Reliability Engineering and System Safety 126 (2014) 25–36

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

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

Global Sensitivity Analysis for multivariate output using Polynomial Chaos Expansion

Oscar Garcia-Cabrejo*, Albert Valocchi

Hydrosystems Laboratory, Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, 61801, USA

ARTICLE INFO

Article history:
Received 27 June 2013
Received in revised form
2 January 2014
Accepted 5 January 2014
Available online 11 January 2014

Keywords:
Global Sensitivity Analysis
Sobol indices
Polynomial chaos
Sparse polynomial chaos
Regression

ABSTRACT

Many mathematical and computational models used in engineering produce multivariate output that shows some degree of correlation. However, conventional approaches to Global Sensitivity Analysis (GSA) assume that the output variable is scalar. These approaches are applied on each output variable leading to a large number of sensitivity indices that shows a high degree of redundancy making the interpretation of the results difficult. Two approaches have been proposed for GSA in the case of multivariate output: output decomposition approach [9] and covariance decomposition approach [14] but they are computationally intensive for most practical problems. In this paper, Polynomial Chaos Expansion (PCE) is used for an efficient GSA with multivariate output. The results indicate that PCE allows efficient estimation of the covariance matrix and GSA on the coefficients in the approach defined by Campbell et al. [9], and the development of analytical expressions for the multivariate sensitivity indices defined by Gamboa et al. [14].

1. Introduction

Important models of interest in engineering produce vector or multivariate output, such as partial differential equations (PDE) that describe spatial and temporal changes of variables of interest. Conventional methodologies for Global Sensitivity Analysis (GSA), including Sobol decomposition [24] and moment-independent approach [8], were designed for scalar output. However, these methodologies are applied to each one of the variables comprising the multivariate output resulting in a large number of sensitivity measures. If the correlation in the output is strong, there is a high degree of redundancy in the estimated indices, situation in which it is difficult to interpret the results of the sensitivity analysis. Saltelli and Tarantola [23] warned about this problem *"A possible cause of difficulty for SA is when the model consists of a large set (k) of input factors and, simultaneously, has many output variables (e.g., m). In such a case, a complete analysis would require the estimation of the sensitivity of each output to every input, thus returning $m \times k$ indices. We believe that the analysis can be made more effective by focusing not on the model output per se but on the problem that such output is supposed to solve. To this end, model use should be declared before uncertainty and sensitivity analyses are performed"*. In other words, the suggestion proposed by Saltelli and Tarantola [23] is to simplify the original problem defining a scalar variable of interest to apply the GSA. 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 the case of PDE, the use of the sensitivity measures designed for scalar output ignores the important characteristics of spatial and/or temporal correlation that is generated from the physical processes encoded by mathematical model. There is a growing need for GSA methodologies specifically designed for multivariate output due to the increasing role of complex quantitative models used by engineers/scientists to support decision making. There are two approaches for the application of GSA in the case of multivariate output. In the first approach, Campbell et al. [9] proposed a methodology for GSA when the model output can be represented as functions that can be extended to multivariate output. In this approach the model output is decomposed in an orthogonal basis and then GSA is applied to the coefficients of this expansion.

Lamboni et al. [18] applied this approach to mathematical models of crop growth, where the output displayed temporal variations. The orthonormal basis used in this case was the eigenvectors of the covariance matrix, and the sensitivities of the coefficients were estimated using conventional ANOVA decomposition. Following this work, Lamboni et al. [19] proposed a new set of sensitivity indices for multivariate output that can be applied in the approach defined by Campbell et al. [9]. In the second approach, Gamboa et al. [14] 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

* Corresponding author.
E-mail address: garcia30@illinois.edu (O. Garcia-Cabrejo).

0951-8320/$ - see front matter © 2014 Published by Elsevier Ltd.
http://dx.doi.org/10.1016/j.ress.2014.01.005

approach [18,19] and therefore it is expected to be more efficient in computational terms.

In the previous approaches, estimation of sensitivity measures is based on Monte Carlo simulation [24,21,22], an approach that is simple to apply but in some cases requires a large number of model evaluations. Metamodels are good alternatives for estimation of sensitivity indices because they are simpler to evaluate than the original model and can be more accurate than the conventional Monte Carlo simulation in the case of small to moderate sample sizes [26]. There are different types of metamodels such as multiple linear and nonlinear regression, cubic splines, Artificial Neural Networks, Gaussian Processes and orthogonal polynomials. A specific type of orthogonal polynomial metamodel is the so-called Polynomial Chaos Expansion (PCE), that is, a series expansion of a random variable using orthogonal basis that depends on the predetermined Probability Density Functions (PDF) [17]. A typical example is the use of Hermite polynomials to represent normal random variables.

The importance of PCE for GSA of scalar output is that analytical expressions for Sobol indices can be obtained from the coefficients of PCE, a result established by Sudret [27]. However, the use of PCE in the case of GSA for multivariate output has not been studied to date. Therefore in this paper, PCE is applied to two approaches of GSA for multivariate output:

1. In the decomposition of output approach, PCE is used to estimate efficiently the covariance matrices used to define the orthogonal decomposition of the output, and to obtain the Sobol decomposition of the coefficients in the resulting expansion.
2. In the covariance decomposition approach, PCE allows the development of analytical expressions for the multivariate sensitivity indices proposed by Gamboa et al. [14] in a similar way to the approach proposed by Sudret [27] for the scalar case.

In addition, these two approaches are applied to a simple problem of reactive transport in porous media, allowing for a comparison with the goal of identifying advantages and disadvantages of these methodologies. This paper is organized as follows: Section 2 includes a description of the main concepts used in this paper (scalar and multivariate GSA in Section 2.1, a review of the PCE in Section 2.3, the application of PCE to scalar and multivariate GSA in Section 2.4), the proposed approach is tested on a simple problem of multicomponent transport in porous media in Section 3, and finally results and discussion are included in Section 4.

2. Methodology

2.1. Global Sensitivity Analysis for scalar and multivariate output

The Global Sensitivity Analysis (GSA) is defined as the determination of how "*uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input*" [23,22], and one of the main methodologies for GSA is called Sobol decomposition [24,2].

Let $X_i,\ i=1,\dots,n$ be a set of independent input random variables (RV) defined on some probability space $(\Omega, F, \mathbb{P})$ (with Ω as the sample space, F is the σ-algebra of events, and $\mathbb{P}$ is the probability measure); the model output $\boldsymbol{Y}=(Y_1,\dots,Y_m)$ defined as $Y_r=g(X_1,\dots,X_n;r);\ r=1,\dots,m$ where $g(X_1,\dots,X_n;r)$ is a deterministic function; and with a positive semidefinite (PSD) covariance matrix $\boldsymbol{C}(Y_1,\dots,Y_m)$. In the case of scalar output $r=1$, the Sobol decomposition of the function $Y_1=g(X_1,\dots,X_n)$ is given by

$$Y_1=g_0+\sum_{i=1}^{n} g_i(X_i)+\sum_{1\le i<j\le n} g_{i,j}(X_i,X_j)+\dots+g_{i,j,\dots,n}(X_1,X_2,\dots,X_n) \tag{1}$$

where $g_i(X_i)$ represents the variation of Y_1 due to change of variable X_i only when the mean g_0 has been considered, and in the same way, $g_{i,j}$ represents the variation of Y_1 that is not accounted by the changes in variables i and j taken separately. This gives information on how the variables i and j are interacting. The terms in the expansion $g_{i,j,\dots}$ are orthogonal components. Due to the independence of the input variables $X_i,\ i=1,\dots,n$, the variance of output variable Y_1 is given by

$$V[Y_1]=\sum_{i=1}^{n} V_i+\sum_{1\le i<j\le n} V_{i,j}+\dots+V_{1,2,\dots,n} \tag{2}$$

The variation of the output Y_1 associated with variations in input variable X_i with no reference to other variables is given by the ratio of $V_i/V[Y_1]$, and this leads to the definition of the Single Effect Index [24] as

$$S1_i=\frac{V_i}{V[Y_1]} \tag{3}$$

The variation of output Y_1 associated with changes in input variable X_i interacting with other variables requires consideration of the variances associated with terms (single, pairs, triplets, etc) where X_i appears, and this defines the Total Effect Index [16]:

$$ST_i=\frac{V_i+\sum_{1\le i<j\le n} V_{ij}+\dots+V_{1,2,\dots,n}}{V[Y_1]}=1-\frac{V[Y_1|X_{\sim i}]}{V[Y_1]} \tag{4}$$

where $\sim i$ indicates fixing all input variables except variable i. In the case of multivariate output, the Sobol decomposition is obtained for each component $Y_r,\ r=1,\dots,m$ of the model output, leading to a large number of sensitivity measures for each output variable. In general these sensitivity measures are redundant if the correlation in the model output is important [18,19] leading to difficulties for the interpretation of these results. To deal with this problem, two different alternatives have been proposed for GSA in the case of multivariate output. These are called output decomposition [9] and covariance decomposition approaches [14].

2.1.1. Output decomposition approach

Campbell et al. [9] proposed an approach for GSA of functional output that is based on two steps:

1. Decomposition of the model output Y_r in terms of orthonormal basis $\varphi(\cdot)$:
$$Y_r=\overline{Y}_r+\sum_{k=1}^{K} h_{r,k}\varphi_k, \quad r=1,\dots,m \tag{5}$$
where $\overline{Y_r}$ is the mean of Y_r, $h_{r,k}$ are the coefficients and K is the number of basis used (in general $K \ll m$). The orthonormal basis can be obtained using methods such as Principal Component Analysis (PCA), and orthogonal polynomials.
2. Application of any approach of GSA on the coefficients $h_{r,k}$ to identify and rank the input variables associated with each orthonormal basis in the previous expansion (using the single effect index, see Eq. (3)) and the presence of interactions between these input variables in each basis (using the total effect index, see Eq. (4)). The combination of projection on a orthonormal basis and GSA is an useful tool to separate and identify the effect of specific physical process in the model response $Y_r,\ r=1,\dots,m$ being analyzed.

Lamboni et al. [19] defined the generalized sensitivity indices from the ANOVA decomposition of the coefficients in the expansion given in Eq. (5), where the orthogonal basis φ are the eigenvectors of the covariance matrix $C(Y_1,\dots,Y_m)$. The variation of the coefficients $h_{i,k}$ of the k-th eigenvector associated with variations in the