Sensitivity analysis for complex ecological models − A new approach

Vardit Makler-Pick [a,*], Gideon Gal [b], Malka Gorfine [c], Matthew R. Hipsey [d], Yohay Carmel [a]

[a] Faculty of Civil and Environmental Engineering, Technion-Israel Institute of Technology, Technion City, Haifa 32000, Israel
[b] Y. Allon Kinneret Limnological Laboratory, IOLR, P.O. Box 447, Migdal 14950, Israel
[c] William Davidson Faculty of Industrial Engineering and Management Technion, Israel Institute of Technology, Israel
[d] School of Earth & Environment, University of Western Australia, 35 Stirling Highway, Crawley, Western Australia 6009, Australia

ARTICLE INFO

ABSTRACT

A strategy for global sensitivity analysis of a multi-parameter ecological model was developed and used for the hydrodynamic-ecological model (DYRESM−CAEDYM, **DY**namic **RE**servoir **S**imulation **M**odel-**C**omputational **A**quatic **E**cosystem **D**ynamics **M**odel) applied to Lake Kinneret (Israel). Two different methods of sensitivity analysis, RPART (Recursive Partitioning And Regression Trees) and GLM (General Linear Model) were applied in order to screen a subset of significant parameters. All the parameters which were found significant by at least one of these methods were entered as input to a GBM (Generalized Boosted Modeling) analysis in order to provide a quantitative measure of the sensitivity of the model variables to these parameters. Although the GBM is a general and powerful machine learning algorithm, it has substantial computational costs in both storage requirements and CPU time. Employing the screening stage reduces this cost. The results of the analysis highlighted the role of particulate organic material in the lake ecosystem and its impact on the over all lake nutrient budget. The GBM analysis established, for example, that parameters such as particulate organic material diameter and density were particularly important to the model outcomes. The results were further explored by lumping together output variables that are associated with sub-components of the ecosystem. The variable lumping approach suggested that the phytoplankton group is most sensitive to parameters associated with the dominant phytoplankton group, dinoflagellates, and with nanoplankton (Chlorophyta), supporting the view of Lake Kinneret as a bottom−up system. The study demonstrates the effectiveness of such procedures for extracting useful information for model calibration and guiding further data collection.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Computer models of ecosystems are increasingly used in order to predict possible impacts of policy measures prior to their implementation and to achieve a better understanding of these ecosystems (Ford, 1999). Success of ecosystem models is generally examined through comparisons to time-series of field data. However, when such comparisons are conducted, model predictions do not always match the observed data. The discrepancies can be attributed to various sources of error, such as estimation error of the initial conditions, sampling errors in the field data and errors in the model equations and parameters (Loehle, 1997). The considerable complexity of these models often requires the inclusion of a large number of parameters, many of whose values are uncertain.

Uncertainty in parameter values is attributed to the complexity of natural ecosystems and to the measures by which the parameters are obtained. Parameter values can be obtained from empirical observations or experiments, where the degree of uncertainty around the estimated value can be assessed and even reduced in most cases (Fieberg and Jenkins, 2005). If observations or experiments are not available, parameters can be derived from expert opinion or other models, yet such means are typically characterized by large uncertainty (Ray and Burgman, 2006). Moreover, models have various sensitivities to the different parameters. A parameter that the model is sensitive to is one that minor changes in its value would result in major changes in model output or inference. When high uncertainty in the value of a parameter coincides with high sensitivity of the model to that parameter, the reliability of model predictions may be very low (Bar Massada and Carmel, 2008).

In order to reduce the uncertainty associated with parameter values, considerable effort must typically be invested by the modeler. A prioritized list of influential parameters may be compiled. Such a list can be used to determine the parameters in

* Corresponding author. Tel.: +972 4 9834889; fax: +972 4 6724627.
E-mail address: vardit@oranim.ac.il (V. Makler-Pick).

which the reduction of uncertainty would result in the greatest increase in model accuracy and thus help prescribe resource allocation into further research (Thornton et al., 1979).

Sensitivity analysis (SA) may be used to qualitatively or quantitatively apportion the variation of the model outputs to different sources of variation in model components such as parameters, submodels and forcing data (Brugnach, 2005; Frey et al., 2004; Saltelli et al., 2000, 2008; Helton et al., 2006). Although SA is an optional element within the modeling process (Jorgensen, 1994), several modeling guidelines such as the EPA guidance document (2003) or the European Commission Impact assessment guidelines (2005) prescribe sensitivity analysis as a tool to ensure the modeling quality. SA is therefore considered an important stage in development of ecological models (Ravalico et al., 2005; Saltelli et al., 2000; de Young et al., 2004). In addition, SA can also have ecological importance by identifying the governing parameters and processes in a certain ecological system or even to improve model formulations (Thornton et al., 1979; Cariboni et al., 2007). For example, Cossarini and Solidoro (2008) found that the most relevant parameters in the trophdynamic model of the Gulf of Trieste (Northern Adriatic Sea, Italy) are those related to the growth formulation of the phytoplankton group, the decay rate of particulate organic phosphorus and the mortality rate of bacteria. Cariboni et al. (2007) applied a SA to a pelagic fish population model, revealing that the total order sensitivity index for larvae was ten times more than the total order of sensitivity index estimated for adult fish. These results indicate that from the fishing regulatory point of view the main effort has to be put into developing strategy for protecting young individuals.

Sensitivity analysis of model parameters is carried out by changing them and observing the corresponding response in the output variables. The change in the parameters is chosen on the basis of our knowledge of their acceptable ranges. In local SA, parameter values are changed one at a time, while fixing all other parameter values (Bar Massada and Carmel, 2008). Global SA is a group of techniques that alter a subset or all the parameters simultaneously in a given model simulation (Helton et al., 2006; Helton and Davis, 2003; Fieberg and Jenkins, 2005; Ginot et al., 2006; Chu et al., 2007; Marino et al., 2008). Global SA should probably be preferred in most situations, since (1) it accounts for the effects of interactions between different parameters, and (2) as ecological models are rarely linear, global SA does not assume a linear relationship between the parameters and state variables (Saltelli et al., 2000; Cariboni et al., 2007). Moreover, one may be interested in the relative impact of a group of parameters, a submodel or a process, which local SA is incapable of addressing.

A known shortcoming of global SA is the heavy computational demands (Hamby, 1994; Ascough et al., 2005; Moore and Ray, 1999). These become particularly limiting in models with tens or hundreds of parameters. Such complex models are ubiquitous in ecology, and it is not uncommon to find ecological models with 200 parameters or more. In such models, a single simulation run may last hours, even on powerful computers, and the number of simulations required for a significant global SA may be prohibitively large. SA of such models becomes an intricate and complex task which needs to be well thought out. Furthermore, sensitivity analysis outputs do not always provide the modeler with information on the effect of small changes (e.g. when the parameter is changed within its allowable domain) or how exactly several parameters interact with each other to effect a certain output variable.

Various criteria should therefore be considered when selecting an appropriate SA method (Ravalico et al., 2005; Ascough et al., 2005). The key criteria are: (1) the computational cost associated with an extensive SA (Hamby, 1994; Ascough et al., 2005; Moore

and Ray, 1999), (2) the ability of the method to account for interactions between parameters, (3) the ability of the method to account for non-linearities and non-monotonicity often present in ecological models, (4) the input data required for the analysis, for example in many cases knowledge of parameter probability distributions is required but this knowledge is not always available, and (5) the ability to understand and use the output of the SA.

In this paper, a new global SA approach, applicable to multiparameter models, was developed in order to satisfy the abovementioned criteria. The approach combines several analysis methods. In the first step, two separate and independent analyses methods were performed: (1) based on general linear models (GLM) with random effects and with correction for multiple comparisons (i.e. a least squares method for fitting models that involves continuous and discrete variables); and (2) based on recursive partitioning and regression trees (RPART) which builds classification or regression models of a very general structure using a two stage procedure; the resulting models can be represented as binary trees. The outcomes of these two methods (i.e. the most sensitive parameters selected based on these two methods) were combined to generate a subset of parameters (for each output variable) to which the model was most sensitive. In the second step a more intricate quantitative method, a generalized boosted regression model (GBM, Friedman, 2001, 2002), was applied to the subset of parameters defined in the first stage. The GBM is a general, automated, data-adaptive modeling algorithm that can estimate the non-linear relationship between a variable of interest and a large number of covariates. The impact of the selected parameters on the output variables was estimated and the estimates were used to construct, for each one of the output variables, a final ordered list of parameters with a quantitative measure of the sensitivity of the output variables to the parameters.

The method was applied to a complex hydrodynamic-ecological model, **DY**namic **RE**servoir **S**imulation **M**odel-**C**omputational **A**quatic **E**cosystem **D**ynamics **M**odel, (DYRESM−CAEDYM and DYCD hereafter) used to study Lake Kinneret (Israel). In previously studies of DYCD performance, the seasonal variability and vertical variation in temperature, oxygen, and nutrients were successfully captured (Bruce et al., 2006; Gal et al., 2009), however, these studies also highlighted that much uncertainty exists in predicting nutrient−planktonic interactions that are highly non-linear and are less understood. Therefore the motivation of this analysis was centered on gaining deeper insights into these non-linear interactions. Although inference is not typically mentioned as a specific goal of sensitivity analyses, in this particular application the SA results were also used to derive insights into the model and into the properties of the actual ecosystem of Lake Kinneret.

2. Methods

2.1. DYRESM−CAEDYM (DYCD)

The 1-D hydrodynamic-ecological model, DYCD, developed at the Centre for Water Research, University of Western Australia (Hamilton, 1999; Imberger and Patterson, 1981) simulates the hydrodynamic and biogeochemical dynamics for aquatic ecosystems. DYRESM uses a Lagrangian approach for simulation of the hydrodynamics of aquatic ecosystems (Imberger and Patterson, 1981, 1989). Based on inflows, withdrawals, and meteorological conditions, it calculates the water level and changes to water temperature, salinity and stratification dynamics over time. The water column is represented by set of layers whose thickness ranges between 0.65 m and 2 m. DYRESM has been applied to lakes of varying types (Hamilton, 1999; Horn et al., 2001) including to Lake Kinneret (Gal et al., 2003).

CAEDYM dynamically couples with DYRESM to simulate nutrient cycling and various plankton groups and is the focus of the present SA analysis. CAEDYM consists of a series of partial differential equations to simulate time-varying concentrations of biogeochemical variables accounting for processes such as primary production, secondary production, nutrient cycling, oxygen dynamics and sediment-water interactions (Hipsey and Hamilton, 2008). Various configurations of the model have previously been validated in reservoirs, lakes (Romero et al., 2004;