Reliability Engineering and System Safety 107 (2012) 35–43

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

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

An efficient computational method for global sensitivity analysis and its application to tree growth modelling

Qiong-Li Wu [a,b,c,*], Paul-Henry Cournède [a,b], Amélie Mathieu [d]

[a] INRIA Saclay Île-de-France, EPI Digiplante, 91893 Orsay Cedex, France
[b] Ecole Centrale de Paris, Laboratory MAS, 92290 Châtenay-Malabry Cedex, France
[c] School of Electronic and Information Engineering, Wuhan University, Wuhan, China
[d] UMR INRA/AgroParisTech Environnement et Grandes Cultures, 78850 Thiverval-Grignon, France

ARTICLE INFO

Article history:
Received 3 November 2010
Received in revised form
29 June 2011
Accepted 6 July 2011
Available online 19 July 2011

Keywords:
Sensitivity analysis
Sobol indices
Error estimation
Computing efficiency
Functional structural plant models
GreenLab

ABSTRACT

Global sensitivity analysis has a key role to play in the design and parameterisation of functional–structural plant growth models which combine the description of plant structural development (organogenesis and geometry) and functional growth (biomass accumulation and allocation). We are particularly interested in this study in Sobol's method which decomposes the variance of the output of interest into terms due to individual parameters but also to interactions between parameters. Such information is crucial for systems with potentially high levels of non-linearity and interactions between processes, like plant growth. However, the computation of Sobol's indices relies on Monte Carlo sampling and re-sampling, whose costs can be very high, especially when model evaluation is also expensive, as for tree models. In this paper, we thus propose a new method to compute Sobol's indices inspired by Homma–Saltelli, which improves slightly their use of model evaluations, and then derive for this generic type of computational methods an estimator of the error estimation of sensitivity indices with respect to the sampling size. It allows the detailed control of the balance between accuracy and computing time. Numerical tests on a simple non-linear model are convincing and the method is finally applied to a functional–structural model of tree growth, GreenLab, whose particularity is the strong level of interaction between plant functioning and organogenesis.

1. Introduction

Sensitivity analysis (SA) is a fundamental tool in the building, use and understanding of mathematical models [1]. Sampling-based approaches to uncertainty and sensitivity analysis are both effective and widely used [2]. For this purpose, Sobol's method is a key one [3]. Since it is based on variance decomposition, the different types of sensitivity indices that it estimates can fulfill different objectives of sensitivity analysis: factor prioritisation, factor fixing, variance cutting or factor mapping [4]. It is a very informative method but potentially computationally expensive [2]. Besides the first-order effects, Sobol's method also aims at determining the levels of interaction between parameters [5]. In [6], the authors also devised a strategy for sensitivity analysis that could work for correlated input factors, based on the first-order and total-order index from variance decomposition.

Such type of global sensitivity analysis method has a key role to play in functional–structural plant growth modelling. In this recent field of research in plant biology [7] models are not yet stable. They aim at combining the description of both plant structural development and eco-physiological functioning at a very detailed scale, typically that of organs (leaves, internodes, *etc.*). The complexity of the underlying biological processes, especially the interaction between function and structure [8] usually makes parameterisation a complex step in modelling, and the analysis of model sensitivity to parameters and of their levels of interactions provides useful information in this process [9].

Computational methods to evaluate Sobol indices sensitivity rely on Monte Carlo sampling and re-sampling [3,10]. For k-dimensional factor of model uncertainty, the k first-order effects and the 'k' total-order effects are rather expensive to estimate, needing a number of model evaluations strictly depending on k [11]. Especially, for individual-based tree growth models, at organ level, the cost of model evaluation can be very heavy [7]. Therefore, it is crucial to not only devise efficient computing techniques, in order to make best use of model evaluations [12], but also to have a good control of the estimation accuracy with respect to the number of samples.

* Corresponding author at: INRIA Saclay Île-de-France, EPI Digiplante, 91893 Orsay Cedex, France.
E-mail address: qiongli.wu@gmail.com (Q.-L. Wu).

0951-8320/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ress.2011.07.001

The objective of this paper is to study these two aspects. First, we propose a computing method inspired by [10], which improves slightly their use of model evaluations, and then derive an estimator of the error of sensitivity indices evaluation with respect to the sampling size for this generic type of computational methods. Numerical tests are then shown to illustrate the results. Finally, the method is applied to a functional–structural model of tree growth, whose particularity is the strong level of interaction between plant functioning and organogenesis [13].

2. Computing sensitivity analysis

2.1. General concepts of Sobol's sensitivity analysis

We recall here the basic concepts of Sobol's method [3] to present the original work about sensitivity analysis indices. The function $f(\mathbf{X}) \equiv f(X_1,X_2,\ldots,X_k)$ under investigation is defined in the k-dimensional cube $\mathbf{K}^k$. If the input factors are mutually independent then there exists a unique decomposition of $f(\mathbf{X})$:

$$f(X_1,\ldots,X_k) = f_0 + \sum_{i=1}^{k} f_i(X_i) + \sum_{1\le i<l\le k} f_{il}(X_i,X_l) + \cdots + f_{1,2,\ldots,k}(X_1,\ldots,X_k) \tag{1}$$

The basic idea of Sobol's method (see [3]) is to decompose the function of interest into terms of increasing dimensionality as in Eq. (1), such that all the summands are mutually orthogonal. The variance of the output variable Y can thus be decomposed into

$$V = \sum_{i=1}^{k} V_i + \sum_{1\le i<l\le k} V_{il} + \cdots + V_{1,2,\ldots,k} \tag{2}$$

where V_i, V_{il}, $V_{1,2,\ldots,k}$ denote the variance of f_i, f_{il}, $f_{1,2,\ldots,k}$, respectively. In this approach the first-order sensitivity index for factor X_i is given by

$$S_i = \frac{V(E(Y|X_i))}{V(Y)} = \frac{V_i(E_{-i}(Y|X_i))}{V} \tag{3}$$

where E and V indicate, respectively, the mean and variance operators, $-i$ indicates all factors but i. The inner expectation is taken at a generic point in the space of variable X_i, while the outer variance is over all possible values of this generic point.

The higher order sensitivity indexes $S_{i_1,\ldots,i_s}$ are given by

$$S_{i_1,\ldots,i_s} = \frac{V_{i_1,\ldots,i_s}}{V} \tag{4}$$

for $s>1$, Eq. (2) can be rewritten in terms of sensitivity indexes as

$$1 = \sum_{i=1}^{k} S_i + \sum_{1\le i<l\le k} S_{il} + \cdots + S_{1,2,\ldots,k} \tag{5}$$

The total-order effect ST_i is instead given by

$$ST_i = \frac{E_{-i}(V_i(Y|X_{-i}))}{V} \tag{6}$$

Note that this time the inner variance is over all possible generic values of X_i while the outer mean is over the space X_{-i}. If $ST_i=0$, then X_i is non-influent, so the index ST_i is suitable for fixing non-influential factors. Standard Sobol's method was proposed in [3]. In [10], an improved estimator is presented to compensate the system error, completed in [12], and a computationally efficient design is discussed. This method will therefore be called Homma–Saltelli method. We then propose an improvement of this method to promote its convergence characteristics.

2.2. Sobol's computing method and Homma–Saltelli (H–S) improvement

The standard Sobol's method for SA was put forward in [3], as one numerical simulation method to get the conditional expectation value for model output Y. We first decide the base sampling dimension N, then we implement the following steps:

1. Generate a Monte Carlo sampling of dimension N of the input factors according to their random distributions and form the $N\times k$ matrix $U_{N\times k}$ (k being the dimension of the input space) with each row a set of parameters; $U_{N\times k}$ is called the 'sampling matrix'

$$U_{N\times k} = \begin{bmatrix} x_{1^{(1)}} & \cdots & x_{i^{(1)}} & \cdots & x_{k^{(1)}} \\ x_{1^{(2)}} & \cdots & x_{i^{(2)}} & \cdots & x_{k^{(2)}} \\ \vdots & & & \ddots & \vdots \\ x_{1^{(N)}} & \cdots & x_{i^{(N)}} & \cdots & x_{k^{(N)}} \end{bmatrix}$$

2. Generate another sampling matrix of dimension $N\times k$, $W_{N\times k}$, called the 're-sampling matrix'

$$W_{N\times k} = \begin{bmatrix} x_{1^{(N+1)}} & \cdots & x_{i^{(N+1)}} & \cdots & x_{k^{(N+1)}} \\ x_{1^{(N+2)}} & \cdots & x_{i^{(N+2)}} & \cdots & x_{k^{(N+2)}} \\ \vdots & & & \ddots & \vdots \\ x_{1^{(2N)}} & \cdots & x_{i^{(2N)}} & \cdots & x_{2k^{(N)}} \end{bmatrix}$$

3. Define a matrix $W'_{N\times k}$ formed by all columns of $W_{N\times k}$, except the ith column obtained from the ith column of $U_{N\times k}$

$$W'_{N\times k} = \begin{bmatrix} x_{1^{(N+1)}} & \cdots & x_{i^{(1)}} & \cdots & x_{k^{(N+1)}} \\ x_{1^{(N+2)}} & \cdots & x_{i^{(2)}} & \cdots & x_{k^{(N+2)}} \\ \vdots & & & \ddots & \vdots \\ x_{1^{(2N)}} & \cdots & x_{i^{(N)}} & \cdots & x_{2k^{(N)}} \end{bmatrix}$$

4. Compute the model output for each set of input parameters from $U_{N\times k}$ and $W'_{N\times k}$ (that is to say for each row in $U_{N\times k}$ and $W'_{N\times k}$) to obtain two column vectors of model outputs of dimension N: $\mathbf{y} = f(U_{N\times k})$, $\mathbf{y}'_R = f(W'_{N\times k})$.
5. The sensitivity indices are hence computed based on scalar products of the above-defined vectors of model outputs.

The applicability of the sensitivity estimates S_i to a large class of functions $f(\mathbf{X})$ is linked to the possibility of evaluating the multi-dimensional integral associated with these estimates via Monte Carlo methods. For a given sampling size N tending to ∞ the following estimate for the mean value of the output is straightforward:

$$\widehat{f}_0 = \frac{1}{N}\sum_{j=1}^{N} \mathbf{y}^{(j)} \tag{7}$$

where $\mathbf{y}^{(j)}$ is the model output for a sample point in the parameter space $\mathbf{K}^k$. The hat symbol will be used to denote estimates.

To list the estimator for the standard Sobol's in [3], the following notation will be introduced:

$$\overline{V} = \frac{1}{N}\sum_{j=1}^{N} \mathbf{y}^{(j)^2} \tag{8}$$

$$\overline{V}_i = \frac{1}{N}\sum_{j=1}^{N} \mathbf{y}^{(j)}\mathbf{y}'^{(j)}_R \tag{9}$$