Reliability Engineering and System Safety 147 (2016) 123–131

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

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

A new kind of sensitivity index for multivariate output

Luyi Li *, Zhenzhou Lu, Danqing Wu

School of Aeronautics, Northwestern Polytechnical University, Xi'an 710072, PR China

ARTICLE INFO

Article history:
Received 5 February 2015
Received in revised form
1 July 2015
Accepted 8 November 2015
Available online 2 December 2015

Keywords:
Sensitivity analysis
Multivariate output
Multivariate probability integral transformation
Cumulative distribution function

ABSTRACT

Mathematical and computational models with correlated multivariate output are commonly used for risk assessment and decision support in engineering. Traditional methods for sensitivity analysis of the model with scalar output fail to provide satisfactory results for this multivariate case. In this work, we introduce a new sensitivity index which looks at the influence of input uncertainty on the entire distribution of the multivariate output without reference to a specific moment of the output. The definition of the new index is based on the multivariate probability integral transformation (PIT), which can take into account both of the uncertainties and the correlations among multivariate output. The mathematical properties of the proposed sensitivity index are discussed and its differences with the sensitivity indices previously introduced in the literature are highlighted. Two numerical examples and a rotating shaft model of an aircraft wing are employed to illustrate the validity and potential benefits of the new sensitivity index.

1. Introduction

Mathematical and computational models with correlated multivariate output are frequently used for risk assessment and decision support in engineering. Traditional methods for sensitivity analysis of model with scalar output, such as variance-based method and moment-independent approach, can be applied to each output in this situation. However, as pointed out by Campbell et al. [1], it may be insufficiently informative to perform sensitivity analysis on each output separately or on a few context specific scalar functions of the output. Furthermore, this will lead to a large number of redundant sensitivity indices if the correlation in the output is strong, situation in which it is difficult to interpret the results of the sensitivity analysis [2]. It is recommended by Saltelli and Tarantola [3] to simplify the original problem defining a scalar variable of interest to apply the sensitivity analysis. Although this approach can be applied in many cases, there are some situations where this reduction is not possible due to the specific nature of the problem. In this situation, it is more appropriate to apply sensitivity analysis to the multivariate output as a whole. Consequently, there is a need to define criteria and to develop methods specifically adapted to the sensitivity analysis of multivariate output.

Campbell et al. [1] proposed a simple and very useful approach, called output decomposition method, for sensitivity analysis of model with functional output, which was then applied by Lamboni et al. [4] to mathematical models of crop growth with output displaying temporal variations. It consists in (i) performing an orthogonal decomposition of the multivariate output, and (ii) applying sensitivity analysis to the most informative components individually. The method allows restricting attention to a few components rather than a whole dynamic. To summarize the sensitivity over the whole dynamic, Lamboni et al. [5] further proposed a new set of synthetic sensitivity indices for multivariate output which can be applied in the method defined by Campbell et al. [1]. Gamboa et al. [6] defined a new set of sensitivity measures based on decomposition of the covariance of the model output that are equivalent to the Sobol indices in the scalar case. This approach does not require the spectral decomposition of the covariance matrix as in the output decomposition, and thus is considered to be more efficient in computational terms [2]. Detailed comparison and equivalence between the output and covariance decomposition approaches with their efficient polynomial chaos expansion estimates can be found in [2].

All these sensitivity analysis methods for multivariate output are based on the variance of the multivariate output, which implicitly assumes that the variance is sufficient to describe output variability. However, variance only provides a summary of the whole distribution of the output with the inevitable loss of information that occurs when the information contained in the whole distribution is mapped into a single quantity [7]. Therefore, a sensitivity index which refers to the entire distribution of the

* Correspondence to: School of Aeronautics, Northwestern Polytechnical University, Xian City, P.O. Box 120, 710072, Shaanxi Province, PR China.
Tel.: +86 29 88460480.
E-mail address: luyili@mail.nwpu.edu.cn (L. Li).

http://dx.doi.org/10.1016/j.ress.2015.11.006

multivariate output should be used if one wants to assess which input influences the decision-maker state of knowledge. Cui et al. [8] extended the moment-independent sensitivity analysis method for scalar output [9] to the multivariate case, and defined a sensitivity index based on the joint probability density function (PDF) of the multivariate output. This method, although, can take into account both the entire uncertainty and correlation of the multivariate output. It suffers severely from the "curse of dimensionality" for computing the high dimensional integration in the sensitivity index, let alone the difficulty in estimating the joint PDF of the high dimensional variables. The advantages of using the distance between cumulative distribution functions (CDFs) as a measure of the input importance in case of scalar output have been manifested in many literatures [10–12], yet it has not been extended to the case of multivariate output. As pointed output in Ref. [13], the CDF-based method is easier to implement than the PDF-based method, and the computational efficiency of the CDF-based method can be improved as compared with the PDF-based method.

In this work, we introduce a new sensitivity index which looks at the influence of input uncertainty on the entire distribution of the multivariate output without reference to any of its specific moment. The entire uncertainty of the multivariate output is described by its joint CDF in the new method instead of joint PDF, which can be easily approximated with the random samples of the outputs. The definition of the new sensitivity index is then based on the multivariate probability integral transformation (PIT) distribution of the multivariate output, which has been recognized containing valuable information about the correlation structure of the joint CDF of the outputs. With our proposed method, the challenges of considering both uncertainty and correlations of the multivariate output are addressed. Besides, due to the univariate nature of the multivariate PIT, the proposed sensitivity index is evaluated through univariate integration regardless of the dimensionality of the outputs, which significantly cuts down the computational cost compared to the method based on the joint PDF.

This paper is organized as follows: Section 2 reviews existing sensitivity analysis methods for model with multivariate output. A brief introduction of the probability integral transformation theorem is provided at the beginning of Section 3, followed by the definition of the new sensitivity index. The mathematical properties of the new index are also discussed and proved in this section. The differences of the proposed sensitivity index with those previously introduced in the literature are highlighted in Section 4. Two numerical examples and a rotating shaft model of an aircraft wing are employed in Section 5 to illustrate the validity and potential benefits of the new sensitivity index. Conclusions come at the end of the paper.

2. Sensitivity analysis methods for multivariate output

Let $X_i, i=1,...,n$ be a set of independent input variables characterized by the PDFs $f_{X_i}(x_i), i=1,...,n$, and $\mathbf{Y}=(Y_1,...,Y_m)$ be the output vector of the model of interest defined as $Y_r=g(X_1,...,X_n,r)$ $,r=1,...,m$ where $g(X_1,...,X_n,r)$ is a deterministic model function, and with a positive semidefinite covariance matrix $\mathbf{C}(Y_1,...,Y_m)$. Three different methods have been proposed for sensitivity analysis in the case of multivariate output, which are called output decomposition method [1], covariance decomposition approach [6] and PDF based method [8]. The main concepts of these methods are briefly reviewed in this section.

2.1. Output decomposition method

Campbell et al. [1] proposed a method for sensitivity analysis of functional output that is based on two steps:

(1) Performing an orthogonal decomposition as shown in Eq. (1) of the multivariate output

$$Y_r-\overline{Y}_r=\sum_{k=1}^{K}h_{r,k}\varphi_k, r=1,...,m \tag{1}$$

where $\overline{Y}_r$ is the mean of Y_r, φ_k are the orthonormal bases, and $h_{r,k}$ are the corresponding coefficients, with K the number of basis used.

(2) Applying sensitivity analysis on the coefficients $h_{r,k}$ to identify and rank the input variables associated with each orthonormal basis in the previous expansion and the presence of interactions between these input variables in each basis.

As indicated in [1], there is a large collection of available methods for the first step: it can be based either on a data driven method such as principal component analysis (PCA), or on the projections of output on a polynomial, spline, or Fourier basis defined by the user. The second step can also be performed by many sensitivity analysis methods, such as factorial design, Fourier amplitude sensitivity test (FAST), or Sobol' analysis of variance (ANOVA) decomposition and its most recent versions developed by Saltelli et al. [14].

By combining the PCA and ANOVA decompositions, Lamboni et al. [5] defined a set of generalized sensitivity indices for multivariate case as shown in Eqs. (2)–(5), where the orthogonal bases φ_k are the eigenvectors of the covariance matrix $\mathbf{C}(Y_1,...,Y_m)$.

The first order sensitivity index of the input variable X_i for the coefficient $h_{r,k}$ of the kth eigenvector, i.e. the kth principal component of the multivariate output, is defined by [5]

$$SI_{i,k}=\frac{V_{i,k}}{V_k} \tag{2}$$

where $V_{i,k}$ is the variance of the kth eigenvector due to changes in input variable X_i and $V_k=\lambda_k$ is the variance of the kth eigenvector that is equal to the kth eigenvalue of the spectral decomposition of the covariance matrix. Similarly, the total sensitivity index of the input variable X_i, including all the interaction of X_i with the other inputs, for the kth principal component $h_{r,k}$ is defined by Eq. (3) [5],

$$TSI_{i,k}=\frac{\sum_{\omega_i}V_{\omega_i}}{V_k} \tag{3}$$

where ω_i includes all terms in the ANOVA decomposition that include the input variable X_i.

To further quantify the contribution of each input variables X_i, $i=1,...,n$ to the total variance of the multivariate output $Y_r, r=1,...,m$, two generalized sensitivity indices are defined in [5], and they are generalized first order sensitivity index

$$GSI_i=\sum_{k=1}^{K}\frac{\lambda_k}{V[\mathbf{Y}]}SI_{i,k} \tag{4}$$

and generalized total sensitivity index

$$GTSI_i=\sum_{\omega_i}GSI_{\omega_i} \tag{5}$$

where λ_k is the eigenvalue (variance) of the kth eigenvector.

While the generalized first order sensitivity index GSI_i measures the contribution of the input variable X_i on the total variance of the output , the generalized total sensitivity index $GTSI_i$ quantifies the overall contribution of X_i on $V[\mathbf{Y}]$, including the unique contribution of X_i as well as the contributions by all the interaction