GRASS/ OSGeo-News

Open Source GIS and Remote Sensing informations Volume 1, June 2004

Spatial Analysis and Modeling Tool

A tool for spatial simulation

Ralf Wieland

Abstract

This paper presents an overview of the integrated spatial simulation tool SAMT. SAMT was designed to integrate models from different sciences such as economic and ecology, necessary for the evaluation of sustainability of landscapes. Traditional geographic information and modeling systems have specific limitations. To overcome such limitations, a new tool was developed to focus specifically on interactive spatial simulation. Of particular interest are methods for incorporating expert knowledge (fuzzy models) and for extracting models from examples (e.g., using neural networks).

Structure and Main Functions

Modern GIS tools like ARGGIS (ARCGIS , 2008) and GRASS (GRASS , 2008) are designed to perform the following functions:

- Store spatial information
- Perform spatial analyses
- Present spatial information (view maps, print maps, 3D-view etc.)

These software packages are large, and in the case of ARGGIS, expensive. Tools like MATLAB (MATLAB , 2008) or its free counterpart OCTAVE (OCTAVE , 2008) seem better-suited for numerical simulations; using an extra toolbox with commensurate expense and programmer skill, MATLAB also handles spatial data sets. SAMT fills the gap between GIS and MATLAB. As a small, grid-based simulation system, SAMT was designed to be interactive and fast. Among its strengths, SAMT can

- Read and write ASCII grids from ARGGIS and efficiently store the grids using HDF (HDF , 2008). Compared to an ASCII grid, map I/O is about 15 times faster using HDF.
- Apply modern algorithms to grid data sets (fuzzy, neural networks, wavelets, etc.).

Contents of this volume:

- Integrate different models into a system.
- Run under the Linux operating system.

Figure 1 shows the structure of SAMT. At its core are general functions like simple grid operations (generate a random grid, normalize a grid, exchange one selected value of a grid with another, etc.) and analytical functions (statistics, histograms, etc.) Additional modules have been implemented for fuzzy modeling, neural networks and dynamic simulation; there is also a module designed to link user-defined (C++) models to SAMT. Each of these modules is called as a separate process, which allows for rapid model development; an error in a given module has no influence on the core. Data transfer between core and modules is based on the fast HDF technology.

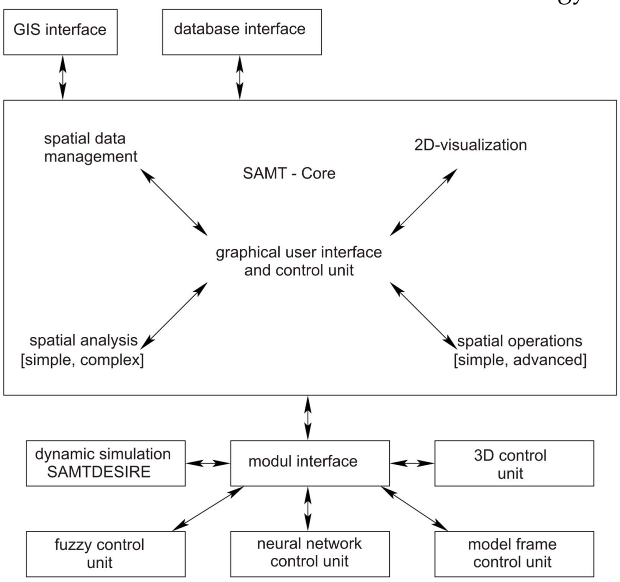

Figure 1: The SAMT structure consisting of the core and different modules

SAMT is implemented as a graphical user interface using the Qt library (QT , 2008). Menu options include:

- I/O Operations - Transfer ASCII Grid files between SAMT and ARCGIS, access HDF technology, export grids as a .png image, import point themes stored in CSV (comma separated value) format, import fuzzy models and neural networks
- Project Management - Create, open and store projects (which are essentially an aggregation of grids, points and models)
- Model Operations - The fuzzy development system, the neural network development tool, and some export functions to the statistical computing software R (R , 2008)
- Analysis - Elementary analysis functions (e.g., histogram of the values in a selected map, charts of a freely-positioned transect like elevation, or power spectra over this transect), statistics (min, max mean, standard deviation of a selected map), and enhanced methods such as the analysis of a fuzzy model using grid data
- SGRID - Simple operations over one grid (normalization of a grid, different cut function, etc.), as well as two more demanding functions: the multiple subarray algorithm (Takaoka, 2002) (used to find changes between simulation results) and a tree algorithm that aggregates integer-coded grid cells according to neighborhood statistics (used, for example, to find soil communities)
- AGRID - Operations over two grids (add, multiply, etc.), as well as more complicated operations such as moving window-based functions
- Special Operations - Some operations from image processing (wavelets, Fourier transformation, etc.), functions to transform point themes in grids (Voronoi maps, spatial interpolation, etc.), and some preliminary watershed algorithms including flood fill
- AddON-View - Provides options for controlling the view (black and white vs. color, grid lines, etc.)
- 3D - Three-dimensional elevation and splatter views (Schroeder et al., 1998) to visualize four input grids (one per dimension plus color). The splatter view can be used as an analytical tool before training a neural network. SAMT developers intend to enhance 3-d functionality in future.
- TABLE - Read and store tabular information using either CSV or database tables (MySQL , 2008)

Figure 2 shows a screenshot of SAMT:

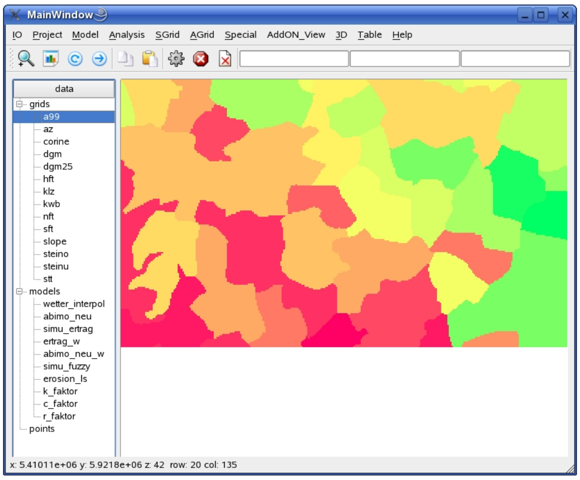

Figure 2: Screenshot of SAMT

From the toolbar, the user can copy or remove a grid, start a view, delete a grid, etc. To the right of the toolbar are three line inputs used as parameters for the models. Parameter usage is explained in the status line. This is also available for user-defined models imported into SAMT, and can be used as a reference when calling the models. Help is provided by the help system.

SAMT extends its selection of mathematical grid functions with the mathematical parser muParser (muParser , 2006). This parser supports the application, on up to three grids, of mathematical operations like sin, cos, tan, shinh, cosh, tanh, sqrt, log, exp, etc. After parsing, a virtual machine does the calculation. For example, two grids (a and b) will be combined cell by cell using $resultgrid_i = \sqrt{a_i * a_i + b_i * b_i}\ \forall i$. For two grids with 5000*3651 cells on an Intel core, 2 cpu 6400@2.13GHz processor with 2 GB main memory, this takes about 1 s. Figure 3 shows a screenshot of the mathematical parser:

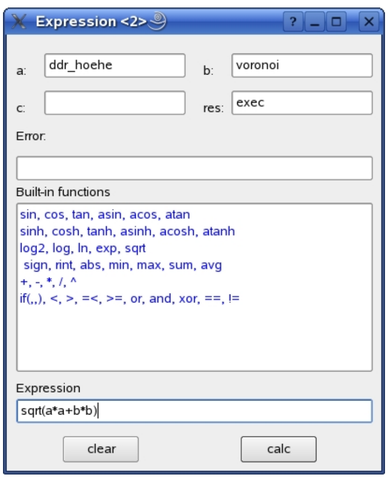

Figure 3: Screenshot of the parser

Fuzzy Modelling

One of the main reasons for developing SAMT was the need to apply fuzzy models to maps for calculating habitat quality Wieland et al. (2002). Such modeling was used, for example, to optimize the location of wind power stations with respect to the habitat of protected birds. The expert (a biologist, in this case) wanted to apply his knowledge to the maps in a natural way. SAMT provides a separate development toolbox SAMT_FUZZY to help the modeler transform expert knowledge into fuzzy sets and fuzzy rules. This toolbox can be used to evaluate the fuzzy model using data sets provided by the modeler, as well as to train fuzzy rules using such data sets. A graphical user interface allows the input and the adaptation of the fuzzy sets and makes it easy for the modeler to optimize his or her model. The following Figure 4 shows the analysis window of the SAMT_FUZZY:

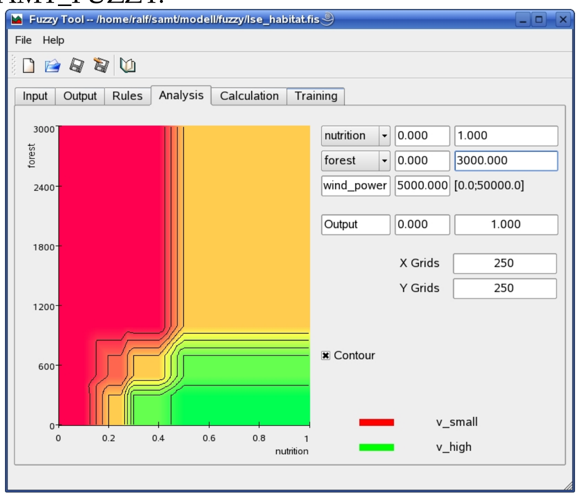

Figure 4: Analysis window of SAMT_FUZZY

The modeler can click in the graphic and get a table with all information about the selected point: The values of the inputs, the rule number, the rule text, the strength of the rules, the output of the rules and the calculated combined output of the active rules. This allows for a fast check of the rules and a possible adaptation. The fuzzy sets of all inputs can be adapted in another window using the mouse or keyboard. During development, this procedure of rule checking and adaptation is quite reasonable, but it is not sufficient for evaluating fuzzy models with real spatial data sets. To perform such an analysis SAMT has included two additional tools, one for fuzzy rule checking as described (but now with the real spatial data sets) and a second for sorting the rules according to their frequency of use, with provision for selected rule modification. This last tool can be used to check the region of influence of selected rules in the real spatial data set.

Neural Networks

An artificial neural network (ANN) is used to create a nonlinear regression or classification model using a set of data provided by the user. As an alternative to user-provided training data, SAMT can generate training data from maps. Up to three maps may be used as inputs; another map is generated as output. SAMT provides a variety of ANNs: Self-organizing maps (SOM) (Kohonen, 2001) that classify data using unsupervised training, radial basis function (RBF) (Korn, 2004), and the feed forward networks (Principe et al., 2000). The feed forward networks are the most common ANNs, inspiring the development of a separate toolbox SAMT_NN with the following features:

- Up to three inputs within SAMT (if used standalone, up to 30 inputs)
- One hidden layer (maximum 50 nodes), one output
- Training algorithm: Back-propagation (Freemann, 1992), Levenberg-Marquardt (Gershenfeld, 2002) and Levenberg-Marquardt with regularization (Bishop, 1995)

Regularization is useful to avoid overtraining, where an ANN can reproduce the training results quite well but has lost its ability to generalize (e.g., it cannot be used with data sets from different locations or time frames).

For input, SAMT_NN supports comma-separated values (CSV), a compromise with respect to the variety of common data formats. The input data can be used as a training set, or can be randomly split into training and validation data sets. The user can select the number of hidden nodes, the training algorithm, the number of training steps, etc. After training, SAMT_NN provides some tools for evaluating the results. A plot of model output versus target values provides initial feedback about the trained ANN. A functional plot of one selected input with respect to as many as 10 classes of another input is important for comparing the behavior of the trained ANN with the experience of the modeler. This is important in order to find errors in the trained ANN (an ANN starts training with random weights and can get caught in a local optimum, so sometimes retraining with different starting weights can be helpful). An example of such a functional plot is given in Figure 5.

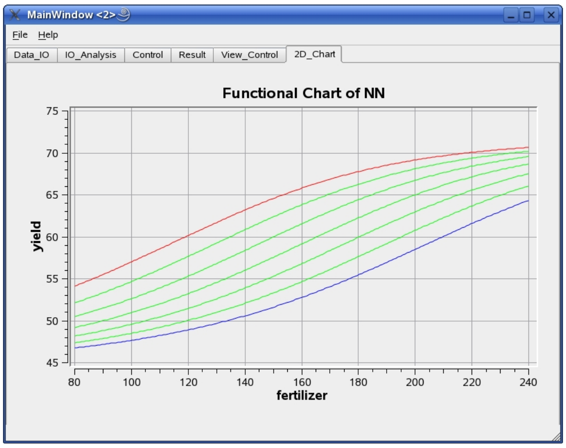

Figure 5: Functional plot of an ANN for yield=f(soil,fertilizer,climate) with $soil \in (32, 35.5, 39, 42.5, 46, 49.5)$ soil classes and climate representing the potential climate water balance of one year (-32 mm in this example)

The ANNs trained with SAMT_NN can be exported to SAMT and used as models (SAMT includes a library that provides the methods to open and execute a trained ANN). Alternatively, trained ANNs from SAMT_NN can be used in C++ outside of SAMT.

SAMT_DESIRE

SAMT_DESIRE (Korn, 2007) is a dynamic simulation system like MATLAB's Simulink (MATLAB , 2008). SAMT_DESIRE can read typed or programmed equations and differential equations in a natural mathematical notation. On a "drun" command, SAMT_DESIRE then immediately compiles and solves the problem with a choice of integration rules, producing time-history graphs or listings. Similar programs have long been used for simulating dynamic systems like aerospace vehicles, but SAMT_DESIRE is particularly convenient for interactive modeling. Fast code for time-history solutions is compiled at runtime, but interpreted commands like

```
reset
gain=gain + 0.2
drun
```

let you modify your model on the fly and then try a new simulation run. Such interpreted commands can also be combined into useful experiment-protocol programs (scripts) that control "multirun simulation studies", e.g. for statistics evaluation (Monte Carlo simulation), and for parameter optimization or model identification.

The dynamic simulation module SAMT_DESIRE opens up a new set of possibilities within SAMT:

- Provides a sophisticated modeling system for dynamic models
- The included vector compiler can be used for model replication and for Monte Carlo simulation.
- Using an interface between SAMT and SAMT_DESIRE allows the user to copy results of a dynamic simulation to maps at any point during a simulation. time. Korn et.al. (2005)
- SAMT_DESIRE supports rapid prototyping of models; its outputs can be used to check those of simulation modules coded in C++.

Additionally, SAMT_DESIRE allows the use of fuzzy models or ANNs produced by SAMT_FUZZY or SAMT_NN in a dynamic simulation. As such, SAMT_DESIRE is the dynamic counterpart to SAMT.

Conclusions and Future Work

SAMT is an open source software package using open source libraries (gnu scientific library (GSL , 2008), Qwt (QWT , 2008), etc.) and is implemented under the Linux operating system. A test CD is available at the project home page (SAMT , 2008), and contains a copy of SAMT, all the libraries, and the SAMT source code. This CD can be used as a live system, with data accessed from a USB stick or via a network connection. Another option is to copy the live CD on the hard disk (this can damage your file system - please be careful). Another way to install SAMT on your computer is to compile and install all libraries before installing SAMT. This is not difficult, but requires good Linux skills. A Windows version is also planned.

SAMT and its components are used by small groups of modelers in a variety of different countries. The graphical user interface helps beginners use this software (SAMT was developed as an interface around a spatial library). It is still a simulation system, as opposed to a GIS package. SAMT includes some GIS methods in order to facilitate working with spatial data sets, but is limited in its presentation of spatial data sets and handling of vector data sets.

Plans for the near future include implementation of a new neural network toolbox based on the Fast Artificial Neural Network library (FANN) (FANN , 2008) and inclusion of a support vector machine (SVM) (Chang and Lin , 2007) as an alternative method for modeling. The SAMT_NN and the new neural network toolbox will exist independently. Subsequently, fuzzy training will be improved and new three-dimensional views will be added.

Many thanks to all the developers of the open source software upon which SAMT is constructed, and to the developers of the open source software that was used during SAMT development (gcc, emacs, Linux etc.)

This contribution was supported by the German Federal Ministry of Consumer Protection, Food and Agriculture, and by the Ministry of Agriculture, Environmental Protection and Regional Planning of the Federal State of Brandenburg (Germany).

Bibliography

ARCGIS Available from: http://www.esri.com/software/-arcgis/index.html

GRASS Available from: http://grass.osgeo.org/

MATLAB Available from: http://www.mathworks.com/

OCTAVE Available from: http://www.gnu.org/software/octave/

HDF (Version 1.4.5) Available from: http://hdf.ncsa.uiuc.edu/HDF5/

QT-Libary Available from: http://trolltech.com/products/qt/

statistic software R Available from: http://www.r-project.org/

Takaoka, T. (2002): Efficient Algorithms for the Maximum Subarray Problem by Distance Matrix Multiplication, Proceedings of the Australasian Theory Conference, CATS02, (2002) 189-198

Schroeder, W.; Martin, K.; Lorenzen, B., 1998. The Visualization Toolkit. Prentice Hall PTR.

MySQL Available from: http://dev.mysql.com/downloads/

muParser Available from: http://muparser.sourceforge.net/

Wieland, R., Voss M., 2002. Land use change and habitat quality in Northeast German agro-landscapes. In: Tenhunen, J. D., R. Lenz, R. Hantschel (Eds), Ecosystem approaches to landscape management in Central Europe, Springer: Berlin 341-346.

Kohonen, T., 2001. Self-Organizing Maps. Springer

Korn, G.A., 2004. Model Replication Techniques for Parameter-influence Studies and Monte Carlo Simulation with Random Parameters. Math. and Computers in Simulation 67 (6) 501-513.

Principe, J.; Euliano, N.; Letebvre, C., 2000. Innovating Adaptive and Neural Systems Instructions with Interactive Electronic Books. Proceedings of the IEEE, special issue on engineering.

Freemann, J.A., 1992. Simulating Neural Networks with Mathematica. Addison-Wesley.

Gershenfeld, N., 2002. The Nature of Mathematical Modeling. Cambridge University Press.

Bishop, Ch., M., 1995. Neural Networks for Pattern Recognition. Oxford University Press.

Korn, G.A., 2007. Advanced Dynamic-system Simulation. Wiley, N.Y.

Korn, G.A.; Wieland, R.; Holtmann, X., 2005 Spatial Analysis and Developing Tool - a New Software for Landscape Analysis. Simulation News Europe 44/45 2005

SAMT (Version 2.8.x) Available from: http://www.samt-lsa.org

FANN (Version 2.1) Available from: http://leenissen.dk/fann/

Chang, Ch.-Ch., Lin, Ch-J., LIBSVM: a Library for Support Vector Machines, Available from: http://www.csie.ntu.edu.tw/ cjlin , 2007.

Gnu scientific library Available from: http://www.gnu.org/software/gsl/

Qwt-5.0.2 Available from: http://sourceforge.net/projects/qwt

Ralf Wieland
ZALF Institute for Landscape Analysis
`http://www.samt-lsa.org`
`rwieland@zalf.de`

New Publications

the title of the publication

Herborg et al. (2003)

the abstract/description goes here

the title of the next publication

Lurz et al. (2001)

the abstract/description goes here

Bibliography

L.M. Herborg, M.G. Bentley, A.S. Clare, S.P. Rushton (2003) The spread of the Chinese mitten crab (Eriocheir sinensis) in Europe; the predictive value of an historical data set. *Hydrobiologia* 503: 21-28.

P.W.W. Lurz, S.P. Rushton, L.A. Wauters, I. Currado, P. Mazzoglio, M.D.F. Shirley (2001) Predicting grey squirrel expansion in North Italy: a spatially explicit modelling approach. *Landscape Ecology* 16: 407-420.

Recent and Upcoming Events

the title of the event

a possible subtitle of the event

City, country, date

the text goes here

Your Name
Your organization
`yourhomepage`
`mail AT mail net`