Reliability Engineering and System Safety 170 (2018) 20–30

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

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

Multivariate global sensitivity analysis for dynamic models based on wavelet analysis

Sinan Xiao [a], Zhenzhou Lu [a,*], Pan Wang [b]

[a] School of Aeronautics, Northwestern Polytechnical University, Xi'an, 710072, Shaanxi, China
[b] School of Mechanics and Civil & Architecture, Northwestern Polytechnical University, Xi'an, 710129, Shaanxi, China

ARTICLE INFO

Keywords:
Global sensitivity analysis
Dynamic model
Energy distribution
Wavelet analysis

ABSTRACT

Dynamic models with time-dependent output are widely used in engineering for risk assessment and decision making. Global sensitivity analysis for these models is very useful for simplifying the model, improving the model performance, etc. The existent covariance decomposition based global sensitivity analysis method combines the variance based sensitivity analysis results of the model output at all the instants, which just utilizes the information of the time-dependent output in time domain. However, many significant features of time-dependent output may not be obtained from the time domain. Thus, performing global sensitivity analysis for dynamic models just with the information in time domain may be incomplete. In this paper, a new kind of sensitivity indices based on wavelet analysis is proposed. The energy distribution of model output over different frequency bands is extracted as a quantitative feature of the time-dependent output, and it contains the information of model output in both time and frequency domains. Then, a vector projection method is utilized to measure the effects of input variables on the energy distribution of model output. An efficient algorithm is also proposed to estimate the new sensitivity indices. The numerical examples show the difference between the new sensitivity indices and the covariance decomposition based sensitivity indices. Finally, the new sensitivity indices are applied to an environmental model to tell the relative importance of the input variables, which can be useful for improving the model performance.

1. Introduction

In practical engineering systems and mathematical models, uncertainties are often encountered in the input factors or the parameters [1–3], which will lead to uncertain performance. Uncertainty analysis has been widely used to help decision makers understand the degree of confidence in the decision they made and assess the risk [4,5]. However, many applications of uncertainty analysis just focus on the propagation of uncertainty from model inputs to output, but they usually do not provide information on how the uncertainty of model output can be apportioned to the uncertainty of model inputs, and therefore, on which factors to devote data collection resources so as to reduce the uncertainty of model output most effectively [5,6]. Global sensitivity analysis (GSA) has been widely used to solve this problem and can effectively apportion the uncertainty of model output to different sources of uncertainty in the model inputs [7,8]. Thus, GSA can help researchers find the important and unimportant input factors, measure the relative contributions of the uncertainties of input factors to the uncertainty of model output or detect the interaction effect between different input factors. Then, researchers can reduce the uncertainty of model output effectively through the calibration of the most important input factors and simplify the model by fixing the unimportant input factors into nominal values. Due to these significant advantages, GSA has been widely used in risk assessment, decision making, etc. For example, Saltelli & Tarantola [9,10] applied GSA to the safety assessment of nuclear waste disposal, Frey & Patil [11–13] used GSA for the risk assessment of food safety, Iman et al. [14] applied GSA to the analysis of projecting losses associated with hurricanes, Borgonovo & Peccati [5] used GSA techniques in the investment decisions, Herman et al. [15,16] used GSA to measure the effects of parameter uncertainty on spatially distributed hydrologic models and facilitate the corresponding diagnostic analyses, Lamboni et al. [17] used GSA for dynamic crop models to help researchers make better decisions in the growing season of crops. More details of GSA can be found in the reviews about sensitivity analysis [18,19].

The traditional GSA methods, such as elementary effect method [20–22], variance based method [23–25], derivative based method [26–28] and moment independent method [29–31], etc. mainly focus on the models with scalar output. However, many practical models with dynamic (multivariate) output are widely used for risk assessment or decision making in engineering [32–34]. In these models, time-dependent model output is often considered. Usually, an appropriate scalar

* Corresponding author.
E-mail addresses: bruce1209@163.com (S. Xiao), zhenzhoulu@nwpu.edu.cn (Z. Lu), panwang@nwpu.edu.cn (P. Wang).

https://doi.org/10.1016/j.ress.2017.10.007
Received 5 March 2017; Received in revised form 10 October 2017; Accepted 15 October 2017
Available online 16 October 2017
0951-8320/

objective function of the model output (such as an aggregated statistic like sum, average or maximum value) is selected, then GSA is performed on the selected objective function [35]. If no appropriate scalar objective function can be obtained, GSA can also be performed on the model output at each instant respectively [36]. Although performing GSA on a pre-defined scalar objective function of the dynamic output can be convenient and useful when the selected scalar objective function has a meaningful interpretation, it may lead to too many potential scalar objective functions, and there are some situations where this reduction is not possible due to the specific nature of the problem [37]. On the other hand, conducting GSA on the model output at each instant can detect how the sensitivities of model inputs change over time, however, it cannot provide the information how the input variables affect the whole mode output in an entire time interval. As mentioned by Campbell et al. [38], it may be insufficiently informative to perform GSA on the output at each instant separately or on a few context specific scalar functions of the model output. Thus, conducting GSA on the whole dynamic model output will be an important complementarity to multivariate global sensitivity analysis.

For models with dynamic (multivariate) output, Campbell et al. [38] proposed an output decomposition method to identify the most interesting features of the model output and then perform GSA on these features. This method can be carried out by performing an orthogonal decomposition of the dynamic model output, then applying GSA on the most informative components individually. However, this method still focuses on a few components of model output rather than the whole model output. Based on this idea and the principal component analysis (PCA), Lamboni et al. [32] further proposed a set of generalized global sensitivity indices which can focus on the whole model output. Later, Gamboa et al. [39] defined a set of global sensitivity indices on the whole model output based on the decomposition of the covariance matrix of model output. These sensitivity indices are equivalent to the Sobol' indices [23] when the dynamic output degenerates into scalar output. Garcia-Cabrejo et al. [40] pointed out that the sensitivity indices based on covariance decomposition and the sensitivity indices based on output decomposition with PCA are equivalent if the selected eigenvectors in the principal component decomposition preserve the original total variance of model output, and the sensitivity indices based on covariance decomposition are more computationally efficient since they do not require spectral decomposition of the covariance matrix.

For the dynamic models, the output usually is a time-dependent variable, which can be represented as a stochastic process. The analysis of a time-dependent variable is often performed in the time domain, and the corresponding features in the time domain can be obtained. However, sometimes, the features of a time-dependent variable in the time domain are difficult to obtain. In this situation, it would be desirable to analyze time-dependent output in other domains, such as the frequency domain or time-frequency domain, in which some simple and useful features of the time-dependent output may be obtained easily. For example, in the field of signal processing and fault diagnose, time-dependent variables are often analyzed in the frequency domain [41] or time-frequency domain [42] to obtain important information for decision making. Fourier analysis [43] is a widely used tool for frequency analysis, and it can help researchers easily obtain the features of dynamic output in the frequency domain. However, there is a shortage of Fourier analysis, i.e. it is not able to reveal the inherent information of non-stationary time-dependent variables. This is caused by the fact that Fourier analysis is a pure tool for frequency analysis, i.e. it can just obtain the frequency components of a time-dependent variable but does not know when these frequency components occur. Wavelet analysis [44,45] is a powerful tool for time-frequency analysis. It not only can obtain the frequency bands of a time-dependent variable, but also can know the time intervals in which these frequency bands occur. Through wavelet analysis, researchers can obtain the features of the dynamic output at time-frequency domain, thus a more comprehensive understanding of the dynamic output can be obtained. The existent multivariate GSA methods (output decomposition method and covariance decomposition method) can be regarded as measuring the effects of model inputs on the dynamic output in the time domain. Since many useful information of dynamic output can also be obtained in the time-frequency domain, multivariate GSA can also be performed in the time-frequency domain.

In this work, wavelet analysis is used to extract a significant feature, i.e. energy distribution, of the dynamic output in the time-frequency domain, then multivariate GSA is performed based on the energy distribution of dynamic output through using a vector projection method, which measures the similarity between the unconditional variance vector and the conditional variance vector of the energy distribution of dynamic output. Compared to the output decomposition method with PCA and covariance decomposition method, the proposed method can obtain the information of dynamic output in the time-frequency domain, which can include more information of model output. The proposed method will be tested on several numerical examples with both stationary and non-stationary output variables. Finally, the proposed method is applied to a complex environmental model called LevelE, which is used for predicting the radiologic release to humans due to the underground migration of radionuclides. The LevelE model has become the benchmark model in GSA studies [10,24,46,47]. However, the previous studies usually focused on a single output, i.e. the maximum radiological dose simulated over a time period or the radiological dose at a given time. Here, the time-dependent output over a time interval is considered, and the proposed GSA method is conducted on this whole time-dependent output.

The rest of this paper is organized as follows. Section 2 reviews the multivariate GSA method based on the covariance decomposition of model output. In Section 3, the wavelet analysis is briefly introduced and multivariate GSA in the time-frequency domain is proposed. The discussion and estimation of the new sensitivity indices are shown in Section 4. In Section 5, the proposed method is compared with the covariance decomposition based method on several numerical examples. The application of the proposed method to the LevelE model is conducted in Section 6. Conclusions are given in Section 7.

2. Review of covariance decomposition based multivariate global sensitivity analysis

Let $X_i (i = 1, 2, \dots, d)$ be a set of independent random input variables with probability density function (PDF) $f_{X_i}(x_i) (i = 1, 2, \dots, d)$. The output of the dynamic model is defined as

$$Y(t) = g(X_1, X_2, \dots, X_d, t), \quad t \in T \tag{1}$$

where $Y(t)$ is the model output at time t and $g(X_1, X_2, \dots, X_d, t)$ is a deterministic model response function. The output becomes a vector $\mathbf{Y} = (Y(t_1), Y(t_2), \dots, Y(t_m))$ if the time domain T is discrete or more generally a function $Y(t)$ $(t \in T)$ if T is continuous.

The covariance decomposition based multivariate global sensitivity analysis method was proposed by Gamboa et al. [39]. This method is based on the high dimensional model representation [23] of the outputs, i.e.

$$\begin{aligned} Y(t_r) = {} & g_{0,t_r} + \sum_{i=1}^{d} g_i(X_i, t_r) + \sum_{1 \le i < j \le d} g_{i,j}(X_i, X_j, t_r) \\ & + \dots + g_{1,2,\dots,d}(X_1, X_2, \dots, X_d, t_r), \quad r = 1, \dots, m \end{aligned} \tag{2}$$

where

$$\begin{aligned} & g_{0,t_r} = E(Y(t_r)) \\ & g_i(X_i, t_r) = E\left(Y(t_r) \middle| X_i\right) - g_{0,t_r} \\ & g_{i,j}(X_i, X_j, t_r) = E\left(Y(t_r) \middle| X_i, X_j\right) - g_i(X_i, t_r) - g_j(X_j, t_r) - g_{0,t_r} \\ & \dots \end{aligned} \tag{3}$$