

Modelling water use intensity at farm scale – an interdisciplinary study of agriculture and hydrology (AgroHyd)

Simone Kraatz^{1a}, Uwe Hunstock², Judy A. Libra^{1b}, Katrin Drastig^{1c}

¹Leibniz Institute for Agricultural Engineering Potsdam-Bornim

^{1a}(sikraatz@atb-potsdam.de)

^{1b}(jlibra@atb-potsdam.de)

^{1c}(kdrastig@atb-potsdam.de)

²runlevel3 GmbH Berlin (u.hunstock@runlevel3.de)

Abstract: Efficient water management strategies based on knowledge about how agricultural and hydrological processes affect local and regional water flows are needed to deal with changes in water availability due to global change. Transparent and flexible tools that can predict the effect of farm systems on these water flows are required. Such a new modelling tool is being built based on the database platform MongoDB. It will be able to calculate the influence of natural processes and of management decisions on the water use intensity in a variety of farm systems for the production of food and biomass for energy. Instead of treating the farm scale as a “black box”, the tool models the local agriculture and hydrologic processes, allowing the calculation of water-based indicators for various scenarios. Such forward-looking environmental indicators can function as early-warning signals, and support early intervention in the processes by enabling resource efficient process changes to be identified. The introduced software “ATB Modelling Database” is flexible enough to allow the addition of new processes and datasets as the database is expanded to reflect the diversity of the worldwide farming processes.

Keywords: water intensity; agriculture; hydrology; modelling; database.

1 Introduction

The decreasing availability of water caused by overuse and climate change, together with a growing world population, requires increased efficiency in water use now and in the future. The agri-food sector is of particular importance in understanding and addressing the challenges of water scarcity since agriculture is known to be the largest freshwater consumer globally. It accounts for 70% of freshwater withdrawals (UNESCO-WWAP, 2009). In order to develop recommendations for efficient water management strategies for stakeholders at the local and regional levels, more knowledge is needed about the water flows at the farm scale and how they can be managed. This information needs to be translated into sustainability indicators, providing decision makers a basis on which to evaluate the local efficiency of green water use for rainfed crop production and blue water use for irrigation and technical water. Currently, water uses such as in-stream and degradative uses are still underrepresented in the methodological development of water footprinting (Berger and Finkbeiner, 2010). The consumption of green water is only considered in a few methodological procedures (Allan, 2008; Hoekstra and Hung, 2002; Mila i Canals et al., 2008). But the accounting of green water use in agricultural processes is necessary to assess effects like water deprivation for ecosystems and the reduction in the renewability of ground and surface water. Existing studies about water use in agriculture are often focussed on separate parts of the system, either on plant production (Molden et al., 2003) or livestock production (Peden et al., 2007). Linking the production subsystems to produce a holistic view of the farm system with its interactions as a function of management

systems can be an enormous undertaking. It requires the coupling of a suite of biophysical models to produce large, complex sets of output data that are hard to interpret (Moore et al., 2011). To aggregate and present such complex datasets in an understandable way, Moore et al., (2011) introduce a conceptual framework based on multiple water use efficiencies that can be used to compare the productivity and sustainability of farming practices. Included are direct water flows within the system and other material flows (e.g. biomass, money). However, the underlying task of simulating water flows (including direct and indirect water flows) in coupled agricultural and hydrological processes at the farm scale with a flexible and manageable model remains a challenge.

This need for the development of improved models with a holistic representation of coupled hydrologic processes in aquatic and terrestrial ecosystems across landscapes has been the topic of many recent national and international publications and meetings (Jaeger and Henrichs, 2008, Famiglietti et al., 2010, Teutsch and Krueger 2010, Wood et al., 2011). Various groups have called for community efforts to overcome the gaps in our knowledge and in our tools for modelling integrated systems. They have outlined steps to develop such new integrated models with predictive capabilities. The groups have in common that they call for coordinated model development and integration across subdisciplines to develop models with open access frameworks and highly flexible model structures that can use efficient parallel computing. Existing and new data sources should be easily incorporated, and sensitivity and uncertainty analyses integrated in the modelling system. The coordinated development and use of such predictive models requires increased transparency of the modelling tools, their requirements and outcomes.

Currently the software “ATB Modelling Database” is being developed to calculate the water flows associated with the production of plant and animal products at the farm scale. Instead of treating the farm scale as a “black box”, the modelling system is designed as a tool which combines agriculture and hydrologic processes to determine the water use intensity. The goal is to develop a holistic on-farm model of water flows which incorporates algorithms to describe how farm practices, local soil conditions, climate and hydrological processes influence the complex water-soil-plant-livestock interactions, with a transparent and expandable model structure. As starting point for the model development, the geographical location northern Germany is used.

In order to cope with the large amount of data and calculations involved, the implementation of the model requires efficient and expandable software. The non-relational database platform MongoDB offers features that meet the project software development requirements and supports the development of a flexible modelling system. This schema-free open source database developed mainly for web applications and is capable of handling diverse formats and large datasets. The ability to easily implement changes in the database allows a stepwise expansion of the model to broaden its application to different farm types, products and regional conditions as the project knowledge base expands. The main idea of this concept is to run all calculations as embedded JavaScript code inside the database. MongoDB provides parallelization, sharding (partitioning of data across servers), geographical functions, embedded objects and map/reduce function out of the box. This will enable the research team to access a broad amount of data and run complex models in a reasonable amount of time.

The goal of this study is:

- To introduce a methodical approach which combines agricultural processes and hydrological parameter for calculating water use intensity at farm scale.
- To describe a new modelling tool for water use intensity in agricultural production, which uses the advantages of MongoDB to build a flexible and transparent structure.

2 System description “AgroHyd”

2.1 Goal and scope

The water intensity, which is defined as the ratio of the agricultural production to the farm water use, is calculated at the farm-scale. The water intensity is based on the entire amount of freshwater required for producing the agricultural products and determined by using the method of life cycle analysis. The linked agriculture and hydrologic processes on-farm are considered within the system boundaries shown in figure 1. The system boundaries of the model are defined from “cradle to farm gate”. The model includes direct water flows, e.g. precipitation, tap water, irrigation water, transpiration, interception losses from plant leaves and mulch as well as evaporation from soil. The system also includes indirect water flows in background processes, i.e. mining of raw materials, production of materials, construction of farm buildings and machines, generation of electricity and imported feed. The cumulative water use includes the water used during the products use, disposal and recycling.

The farm system is divided into separate production units. That allows the assessment of different management strategies of farming on the water intensity of each unit. The main units are plant production and livestock farming, which are considered to be linked to each other but also linked within themselves. Plant production contains the processes cultivation, fertilisation, plant protection and harvest. Livestock farming includes the processes feed-supply, replacement, harvest. The direct water use for machines and technical facilities, buildings and storage is considered, as well as the indirect water used in their production.

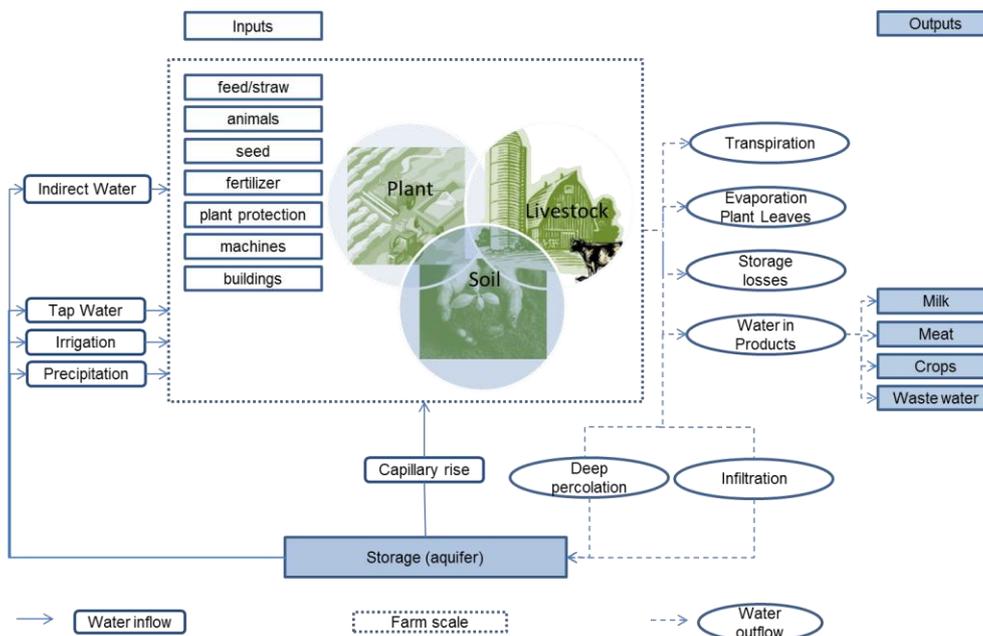


Figure 1. System boundaries and water flows

The water flows are currently calculated on a daily time step and aggregated over the period of time specific to the indicator of interest, e.g. water use intensity for vegetation periods or fiscal years. Gradual changes due to changes in management practices over longer time periods, e.g. the accumulation of humus in the field, can also be simulated with the ATB modelling database. In order to evaluate the indicators, a range of farm systems and regions will be investigated over several years to develop a basis for analysis.

2.2 Requirements for the modelling tool

The interdisciplinary nature of this study places extensive demands on the modelling system. Data input and process models from the hydrological perspective are necessary to determine the water flows within the farm production units. This requires that substantial data on local climate conditions be integrated into the model.

In order to calculate water flows from the agricultural perspective, the model includes specific information and process models to describe the plant and livestock production units as well as the interactions in the integrated farm system. The model database includes information on site and soil conditions of each farm as well as process models that can describe the influence of farm management strategies. The main management decisions considered are the plant species, seeding, cultivation, fertilisation, plant protection, expected yields and quality of crops, animal species, feeding rations, livestock housing, expected yield, replacement and service life of animals.

The compatibility to external databases (e.g. agricultural management software) and the easy integration of data from other databases (e.g. German Weather Service) has to be assured.

In order to evaluate farming processes worldwide, the modelling system has to be flexible enough to allow the addition of new processes and datasets to the database to reflect the diversity of the worldwide farming processes. The integration of GIS data has to be possible as well as the extension of the modelling system to cover aspects of water quality.

3 Software development - modelling tool

3.1 The basic principles

Despite the existence of many modelling systems already, a complete new software is being created to meet the needs of this interdisciplinary study and some of the criteria for new modelling systems discussed above. The technical basis for this modelling system is the quite new database MongoDB (MongoDB, 2012). A major advantage delivered by this database is its expandability: the first time data are written at runtime to the database, the structures are implicitly created. Therefore there is no need to predefine the modelling targets in advance. Another key advantage of this database is the use of embedded functions implemented with Javascript code, which will be used to a large extent within this project. No calculation is made by the client application (or data browser) on the local PC. Every computation takes place inside the database objects. Only one database query has to be performed by the client application to get the result of the model output.

3.2 Concepts for abstraction, model implementation and computation

The software is designed to be a generic modelling system, which can be used for any calculation where the system can be abstracted like:

- objects as datasets or functions
- processes as functions

Nearly any real world process can be mapped to these two basic elements because of the universality of a function. As a consequence, the data model of the complete modelling system can be divided into two parts: storage and computing (figure 2).

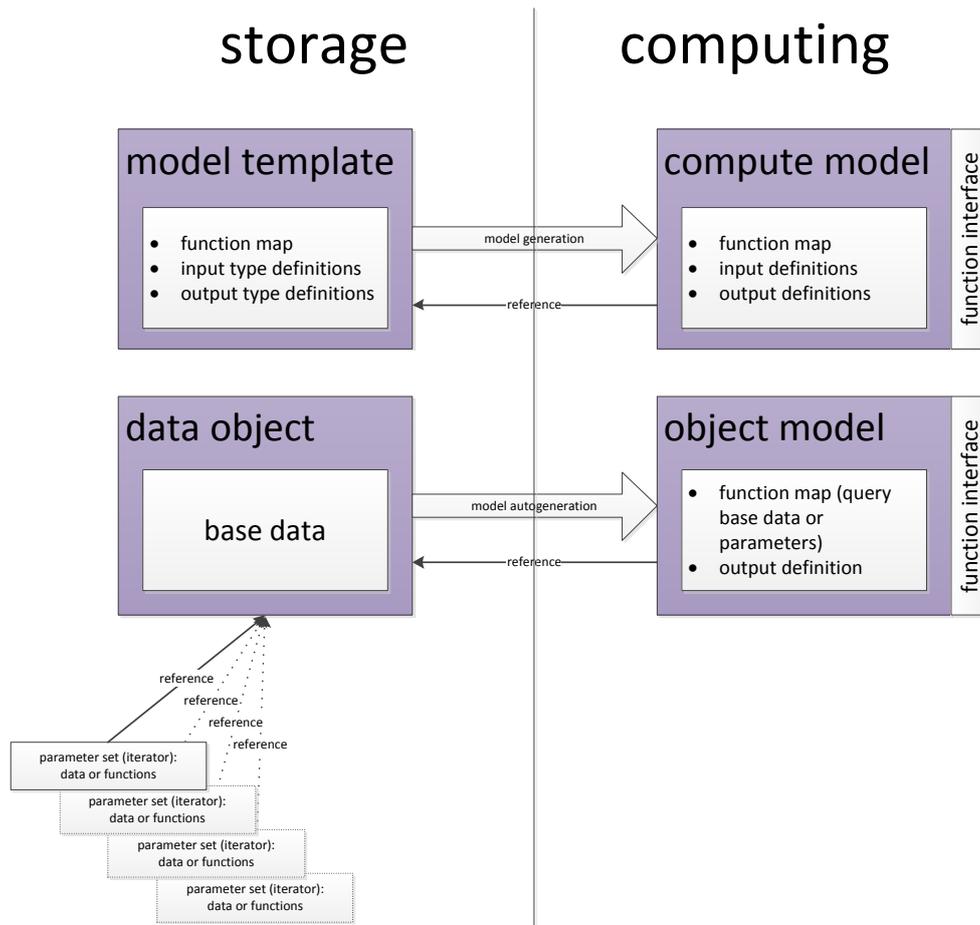


Figure 2: data model of the modelling system

Storage holds model templates and data objects. A model template is written for each equation in the farm model. Data objects are data collections associated with real world objects and made up of parameter sets. The client application (symbolized by the wide arrows in the figure 2) then generates the “compute model” from model template during the build of a concrete modelling run and automatically generates an “object model” for the case of a new data object. How these pieces fit into the database and how the modelling system is built is described in detail in the following sections.

3.2.1 data objects

Data objects of the present model are data collections such as climate values, soil profiles, statistical food data, geographical maps and so on. Data objects have in common that there is no input to this object. Data can be organized in any dimension (i.e. time, location). The argument to access this dimension is called an “iterator”. When creating a data object, an associated “object model” is automatically generated with the advantage that the data can be accessed via function calls. The arguments for these calls contain at least the iterator to get the associated data for this point (or date or location). The result is a completely transparent database object which carries out the database query automatically. If the amount of data for one object is too large to embed the data, it can be swapped to assigned parameter objects.

Correctional functions can be integrated into the data (parameter-) object. This is very useful to equalize the data and simplifies our application further, especially when using data from different sources in several formats or in different units. Since the compute model needs an exact defined format, a corrector function can be written and embedded in the data object. This enables the data to be converted

at runtime to fit the compute model requirements. The goal of this technique is to insure the integrity of the original data, which stay untouched. The conversion is runtime evaluated and the corrector function only has to be stored once to allow instant updates. Original data is never compromised and it is possible to add new datasets in the format of the source at any time.

Additionally alien data can be accessed as source with the embedded functions. This enables complex queries to remote databases. For example, an internet search engine query as data source may be exotic but it is possible.

3.2.2 model templates and computation

The backbone for the computations is the (growing) collection of model templates. These database objects are based on a skeleton with definitions for inputs, outputs and arguments which contain Javascript code for calculating the model functions. The embedded functions are implemented through the client application. In this way, common functions can be defined in model templates and combined as needed with other database objects. MongoDB built-in commands can be used within functions. Debugging is not as easy compared to Integrated Development Environment (IDEs) but possible. The result is a database object with all the model computations embedded with a well-defined function interface.

When this model template is used for an actual calculation, the input and output definitions are connected to real data models or other compute models and the template is copied to a compute model. At this point the client application is responsible for proper compute model generation and tying the connections together. In detail this means, that functions inside the compute model are generated to access the inputs.

There is one of the rare references in the system from the generated model to the template. It is used to be able to back annotate advancements from the compute model to the template.

The Javascript code i.e. the compute models are Just-in-time (JIT)-compiled and server side executed at runtime by the Javascript engine. This guarantees excellent performance. MongoDB can be installed with Mozilla's mature engine "spider monkey" or Google's "V8" engine.

3.2.3 function interfaces

Common to both model types (object and compute models, refer to figure 2) is the function interface. It can connect the models in the required way. It does not matter if data or computed results are accessed. A function call is universal and is possible to diversify inside the function. No differentiation of the access methods is used between the two model types, which leads to a further simplification of the modelling system.

Chains or networks directly out of our database objects can be built by using the instrument "function interfaces". The client application is the tool directing this process. Upon trying to connect inputs to outputs the function interfaces are verified for matching function- result- and argument types. This match is essential for a working function call. Therefore this process is done "type safe". Another benefit of function interfaces is that one can simply relay the function call and, for instance, do analysis work with the intercepted call.

3.3 Uncertainty in model calibration and validation

The modelling software is suited to analyze the effect of uncertainty on the model behavior. There are several ways to achieve this. In all cases, a compute model has to be created with inputs and outputs equal to the goal of the analysis, but with additional arguments. It is called an "analysis model". It is handled and integrated into the system like all other parts. There will be a template and, if used for computations, a compute model with connected inputs and outputs. When calling

an output function, the call is relayed to the input, but in between we can intercept and modify data to analyze our system.

At the current stage of software development, it has not been decided, if an external analysis framework will be called or if self-created algorithms will be implemented.

4 Conclusions

The ATB modelling database based on MongoDB is a flexible tool that can be used to calculate and predict the effect of farm systems on water use intensity. It is able to handle massive amounts of data, diverse formats and carry out parallel computing, which is needed for the interdisciplinary consideration of agricultural and hydrological processes within the farm system. Some of the innovations implemented in the modelling system that have proven their value are:

- accessing data via function call - using this approach insures the simplicity and universality of the modelling system;
- integrating correctional functions into the data object – this allows easy and consistent equalization of data with different formats or units, making it possible to add new datasets in the format of the source at any time;
- using function interfaces for the object model and the compute model – this results in even more simplification of the modelling system and gives the ability to intercept, manipulate and analyse object data;
- developing an “analysis” model on the basis of the compute model – this can easily be used to analyze the effect of uncertainty on the model behavior.

ACKNOWLEDGMENTS

The authors gratefully acknowledge financial support by the Senate Competition Committee (SAW) within the Joint Initiative for Research and Innovation of the Leibniz Association.

REFERENCES

- Allan, J.A., Virtual water: a strategic resource, global solutions to regional deficits, *Ground Water*, 36, 545-546, 1998.
- Berger, M., and M. Finkbeiner. Water Footprinting: How to Address Water Use in Life Cycle Assessment?, *Sustainability*, 2, 919-944, 2010.
- Famiglietti, J.S., Murdoch, L., Lakshmi, V., and R.P. Hooper, Towards a Framework for Community Modelling in Hydrologic Science. Report on 2nd Workshop: Blueprint for a Community Hydrologic Modelling Platform, March 31-April 1, 2009, Univ. of Memphis, Memphis, TN, 2010.
- Hoekstra, A.Y., and P.Q. Hung, Virtual Water Trade: A Quantification of Virtual Water Flows between Nations in Relation to International Crop Trade. In: *Value of Water Research Report Series 11*, UNESCO-IHE: Delft, The Netherlands, 2002.
- Jaeger, A., and T. Henrichs, Modelling environmental change in Europe: towards a model inventory (SEIS/Forward). European Environment Agency, 2008..
- Mila i Canals, L., J. Chenoweth, A. Chapagain, S. Orr, A. Anton, and R. Clift, Assessing fresh water use in LCA: Part I – inventory modelling and characterisation factors for the main impact pathways. *Int. J. LCA*, 14, 28-42, 2008.
- Molden, D., H. Murray-Rust, R. Sakthivadivel, and I. Martin, A water-productivity framework for understanding and action. In: *Kijne, J.W., R. Barker, and D. Molden (Eds.), Water productivity in Agriculture: Limits and Opportunities for Improvements*. CABI Publishing, Wallingford, UK, 1-18, 2003.
- MongoDB, <http://www.mongodb.org/> accessed 02/27/2012

- Moore, A.D., M.J. Robertson, and R.Routley, Evaluation of water use efficiency of alternative farm practices at a range of spatial and temporal scales: A conceptual framework and a modelling approach. *Agricultural Systems* 104.162-174, 2011
- Peden, D., G. Tadesse, and A. Misra, Water and livestock for human development. In: *Molden, D. (Ed.), Water for Food, Water for Life: A Comprehensive assessment of Water management in Agriculture. International Water Management Institute, Columbo, Sri Lanka and Earthscan, London, 485-514, 2007.*
- Teutsch, G. and E. Krueger (Eds.), Water Science Alliance White Paper: Priority Research Fields. Helmholtz-Zentrum für Umweltforschung – UFZ, Leipzig, 2010.
- UNESCO-WWAP, The United Nations world water development report 3: water in a changing world. The United nations Educational, Scientific and Cultural Organization, Paris, 2009.
- Wood, E.F., Roundy, J.K., Troy, T.J., Beek, L.P.H. van, Bierkens, M.F.P., Blyth, E., Roo, A. de, Döll, P., Ek, M., Famiglietti, J., Gochis, D., Giesen, N. van de, Houser, P., Jaffé, P.R., Kollet, S., Lehner, B., Lettenmaier, D.P., Peters-Lidard, C., Sivapalan, M., Sheffield, J., Wade, A., and P. Whitehead, Hyperresolution global land surface modelling: Meeting a grand challenge for monitoring Earth's terrestrial water. *Water Resources Research* 47. 10 PP. 2011.