Ecological Complexity 20 (2014) 219–232

Contents lists available at ScienceDirect

Ecological Complexity

journal homepage: www.elsevier.com/locate/ecocom

Original Research Article

A comprehensive methodology of global sensitivity analysis for complex mechanistic models with an application to plant growth

Qiongli Wu *, Paul-Henry Cournède

Ecole Centrale de Paris, Laboratory MAS, 92290 Châtenay-Malabry Cedex, France

ARTICLE INFO

Article history:
Received 24 May 2013
Received in revised form 5 December 2013
Accepted 21 December 2013
Available online 21 January 2014

Keywords:
Global sensitivity analysis
Mechanistic model
Plant growth

ABSTRACT

A good sensitivity analysis (SA) practice must consider the diverse requirements and limitations of the practice both for the purpose of the analysis in the model design process and from a methodology perspective. Complex mechanistic models are often organized with several modules dynamically describing the diverse multi-physical processes. They are also characterized by a significant number of control factors. The strong interactions between the factors or the modules are crucial for understanding the model complexity. A comprehensive methodology must be devised to meet not only the classical objective of parameter screening for parameter estimation but also the objective of performing model diagnosis by qualitatively and quantitatively checking the module importance and interactions. In this paper, we proposed a comprehensive SA methodology adapted to complex mechanistic models characterized by several interacting processes with modules describing each of them. In this methodology, we successively perform the analysis of model nonlinearity, module importance ranking and its evolution with time, module-by-module parameter screening, quantitative analysis of both intra- and inter-module interactions, and the analysis of the complete model with a reduced number of parameters due to parameter screening. The numerical implementation strategy and computational cost analysis are also presented. A case study is presented on the Nitrogen Economy Model within plant Architecture (NEMA), which is a typical model organized into modules describing the multi-biophysical processes of plant growth. The results demonstrate that our methodology can help to reveal the importance evolution and interactions between biophysical processes described by the model modules. The reduction in the number of influential parameters to estimate from 83 to 17 by SA is also a significant step forward for the NEMA model parameterization improvement process.

1. Introduction

Mathematical models are increasingly used to simulate and predict the behavior of physical or biological systems, partly due to the ability of models to incorporate high levels of complexity that are characteristic of these systems (Song et al., 2012). In ecological modelling, the complexity results mainly from the different scales involved in the described physical or biological processes and their interactions. Studying these different interactions and determining how the interactions work together pose a significant challenge. Ecologists refer to this phenomenon as ecological complexity.

Ecological models are often characterized by many modules describing the corresponding biophysical processes and by many parameters serving as control factors. Mechanistic models (process driven) are more powerful than empirical models (data driven) for describing this type of system because they tell us about the underlying processes that drive patterns. In mechanistic models, the mathematical functions connecting inputs and outputs can be attached to the physical meanings of processes.

The different steps in the development of mechanistic models can be summarized as follows. (1) Conceptual work for the definition of the model objectives and the scientific hypotheses for the description of all processes involved: this step also involves defining the model structure and the mathematical formalisms to write the process equations. (2) Mathematical and numerical analysis for conditional verification: the main aim of this step is to study the general and limit model behavior, identifiability, and model stability. (3) Experimentation, collection and analysis of data. Specific protocols need to be designed about how unknown parameters will be measured in experiments or be estimated from the collected data. (4) Model identification: model structure identification, comparing the model to experimental data, and parametric identification. (5) Model validation: this step is performed to qualitatively and quantitatively check whether the resulting model achieves the modelling objectives.

* Corresponding author. Tel.: +33 612991555.
E-mail addresses: qiongli.wu@gmail.com, wqlgirl@163.com (Q. Wu).

1476-945X/$ – see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ecocom.2013.12.005

Several complications arise in the design process of a mechanistic model. First, it is generally impossible to identify a single model structure for a physiological system because the system is never closed and more than one model will be possible (Aronica et al., 1998). Moreover, uncertainty can be related to the interactions between the multiphysical processes described by the model modules (De Rocquigny et al., 2008). If we consider plant growth models, for example, these models are developed to simulate biomass production and distribution among organs, in interaction with the environment (De Reffye et al., 2008). The complicated and interacting physiological processes governing plant growth bring a large amount of uncertainty into their modelling. Moreover, field surveys for collecting the necessary data for the development and parameterization of models are difficult and expensive. Even though high-throughput phenotyping techniques have strongly developed in the recent years, these techniques remain expensive and are difficult to reproduce in diverse environmental conditions (Tsaftaris and Noutsos, 2009; Walter et al., 2012). Moreover, the data analysis to extract the necessary data for model parameterization is far from straightforward.

As a consequence, input data (environmental factors) and experimental data from which model parameters are estimated are characterized by significant uncertainties.

How do the uncertainties of model structure, parameters, and multiphysical process modules and the interactions between them contribute to the total model uncertainty? It is important to answer this question both qualitatively and quantitatively for the application and development of complex mechanistic models (Harper et al., 2011). In our case, the types of models that we focus on have many parameters with interactions between them and may have submodels (modules) describing the multiphysical processes; hence, there are interactions between these modules. Sensitivity analysis (SA) is the study of how the variation (uncertainty) in the output of a mathematical model can be apportioned, qualitatively or quantitatively, to different sources of variation in the input factors of the model (Saltelli et al., 2008). Global SA methods can help to identify influential model parameters, processes, and the interactions between them, which constitute the major portion of the uncertainties in model outputs. It is expected that the analysis result can facilitate the design of such complex models in two ways. First, quite classically, in the parameterization process, it helps to simplify parameter estimation by screening the noninfluential parameters for the model outputs corresponding to the experimental data available for parameter estimation. The complexity of parameter estimation can be reduced by removing the noninfluential parameters. Because the interactions between parameters can deeply affect the strategy of parameter estimation, we are also obliged to evaluate the interactions between parameters (Varella et al., 2010). Variance decomposition-based SA methods, such as Sobol's method, can be used to study such interactions by the high-order sensitivity indices (Wu et al., 2012). Second, SA helps by providing new biological insights and diagnosis based on the importance of processes and their interactions in the different stages of growth. From a methodology perspective, there are different SA methods that have different objectives and are adapted to different types of models. First, the aims of SA must be considered, namely, factor prioritization, factor fixing, variance cutting, or factor mapping (Saltelli et al., 2008). Each aim involves different methods. Moreover, the computing cost issue related to the complexity of the models must also be seriously considered. For instance, the standardized regression coefficients (SRCs) can be viewed as an interesting tradeoff between the accuracy of the analysis and the computing cost, but they are only validated when the model's linearity is high. Sobol's method can help us to obtain the interaction information for parameters and functional modules, but in some cases, the sampling-based Monte Carlo simulation makes Sobol's method prohibitive from a computational perspective. In particular, when we perform the module analysis, the interactions between modules involve diverse combinations of interactions between factors within the modules, as revealed by a number of higher-order Sobol's indices. In this case, it is not practical to compute all of the higher-order indices individually. However, it is possible to obtain the portion of uncertainty contributed by the intra- and inter-module interactions using the decomposition of total-order indices (Chen et al., 2004) for group factors associated with the modules of the model without additional computing cost. A methodology must be devised to make full use of such characteristics of Sobol's method.

To summarize, both in the model design procedure (parameterization and model diagnosis) and from a methodology perspective, a good SA practice must consider the diverse requirements and limitations of the practice. It is increasingly common for comprehensive SAs to involve several complementary methods. Flexibility in the use of different SA techniques is crucial (Sun et al., 2012). Several works have been performed on the complementary strategy of SA techniques (Sun et al., 2012; Makler-Pick et al., 2011), and most of these studies are based on machine learning algorithms. In the case of complex mechanistic models, with many modules describing diverse multiphysical processes, a significant number of control factors, and interactions between factors or modules, it is necessary to emphasize both the 'factor screening' for parameter estimation and the model diagnosis by establishing the module importance and interactions while avoiding prohibitive computing costs. For this objective, variance-based methods, particularly Sobol's method, have the advantage of providing a detailed quantitative analysis of the interactions compared with machine learning algorithms. Thus, a comprehensive methodology must be devised to fulfill the objectives by fully utilizing such complementary SA methods.

In this paper, we propose a comprehensive SA methodology adapted to complex mechanistic models characterized by several interacting processes with modules describing each of them. This procedure combines both SRC and Sobol's method so that both the advantages of the SRC method's computing efficiency and Sobol's quantitative analysis can be used. We also discuss the numerical implementation strategy and the computational cost issue of the methodology. A case study is presented on the Nitrogen Economy Model within plant Architecture (NEMA) (Bertheloot et al., 2011a), which is a typical model with several distinct biological function modules with interactions describing carbon (C) and nitrogen (N) acquisition by a wheat plant and C and N distributions between plant organs after flowering. This model has the specificity to integrate physiological processes governing the N economy within the plant architecture.

We first recall the principles of the classical SA methods in this study and detail their extensions specifically deduced for the analysis of complex models in Section 2. Then, the steps of our methodology are given in Section 3. Numerical implementation and the computing cost issue are discussed in Section 4. The case study of NEMA is illustrated in Section 5, followed by a discussion in Section 6.

2. Background methods

2.1. Standardized regression coefficient (SRC) method and nonlinearity assessment

The SRC method (Cariboni et al., 2007) is based on the linear approximation of the model and Monte Carlo simulations. One