower NSE and KGE values, highlighting the potential of this information-based OF to represent the hydrograph properties. However, also as expected, NSE and KGE performance is further reduced when US-Ent is used, probably because this OF maximises information ignoring the effect of data scale.

Table 1. Model performance (calibration and validation) using the optimum parameter set (presented horizontally) for each OF. [ ] is used for the validation period.

<table>
<thead>
<tr>
<th></th>
<th>Bucket model</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NSE</td>
<td>KGE</td>
<td>US-Ent</td>
<td>SUS-Ent</td>
<td>NSE</td>
<td>KGE</td>
<td>US-Ent</td>
<td>SUS-Ent</td>
</tr>
<tr>
<td>NSE</td>
<td>0.83 [0.81]</td>
<td>0.79 [0.82]</td>
<td>0.024 [0.036]</td>
<td>0.070 [0.066]</td>
<td>0.84 [0.79]</td>
<td>0.68 [0.74]</td>
<td>0.087 [0.079]</td>
<td>0.087 [0.079]</td>
</tr>
<tr>
<td>KGE</td>
<td>0.80 [0.77]</td>
<td>0.89 [0.83]</td>
<td>0.077 [0.114]</td>
<td>0.077 [0.129]</td>
<td>0.84 [0.78]</td>
<td>0.89 [0.84]</td>
<td>0.026 [0.008]</td>
<td>0.029 [0.052]</td>
</tr>
<tr>
<td>US-Ent</td>
<td>0.79 [0.81]</td>
<td>0.79 [0.83]</td>
<td>0.001 [0.071]</td>
<td>0.056 [0.071]</td>
<td>0.73 [0.77]</td>
<td>0.64 [0.78]</td>
<td>0.001 [0.037]</td>
<td>0.024 [0.037]</td>
</tr>
<tr>
<td>SUS-Ent</td>
<td>0.81 [0.77]</td>
<td>0.81 [0.80]</td>
<td>0.005 [0.018]</td>
<td>0.014 [0.039]</td>
<td>0.81 [0.76]</td>
<td>0.70 [0.76]</td>
<td>0.003 [0.047]</td>
<td>0.003 [0.047]</td>
</tr>
</tbody>
</table>

A graphical illustration of the bucket model behaviour is presented in Figure 4 (the PDM model showed similar behaviour, so results are not presented to avoid repetition). Overall, the four OFs simulate flow which fits well the observed data. High flow values seem to be underestimated by the NSE and better represented by the KGE and the entropy-based
measures; however, fitting using US-Ent is poor during the recession. This appears to be mostly due to the low optimised $K_q$ value (2.4 hours); parameter identifiability plots suggested US-Ent was not capable of identifying this parameter as robustly as the other three measures could. This is almost certainly because of its insensitivity to range and mass.

![Figure 4](image)

Figure 4. Time series fits depicting simulated and observed runoff in the Mahurangi catchment using different OFs with the bucket model.

Models are further evaluated using the flow duration (FD) curve as a diagnostic measure (Figure 5). Results from the SUS-Ent calibration provide a very close representation of the FD curve, which supports our insight that entropy’s probabilistic derivation is closely related to FD. Low flow is not fitted well by this measure, but we believe this is an artefact of the simplistic linear binning technique we used in this preliminary study versus the logarithmic scales we are examining in Figure 5. Further work will explore logarithmic and other binning techniques. Visual analysis suggests that the KGE fits high and low flows well; however the fitting of medium flows is poor. NSE seems able to capture the medium flows better, however high and low flow values are not accurately represented.

![Figure 5](image)

Figure 5. Flow duration curves in the Mahurangi catchment using different OFs and rainfall-runoff models: (a) bucket, and (b) PDM.

Simulated runoff time series are further analysed in a diagnostic manner. Table 2 examines performance of each of the four OF model calibrations to seven static and dynamic components of the observed flow data. These include errors in time lag, FD curve (normalised error), mass balance, peak and mean runoff respectively, correlation between modelled and observed values, and the variability measure described in Gupta et al. [2009] (standard deviation of simulated flow divided by standard deviation of observed flow-optimal at 1). The values of each of these errors or deviations are written within each cell for both calibration and validation period, while the sensitivity of the OFs to the seven individual components is qualitatively presented with colour: light grey showing low sensitivity, dark grey representing high sensitivity.
Table 2. Evaluation using the 4 OFs. [ ] denotes performance within the validation period.

<table>
<thead>
<tr>
<th>Bucket model</th>
<th>Time lag (hours)</th>
<th>Shape-FDC (-)</th>
<th>Variability (%)</th>
<th>Mass balance (%)</th>
<th>Peak flow (mm)</th>
<th>Mean flow (x10^{-1} mm)</th>
<th>Correlation (-)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSE</td>
<td>1 [1]</td>
<td>0.43 [0.27]</td>
<td>0.82 [0.87]</td>
<td>-4.2 [-7.8]</td>
<td>4.89 [5.38]</td>
<td>0.03 [0.08]</td>
<td>0.91 [0.90]</td>
</tr>
<tr>
<td>KGE</td>
<td>0.10 [0.26]</td>
<td>0.30 [0.16]</td>
<td>1.07 [1.15]</td>
<td>0.32 [1.35]</td>
<td>2.45 [2.29]</td>
<td>0.01 [0.01]</td>
<td>0.91 [0.90]</td>
</tr>
<tr>
<td>US-Ent</td>
<td>10 [10]</td>
<td>0.53 [0.28]</td>
<td>0.84 [0.92]</td>
<td>-1.2 [-6.1]</td>
<td>3.40 [2.95]</td>
<td>0.01 [0.01]</td>
<td>0.89 [0.90]</td>
</tr>
<tr>
<td>SUS-Ent</td>
<td>3 [3]</td>
<td>0.35 [0.23]</td>
<td>1.03 [1.07]</td>
<td>15.8 [13.7]</td>
<td>3.58 [3.89]</td>
<td>0.13 [0.13]</td>
<td>0.90 [0.89]</td>
</tr>
<tr>
<td>PDM model</td>
<td>Time lag (hours)</td>
<td>Shape-FDC (-)</td>
<td>Variability (%)</td>
<td>Mass balance (%)</td>
<td>Peak flow (mm)</td>
<td>Mean flow (x10^{-1} mm)</td>
<td>Correlation (-)</td>
</tr>
<tr>
<td>NSE</td>
<td>1 [1]</td>
<td>0.50 [0.29]</td>
<td>0.81 [0.81]</td>
<td>18.8 [16.3]</td>
<td>4.41 [5.04]</td>
<td>0.13 [0.16]</td>
<td>0.92 [0.89]</td>
</tr>
<tr>
<td>KGE</td>
<td>0.10 [0.24]</td>
<td>0.21 [0.24]</td>
<td>1.02 [1.08]</td>
<td>8.2 [8.3]</td>
<td>3.29 [3.95]</td>
<td>0.07 [0.08]</td>
<td>0.92 [0.90]</td>
</tr>
<tr>
<td>US-Ent</td>
<td>9 [9]</td>
<td>0.36 [0.23]</td>
<td>1.03 [1.03]</td>
<td>27.3 [21.4]</td>
<td>1.73 [0.61]</td>
<td>0.22 [0.21]</td>
<td>0.88 [0.88]</td>
</tr>
<tr>
<td>SUS-Ent</td>
<td>10 [10]</td>
<td>0.25 [0.30]</td>
<td>1.02 [0.99]</td>
<td>26.4 [21.7]</td>
<td>2.85 [4.16]</td>
<td>0.21 [0.21]</td>
<td>0.91 [0.88]</td>
</tr>
</tbody>
</table>

As explained earlier, NSE and KGE, in contrast to US-Ent and SUS-Ent, are sensitive to time to peak. However, SUS-Ent introduces less error than the NSE in the FD curve, further supporting our theoretical observation that Shannon entropy is strongly related to FD. KGE also fits the FD curve well (0.30 and 0.21 normalised error using the bucket and PDM model respectively). Both entropy-based measures are sensitive to the variability and peak flows (1.03 and 3.58 mm difference respectively using the bucket model, and 1.02 and 2.85 mm respectively using the PDM model). SUS-Ent failed to conserve the mass balance (15.8 and 26.4 % error using the bucket and PDM model respectively). This indicates that although the current scaling approach is able to represent the data range, it is unable to conserve the mass. It is important to note that the linear correlation is equally well represented for all four OFs. NSE introduced the highest errors in variability and peak runoff, probably due to its over-emphasis on obtaining high linear correlation, as explained in Gupta et al. [2009]. In contrast, the models optimised with KGE represented variability, mean flow and linear correlation well.

6. CONCLUSIONS

The potential of information entropy measures as objective functions and the use of entropy with other measures as a diagnostic in rainfall-runoff modelling was demonstrated in the present study. Two rainfall-runoff models, a simple bucket model and the PDM model, were calibrated using four objective functions (NSE, KGE, and our unscaled and scaled versions of Shannon entropy). Output was evaluated in terms of both static and dynamic properties of the streamflow data (i.e. FD curve, variability, mass balance, time to peak, peak runoff, mean and correlation). As expected, results are consistent with the many previous studies showing that not all the static and dynamic properties of the flow series can be adequately captured by a single objective function. NSE is able to capture the time to peak and linear correlation with observed flow. However, it underestimates the variability and mean of flows, and produces (at least from the four objective functions tested) the largest error in peak runoff. The new KGE measure proposed recently by Gupta et al. [2009] seems able to overcome some limitations of NSE. Variability and mean flows are well matched, while keeping the linear correlation between modelled and observed high. The mass balance and peak runoff error is also decreased. These conclusions are consistent for both rainfall-runoff models on this one catchment; however studies on further catchments and with further model structures are required to generalise the conclusions.

Results support our theoretical observations that Shannon entropy is strongly related to the FD relationship, and we suggest that this is likely to provide a more robust measure of FD curve fit than those in current use. Entropy is insensitive to timing errors. This makes it dangerous as a stand-alone measure, but potentially provides a useful diagnostic whereby
(in combination with other measures) timing errors could be decoupled from other errors. It can be scaled to capture the range of data as we demonstrated; our preliminary approach is not sufficiently sensitive to mass errors. Further work will address more appropriate scaling methods. As entropy provides an OF measure with very different sensitivities and insensitivities to those currently in use, it also has obvious potential in combination with other measures in a more traditional multi-objective calibration framework.

REFERENCES

Jackson, B., B. Schaefli, H.V. Gupta and others, What is the information content of hydrologic data?, Advances in Water Resources, In Review.
Ruddell, B.L., and P. Kumar, Ecohydrologic process networks: 1. Identification, Water Resources Research, 45(W03419), 2009.
Propagating Data Uncertainty and Variability into Flow Predictions in Ungauged Basins

Andrew D. Gronewold a, Ibrahim M. Alameddine b

aUSEPA, National Exposure Research Laboratory, Research Triangle Park, North Carolina, USA (gronewold.andrew@epa.gov)
bDuke University, Nicholas School of the Environment, Durham, North Carolina, USA (ima4@duke.edu)

Abstract:
Explicitly acknowledging uncertainty and variability in model-based hydrological forecasts is a challenging task. Many basins are either ungauged, are undergoing rapid land use change, or are in regions expected to experience significant climate change. These factors, in addition to uncertainty in monitoring data and model structure, collectively contribute to discrepancies between model predictions and observations. Few hydrological modeling studies, however, routinely quantify data uncertainty. Furthermore, few studies compare model forecasts to observations while considering intrinsic uncertainty in the model itself. To bridge this research gap, we test a series of rainfall-runoff models within gauged and ungauged basins in Eastern North Carolina (US). In the model calibration phase, we propagate data uncertainty into model forecasts within a Bayesian framework. We then assess model suitability by examining the distribution of Bayesian posterior p-values (defined as the model-derived probability of a flow measurement as or more extreme than that observed). Evaluating model performance in this way helps identify potential sources of model bias and error, and clearly demonstrates the magnitude of those errors relative to the various potential sources of variability and uncertainty in the model forecast.

Keywords: rainfall-runoff model; ungauged basins; IHACRES; data variability; Bayesian

1 INTRODUCTION

Calibrating continuous rainfall-runoff models is a complicated and often arduous task. Notable challenges include selecting a suitable model (and model error) structure [Young and Beven, 1994; Wagener et al., 2001], identifying a robust set of model parameters [Beven, 1989; Beven and Freer, 2001] and quantifying potential correlation between (and uncertainty within) those parameters [Duan et al., 1992; Kuczera and Parent, 1998; Montanari and Brath, 2004]. Addressing these challenges becomes particularly important when considering how uncertainty and variability might impact (and be incorporated into) hydrological model-based studies.

For example, it is widely recognized that hydrological modeling tools are needed to forecast flows under future land use and climate change scenarios [Nandakumar and Mein, 1997; Anderson et al., 2006], or in basins which are ungauged and for which a model can not be calibrated directly [Seibert, 1999; Kokkonen et al., 2003]. Yet despite the broad range of research on the importance of uncertainty and change in hydrological modeling and water resources research [for further discussion, see Milly et al., 2008], we find that only recently has hydrological modeling research begun explicitly focusing on quantifying forcing data variability and propagating that variability into model parameter estimates and flow forecasts using robust (such as probabilistic and Bayesian) procedures [Vrugt et al., 2008; McMillan and Clark, 2009]. Furthermore, we
find that few modeling studies, if any, compare model forecasts to observations while considering intrinsic uncertainty in the model itself.

To bridge these research gaps, we apply a well-known conceptual rainfall-runoff (CRR) model to gauged basins along the Eastern coast of the United States (U.S.) within a Bayesian framework in order to forecast flows in nearby ungauged basins. We begin by developing parameter probability distributions using an ensemble modeling approach, and then apply the calibrated CRR models to an ungauged basin using recently obtained field data by first assuming that the field data is deterministic (i.e. certain) and then, for comparison, allowing for uncertainty and variability in the forcing data. We apply the derived parameter distributions to generate probabilistic flow forecasts in the ungauged basin, and assess the suitability of the two different assumptions regarding uncertainty in forcing data by comparing the forecasts to field observations using the distribution of Bayesian posterior $p$-values.

2 Model and Data

2.1 Model

To address the goals of our study, we apply the IHACRES model, a well-known [see, for example Dye and Croke, 2003; Croke and Jakeman, 2004] version of the more general class of data-based mechanistic (DBM) rainfall-runoff models [Young and Beven, 1994] to coastal watersheds in the Eastern U.S. The IHACRES rainfall-runoff model has been described extensively in previous works, including those providing an introduction to DBM rainfall-runoff models [Whitehead et al., 1979; Jakeman et al., 1990] as well as those which describe and apply the IHACRES graphical user interface software package [Littlewood et al., 1997; Jakeman and Letcher, 2003; Kokkonen et al., 2003; Anderson et al., 2006] and its recent developments [Croke and Jakeman, 2004; Croke et al., 2006]. We provide a brief description of the model here (and in the Appendix) for reference, and direct readers interested in a more detailed description of IHACRES to these earlier works.

The IHACRES model is divided into two components. The first is a nonlinear loss module which uses a measure of evaporation (such as temperature $t$ or pan evaporation) to translate incident rainfall ($r_k$, in units of mm) at time $k$ into effective rainfall ($u_k$, also in mm). The second component is a linear unit hydrograph-based module which translates effective rainfall ($u_k$) into streamflow ($x_k$). The IHACRES model (like many other CRR models) can divide flow into a “quick” and “slow” component. In initial attempts to calibrate the IHACRES model, however, we found that representing flow through a single flow path provided as good or better model performance than representing flow through two parallel flow paths. Furthermore, we do...
not expect the catchments in our study area (as described in the following section) to generate a significant base flow, and implement a version of the IHACRES model with only four parameters; $c$ (a mass balance parameter, in l/mm, often described as an index of watershed wetness capacity), $f$ (a temperature modulation parameter, in l/deg C), $\tau_w$ (the time constant of wetness decline, in days), and $\tau_f$ (the flow response time constant, in days).

2.2 Data

The watersheds selected for this study drain into some of the most sensitive coastal embayments in the world, including the Chesapeake Bay (VA) and the Neuse River Estuary (NC), which collectively host a wide range of both natural resources and recreational and commercial uses. Unfortunately, water quality in these embayments is declining due, in part, to elevated pollutant loading levels [see, for example Borsuk et al., 2003; Fries et al., 2007; Gronewold et al., 2008]. Understanding the dynamics of these hydrological systems and applying that understanding to model forecasts is critical to the success of ongoing studies and large-scale planning initiatives addressing these water quality problems, including those being conducted through the United States Environmental Protection Agency (USEPA) total maximum daily load (TMDL) program [National Research Council, 2001; Houck and Environmental Law Institute, 2002; Reckhow, 2003], the most comprehensive and far-reaching water quality management program in the U.S.

Table 1: Summary of land use characteristics for each watershed in the eastern North Carolina and Virginia study area.

<table>
<thead>
<tr>
<th>Watershed</th>
<th>Area (km²)</th>
<th>Land use land cover percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bear Creek</td>
<td>149.4</td>
<td>Agricultural 41.4 Forested 26.3 Urban 0.8 Other 31.5</td>
</tr>
<tr>
<td>Chicod Creek</td>
<td>116.6</td>
<td>Agricultural 56.9 Forested 24.9 Urban 3.7 Other 14.5</td>
</tr>
<tr>
<td>Contentnea Creek</td>
<td>1898.5</td>
<td>Agricultural 42.2 Forested 34.9 Urban 3.0 Other 20.5</td>
</tr>
<tr>
<td>Durham Creek</td>
<td>67.3</td>
<td>Agricultural 11.4 Forested 43.9 Urban 0.4 Other 44.2</td>
</tr>
<tr>
<td>Hood Creek</td>
<td>55.9</td>
<td>Agricultural 5.1 Forested 69.5 Urban 0.3 Other 25.1</td>
</tr>
<tr>
<td>Moccasin Creek</td>
<td>4.9</td>
<td>Agricultural 35.5 Forested 49.6 Urban 0.0 Other 14.9</td>
</tr>
<tr>
<td>Nahunta Swamp</td>
<td>208.2</td>
<td>Agricultural 52.4 Forested 29.3 Urban 1.6 Other 16.7</td>
</tr>
<tr>
<td>Piscataway Creek</td>
<td>72.3</td>
<td>Agricultural 26.0 Forested 67.8 Urban 0.1 Other 6.1</td>
</tr>
<tr>
<td>Potecasi Creek</td>
<td>582.8</td>
<td>Agricultural 31.4 Forested 49.4 Urban 2.6 Other 16.6</td>
</tr>
<tr>
<td>Swift Creek</td>
<td>696.7</td>
<td>Agricultural 34.5 Forested 26.7 Urban 2.5 Other 36.3</td>
</tr>
<tr>
<td>Van Swamp</td>
<td>59.6</td>
<td>Agricultural 5.7 Forested 34.1 Urban 0.0 Other 60.1</td>
</tr>
</tbody>
</table>

We delineated watersheds for eleven streams and creeks in this region (figure 1) for which the United States Geological Survey (USGS) maintains a permanent flow gauge with a relatively long (i.e. approximately 2-10 years) uninterrupted flow record. Land use and land cover (LULC) information for each contributing watershed (based on 2001 imagery) was obtained from Homer et al. [2004]. A summary of the characteristics of each watershed, including total land area and LULC data, is included in table 1).

Daily precipitation and temperature measurements in the vicinity of the watersheds were collected from the National Oceanic and Atmospheric Administration (NOAA) National Climatic Data Center (NCDC) network of monitoring stations (figure 1). Average daily rainfall for each watershed was calculated using the Kriging model in the fields package within the statistics and graphics software program R [Ihaka and Gentleman, 1996]. Average daily temperature in each watershed was assumed equal to the average daily temperature recorded at the nearest NCDC weather station.

3 METHODOLOGY

3.1 Parameter estimation

We calibrate the IHACRES model to the eleven watersheds using the IHACRES v2.1 software package [Littlewood et al., 1997; Croke et al., 2005] and flow, temperature, and precipitation data.
as described above (figure 1). Here, we evaluate the model over a uniform sampling grid, similar to the exhaustive gridding (EG) procedures introduced in Duan et al. [1992], with $\tau_w \in [2, 300]$ and $f \in [0, 12]$. A common criticism of this approach, of course, is the potential computational effort of evaluating the model over a multi-dimensional grid. A distinct advantage of using the IHACRES v2.1 software package to implement this procedure, however, is that for any given pair of parameters $\tau_w$ and $f$, IHACRES implements an instrumental variable (IV) procedure to calculate the other two parameters (i.e. $c$ and $s$), thus greatly reducing the dimensionality of the problem.

Figure 2: Histograms of simulated samples from marginal prior distributions (first row), normalized likelihood functions (second and third rows), and posterior probability density functions (fourth and fifth rows) for four IHACRES model parameters. Likelihood functions and posterior probability density functions are presented based on both fixed (second and fourth rows) and variable (third and fifth rows) data inputs. Vertical dashed lines in the fourth and fifth rows indicate 95\% credible intervals.

We then, following “ensemble modeling” procedures outlined in McIntyre et al. [2005] and McMillan and Clark [2009], combine the resulting parameter sets to form a joint probability distribution potentially suitable for application to both gauged and ungauged watersheds throughout the region. Following guidance presented in similar studies on hydrological model uncertainty [for example, the rainfall-runoff models presented in Duan et al., 1992; Beven, 2001; McMillan and Clark, 2009], we remove from the IHACRES-generated ensemble of calibrated parameter sets those with a Nash-Sutcliffe index of model efficiency (NSE) less than 0.6 [Nash and Sutcliffe, 1970]. We repeat this calibration procedure by modifying the rainfall data using an event multiplier [for details, see Vrugt et al., 2009]. While some recent studies [Kavetski et al., 2006; Stedinger et al., 2008, for example] promote more formal Bayesian approaches to addressing uncertainty (which, among other differences, address variability from all forcing data), our focus on variability and uncertainty in precipitation data is a reasonable simplifying step for this particular
study. This approach leads to two joint parameter posterior distributions, each based on a different assumption regarding input data variability, which can subsequently be used to forecast flows in Ware Creek (and, potentially, in similar ungauged basins).

3.2 Assessing model performance

In order to evaluate potential benefits of propagating forcing data uncertainty into IHACRES model parameters (and model forecasts), we apply the parameter sets from each approach to generate 100,000 daily simulations of flow in Ware Creek, a small tributary of the Newport River Estuary in Eastern NC (see figure 1). Ware Creek does not have a permanent flow gauge, however field-scale flow measurements were collected for model validation between 2007 and 2008 [Kirby-Smith, 2008]. We assess the impact of each assumption regarding forcing data variability using the Bayesian posterior predictive p-value, calculated for each flow observation as the area under the curve of the predictive probability distribution (for a particular observation) which equals or exceeds the observed value [for details, and for a similar application, see Gelman et al., 2004; Gronewold et al., 2009].

4 RESULTS, DISCUSSION, AND CONCLUSIONS

The results of our parameter estimation procedure indicate that the joint parameter likelihood function (and posterior probability density function) derived from a stochastic representation of forcing data differs considerably from the joint parameter likelihood function based on an assumption of “exact” forcing data (figure 2). For example, the marginal normalized likelihood for $\tau_f$ based on an assumption of no data uncertainty (second row, fourth column) is noticeably different from the marginal normalized likelihood for $\tau_f$ based on an assumption of data variability (third row, fourth column). Similar differences are noticeable for other model parameters as well and propagate into differences in marginal posterior probability density functions for each parameter (fourth and fifth row in figure 2, with dashed lines indicated 95% credible intervals) as well as (see following paragraph) into flow forecasts. While not an explicit goal of this paper, these results
support the widely recognized view [see, for example Kuczera and Parent, 1998] that uniquely
determined model parameter values are effectively unreliable, and that appropriate assessment of
model parameter uncertainty is critical to model performance.

Our analysis of flow forecasts in Ware Creek (figure 3) indicates that 95% prediction intervals derived from
a modeling approach which explicitly acknowledges forcing data uncertainty (black lines in figure 3) are con-
siderably narrower than those derived from a modeling approach which assumes invariable data (grey region in
figure 3). Furthermore, our results indicate that the prediction intervals derived from the model acknowledging
data variability may, in fact, fail to include a significant portion of the observed flow measurements (red line in
figure 3). Our analysis of Bayesian posterior p-values (figure 4) further supports this observation (indicated
by the relative weight of each histogram at a p-values of 0, and of p-values less than 0.5).

While these results suggest that the approach of ignoring precipitation data variability might pro-
vide a better explanation of observed flow, we suspect that our results may be somewhat biased
based on our exclusive focus on variability in precipitation measurements alone [an approach cons-
sistent with similar studies by Nandakumar and Mein, 1997; Vrugt et al., 2008; Biemans et al.,
2009]. We leave analysis of flow measurement error and other forcing data for future research, and
conclude by acknowledging the potential advantages of our proposed model evaluation procedure
which, unlike more common approaches based on point estimates of flow [such as the Nash-
Sutcliffe index of model performance, Nash and Sutcliffe, 1970], highlights potential sources of
model bias and error and indicates the magnitude of those errors relative to the various potential
sources of variability and uncertainty in the model forecast.

ACKNOWLEDGMENTS

The authors thank Bill Kirby-Smith for providing field data for Ware Creek. Although this work
was reviewed by EPA and approved for publication, it may not necessarily reflect official Agency
policy.

REFERENCES

Anderson, R. M., B. F. Hobbs, and J. F. Koonce. Modeling effects of forest cover reduction
on larval walleye survival in Lake Erie tributary spawning basins. *Ecosystems*, 9(5):725–739,
2006.
Beven, K. How far can we go in distributed hydrological modelling? *Hydrology and Earth System
Beven, K. Changing ideas in hydrology: the case of physically-based models. *Journal of Hydro-
Beven, K. and J. Freer. Equifinality, data assimilation, and uncertainty estimation in mechanis-
tic modelling of complex environmental systems using the GLUE methodology. *Journal of
uncertainty on discharge calculations for main river basins. *Journal of Hydrometeorology*, 10
Andrew D. Gronewold and Ibrahim M. Alameddine/ Propagating Uncertainty and Variability into Flow Predictions


Kirby-Smith, W. W. Study on flow patterns in Ware Creek, North Carolina, 2008.


**APPENDIX A**

The IHACRES model derives effective rainfall $u_k$ (in mm) at time $k$ from incident rainfall $r_k$ (in mm) through a unitless catchment wetness index $s_k$ as follows [Jakeman et al., 1990]:

$$u_k = r_k s_k$$

$$s_k = cr_k + \left(1 - \frac{1}{\tau_w(T_k)}\right) s_{k-1} ; s_0 = 0 \text{ and (ideally) } 0 < s_k < 1$$

$$\tau_w(T_k) = \tau_w \exp\{(R - T_k)f\}$$

where,

- $c$ = volume-forcing constant (1/mm)
- $\tau_w(T_k)$ = mean soil storage residence time at temperature $T_k$ (unitless)
- $T_k$ = mean daily temperature (deg C)
- $\tau_w$ = catchment drying time constant at reference temperature $R$
- $R$ = reference temperature = 20 (deg C)
- $f$ = temperature modulation factor (1/deg C)

Streamflow $x_k$ at time $k$ is then calculated from effective rainfall $u_k$ through recursive application of the following [Young, 2003]:

$$x_k = \alpha x_{k-1} + \beta u_k$$

where $\alpha$ and $\beta$ are model coefficients such that $\beta = 1 - \alpha$ and $\tau_f = \frac{1}{\ln \alpha}$. 

1802
## Quick Index by Author

<table>
<thead>
<tr>
<th>Author</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adam</td>
<td>1033</td>
</tr>
<tr>
<td>Adams</td>
<td>530</td>
</tr>
<tr>
<td>Agúndez</td>
<td>666</td>
</tr>
<tr>
<td>Agúndez</td>
<td>995</td>
</tr>
<tr>
<td>Agúndez</td>
<td>2106</td>
</tr>
<tr>
<td>Ahmad</td>
<td>1240</td>
</tr>
<tr>
<td>Ahuja</td>
<td>1073</td>
</tr>
<tr>
<td>Ahuja</td>
<td>1190</td>
</tr>
<tr>
<td>Ahuja</td>
<td>1868</td>
</tr>
<tr>
<td>Alameddine</td>
<td>1795</td>
</tr>
<tr>
<td>Alexandrov G.</td>
<td>149</td>
</tr>
<tr>
<td>Alexandrov G.</td>
<td>966</td>
</tr>
<tr>
<td>Alexandrov G.</td>
<td>1549</td>
</tr>
<tr>
<td>Al-Hassini</td>
<td>2316</td>
</tr>
<tr>
<td>Altartouri</td>
<td>213</td>
</tr>
<tr>
<td>Amirthananth</td>
<td>1621</td>
</tr>
<tr>
<td>Andales</td>
<td>2548</td>
</tr>
<tr>
<td>Andersen</td>
<td>1849</td>
</tr>
<tr>
<td>Andrade</td>
<td>1397</td>
</tr>
<tr>
<td>Andrew</td>
<td>633</td>
</tr>
<tr>
<td>Andrews</td>
<td>1779</td>
</tr>
<tr>
<td>Anghileri</td>
<td>2166</td>
</tr>
<tr>
<td>Anputhas</td>
<td>1638</td>
</tr>
<tr>
<td>Anquetin</td>
<td>1003</td>
</tr>
<tr>
<td>Antonello</td>
<td>1081</td>
</tr>
<tr>
<td>Antonello</td>
<td>1174</td>
</tr>
<tr>
<td>Antoniuk</td>
<td>1257</td>
</tr>
<tr>
<td>Aparecida de Lima</td>
<td>162</td>
</tr>
<tr>
<td>Araújo</td>
<td>1649</td>
</tr>
<tr>
<td>Arbuthnott</td>
<td>1917</td>
</tr>
<tr>
<td>Arnold</td>
<td>1499</td>
</tr>
<tr>
<td>Arnold</td>
<td>2393</td>
</tr>
<tr>
<td>Ascoough</td>
<td>1073</td>
</tr>
<tr>
<td>Ascoough</td>
<td>1096</td>
</tr>
<tr>
<td>Ascoough</td>
<td>1190</td>
</tr>
<tr>
<td>Ascoough</td>
<td>1868</td>
</tr>
<tr>
<td>Ascoough</td>
<td>2548</td>
</tr>
<tr>
<td>Atanasova</td>
<td>2212</td>
</tr>
<tr>
<td>Athanasiadis</td>
<td>1199</td>
</tr>
<tr>
<td>Athanasiadis</td>
<td>1849</td>
</tr>
<tr>
<td>Athanasiadis</td>
<td>1858</td>
</tr>
<tr>
<td>Athavale</td>
<td>1933</td>
</tr>
<tr>
<td>Aubert</td>
<td>801</td>
</tr>
<tr>
<td>Aumann</td>
<td>1257</td>
</tr>
<tr>
<td>Ausseil</td>
<td>633</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Author</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ausseil</td>
<td>641</td>
</tr>
<tr>
<td>Bacher</td>
<td>666</td>
</tr>
<tr>
<td>Bacher</td>
<td>995</td>
</tr>
<tr>
<td>Bacher</td>
<td>1166</td>
</tr>
<tr>
<td>Bacher</td>
<td>2106</td>
</tr>
<tr>
<td>Bai</td>
<td>1578</td>
</tr>
<tr>
<td>Baker</td>
<td>296</td>
</tr>
<tr>
<td>Baker</td>
<td>1023</td>
</tr>
<tr>
<td>Balan</td>
<td>716</td>
</tr>
<tr>
<td>Balaž</td>
<td>356</td>
</tr>
<tr>
<td>Balbi</td>
<td>700</td>
</tr>
<tr>
<td>Ballé-Béganton</td>
<td>666</td>
</tr>
<tr>
<td>Ballé-Béganton</td>
<td>995</td>
</tr>
<tr>
<td>Ballé-Béganton</td>
<td>1166</td>
</tr>
<tr>
<td>Ballé-Béganton</td>
<td>2106</td>
</tr>
<tr>
<td>Banovec</td>
<td>2212</td>
</tr>
<tr>
<td>Barbetto</td>
<td>436</td>
</tr>
<tr>
<td>Bari</td>
<td>1621</td>
</tr>
<tr>
<td>Basset</td>
<td>716</td>
</tr>
<tr>
<td>Bauer</td>
<td>980</td>
</tr>
<tr>
<td>Beaumont</td>
<td>675</td>
</tr>
<tr>
<td>Beevers</td>
<td>2174</td>
</tr>
<tr>
<td>Bek</td>
<td>934</td>
</tr>
<tr>
<td>Benedetti</td>
<td>2212</td>
</tr>
<tr>
<td>Bengochea</td>
<td>372</td>
</tr>
<tr>
<td>Escalona</td>
<td>372</td>
</tr>
<tr>
<td>Bennett</td>
<td>1011</td>
</tr>
<tr>
<td>Bennett</td>
<td>1710</td>
</tr>
<tr>
<td>Bennett</td>
<td>2343</td>
</tr>
<tr>
<td>Benoy</td>
<td>860</td>
</tr>
<tr>
<td>Benoy</td>
<td>911</td>
</tr>
<tr>
<td>Berger</td>
<td>791</td>
</tr>
<tr>
<td>Berger</td>
<td>1499</td>
</tr>
<tr>
<td>Berger</td>
<td>569</td>
</tr>
<tr>
<td>Bert</td>
<td>724</td>
</tr>
<tr>
<td>Beulens</td>
<td>1858</td>
</tr>
<tr>
<td>Beverly</td>
<td>296</td>
</tr>
<tr>
<td>Biermann</td>
<td>1300</td>
</tr>
<tr>
<td>Bindoff</td>
<td>2343</td>
</tr>
<tr>
<td>Bjornsson</td>
<td>2139</td>
</tr>
<tr>
<td>Black</td>
<td>539</td>
</tr>
<tr>
<td>Blackstock</td>
<td>1231</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Author</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blackwell</td>
<td>675</td>
</tr>
<tr>
<td>Blanchard</td>
<td>562</td>
</tr>
<tr>
<td>Bohnet</td>
<td>609</td>
</tr>
<tr>
<td>Bohnet</td>
<td>2425</td>
</tr>
<tr>
<td>Bol</td>
<td>675</td>
</tr>
<tr>
<td>Bordenave</td>
<td>1166</td>
</tr>
<tr>
<td>Borkowski</td>
<td>2299</td>
</tr>
<tr>
<td>Bordenave</td>
<td>902</td>
</tr>
<tr>
<td>Bursuk</td>
<td>1745</td>
</tr>
<tr>
<td>Boudens</td>
<td>2257</td>
</tr>
<tr>
<td>Brando</td>
<td>237</td>
</tr>
<tr>
<td>Branger</td>
<td>1003</td>
</tr>
<tr>
<td>Branger</td>
<td>1058</td>
</tr>
<tr>
<td>Braud</td>
<td>1003</td>
</tr>
<tr>
<td>Brag</td>
<td>1058</td>
</tr>
<tr>
<td>Brewer</td>
<td>2503</td>
</tr>
<tr>
<td>Brondizio</td>
<td>774</td>
</tr>
<tr>
<td>Brooks</td>
<td>1762</td>
</tr>
<tr>
<td>Brown</td>
<td>683</td>
</tr>
<tr>
<td>Brown</td>
<td>782</td>
</tr>
<tr>
<td>Bru</td>
<td>153</td>
</tr>
<tr>
<td>Bru</td>
<td>2131</td>
</tr>
<tr>
<td>Brüggemann</td>
<td>499</td>
</tr>
<tr>
<td>Brugnach</td>
<td>547</td>
</tr>
<tr>
<td>Bryan</td>
<td>2139</td>
</tr>
<tr>
<td>Buchan</td>
<td>1231</td>
</tr>
<tr>
<td>Budinčević</td>
<td>356</td>
</tr>
<tr>
<td>Bursik</td>
<td>418</td>
</tr>
<tr>
<td>Bursik</td>
<td>427</td>
</tr>
<tr>
<td>Butler</td>
<td>1868</td>
</tr>
<tr>
<td>Cabrera</td>
<td>774</td>
</tr>
<tr>
<td>Cakmak</td>
<td>625</td>
</tr>
<tr>
<td>Calanca</td>
<td>658</td>
</tr>
<tr>
<td>Calanca</td>
<td>1657</td>
</tr>
<tr>
<td>Calanca</td>
<td>1671</td>
</tr>
<tr>
<td>Calvet</td>
<td>322</td>
</tr>
<tr>
<td>Canu</td>
<td>1925</td>
</tr>
<tr>
<td>Capiral</td>
<td>224</td>
</tr>
<tr>
<td>Carlson</td>
<td>1073</td>
</tr>
<tr>
<td>Carlson</td>
<td>1257</td>
</tr>
<tr>
<td>Carlson</td>
<td>1868</td>
</tr>
<tr>
<td>Author</td>
<td>Page</td>
</tr>
<tr>
<td>-----------------</td>
<td>------</td>
</tr>
<tr>
<td>Carpani</td>
<td>617</td>
</tr>
<tr>
<td>Carraro</td>
<td>1310</td>
</tr>
<tr>
<td>Carretero</td>
<td>1957</td>
</tr>
<tr>
<td>Carvalho</td>
<td>483</td>
</tr>
<tr>
<td>Carvalho, Eduardo</td>
<td>841</td>
</tr>
<tr>
<td>Castellazzi</td>
<td>683</td>
</tr>
<tr>
<td>Castelletti</td>
<td>67</td>
</tr>
<tr>
<td>Castelletti</td>
<td>96</td>
</tr>
<tr>
<td>Castleton</td>
<td>1223</td>
</tr>
<tr>
<td>Cechet</td>
<td>2317</td>
</tr>
<tr>
<td>Cechet</td>
<td>2343</td>
</tr>
<tr>
<td>Cerk</td>
<td>2212</td>
</tr>
<tr>
<td>Chaloupka</td>
<td>1412</td>
</tr>
<tr>
<td>Chan, C.</td>
<td>2326</td>
</tr>
<tr>
<td>Chang</td>
<td>2196</td>
</tr>
<tr>
<td>Characklis</td>
<td>2160</td>
</tr>
<tr>
<td>Chavent</td>
<td>902</td>
</tr>
<tr>
<td>Chen B.</td>
<td>2557</td>
</tr>
<tr>
<td>Chen, B.</td>
<td>2564</td>
</tr>
<tr>
<td>Chen, B.</td>
<td>2627</td>
</tr>
<tr>
<td>Chen, B.</td>
<td>2635</td>
</tr>
<tr>
<td>Chen, B.</td>
<td>2643</td>
</tr>
<tr>
<td>Chen, B.</td>
<td>2651</td>
</tr>
<tr>
<td>Chen, B.</td>
<td>2667</td>
</tr>
<tr>
<td>Chen, D.</td>
<td>2265</td>
</tr>
<tr>
<td>Chen, J.</td>
<td>1240</td>
</tr>
<tr>
<td>Chen, S.</td>
<td>2493</td>
</tr>
<tr>
<td>Chen, S.</td>
<td>2557</td>
</tr>
<tr>
<td>Chen, S.-T.</td>
<td>1703</td>
</tr>
<tr>
<td>Chen, Serena</td>
<td>170</td>
</tr>
<tr>
<td>Chen, Y.</td>
<td>1240</td>
</tr>
<tr>
<td>Cheng, C.-t.</td>
<td>2513</td>
</tr>
<tr>
<td>Cheng, L.</td>
<td>2571</td>
</tr>
<tr>
<td>Cheng, S. K.</td>
<td>2503</td>
</tr>
<tr>
<td>Cherukuru</td>
<td>237</td>
</tr>
<tr>
<td>Chi</td>
<td>2651</td>
</tr>
<tr>
<td>Chsholm</td>
<td>205</td>
</tr>
<tr>
<td>Chua</td>
<td>1688</td>
</tr>
<tr>
<td>Clark</td>
<td>2288</td>
</tr>
<tr>
<td>Codina</td>
<td>1940</td>
</tr>
<tr>
<td>Codina</td>
<td>2457</td>
</tr>
<tr>
<td>Coletti</td>
<td>716</td>
</tr>
<tr>
<td>Collins</td>
<td>633</td>
</tr>
<tr>
<td>Colprim</td>
<td>2237</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Quick Index by Author

<table>
<thead>
<tr>
<th>Author</th>
<th>Pages</th>
<th>Author</th>
<th>Pages</th>
<th>Author</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fan</td>
<td>2378</td>
<td>Galelli</td>
<td>67</td>
<td>Graveline</td>
<td>2131</td>
</tr>
<tr>
<td>Fan</td>
<td>2672</td>
<td>Gao</td>
<td>2579</td>
<td>Green</td>
<td>1073</td>
</tr>
<tr>
<td>Farr</td>
<td>1257</td>
<td>Gao</td>
<td>2672</td>
<td>Greenhalgh</td>
<td>633</td>
</tr>
<tr>
<td>Feikema</td>
<td>296</td>
<td>Garcia-Gasulla</td>
<td>1841</td>
<td>Greenwood</td>
<td>2401</td>
</tr>
<tr>
<td>Feng</td>
<td>2378</td>
<td>Garcia-Mares</td>
<td>304</td>
<td>Gregersen</td>
<td>1081</td>
</tr>
<tr>
<td>Feng</td>
<td>2385</td>
<td>Garcia-Mares</td>
<td>569</td>
<td>Gregersen</td>
<td>1174</td>
</tr>
<tr>
<td>Fenton</td>
<td>1283</td>
<td>Gargoum</td>
<td>444</td>
<td>Grégoire</td>
<td>2044</td>
</tr>
<tr>
<td>Ferencik</td>
<td>196</td>
<td>Garnier</td>
<td>2044</td>
<td>Griffith</td>
<td>675</td>
</tr>
<tr>
<td>Fernando</td>
<td>224</td>
<td>Garrido</td>
<td>1999</td>
<td>Gronewold</td>
<td>1795</td>
</tr>
<tr>
<td>Ferracini</td>
<td>841</td>
<td>Gebreselasie</td>
<td>224</td>
<td>Groos</td>
<td>1081</td>
</tr>
<tr>
<td>Ferrer</td>
<td>322</td>
<td>Geelhoed</td>
<td>1357</td>
<td>Groos</td>
<td>1174</td>
</tr>
<tr>
<td>Ferrero</td>
<td>2007</td>
<td>Geiger</td>
<td>460</td>
<td>Gualtieri</td>
<td>399</td>
</tr>
<tr>
<td>Fiandrino</td>
<td>995</td>
<td>Ger</td>
<td>625</td>
<td>Gruber</td>
<td>692</td>
</tr>
<tr>
<td>Figueiredo</td>
<td>841</td>
<td>Geraldo</td>
<td>2229</td>
<td>Gu</td>
<td>2571</td>
</tr>
<tr>
<td>Filatova</td>
<td>782</td>
<td>Ghosh</td>
<td>2025</td>
<td>Gualtieri</td>
<td>399</td>
</tr>
<tr>
<td>Fink</td>
<td>1096</td>
<td>Gianninni</td>
<td>593</td>
<td>Guariso</td>
<td>1430</td>
</tr>
<tr>
<td>Fiorese</td>
<td>1430</td>
<td>Gibert</td>
<td>1940</td>
<td>Guergachi</td>
<td>1292</td>
</tr>
<tr>
<td>Firanj</td>
<td>356</td>
<td>Gibert</td>
<td>2015</td>
<td>Guergachi</td>
<td>1933</td>
</tr>
<tr>
<td>Firbank</td>
<td>675</td>
<td>Gibert</td>
<td>2079</td>
<td>Guerra</td>
<td>1902</td>
</tr>
<tr>
<td>Fitch</td>
<td>1562</td>
<td>Gibert</td>
<td>2455</td>
<td>Guesdon</td>
<td>2106</td>
</tr>
<tr>
<td>Flacke</td>
<td>708</td>
<td>Gibert</td>
<td>2457</td>
<td>Guesmi</td>
<td>153</td>
</tr>
<tr>
<td>Fletcher</td>
<td>1812</td>
<td>Gijsbers</td>
<td>1081</td>
<td>Guezennec</td>
<td>153</td>
</tr>
<tr>
<td>Flores-Alsina</td>
<td>1999</td>
<td>Gijsbers</td>
<td>1174</td>
<td>Guillard</td>
<td>995</td>
</tr>
<tr>
<td>Flores-Alsina</td>
<td>2204</td>
<td>Gijsbers</td>
<td>1587</td>
<td>Gulden</td>
<td>716</td>
</tr>
<tr>
<td>Flores-Alsina</td>
<td>2221</td>
<td>Gill</td>
<td>205</td>
<td>Gu, A.</td>
<td>868</td>
</tr>
<tr>
<td>Fontenelle</td>
<td>562</td>
<td>Gill</td>
<td>1438</td>
<td>Gurka</td>
<td>313</td>
</tr>
<tr>
<td>Forax</td>
<td>188</td>
<td>Gilson</td>
<td>2044</td>
<td>Guru</td>
<td>1554</td>
</tr>
<tr>
<td>Forbes</td>
<td>1688</td>
<td>Gimona</td>
<td>683</td>
<td>Gutiérrez-Pérez</td>
<td>1990</td>
</tr>
<tr>
<td>Francis</td>
<td>179</td>
<td>Gimona</td>
<td>757</td>
<td>Haase</td>
<td>817</td>
</tr>
<tr>
<td>Francis</td>
<td>1257</td>
<td>Gimona</td>
<td>809</td>
<td>Haase</td>
<td>1265</td>
</tr>
<tr>
<td>Frangoudès</td>
<td>562</td>
<td>Giordano</td>
<td>507</td>
<td>Haase</td>
<td>1507</td>
</tr>
<tr>
<td>Frank</td>
<td>1532</td>
<td>Giove</td>
<td>2073</td>
<td>Hailegiorgis</td>
<td>716</td>
</tr>
<tr>
<td>Frantz</td>
<td>1300</td>
<td>Giove</td>
<td>2409</td>
<td>Hakim</td>
<td>1292</td>
</tr>
<tr>
<td>Frasheri</td>
<td>927</td>
<td>Giupponi</td>
<td>593</td>
<td>Halachev</td>
<td>1933</td>
</tr>
<tr>
<td>Freni</td>
<td>1727</td>
<td>Giupponi</td>
<td>700</td>
<td>Hamza</td>
<td>2316</td>
</tr>
<tr>
<td>Freni</td>
<td>2245</td>
<td>Goethals</td>
<td>876</td>
<td>Hansen</td>
<td>1041</td>
</tr>
<tr>
<td>Friesen</td>
<td>2299</td>
<td>Goethals</td>
<td>1454</td>
<td>Hao</td>
<td>2385</td>
</tr>
<tr>
<td>Fu</td>
<td>2087</td>
<td>Goethals</td>
<td>1948</td>
<td>Hargreaves</td>
<td>934</td>
</tr>
<tr>
<td>Fuertes-Miquel</td>
<td>304</td>
<td>Gomes</td>
<td>841</td>
<td>Harou</td>
<td>1041</td>
</tr>
<tr>
<td>Fuhrer</td>
<td>658</td>
<td>Gonzalez</td>
<td>1454</td>
<td>Harou</td>
<td>2184</td>
</tr>
<tr>
<td>Fuhrer</td>
<td>1657</td>
<td>Gosselin</td>
<td>2257</td>
<td>Harper</td>
<td>1081</td>
</tr>
<tr>
<td>Fuhrer</td>
<td>1671</td>
<td>Gottardo</td>
<td>2073</td>
<td>Harper</td>
<td>1174</td>
</tr>
<tr>
<td>Fürst</td>
<td>919</td>
<td>Gotts</td>
<td>757</td>
<td>Hartig</td>
<td>1481</td>
</tr>
<tr>
<td>Gaddis</td>
<td>530</td>
<td>Gotts</td>
<td>809</td>
<td>Harvey</td>
<td>179</td>
</tr>
<tr>
<td>Gairola</td>
<td>80</td>
<td>Grangere</td>
<td>562</td>
<td>Hasan</td>
<td>1973</td>
</tr>
<tr>
<td>Gairola</td>
<td>1249</td>
<td>Graveline</td>
<td>153</td>
<td>Hasbani</td>
<td>868</td>
</tr>
<tr>
<td>Author</td>
<td>Page</td>
<td>Author</td>
<td>Page</td>
<td>Author</td>
<td>Page</td>
</tr>
<tr>
<td>-------------</td>
<td>------</td>
<td>-------------</td>
<td>------</td>
<td>-------------</td>
<td>------</td>
</tr>
<tr>
<td>Hasse, A.</td>
<td>577</td>
<td>Jackson</td>
<td>1787</td>
<td>Kigobe</td>
<td>825</td>
</tr>
<tr>
<td>Hasse, A.</td>
<td>2434</td>
<td>Jaffe</td>
<td>1917</td>
<td>Kigobe</td>
<td>2096</td>
</tr>
<tr>
<td>Hasse, D.</td>
<td>577</td>
<td>Jagers</td>
<td>1050</td>
<td>Kiiveri</td>
<td>2326</td>
</tr>
<tr>
<td>Hasse, D.</td>
<td>2434</td>
<td>Jagers</td>
<td>1150</td>
<td>King</td>
<td>2139</td>
</tr>
<tr>
<td>Hassell</td>
<td>1663</td>
<td>Jakeman</td>
<td>1011</td>
<td>Kinzelbach</td>
<td>2174</td>
</tr>
<tr>
<td>Hastings</td>
<td>2053</td>
<td>Jakeman</td>
<td>1710</td>
<td>Kipka</td>
<td>1096</td>
</tr>
<tr>
<td>Hawker</td>
<td>2025</td>
<td>Jakeman</td>
<td>1719</td>
<td>Kirilenko</td>
<td>1630</td>
</tr>
<tr>
<td>Hayes</td>
<td>2503</td>
<td>Jakoby</td>
<td>1532</td>
<td>Kirsch</td>
<td>126</td>
</tr>
<tr>
<td>Hayter</td>
<td>380</td>
<td>Janisch</td>
<td>980</td>
<td>Kirsch</td>
<td>2160</td>
</tr>
<tr>
<td>Hehir</td>
<td>1578</td>
<td>Janjic</td>
<td>364</td>
<td>Kjeldsen</td>
<td>650</td>
</tr>
<tr>
<td>Hennickler</td>
<td>980</td>
<td>Jankowfsky</td>
<td>1003</td>
<td>Klein</td>
<td>1404</td>
</tr>
<tr>
<td>Hepting</td>
<td>1910</td>
<td>Jankowfsky</td>
<td>1058</td>
<td>Klepeis</td>
<td>205</td>
</tr>
<tr>
<td>Hepting</td>
<td>1917</td>
<td>Janmaat</td>
<td>1638</td>
<td>Klotz</td>
<td>179</td>
</tr>
<tr>
<td>Herrera</td>
<td>1925</td>
<td>Janssen</td>
<td>1894</td>
<td>Knapen</td>
<td>1081</td>
</tr>
<tr>
<td>Herrere</td>
<td>1980</td>
<td>Janssesen</td>
<td>1300</td>
<td>Knapen</td>
<td>1174</td>
</tr>
<tr>
<td>Herzig</td>
<td>633</td>
<td>Jansson</td>
<td>1273</td>
<td>Knapen</td>
<td>1894</td>
</tr>
<tr>
<td>Higgins</td>
<td>2139</td>
<td>Jarvis</td>
<td>757</td>
<td>Knetsch</td>
<td>492</td>
</tr>
<tr>
<td>Higuchi</td>
<td>952</td>
<td>Jeanes</td>
<td>1779</td>
<td>Kobayashi</td>
<td>1397</td>
</tr>
<tr>
<td>Hirsch</td>
<td>959</td>
<td>Jessen</td>
<td>1421</td>
<td>Koenderik</td>
<td>1207</td>
</tr>
<tr>
<td>Hoanh</td>
<td>2114</td>
<td>Jirayoot</td>
<td>2114</td>
<td>Kok</td>
<td>1300</td>
</tr>
<tr>
<td>Holland</td>
<td>2139</td>
<td>Johst</td>
<td>748</td>
<td>Kokoszkiewicz</td>
<td>1182</td>
</tr>
<tr>
<td>Holmes</td>
<td>1812</td>
<td>Jolma</td>
<td>196</td>
<td>Kolega</td>
<td>1820</td>
</tr>
<tr>
<td>Holz</td>
<td>2343</td>
<td>Jolma</td>
<td>213</td>
<td>Kollat</td>
<td>1379</td>
</tr>
<tr>
<td>Holzkämper</td>
<td>658</td>
<td>Jones</td>
<td>894</td>
<td>Kopp</td>
<td>313</td>
</tr>
<tr>
<td>Holzkämper</td>
<td>1105</td>
<td>Jones</td>
<td>2503</td>
<td>Korfmancher</td>
<td>2025</td>
</tr>
<tr>
<td>Holzkämper</td>
<td>1657</td>
<td>Jones, M.</td>
<td>418</td>
<td>Korpinnen</td>
<td>213</td>
</tr>
<tr>
<td>Hood</td>
<td>131</td>
<td>Joschko</td>
<td>474</td>
<td>Kragt</td>
<td>1011</td>
</tr>
<tr>
<td>Hou</td>
<td>2579</td>
<td>Ju</td>
<td>2564</td>
<td>Kralisch</td>
<td>1096</td>
</tr>
<tr>
<td>Hou</td>
<td>2619</td>
<td>Ju</td>
<td>2651</td>
<td>Krause</td>
<td>1073</td>
</tr>
<tr>
<td>Hoyle</td>
<td>1745</td>
<td>Juizo</td>
<td>2174</td>
<td>Krause</td>
<td>1096</td>
</tr>
<tr>
<td>Hrebicek</td>
<td>1199</td>
<td>Jusoff</td>
<td>280</td>
<td>Krause</td>
<td>1858</td>
</tr>
<tr>
<td>Hřebiček</td>
<td>1113</td>
<td>Kabisch, N.</td>
<td>577</td>
<td>Krishnan</td>
<td>1933</td>
</tr>
<tr>
<td>Hu, B.-G.</td>
<td>1965</td>
<td>Kabisch, N.</td>
<td>2434</td>
<td>Krol</td>
<td>1649</td>
</tr>
<tr>
<td>Hummell</td>
<td>1081</td>
<td>Kabisch, S.</td>
<td>577</td>
<td>Kropp</td>
<td>1885</td>
</tr>
<tr>
<td>Hummell</td>
<td>1174</td>
<td>Kabisch, S.</td>
<td>2434</td>
<td>Kumar</td>
<td>1105</td>
</tr>
<tr>
<td>Husain</td>
<td>75</td>
<td>Karaouli</td>
<td>153</td>
<td>Kumar, V.</td>
<td>1438</td>
</tr>
<tr>
<td>Husain</td>
<td>88</td>
<td>Karatzglou</td>
<td>1925</td>
<td>Kumar, V.</td>
<td>1446</td>
</tr>
<tr>
<td>Huser</td>
<td>1283</td>
<td>Kasprzyk</td>
<td>2160</td>
<td>Kumar, V.</td>
<td>1762</td>
</tr>
<tr>
<td>Hutchings</td>
<td>650</td>
<td>Kassahun</td>
<td>1858</td>
<td>Kuo</td>
<td>112</td>
</tr>
<tr>
<td>Hutchins</td>
<td>782</td>
<td>Kearney</td>
<td>1595</td>
<td>Laciana</td>
<td>724</td>
</tr>
<tr>
<td>Izquierdo</td>
<td>1925</td>
<td>Keesman</td>
<td>1719</td>
<td>Lacombe</td>
<td>2114</td>
</tr>
<tr>
<td>Izquierdo</td>
<td>1980</td>
<td>Kennedy</td>
<td>716</td>
<td>Lahcen</td>
<td>188</td>
</tr>
<tr>
<td>Izquierdo</td>
<td>1990</td>
<td>Khaiter</td>
<td>272</td>
<td>Laín</td>
<td>1885</td>
</tr>
<tr>
<td>Izquierdo</td>
<td>2035</td>
<td>Khaiter</td>
<td>1387</td>
<td>Lamé</td>
<td>1357</td>
</tr>
<tr>
<td>Izquierdo</td>
<td>2465</td>
<td>Khaiter</td>
<td>2452</td>
<td>Lample</td>
<td>666</td>
</tr>
<tr>
<td>Izquierdo</td>
<td>2466</td>
<td>Khandokar</td>
<td>75</td>
<td>Lample</td>
<td>995</td>
</tr>
<tr>
<td>Author</td>
<td>First Page</td>
<td>Last Page</td>
<td>Middle Name</td>
<td>Last Name</td>
<td>Title</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
<td>-----------</td>
<td>-------------</td>
<td>-----------</td>
<td>-------</td>
</tr>
<tr>
<td>Lane</td>
<td>296</td>
<td></td>
<td></td>
<td>Limosani</td>
<td></td>
</tr>
<tr>
<td>Lane</td>
<td>894</td>
<td></td>
<td></td>
<td>Lin</td>
<td></td>
</tr>
<tr>
<td>Lane</td>
<td>1762</td>
<td></td>
<td></td>
<td>Lin</td>
<td></td>
</tr>
<tr>
<td>Langan</td>
<td>1231</td>
<td></td>
<td></td>
<td>Li</td>
<td></td>
</tr>
<tr>
<td>Lange</td>
<td>1438</td>
<td></td>
<td></td>
<td>Lindim</td>
<td></td>
</tr>
<tr>
<td>Lange</td>
<td>1446</td>
<td></td>
<td></td>
<td>Lissner</td>
<td></td>
</tr>
<tr>
<td>Laniado</td>
<td>584</td>
<td></td>
<td></td>
<td>Liu, C.-T.</td>
<td>1965</td>
</tr>
<tr>
<td>Lankhorst</td>
<td>601</td>
<td></td>
<td></td>
<td>Liu, D.</td>
<td></td>
</tr>
<tr>
<td>Lanuza</td>
<td>117</td>
<td></td>
<td></td>
<td>Liu, D.</td>
<td></td>
</tr>
<tr>
<td>Lanzi</td>
<td>1310</td>
<td></td>
<td></td>
<td>Liu, G.-Y.</td>
<td>2627</td>
</tr>
<tr>
<td>Lapola</td>
<td>1421</td>
<td></td>
<td></td>
<td>Liu, J.</td>
<td></td>
</tr>
<tr>
<td>Last</td>
<td>2503</td>
<td></td>
<td></td>
<td>Liu, L.</td>
<td></td>
</tr>
<tr>
<td>Latchininsky</td>
<td>2053</td>
<td></td>
<td></td>
<td>Liu, L.</td>
<td></td>
</tr>
<tr>
<td>Latynskiy</td>
<td>791</td>
<td></td>
<td></td>
<td>Liu, W.</td>
<td></td>
</tr>
<tr>
<td>Laugier</td>
<td>995</td>
<td></td>
<td></td>
<td>Liu, X.</td>
<td></td>
</tr>
<tr>
<td>Laurain</td>
<td>2044</td>
<td></td>
<td></td>
<td>Llinás</td>
<td></td>
</tr>
<tr>
<td>Lautenbach</td>
<td>692</td>
<td></td>
<td></td>
<td>Lloyd, W.</td>
<td>1073</td>
</tr>
<tr>
<td>Lautenbach</td>
<td>1489</td>
<td></td>
<td></td>
<td>López</td>
<td></td>
</tr>
<tr>
<td>Lautenbach</td>
<td>2454</td>
<td></td>
<td></td>
<td>López-Jiménez</td>
<td>372</td>
</tr>
<tr>
<td>Le</td>
<td>1524</td>
<td></td>
<td></td>
<td>López-Patiño</td>
<td>372</td>
</tr>
<tr>
<td>Le Mao</td>
<td>562</td>
<td></td>
<td></td>
<td>Lorenz</td>
<td></td>
</tr>
<tr>
<td>Le Moine</td>
<td>2106</td>
<td></td>
<td></td>
<td>Lorz</td>
<td></td>
</tr>
<tr>
<td>Leavesley</td>
<td>1073</td>
<td></td>
<td></td>
<td>Losekann</td>
<td></td>
</tr>
<tr>
<td>Leavesley</td>
<td>1190</td>
<td></td>
<td></td>
<td>Lotov</td>
<td></td>
</tr>
<tr>
<td>Leavesley</td>
<td>1868</td>
<td></td>
<td></td>
<td>Lowndes</td>
<td></td>
</tr>
<tr>
<td>Leccia</td>
<td>902</td>
<td></td>
<td></td>
<td>Lozano</td>
<td></td>
</tr>
<tr>
<td>Lee, B.</td>
<td>1554</td>
<td></td>
<td></td>
<td>Ludwig</td>
<td></td>
</tr>
<tr>
<td>Lee, C.-S.</td>
<td>2196</td>
<td></td>
<td></td>
<td>Luè</td>
<td></td>
</tr>
<tr>
<td>Lee, E.E.</td>
<td>2025</td>
<td></td>
<td></td>
<td>Lung</td>
<td></td>
</tr>
<tr>
<td>Lee, Y.-C.</td>
<td>2196</td>
<td></td>
<td></td>
<td>Lutz</td>
<td></td>
</tr>
<tr>
<td>Leirens</td>
<td>2229</td>
<td></td>
<td></td>
<td>Lyne</td>
<td></td>
</tr>
<tr>
<td>Leon</td>
<td>860</td>
<td></td>
<td></td>
<td>Ma</td>
<td></td>
</tr>
<tr>
<td>Leon</td>
<td>911</td>
<td></td>
<td></td>
<td>Macal</td>
<td></td>
</tr>
<tr>
<td>Lerner</td>
<td>1105</td>
<td></td>
<td></td>
<td>MacDonald</td>
<td></td>
</tr>
<tr>
<td>Lerner</td>
<td>1438</td>
<td></td>
<td></td>
<td>Maciag</td>
<td></td>
</tr>
<tr>
<td>Lerner</td>
<td>1446</td>
<td></td>
<td></td>
<td>MacLellan</td>
<td></td>
</tr>
<tr>
<td>Lesins</td>
<td>952</td>
<td></td>
<td></td>
<td>Macleod</td>
<td></td>
</tr>
<tr>
<td>Lester</td>
<td>2139</td>
<td></td>
<td></td>
<td>MacLeod</td>
<td></td>
</tr>
<tr>
<td>Lester</td>
<td>2476</td>
<td></td>
<td></td>
<td>Madani</td>
<td></td>
</tr>
<tr>
<td>Levrer</td>
<td>666</td>
<td></td>
<td></td>
<td>Maes</td>
<td></td>
</tr>
<tr>
<td>Lewis</td>
<td>1688</td>
<td></td>
<td></td>
<td>Magliocca</td>
<td></td>
</tr>
<tr>
<td>Li, B.</td>
<td>2025</td>
<td></td>
<td></td>
<td>Magness</td>
<td></td>
</tr>
<tr>
<td>Liao</td>
<td>2513</td>
<td></td>
<td></td>
<td>Maheepala</td>
<td></td>
</tr>
<tr>
<td>Liao</td>
<td>2635</td>
<td></td>
<td></td>
<td>Makeschin</td>
<td></td>
</tr>
<tr>
<td>Lim</td>
<td>1663</td>
<td></td>
<td></td>
<td>Makowski</td>
<td></td>
</tr>
<tr>
<td>Lima</td>
<td>841</td>
<td></td>
<td></td>
<td>Makropoulos</td>
<td>1023</td>
</tr>
</tbody>
</table>
Quick Index by Author

Monclús 2007 Nieves 1841 Pérez-García 1925
Mongruel 562 Nishi 1257 Pérez-García 1980
Mongruel 666 Nativattananon 1335 Pérez-García 1990
Mongruel 995 Nolan 2139 Pérez-García 2035
Mongruel 1166 North 724 Perišić 356
Mongruel 2106 Norton 1710 Perraud 1578
Monistrol 851 Norton 1770 Perumal 436
Montalvo 1990 Oaks 1868 Pessoa 841
Montalvo 2035 Olaya 1473 Pessoa 1902
Mora Rodriguez 304 Oliver 2139 Peters 1554
Mora-Lopez 1957 Onurlu 625 Petersen 1300
Moramarco 436 Orr 675 Petit 902
Mora-Pérez 372 Otero 1473 Phakat 2326
Moreira 1352 Ouammi 972 Phyn 1283
Moreira 1397 Overton 2139 Pianosi 2123
Morgan 633 Oxley 1129 Pianosi 2166
Morgan 1438 Padula 1041 Pieri 427
Morris 296 Page 474 Pietzsch 919
Moshirvaziri 894 Paiva 1352 Pilgrim 675
Moulin 2257 Pan 1257 Piliouigne 1957
Mu, Z. 2571 Paningbatan 117 Pillmann 452
Mukai 232 Pano 927 Pinedo-Vasquez 774
Mukhi 2299 Parigot 188 Pinho 380
Müller 801 Parker 782 Pinte 1041
Müller 1273 Parrado 1310 Piorr 1265
Müller 1532 Patra 418 Pitman 418
Müller 1877 Patra 427 Pizzol 2073
Munro 340 Pattberg 1300 Pizzol 2409
Munro 410 Pauwels 876 Poch 851
Muratori 584 Pauwels 1948 Poch 1999
Murlà 851 Paxton 2416 Podestá 724
Murphy 1066 Payraudeau 2044 Poggio 683
Nakahai 245 Pazzianotto 1902 Polhill 757
Narayanan 1463 Pechhlivanidis 1787 Polhill 809
Narayanan 2334 Pelton 1158 Pollino 170
Nassauer 782 Pelton 1223 Pollino 516
Nazeri 280 Pendelberry 1327 Pollino 2087
Neighbour 2299 Pereira 841 Pollino 2493
Neveu 1166 Perera 262 Pratt 1554
Newham 1710 Perez Young 162 Preiss 2454
Newham 1719 Perez, J. 562 Preston 2401
Ngenyama 1292 Perez, J. 1166 Preziosi 507
Nicolas 322 Pérez, L. 766 Price 1868
Niehaus 2307 Perez, Pascal 700 Priess 1489
Niemi 196 Pérez-Bonilla 2079 Princz 139
Nieuwenhuizen 601
Quick Index by Author

Prokop 2025  Rodriguez 1003  Schilling 791
Prou 666  Rodriguez 1058  Schimak 1199
Prowse 885  Rodriguez 2229  Schmitt 460
Pulido-Velazquez 1041  Rodriguez-Roda 1999  Schmitz 474
Puotinen 205  Rodriguez-Roda 2007  Scholten 1858
Purucker 1121  Rodriguez-Roda 2204  Scholz 1524
Quaas 1532  Rojas 1073  Schreiber 469
Quach 96  Rojas 1868  Schut 1089
Queiroz 841  Romano 507  Schwartz 1265
Queiroz 1902  Romano 1438  Schwarz 817
Quinn 2533  Rosenberg 1041  Schwarz 1489
Qureshi 2467  Rouleau 716  Schwarz 2454
Rajkovic 364  Roux 1033  Schwarz 2035
Ralihalizara 801  Rovere 724  Schweitzer 1516
Ramirez 2237  Rowe 1129  Sechi 2150
Ratcliffe 1595  Rutledge 633  Seen 188
Red 1041  Rutledge 1283  Seiber 1273
Reed 1379  Sacile 972  Seidl 1524
Reed 2160  Safiolea 1023  Seixas 1902
Reichert 1735  Safrova 732  Semenzin 2073
Reis 1129  Sahoo 436  Seppelt 692
Rethoret 1166  Saint 1066  Serra 483
Réthoret 666  Sallis 1463  Seth 1343
Reutemann 1812  Sallis 2334  Sevilla 2015
Reynaud 569  Salvetti 2123  Sevilla 2063
Reynaud 1041  Sammartino 332  Sevilla 2455
Rheinheimer 1041  Sammartino 348  Shah 1379
Richards 1412  Sanabria 2317  Shahadi 245
Richards 2288  Sanabria 2343  Shanmuganathan 1463
Rink 577  Sánchez-Marré 1940  Shanmuganathan 2334
Rink 2434  Sánchez-Marré 2015  Shariff 80
Riolo 782  Sánchez-Marré 2063  Shariff 1249
Riu 1999  Sánchez-Marré 2455  Shaw 1438
Rivington 1231  Sánchez-Marré 2457  Shaw 1446
Rizk 2316  Sanchis 322  Shen 2513
Rizzi 2073  Sancho 2007  Shepherd 675
Rizzi 2409  Sanguino 1473  Sheridan 427
Rizzoli 1199  Sarathi 313  Sheridan 894
Rizzoli 1849  Savitskaya 1680  Shogren 748
Rizzoli 1858  Schaab 288  Shrestha, D. 825
Robba 972  Schade 1081  Shrestha, R. 885
Robert 2106  Schade 1174  Sidrach-de-Cardona 1957
Robin 562  Schaldach, A. 1421  Silvert 523
Robinson 782  Schaldach, R. 1421  Simon, K-H 452
Robson 237  Schell 2053  Sinclair 633
Robson 2359  Scherb 499  Sinclair 633
<table>
<thead>
<tr>
<th>Author</th>
<th>Page No.</th>
<th>Co-Author</th>
<th>Journal Name</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skerjanec</td>
<td>2212</td>
<td></td>
<td>Van Ittersum</td>
<td>1849</td>
</tr>
<tr>
<td>Skewes</td>
<td>2503</td>
<td></td>
<td>Van Oel</td>
<td>1649</td>
</tr>
<tr>
<td>Slimani</td>
<td>153</td>
<td></td>
<td>Van Ruijven</td>
<td>1207</td>
</tr>
<tr>
<td>Sloboda</td>
<td>833</td>
<td></td>
<td>Vanecek</td>
<td>1081</td>
</tr>
<tr>
<td>Smith</td>
<td>1671</td>
<td></td>
<td>Vanecek</td>
<td>1174</td>
</tr>
<tr>
<td>Sojda</td>
<td>126</td>
<td>Tan, X.</td>
<td>Vannier</td>
<td>1003</td>
</tr>
<tr>
<td>Sojda</td>
<td>2401</td>
<td>Tan, X.</td>
<td>Vanrobaeys</td>
<td>860</td>
</tr>
<tr>
<td>Sojda</td>
<td>2416</td>
<td>Tan, X.</td>
<td>Vanrobaeys</td>
<td>911</td>
</tr>
<tr>
<td>Sorektechnik</td>
<td>2106</td>
<td>Tan, X.</td>
<td>Vanrolleghem</td>
<td>2221</td>
</tr>
<tr>
<td>Soller</td>
<td>1223</td>
<td>Tang</td>
<td>Vernier</td>
<td>666</td>
</tr>
<tr>
<td>Solomatine</td>
<td>825</td>
<td>Tang</td>
<td>Vernier</td>
<td>902</td>
</tr>
<tr>
<td>Soncini-Sessa</td>
<td>67</td>
<td>Tanner</td>
<td>Vernier</td>
<td>1166</td>
</tr>
<tr>
<td>Soncini-Sessa</td>
<td>96</td>
<td>Teng</td>
<td>Verweij</td>
<td>601</td>
</tr>
<tr>
<td>Soncini-Sessa</td>
<td>2123</td>
<td>Terhorst</td>
<td>Verweij</td>
<td>1273</td>
</tr>
<tr>
<td>Soncini-Sessa</td>
<td>2166</td>
<td>Termini</td>
<td>Verweij</td>
<td>1894</td>
</tr>
<tr>
<td>Soukopová</td>
<td>1113</td>
<td>Termini</td>
<td>Vescoutis</td>
<td>1820</td>
</tr>
<tr>
<td>Soulis</td>
<td>153</td>
<td>Thees</td>
<td>Viallet</td>
<td>1003</td>
</tr>
<tr>
<td>Soulis</td>
<td>253</td>
<td>Therond</td>
<td>Viallet</td>
<td>1058</td>
</tr>
<tr>
<td>Soulis</td>
<td>1611</td>
<td>Thurston</td>
<td>Vieira</td>
<td>380</td>
</tr>
<tr>
<td>Soulis</td>
<td>2131</td>
<td>Ticehurst</td>
<td>Viera Vilella</td>
<td>162</td>
</tr>
<tr>
<td>Sousa</td>
<td>841</td>
<td>Ticehurst</td>
<td>Villanueva</td>
<td>2063</td>
</tr>
<tr>
<td>Squire</td>
<td>1595</td>
<td>Tilmant</td>
<td>Viviani</td>
<td>1727</td>
</tr>
<tr>
<td>Srinetra</td>
<td>2114</td>
<td>Tilmant</td>
<td>Viviani</td>
<td>1753</td>
</tr>
<tr>
<td>Stacey</td>
<td>179</td>
<td>Timbal</td>
<td>Voigt</td>
<td>499</td>
</tr>
<tr>
<td>Stamou</td>
<td>1023</td>
<td>Tomberlin</td>
<td>Voinov</td>
<td>530</td>
</tr>
<tr>
<td>Stefanescu</td>
<td>418</td>
<td>Tomlinson</td>
<td>Voinov</td>
<td>1139</td>
</tr>
<tr>
<td>Stefanescu</td>
<td>427</td>
<td>Tonguc</td>
<td>Volk</td>
<td>692</td>
</tr>
<tr>
<td>Stefonopoulo</td>
<td>153</td>
<td>Top</td>
<td>Vukovic</td>
<td>364</td>
</tr>
<tr>
<td>Stefopoulo</td>
<td>2131</td>
<td>Torresan</td>
<td>Waanders</td>
<td>2139</td>
</tr>
<tr>
<td>Stelfox</td>
<td>1257</td>
<td>Torresan</td>
<td>Walker</td>
<td>1240</td>
</tr>
<tr>
<td>Strauch</td>
<td>692</td>
<td>Troost</td>
<td>Wall</td>
<td>732</td>
</tr>
<tr>
<td>Struski</td>
<td>562</td>
<td>Troost</td>
<td>Wang, J.H.</td>
<td>2579</td>
</tr>
<tr>
<td>Stuut</td>
<td>2316</td>
<td>Tryby</td>
<td>Wang, L.</td>
<td>2265</td>
</tr>
<tr>
<td>Su</td>
<td>1327</td>
<td>Tryby</td>
<td>Wang, L.</td>
<td>2587</td>
</tr>
<tr>
<td>Suarez</td>
<td>126</td>
<td>Tscherning</td>
<td>Wang, S.</td>
<td>2571</td>
</tr>
<tr>
<td>Sulis</td>
<td>2150</td>
<td>Tseng</td>
<td>Wang, Y.</td>
<td>2378</td>
</tr>
<tr>
<td>Sullivan</td>
<td>313</td>
<td>Ubeda</td>
<td>Wang, Y.</td>
<td>2571</td>
</tr>
<tr>
<td>Sun</td>
<td>2366</td>
<td>Uribe</td>
<td>Wang, Y.</td>
<td>2579</td>
</tr>
<tr>
<td>Sun</td>
<td>2372</td>
<td>Valeo</td>
<td>Wang, Y.</td>
<td>2619</td>
</tr>
<tr>
<td>Sun, S.</td>
<td>740</td>
<td>Valiantzas</td>
<td>Wang, L.</td>
<td>2595</td>
</tr>
<tr>
<td>Sun, S.</td>
<td>782</td>
<td>Van Delden</td>
<td>Wattenalapa</td>
<td>1335</td>
</tr>
<tr>
<td>Sun, X.-M.</td>
<td>2672</td>
<td>Van Delden</td>
<td>Wätzold</td>
<td>748</td>
</tr>
<tr>
<td>Sunyer</td>
<td>851</td>
<td>Van Griensven</td>
<td>Weber</td>
<td>724</td>
</tr>
<tr>
<td>Sutherland</td>
<td>1257</td>
<td>Van Griensven</td>
<td>Weber</td>
<td>2123</td>
</tr>
<tr>
<td>Swayne</td>
<td>131</td>
<td>Van Griensven</td>
<td>Weber</td>
<td>2166</td>
</tr>
</tbody>
</table>
Quick Index by Author

Wedderburn 1283   Zabeo 2409
Weiß 1885   Zakaria 80
Welsh 539   Zasada 1265
Wery 1033   Zeng 232
Wetzel 1096   Zhang, H. 1828
Whelan 1121   Zhang, X. 1696
Whelan 1158   Zhang, X.J. 2619
Whelan 1223   Zhang, X.-Y. 2672
White 2343   Zhang, Z. 2265
Whitten 2467   Zhao, R. 2372
Wickramasuriya 205   Zheng 2378
Widdowson 1343   Zhou Qiqi 554
Wien 1849   Zhu, T. 1041
Wijesekara 868   Zhuang 2571
Wills 2541
Winebrake 2025
Wohlgemuth 483
Wong, Isaac 833
Wong, Isaac 860
Wong, Isaac 911
Wrobel 1885
Wu 675
Wu, R. 2587
Wu, X. 2272
Wu, X.-G. 2351
Wu, X.-y. 2513
Wu, J. 2272
Xevi 1240
Xu, J. 2603
Xu, J. 2659
Yalew 1182
Yang, C.-Y. 2196
Yang, J. 2635
Yang, X. 2587
Yang, X. 2595
Yang, X. 2595
Yang, Z.-F. 2366
Yang, Z.-F. 2372
Yang, Z.-F. 2627
Yang, Z.-F. 2667
Yang, J. 2557
Yarmoloy 1257
Ying 1327
Young, G.F. 708
Yu 1703
Zabeo 2073