# From Complex to Simple in Environmental Simulation Modelling

**Khaiter, P.A.**[1] **and M.G. Erechtchoukova**[1]

[1] The Atkinson Faculty of Liberal and Professional Studies, York University, CANADA
Email: pkhaiter@yorku.ca

***Keywords:*** *Simulation modelling, Ecological system, Experiment design, Response surface method, Environmental management*

## EXTENDED ABSTRACT

Decision making in environmental management requires a reliable prediction of the ecosystems dynamics. The ecological objects, however, belong to the class of complex dynamic systems whose behaviour in general and responses to human-caused impact in particular are determined by the interplay of a number of factors, constituting components and processes as well as by the interrelated positive and negative feedbacks spanning multiple spatial and temporal scales. Given also obvious restrictions on the experiments with ecosystems, it becomes clear that the problem could only be dealt by means of quantitative models.

There is a great deal of discussions among the researchers on the issue of advantages/disadvantages of simple or parsimonious models versus large and comprehensive models in ecosystem science.

The paper presents an approach (Fig. 1) to the construction of a simple model based on the complex simulation model developed for a particular environmental problem. The approach suggested does not eliminate the necessity to build a complex model. Moreover, a complex model plays an important role in the approach. But parameterisation and other required in the model building steps are performed only once for a large model and are not needed in order to apply a simple model to various management tasks related to the investigated real world system.

The approach is based on an idea of classification of the components of the model state vector into primary and secondary variables. The former are used as the factors in response surface design aimed at the construction of a resulting function as a statistically valid approximation of the solutions delivered by the complex model. The resulting response surface equation serves as a simple model suitable for practical calculations. It is important to note that the simple model so obtained retains the descriptive power of the initial comprehensive simulation model because it does not ignore secondary variables and, thus, does not introduce additional uncertainty to the modelling results.

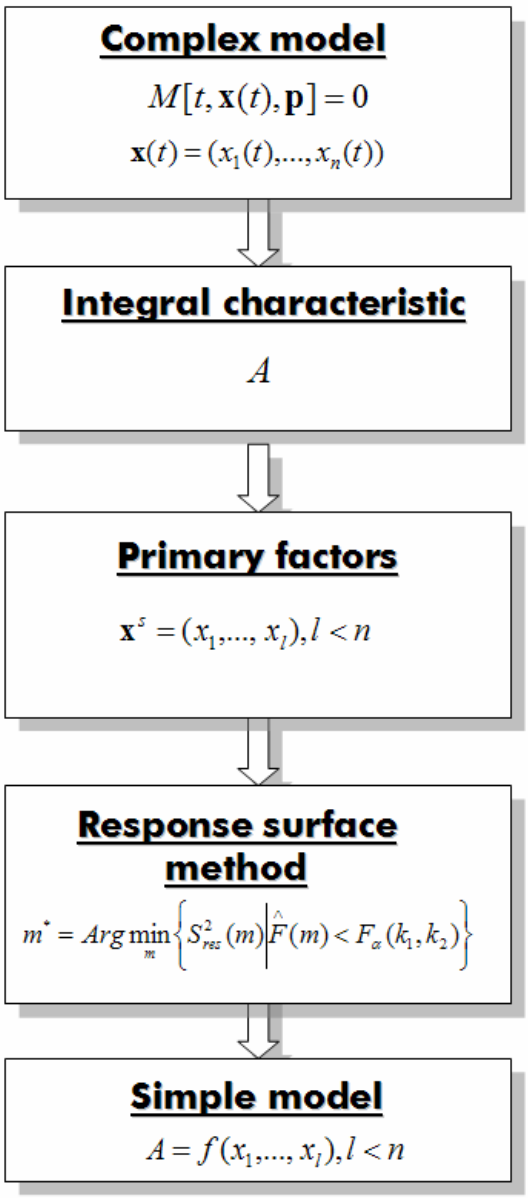


**Figure 1**. Model simplification process.

## 1. INTRODUCTION

In environmental decision making, we are dealing with ecological systems whose behaviour is highly complex, with dynamics and feedbacks spanning multiple spatial and temporal scales (e.g., Levin 1999). In addition, all ecological systems are now impacted by human actions. As noted by Vitousek *et al*. (1997), human alteration of Earth is substantial and growing. Thus, according to the most recent estimates by the World Wildlife Fund (WWF 2006), the demand on planet's ecosystems (the so-called ecological footprint index) has more than tripled since 1961 and now exceeds the world's ability to regenerate by about 25 percent. There is an obvious and urgent need to carefully foresee the likely consequences of societal development for the planetary leaving systems (e.g., Clark *et al*. 2001).

Predictions of this kind are, however, possible only if there is an adequate model describing ecosystem dynamic behaviour and its variations as a result of anthropogenic stresses. It is true that quantitative models play now an important role in all of the sciences. But the importance of models in environmental management, where experiments on real world objects are significantly limited, if not entirely forbidden, is hard to overestimate. Models in ecosystem science serve a number of functions. They allow the investigators to test hypothesis, to uncover patterns embedded in observation data, to synthesize data on disparate components into an integrated view of ecosystem functions, and ultimately to predict the future behaviour of some aspects of the ecosystem under given scenarios of future external drivers (Canham *et al*. 2003).

There is a great deal of discussions among the researchers on the issue of advantages/disadvantages of simple or parsimonious models versus complex models in ecosystem science. In this paper, we review the main pros and cons of either approach to modelling philosophy.

The main issue is to provide practitioners with a model of a descriptive and forecasting power sufficient for the purposes of environmental decision making and management. In this connection, a problem of the transition from complex sophisticated models to their simplified versions has a practical significance for ecological science. The paper presents an approach to the development of a simple model based on the methods of the theory of experimental design. It is demonstrated that using the response surface method enables to preserve the descriptive capabilities of a complex simulation model.

The problem of environmental management in a forested watershed and simulation model "Forest hydrology" (SMFH) are used as a case study and illustration to the approach suggested.

## 2. DEFINITIONS

The state of an ecological system at a given moment in time, $t$, can be described by a phase vector $\mathbf{x}(t) = (x_1(t),\ldots, x_n(t))$. The coordinates of vector $\mathbf{x}$ represent quantitatively the components (or sub-systems) of the ecosystem, such as species numbers, concentrations of substances, biomass, primary production/destruction/respiration, dissolved oxygen, nutrients, *etc*. Following a unified notation proposed by Ide *et al.* (1997), a model for the natural evolution of the system is governed by an equation:

$$M[t, \mathbf{x}(t), \mathbf{p}] = 0 \qquad (1)$$

with the initial conditions $\mathbf{x}(0) = \mathbf{x}_0$. Here $M$ is the model dynamics operator and $\mathbf{p}$ is the vector of model parameters. Depending on the aim of the research, a particular ecosystem being modelled and observation data available, the operator $M$ may be in a form of an algebraic expression, or differential or integral operator. In many cases, the operator $M$ of the system evolution (1) in a computer simulation or prediction is nonlinear and deterministic, while the true trajectories of the modelled system may differ from Eq. (1) by random or systematic errors.

## 3. STEPS IN MODEL BUILDING

Any model in environmental management is expected to represent a real world ecological system or some of its aspects of a particular interest. At the same time, a model is unavoidably a simplification of objective reality (Straškraba and Gnauck 1985). Due to the complexity of the real ecological systems, the model always reflects only substantial properties of the system rather than all its details. Generally, model development involves the sequence of required stages (Fig. 1). The stage of model design includes: (1) development of a conceptual model (i.e., selection of major variables and processes); (2) the model mathematical description in the form of equations; (3) parameterization (i.e., determination of quantitative values of model parameters); and (4) coding (i.e., translation of the mathematical equations into computer based software). The model testing stage includes: (1) simulation runs; (2) verification (i.e., comparison of results obtained from model simulation with values observed in the system); (3) validation (i.e., a proof that the reactions and dynamics generated by

the model are similar with the behaviour of the real system); (4) stability analysis of simulation results (i.e., investigation of model reactions to perturbations of initial values of model variables), mostly in a sense of the Liapunov stability; (5) sensitivity analysis to major parameter (i.e., a series of tests in which the modeller varies the values of model parameters to see the corresponding changes in model outputs); and (6) uncertainty analysis (i.e., a measure of an error in model simulation of given observations due to parameters, state variables and model structure).

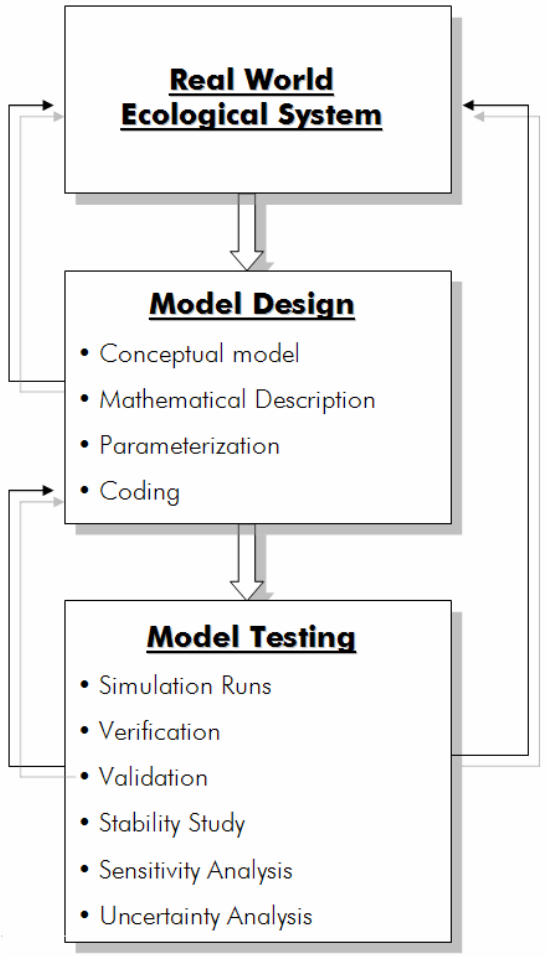


**Figure 1**. Model building stages.

## 4. PROBLEM OF MODEL COMPLEXITY

### 4.1. Large vs. small models

Environmental modelling deals with complex dynamic systems whose behaviour is determined by many interrelated phenomena. A modelling description of such a system may require hundreds of equations using more than thousand parameters. The main drawbacks of huge simulation models are: (1) their high cost and long term of development and (2) difficulties with communicating to and understanding by decision and policy makers and general public. There is also a fundamental problem of parameterization of models (Jørgensen 1994) which becomes even more complicated the larger the model is and the more parameters it, therefore, involves. Many parameters are usually unknown, or their literature values are highly variable (Van Nes and Scheffer 2005). In addition, a non-linear model has more than one steady state solution. Different values of parameters can alter the stability properties of the solutions making the application of automatic parameterization procedures difficult or even impossible. On the top of uncertainty in parameters (Lindenschmidt 2006, Snowling and Kramer 2001), validation of large models is also problematic (Oreskes *et al*. 1994; Rykiel 1996). The uncertainty in the initial and boundary conditions, though being usually neglected (Van der Molen 1999), also contributes to the overall model uncertainty. The interplay between model complexity and overall uncertainty delivered by all model components should also be taken into account (Erechtchoukova 2005, Reichert and Borsuk 2005).

Interestingly to note that ecological models are often validated only by comparison with empirical data. However, the ability to predict independent past or future data alone does not mean that the model adequately describes the cause and effect relations of the real system (Rykiel 1996).

In contrast to the large models that are meant to represent many features of the complex dynamics of real world ecosystems, small simple models often focus on a single aspect or phenomenon of the system in study and are based on clear assumptions. They usually consist of a few equations with little parameters (e.g., Lotka-Volterra model to describe the predator-prey dynamics of biological systems). It is, of course, easier to present and explain the results generated by minimal models. The main problem with small models in environmental simulation modelling, however, is that they are too generic and neglect many essential aspects of the real ecosystems. As a consequence, minimal models are usually very hard to test and to validate with empirical data as noted by Van Nes and Scheffer (2005). In this connection, an opinion had been expressed that complex systems need complex solutions (Logan 1994). Therefore, the task is to find a way to retain a descriptive power of a large complex model when replacing it by a simplified version. The latter goal can be achieved using, for example, methods of the statistical design of experiments (Lorenzen and Anderson 1993) as will be further demonstrated in this paper.

### 4.2 Design of experiments and simplification

Application of a simulation model requires multiple computer experiments with various initial values of the input variables $\mathbf{x}_0$ and model parameters $\mathbf{p}$ in Eq. (1) as well as computations over a number of modelling steps. This circumstance leads to the necessity to store huge sets of intermediate simulation results and their subsequent processing.

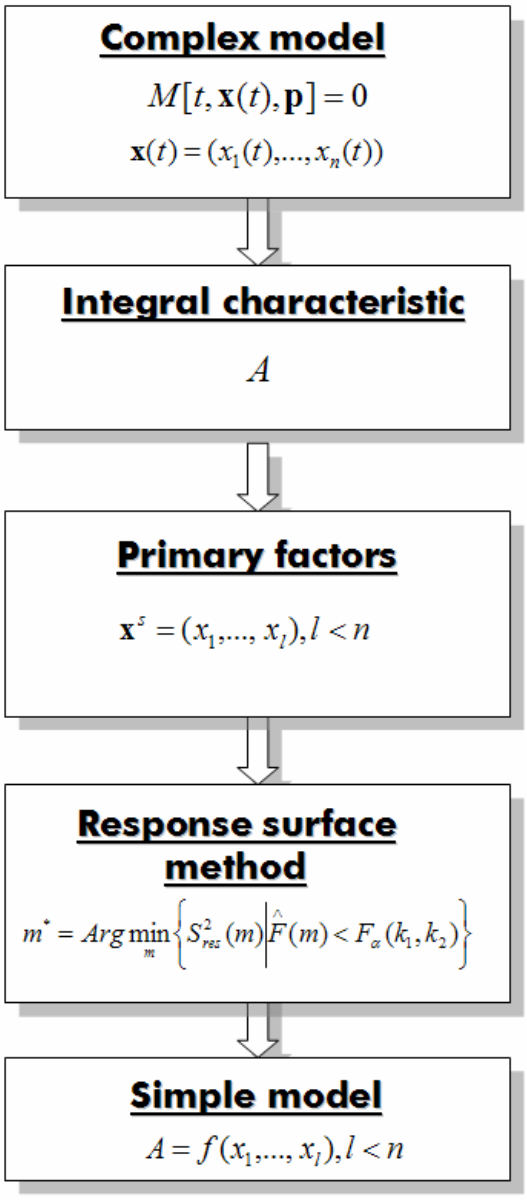


**Figure 2**. Model simplification process.

The complexity of simulation models can be reduced. Usually, there is no practical need to store all the values generated in each simulation run, but only some aggregated data of particular interest. Moreover, in many practical cases of environmental decision-making, it is important to trace and predict a certain integral indicator of an ecosystem in study denoted hereafter as $A$. For example, in the lake ecosystem management when dealing with the eutrophication problem, such an integral indicator is water turbidity.

It can be further concluded that, depending on the task of modelling, some of the state variables are more important for the decision makers while others play a secondary role. In the already mentioned problem of lake eutrophication, concentrations of nutrients (first of all, phosphorus and nitrogen) or total phytoplankton biomass are important state variables while hydrodynamic characteristics of water flow are secondary factors. We may always assume that the first $l$ components of vector $\mathbf{x}$ are primary state variables. It is possible to consider the integral indicator $A$ as a resulting function of just a limited number of the most important (primary) state variables:

$$A = f(x_1, ..., x_l), l < n \qquad (2)$$

Such a resulting function can be determined using the methods of experimental design as a response surface in the space of selected primary state variables: $\boldsymbol{x}^s = (x_1, \ldots, x_l)^T$, $l < n$. In this case, from the mathematical point of view, the problem is to choose the response surface which will best approximate the model outputs generated by the initial large simulation model (1).

Statistical techniques of orthogonal designs, introduced by Box and Wilson (1951) for second-order response surfaces, can be used in environmental simulation modelling. The logical scheme of the approach is shown in Fig. 2.

## 5. CASE STUDY: FOREST HYDROLOGY

### 5.1. Background

It is important for environmental management to predict the possible changes in the hydrological regime of a given forested watershed that may occur as a result of projected anthropogenic activities.

Ideally, in order to address this problem, it is necessary to compare the differences between the components of the water balance in an experimental watershed before and after a particular management practice has been carried out. But it is clear that such data cannot be obtained for most forested watersheds, since each

one is unique in terms of its morphology, hydrology, landscape, and vegetation. Hence the results obtained on some experimental watershed are not always directly applicable to other watersheds even if they are located within the same geographical zone, on comparable soils and are approximately of the same size.

The only way to overcome this methodological obstacle is to use a comprehensive simulation model of the processes of moisture transformation in a studied forested watershed (Khaiter 1993).

We consider the simulation model "Forest hydrology" (SMFH) as a case study to illustrate the suggested approach to simplification in environmental modelling.

### 5.2. Model description

The SMFH takes as its inputs a limited set of parameters and initial values of the state variables, such as forest vegetation type and age, percent forested area, meteorological data, soil properties, and projected management activity. Based on this input information, the model simulates the processes of moisture transformation in a boreal forested watershed and calculates precipitation interception, evaporation from snow and water, throughfall, snowmelt, water release from snow, freezing and thawing of soil-grounds, infiltration, transpiration and the formation of all kinds of runoff. These processes are modelled at three levels: tree crown, forest floor, and a specified soil layer. The model produces as outputs the values of the water balance components and provides a quantitative assessment of the hydrological role of forest as a result of the management activity being studied.

The SMFH represents the distribution of precipitation using the following water balance equation:

$$PR = EVC + EVF + EVS + Q_{SUR} + Q_{SUB} + TR + \Delta SM + Q_{GR} \quad (3)$$

where $PR$ is atmospheric precipitation; $EVC$, $EVF$, and $EVS$ are evaporation from canopy, floor and soil, respectively; $Q_{SUR}$, $Q_{SUB}$ are surface and sub-surface fluxes, respectively; $TR$ is transpiration; $\Delta SM$ is the variation of soil moisture contents; and $Q_{GR}$ is water recharge to the groundwater table. The model considers moisture transformation at three levels (or *hydrological niches*): (1) tree crown, (2) forest floor, and (3) soil layer of a given thickness, $Z$. The balance condition should obviously be satisfied for each of the hydrological niches:

$$\frac{dN^j}{dt} = \sum_i INC_i^j - \sum_k OUT_k^j \quad (4)$$

where $j$ denotes a hydrologic niche ($j$ = 1, 2, 3); $N^j$ is the moisture contents in the $j$th hydrological niche; $INC_i^j$, $OUT_k^j$ are the $i$th income and $k$th outcome water balance item, respectively, for the $j$th hydrological niche.

The formulation of the notion of the *hydrological function of a forest* and its estimate, $\Delta QUSE$, was proposed (Khaiter 1993) in the form of the following expression:

$$\Delta QUSE = \sum_{t=1}^{T} \left\{ \left[ \Delta SM^f(t) + Q_{SUB}^f(t) + Q_{GR}^f(t) \right] - \left[ \Delta SM^o(t) + Q_{SUB}^o(t) + Q_{GR}^o(t) \right] \right\} \quad (5)$$

where the superscripts $f$ and $o$ denote forested and open (forestless) watersheds, respectively; $t$ is the time variable; $T$ is duration of a specified time interval.

The SMFH described above computes the values of major components of the water balance (3), and in the current implementation has an iteration step $\Delta t$ = 1 day.

### 5.3. Model simplification

In many practical cases, management of a forested watershed requires to predict the changes of the integral annual indicator $\Delta QUSE$ from (5) that might result from one anthropogenic impact or another. To achieve the necessary aggregation, $\Delta QUSE$ can be considered as an integral function of the primary state variables as described in (2):

$$\Delta QUSE = f(x_1, ..., x_l). \quad (6)$$

These primary state variables have been used as factors in the design of computer simulation experiments with the SMFH. Two factors (soil density, $SD$, and percent forested area, $F\%$) were taken into account from the practical point of view. This selection was determined by the fact that forest-management activities will first affect these two characteristics of an ecosystem (cf. Khaiter, 1991). The algorithm developed builds the response surface [i.e., function $f$ from (6)] which summarizes the data generated by simulation

experiments with the large initial model in the best possible way. First- and second-order models with and without autocorrelation of factors have been investigated as a suitable approximation to the true response surface. For statistical evaluation, the series of parallel simulation runs have been done for each design point by varying an input variable. The order of the best suitable polynomial, $m^*$, was selected using the condition:

$$m^* = Arg \min_m \left\{ S^2_{res}(m) \middle| \hat{F}(m) < F_\alpha(k_1, k_2) \right\} \quad (7)$$

where $S^2_{res}$ is the residual variance; $\hat{F}$ and $F_\alpha$ are the sampling and the critical values of the *F*-test, respectively, with Fisher's distribution; $k_1$ and $k_2$ are degrees of freedom; and α is the significance level. As a result, the following response function was obtained:

$$\Delta QUSE = -542.9 SD + 31.8396 F\% - 14.378\ SD\ F\% + 922.9 \quad (8)$$

The resulting response surface is shown in Fig. 3.

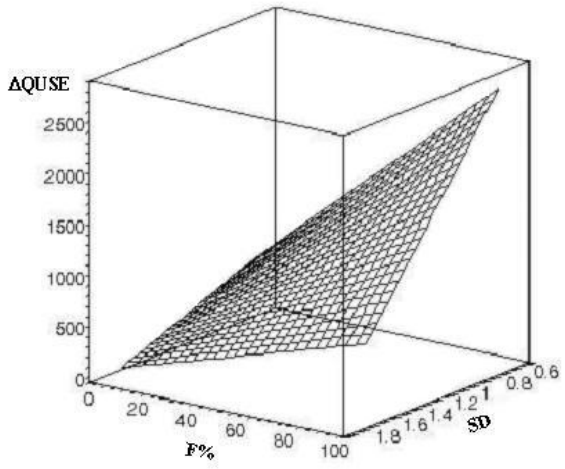


**Figure 3**. Response surface of the function Δ*QUSE*.

Eq. (8) can be recommended as a simplified model for practical calculations.

## 6. CONCLUSION

In the paper, we discussed the application of simulation modelling for the practices of environmental management and the stages in a simulation model design and development. It is expected from any model used in environmental decision making to represent the real world ecological systems.

The accuracy of simulation results is important for an informed decision making and in many cases can only be achieved through the experiments with a sophisticated model involving a large number of state variables and even larger number of parameters. At the same time, there are serious drawbacks of the large simulation models which we analysed in the paper. Because of these problems with the sophisticated models, decision makers very often prefer to rely on simpler models as they are easier to utilize and present to the general public, of course, if a smaller model is able to describe the behaviour of the real world systems sufficiently well. Therefore, one of the key issues in the transition from a complex model to its simplified version is to retain the descriptive capabilities of the large simulation models.

The study presented an approach to the construction of a smaller model based on the complex simulation model developed for a particular environmental problem. The approach suggested does not eliminate the necessity to build a complex model. Moreover, a complex model plays an important role in the approach. But parameterisation and other required in the model building steps are performed only once for a large model and are not needed in order to apply a simple model to various management tasks related to the investigated real world system.

The approach is based on the idea of classification of the components of the model state vector into primary and secondary variables. The former are more important for a given practical task of environmental decision making and are used as the factors in response surface design aimed at the construction of a resulting function which is a statistically valid approximation of the solutions delivered by the complex model. The resulting response surface equation serves as a simple model suitable for practical calculations. It is important to note that the simple model so obtained retains the descriptive power of the initial comprehensive simulation model because it does not ignore secondary variables and, thus, does not introduce additional uncertainty to the modelling results.

## 7. ACKNOWLEDGEMENT

The authors wish to thank anonymous reviewers for their helpful suggestions and comments on the earlier versions of the manuscript. Part of this work was supported by the Atkinson Faculty

Council through the 2006-2007 and 2007-2008 individual grants to each co-author.

## 8. REFERENCES

Box, G.E.P., and K.B. Wilson (1951), On the experimental attainment of optimum conditions, *Journal of the Royal Statistical Society*, Ser. B., 13(1), 1-45.

Canham, C.D., J.J. Cole, and W.K. Lauenroth (2003), Models in ecosystem science, *Models in ecosystem science*, Princeton University Press, pp. 1-10, Princeton, New Jersey.

Clark, J., S. Carpenter, M. Barber, S. Collins, A. Dobson, J. Foley, D. Lodge, M. Pascual, R. Pielke Jr., W. Prizer, C. Pringle, W. Reid, K. Rose, O. Sala, W, Schlesinger, D. Wall, and D. Wear (2001), Ecological forecast: an emerging imperative, *Science*, 293(5530), 657-660.

Erechtchoukova, M.G. (2005), Uncertainty transformation in ecological simulation models, International Congress on Modelling and Simulation MODSIM 2005 (A. Zerger, and R.M. Argent – eds.), December 2005, Modelling and Simulation Society of Australia and New Zealand, pp. 2477-2483, Melbourne.

Ide, K., P. Clourtier, M. Ghil, and A.C. Lorenc (1997), Unified notation for data assimilation: operational, sequential and variational, *Journal of the Meteorological Society of Japan*, 75(1B), 181-189.

Jørgensen, S.E (1994), *Fundamental of Ecological modelling*. *Developments in Environmental Modelling*, 19, 2nd ed., Elsevier, 628 pp., Amsterdam.

Khaiter, P.A. (1991), *Modelling of the anthropogenic dynamics of forest biogeocenosis*, Znaniye, 24 pp., Kiev.

Khaiter, P.A. (1993), Mathematical modelling in the study of the hydrological regime in 'Forest-Watershed' system. Proc. Int. Conf. on Computational Methods in Water Resources, Boulder, Colorado, CMP, pp. 789-794, Southampton, UK.

Levin, S.A. (1999), *Fragile dominion: complexity and the commons*, Perseus Books, 250 pp., Reading, Mass.

Lindenschmidt, K.-E. (2006), The effect of complexity on parameter sensitivity and model uncertainty in river water quality modelling, *Ecological Modelling*, 190, 72-86.

Logan, J.A. (1994), In defence of big ugly models, *Transactions of the American Entomological Society*, 40, 202-207.

Lorenzen, T.J., and V.L. Anderson (1993), *Design of experiments: a no-name approach*, Marcel Dekker, 414 pp., New York.

Oreskes, N., K. Shraderfrechette, and K. Belitz (1994), Verification, validation, and confirmation of numerical models in the earth sciences, *Science*, 263, 641-646.

Reichert, P., and M.E. Borsuk (2005), Does high forecast uncertainty preclude effective decision support? *Environmental Modelling and Software*, 20, 991-1001.

Rykiel Jr., E.J. (1996), Testing ecological models: the meaning of validation, *Ecological Modelling,* 90, 229-244.

Snowling, S.D., and J.R. Kramer (2001), Evaluating model uncertainty for model selection, *Ecological Modelling*, 138, 17-30.

Straškraba, M., and A.H. Gnauck (1985), *Freshwater ecosystems. Modelling and simulation. Developments in Environmental Modelling*, 8, Elsevier, 309 pp., Amsterdam.

Van der Molen, D.T. (1999), *The role of eutrophication models in water management,* Wageningen University, 167 pp., Wageningen, The Netherlands.

Van Nes, E.H., and M. Scheffer (2005), A strategy to improve the contribution of complex simulation models to ecological theory, *Ecological Modelling*, 185, 153-164.

Vitousek, P.M., H.A. Mooney, J. Lubchenko, and J.M. Melillo (1997), Human domination of Earth's ecosystems, *Science*, 277(5325), 494-499.

World Wildlife Fund For Nature (WWF) (2006), *Living Planet Report*, WWF International, 44 pp., Gland, Switzerland.