Reliability Engineering and System Safety 162 (2017) 28–39

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

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

Extending Morris method for qualitative global sensitivity analysis of models with dependent inputs

Qiao Ge*, Monica Menendez

Institute for Transport Planning and Systems, ETH Zurich, HIL F 37.3, Stefano-Franscini-Platz 5, 8093 Zurich, Switzerland

ARTICLE INFO

Keywords:
Dependent inputs
Screening
Global sensitivity analysis
Dependent contributions
Independent contributions
Variance-based sensitivity indexes

ABSTRACT

Global Sensitivity Analysis (GSA) can help modelers to better understand the model and manage the uncertainty. However, when the model itself is rather sophisticated, especially when dependence exists among model inputs, it could be difficult or even unfeasible to perform quantitative GSA directly. In this paper, a non-parametric approach is proposed for screening model inputs. It extends the classic Elementary Effects (i.e., Morris) method, which is widely used for screening independent inputs, to enable the screening of dependent model inputs. The performance of the proposed method is tested with three numerical experiments, and the results are cross-compared with those from the variance-based GSA.

It is found that the proposed method can properly identify the influential and non-influential inputs from a complex model with several independent and dependent inputs. Furthermore, compared with the variance-based GSA, the proposed screening method only needs a few model runs, while the screening accuracy is well maintained. Therefore, it can be regarded as a practical tool for the initial GSA of high dimensional and computationally expensive models with dependent inputs.

1. Introduction

Along with the continuous development of computational techniques, and the increasing power of computers, simulation models are more and more advanced and powerful nowadays, and they have become one essential resource for scientific research and practical applications. However, the complexity of simulation models has also significantly increased, especially due to the fact that they now often include a large number of parameters. To aid modelers to better understand the model, and to manage the uncertainties in model computation, it is necessary to further investigate the relationship between model inputs and output(s), especially when the model is extremely sophisticated and/or treated like a black box. One commonly used tool for such task is Global Sensitivity Analysis (GSA) [1].

The sensitivity information can be obtained through performing GSA in the input space, and analyzing the impacts of variations of inputs on the variations of output(s). Such information can be qualitative (e.g., the sets of influential and non-influential inputs) or quantitative (e.g., the variance contributions of model inputs to the total variance of model output(s)). They can be used for e.g., identifying the key inputs, reducing the uncertainties of model response, setting priorities for model calibration. Due to its importance, several GSA approaches have been extensively developed in the past few decades. In general, these approaches can be classified according to the sensitivity indexes to be assessed: derivative-based sensitivity measure [2–5], regression-based sensitivity measure [6,7], qualitative (or screening) sensitivity measure [8–12], variance-based sensitivity measure [13–17], and moment independent sensitivity measure [18–20]. More information regarding the different GSA approaches can be found in [1,21].

All these sensitivity measures, however, are based on the independence assumption of the model inputs. In practice, due to certain constrains in the input space, and/or the intricate relations among the inputs obtained from empirical experiments (e.g., some inputs are the outcome of another model or experiment) [22], it is very likely that the model inputs are actually dependent inputs, or mixtures of both independent and dependent inputs. In these cases, simply assuming that all model inputs are independent, and directly applying the aforementioned GSA methods can be fallacious, and consequently can lead to incorrect inferences (see e.g., the numerical experiments in Section 4).

To the authors' knowledge, only a few recent studies that are able to perform GSA of models with dependent inputs can be found in the literature. For example, studies such as [23–27] discussed the variance decomposition with parametric or non-parametric methods for models with dependent inputs, while [28–30] extended the analytical formula-

* Corresponding author.
E-mail addresses: qiao.ge@ivt.baug.ethz.ch (Q. Ge), monica.menendez@ivt.baug.ethz.ch (M. Menendez).

http://dx.doi.org/10.1016/j.ress.2017.01.010
Received 22 December 2015; Received in revised form 8 December 2016; Accepted 13 January 2017
Available online 15 January 2017
0951-8320/

tions and the corresponding numerical estimators to compute the extended Sobol' sensitivity indexes [13] for dependent inputs. All these methods are quantitative, i.e., they quantitatively decompose and analyze the variance contributions of dependent model inputs. Yet, for complex models in practice, the application of quantitative GSA might still be difficult or unfeasible when the model itself contains many inputs (i.e., high dimensional) and/or is computationally expensive.

Therefore, in this paper we recall the application of inputs screening before performing any quantitative GSA (see [21] for more details). The main goal is to identify the least influential inputs, so that they can be fixed at their nominal values without significantly influencing the model output. This approach is also known as the *Factor Fixing (FF) setting* [31,1]. It is found in [21,32,33] that for high dimensional and computationally expensive models with independent inputs, the inputs screening can reduce the complexity of the model, and hence enhance the efficiency of the quantitative GSA without losing accuracy. Therefore, it is expected that a proper screening method may also provide such benefits for models with dependent inputs.

The objective of this paper is thus to develop an efficient, non-parametric approach for screening dependent model inputs. The performance of the proposed screening method is evaluated through numerical experiments. The screening results from each experiment are cross-compared with the reference results obtained using the variance-based GSA [29,30]. This analysis shows that the proposed screening approach can efficiently identify the non-influential inputs with satisfactory accuracy at a low computational cost.

The paper is organized as follows. A brief review of the screening method, specifically, the Elementary Effects (EE) method and its recent extensions, as well as the variance decomposition for dependent inputs is given in Section 2. The methodology for the proposed screening method is presented in Section 3. The details about the numerical experiments, and the corresponding results are introduced and discussed in Section 4. Conclusions are given in Section 5.

2. Literature review

2.1. Elementary effects method

The Elementary Effects (EE) method, which is also known as the Morris method, was first introduced by Morris in [8]. Let f be a model with k independent inputs, i.e., $\mathbf{X} = \{X_1, X_2, \ldots, X_k\}$, which are defined in the k-dimensional input space Ω^k ($\Omega^k \subset \mathbb{R}^k$). Let Y be the model output, i.e., $Y = f(\mathbf{X})$. Moreover, let $\mathbf{x} = \{x_1, x_2, \ldots, x_k\}$ be the values assigned to $\mathbf{X}$ in Ω^k, i.e., $\mathbf{x} \in \Omega^k$. If only x_i (i.e., the value of input X_i, $i \in [1, k]$) is varied by a given value Δ, while the values of all other inputs remain unchanged, the corresponding model output is $f(x_1, \ldots, x_{i-1}, x_i + \Delta, x_{i+1}, \ldots, x_k)$, with $\{x_1, \ldots, x_{i-1}, x_i + \Delta, x_{i+1}, \ldots, x_k\} \in \Omega^k$. The formula for computing the EE of X_i (i.e., EE_i) is given below:

$$EE_i = \frac{f(x_1, \ldots, x_{i-1}, x_i + \Delta, x_{i+1}, \ldots, x_k) - f(\mathbf{x})}{\Delta}. \tag{1}$$

The above definition employs the One-At-a-Time (OAT) design. To investigate the global sensitivity of X_i, the OAT experiment is repeated using N different values of $\mathbf{x}$ that are randomly sampled in Ω^k. Accordingly, N EEs can be obtained for input X_i. In [8], the mean (i.e., μ_i) and the standard deviation (i.e., σ_i) of these N EEs are used as the screening measures:

$$\mu_i = \frac{1}{N}\sum_{r=1}^{N} EE_{i,r}, \tag{2}$$

$$\sigma_i = \sqrt{\frac{1}{N-1}\sum_{r=1}^{N} (EE_{i,r} - \mu_i)^2}, \tag{3}$$

where $EE_{i,r}$ corresponds to the r-th EE of X_i.

In [9], the absolute mean (i.e., μ_i^*) is proposed to replace μ_i:

$$\mu_i^* = \frac{1}{N}\sum_{r=1}^{N} |EE_{i,r}|, \tag{4}$$

where $|\cdot|$ stands for the absolute value. As discussed in [9], when the model contains many interactive inputs and/or the model is not monotonic, using μ_i^* as a screening measure can significantly reduce the Type II error (i.e., considering an influential input as non-influential [1]). According to [8,9], an input X_i is identified as:

a) a non-influential input, if μ_i^* is close to zero;
b) an influential input with negligible non-linear effects, if μ_i^* is high but σ_i is low; or
c) an influential input with non-linear effects and/or strong interactions with other inputs, if both μ_i^* and σ_i are high.

It was found in [2,3] that for non-monotonic models, the EE method is less accurate than the Monte Carlo/Quasi-Monte Carlo integration method in estimating the derivative-based sensitivity measures [2,4]. However, the same studies [2,3] also emphasized that such high accuracy of estimates may not be required for inputs screening, where the aim is to identify influential and non-influential inputs with a low computational cost. Thus, the EE method can still be considered as a good compromise between accuracy and efficiency, especially for the SA of high dimensional and computationally expensive models.

It is worth mentioning that although the EE method was developed for screening purposes [8], this approach can also be used for ranking the inputs in order of importance [34,35,3]. In particular, Saltelli et al. [35] recommended to use μ^* for inputs ranking when the SA is FF setting. Moreover, the empirical study in [9] showed that using μ^* for ranking independent model inputs could achieve similar results as using the Sobol' total sensitivity index (see Section 2.2.1). Due to its high efficiency, the EE method has been successfully employed for ranking independent inputs by many researchers from different disciplines (e.g., [9,36–38,33,39,40]).

To enhance its computational efficiency, the classic Morris EE method has been extended by adopting different sampling designs, for example, the trajectory design[1] [8,9,37,11], the cell design [10], and the radial design [42]. The trajectory design is the most commonly used sampling design for computing EE, while the radial design shows the best performance (see the experiments performed in [42] for more details). These two designs are therefore implemented in the screening approach proposed in this paper, and will be introduced in Section 3.2.1.

When the model contains dependent inputs, the classic EE method's biggest drawback is that it does not account for the impacts from inputs dependence. For example, when an input is changed by Δ, any other correlated input should simultaneously show some variation. However, such variation is are not included in Eq. (1). Hence, when dependence exists among model inputs, the application of the classic EE method could yield incorrect screening results. To take the effects of inputs dependence into account, it is necessary to recall the variance decomposition approach from the variance-based GSA. The corresponding information is presented in the following section.

2.2. Variance decomposition and sensitivity indexes

2.2.1. Model with independent inputs

We consider a square integrable function f with k independent inputs $\mathbf{X} = \{X_1, X_2, \ldots, X_k\}$ defined in $\mathbb{R}^k$. According to [13], f can be

[1] The trajectory design is also known as winding stairs in [41]. The difference is the trajectory design always produces random trajectories separately, while the winding stairs design joins all random trajectories together.