CHAPTER SIXTEEN

REGIONAL MODELS OF INTERMEDIATE COMPLEXITY (REMICs) – A NEW DIRECTION IN INTEGRATED LANDSCAPE MODELLING

K.-O. Wenkel a, R. Wieland a, W. Mischel a, A. Schultz b, C. Kampichler c, A. Kirilenko d, and A. Voinov e

Contents

16.1. Why Do We Need Better Models on a Landscape Scale? 285
16.2. The Way Forward 287
16.3. Landscape Models 288
16.3.1 Selection of landscape indicators 288
16.3.2 REMICs 288
16.3.3 Hybrid models 289
16.3.4 Complexity in landscape modelling 289
16.4. A Sample Modelling Tool 291
16.5. Conclusions 293
References 294

16.1. WHY DO WE NEED BETTER MODELS ON A LANDSCAPE SCALE?

Over recent decades marked progress has been achieved in ecosystem modelling. Traditionally, process and ecosystem models in ecology are developed for the one-dimensional case (point model), i.e. only for selected points within an area. Examples of such process models are soil-plant-atmosphere models such as CERES (Ritchie and Godwin, 1993), WOFOST (Supit et al., 1994), DAISY

---

*Environmental Modelling, Software and Decision Support* © 2008 Elsevier B.V. 1574-101X. All rights reserved.

---

a Institute of Landscape Systems Analysis of the Leibniz Centre for Agricultural Landscape Research (ZALF) Müncheberg, Eberswalder Str. 84, 15374 Müncheberg, Germany

b University of Applied Sciences Eberswalde, Faculty of Forest and Environment, Alfred-Möller-Str. 1, 16225 Eberswalde, Germany
c Universidad Juarez Autónoma de Tabasco, Tabasco 2000, Apartado Postal 984, 86031 Villahermosa, Tabasco, Mexico
d Department of Earth Systems Science and Policy, University of North Dakota, Grand Forks, ND 58202-9011, USA
e Chesapeake Research Consortium, 645 Contees Wharf Road, PO Box 28, Edgewater, MD 21037, USA
(Hansen et al., 1990), STICS (Brisson et al., 2003), AGROTOOL (Poluektov et al., 2002), HERMES (Kersebaum, 1995), APSIM (Keating et al., 2003), THESEUS (Wegehenkel et al., 2004) or AGROSIM (Mirschel et al., 2001; Wenkel and Mirschel, 1995) each with different foci. An overview of agroecosystem and ecological models is given in CAMASE (2005).

These developments have stimulated initial attempts to put together spatial ecosystem models using Geographic Information Systems (GIS) to mechanically connect local ecosystem dynamics. Prototypes of integrated landscape models at first appeared as coupled comprehensive process models replicated over a network of points. These were either grid points, like in raster-based GIS, or a linked network of compartments, like in vector-based GIS.

Local point models describe processes in great detail and demand a lot of information as driving forces or as parameters. At a regional scale it is difficult to provide all this information for all the spatial nodes. Here the available information is usually characterised by significant fuzziness, and heterogeneity. Monitoring is usually conducted only at several points in the landscape, and vast areas are not supported by any data at all. At the same time, remote sensing information that can cover considerable areas and could provide good spatial information, are usually available only for a few points in time, and rarely provide any understanding of the dynamics of processes. It is almost a given that the GIS-based models will be under-parameterised with much uncertainty about the spatiotemporal dynamics. Another limitation of such comprehensive ecosystem models is their high computational requirements, making it very difficult to calibrate and analyse them.

In a region there are many processes that are related to different spatial and temporal scales. For example, biotic processes described by habitat or population models usually are quite local. On the other hand, some abiotic processes such as flow of material driven by wind, water or diffusion are strongly influenced by spatial features within catchments such as the heterogeneity of the land use types or elevation. On the policy-making side we also find that while most of the concerns of people are quite local, and focus on local effects of various policies, really effective plans and regulations require that regional catchment-wide processes and trends are taken into account.

An example is energy production by wind or from crops. These types of land use can provide new benefits for a whole region, but the associated ecological problems can be quite local (vulnerability of bird habitat, or increased nutrient pollution and water withdrawals for irrigation to produce biofuels). In all cases it is necessary to investigate the spatiotemporal impacts of different locally-realised land use activities on the environment and at the regional scale.

Landscapes, the subject of landscape management, are complex, spatially and temporally multi-layered systems, which change and develop naturally but are also subject to anthropogenic changes (Lausch, 2003). To analyse, evaluate and manage them, tools and models are required that can represent and interpret the variety and complexity of the connections between biotic and abiotic landscape structures and functions. The more recent growing concerns about climate change and the end
of cheap oil will only add to the mix of problems we face at the regional scale and require even more urgent and reliable decision making and policy support.

In spite of substantial progress in the development of mechanistic, integrated ecosystem models in recent years, integrated space and time-related dynamic landscape models are exceptions to date. The development of such integrated regional models is a very ambitious task and requires its “own inter-methodical approach” (Lausch, 2003). However this is still a challenge as the variety and diversity of knowledge from various scientific disciplines is increasing, requiring ever more creativity, versatility and openness in research and communication.

16.2. THE WAY FORWARD

The development of integrated dynamic models and their use for sustainable landscape management is one of the most important challenges for landscape ecology and landscape sciences for the next decade. As already mentioned, marked progress has been achieved in ecosystem modelling. On the other hand, there are only a few models that can be used for sustainable resource management at regional or landscape scale. New frameworks and methods are necessary in landscape and regional environmental modelling. From the scientific point of view we are confronted with the following problems:

• How do we reduce the complexity of landscape processes within models? How much detail in process dynamics can we incorporate in regional models still capturing the most important feedback loops in regional landscape systems?
• How can we adequately capture the functional consequences of structural diversity and spatial heterogeneity in landscape models?
• How can we compromise between short and long term developments adequately in landscape models? What do we mean by “adequate”?  
• Which model types and tools are best suited to landscape modelling, considering the fuzziness and uncertainties in ecological functions and data and the dynamic nature of processes?
• How can we bring together conceptually different ecological and economic models?
• Which spatial resolution is required to appropriately capture processes with regional significance? How can we adequately include spatial ecological heterogeneity and its influences on ecosystem functions?

To find answers to these problems we need new ideas, new thinking and a paradigm shift in integrated regional modelling. We propose the following steps as crucial ones, not necessarily in the order below:

• Select and discuss a set of landscape indicators for assessment of sustainability of land use systems and landscape development at the regional scale.
• Develop a new generation of integrated regional models of intermediate complexity, perhaps following some of the concepts developed by the climate modelling community.
• Create hybrid models by combining simplified or aggregated process-oriented dynamic models invoking statistical, fuzzy and other model types.
• Develop and use new software toolboxes for the efficient development and simulation of dynamic landscape simulation models (see Chapter 7).
• Promote open source in program development to enhance free flow of ideas, models and communication (see Chapter 20).
• Develop a new generation of tools for multi-criteria evaluation and visualisation.
• Improve validation of regional models using long term monitoring data from landscape studies around the world (e.g. the LTER data sets).

16.3. Landscape Models

16.3.1 Selection of landscape indicators

Because landscapes are highly complex systems, one possible approach is to depart from process-based modelling and use indicators instead. Landscape indicators describe the landscape in a more aggregated way. They may come from a variety of disciplines such as geology, economics, ecology or soil science. For example, there is the well-known Normalised Difference Vegetation Index (NDVI) used in remote sensing as a proxy for the state of vegetation and photosynthetic activity. Many other indices can be developed, but the real challenge is to select the most representative ones and learn to quantify and measure them. Additionally, we need to remember that these indices represent processes that are occurring in space and time, and need to be scalable in both dimensions.

It is also clear that no one set of indices can be universal to serve all needs. As with other modelling efforts it is a problem-specific process that is driven by the goals of the individual project. The selection of a problem and scale-oriented indicator system cannot be completely free from subjective preferences. To describe a region by indicators derived from the main regional processes and functions, the indicator selection should be undertaken in participative discussions with scientists and stakeholders. In this process the chosen indicators and their limitations should be critically evaluated and determined. Yet another challenge is the interpretation of indices and model results and translation of these results into meaningful measures to influence decision and policy making.

16.3.2 REMICs

As in other models it is always an open question as to how much detail in process dynamics and which spatial resolution is required to get sufficiently accurate answers on the socioecological impacts of land use changes at the regional scale to enable decision support for landscape management. Neither conceptual models on the one side, nor three-dimensional comprehensive models on the other side, seem to be the right solution for all problems. The answer to this question is always going to be determined by the specific problems that are to be. For most landscape-scale regional models, however, there seems to be a common level of detail that
may be required. Regional Models of Intermediate Complexity (REMICs) seem to be the kind of models that would address most of the management problems at the regional scale. REMICs describe most of the processes in a more simplified, reduced form. They explicitly simulate the interactions between different components of the landscape system, including the necessary feedback loops (Claussen et al., 2002). However, they are simple enough to allow long-term simulations at the landscape scale. An important feature of REMICs is that they are characterised by a lower degree of detail in their description of process dynamics, but may have more variables to parameterise some of these processes.

16.3.3 Hybrid models

In many cases it is unlikely that one model will be sufficient to address all the needs of landscape or watershed management. There are many processes and many important feedbacks that may need to be considered. All these processes not only differ in their functional content, but also differ in their time and space resolution. At the same time experimental data and information is likely to be sparse, patchy and/or uncertain. Depending on the particular goals of the study and the data availability it is necessary to have a large model base, a repository of modules and tools that can be pulled out to describe various landscape variables and indicators at different levels of complexity and for different types of input and output data requirements. In most cases for a comprehensive landscape study we will need hybrid models that will be an intelligent combination of simplified and/or aggregated process-oriented, robust, dynamic models with static and/or stochastic simulation models (e.g. empirical, statistical model, or a neural network or a fuzzy model). These hybrid models have to organise and manage the data flows between the sub-models via data interfaces for both spatial (maps) and temporal (timeseries) data types as well as combinations of those (dynamic maps, animations).

16.3.4 Complexity in landscape modelling

The term “complex system” usually refers to a system of many parts which are interactive. Natural complex systems like landscapes, for instance, are modelled using mathematical techniques of dynamical systems, which include differential equations, difference equations and also maps. If we are striving for intermediate complexity for landscape models, what are the metrics that can be used to estimate complexity? There are several sources of complexity that we need to take into account. These would include: the model structure, the number of model parameters and variables, the spatial and temporal resolution and extent, accuracy of measurements, the number of observations, etc.

Uncertain parameters of models are approximated on the basis of available datasets and therefore they depend on the quality of the empirical data. The quality of spatial data varies strongly between different data sources. Common to all these data sets is their noise and that they often are “snapshots in time” that cannot be repeated. For example, a satellite image depends on the position of the satellite and the weather conditions, so for a new image of the same area a time lag up to a
month is not uncommon; besides the cloud conditions may be very different by then and the two images may be very different. A modeller has to cope with this poor data quality. Most other data sets for spatial modelling are limited in a similar way. For example, soil maps containing data about soil texture, or moisture content, or porosity, are generalisations over considerable spatial units. This means that for every point on the map a discrete value of, say, soil texture is given as a mean value for a whole neighbourhood of the point. This is rarely the value that can be measured in reality. This generalisation introduces an additional error (noise) in the simulation because of its quantisation. All data sets are quantified in some respects. It is important to be aware that all data used in a simulation have a quantisation error or, in other words, have limited information content.

The information that can be used from a real data source is limited by the signal to noise ratio. Following Shannon (1948) we can define the information content \( I \) of a data source as:

\[
I = \log_2(1 + S/N)
\]

where \( S \) is the signal strength and \( N \) is the noise. For example, in an elevation model the noise can be the accuracy of the measurement, say 25 m. Suppose the signal has a maximum value of 175 m in our region. Then the information can be estimated as:

\[
I = \log_2(1 + 175/25) = \log_2(8) = 3 \text{ bit.}
\]

This low resolution \( I \) would preclude the use of highly sophisticated procedures of parameter fitting due to over-fitting. On the other hand, there is another kind of uncertainty involved. This uncertainty stems from the granularity size used in the model. On the one hand, there are common, very detailed models to calculate, for example crop yield. On the other hand, there are simple but robust fuzzy models or neural networks that can do the same, but in a different way. The question is, how complex may landscape models (artificial neural networks, fuzzy systems, regression equations, etc.) be to take into account the uncertainty of inputs? One approach, which illustrates the difficulty of applying single metrics, is to approximate the complexity of models by the number of their parameters. This is what we find in the Akaike Information Criterion (AIC) (Burnham and Anderson, 2004):

\[
AIC = 2k - 2 \ln(L) = 2k + n \cdot \ln\left(\frac{RSS}{n}\right)
\]

where \( k \) is the number of parameters, \( L \) is the likelihood function, \( n \) is the number of observations and \( RSS \) is the residual sum of squares. Note that this equation holds only for Gaussian distributed noise.

This criterion can be improved either by making \( RSS \) smaller, or by reducing the number of parameters, or by increasing the number of observations. For a model with a high number of input parameters, as well as a high level of detail (which often implies a lot of parameters), the AIC will be high. If two or more models are available the AIC can be used to compare their complexity and select the appropriate one. However this formal approach does not work in all cases. The reaction of a model to an input that is beyond its defined input range can be fundamental. This is important for the use of models in similar but not identical landscapes (portability
of a model). Another difficulty is that AIC does not consider the different meaning of parameters for different modelling techniques. For example, there is a big difference between parameters of a feedforward artificial neural network and a counter propagation network in the sense of the sensitivity to small variations in parameters. Also it is not clear how to use the index for hybrid models, which may have several models of various degrees of complexity interacting among themselves. The complexity of such models is more than the sum of their individual AICs. In any case choosing the right model complexity remains mostly an art and can be improved by engagement of stakeholders at an early model development stage.

16.4. A Sample Modelling Tool

Simulation systems, like Matlab or ModelMaker for example, can handle simulations quite well, but have shortcomings in spatial data handling. GIS platforms like ArcGIS, can handle spatial data well but are not very efficient for dynamic simulations. What is needed is an open source toolbox which is able to both run simulations and handle spatial datasets. This toolbox should include:

- an interface to the GIS especially to read and write grids in ASCII ArcGIS format;
- a spatial data management system to store and load big spatial data sets with high speed (HDF-Format for example) and to exchange data with other simulation software;
- a connection to a database management system to store all information (specific model parameters, parameters of the simulation, measured values to validate the models etc.) in a flexible way;
- a basic set of grid operations (normalise grids, multiply grids, add grids, etc.);
- a basic set of analysis functions, like histograms, statistical routines, etc.;
- tools to handle expert knowledge (e.g. fuzzy models) and to build models from empirical data sets (e.g. neural networks);
- a framework to include models or other open source software components in the toolbox; and
- a graphical user interface to control the simulation and write reports, visualise results as maps, charts or in a three-dimensional form and include methods to handle multicriteria results.

SAMT (Spatial Analysis and Modeling Tool) is an example of such an integrated spatial simulation toolbox that can combine dynamic and spatial computations (Wieland et al., 2006). SAMT is an interactive spatial simulation system and it provides a framework to include different other models. In SAMT, there is a link to a neural networks module (SAMT_NN) to extract models from data, and a module for fuzzy modelling (SAMT_FUZZY) to take into account expert knowledge. The neural networks and fuzzy models have their own graphical user interfaces; the other models are controlled by SAMT. One of the most powerful extensions of SAMT is the simulation toolbox SAMTDESIRE for development of dynamic
models. SAMT is a framework that combines all the models and data for a simulation and provides basic functions for spatial operations and analysis. Figure 16.1 shows the structural scheme of SAMT. Some applications performed with SAMT are given by Aijbefun et al. (2004) and Mirschel et al. (2006).

Another advantage of SAMT over other packages with similar functionality is that it is open source. The whole SAMT toolbox with SAMT-NN, SAMT-FUZZY and SAMTDESIRE is freely available from http://www.samt-lsa.org. Making the source code available to the public enables anyone to copy, modify and redistribute it without paying royalties or fees. Open source code evolves through community cooperation. Scientific open source software means that everybody has the opportunity to understand the algorithms behind the program, is able to enhance it and return it to the community. This can improve software quality, with better error control, and by offering a framework where new scientific methods and algorithms can be easily incorporated, expanding the existing tool instead of developing entirely new ones. Another advantage of this strategy is that the data format specifications are always available – an indispensable precondition for combining different models.
16.5. Conclusions

As we have seen, the development of dynamic landscape simulation models, which can be used to support sustainable resource management at a regional scale, is a great challenge for landscape research in the next decade. We advocate the development of new models of intermediate complexity that would operate with indices as well as landscape variables and would allow modularity and flexibility to recombine these modules in various ways, depending on the type of study and its goals. SAMT offers promise as one of the toolboxes that can support this kind of modelling.

There are many ambitious tasks yet to be solved. Among them are the following:

- **Matching scales.** Temporal scales of ecological processes and models are highly variable. Most environmental management decisions are designed to span years or decades, whereas various critical regimes and threshold conditions may occur over weeks or even days. The time scale of the models needs to relate to the time scale of the management questions and their implications. Therefore for purposes of ecosystem management we need the capability to run landscape models over sufficiently long periods of time, with small enough time steps to track the critical conditions.

- **Improved validation.** A prerequisite for scientifically rigorous validation of complex integrated landscape simulation models is a network of long-term experimental study areas or, if feasible, specially designed landscape experiments. Backcasting and comparing model results to historical conditions offers a useful way to validate a model. The core research areas within the Long Term Ecological Research Network (LTER) (LTERNET, 2006) could be the basis for the much needed data sets and studies.

- **Stakeholder involvement.** Development of sophisticated simulation models is an integrative, interactive and iterative process. The development of landscape models is a powerful process for synthesis of data, theories and opinions over scales of space, time and biological organisation. It is also a process for creating new insights and questions for new experimental studies at regional scale. By incorporating stakeholder knowledge we can improve understanding of landscape dynamics and assure that the model results will be better accepted by the decision makers.

- **Improved software tools.** Integrating models into decision making requires specialised, flexible, user-friendly and transparent tools for environmental management and stakeholder participation. For such applications, certain modules and analytical tools need to be designed upfront, while integrative and interactive graphic user interfaces can help combine modules and produce compelling visualisations.

- **Community modelling.** The process of a group of scientists collaborating and sharing their expertise to develop integrated landscape simulation models can be a worthwhile scientific accomplishment. Open source software and models is an almost essential prerequisite of success in such community efforts.
• **Support for multicriteria decision-making.** If we assume Pareto optimality as a guiding principle for decision making ("given a set of alternative allocations and a set of individuals, we choose the allocation that can make at least one individual better off, without making any other individual worse off"), we need to be able to provide information about alternatives to be able to run the optimisation. A promising approach is to use genetic algorithms to control the search procedure for optimising.

• **Dealing with uncertainties.** Model results always contain uncertainties because they are based on (1) incomplete understanding of interactions and processes, and (2) field and laboratory studies that are always approximate (Dale, 2003). Model results are projections (i.e. estimates of future possibilities) rather than predictions, something that is declared in advance (Dale and van Winkle, 1998). Models produce approximations to real situations and are only as good as the assumptions upon which they are based. Model results therefore should be considered with caution; they are the logical extension of existing data produced via a process that assimilates and applies current understanding. Current understanding of complex environmental systems, as reflected in models, will rarely be adequate alone to provide simple answers to environmental questions. Often incomplete information must be accepted, and decisions must be made with the best available information. However, the absence of full information does not imply that there is no scientific value in developing models.

**REFERENCES**


