International Journal of Approximate Reasoning ••• (••••) •••–•••

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

A probabilistic procedure for quantifying the relative importance of model inputs characterized by second-order probability models

Pengfei Wei [a,*], Fuchao Liu [a], Zhenzhou Lu [b], Zuotao Wang [a]

[a] School of Mechanics, Civil Engineering and Architecture, Northwestern Polytechnical University, 710072 Xi'an, People's Republic of China
[b] School of Aeronautics, Northwestern Polytechnical University, 710072 Xi'an, People's Republic of China

ARTICLE INFO

Article history:
Received 9 November 2017
Received in revised form 20 April 2018
Accepted 20 April 2018
Available online xxxx

Keywords:
Sobol' indices
Relative importance
Second-order probability model
Extended Monte Carlo simulation
Kriging

ABSTRACT

This paper develops a new global sensitivity analysis (GSA) framework for computational models with input variables being characterized by second-order probability models due to epistemic uncertainties. Firstly, two graphical tools, called individual effect (IE) function and total effect (TE) function, are defined for identifying the influential and non-influential input variables. Secondly, two probabilistic GSA indices, called T-indices, are introduced for comparing the relative importance of pairwise influential input variables. Thirdly, the expected Sobol' indices are introduced for ranking the importance of the input variables. For efficiently estimating the proposed GSA indices, the extended Monte Carlo simulation (EMCS), whose computational cost is the same as the Monte Carlo simulation for estimating the Sobol' indices, is firstly introduced, and then a procedure combining Kriging surrogate model and EMCS procedure is introduced for further reducing the computational cost. Three numerical examples and a ten-bar structure are introduced for illustrating the significance of the proposed GSA framework and demonstrating the effectiveness of the computational methods.

1. Introduction

Global sensitivity analysis (GSA) aims at quantifying the different types of effects of uncertain model input variables on the uncertainty of model output, and measuring the relative importance of each model input. It has been widely used in many fields involving computational models such as computational physics and chemistry [1,2], environmental science [3] and reliability engineering [4,5]. Till now, many GSA techniques have been developed by researchers [6,7]. The screening method was developed by Morris [8] for identifying a small number of influential input variables for a high-dimensional computational model, with acceptable computational cost. The variance-based GSA was developed by Sobol' [9] as well as Homma and Saltelli [10] for identifying the individual and interaction contributions of each input variable on the model output variance. With the consideration that variance is not affordable for characterizing uncertainty, the moment-independent GSA was developed by Borgonovo [11,12] for estimating the effect of each uncertain input variable on the probability density function (PDF) of model output, and it was later proved to be monotonic transformation invariant [13] and suitable for measuring the strength of dependence between model input and output variables [14]. Among all the available GSA techniques,

* Corresponding author.
E-mail address: pengfeiwei@nwpu.edu.cn (P. Wei).

https://doi.org/10.1016/j.ijar.2018.04.007
0888-613X/

the variance-based one has drawn the most attentions, and many computational methods, such as Monte Carlo Simulation (MCS) [15,16], Fourier Amplitude Sensitivity Test (FAST) [17,18] and surrogate models [19–22], have been developed for estimating the variance-based GSA indices (also called Sobol' indices).

All the above-mentioned GSA techniques are based on the assumption that the PDFs of the model input variables are precisely determined. However, this is only applicable when the size of data used for identifying the PDFs of these variables is sufficiently large and the available data is accurate enough. When the input variables are deterministic, but there is uncertainty about their true values, they can be characterized by subjective probability distributions, and all the above-mentioned methods can be used to this kind of problems, as shown in Ref. [23]. However, if there is aleatory uncertainty about the input variables, and due to the epistemic uncertainty resulting from incompleteness and/or inaccuracy of the available data, the PDFs of the input variables cannot be precisely determined [24,25]. In this situation, the distribution parameters can be characterized either by confidence intervals or by random numbers instead of constant values. When the Bayesian updating approach is applied, the distribution parameters are characterized as random variables, and it is called second-order probability model [26]. In this situation, the possible PDF of each input variable degrades into its true PDF when the available data is sufficiently big and accurate.

When the input variables are characterized by the second-order probability models, three strategies have been developed for implementing the GSA analysis. In the first strategy, the sensitivity indices are defined as the partial derivatives of probabilistic responses (e.g., model output variance and expectation) w.r.t. the distribution parameters, and they reflect the changes of the probabilistic responses due to small changes of distribution parameters. These local sensitivity indices can be efficiently computed by the score function method [27]. In the second strategy, the GSA indices are defined by decomposing the variance of some probabilistic responses into partial variances of increasing orders governed by the random distribution parameters instead of the random input variables, so as to estimate the individual, interaction and total effects of each random distribution parameter on the probabilistic responses in a global sense [28,29]. The main disadvantage of the former two strategies is that the sensitivity indices measure the relative importance of the distribution parameters instead of the random input variables, thus cannot be used for ranking the random input variables. In the third strategy, the GSA indices are defined simultaneously for the random input variables as well as their distribution parameters by introducing the probability integral transformation (PIT) to each of the random input variables [30,31]. The main disadvantage of this strategy is that, when the data size is sufficiently large and the uncertainties of the distribution parameters disappear, the GSA indices will not degrade into the classical Sobol' indices, due to the fact that Sobol' indices are not monotonic transformation invariant [13].

In this paper, we propose a new framework for quantifying the relative importance of each input variable when they are characterized by the second-order probability models. Firstly, we propose two graphical tools for identifying the influential and non-influential input variables, and then we propose two probabilistic GSA indices, based on Sobol' indices, to compare the relative importance of each pair of input variables. At last, the expected Sobol' indices are introduced for ranking the input variables. For estimating the proposed GSA indices, a numerical method called extended Monte Carlo simulation (EMCS) is introduced, whose computational cost is the same with the classical MCS procedure for estimating the Sobol' indices. Then, for further reducing the computational burden, the Kriging surrogate model is introduced by combining with the EMCS procedure.

The rest of this paper is organized as follows. Section 2 reviews the second-order probability model. Second 3 recalls the classical Sobol' indices. Section 4 proposes the new GSA framework. Section 5 develops EMCS and Kriging-EMCS methods for estimating the proposed GSA indices. Section 6 introduces three numerical examples to illustrate the proposed method. Section 7 applies the proposed methods to a planar ten-bar structure. Section 8 gives conclusions.

2. Problem statement

For a random input variable x with PDF $p(x;\boldsymbol{\theta})$ and a set of scarce data $x^{(j)}$ ($j = 1, 2, \ldots, N$), many methods in statistics are available for estimating the distribution parameters $\boldsymbol{\theta}$. The commonly used ones include interval estimation approach and Bayesian updating approach. When the Bayesian updating approach is applied, the likelihood function should be firstly derived, i.e.,

$$L(\boldsymbol{\theta}) \propto \prod_{j=1}^{N} p\left(x^{(j)};\boldsymbol{\theta}\right) \tag{1}$$

If the available data are inaccurate and characterized by intervals, one can refer to Ref. [26] for a more general form of the likelihood function. The maximum likelihood estimates (MLEs) of $\boldsymbol{\theta}$ can then be obtained by maximizing the likelihood function in Eq. (1). However, the MLE cannot reflect the epistemic uncertainty of x resulting from incompleteness of data. The epistemic uncertainty of x can only be reflected by the uncertainty presented in $\boldsymbol{\theta}$. By choosing a prior PDF $f'_{\Theta}(\boldsymbol{\theta})$, following the Bayes' formula, the posteriori PDF $f_{\Theta}(\boldsymbol{\theta})$ can be derived as:

$$f_{\Theta}(\boldsymbol{\theta}) = \frac{L(\boldsymbol{\theta}) f'_{\Theta}(\boldsymbol{\theta})}{\int L(\boldsymbol{\theta}) f'_{\Theta}(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta}} \tag{1}$$