International Environmental Modelling and Software Society (iEMSs)
2012 International Congress on Environmental Modelling and Software
Managing Resources of a Limited Planet, Sixth Biennial Meeting, Leipzig, Germany
R. Seppelt, A.A. Voinov, S. Lange, D. Bankamp (Eds.)
http://www.iemss.org/society/index.php/iemss-2012-proceedings

The Spatial Data Analyzer© as a model visualization and analysis tool

Qimiao Lu
Baird & Associates, 1267 Cornwall Rd, Ste 100, Oakville, Ontario, Canada L6J 7T5, qlu@baird.com
Thorsten Arnold
Drinking Water Source Protection, Saugeen, Grey Sauble, NBP Region
thorsten.r.arnold@gmail.com

Abstract: Landscape processes are dynamic in time. Models are readily available to describe these processes in time and space, and watershed managers worldwide are handling a great number and variety of models and high-dimensional data. With a variety of scenarios of each model type, and with increasing temporal resolution, model outputs quickly exceed the quantity that GIS applications can handle conveniently. Tools are needed that allow analyzing diverse models and their scenarios within a single software front-end, such that modelers minimize the time spent on quality control, peer review, data reformatting, error tracking and software handling.

The Spatial Data Analyzer© (SDA), developed by Baird & Associates, uses a toolbox for importing measurement data and model outputs into a format that allows rapid accessing of any amount of data, based on smart indexing and file access. The software facilitates model integration across disciplines, as it allows synchronous visualization, scenario analysis and translation of data between grids. We are looking for partners to identify needs of the integrated environmental modeling community.

Keywords: Visualization; Integrated Modeling; Data management; Knowledge management

1 INTRODUCTION

Landscape processes are dynamic in time. Data that describe these processes are time series of measurement stations (0D), streams or roads (1D), land use and other spatially distributed processes (2D), or volumetric processes (3D), such as the distribution of pore water content in soils, groundwater flow, surface currents, chemical concentrations, or met-oceanic data. Models are readily available to describe these processes in time and space with geo-temporal (or “dynamic”) data, and natural resource managers (NRMs) are handling an increasing number and variety of models.

In the last decade, Geographic Information System (GIS) has become an accepted tool in environmental management and related research. Centralized data storage, backup, access, and version control are key benefits within modern work processes. However, GIS databases are not optimized to handle frequent access to spatially-resolved time series data, for example the hourly output of a 3D hydrodynamic model.

The management of numerical models, especially for high-dimensional model data, is posing an increasing challenge to NRMs, who have to administer and supervise modeling contracts, ensure quality control, manage model updates, and sometimes have to defend information generated with models in court [Arnold, 2012, this conference). Environmental researchers must increasingly integrate multiple models from different knowledge domains, to analyze feedback between separate sub systems. Visualization and analysis software exists, but most programs are targeted to the needs of one disciplinary domain only, such that projects use multiple software tools for error checking, interpretation, analysis, and visualization. The need for such a multitude of tools has become a barrier for knowledge flow across departments because their use requires technical knowledge and licenses.

Spatial Data Analyzer© (SDA), developed by Baird & Associates [Baird, 2012, Lu, 2006], is a knowledge management tool for interdisciplinary model handling and analysis, for debugging and quality control, and for the visualization of data, as movies, and publication, as maps. The SDA software allows simultaneously importing and visualizing a wide range of GIS and time series data formats and virtually any amount of model output. It allows synchronous visualization, scenario analysis, and translation between grids. SDA uses a toolbox to transform model output into its generic data format that allows smart indexing of data files, such that file access is rapid and memory usage remains moderate. In this paper, we summarize the principles and functionality of the SDA software and present the generic data format. Then, we discuss potential uses of SDA for model management, integration, and model coupling.

2 SOFTWARE TOOLS FOR INTEGRATED MODELING

Environmental research increasingly relies on system thinking, and on the analysis of interactions between multiple system components. The integration of multiple quantitative models has become a preferred method of environmental scientists to apply this approach in research. Until now, the model integration process challenges researchers semantically, conceptually, technically and procedurally [Liu et al., 2008; Arnold, in press].

Expected potential benefits from integrated environmental modeling are great, so this method of environmental systems research has created a great diversity of pilot research projects. However, their use in mainstream praxis remains limited [Evert et al., 2005]. With prototype software, the costs for conceptual and technical integration remain prohibitive and results are neither timely nor of the quality anticipated [Argent, 2004].

Modeling teams must ensure conceptual and technical consistency across disciplines: i) scientific methods must be aligned coherently, ii) linkages of sub models across scales must be meaningful, and iii) model outputs must be interpreted in an integrated manner. Eventually, models can be coupled loosely based on data exchange, or in a more automated manner.

Information technology (IT)-overhead cost for handling multiple models and data formats can create a practical barrier for model users in this process. Until today, many projects rely on a multitude of data manipulation tools, data viewers and analyzing tools. Re-programming of similar tools and the knowledge needs for understanding these tools remain barriers to integration. Efficiency gains within the model integration process could greatly improve outputs and increase practical benefits from integrated environmental modeling research.

Streamlined software tools can simplify conceptual as well as technical steps, relieving modelers of tedious tasks like data parsing that constitute additional sources of error. Efforts to simplify the technical integration process have created technical interoperability solutions, such as the OpenMI, and data exchange standards, such as the self-documented NetCDF format. Users welcome that knowledge and experience gained with such standard can be reused in future projects [Knappen et al., 2011], even if totally different models are used. However, modelers consistently miss an environment for data analysis and visualisation of multiple model components [Knappen et al., 2011].

Recently, a more integrated generation of tools is emerging that streamlines tasks related to the handling of model input and output. These range from data downloading, data editing and re-formatting, data analysis, generic visualization of spatial time series, GIS functionality, diagnostic testing, verification, and even include higher-level modeling interface:

- The **HydroDesktop** software offers convenient access to hydrological data web services. For measurement station time series, HydroDesktop covers "data discovery, download, visualization, editing, and integration with other analysis and modeling tools" (Website). A recent extension of this package, the HydroModeler, offers a component-based modeling environment for loose coupling, using the OpenMI [Goodall et al., 2011, Amesa et al., in press].

- **SimEnv** [Flechsig et al., 2008] is a parameter sampling-based multi-run simulation environment for models with multi-dimensional output. A beta version of **SimEnvVis** now offers an impressive diversity of visualization options for the analysis of model results from multi-scenario runs [Nocke, 2007].

- **DelftShell** offers a "higher level integrated modeling system targeting the interaction with the end user via a graphical user interface" [Donchyts et al., 2008], as a modeling environment that combines elements of GIS and Data Management, but also provides direct linkage to models through an OpenMI interface.

- The **OpenEarth** offers a platform for research, consultancy and construction projects that integrates data, models and tools in marine and coastal engineering, in order to reduce project costs for basic software infrastructure. For visualization of spatial data, it offers an interface to GoogleEarth©.

The Spatial Data Analyzer© was designed as a one-stop-shop tool for efficient post-processing of high-dimensional model outputs. Tasks range from debugging of modeling software, error analysis and quality control of model applications, data analysis, simultaneous visualization of multiple data types, animations, and (template-based) publishing of maps. SDA uniquely combines several features for model integration. SDA's strength is a user interface that resembles GIS software, which allows visualizing (and publishing) several large, 3D dynamic data sets simultaneously. As is, SDA offers several advantages to environmental modelers who handle complex, coupled models, especially as single model output and data viewer.

3 THE SPATIAL DATA ANALYZER© (SDA)

3.1 The Philosophy of SDA

The SDA offers a graphical user interface similar to conventional GIS tools, with added functionalities for time series visualization and data analysis functions that modelers frequently utilize (Figure 1). The drag-and-drop user interface reduces the entry barrier for new users. Similar to GIS software, SDA organizes multiple layers of vector and image files in a project organizer window and visualizes these in a separate frame, allowing modifications to color palettes, labeling, and other

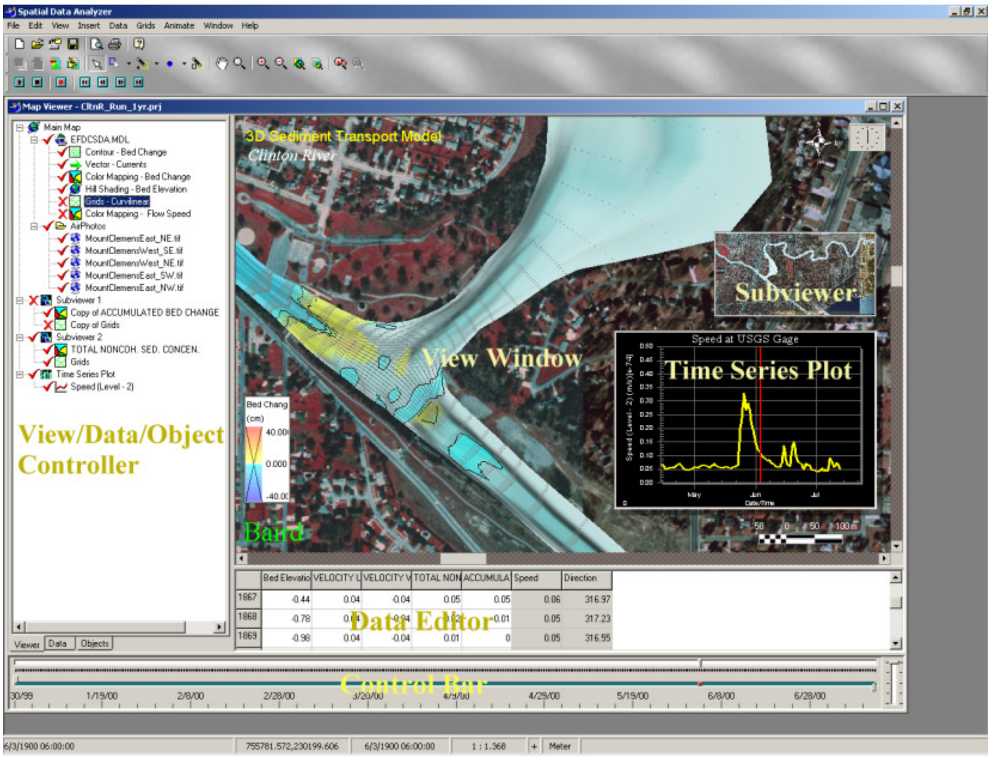

Figure 1 Graphical user interface of SDA

attributes. A toolbar offers zooming, measurements of distance and area, and other functions known from GIS.

Multi-dimensional time series data is managed in a separate data organizer. The data organizer provides the graphic user interface for users to perform a variety of data analysis. Geo-temporal data can be selected and visualized together with "static" GIS data, as color-coded grid and concentration map, contour lines, vectors, barbs, or hill shading. The user can scroll through the data time series via a time bar, such that geo-temporal data from one or multiple models are visualized immediately. A second vertical bar allows scrolling through horizontal layers of 3D grids in vertical direction. The users can narrow the time periods of datasets for the animations and movies it records using an additional temporal control bar. Any

number of variables can be visualized simultaneously, for example using different coloring or display modes. Furthermore, the user can compute custom variables from existing variables using of a wide range of mathematical operations. Just as GIS systems can open many layers, the SDA links to multiple geo-temporal databases simultaneously.

Time series graphs are X-Y-plots at a point location. Such points can be extracted ("probed") from these multi-dimensional time series, or imported as time series at a measurement station. Vast options exist for controlling graph properties, the time intervals, including the use of multiple y-axes and staggered plots. X-Y-plots are useful for comparing model data with measurement data at fixed locations, for example during calibration and quality control.

Tracking data, as a separate data format, describes moving objects and its associated information, such as ship observation data, air traffic data, and fish tracking telemetry data. Tracking data is defined as an ASCII table with location (x,y,z), time and date, and the quantity of one or multiple variables.

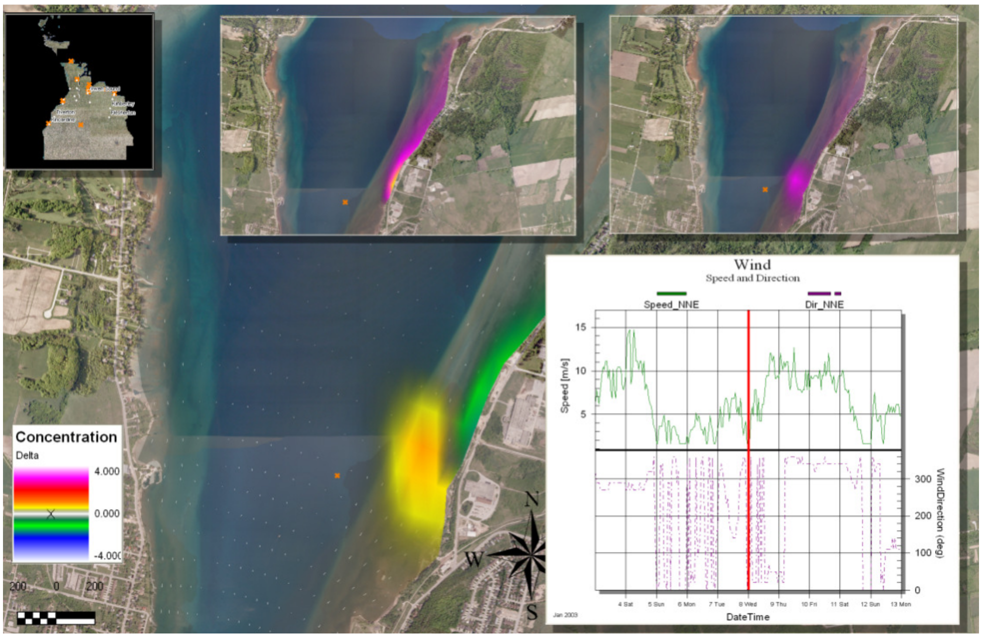

Figure 2 Difference between two flow scenarios computed and visualized with SDA

The visualization frame can be subdivided into several components or submaps. Four types of submaps are available: the regular sub viewer, an index map, profile viewers, and time series plots. Each submap can contain a GIS background plus any geo-temporal data, and is timed simultaneously with the main map via the time bar. Sub viewers can simultaneously depict different geographical scales, multiple scenarios of the same model, or totally different models.

An object toolbar offers specialized objects (north arrow, legend, scale, clock, icons, text box), control of submaps, specialized plots (rose plots, splatter plots) and profiles (grid cross-sections). For the latter, a probing function allows extracting points or cross-sectional profiles from high-dimensional data along drawn lines or along the model grid.

Graphical capabilities have been described in Lu [2006]. SDA offers color mapping with partial transparency, hill shading, contour maps, vector layers, tracking and barb layers, overlay with GIS shape files and images, as well as any combinations of these. For example, Figure 2 shows two spill scenarios (top submaps), a difference layer (main map), in combination with an index map (top left) and a XY-plot of wind speed and direction.

3.2 Data analysis options

SDA offers several tools that conveniently help users to edit, extract, analyse, and process dynamic data in space and time. From a geo-temporal and high-dimensional dataset, users can extract time series at any grid points, or at profiles using the drag-and-drop operation, without re-running the model. To save disk space, probed data remains stored in a memory buffer until it is explicitly exported. Furthermore, users can define a new variable using mathematical equations with a math parser.

The SDA data analysis toolbox offers several complex functionalities which are common to many model applications, and was designed such that future applications can be integrated flexibly. For time series analysis, SDA offers aggregation using mean, count and sums, basic statistical and frequency distribution, correlation between two dataset, and a flexible math editor for user-defined analysis. For hydrological applications, it analyses flow time series and provides flow duration curves. The dynamic data tools for geo-temporal data offer basic statistical analysis and several complex functions, such as balancing flow through a cross-section, or storm analysis. Furthermore, custom-designed analysis allows combining several model scenarios that use the same grid, for scenario comparison or to derive time series of complex index variables. For nested modeling with multiple grids or file-based model coupling, a grid-to-grid extraction tool translates data conveniently. Finally, forward and backward particle tracking allows tracing spills within vector fields. Particle properties can be specified such that no further modeling is necessary to delineate time-of-travel zones.

3.3 Other features

SDA projects can be exported into a variety of media, such as on-screen play, off-screen media record, high-resolution image export, and high-quality print. On-screen animations are either played automatically at a determined speed, or interactively by moving the time slider forward and backward. Off-screen movies can be recorded in AVI and MPEG format, and played outside of the SDA environment during presentations. The visualization frame can be exported as an image with variable resolution, for posters or publications. Map templates can be customized for automated creation of maps in a consistent layout.

3.4 The SDA data format

The SDA data format for storing spatially resolved time series is optimized for rapid, index–based access needed for visualization. Each model run is stored as a bundle of three file types, described below (see Figure 3):

1. The model description file (see component 1 in Figure 3) references grid type, definition of variables (number of variables, their units, min, max), and the location of the grid file and main binary data file;
2. A grid file (see components 2, 3a and 3b in Figure 3) is a text meta file that specifies the model grid and associated static (not varied with time) variables. For hydrological model output, the grid can be associated with the shape file, which can be specified in the grid file;
3. The binary data file (see component 4 in Figure 3) stores equally-sized blocks of data for each time step, preceeded with a date number.

This simple format allows rapid data access using index pointers with less memory consuming, such that data is loaded only as needed. Data can be stored locally on hard drives, or on network drives. Locations of all files used in an SDA project are referenced within the .xml project files.

Model output must be transformed into this SDA format. There are three options: the SDA toolbox automates the conversion of several registered models and common data formats, such as netCDF, a text wizard allows to custom-define import rules for ASCII tables, otherwise, users can program stand-alone scripts that read model input and output, and write it as binary data. Once imported, SDA offers an export function of selected time step and variables into ASCII tables.

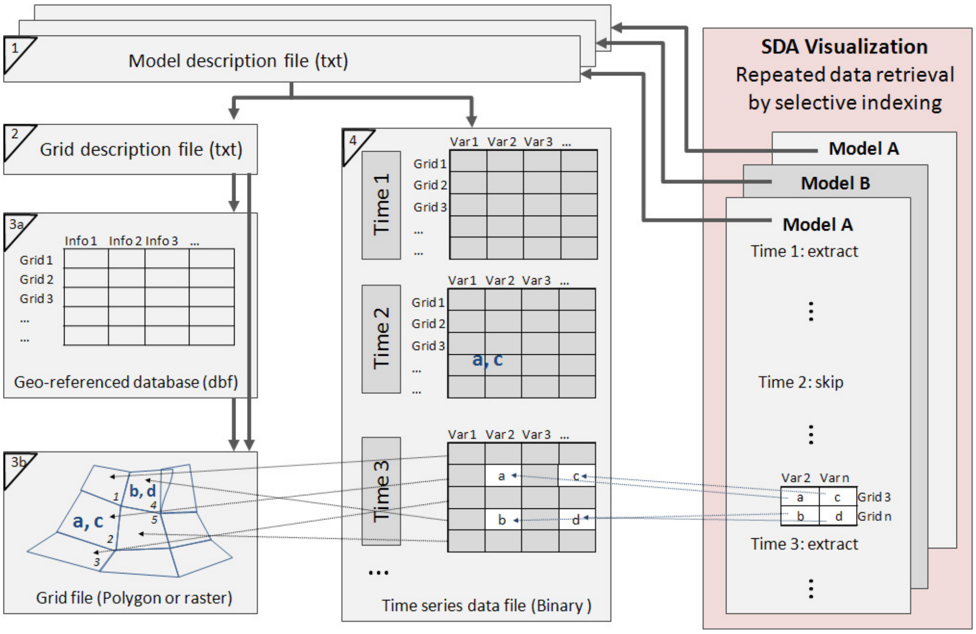

Figure 3 The SDA Data format for model data is optimized for index-based access for rapid visualization

1 OUTLOOK

The authors are currently assessing options to use the SDA software as a model management tool for public managers of natural resources, as well as assessing the requirements and needs from the integrated environmental modeling community.

2 REFERENCES

Amesa, D.P. and J.S. Horsburgh, Y.Cao, J. Kadlec, T. Whiteaker, D. Valentine HydroDesktop: Web services-based software for hydrologic data discovery, download, visualization, and analysis. *Environmental Modelling & Software*, in press

Arnold, T. Procedural knowledge for integrated modelling: Towards the Modelling Playground. Accepted for the thematic issue on the future of integrated modelling science and technology. *Environmental Modeling & Software* , in press

Arnold, T. Model Management for watershed practitioners: Options for a rural watershed in Ontario, Canada. 2012 Int. Congress on Env. Modelling and Software: Managing Resources of a Limited Planet, Leipzig, Germany. R. Seppelt, A.A. Voinov, S. Lange, D. Bankamp (Eds.)

Argent R.M., An overview of model integration for environmental applications – components, frameworks and semantics. *Environmental Modeling & Software 19(3).* 219–234, 2004.

Baird. Product Website; http://www.bairdsoftware.com/bairdsoftware/en_html, 2004.

Donchyts, G.; Baart, F.; Jagers, B., DelftShell - Integrated Modeling Environment with elements of GIS, Data Management and OpenMI support. *American Geophysical Union*, Fall Meeting 2008.

Evert F.K. v., Holzworth D., Muetzelfeldt R., Rizzoli A. and Villa F., Convergence in integrated modeling frameworks. *Tech. rep., Wageningen University* [http://library.wur.nl/oai] (Netherlands), 2008.

Flechsig, M., Böhm, U., Nocke, T., Rachimow, C.,The Multi-Run Simulation Environment SimEnv. Version 2.01. See http://www.pik-potsdam.de/software/simenv/, 2008.

Goodall, Jonathan L. and Anthony M. Castronova., Component-based Modeling using the HydroModeler Plug-in. Civil and Environmental Engineering University of South Carolina, *CUAHSI HIS Webinar Series*, November 7, 2011.

Knapen, M.J.R. and S.J.C. Janssen, O.R. Roosenschoon, P.J.F.M. Verweij., Evaluating and improving OpenMI as a model integration platform across disciplines *19th International Congress on Modeling and Simulation, Perth, Australia, 12–16* 2011. http://mssanz.org.au/modsim2011, 2011.

Liu, Y., Gupta, H., Springer, E., and Wagener, T., Linking science with environmental decision making: Experiences from an integrated modeling approach to supporting sustainable water resources management. *Environmental Modeling & Software 23 (7)* 846-858, 2008.

Lu, Q., SDA: An efficient GIS-based animation tool for modeling studies. In: Coastal hydrology and processes: Proceedings of the AIH 25th Anniversary. Ed. Vijay P. Singh, Y. Jun Xu, Water Resources Publications LLC, 2006.

Nocke, T., Visuelles Data Mining und Visualisierungsdesign für die Klimaforschung, *PhD thesis,* University of Rostock (in German), 2007.