Reliability Engineering and System Safety 94 (2009) 1194–1204

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

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

Global sensitivity analysis of computer models with functional inputs

Bertrand Iooss [a,*], Mathieu Ribatet [b]

[a] CEA, DEN, DER/SESI/LCFR, F-13108 Saint-Paul-lez-Durance, France
[b] École Polytechnique Fédérale de Lausanne, Chair of Statistics, STAT-IMA-FSB-EPFL, Station 8, CH-1015 Lausanne, Switzerland

ARTICLE INFO

Article history:
Received 26 October 2007
Received in revised form
4 June 2008
Accepted 16 September 2008
Available online 19 October 2008

Keywords:
Sobol's indices
Joint modeling
Generalized additive model
Metamodel
Stochastic process
Uncertainty

ABSTRACT

Global sensitivity analysis is used to quantify the influence of uncertain model inputs on the response variability of a numerical model. The common quantitative methods are appropriate with computer codes having scalar model inputs. This paper aims at illustrating different variance-based sensitivity analysis techniques, based on the so-called Sobol's indices, when some model inputs are functional, such as stochastic processes or random spatial fields. In this work, we focus on large cpu time computer codes which need a preliminary metamodeling step before performing the sensitivity analysis. We propose the use of the joint modeling approach, i.e., modeling simultaneously the mean and the dispersion of the code outputs using two interlinked generalized linear models (GLMs) or generalized additive models (GAMs). The "mean model" allows to estimate the sensitivity indices of each scalar model inputs, while the "dispersion model" allows to derive the total sensitivity index of the functional model inputs. The proposed approach is compared to some classical sensitivity analysis methodologies on an analytical function. Lastly, the new methodology is applied to an industrial computer code that simulates the nuclear fuel irradiation.

1. Introduction

Modern computer codes that simulate physical phenomenas often take as inputs a high number of numerical parameters and physical variables, and return several outputs–scalars or functions. For the development and the use of such computer models, sensitivity analysis (SA) is an invaluable tool. The original technique, based on the derivative computations of the model outputs with respect to the model inputs, suffers from strong limitations for computer models simulating non-linear phenomena. More recent global SA techniques take into account the entire range of variation of the model inputs and aim to apportion the whole output uncertainty to the model input uncertainties [1]. The global SA methods can also be used for model calibration, model validation, decision making process, i.e., any process where it is useful to know which are the variables that mostly contribute to the output variability.

The common quantitative methods are applicable to computer codes with scalar model inputs. For example, in the nuclear engineering domain, global SA tools have been applied to numerous models where all the uncertain model inputs are modeled by random variables, possibly correlated—such as thermal-hydraulic system codes [2], waste storage safety studies [3], environmental model of dose calculations [4], reactor dosimetry processes [5]. Recent research papers have tried to consider more complex model inputs in the global SA process, especially in petroleum and environmental studies:

- Tarantola et al. [6] work on an environmental assessment on soil models that use spatially distributed maps affected by random errors. This kind of uncertainty is modeled by a spatial random field (following a specified probability distribution), simulated at each code run. For the SA, the authors propose to replace the spatial model input by a "trigger" random parameter ξ that governs the random field simulation. For some values of ξ, the random field is simulated and for the other values, the random field values are put to zero. Therefore, the sensitivity index of ξ is used to quantify the influence of the spatial model input.
- Ruffo et al. [7] evaluate an oil reservoir production using a model that depends on different heterogeneous geological media scenarios. These scenarios, which are of limited number, are then identified by a discrete trigger spanning randomly the set of scenarios. Therefore, the sensitivity index of the trigger is used to identify the influence of the alternative scenarios.
- Iooss et al. [8] study a groundwater radionuclide migration model which depend on several random scalar parameters and on a spatial random field (a geostatistical simulation of the hydrogeological layer heterogeneity). The authors propose to consider the spatial model input as an "uncontrollable"

* Corresponding author. Tel.: +33 4 42 25 72 73; fax: +33 4 42 25 24 08.
E-mail address: bertrand.iooss@cea.fr (B. Iooss).

0951-8320/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ress.2008.09.010

parameter. Therefore, they fit on a few simulation results of the computer model a double model, called a joint model: the first component models the effects of the scalar parameters while the second models the effects of the "uncontrollable" parameter.

In this paper, we tackle the problem of the global SA for numerical models and when some model inputs ε are functional. $\varepsilon(\boldsymbol{u})$ is a one or multi-dimensional stochastic function where $\boldsymbol{u}$ can be spatial coordinates, time scale or any other physical parameters. Our work focuses on models that depend on scalar parameter vector $\boldsymbol{X}$ and involve some stochastic process simulations or random fields $\varepsilon(\boldsymbol{u})$ as inputs. The computer code output Y depends on the realizations of these random functions. These models are typically non-linear with strong interactions between model inputs. Therefore, we concentrate our methodology on the variance based sensitivity indices estimation; that is, the so-called Sobol's indices [1,9].

To deal with this situation, a first natural approach consists in using either all the discretized values of the functional model input $\varepsilon(\boldsymbol{u})$ or its decomposition into an appropriate basis of functions. Then, for all the new scalar variables related to $\varepsilon(\boldsymbol{u})$, sensitivity indices are computed. However, in the case of complex functional model inputs, this approach seems to be rapidly intractable as these variables cannot be represented by a small number of scalar variables [6]. Moreover, when dealing with non-physical variables (for example, coefficients of orthogonal functions used in the decomposition), sensitivity indices interpretation may be laborious. Indeed, most often, physicists would prefer to obtain one global sensitivity index related to $\varepsilon(\boldsymbol{u})$. Finally, a major drawback for the decomposition approach is related to the uncertainty modeling stage. More precisely, this approach needs to specify the probability density functions for the coefficients of the decomposition.

The following section presents three different strategies to compute Sobol's indices with functional model inputs: (a) the macroparameter method, (b) the "trigger" parameter method, and (c) the proposed joint modeling approach. Section 3 compares the relevance of these three strategies on an analytical example: the WN-Ishigami function. Lastly, the proposed approach is illustrated on an industrial computer code simulating fuel irradiation in a nuclear reactor.

2. Computational methods of Sobol's indices

First, let us recall some basic notions about Sobol's indices. Let define the model

$$\begin{aligned} &f : \mathbb{R}^p \to \mathbb{R} \\ &\boldsymbol{X} \mapsto Y = f(\boldsymbol{X}) \end{aligned} \tag{1}$$

where Y is the code output, $\boldsymbol{X} = (X_1, \ldots, X_p)$ are p independent inputs, and f is the model function. f is considered as a "black box", i.e., a function whose analytical formulation is unknown. The main idea of the variance-based SA methods is to evaluate how the variance of a model input or a group of model inputs contributes to the output variance of f. These contributions are described using the following sensitivity indices:

$$S_i = \frac{\mathrm{Var}[\mathbb{E}(Y|X_i)]}{\mathrm{Var}(Y)}, \quad S_{ij} = \frac{\mathrm{Var}[\mathbb{E}(Y|X_iX_j)]}{\mathrm{Var}(Y)} - S_i - S_j, \quad S_{ijk} = \cdots \tag{2}$$

These indices, namely Sobol's indices, can be used for any complex model functions f. The second order index S_{ij} expresses the model sensitivity to the interaction between the variables X_i and X_j (without the first order effects of X_i and X_j), and so on for higher orders effects. The interpretation of these indices is natural as all indices lie in $[0, 1]$ and their sum is equal to 1. The larger an index value is, the greater is the importance of the variable or the group of variables related to this index.

For a model with p inputs, the number of Sobol's indices is $2^p - 1$; leading to an intractable number of indices as p increases. Thus, to express the overall output sensitivity to an input X_i, Homma and Saltelli [10] introduce the total sensitivity index:

$$S_{T_i} = S_i + \sum_{i<j} S_{ij} + \sum_{i<j<k} S_{ijk} + \cdots = \sum_{l \in \#i} S_l \tag{3}$$

where $\#i$ represents all the "non-ordered" subsets of indices containing index i. Thus, $\sum_{l\in\#i} S_l$ is the sum of all the sensitivity indices having i in their index. The estimation of these indices (Eqs. (2) and (3)) can be performed by simple Monte Carlo simulations based on independent samples [9,11], or by refined sampling designs introduced to reduce the number of required model evaluations significantly, for instance FAST [12] and quasi-random designs [13].

Let us now consider a supplementary model input which is a functional input $\varepsilon(\boldsymbol{u}) \in \mathbb{R}$ where $\boldsymbol{u} \in \mathbb{R}^d$ is a d-dimensional location vector. $\varepsilon(\boldsymbol{u})$ is defined by all its marginal and joint probability distributions. In this work, it is supposed that random function realizations can be simulated. For example, these realizations can be produced using geostatistical simulations [14] or stochastic processes simulations [15]. Our model writes now

$$Y = f(\boldsymbol{X}, \varepsilon) \tag{4}$$

and in addition to Sobol's indices related to $\boldsymbol{X}$, our goal is to derive methods to compute the sensitivity indices relative to ε, i.e., S_ε (first order index), S_{T_ε} (total sensitivity index), $S_{i\varepsilon}$ (second order indices), $S_{ij\varepsilon}, \ldots$.

2.1. The macroparameter method

With the macroparameter method, the functional input is not seen as a functional one by the computer code. It is discretized in a potentially large number (say n_{dis}) of values (for example, several thousands), each of them being a scalar model input of the computer code. As all these values come from the functional input (which possesses a specific correlation structure), they can be considered as an ensemble of correlated inputs. Taking into account correlation between model inputs in SA has been a challenging problem, recently solved by a few authors (see [16] for a recent review).

One solution, proposed by Jacques et al. [5], to deal with correlated model inputs, is to consider multi-dimensional sensitivity indices [17]: each group of correlated inputs is considered as a multi-dimensional input or macroparameter. One therefore performs an SA by groups of correlated inputs. To estimate Sobol's indices (first order, second order,..., total), a large number of model inputs (correlated and non-correlated) have to be generated. As we know how to generate independent samples of a correlated variables group, the simple Monte Carlo sampling technique can be used (Sobol' [9] algorithm or Saltelli' [11] algorithm). However, more efficient techniques than simple Monte Carlo (in terms of the required size sample), as FAST or quasi Monte Carlo which use deterministic samples, are prohibited with correlated inputs.

In our context, this approach, using the simple Monte Carlo algorithm, seems to be relevant as the functional model input $\varepsilon(\boldsymbol{u})$ can be considered as a single multi-dimensional input (i.e., a macroparameter). For instance, the first order Sobol's index related to $\varepsilon(\boldsymbol{u})$ is defined as previously by

$$S_\varepsilon = \frac{\mathrm{Var}[\mathbb{E}(Y|\varepsilon)]}{\mathrm{Var}(Y)} \tag{5}$$