Reliability Engineering and System Safety 165 (2017) 1–10

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

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

Multivariate sensitivity analysis based on the direction of eigen space through principal component analysis

Sinan Xiao, Zhenzhou Lu*, Liyang Xu

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

ARTICLE INFO

Keywords:
Sensitivity analysis
Multivariate outputs
Principal component analysis
Eigen space
Hydrological model

ABSTRACT

In this paper, a new kind of sensitivity indices based on the principal component analysis (PCA) is proposed to measure the effects of input variables on multivariate outputs. Through PCA, the outputs are projected onto a new coordinate system (eigen space), which is constructed by the eigenvector (principal components). The existent sensitivity indices based on PCA focus on the variance of principal components, which can be considered as a magnitude of the uncertainty in the corresponding coordinate axes. In addition, the direction of the coordinate axes in the eigen space also contains another part of uncertainty of outputs (the direction of the uncertainty). The new sensitivity indices measure the effect of input variables on the direction of the coordinates axes through the angles between the unconditional and conditional eigenvectors. Thus, the new sensitivity indices can reflect different effect of input variables on the output compared to the existent sensitivity indices. The results of three numerical examples and an environmental model show the difference between the new sensitivity indices and the existent sensitivity indices. Since the new sensitivity indices measure the effects of input variables on the multivariate outputs from a different perspective compared to the existent sensitivity indices, they should be mutually complementary to each other.

1. Introduction

Uncertainties are often encountered in the practical systems and models [1–3], which lead to uncertain performance. Uncertainty analysis is widely used to help decision makers to understand the degree of confidence of the model results so that they can know the degree of confidence in the decision they made and assess the risk [4,5]. However, the results of uncertainty analysis can't provide information on how the uncertainty of the output can be apportioned to the uncertainty of inputs, and therefore, on which factors to devote data collection resources so as to reduce the uncertainty most effectively [6,7]. Global Sensitivity analysis (GSA) is a scientific analysis tool to solve this problem, which can apportion the uncertainty of model output to different sources of uncertainty in the model input [8,9]. Thus, it can help researchers find the important model inputs, i.e., the inputs which have a significant influence on output [10]. More details of GSA can be found in the reviews of other researchers, such as reference [11]. GSA has been widely used in risk assessment and decision making. For example, Saltelli & Tarantola [12,13] used GSA on the safety assessment for nuclear waste disposal, Herman et al. [14,15] used GSA on the hydrologic models to detect the sensitive factors affecting the model performance, then the model performance can be improved more efficiently, Frey & Patil [16,17] used GSA for the food safety risk assessment, Cuntz et al. [18] applied GSA through the sequential screening for the identification of non-informative hydrological model parameters with low computational cost, Borgonovo & Peccati [19] used GSA techniques in the investment decisions, Lamboni et al. [20] used GSA for dynamic crop models to help researchers make better decisions in the growing season of crops.

The traditional GSA methods, such as the elementary effect method [10,21], variance based method [22–24], derivative based method [25,26] and moment dependent method [27–29], focus on the model with single output. However, practical models with multivariate outputs are widely used for risk assessment and decision making in practical engineering [30]. A direct way to perform sensitivity analysis for models with multivariate outputs is to perform sensitivity analysis for each output separately. However, this way is just a repetition of the traditional GSA and it ignores the correlations among the multivariate outputs. Thus, it may be insufficient to perform sensitivity analysis on each output separately or on a few context specific scalar functions of the output [31]. In addition, redundant sensitivity indices will be generated if a strong correlation exists among the outputs, and it is difficult to interpret the results of sensitivity analysis [32]. It is recommended in [31] to apply sensitivity analysis to the multivariate

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

http://dx.doi.org/10.1016/j.ress.2017.03.011
Received 13 July 2016; Received in revised form 8 December 2016; Accepted 1 March 2017
Available online 21 March 2017
0951-8320/

output as a whole, and criteria and methods need to be developed for the sensitivity analysis of multivariate outputs.

Campbell et al. [33] proposed that sensitivity analysis of multivariate outputs can be carried out by (1) expanding the multivariate outputs in terms of an appropriate set of basis function, (2) performing sensitivity analysis on most informative components separately. According to this method, researches can focus on a few components rather than the whole outputs. Lamboni et al. [31] utilized the principal component analysis (PCA) [34,35] to perform the expansion of the outputs and carried out sensitivity analysis on the principal components. In addition, they also proposed generalized sensitivity indices for the multivariate outputs as a whole. Gamboa et al. [36] defined generalized Sobol' sensitivity indices for multivariate outputs based the decomposition of covariance matrix of model outputs. This method doesn't need spectral decomposition compared to the output decomposition method [31], thus it is more computational efficient. However, the output decomposition method can focus on the most informative components, which can reduce the output dimension. Garcia-Cabrejo et al. [32] pointed out that the output decomposition method and the covariance decomposition method are equivalent if the first K eigenvectors in the principal component decomposition preserve the original variance of outputs.

The covariance decomposition method [36] provides a generalized form of the Sobol indices [24], which can be regarded as measuring the effects of inputs on each outputs and then taking the weighted average of these effects to measure the effects of inputs on the whole output. However, it just consider the variance of each output and ignore the covariance among the outputs. For the PCA-based output decomposition method [31], the original outputs are projected into a new coordinate system (eigen space), which is constructed by the eigenvectors (principal components). Then, the influence of the model inputs on the variance of the principal components measures the importance of each input, and the weighted average is taken to measure the effects of inputs on the whole output. Since the principal components are mutually orthonormal to each other, there are no covariance among them, i.e., it is enough to just consider the variance of each component. Through the PCA decomposition, the uncertainty of model outputs are transformed into the principal components. The variance of the principal components just contains a part of the original uncertainty, and the direction of the eigenvectors contains another part of the uncertainty. Thus, measuring the effect of model inputs on the direction of the eigenvectors will be complementary to the existent sensitivity analysis method.

In this work, a new kind of sensitivity indices is proposed which measure the importance of model inputs through the influence of the inputs on the direction of the transformed output space obtained by PCA. PCA is multivariate statistical method [34,35], which can transform the original variables into a set of new orthogonal variables which are sorted according to their variance. Through the principal component decomposition, a set of eigenvectors (basis vectors), representing the directions of the dimensions in the transformed output space, can be obtained. The new kind of sensitivity indices, which contains the sensitivity indices on the principal components and the generalized sensitivity index, measures the effects of inputs on the outputs through the angles between the unconditional eigenvectors and the conditional eigenvectors.

The rest of this paper is organized as follows: Section 2 reviews the basic theory of PCA and the original sensitivity analysis method for multivariate outputs based on PCA. The new sensitivity index is defined in Section 3, followed by the estimation of the new sensitivity indices. Section 4 compares the new sensitivity indices with the original sensitivity indices through two numerical examples. A hydrological model is studied in Section 5 to measure effects of the uncertainties of different parameters on the model performance. Discussion and conclusion are given in Section 6.

2. Original sensitivity analysis method for multivariate outputs based on principal component analysis

Consider the model response function represented as $\mathbf{Y} = g(\mathbf{X})$, where $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ represents the n-dimensional vector of model input variables, $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_m)$ represents the m-dimensional vector of model output variables. Input variables are independent to each other and are characterized by the probability density function (PDF) $f_{X_i}(x_i)(i = 1, 2, \ldots, n)$.

2.1. Principal component analysis

Principal component analysis (PCA) is a widely used multivariate statistical method, which can transform the original variables into a set of new orthogonal variables, so that most information is contained in the first few components with the largest variance [34,35].

PCA can be performed through the eigenvalue decomposition of the covariance matrix of the outputs. First, center the outputs $\mathbf{Y}$ by subtracting the mean and denote the centered outputs as $\mathbf{Y}^c$, i.e.,

$$\mathbf{Y}^c = \mathbf{Y} - \boldsymbol{\mu}_\mathbf{Y} \tag{1}$$

where $\boldsymbol{\mu}_Y$ is the mean vector of the outputs $\mathbf{Y}$.

Then, perform the eigenvalue decomposition of the covariance matrix, i.e.,

$$\boldsymbol{\Sigma}_\mathbf{Y} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}^T \tag{2}$$

where $\boldsymbol{\Sigma}_\mathbf{Y}$ is the covariance matrix of outputs $\mathbf{Y}$, $\boldsymbol{\Lambda} = diag(\lambda_1, \lambda_2, \ldots, \lambda_m)$ is the diagonal eigenvalue matrix ($\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$ are the eigenvalues), $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_m)$ is the eigenvector matrix ($\boldsymbol{\gamma}_i (i = 1, 2, \ldots, m)$ represent the normalized and mutually orthogonal eigenvectors associated to the eigenvalues). Then the centered outputs $\mathbf{Y}^c$ are transformed into independent variables $\mathbf{H}$ through

$$\mathbf{H} = \mathbf{Y}^c\boldsymbol{\Gamma} \tag{3}$$

where $\mathbf{H} = (H_1, H_2, \ldots, H_m)$ contains the principals components, which are orthogonal to each other.$H_k (k = 1, 2, \ldots, m)$ is centered with variance λ_k, i.e.,$V(H_k) = E(H_k^2) = \lambda_k$. It can also be gotten that $\sum_{k=1}^{m} \lambda_k = trace(\boldsymbol{\Sigma}_\mathbf{Y})$. Through Eq. (3), $\mathbf{Y}^c$ can also be expressed by $\mathbf{H}$ as

$$\mathbf{Y}^c = \mathbf{H}\boldsymbol{\Gamma}' \tag{4}$$

where $\boldsymbol{\Gamma}'$ represents the inverse matrix of $\boldsymbol{\Gamma}$. Thus, the original outputs $\mathbf{Y}$ can be expanded by the mutually orthogonal principal components in $\mathbf{H}$ by

$$\mathbf{Y} = \boldsymbol{\mu}_\mathbf{Y} + \mathbf{H}\boldsymbol{\Gamma}' \tag{5}$$

Usually, the first K principal components containing the most variance of the original outputs are selected, then $\mathbf{Y}$ can be approximately expressed as

$$\mathbf{Y} = \boldsymbol{\mu}_\mathbf{Y} + \mathbf{H}_K\boldsymbol{\Gamma}'_K \tag{6}$$

where $\mathbf{H}_K$ contains the first K principal components of $\mathbf{H}$, $\boldsymbol{\Gamma}_K$ denotes the first K eigenvectors in $\boldsymbol{\Gamma}$.

2.2. Original sensitivity indices based on principal component analysis

By combining the principal component decomposition of model outputs $\mathbf{Y}$ in Eq. (6) and the analysis of variance (ANOVA) decomposition [37], Lamboni et al. [31] defined a set of sensitivity indices on the principal components $H_k (k = 1, 2, \ldots, m)$ and generalized sensitivity indices on the whole multivariate outputs.

The first order sensitivity index of model input variable X_i on the kth principal component H_k of the multivariate outputs $\mathbf{Y}$ is defined as [31]

$$SI_{i,k} = \frac{V_{i,k}}{V_k} \tag{7}$$