Reliability Engineering and System Safety 181 (2019) 97–115

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

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

Efficient method for variance-based sensitivity analysis

Xin Chen, Arturo Molina-Cristóbal, Marin D. Guenov*, Atif Riaz

School of Aerospace, Transport, and Manufacturing, Cranfield University, Cranfield, Bedfordshire, MK43 0AL, United Kingdom

ABSTRACT

Presented is an efficient method for variance-based sensitivity analysis. It provides a general approach to transforming a sensitivity problem into one uncertainty propagation process, so that various existing approximation techniques (for uncertainty propagation) can be applied to speed up the computation. In this paper, formulations are deduced to implement the proposed approach with one specific technique named Univariate Reduced Quadrature (URQ). This implementation was evaluated with a number of numerical test-cases. Comparison with the traditional (benchmark) Monte Carlo approach demonstrated the accuracy and efficiency of the proposed method, which performs particularly well on the linear models, and reasonably well on most non-linear models. The current limitations with regard to non-linearity are mainly due to the limitations of the URQ method used.

1. Introduction

In the context of Uncertainty Quantification and Management (UQ&M), sensitivity analysis is used to identify the contribution of different uncertainty sources on the total variance of system/model outputs [1]. This is particularly useful for large scale simulation or design problems, where it is normally impractical to consider all the factors, especially at the outset. Various techniques have been developed for sensitivity analysis. Systematic reviews can be found in [1–4] . Among these techniques, the variance-based method, also referred to as the Sobol' Indices, is widely used. It has the benefits of being 'global' and 'model-independent' [1]; where 'global' refers to analysing all the factors simultaneously over the entire region of interest, while 'model-independent' means that the approach is sufficiently general to handle different problems, without the need of knowing the inner structure of the models (i.e. models are treated as ''black-boxes'').

The development of variance-based sensitivity analysis dates back to 1970s, when Cukier et al [5–7], Schaibly and Shuler [8], proposed the method of Fourier Amplitude Sensitivity Test (FAST), in which the Fourier Transformation and searching curves were used to decompose the output variances. Similar problems were also referred to as 'Importance Measure' by Hora and Iman [9,10], Ishigami and Homma [11], and Saltelli et al [12,13]; or 'Top/Bottom Marginal Variance' by Jansen [14]. In parallel, Sobol' adopted the so-called ANOVA (Analysis of Variance)-representation to decompose a function, so that the portions of total variance caused by different factors can be formulated separately [15–19]. The numerical implementation is based on Monte Carlo Simulation along with multiple sampling sets (also referred as pick-freeze scheme [20]). It was later pointed out by Saltelli that all these methods calculate an equivalent statistical quantity [21], and that with this regard, the Sobol's approach is the most general one [22].

Further research has been focusing on the computational efficiency, which includes: improved sampling strategies (Sobol' sequences [23], Latin Hyper Cube [24], and Random Balance Design (RBD) [25–27]); improved formulation of estimators (Jansen [28], Saltelli [29], Sobol' et al [30]); approximation techniques (quadrature plus Latin Hyper Cube [31], grid quadrature [32]); Bayesian approach based on Gaussian processes (Oakley and O'hagan [33]); and Polynomial Chaos Expansion (PCE) [34–40] (where the polynomial coefficients are used to obtain the Sobol' indices), etc.

In general, for most of the aforementioned techniques (except [26,27]), the computational cost is related to the number of uncertainty sources, and becomes very expensive for high dimensional problems. Thus improving efficiency (i.e. the calculation speed), is still an area requiring further research, especially for early stage computational design, where the problem scale is large, and fast assessments are required.

In this research, a general approach is proposed to approximate the sensitivity indices based on the formulation from Saltelli [1,2,29]. In particular, we propose one implementation of the proposed approach, using the Univariate Reduced Quadrature (URQ) method [41], which was originally developed for uncertainty propagation.

The remaining part of the paper is structured as follows. Section 2 contains a background on variance-based sensitivity analysis and a brief description of the URQ method. In Section 3, the general approach for approximation is presented, followed by the detailed formulations incorporated with URQ, which include: the first order, second order, and total effect indices. The method is evaluated in Section 4, using a number of test-cases and is compared to the traditional (benchmark) MCS approach. Finally conclusions and future work are presented in

* Corresponding author at: School of Aerospace, Transport, and Manufacturing, Cranfield University, Cranfield, Bedfordshire, MK43 0AL, United Kingdom.
E-mail addresses: xin.chen@cranfield.ac.uk (X. Chen), m.d.guenov@cranfield.ac.uk (M.D. Guenov).

https://doi.org/10.1016/j.ress.2018.06.016
Received 19 December 2017; Received in revised form 2 May 2018; Accepted 27 June 2018
Available online 05 July 2018
0951-8320/

Section 5.

2. Background

The rationale and the derivation of the variance-based sensitivity analysis method is given by Saltelli in [1,2,29]. In this section, only a brief overview is presented, along with a short description of the URQ technique, which forms a part of the method proposed in Section 3.

2.1. Variance-based sensitivity indices

Consider a computational model with n input variables: $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$. It can be written in the form of a function:

$$y = f(\boldsymbol{x}) \tag{2.1}$$

Here, y is assumed to be the only output variable, while a vector $\boldsymbol{y} = (y_1, y_2, \cdots, y_m)$ can be used for multivariate output functions. In the original definition of Sobol' indices, each output is regarded as a separate scalar and the calculation process should be repeated for each of those. Recent research [42–49] has proposed several generalised sensitivity indices which are dedicated to the case of multivariate outputs, based on decomposition or covariance of the outputs. Such an extension is beyond the scope of the current research. Also, the input variables are assumed to be independent in this work. The reader is referred to [32,50–52] for further information regarding sensitivity analysis with correlated input variables.

2.1.1. First-order indices

A first-order index accounts for the portion of variance caused by uncertainty from only one of the inputs. For instance the sensitivity index of x_i, can be defined as [2]:

$$S_i = \frac{V(y) - E_{X_i}(V_{x\sim i}(y|_{x_i=X_i}))}{V(y)} \tag{2.2}$$

Here $V(y)$ is the total variance, while $V_{x\sim i}(y|_{x_i=X_i})$ is the conditional variance with x_i temporarily fixed as a constant X_i. The expectation $E_{X_i}(V_{x\sim i}(y|_{x_i=X_i}))$ is with regard to the randomness of X_i (which is equivalent to the randomness of x_i, as X_i is a realization of x_i).

By further expansion and derivation, Eq. (2.2) could be reformulated to the following forms:

$$S_i = \frac{E_{X_i}(E^2_{x\sim i}(y|_{x_i=X_i})) - E^2(y)}{V(y)} \tag{2.3}$$

$$S_i = \frac{V_{X_i}(E_{x\sim i}(y|_{x_i=X_i}))}{V(y)} \tag{2.4}$$

$$S_i = \frac{E(f_*(\boldsymbol{x}_{*i})) - E^2(y)}{V(y)} \tag{2.5}$$

The reader is referred to [1,2] for more details on the derivation of Eq. (2.4), and to [11,29] for the derivation of Eqs. (2.3) and (2.5). It should be noted that in Eq. (2.5), the problem is converted into a single loop expectation of the new function $f_*(\boldsymbol{x}_{*i})$, which is defined by multiplying the original function $f(\boldsymbol{x})$ with itself:

$$f_*(\boldsymbol{x}_{*i}) = f(x_1, x_2, \cdots, x_i, \cdots x_n)\cdot f(x'_1, x'_2, \cdots x'_{i-1}, x_i, x_{i+1}', \cdots x_n'), \tag{2.6}$$

where $\boldsymbol{x}_{*i}$ is the new input vector, which consists of $2n - 1$ variables. In this vector, x_k and x_k' are considered as independent variables for each $k = 1, 2, \cdots, n$; $k \neq i$, but with the same Probability Density Function (PDF). Also note that there is no x_i' in vector, $\boldsymbol{x}_{*i}$:

$$\boldsymbol{x}_{*i} = [x_1, x_2, \cdots, x_i, \cdots x_n, x_1', x_2', \cdots x_{i-1}', x_{i+1}', \cdots x_n'] \tag{2.7}$$

2.1.2. Second-order indices

A high order index captures the portion of variance caused by particular combinations (interaction effects) of the input variables [1,2]. For example, the second order index S_{ij} refers to the interaction effect caused by the combination of the i_{th} and j_{th} input variables. Note that this interaction effect leads to a portion in the output variance, while x_i and x_j are still independent inputs. In this research, only the second order indices are considered, but the same principle can be applied to calculate higher order indices as well.

Similar to the first order indices, S_{ij} can be calculated by solving the expectation of conditional variance with regard to two input variables. It can be proven that this formulation also includes the first order effects [1,2], therefore the first order indices need to be subtracted:

$$S_{ij} = \frac{V(y) - E_{X_{i,j}}\left(V_{x\sim i,j}\left(y|_{x_i=X_i, x_j=X_j}\right)\right)}{V(y)} - S_i - S_j \tag{2.8}$$

Some alternatives formulations [1,2] include,

$$S_{ij} = \frac{E_{X_{i,j}}\left(E^2_{x\sim i,j}\left(y|_{x_i=X_i, x_j=X_j}\right)\right) - E^2(y)}{V(y)} - S_i - S_j \tag{2.9}$$

$$S_{ij} = \frac{V_{X_{i,j}}\left(E_{x\sim i,j}\left(y|_{x_i=X_i, x_j=X_j}\right)\right)}{V(y)} - S_i - S_j \tag{2.10}$$

$$S_{ij} = \frac{E(f_{**}(\boldsymbol{x}_{**ij})) - E^2(y)}{V(y)} - S_i - S_j \tag{2.11}$$

Using similar reasoning as applied to Eqs. (2.5)–(2.7), $f_{**}(\boldsymbol{x}_{**ij})$ is defined by multiplying the original function $f(\boldsymbol{x})$ with itself, taking two different sets of independent inputs, but this time sharing the same x_i and x_j in both sets.

$$f_{**}(\boldsymbol{x}_{**ij}) = f(x_1, x_2, \cdots, x_i, \cdots, x_j \cdots x_n)\cdot f(x_1', x_2', \cdots, x_{i-1}', x_i, x_{i+1}', \cdots, x_{j-1}', x_j, x_{j+1}', \cdots, x_n') \tag{2.12}$$

Here $\boldsymbol{x}_{**ij}$ is the corresponding input vector, consists of $2n - 2$ variables. In this vector, x_k and x_k' are considered as independent variables for each $k = 1, 2, ..., n$; $(k \neq i, j)$, but with the same PDF. However there is no x_i' and x_j' in this vector.

$$\boldsymbol{x}_{**ij} = [x_1, x_2, \cdots, x_i, \cdots, x_j, \cdots, x_n, x_1', x_2', \cdots, x_{i-1}', x_{i+1}', \cdots, x_{j-1}', x_{j+1}', \cdots, x_n'] \tag{2.13}$$

2.1.3. Total effect indices

A total effect index accounts for the variable's first order effect and all its interactions with other variables [1,2]. That is,

$$S_i^T = S_i + \sum_{\substack{j=1\\ j\neq i}}^{n} S_{ij} + \sum_{\substack{j,k=1\\ k\neq j\neq i}}^{n} S_{ijk} + \sum_{\substack{j,k,l=1\\ l\neq k\neq j\neq i}}^{n} S_{ijkl} + \ldots \tag{2.14}$$

Apart from calculating sums using Eq. (2.14), which may become impractical when the number of inputs is high, this index is more widely calculated by using a nested structure as:

$$S_i^T = \frac{E_{X\sim i}(V_{x_i}(y|_{\boldsymbol{x}_{\sim i}=\boldsymbol{X}_{\sim i}}))}{V(y)}, \tag{2.15}$$

where all variables except x_i are first fixed for the calculation of the conditional variance, and then are varied in the expectation loop. The reader is referred to [1,2] for more rigorous mathematical derivation. By expansion and further deduction, Eq. (2.15) could be transferred as following alternatives,

$$S_i^T = \frac{V(y) + E^2(y) - E_{X\sim i}(E^2_{x_i}(y|_{\boldsymbol{x}_{\sim i}=\boldsymbol{X}_{\sim i}}))}{V(y)} \tag{2.16}$$