



POLISH ACADEMY OF SCIENCES
Systems Research Institute

ECO – INFO
AND SYSTEMS RESEARCH

Editors:

Jan Studzinski
Olgierd Hryniewicz



ECO – INFO AND SYSTEMS RESEARCH

Polish Academy of Sciences • Systems Research Institute

Series: SYSTEMS RESEARCH

Vol. 52

Series Editor:

Prof. Jakub Gutenbaum

Warsaw 2006

ECO – INFO AND SYSTEMS RESEARCH

Editors:

**Jan Studzinski
Olgierd Hryniewicz**

The purpose of this publication is to present the information technology (IT) tools and techniques that have been developed at the Systems Research Institute of Polish Academy of Sciences in Warsaw (IBS PAN) and at the German Institute for Landscape System Analysis in Müncheberg (ZALF) in the area of applications of informatics in environmental engineering and environment protection. The papers published in this book were presented in the form of extended summaries during a special workshop organized by IBS PAN in Szczecin in September 2006 together with the conference BOS'2006 organized jointly by IBS PAN, University of Szczecin, and the Polish Society of Operational and Systems Research. In the papers the problems of mathematical modeling, approximation and visualization of environmental variables are described. Moreover, some questions concerning the environmental economy are also presented.

Papers Reviewers:

Prof. Andrzej Straszak

Dr. Lucyna Bogdan

Text Editor: Anna Gostynska

Copyright © Systems Research Institute of Polish Academy of Science,
Warsaw 2006

Systems Research Institute of Polish Academy of Science
Newelska 6, PL 01-447 Warsaw

Section of Scientific Information and Publications
e-mail: biblioteka@ibspan.waw.pl

ISBN 83-894-7509-X

9788389475091

ISSN 0208-8029

CHAPTER 1

Specialized Software



SPATIAL ANALYSIS AND MODELING TOOL V 2.0 - SYSTEM DESIGN

Ralf WIELAND, Wilfried MIRSCHEL, Karl-Otto WENKEL

Centre for Agricultural Landscape Research (ZALF)
Institute of Landscape Systems Analysis,
<rwieland@zalf.de, wmirschel@zalf.de, wenkel@zalf.de>

***Abstract:** This paper presents an overview of the integrated spatial simulation tool SAMT. Land use issues are intrinsically multidisciplinary, involving a number of factors acting interactively, leading to landscape and land use changes. Therefore, an analytical tool for land use simulation must be integrated in nature, combining ecologic and economic variables and must provide framework for their integration. Traditional geographic information systems or modeling systems have specific disadvantages. To overcome these disadvantages, a new tool was developed with the focus of interactive spatial simulation. Included in SAMT is expert knowledge (fuzzy models) or extract models for examples (neural networks). SAMT provides a framework to include dynamic models. One of the most powerful extensions of SAMT is the integration of the simulation toolbox SAMTDESIRE. SAMDESIRE allows the development of models, refinement and parameterization of models and perform simulation. These is combined with a powerful graphic output and a graphical control interface. An interface between SAMT and SAMTDESIRE is under development. SAMT itself and SAMTDESIRE is open source software and freely available.*

Keywords: Geographic Information Systems, Neural Networks, Fuzzy-Modeling, Dynamic Simulation System, Land Use, Ecosystems.

1. Introduction

Landscapes are important for social, economic and industrial development of a nation. The usefulness of landscapes includes habitat for human beings, flora and fauna, sources for underground water both in quantity and in quality, as well as medium for agricultural production. These landscape functions are strongly connected with the type and intensity of land use,

as a result of complex network of economic, social, biotic and abiotic interactions.

Land use issues are intrinsically multidisciplinary, involving a number of factors acting interactively, leading to landscape and land use changes (<http://portal.mtt.fi/pls/> ...). Therefore, an analytical tool for land use simulation must be integrated in nature, combining ecologic and economic variables and must provide framework for their integration. The representation of data from different disciplines in a common methodological frame of reference is a good way of ensuring this integration. Given the multidisciplinary nature of landscape and land use issues, and given that these issues involve the interplay of ecologic and economic factors, the challenge therefore, is how to couple ecologic and economic models and analyze the multi-dimensional issues and data involved in land use studies. To perform such an model integration a modeling and simulation toolbox is necessary. Such a toolbox must be able to handle spatial and temporal data sets, execute the simulation and provide methods for interpretation of the results.

1.1. Models in GIS - spatial data in modeling systems

Traditional GIS like ARCGIS (www.esri.com/software/arcgis/) or GRASS (<http://grass.itc.it/>) are specialized to:

- store spatial information
- do spatial analysis
- present spatial information (view maps, print maps, 3D-view etc)

The integration of demanding models is difficult and sometimes very slow due to script language in the GIS.

On the other hand there are powerful modeling softwares like matlab (www.mathworks.com/) or octave (www.octave.org/) as the free counterpart. These softwares can handle complicated models fast, there are also include a lot of modeling techniques but they are not able to handle spatial data. To use modeling software, data must be exported from the GIS, the calculation is performed in the modeling software and the data must be exported to the GIS. This process is difficult. To cope with this situation, a new software „Spatial Analysis and Modeling Tool” (SAMT) (www.samt-lsa.org) was developed. SAMT inherits from a GIS the ability to handle spatial data and from modeling software to simulate different models quickly.

2. System Structure

In SAMT, the main focus is to have an open system to include different models (especially dynamic models, fuzzy models and neural networks). All models run as separate processes. The fuzzy models and neural networks have their own graphical user interface, the other models are controlled by SAMT. SAMT is a framework that combines all the models and data for a simulation and provide a basic set for spatial operations and analysis. Here are the main features of the Spatial Analysis and Modeling System SAMT:

- Interactive,
- Combines dynamic and spatial computations,
- Allows the integration of fuzzy models and neural networks,
- Has its own dynamic model development toolbox „SAMTDESIRE”,
- Designed to integrate different particularly models to one „system“.

The main window of SAMT shows, on the left side the grids, models and point themes. The central element of this view is the map using a color table or a black and white table. The control elements are located in the upper part. The three line inputs (called P1, P2, P3) can be used to provide some parameters for the simulation or visualization. Every command from the control window includes a simple explanation that will be shown in the status line when the command is selected.

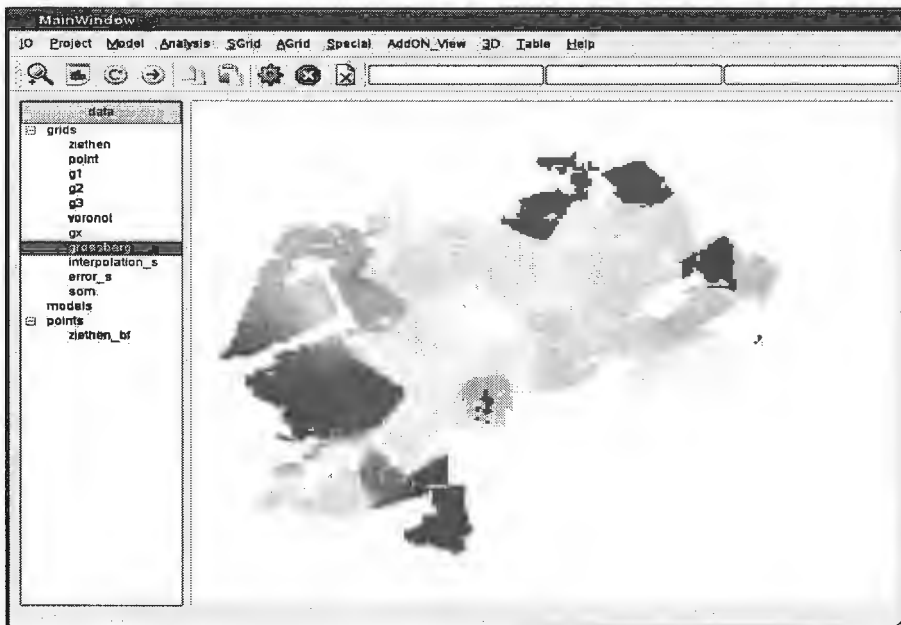


Figure 1. Main window of SAMT (www.samt-lsa.org).

SAMT's core contains all parts for handling, exploration and do spatial operations. That includes operations for database connection and the interface to the GIS (ARCGIS). Additionally was integrated a separate spatatial data management based on HDF-filessystem from NCSA (HDF:<http://hdf.ncsa.uiuc.edu/>). This allows a fast load and store procedure of grids and additional information like the time stamp, spatial information (location, grid size, no-data-code, etc.) name of the model, etc. This filesystem will be also used for the interaction of SAMT with different application models.

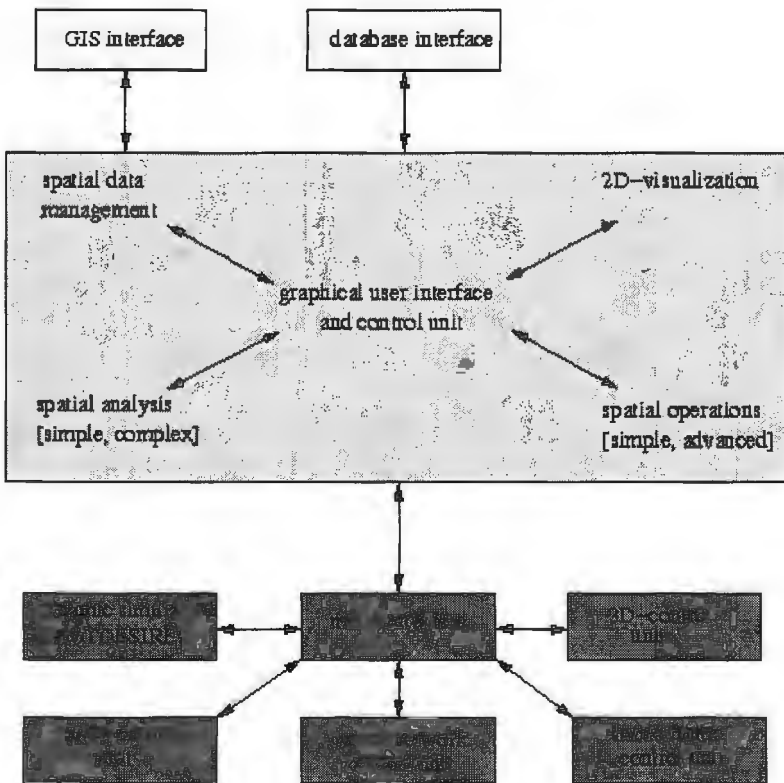


Figure 2. Structure of SAMT.

SAMT includes some tools to analyze map, like a simple statistic, histogram, correlation of two maps etc. Especially interesting is the possibility to define a transect. A transect is line over a grid. The transect can be used to generate a chart with the values under this line. Optionally a semivariogram or a power spectra are available. This transect can be used to analyze and compare the same region of interest in different grids. This procedure is often useful to get an impression about the data sets.

The simple grid operations contains operation over one grid like:

- Set: set the value of P1 to all grid cells
- Rand: overwrites a grid with random values [0..1]
- Exchange: replace grid cells[i][j]=P1 by the value in P2
- Add, Mul: adds, multiplies all grid cells with P1=const
- Normalize:
 - Norm: $cell[i][j] = (cell[i][j] - \min) / (\max - \min)$
 - Norm+-3sd: like above, but cut all values smaller (bigger) 3*standard deviation
 - Norm-Standard: $cell[i][j] = (cell[i][j] - \text{mean}) / \text{std}$
- Inv: $cell[i][j] = \max - cell[i][j]$
- Cut, Cut-Fuzzy: cuts values $\notin [P1, P2]$
- Class, Quad, Tanh, Gauss: functions to a grid

In addition to this simple grid operations there are some „advanced” grid operations like:

- Combine_Add: adds the grid in P1 to the highlighted grid
- Combine_Mul: multiplies the grid in P1 to the highlighted grid
- Combine_Avg, Combine_Min, Combine_Max

Advanced grid operations combine two different grids (overwriting the highlighted). The regional operations perform a moving window technology (Kiesel and Lutze, 2004) over all cells with radius R:

The $cell[i][j] = \iint_A f(x, y) dx dy$ with $f(x, y)$ that can be: mean, median,

min, max, diversity, average,... . The moving window technology can be used to create generalized input grids from maps. This is useful for habitat modeling where a density function of available nutrition must be calculated to estimate the habitat quality.

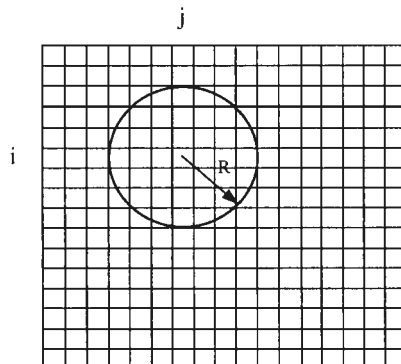


Figure 3. Moving window technology.

SAMT can apply some different view options, like color or black and white scale, contour lines, grid lines, different scales (absolute, relative). An absolute scale is useful to compare different simulation results.

Additionally there are included some simple operations like: sobel filter, laplacian filter, rotation of a grid etc. This operations are sometimes useful and should understand as a basis to expand SAMT.

2.1. Points

Point themes can be used to store measured data. In spatial context a measured data is characterized by a location and a value. (If the data is time depended a time stamp is also necessary.) This data set can be stored using a simple comma separated file format (csv). SAMT can handle points:

- Import points in a grid using Point_2_Grid
- Import points using voronoi diagrams
- Import points using a Shepard algorithm for spatial interpolation
- Import points using ordinary kriging

The first option is very simple and means the creation of a new grid wich includes only the measured data at the appropriate grid cell. This is useful if a distance between this datapoints has to be calculated for example.

Much more interesting is however an interpolation of the measured data points over the grid. The simplest method to do this is the so called voronoi diagram. Algorithm:

$$cell[i][j] = point[k] \quad \text{if } distance (cell[i][j], point [k]) = min$$

Every $cell[i][j]$ gets a value from the nearest point. This is useful to find the shortest distance to an object (like nutrition location) from any point of the area. It is also a first impression for a interpolation of points in an area.

More elaborated is real interpolation routine like the shepard interpolation or ordinary kriging. The shepard interpolation based on a inverse distance:

$$\Phi(x, y) = \frac{\sum_{j=0}^N \left(1 - \frac{r_j(x, y)}{R}\right)^\mu f_j}{\sum_{j=0}^N \left(1 - \frac{r_j(x, y)}{R}\right)^\mu}$$

The parameter μ and the interpolation radius R allow a variation of the interpolation result. This result looks much better than the patchwork of the voronoi diagram. But there are some artifacts in the result. A more powerful method is the widely used ordinary kriging (www.nbb.cornell.edu ...).

The word „kriging” is synonymous with „optimal prediction”. It is a method of interpolation which predicts unknown values from data observed at known locations. This method uses variogram to express the spatial variation, and it minimizes the error of predicted values which are estimated by spatial distribution of the predicted values.



Figure 4. Points and voronoi diagram.



Figure 5. Voronoi diagram versus Shepard interpolation.



Figure 6. Voronoi diagram versus Kriging interpolation.

The ordinary kriging is a bit experimental in SAMT. I can handle only two type of semivariogram function: sherical and dampened hole. It will be expanded in the future.

3. Modeling interface

The components of SAMT are grouped around the interactive control center. All operations of SAMT are controlled by this interactive control center. Most of the operations are integrated in this frame, but some main parts of SAMT (fuzzy toolbox, NN toolbox, 3D-visualization) are realized as stand alone programs, started and initialized by the control frame. The realization as of the models as separate process has a big advantage. When a bug occurs in the modeling process the the process will aborted, but the SAMT-core will not be damaged. The modeler can change the code, compile it and is able to start the process at the same point like before the error occurred. That means the simulation environment is stable and must not be reloaded after a crash. This safes a lot of time during the modeling process. Consequently all additional components like the fuzzy development toolbox, the neural network development toolbox and the dynamic simulation system are realized as separate process.

3.1. Cluster algorithms

Complex analysis functions include a cluster algorithm (k-means) and a kohonen feature map (Kohonen, 1982) (alternatively called self organizing map). The k-means algorithms are widly known (Bacher, 1996) but the kohonen feature map has some interesting features.

The kohonen feature map realizes an un-supervised learning:

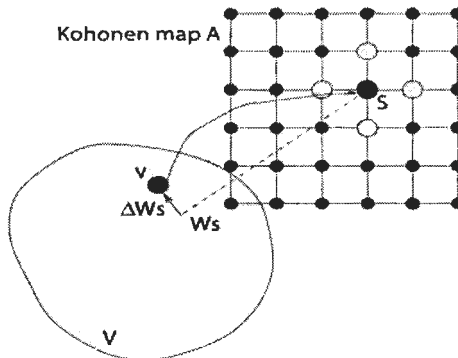


Figure 7. Kohonen feature map.

The training procedure produces a map from the input space V to the kohonen map. The algorithm in SAMT can handle linear som and som organized as a map $\{A\}$.

- Initialization: every node of the kohonen map gets a random starting vector $\vec{w}_{r_i} = rand()$
- Select a vector \vec{v} randomly from V
- Response: determine the winner $w_{r'}$ with: $\|\vec{v} - \vec{w}_{r'}\| \leq \|\vec{v} - \vec{w}_r\| \quad \forall r \in A$
- Adaptation: $\vec{w}_{r'}^{neu} = \vec{w}_{r'}^{alt} + \epsilon h_{r,r'} (\vec{v} - \vec{w}_{r'}^{alt})$ with: $h_{r,r'} = \exp\left(\frac{-(r - r')^2}{2\sigma^2}\right)$

A trained kohonen feature map can reproduce the statistics of V into the map A .

The result of the kohonen feature map or the cluster algorithm can be used together with a radial basis function network (rbf) as a model alternative to the feedforward neural networks. Sometimes the rbf gets better results as the feed forward net.

3.2. Radial basis function network

The output of the cluster (or kohonen) algorithm can be used as a center of a radial basis function network (Gershenfeld, 2002). SAMT provides a selection procedure to get interesting values from a target grid. This grid can contain information about observations of plants or animals for example. The question is: can we predict this occurrence with information of other grids (like soil characteristics, elevation, climate zones etc.)? A selection procedure of SAMT uses a histogram to get the examples for the training procedure (about 100 classes with 20 examples each). This values are the y_i . The \vec{c}_i means the class center from the cluster algorithm. The \vec{x}_i are the location of the selected y_i . The aim of the algorithm is to determine the coefficient a_i . With this a_i a grid of an unknown region can be calculated. That means this type of neural networks will be able to predict the occurrence with un-supervised training.

$$f(\vec{x}, \vec{c}) = \exp(-scale(x) \times |\vec{x} - \vec{c}|) \quad (1)$$

$$y(\vec{x}) = \sum_{i=0}^M a_i \times f(\vec{x}, \vec{c}_i) \quad (2)$$

$$\begin{pmatrix} f(\bar{x}_1, \bar{c}_1) & \cdots & f(\bar{x}_1, \bar{c}_m) \\ f(\bar{x}_2, \bar{c}_1) & \cdots & f(\bar{x}_2, \bar{c}_m) \\ \vdots & \vdots & \vdots \\ f(\bar{x}_n, \bar{c}_1) & \cdots & f(\bar{x}_n, \bar{c}_m) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad (3)$$

4. Integrated models

4.1. Neural networks

Neural networks can be used to extract modeling knowledge from data by supervised training (Schultz, Wieland and Lutze, ?). The common situation is, that a lot of data are available, but no idea of functional model. In this situation a regression approach is often used. For such an approach a regression function must be given. But often there is a lack of knowledge about the functional behavior. In this cases a neural network can be used as a universal regression model. The in SAMT integrated neural network has the following features:

- up to three inputs
- one hidden layer
- one output
- training algorithm: backpropagation, backpropagation with momentum (Freemann, 1994), levenberg-marquardt (www.eng.auburn.edu ...)

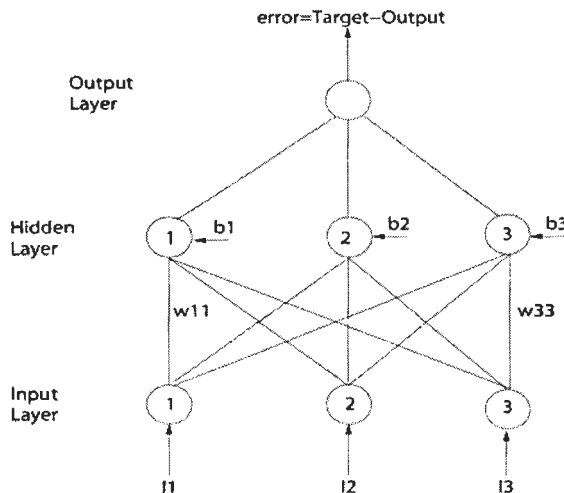


Figure 8. Feed forward neural network.

A simple neural network structure is shown in Figure 8. This network consists of an input layer with three nodes $\{i_1, i_2, i_3\}$, a hidden layer with the nodes $\{h_1, h_2, h_3\}$ and one output layer o1. The input layer is connected with the three inputs $\{x_1, x_2, x_3\}$. The input nodes have an identity function between the inputs $\{x_1, x_2, x_3\}$ and the outputs of this nodes. The hidden nodes and the output nodes use a sigmoid function like:

$$o_i = \frac{2}{1 + \exp(-x_i * w_{i1} - x_2 * w_{i2} - x_3 * w_{i3} + b_i)} - 1 \quad (4)$$

During the training process all weights w_{ij} and biases b_i will be updated so that the mean error will be minimized.

$$E = \sum_{k=1}^M (y_k - o_k)^2 \Rightarrow \text{Min!} \quad (5)$$

With M as the number training patterns. The convergence of the training depends on the structure of input data, the starting point (all weights will be set to small random values between -1,1) and the structure of the network. In SAMT are integrated a set of training algorithms. The well known back propagation algorithm is the default. Some improvement can be done using momentum in the back propagation algorithm. The levenberg-marquardt algorithm is much faster than the back propagation, but it is more sensible from the starting point as the backprop algorithm. A good use of levenberg-marquardt algorithm would be to try it many times one after other. If a good result was found, it can be stored and used as a spatial model. SAMT also provides a graphical toolbox with a colored two dimensional view of the result of a neural network.

4.2. Fuzzy models

Another strong feature of SAMT is its possibility to integrate expert knowledge in form of fuzzy models. Fuzzy models run often as interpreted code and are very slow. For spatial application fuzzy models have to cope with many hundred thousands of cells. In other words, there is a need for a fast fuzzy implementation. This explains the reason for incorporating fuzzy compiler. This compiler is characterized by the following:

- maximum number of inputs are 3 grids,
- the output is of type singleton,
- the fuzzy compiler uses a special structure file to change the behavior.

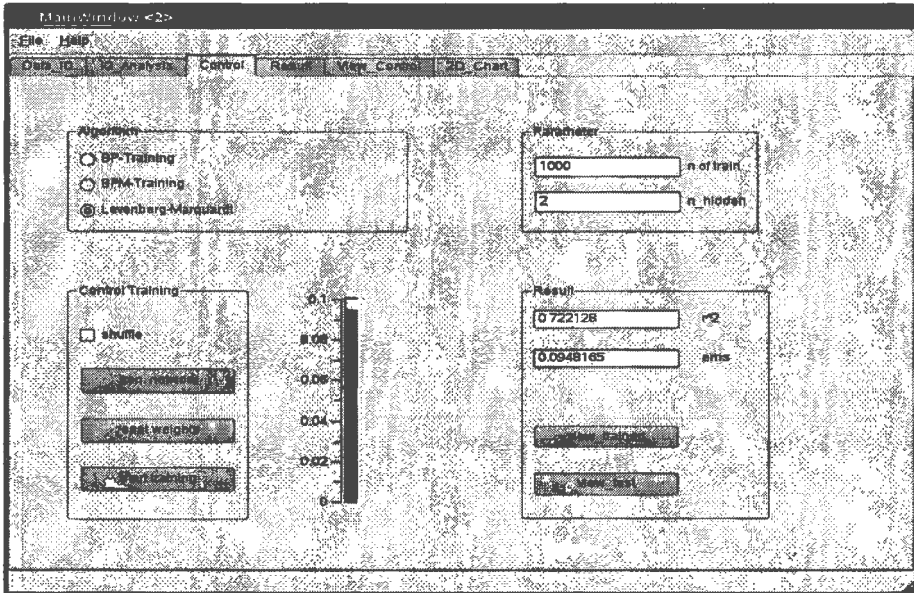


Figure 9. SAMT_NN control window.

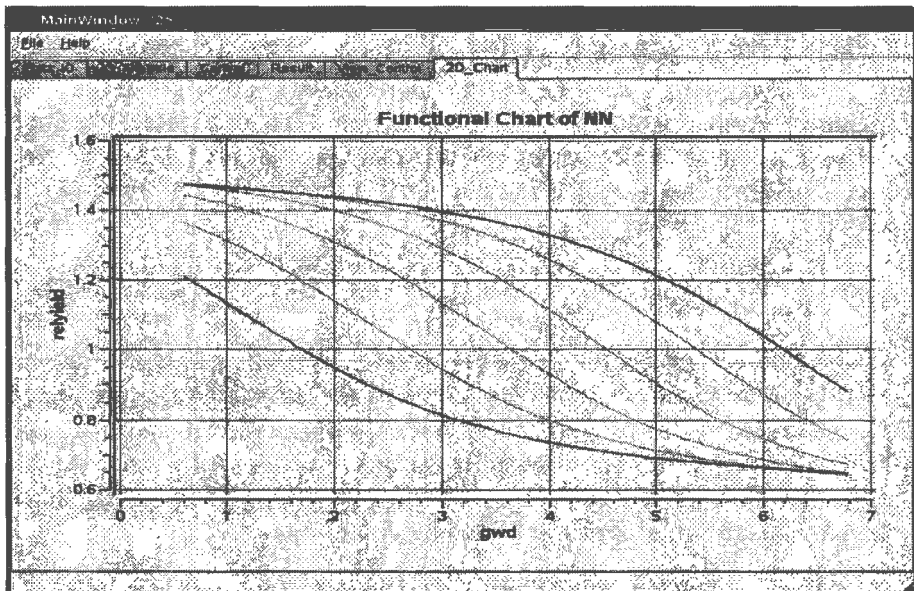


Figure 10. SAMT_NN analysis window.

The standard used min/max operator can be changed by prod/max operator. Additionally a weight value c_R can be used to weigh the belief of a rule. This is managed by a fuzzy modeling toolbox with a sophisticated graphical user interface. With this toolbox it is easy to develop fuzzy models and to test the model with set of tools like diagrams and a powerful rule inspector. The number of inputs is not a restriction in practice. It is easy to divide a big problem into a hierarchy of some small fuzzy models. Also the use of singleton of the outputs are not negative, but it increases the velocity enormously. Tests using computer of a 1.500 MHz P4 with 1GB memory showed that a simple problem using 2 input grids with 54000 cells needs only one second of computation time. A more complex system of about one million cells needs not more than twenty seconds. Handling the same problem using the fuzzy toolbox of matlab will take several hours (up to ten hours). Fuzzy modeling technique is usefully applied in different types of ecologic and landscape modeling, such as habitat modeling, (Wieland and Voss, 2001) yield modeling for agricultural crops.

In SAMT was included a powerful fuzzy development toolbox: SAMT_FUZZY:

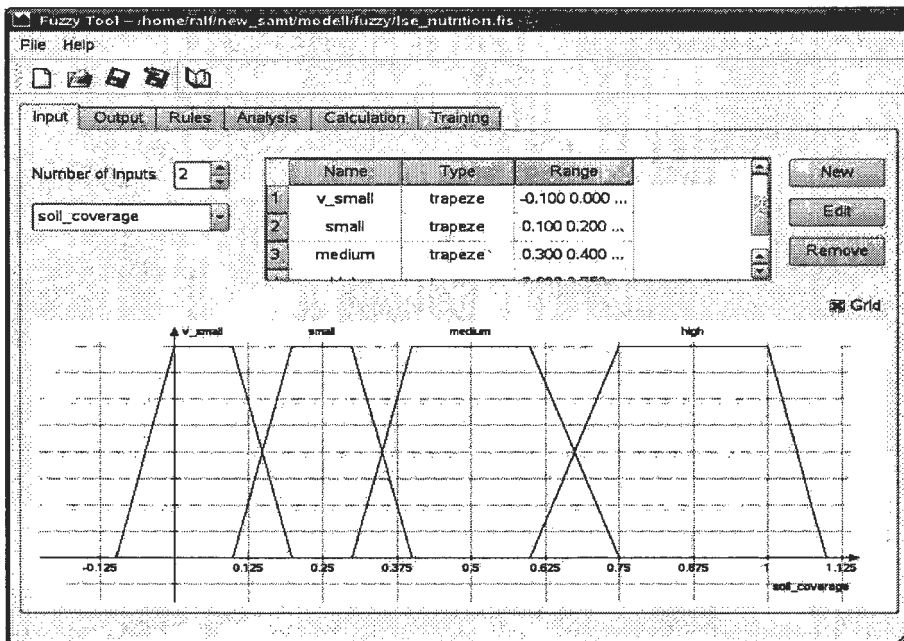


Figure 11. SAMT_FUZZY input window.

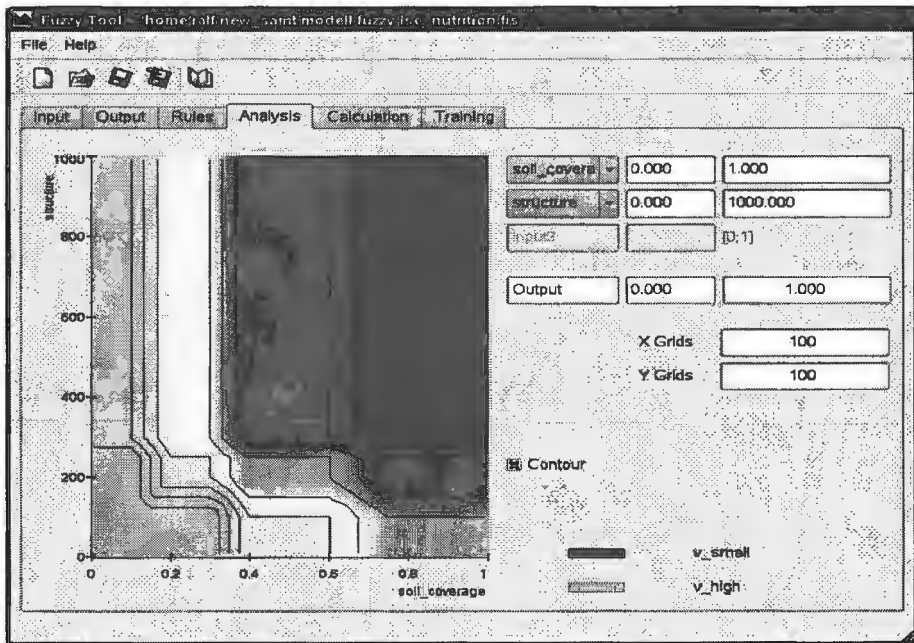


Figure 12. SAMT_FUZZY analysis window.

5. SAMTDESIRE

SAMT can handle dynamic simulation using C++ code. This will be supported by in template to build an model using SAMT. C++ is very fast and should be used together with numeric libraries like gnu scientific library (www.gnu.org/software/gsl/). It should be mentioned that a model in C++ can used the same grid library that SAMT uses. This simplifies the model integration in SAMT. To investigate in dynamic processes for example to adapt parameter using a Monte Carlo simulation C++ is too difficult. This was the reason for the inclusion of a dynamic simulation system called SAMTDESIRE (<http://members.aol.com/gatmkorn/>):

- Solves systems of differential equations,
- Is used for development of models (rapid prototyping),
- Does parameterization of models,
- Uses fuzzy models and neuronal networks from SAMT-Fuzzy and SAMT-NN as part of dynamic models,
- Interacts with SAMT.

The use of SAMTDESIRE is simple due to the excellent graphical interface and the amount of examples that comes with SAMTDESIRE. Here are only the main features listed:

- Has an experiment protocol script (interpreter) which sets up the parameters and calls the dynamic part (compiler),
- The dynamic part is very fast and has a good numeric (many solvers for differential equations),
- It has a powerful vector compiler for model replication,
- It has an excellent graphic,
- An additional documentation is available,
- It allows an interactive and fast model development (Korn, 2004).

Here are two screenshot:

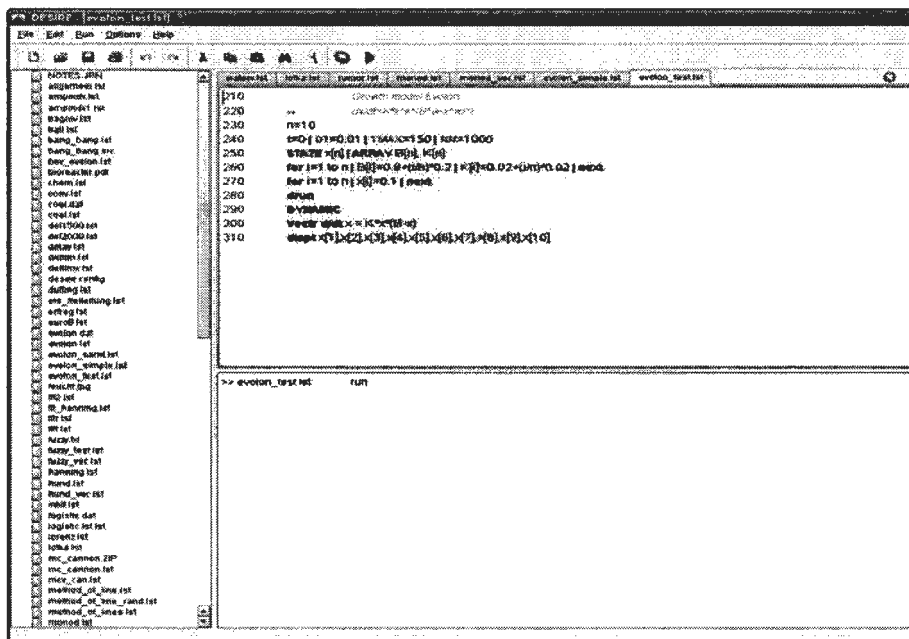


Figure 13. SAMTDESIRE main window.

6. Software version and further development

SAMT is free software using the gnu public license (www.gnu.org/licenses/gpl.html). That means the source code is available, can be changed but it must be free also in future. It runs under the operating

system Linux and uses the following main components: compiler gcc, glibc; qt as gui-library, 3D-library vtk; gnu scientific library, hdf-library, and MySQL as database management tool. The actual version of SAMT is 2.0. For test reason there is available a self booting CD, based on Knoppix-Linux (www.knopper.net/knoppix). In the future there are a lot of things to do. For example: to include a tool for cellular automata, add some tools for analysis and to actualize the help file.

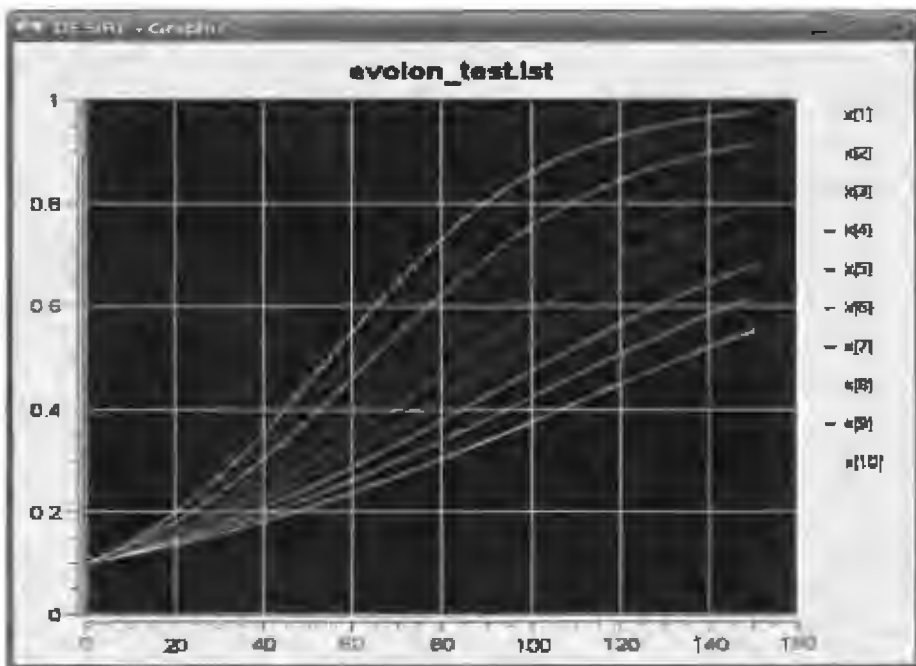


Figure 14. SAMTDESIRE simulation output.

7. Conclusion

SAMT is a new open source software to handle spatial simulations. The main features of SAMT are the combination of different models and the extensive usage of expert knowledge (fuzzy and neural networks). This combination has proven powerful and usable in set of applications. One short example of such a model combination is provided.

8. Acknowledgements

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

Many thanks to Mr. G.A. Korn for the source code of the wonderful dynamic simulation system DESIRE that we used as part of SAMT.

References

- <http://portal.mtt.fi/pls/portal30/url/ITEM/D3036A3D11057223E0340003BA16D803>
<http://www.esri.com/software/arcgis/>
<http://grass.itc.it/>
<http://www.mathworks.com/>
<http://www.octave.org/>
<http://www.samt-lsa.org>
HDF: <http://hdf.ncsa.uiuc.edu/> (version: hdf5-1.4.5-post2)
Kriging: [http://www.nbb.cornell.edu/neurobio/land/OldStudentProjects/cs490-94to95/ clang/kriging.html](http://www.nbb.cornell.edu/neurobio/land/OldStudentProjects/cs490-94to95/clang/kriging.html)
Wieland R., Voss M. (2001) Land use change and habitat quality in Northeast German agro-landscapes. In: *Tenhunen, J.D., R. Lenz, R. Hantschel (Hrsg.): Ecosystem approaches to landscape management in Central Europe*: 341-346; Springer Berlin, (*Ecological Studies*; 147).
Kiesel J., Lutze G. (2004) Skalierbare Generalisierungsmethoden zur Ausgrenzung von Landschaftseinheiten. In: Wittmann J, Wieland R. (*Hrsg*) Simulation in Umwelt- und Geowissenschaften. Shaker Verlag.
Kohonen T. (1982) Self-organized formation of topologically correct feature maps. *Biol. Cybern.*, **43**: 59-69.
Bacher J. (1996) *Clusteranalyse*. Oldenbourg.
Gershenfeld N. (2002) *The Nature of Mathematical Modeling*. Cambridge UP.
Schultz A., Wieland R. and Lutze G. (2000) Neural networks in agroecological modelling - stylish application or helpful tool ? *Computers and Electronics in Agriculture*, **29**: 73-97.
Freemann J.A. (1994) Simulating neural networks with Mathematica. Addison Wesley. 71-96, 98-101.

http://www.eng.auburn.edu/~wilambm/pap/2001/FastConv_IJCNN01.PDF

<http://www.gnu.org/software/gsl/>

<http://members.aol.com/gatmkorn/>

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.

<http://www.gnu.org/copyleft/gpl.html>

<http://www.knopper.net/knoppix/>

Jan Studzinski, Olgierd Hryniewicz (Editors)

ECO – INFO AND SYSTEMS RESEARCH

This book presents the papers that describe the most interesting results of the research that have been obtained during the last few years in the area of applications of informatics in environmental engineering and environment protection at the Systems Research Institute of Polish Academy of Sciences in Warsaw (IBS PAN) and at the German Institute for Landscape System Analysis in Müncheberg (ZALF). The papers were presented in the form of extended summaries during a special workshop organized by IBS PAN in Szczecin in September 2006 together with the conference BOS'2006 dedicated to the applications of systems research in science, technology and economy and organized jointly by IBS PAN, University of Szczecin, and the Polish Society of Operational and Systems Research. They deal with mathematical modeling, approximation and visualization of environmental variables and with development of computer aided decision making systems in the area of environmental informatics.

ISBN 83-894-7509-X
9788389475091
ISSN 0208-8029
