Uncertainty Transformation in Ecological Simulation Models

[1]Marina G. Erechtchoukova

[1]York University, Canada, E-Mail: marina@yorku.ca

Keywords: *Model uncertainty; Sensitivity analysis; Compartment model; Water flow; Water quality.*

EXTENDED ABSTRACT

Being simplified representations of reality, simulation models can never be perfect and their results are always somewhat uncertain. That is why quantification of model uncertainty is important during interpretation of simulation results in decision making process. Uncertainty describes deviations of simulated ecosystem's characteristics from known or observed values. Several sources contribute to such deviations including those associated with main model components such as forcing functions, mathematical formulae, parameters and universal constants and intrinsic model features. Model uncertainty can be evaluated based on its linear estimate under the assumption that all sources of uncertainty are independent. Traditional approaches to investigating model uncertainty consider individual sources whose contribution to the uncertainty can be quantified for a given task. Although the result is incomplete it helps to improve the understanding of the model and increase the confidence in simulation results.

One of the most investigated sources of model uncertainty is errors in a model's supporting data. The paper describes this type of uncertainty and how it is transformed into simulation results. Propagation of this type of model uncertainty through the model was investigated based on a simulation model developed for water quality assessment in the mouth of a large river.

The model was created using simulation framework which consists of at least two independent modules. Each module describes specific groups of processes, e.g., hydrological processes or hydrochemical and hydrobiological processes. Interactions of processes from different groups are modeled by passing simulation results from one module to another.

The hydrodynamic module is built as channel – junction computational network based on equations of continuity and momentum in integral form. Each junction in this module corresponds to a homogeneous compartment of water quality module where chemical and biological processes take place. The water quality module describes phytoplankton dynamics, organic matter and nutrients transformation and uptake. These processes are modeled by ordinary differential equations. Channels carry interactions of compartments, that involve water and ingredients exchange. Simulations were aimed at detection of crucial parameters of the ecosystem and evaluation of uncertainty in simulation of phytoplankton biomass during vegetation.

A sensitivity analysis was conducted to reveal the most critical parameters of hydrodynamic and water quality modules. It appeared that hydrodynamic module is sensitive to the lower boundary condition of the modeled portion of the river. Variations of this condition within 10% caused up to 60% variations in water redistribution between river branches. In general, it is expected that if model components are sensitive to some subcomponents, then model state variables are also sensitive to these components. Therefore, the high uncertainty of the hydrodynamic module raised an expectation that water quality characteristics in the investigated portion of the river, and particularly phytoplankton biomass, are also simulated with a high level of uncertainty caused by errors in lower boundary hydrodynamic condition. Simulation experiments showed, however, that propagated variations in water flow characteristics were reduced to 2% deviations of phytoplankton biomass in water quality module.

The case study presented illustrates the ability of one module of the model to absorb the uncertainty of another module. A similar effect can be observed for another dynamic model built upon differential equations, and applied within the same or similar framework. The structural stability of the module accepting parameters supplied by other modules (from the same model) can improve simulation results significantly. Results also suggest an approach to evaluation of model uncertainty that can be computationally efficient.

1. INTRODUCTION

The term “uncertainty” in its broader sense is reserved in scientific literature to describe the imperfection of human knowledge about the reality. In this paper, the term “uncertainty” is used specifically with respect to simulation models and their properties. The research is based on the acceptance of two facts. First, that model uncertainty is unavoidable, and second, that uncertainty of results can be minimized by selecting the model with specific properties.

Simulation models are powerful tools in investigation of natural systems, specifically predictions of natural disasters and evaluation of sustainable management decisions of various scales. Ecological models form a diversified set of techniques based on different mathematical and computational methods. Methodologies for model application have also been proposed and successfully applied (Van Nes and Schefer, 2005; Argent, 2004). The experience in dealing with complex simulation models have brought up issues associated with model complexity and imperfection of the models and available observation data (Van Nes and Scheffer, 2005; Beck, 1999; Reichert and Omlin, 1997).

The key question of any application of a simulation model is the extent to which users can trust the model and the results it produces. Model suitability is normally studied on the step of model validation. Along with qualitative assessment of how well the model describes cause-effect relationships of the real system, quantitative estimations of model suitability are highly desired. One of the possible articulations of the problem is an evaluation of model uncertainty. Intuitively, model uncertainty is understood as a possible deviation of predicted values of system characteristics obtained via simulations from their observed values.

The evaluation of model uncertainty can be done in different ways, e.g., using Monte Carlo simulations (Waller *et. al.*, 2003), Bayesian statistics (Borsuk *et.al.*, 2004), or differential analysis (Turanyi and Rabitz, 2000). Selection of a method is predetermined by the type of a model being investigated. Numerous publications provide us with taxonomies of ecological simulation models based on various criteria (e.g. Strashcraba and Gnauck, 1985). Among main classification criteria one can find types of mathematical methods utilized, dynamic features of models, model spatial characteristics or intended nature of application. Considerable part of simulation models designed for an assessment of ecological conditions of natural objects and evaluation of various management strategies in terms of their sustainability is constructed based on differential equations. The paper presents results of uncertainty analysis of a dynamic spatially distributed simulation model built for water quality assessment in streams and reservoirs. The state of aquatic ecosystem is described by a compartment model where ecosystem characteristics in each compartment are modeled using differential equations. The modeling framework adopted supports a wide range of simulation models that can be applied to numerous tasks of water resource evaluation and management. Therefore, conclusions derived from the investigated case study can be extrapolated onto other models of the same type employed under the same framework. The case study exploring causes of eutrophication in a branched river system demonstrates the propagation of uncertainty in model supporting data through the model and the transformation of the uncertainty in simulation results.

2. ISSUES OF UNCERTAINTY ANALYSIS

Commonly recognized sources of model uncertainty include main model components identified based on its mathematical formulation (Jorgensen and Bendoricchio, 2001) and the model itself as an abstract representation of reality. Model components include state variables, forcing functions, mathematical equations and formulae, parameters and universal constants. State variables represent quantitative characteristics describing ecosystem state. A set of such characteristics depends on an ecosystem modeled and a particular task utilizing the model. Forcing functions have an external nature with respect to state variables and affect state variables. Mathematical equations, formulae and methods are tools selected to describe processes associated with state variables and deemed important for practical application. These tools contain parameters or coefficients that may vary in time and space, but are independent from state variables. All five components cannot be determined precisely or unambiguously. Any attempt to assign a value or to specify a mathematical expression introduces an error thus influencing simulation results and contributing to the model uncertainty.

Another commonly accepted source of uncertainty is an intrinsic model feature. As a simplified representation of reality, any ecological model never contains all characteristics of a real ecosystem and never describes all processes that take place in the ecosystem. In other words, the lack of our knowledge and expressiveness of selected tools also contribute to model uncertainty.

All these sources are sometimes enumerated under the term "epistemic" uncertainty that arises from the ignorance of a true fact or a value (Regan *et.al.*, 2003).

Given that model uncertainty is unavoidable, evaluation of its extent is required, and intuitive definition of model uncertainty mentioned above is of great help. Uncertainty from a particular source can be estimated based on an approximation of the deviation of output variables simulated with disturbed components from unperturbed state (Turanyi and Rabitz, 2000). Similar to computational methods, one can try to find upper boundary of possible deviations in simulation results from corresponding observed values, although the upper boundary may be never achieved. Then, enumeration of main sources of model uncertainty may give a straightforward algorithm for uncertainty evaluation, namely, to evaluate model uncertainty from each source separately, if it is possible, and add them altogether. Such an algorithm provides a linear estimate of model uncertainty and is widely applied when model uncertainty is investigated under the assumption that all sources of uncertainty are independent.

Unfortunately, practical application of this type of algorithm to ecological models is very limited. Model uncertainty cannot always be apportioned to each source separately. In this case, an attempt to evaluate uncertainty from each source separately results in overestimated values. Moreover, different sources of model uncertainty are not independent and therefore the total uncertainty is not an additive function of uncertainties from enumerated sources. Finally, not each source of model uncertainty can be evaluated. While intrinsic model uncertainty is very well understood, quantitative assessment of this source is hardly possible.

Evaluation of model uncertainty is often conducted along with sensitivity analysis. According to (Jorgensen and Bendoricchio, 2001) "sensitivity analysis attempts to provide a measure of the sensitivity of either parameters, forcing functions, initial values of the state variables or submodels to the state variables of greatest interest in the model." Sensitivity analysis identifies specific model components whose contribution to model uncertainty dominates others. There are several approaches to estimate component sensitivity. The simplest one is a first order local sensitivity analysis which is based on the investigation of first-order partial derivatives. If more than one state variable is investigated, they may have significantly different scales. To compare the reaction of different state variables to perturbations in model components, the sensitivity is calculated based on relative deviations in parameter and state variable values. Mathematical definition of component sensitivity clears relationships between the latter and model uncertainty due to a certain model component. Component sensitivity can be interpreted as relative rate of the model uncertainty due to the component.

Evaluation of model uncertainty caused by different sources improves the understanding of model features and lifts confidence in simulation results even if it cannot be done for all sources. Component uncertainty propagating through the model transforms into model uncertainty. The result of transformation is obviously predetermined by the model's mathematical features. With this respect it seems reasonable to investigate specific features of a given model to obtain reasonable evaluation of at least some portions of model uncertainty. Estimates of uncertainty in model input data can be obtained relatively easy when the sampling and analytical procedures are known. Propagation of the input data uncertainty through the model and quantification of its contribution to model uncertainty seems to be important especially because it is expected that if model components are sensitive to some subcomponents, then model state variables are sensitive to these components (Jorgensen and Bendoricchio, 2001).

3. CASE STUDY

The study of uncertainty propagation and transformation through the compartment model was conducted based on the model developed for lower portion of one of the largest Eastern European rivers, the Don River. The Lower Don lies in the Rostov Region from Tsimlyansk Reservoir to Taganrog Gulf, with a watershed area of more than 160,000 km^2 and a length of 313 km (Figure 1). The river width varies from 400 to 600 m in the lower part. Average water depth during low-water season is 2-6 m in the main channel, and decreases to 0.7 m within the shoals. High-water events increase water depth up to 9 m. The mouth of the Don River lies downstream from Rostov-on-Don. Its watershed covers an area of about 340 km^2, and includes many branches and creeks, the biggest of which are Stary Don and Bolshaya Kalancha. This portion of the Don River is the main source for municipal and industrial water supply for the city of Rostov-on-Don (population over 1,000,000 people), and smaller towns situated on the river's banks.

Data from a routine water quality monitoring system were used to sketch water quality conditions in the region. The hydrochemical composition of the water in this portion of the river forms under the influence of natural and anthropogenic factors. The water comes to the system from Tsimlyansk Reservoir and a few tributaries. As the water flows downstream, its chemical composition can be affected by the loading of both dissolved and suspended matter. The dissolved oxygen regime of the river is stable. The concentrations of heavy metals, phenols, and synthetic surfactants slightly exceed corresponding maximum allowable concentrations. Water quality problems of the Lower Don system are associated with the processes of eutrophication and contamination by organic matter. Besides, high turbidity may influence essentially the Don River ecosystem and hence slow down the self-purification processes.

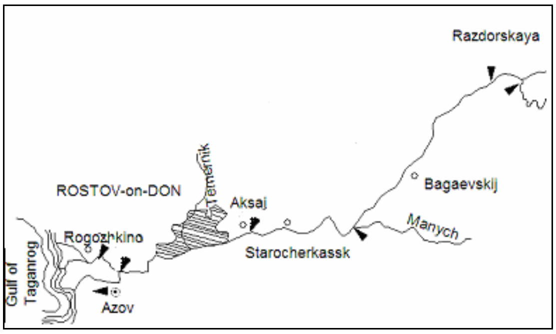

Figure 1. The studied portion of the Don River

4. MODELLING FRAMEWORK AND INVESTIGATED MODEL

Ecological models normally include state variables describing ecosystem characteristics and undergoing physical, chemical and biological processes. While partial differential equations allow modelers to describe chemical and biological transformations along with diffusion and advection processes at the same time (e.g., Markman and Erechtchoukova, 1983), the commonly accepted practice is to separate main groups of processes and to model them using individual modules. Interactions of processes from different groups are modeled by passing the simulation results from one module to another (Ambrose *et.al.*, 1993). In the case of an aquatic ecosystem, water flow is the main transport mechanism playing a critical role in ecosystem's dynamics and viability. That is why water flow characteristics are required for water quality simulation.

The selected class of compartment models is based on the following framework. A water body is mapped into channel – junction computational network. Each junction is a homogeneous compartment where chemical and biological processes take place. These processes are described by ordinary differential equations. Channels carry interactions of compartments that involve water and ingredients exchange. Mass transport between compartments is calculated based on channel flows which are characterized by water discharges and/or velocities. Normally such data are available at certain points in the water body, but temporal and spatial frequencies of observed data are not sufficient to support such model. That is why water flow characteristics are also obtained from simulations based on hydrodynamic module. The resulting unsteady water flow characteristics for a certain period of time are used later for simulation of aquatic ecosystem characteristics.

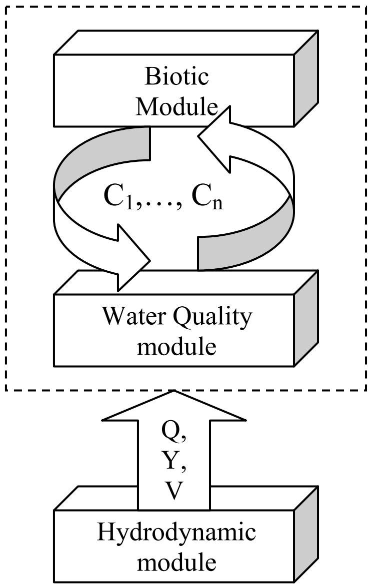

Figure 2. Simulation framework

The framework described may include at least two modules: a module describing hydrodynamic processes and a module simulating chemical and biological processes. Depending on the task, the latter can be further divided into more or less independent modules (Figure 2). The framework presents all the sources of model uncertainty as mentioned above, and raises the question of how the uncertainty propagates from one module to another and transforms through the entire model.

The framework was adopted to simulate primary production growth in the mouth of the Don River described above. Simulations were aimed at detection of crucial parameters of the ecosystem and main source of phytoplankton biomass during vegetation. The model consisted of two modules that computed water flow and water quality characteristics. The area of study was presented as a set of 21 homogeneous compartments – junctions with two branches in the mouth. The hydrodynamic module was designed based on one-dimensional equations of continuity and momentum in integral form. Numerical solution of the equations was obtained using branched junction network and applying four-point approximation scheme with some additional assumptions (Yereschukova, 1997).

Total phytoplankton biomass was selected as a main indicator of the eutrophication process. Other water quality parameters selected as state variables include concentrations of organic nitrogen and phosphorus, inorganic phosphorus, ammonium and nitrate. The selected state variables were modeled using EUTRO module of WASP modeling package (Ambrose *et.al.*, 1993). The model described nutrients dynamics based on Michaelis-Menten kinetic equations including solar radiation and water temperature as forcing functions.

The hydrodynamic module was calibrated for the summer low-water period. After that, the model was run with input data set corresponding to the spring high-water period and fall high-water period. Results showed a good fit with observation data. Thus, the difference between simulated and measured values of water stages did not increase in time and did not exceed 5% of water depth that varied between 4 m and 4.6 m at the observation site.

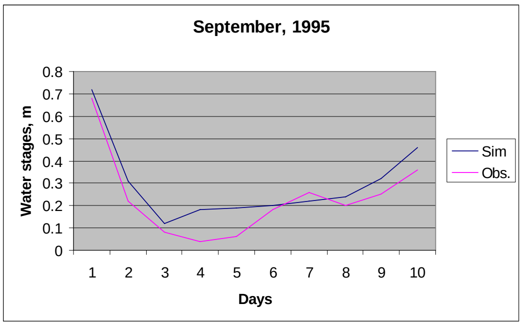

Figure 3. Water stages near Aksai

Water discharges and stages were simulated for the period of time from May to October for the year with average hydrological regime. The results of simulation supported a conclusion that the module describes different hydrological events of the river with the required level of accuracy (Figure 3). The deviation of calculated values from observed data did not exceed guaranteed accuracy for the measurements. Higher deviations at the lower boundary of the portion of the river suggested looking into model uncertainty due to errors in boundary conditions.

The water quality module was calibrated using observation data collected by routine monitoring system. Computed hydrodynamic characteristics of the water flow were passed to the water quality module. The model described seasonal dynamics of nutrients satisfactorily for the whole studied portion of the river. Total calculated phytoplankton biomass changed from a relatively small amount in spring up to 3 mg/l in summer, which corresponds to average measured values. Numerical stability analysis confirmed that selected basic solution is dynamically stable. To investigate sensitivity of this module the following parameters were selected: maximum growth rate, mortality rate, rates of decomposition of organic phosphorus and nitrogen, light saturation coefficient, carbon-to-chlorophyll ratio, and water flow. Results demonstrated that the model is structurally stable. At the same, time phytoplankton biomass is most sensitive to carbon-to-chlorophyll ratio and light intensity.

5. UNCERTAINTY ANALYSIS

Obviously, the constructed model has intrinsic uncertainty. It is expected to imitate only main features of ecosystem characteristics. Global constants, namely acceleration of gravity and Manning's roughness coefficient are also included in the model. But these sources of model uncertainty are not considered in the study.

The next source of the uncertainty – errors in initial values of state variables – needs to be discussed. The underlying Saint Venant equations form a dynamic system that cannot be solved or investigated analytically. The framework described above assumes only numerical analysis which is accomplished as a series of model runs with different initial values. According to the classic Lyapunov definition of stability, the unperturbed state of the system is stable if small perturbations of initial values of state variables result in small deviations of perturbed state from unperturbed one (Leipholz, 1970). Keeping in mind that numerical solutions accumulate computational errors, one can conclude that uncertainty and sensitivity analysis can produce valid results only for stable unperturbed states. In the case of unstable solutions, it is hard to distinguish between the impact of variations in parameter values and the effect of accumulated computational errors.

To evaluate the propagation of errors in the lower boundary condition, the series of computation experiments were conducted based on full factorial design with deviations of ± 10% of the basic values obtained through observations. Results of module runs were compared using the measure of deviations between basic solution and solutions with perturbed input data sets:

$$\delta_X = \frac{\Delta \Big/ \frac{1}{N}\sum_{i=1}^{N} X_b(t_i)}{\Delta p / p} \cdot 100\ \% \qquad (1)$$

where δ_X is the relative measure of changes in a state variable X, %, $X_b(t_i)$ is the value of the state variable at the time t_i, obtained via simulations with basic input data, N is the number of time steps during the simulations, and $\Delta p/p$ is the relative change in the parameter value. Δ is the absolute measure of deviations, and is calculated based on the following formula:

$$\Delta = \frac{1}{N}\sum_{i=1}^{N} \left| (X_b(t_i) - X(t_i) \right| \qquad (2)$$

Here, $X(t_i)$ is the value of the state variable at the time t_i obtained from simulations with modified input data. Results of the hydrodynamic module uncertainty due to the lower boundary condition presented in Table 1 correspond to a hydrological regime with average water discharge about 600m^3/s.

Table 1. Uncertainty in water flow passing through the Stary Don branch.

Run	Water flow Q, %	Δ, m^3/s	δ_Q, %
(-,0)	36.7	30.8	135
(+,0)	27.1	36.4	161
(0,-)	28.0	30.1	133
(0,+)	40.0	53.9	238
(-,+)	43.2	72.2	337
(-,-)	29.7	18.2	80
(+,+)	34.4	14.7	65
(+,-)	21.1	78.4	345

The first position corresponds to the boundary condition of the Stary Don branch, the second to the Bolshaya Kalancha branch. Sign "-" means decreasing value by the 10%, "+" - increasing by the same amount, "0" - the same value as the basic input data set. It should be noted that, water flow characteristics at the upper and middle part of the Lower Don River are not so sensitive to the changes in the lower boundary condition. Deviations of simulation results on modified input data set from simulations on basic input data set increased, but not more than 5%. According to the basic input data set the Stary Don branch gets 32.3% of water flow passing through the Lower Don River System. The formulae (1) and (2) have been applied to estimate changes in water distribution between branches caused by 10% change in the boundary condition. High values of δ_Q exceeding 100% indicate that uncertainty of downstream boundary condition is magnified by the module. This effect is more significant for the cases, when average water discharges are low. It was found during simulations that in July, when the average water discharge has its minimal value of 400m^3/s, the relative measure of deviations δ_Q reaches its maximal value at 600%.

Uncovered sensitivity of the hydrodynamic module to the low boundary water stages may cause an expectation that water quality parameters can be also considerably affected by the same conditions. This hypothesis has been investigated.

The study showed how water flow influences the dynamics of ecosystem characteristics. Simulations revealed that average travel time in the Lower Don River system did not exceed 5 days which implies the dominated portion of phytoplankton biomass comes to the system from upstream. A sharp drop in water volume passing through the system during low-flow events in summer may increase travel time significantly. Simulation runs with input data corresponding average water discharge about 400m^3/s did not reveal dramatic increase in phytoplankton biomass compared to simulation with average water discharge of 600m^3/s. The added biomass was less than 9%. The only notable difference was in the time lag of reaching maximum concentrations and the beginning of rapid concentrations decrease in the fall.

A series of simulation experiments were done to evaluate a propagation of hydrodynamic module uncertainty due to boundary conditions through water quality module. Water redistribution between main branches in the mouth of the river affected concentrations of phytoplankton biomass in these branches insignificantly. Having up to 60% of uncertainty in water discharge corresponding to water flow in the Stary Don branch, the maximum relative deviation of concentrations of phytoplankton biomass in this branch δ_{Ph} was less than 20%. Therefore, water quality state variables were not sensitive to the uncertainty in low boundary hydrological condition.

6. DISCUSSION

The series of simulations demonstrated how 10% deviations of lower boundary hydrodynamic condition were transformed in 60% deviations of water discharge in the river branch and finally were reduced to 2% deviations in phytoplankton biomass. The result of uncertainty transformation deserves attention. In general, those modules that are sensitive to certain parameters are expected to pass the uncertainty on to other modules and to make model state variables also sensitive to the parameters, but simulation experiments did not support this expectation. From stability analysis perspectives (Leipholz, 1970), stability of a solution depends on parameters' values. Changes in model parameters can trigger the alteration of a system state from stable to unstable causing significant increase in model uncertainty. This implication of stability theory is supported by simulation experiments, (e.g. van Nes and Scheffer, 2003). Model insensitivity to perturbations in parameter values means structural stability of the model within the given parameters domain.

Simulation experiments showed that basic solution of the water quality module is stable and the area of attraction of this solution in parameter space is large enough to cover significant variations in computed values of water discharge and water depth. The model structure did not allow us to investigate parameter space analytically to determine the parameter values when the basic solution loses its stability, but computational experiments clearly indicated that these values are beyond the range of water flow characteristics observed in the given natural aquatic ecosystem. Simulation results indicated that water flow in the lower part of the River is still high, and hence travel time is insufficient for phytoplankton to grow notably in the system. It is simply washed out of the system regardless which river branch it follows. Such interpretation also supports the conclusion that the model is not sensitive to the lower boundary condition.

7. CONCLUSION

The case study presented illustrates the ability of one module of the model to absorb the uncertainty of another module. A similar effect can be observed for another dynamic model built upon differential equations and applied within the same or similar framework. Structural stability of a module accepting parameters supplied by other modules can improve simulation results significantly. The practical implication of the results consists in the top-down approach to conduct the uncertainty analysis. The approach, when the uncertainty of the final module is first evaluated based on an average estimation, is expected to be more computationally efficient and is the subject of further investigations.

8. ACKNOWLEDGMENTS

The author is very thankful to Jim Ascough and anonymous reviewers for their valuable comments and helpful editorial remarks.

9. REFERENCES

Ambrose, R.B., Wool, T.A., Martin, J.L. (1993). The Water Quality Analysis Simulation Program, WASP5. Part A: Model documentation. USEPA ERL, Athens, GA 30613.

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

Beck, M.B. (1999), Coping with ever large problems, models, and data bases. *Water Science and Technology*. Vol. 39, No 4 1-11.

Borsuk, M.E., Stow, C.A., Reckhov, K.H. (2004), A Bayesian network of eutrophication models for synthesis, prediction, and uncertainty analysis. *Ecological Modelling*. 173 (2004) 219 – 239.

Jorgensen, S.E. and Bendoricchio, G. (2001), Fundamentals of Ecological Modelling. Elsevier Science, 530 pp.

Leipholz, H. (1970), Stability theory. Academic Press. New York. 277 pp.

Markman, G.S., Erechtchoukova, M.G. (1983), A spatial synchronization of periodic regimes in a distributed biochemical system. Dynamics of Biological Populations. Gorky State University Gorky. 3 – 6.

Regan, H.M., Akcakaya, H.R., Ferson, S., Root, K.V., Carrol, S., Ginzburg, L.R. (2003), Treatments of uncertainty and variability in ecological risk assessment of signle-species populations. *Human and Ecological Risk Assessment*. Vol. 9, No 4 1 – 18.

Reichert, P, and Omlin, M. (1997), On the usefulness of overparameterized ecological models. *Ecological Modelling*. 95 (1997) 289 – 299.

Straskraba, M., Gnauk, A. (1985), Freshwater Ecosystems. Modelling and Simulation. Elsevier, Amsterdam. 309.

Turanyi, T., Rabitz, H. (2000), Local methods. Sensitivity Analysis. Saltelli *et.al.* (Eds.) 81– 99.

Van Nes, E.H. and Scheffer, M. (2003), Alternative attractors may boost uncertainty and sensitivity in ecological models. *Ecological Modelling*, 159 (2003) 117 – 124.

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

Waller, L.A., Smith, D., Childs, J.E., Real, L.A. 2003, Monte Carlo assessment of goodness-of-fit for ecological simulation models. *Ecological Modelling*. 164 (2003) 49 – 63.

Yereschukova, M.G. (1997), Hydrodynamic and Water Quality Modeling of a River Delta with Limited Available Data. Fish Physiology, Toxicology and Water Quality. Proc of the Fourth Int. Symp., Bozeman, Montana, USA, Sept. 19 – 21, 1995. 269 – 280.