Reliability Engineering and System Safety, Volume 181, Issue January, 2019, pp. 97-115
DOI: 10.1016/j.ress.2018.06.016

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

* Corresponding Author, email address: m.d.guenov@cranfield.ac.uk

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 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: $\boldsymbol{x} = (x_1,\ x_2, \dots, x_n)$. 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 $\boldsymbol{y} = (y_1,\ y_2, \dots, y_m)$ 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]:

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

Here $V(y)$ is the total variance, while $V_{x_{\sim i}}\left(y|_{x_i=X_i}\right)$ is the conditional variance with x_i temporarily fixed as a constant X_i. The expectation $E_{X_i}\left(V_{x_{\sim i}}\left(y|_{x_i=X_i}\right)\right)$ 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, equation (2.2) could be reformulated to the following forms:

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

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

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

The reader is referred to [1,2] for more details on the derivation of equation (2.4), and to [11,29] for the derivation of equations (2.3) and (2.5). It should be noted that in equation (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 $f(x)$ with itself:

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

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 $k = 1,2,\dots,n; k \neq i$, but with the same Probability Density Function (PDF). Also note that there is no $x_i{}'$ in vector, $\boldsymbol{x_{*i}}$:

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

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,

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

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

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

Using similar reasoning as applied to equations (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.

$$\begin{aligned} f_{**}(\boldsymbol{x_{**ij}}) = {} & f(x_1, x_2, \dots, x_i, \dots, x_j \dots x_n) \\ & \cdot f(x_1', x_2', \dots, x_{i-1}', x_i, x_{i+1}', \dots, x_{j-1}', x_j, x_{j+1}', \dots, x_n') \end{aligned} \tag{2.12}$$

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 $k = 1,2,\dots,n;\ (k \neq i,j)$, but with the same PDF. However there is no $x_i{}'$ and $x_j{}'$ in this vector.

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

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,

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

Apart from calculating sums using equation (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:

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

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, equation (2.15) could be transferred as following alternatives,

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

$$S_i^T = \frac{V(y) - V_{\boldsymbol{X}_{\sim i}}(E_{x_i}(\, y|_{\boldsymbol{x}_{\sim i} = \boldsymbol{X}_{\sim i}}))}{V(y)} \quad (2.17)$$

$$S_i^T = \frac{V(y) + E^2(y) - E(f_{\sim *}(\boldsymbol{x}_{\sim * \boldsymbol{i}}))}{V(y)} \quad (2.18)$$

Again, the same reasoning as applied to equations (2.5) - (2.7), $f_{\sim *}(\boldsymbol{x}_{\sim * \boldsymbol{i}})$ is defined by multiplying the original function $f(x)$ with itself. This time all the inputs are the same except x_i and ${x_i}'$.

$$f_{\sim *}(\boldsymbol{x}_{\sim * \boldsymbol{i}}) = f(x_1, x_2, \dots, x_i, \dots, x_n) \cdot f(x_1, x_2, \dots, x_{i-1}, {x_i}', x_{i+1}, \dots, x_n) \quad (2.19)$$

Therefore $\boldsymbol{x}_{\sim * \boldsymbol{i}}$ consists of only $n + 1$ variables. In this vector, x_i and ${x_i}'$ are considered independent variables, but with the same PDF.

$$\boldsymbol{x}_{\sim * \boldsymbol{i}} = [x_1, x_2, \dots, x_i, \dots, x_n, {x_i}'] \quad (2.20)$$

2.2 Univariate Reduced Quadrature Method

The Univariate Reduced Quadrature (URQ) method was proposed by Padulo, Campobasso, and Guenov [41] for efficient uncertainty propagation. It uses the first four statistical moments (mean μ_{x_p}, standard deviation σ_{x_p}, skewness γ_{x_p}, and kurtosis Γ_{x_p}) of the stochastic inputs, and produces the mean and variance of the output.

$$E(y) \approx E_{URQ}(y) = W_0 f(\boldsymbol{\mu_x}) + \sum_{p=1}^{n} W_p \left[\frac{f(\boldsymbol{x_p^+})}{h_p^+} - \frac{f(\boldsymbol{x_p^-})}{h_p^-} \right] \tag{2.21}$$

$$\begin{aligned} V(y) \approx V_{URQ}(y) \\ = \sum_{P=1}^{n} \Bigg\{ W_p^+ \left[\frac{f(\boldsymbol{x_p^+}) - f(\boldsymbol{\mu_x})}{h_p^+} \right]^2 + W_p^- \left[\frac{f(\boldsymbol{x_p^-}) - f(\boldsymbol{\mu_x})}{h_p^-} \right]^2 \\ + W_p^{\pm} \frac{[f(\boldsymbol{x_p^+}) - f(\boldsymbol{\mu_x})][f(\boldsymbol{x_p^-}) - f(\boldsymbol{\mu_x})]}{h_p^+ h_p^-} \Bigg\} \end{aligned} \tag{2.22}$$

The coefficients in equation (2.21) and (2.22) are shown in the Table 1. The required number of model evaluation is $2n + 1$ in total.

Table 1. Coefficients and vectors used in URQ

$h_p^+ = \frac{\gamma_{x_p}}{2} + \sqrt{\Gamma_{x_p} - \frac{3\gamma_{x_p}^2}{4}}$	$W_0 = 1 + \sum_{p=1}^{n} \frac{1}{h_p^+ h_p^-}$
$h_p^- = \frac{\gamma_{x_p}}{2} - \sqrt{\Gamma_{x_p} - \frac{3\gamma_{x_p}^2}{4}}$	$W_p = \frac{1}{h_p^+ - h_p^-}$
$\boldsymbol{\mu_x} = [\mu_{x_1}, \mu_{x_2}, \dots, \mu_{x_n}]$	$W_p^+ = \frac{(h_p^+)^2 - h_p^+ h_p^- - 1}{(h_p^+ - h_p^-)^2}$
$\boldsymbol{x_p^+} = [\mu_{x_1}, \mu_{x_2}, \dots, \mu_{x_p} + h_p^+ \sigma_{x_p}, \dots, \mu_{x_n}]$	$W_p^- = \frac{(h_p^-)^2 - h_p^+ h_p^- - 1}{(h_p^+ - h_p^-)^2}$
$\boldsymbol{x_p^-} = [\mu_{x_1}, \mu_{x_2}, \dots, \mu_{x_p} + h_p^- \sigma_{x_p}, \dots, \mu_{x_n}]$	$W_p^{\pm} = \frac{2}{(h_p^+ - h_p^-)^2}$

3 Proposed Method

3.1 General Approach

The formulations reviewed in Section 2.1 are summarized in Table 2, where the equations are categorised into four options. In the traditional Monte Carlo Simulation (MCS) approach [11,16], only option 4 is adopted, because the other three options (1, 2, and 3) are computationally too expensive for MCS, due to the nested integrals in the formulations. The rationale of the proposed approach is that, since the nature of these integrals is to solve nested expectations/variances, the calculation can be reformulated as an uncertainty propagation process. As there are plenty of more efficient uncertainty propagation techniques compared with MCS, options 1, 2, and 3, may become computationally affordable.

Table 2. Equations used in four options to implement the proposed approach

Index	Option 1: Nested Expectation of Variance	Option 2: Nested Expectations	Option 3: Nested Variance of Expectation	Option 4: Single Loop Expectation
First Order: S_i	Eq. (2.2)	Eq. (2.3)	Eq. (2.4)	Eq. (2.5)
Second Order: S_{ij}	Eq. (2.8)	Eq. (2.9)	Eq. (2.10)	Eq. (2.11)
Total Effect: S_i^T	Eq. (2.15)	Eq. (2.16)	Eq. (2.17)	Eq. (2.18)

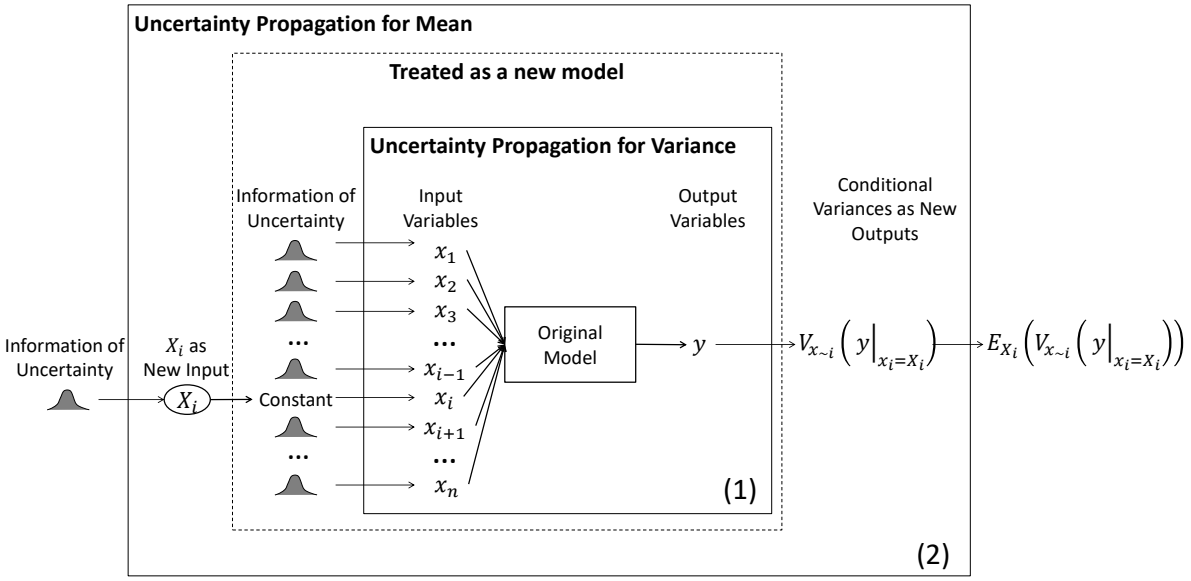

Figure 1. The general process, illustrated with first order index using option 1

The critical part is to construct the nested propagation loops as illustrated in Figure 1, by taking S_i in option 1 as an example. The uncertainty propagation is first applied to calculate the variance of the original model, as indicated by the number (1) in Figure 1. In this process, the i_{th} variable is temporarily fixed as X_i, thus the calculated variance is conditional. Now considering the variation of X_i, the block indicated by dashed lines can be treated as a 'new model', with X_i as its input, and the conditional variances as the outputs. Another propagation loop is conducted on top of this 'new model', regarding the uncertainty of X_i, as indicated by the number (2) in Figure 1. This outer loop provides the means (expectations) of the conditional variances, which could be used to calculate the indices as defined by Eq. (2.2). For Option 2 and 3, similar approaches could be applied by modifying the sequence of calculating the means and variances. Option 4, as discussed previously, does not require any nested loops. The approach is to apply single-loop uncertainty propagation for the mean of $f_*(\boldsymbol{x}_{*\boldsymbol{i}})$ defined by Eq. (2.6).

3.2 Formulation with URQ

Following the general approach, the specific formulation with the URQ method is deduced in this section, regarding the first order, second order, and total effect indices. In this research, we considered

all the four options as listed in Table 2, as it is difficult to predict their performance before numerical test. To avoid repetition, only the derivation of option 1 will be explained in detail.

3.2.1 First-order Indices

Given equation (2.2), as the computation of $V(y)$ via URQ has already been specified in equation (2.22), the following section will focus on the calculation of $E_{X_i}\left(V_{x_{\sim i}}\left(y|_{x_i=X_i}\right)\right)$.

To calculate the inner loop $V_{x_{\sim i}}\left(y|_{x_i=X_i}\right)$, firstly x_i is temporarily fixed as a constant X_i. Compared with the original function f defined in equation (2.1), a new function $f_1^{(\sim i)}$ could be obtained, which has $n-1$ input variables.

$$f_1^{(\sim i)}(x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n) = y|_{x_i=X_i} = f(x_1, x_2, \dots, x_{i-1}, X_i, x_{i+1}, \dots, x_n) \tag{3.1}$$

Equation (2.22) from URQ could be used to calculate the variance of this new function $f_1^{(\sim i)}$.

$$\begin{aligned} V_{x_{\sim i}}\left(y|_{x_i=X_i}\right) = V\left(f_1^{(\sim i)}\right) = \sum_{\substack{p=1 \\ p\neq i}}^{n} \Bigg\{ & W_p^{+}\left[\frac{f_1^{(\sim i)}(\boldsymbol{x}_{\boldsymbol{p}}^{+(\sim i)}) - f_1^{(\sim i)}\left(\boldsymbol{\mu}_{\boldsymbol{x}}^{(\sim i)}\right)}{h_p^{+}}\right]^2 + \\ & W_p^{-}\left[\frac{f_1^{(\sim i)}(\boldsymbol{x}_{\boldsymbol{p}}^{-(\sim i)}) - f_1^{(\sim i)}\left(\boldsymbol{\mu}_{\boldsymbol{x}}^{(\sim i)}\right)}{h_p^{-}}\right]^2 + \\ & W_p^{\pm}\frac{\left[f_1^{(\sim i)}(\boldsymbol{x}_{\boldsymbol{p}}^{+(\sim i)}) - f_1^{(\sim i)}\left(\boldsymbol{\mu}_{\boldsymbol{x}}^{(\sim i)}\right)\right]\left[f_1^{(\sim i)}(\boldsymbol{x}_{\boldsymbol{p}}^{-(\sim i)}) - f_1^{(\sim i)}\left(\boldsymbol{\mu}_{\boldsymbol{x}}^{(\sim i)}\right)\right]}{h_p^{+}h_p^{-}} \Bigg\}, \end{aligned} \tag{3.2}$$

where W_p^{+}, W_p^{-}, $W_p^{\pm}$, h_p^{+}, and h_p^{-} are defined by the original URQ method (as shown in Table 1), while $\boldsymbol{\mu}_{\boldsymbol{x}}^{(\sim i)}$, $\boldsymbol{x}_{\boldsymbol{p}}^{+(\sim i)}$, and $\boldsymbol{x}_{\boldsymbol{p}}^{-(\sim i)}$ are defined by removing the i_{th} variable from the original input vector, as shown in equation (3.3), (3.4), and (3.5) respectively.

$$\boldsymbol{\mu}_{\boldsymbol{x}}^{(\sim i)} = [\mu_{x_1}, \mu_{x_2}, \dots, \mu_{x_{i-1}}, \mu_{x_{i+1}}, \dots, \mu_{x_n}] \tag{3.3}$$

$$\boldsymbol{x}_{\boldsymbol{p}}^{+(\sim i)} = [\mu_{x_1}, \mu_{x_2}, \dots, \mu_{x_p} + h_p^{+}\sigma_{x_p}, \dots, \mu_{x_{i-1}}, \mu_{x_{i+1}}, \dots, \mu_{x_n}], \qquad p \neq i \tag{3.4}$$

$$\boldsymbol{x}_{\boldsymbol{p}}^{-(\sim i)} = [\mu_{x_1}, \mu_{x_2}, \dots, \mu_{x_p} + h_p^{-}\sigma_{x_p}, \dots, \mu_{x_{i-1}}, \mu_{x_{i+1}}, \dots, \mu_{x_n}], \qquad p \neq i \tag{3.5}$$

Recalling the definition of $f_1^{(\sim i)}$, since the value of X_i is still not specified, equation (3.2) now becomes a function of X_i, which could be defined as,

$$g_1(X_i) = V_{x_{\sim i}}\left(y|_{x_i=X_i}\right) \tag{3.6}$$

Then, the expected value of $g_1(X_i)$ can be calculated by employing Equation (2.21).

$$E_{X_i}\left(V_{x_{\sim i}}\left(\left.y\right|_{x_i=X_i}\right)\right) = E_{X_i}\left(g_1(X_i)\right)$$
$$= W_0^{(i)} g_1\left(\mu_{x_i}\right) + W_i\left[\frac{g_1(\mu_{x_i} + h_i^+\sigma_{x_i})}{h_i^+} - \frac{g_1(\mu_{x_i} + h_i^-\sigma_{x_i})}{h_i^-}\right], \tag{3.7}$$

where W_i, h_i^+, and h_i^- are defined in Table 1 and $W_0^{(i)}$ is defined by:

$$W_0^{(i)} = 1 + \frac{1}{h_i^+ h_i^-} \tag{3.8}$$

Substituting Equation (3.7) and (2.22) into Equation (2.2), the first order Sobol' index is obtained by:

$$S_i \approx \frac{V_{URQ}(y) - W_0^{(i)} g_1\left(\mu_{x_i}\right) + W_i\left[\frac{g_1(\mu_{x_i} + h_i^+\sigma_{x_i})}{h_i^+} - \frac{g_1(\mu_{x_i} + h_i^-\sigma_{x_i})}{h_i^-}\right]}{V_{URQ}(y)} \tag{3.9}$$

In this equation, $V_{URQ}(y)$ is the total variance calculated by equation (2.22).

3.2.2 Second Order Indices

Given equation (2.8), the objective is to calculate $E_{X_{i,j}}\left(V_{x_{\sim i,j}}\left(\left.y\right|_{x_i=X_i, x_j=X_j}\right)\right)$, where two variables (noted as the i_{th} and j_{th} input variables) are involved in the nested loop. Similar to the deduction of the first order index, the second order index can be obtained by:

$$S_{ij} \approx \left\{V_{URQ}(y) - W_0^{(i,j)} g_2\left(\mu_{x_i}, \mu_{x_j}\right)\right.$$
$$+ W_i\left[\frac{g_2\left(\mu_{x_i} + h_i^+\sigma_{x_i}, \mu_{x_j}\right)}{h_i^+} - \frac{g_2\left(\mu_{x_i} + h_i^-\sigma_{x_i}, \mu_{x_j}\right)}{h_i^-}\right] \tag{3.10}$$
$$\left. + W_j\left[\frac{g_2\left(\mu_{x_i}, \mu_{x_j} + h_j^+\sigma_{x_j}\right)}{h_j^+} - \frac{g_2\left(\mu_{x_i}, \mu_{x_j} + h_j^-\sigma_{x_j}\right)}{h_j^-}\right]\right\}$$
$$/V_{URQ}(y) - S_i - S_j$$

In this equation, W_i, W_j, h_i^+, h_j^+, h_i^- and h_j^- are defined in Table 1 and $W_0^{(i,j)}$ is defined by:

$$W_0^{(i,j)} = 1 + \frac{1}{h_i^+ h_i^-} + \frac{1}{h_j^+ h_j^-} \tag{3.11}$$

The function g_2 is defined as the conditional variance:

$$g_2(X_i, X_j) = V_{x_{\sim i,j}}\left(y|_{x_i=X_i, x_j=X_j}\right) = V\left(f_2^{(\sim i,j)}\right)$$
$$= \sum_{\substack{p=1 \\ p\neq i,j}}^{n} \left\{ W_p^+ \left[\frac{f_2^{(\sim i,j)}(x_p^{(\sim i,j)+}) - f_2^{(\sim i,j)}\left(\mu_x^{(\sim i,j)}\right)}{h_p^+} \right]^2 \right.$$
$$+ W_p^- \left[\frac{f_2^{(\sim i,j)}(x_p^{(\sim i,j)-}) - f_2^{(\sim i,j)}\left(\mu_x^{(\sim i,j)}\right)}{h_p^-} \right]^2 \qquad (3.12)$$
$$\left. + W_p^{\pm} \frac{\left[f_2^{(\sim i,j)}(x_p^{(\sim i,j)+}) - f_2^{(\sim i,j)}\left(\mu_x^{(\sim i,j)}\right) \right] \left[f_2^{(\sim i,j)}(x_p^{(\sim i,j)-}) - f_2^{(\sim i,j)}\left(\mu_x^{(\sim i,j)}\right) \right]}{h_p^+ h_p^-} \right\},$$

where,

$$\boldsymbol{\mu}_{\boldsymbol{x}}^{(\sim i,j)} = [\mu_{x_1}, \mu_{x_2}, \dots, \mu_{x_{i-1}}, \mu_{x_{i+1}}, \dots, \mu_{x_{j-1}}, \mu_{x_{j+1}}, \dots, \mu_{x_n}] \qquad (3.13)$$

$$\boldsymbol{x}_{\boldsymbol{p}}^{(\sim i,j)+} = \left[\mu_{x_1}, \mu_{x_2}, \dots, \mu_{x_p} + h_p^+ \sigma_{x_p}, \dots, \mu_{x_{i-1}}, \mu_{x_{i+1}}, \dots, \mu_{x_{j-1}}, \mu_{x_{j+1}}, \dots, \mu_{x_n}\right], \quad p \neq i,j \qquad (3.14)$$

$$\boldsymbol{x}_{\boldsymbol{p}}^{(\sim i,j)-} = \left[\mu_{x_1}, \mu_{x_2}, \dots, \mu_{x_p} + h_p^- \sigma_{x_p}, \dots, \mu_{x_{i-1}}, \mu_{x_{i+1}}, \dots, \mu_{x_{j-1}}, \mu_{x_{j+1}}, \dots, \mu_{x_n}\right], \quad p \neq i,j \qquad (3.15)$$

$$f_2^{(\sim i,j)}(x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_{j-1}, x_{j+1}, \dots, x_n) = y|_{x_i=X_i, x_j=X_j}$$
$$= f(x_1, x_2, \dots, x_{i-1}, X_i, x_{i+1}, \dots, x_{j-1}, X_j, x_{j+1}, \dots, x_n) \qquad (3.16)$$

3.2.3 Total Order Indices

Following equation (2.15), the total effect Sobol' index can be obtained by:

$$S_i^T \approx \frac{W_0^{(\sim i)} g_3\left(\boldsymbol{\mu}_{\boldsymbol{x}}^{(\sim i)}\right) + \sum_{\substack{p=1 \\ p\neq i}}^{n} W_p \left[\frac{g_3(\boldsymbol{x}_{\boldsymbol{p}}^{(\sim i)+})}{h_p^+} - \frac{g_3(\boldsymbol{x}_{\boldsymbol{p}}^{(\sim i)-})}{h_p^-} \right]}{V_{URQ}(y)} \qquad (3.17)$$

where $\boldsymbol{\mu}_{\boldsymbol{x}}^{(\sim i)}$, $\boldsymbol{x}_{\boldsymbol{p}}^{(\sim i)+}$, and $\boldsymbol{x}_{\boldsymbol{p}}^{(\sim i)-}$ have been defined in equation (3.3), (3.4), and (3.5), respectively. $W_0^{(\sim i)}$ is given by:

$$W_0^{(\sim i)} = W_0 - \frac{1}{h_i^+ h_i^-} \qquad (3.18)$$

In this equation, g_3 is a function of all the variables, except x_i

$$g_3(X_1, X_2, \dots X_{i-1}, X_{i+1}, \dots X_n) = V_{x_i}\left(\, y|_{\boldsymbol{x}_{\sim i}=\boldsymbol{X}_{\sim i}}\right) = V\left(f_3^{(i)}\right)$$
$$= W_i^+ \left[\frac{f_3^{(i)}(\mu_{x_i} + h_i^+\sigma_{x_i}) - f_3^{(i)}(\mu_{x_i})}{h_i^+}\right]^2 + W_i^- \left[\frac{f_3^{(i)}(\mu_{x_i} + h_i^-\sigma_{x_i}) - f_3^{(i)}(\mu_{x_i})}{h_i^-}\right]^2 \quad (3.19)$$
$$+ W_i^{\pm} \frac{\left[f_3^{(i)}(\mu_{x_i} + h_i^+\sigma_{x_i}) - f_3^{(i)}(\mu_{x_i})\right]\left[f_3^{(i)}(\mu_{x_i} + h_i^-\sigma_{x_i}) - f_3^{(i)}(\mu_{x_i})\right]}{h_i^+ h_i^-}$$

where, $f_3^{(i)}$ is defined by fixing all the variables except the i_{th},

$$f_3^{(i)}(x_i) = y|_{\boldsymbol{x}_{\sim i}=\boldsymbol{X}_{\sim i}} = f(X_1, X_2, \dots, X_{i-1}, x_i, X_{i+1}, \dots, X_n), \quad (3.20)$$

3.3 Algorithm and Computational Cost

In this section, the algorithm of the proposed method is given. Once again to avoid repetition, only option 1 from Table 2 will be discussed, as option 2 and 3 could be implemented in a similar way. The algorithm for option 4 is a straightforward application of URQ.

We start with the first order indices as presented in section 3.2.1; the overall process for calculating $E_{X_i}\left(V_{x_{\sim i}}\left(\, y|_{x_i=X_i}\right)\right)$ is illustrated in Figure 2, which could be considered as one specific realization of the general process in Figure 1.

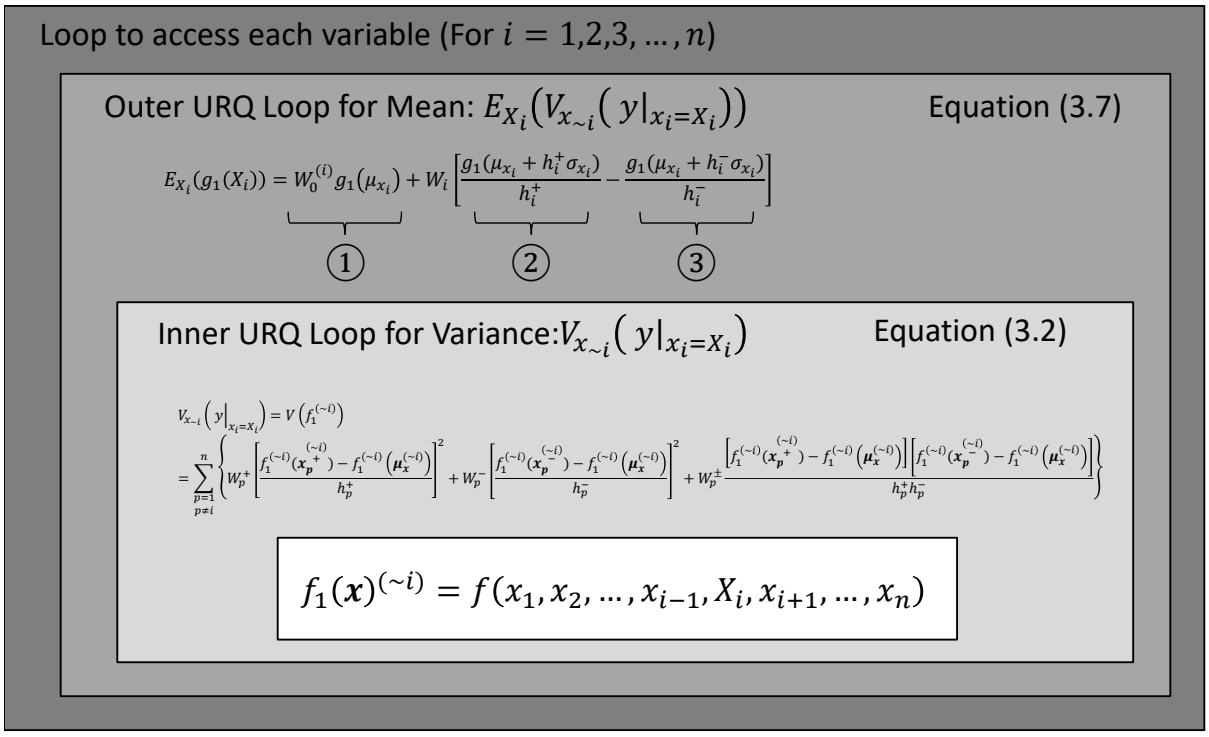

Figure 2. Illustration of the computational algorithm

To execute the outer URQ loop, X_i needs to be sampled at μ_{x_i}, $\mu_{x_i} + h_i^+\sigma_{x_i}$, and $\mu_{x_i} + h_i^-\sigma_{x_i}$, for the term ①, ②, and ③, respectively. For each realization of X_i, the inner URQ loop need to sample the other $n-1$ variables at μ_{x_p}, $\mu_{x_p} + h_p^+\sigma_{x_p}$, and $\mu_{x_p} + h_p^-\sigma_{x_p}$, where $p \neq i$. It should be noted that in

these nested loops, many samples could be reused. Therefore it is beneficial to evaluate the model at all the required points first, then to start the algebraic calculation of equations (3.2) and (3.7). The sampling points are summarized in Table 3.

Table 3. Sampling points

Sampling Point	Number of Evaluations
$(\mu_{x_1}, \mu_{x_2}, \mu_{x_3}, \dots, \mu_{x_n})$	1
$\left(\mu_{x_1}, \mu_{x_2}, \mu_{x_3}, \dots, \mu_{x_p} + h_p^+ \sigma_{x_p}, \dots, \mu_{x_n}\right); p = 1,2,3, \dots, n$	n
$\left(\mu_{x_1}, \mu_{x_2}, \mu_{x_3}, \dots, \mu_{x_p} + h_p^- \sigma_{x_p}, \dots, \mu_{x_n}\right); p = 1,2,3, \dots, n$	n
$\left(\mu_{x_1}, \mu_{x_2}, \mu_{x_3}, \dots, \mu_{x_p} + h_p^+ \sigma_{x_p}, \dots, \mu_{x_q} + h_q^+ \sigma_{x_q}, \dots, \mu_{x_n}\right); p, q = 1,2,3, \dots, n; p \neq q$	$\binom{n}{2} = n(n-1)/2$
$\left(\mu_{x_1}, \mu_{x_2}, \mu_{x_3}, \dots, \mu_{x_p} + h_p^- \sigma_{x_p}, \dots, \mu_{x_q} + h_q^- \sigma_{x_q}, \dots, \mu_{x_n}\right); p, q = 1,2,3, \dots, n; p \neq q$	$\binom{n}{2} = n(n-1)/2$
$\left(\mu_{x_1}, \mu_{x_2}, \mu_{x_3}, \dots, \mu_{x_p} + h_p^+ \sigma_{x_p}, \dots, \mu_{x_q} + h_q^- \sigma_{x_q}, \dots, \mu_{x_n}\right); p, q = 1,2,3, \dots, n; p \neq q$	$n(n-1)$

The total number of evaluations is therefore:

$$N_{total}^{Option\ 1,2,3} = 2n^2 + 1 \tag{3.21}$$

For the second order indices, the process is similar. For $E_{X_{i,j}}\left(g_2(X_i, X_j)\right)$, the outer URQ loop now has five terms. For each $(x_i, x_j) = (\mu_{x_i}, \mu_{x_j})$, $(\mu_{x_i} + h_i^+ \sigma_{x_i}, \mu_{x_j})$, $(\mu_{x_i} + h_i^- \sigma_{x_i}, \mu_{x_j})$, $(\mu_{x_i}, \mu_{x_j} + h_j^+ \sigma_{x_j})$, and $(\mu_{x_i}, \mu_{x_j} + h_j^- \sigma_{x_j})$, the inner URQ loop is calculated while the other $n-2$ variables are set to be μ_{x_p}, $\mu_{x_p} + h_p^+ \sigma_{x_p}$, and $\mu_{x_p} + h_p^- \sigma_{x_p}$ once at a time. It can be shown that the same set of samples used for the first order indices can also be used for the second order indices.

For total effect indices, the outer URQ loop now has $2n-1$ terms, in these terms, each $x_p \neq x_i$ is set as μ_{x_p}, $\mu_{x_p} + h_p^+ \sigma_{x_p}$, and $\mu_{x_p} + h_p^- \sigma_{x_p}$ one at a time. In the inner URQ loop, only x_i is sampled, at the values of μ_{x_i}, $\mu_{x_i} + h_i^+ \sigma_{x_i}$, and $\mu_{x_i} + h_i^- \sigma_{x_i}$. Similar to the second order indices, no further sampling is required for the total effect indices.

The implementations of the option 2 & 3 are similar, and the same sampling points are required. Option 4, as discussed previously, is a straightforward application of single-loop URQ for the mean of $f_*(\boldsymbol{x}_{*\boldsymbol{i}})$ defined by Eq. (2.6). This requires even less points as there is no nested structure in calculation. As $f_*(\boldsymbol{x}_{*\boldsymbol{i}})$ has $2n-1$ input variables, the number of model evaluations for option 4 is:

$$N_{total}^{Option\ 4} = 2(2n-1) + 1 = 4n - 1 \tag{3.22}$$

Because option 1, 2, and 3 require the same set of sampling points (and option 4 requires a subset of these points), for each application, it is recommended to calculate four sets of indices using all the options. A potential way to compare the accuracy of these results (without knowing the true values) is

to sum all the first order and high order indices from each option. Theoretically, for a set of indices from one option, the sum should be one, or slightly smaller if some higher order indices are neglected in implementation (in the case when the high order interactions are not significant). The option which gives the sum closest to one should be selected.

It should be mentioned that, although quadrature techniques are also used in Ref [31] and [32] for approximating integrals, the proposed approach is able to transform the sensitivity problem into an uncertainty propagation process, so that apart from the URQ method, various other efficient propagation techniques can be applied with the same process. Additionally, compared with other quadrature techniques (e.g. the grid quadrature), in which the computational cost grows exponentially with the dimensions, the adopted URQ method requires a much lower number of sampling points. Thus it is especially suitable for analysis of large scale problems.

4 Evaluation

The proposed method is applied on a number of test-cases, and the results are compared with theoretical values (where available) or with estimations from the traditional (benchmark) MCS approach. These include 18 analytical test examples and one practical design case study. The analytical examples are based on single line algebraic equations so that representative mathematical properties can be explored explicitly. By contrast, the practical design case study is based on a complex engineering model. The purpose is to demonstrate the scaling potential of the proposed method. At the end of this section, a comparison is given between the theoretical computational costs of the proposed method and other techniques.

4.1 Analytical Test Cases

4.1.1 Problem Statement

The typical mathematical properties, which may affect and limit the applicability of the proposed method include: non-linearity, non-monotonicity of the model, type of the distribution, multi-modality of the distribution, interaction of variables, and combinations of the above.

In order to take into account these properties, the test-cases are intended to form a “control variable experiment”, which involves five models and various settings of several probability distributions. Although a large amount of experiments have been used for testing and evaluation, only the most representative ones are shown in this paper. A summary of the test-cases is shown in Table 4, where the model and the corresponding distribution settings are indicated by the first and second number in the notation of the test-case. For example T1-3 indicates the combination of the first model and the third group of distribution settings.

The first three test-cases (T1-1, T1-2, and T1-3) are based on a linear model, defined by equation (4.1). In T1-1, all the inputs follow the uniform distribution $U(0,1)$. This is the simplest case which serves as a reference. In T1-2, different distributions are used, including Gaussian, triangular and uniform distributions. Here, the Gaussian distribution is noted as $N(\mu, \sigma)$, where μ and σ are the mean and standard deviation, respectively. The triangular distribution is noted as *Tri*(*ll*, *ul*, *mo*), where *ll*, *ul*, and *mo* are the lower limit, upper limit, and mode, respectively. T1-3 is designed to investigate the influence of multimodal distributions. The selected distribution is the mixed-Gaussian [53], which is represented as $MG(\mu_1, \mu_2, \sigma_1, \sigma_2, p_1, p_2)$, where μ, σ, and p are the mean, standard deviation, and portion of two Gaussian distributions, respectively. It should be noted that in T1-2 and T1-3, the parameters of the distributions are arbitrarily selected. Also, as mentioned above, various other settings have been tested beside the ones reported here, and no influence on accuracy has been found for these linear cases.

Test cases T2-1, T2-2, T2-3, T2-4, T2-5, and T2-6 are based on the non-linear model defined by equation (4.2). In this model, there is no interaction effect between the variables. The influence of non-monotonicity is explored in T2-1, T2-2, and T2-3, in which the distributions of the input variables are gradually expanded from $[0,1]$ to more non-monotonic regions. T2-4 and T2-5 are used to demonstrate the combination of non-monotonicity and multi-modality. Combination of other PDF shapes is used in T2-6.

Test cases T3-1, T3-2, and T3-3 are proposed to evaluate the method on high order indices (for interaction effects). In these three cases, equation (4.3) is used and the expected interaction effects are dependent on the ranges of input variables. In T3-1, first order effect will be the main contributor of the total variance, while in T3-2 and T3-3 more interactions are involved.

Apart from the proposed functions, experiments have also been conducted on the Sobol' G-Function specified in Eq. (4.4) (the original source is [54], while the analytical solution is presented in [23]) and Ishigami-Function in Eq. (4.5) (original source is [11], analytical solution available in [17]). These two functions are considered as classic test-cases for sensitivity analysis.

$$F_1(x_1, x_2, x_3) = x_1 + 2x_2 + 3x_3 \tag{4.1}$$

$$F_2(x_1, x_2, x_3) = x_1^2 + \sin\left(\frac{\pi}{2}x_2\right) + e^{|x_3|} \tag{4.2}$$

$$F_3(x_1, x_2, x_3) = x_1x_3 + x_1\sin\left(\frac{\pi}{2}x_2\right) + x_2e^{|x_3|} + x_1x_2x_3 \tag{4.3}$$

$$F_4(x_1, x_2, x_3) = \prod_{i=1}^{3}\frac{|4x_i - 2| + a_i}{1 + a_i} \tag{4.4}$$

$$F_5(x_1, x_2, x_3) = sin(x_1) + a \cdot sin^2(x_2) + b \cdot x_3^4 sin(x_1) \tag{4.5}$$

Table 4. Summary of the test-cases

Notation	Equation	Distribution Settings	Justification
T1-1:	Linear Model (4.1)	$x_1 \sim U(0,1)$ $x_2 \sim U(0,1)$ $x_3 \sim U(0,1)$	Simplest reference case
T1-2:	Linear Model (4.1)	$x_1 \sim N(1,1)$ $x_2 \sim Tri(0,2,1.5)$ $x_3 \sim U(0,2)$	Linearity and mixture of different distributions
T1-3:	Linear Model (4.1)	$x_1 \sim MG(0,5,0.5,0.7,0.25,0.75)$ $x_2 \sim MG(1,3,0.5,0.5,0.5,0.5)$ $x_3 \sim MG(5,7,0.7,0.5,0.75,0.25)$	Linear and multi-modality
T2-1:	Non-Linear Model (4.2)	$x_1 \sim U(0,1)$ $x_2 \sim U(0,1)$ $x_3 \sim U(0,1)$	Non-linearity and monotonicity
T2-2:	Non-Linear Model (4.2)	$x_1 \sim U(-1,1)$ $x_2 \sim U(0,2)$ $x_3 \sim U(-1,1)$	Non-linearity and non-monotonicity
T2-3:	Non-Linear Model (4.2)	$x_1 \sim U(-1,1)$ $x_2 \sim U(-2.5,2.5)$ $x_3 \sim U(-1,1)$	Non-linearity and non-monotonicity
T2-4:	Non-Linear Model (4.2)	$x_1 \sim MG(-0.5,0.5,0.2,0.2,0.75,0.25)$ $x_2 \sim MG(-2.2,2.2,0.5,0.5,0.5,0.5)$ $x_3 \sim MG(-0.5,0.5,0.2,0.2,0.25,0.75)$	Non-Linear and multi-modality
T2-5:	Non-Linear Model (4.2)	$x_1 \sim MG(-0.5,0.5,0.2,0.2,0.75,0.25)$ $x_2 \sim MG(-1.8,1.8,0.6,0.6,0.5,0.5)$ $x_3 \sim MG(-0.5,0.5,0.2,0.2,0.25,0.75)$	Non-Linear and multi-modality
T2-6:	Non-Linear Model (4.2)	$x_1 \sim N(1,1)$ $x_2 \sim Tri(-2,3,1)$ $x_3 \sim U(-2,2)$	Non-Linear and mixture of different distributions
T3-1:	Non-Linear Model (4.3)	$x_1 \sim U(0,1)$ $x_2 \sim U(0,1)$ $x_3 \sim U(0,1)$	Non-linearity and moderate interaction effect
T3-2:	Non-Linear Model (4.3)	$x_1 \sim U(-0.5,1.5)$ $x_2 \sim U(-0.5,1.5)$ $x_3 \sim U(-0.5,1.5)$	Non-linearity and strong interaction effect
T3-3:	Non-Linear Model (4.3)	$x_1 \sim U(-1,1)$ $x_2 \sim U(-1,1)$ $x_3 \sim U(-1,1)$	Non-linearity and strong interaction effect
T4-1:	Sobol' G Function (4.4)	$x_1 \sim U(0,1);\ a_1 = 0$ $x_2 \sim U(0,1);\ a_2 = 0$ $x_3 \sim U(0,1);\ a_2 = 0$	Classic Test Case
T4-2:	Sobol' G Function (4.4)	$x_1 \sim U(0,1);\ a_1 = 0$ $x_2 \sim U(0,1);\ a_2 = 3$ $x_3 \sim U(0,1);\ a_2 = 5$	Classic Test Case
T4-3:	Sobol' G Function (4.4)	$x_1 \sim U(0,1);\ a_1 = 10$ $x_2 \sim U(0,1);\ a_2 = 30$ $x_3 \sim U(0,1);\ a_2 = 50$	Classic Test Case
T5-1:	Ishigami Function (4.5)	$x_1 \sim U(0,1);\ a = 7$ $x_2 \sim U(0,1);\ b = 0.1$ $x_3 \sim U(0,1)$	Classic Test Case: (a and b from [11])
T5-2:	Ishigami Function (4.5)	$x_1 \sim U(0,1);\ a = 7$ $x_2 \sim U(0,1);\ b = 0.05$ $x_3 \sim U(0,1)$	Classic Test Case: (a and b from[17])

T5-3:	Ishigami Function (4.5)	$x_1 \sim U(0,1)$; $a = 7$ $x_2 \sim U(0,1)$; $b = 0.01$ $x_3 \sim U(0,1)$	Classic Test Case

4.1.2 Results

For each test-case, all the four options (as summarized in Table 2) have been tested and the results are plotted in Figure 3 to 7, with comparison to the theoretical/reference values. For most of the cases, theoretical values were obtained by using the equations from [16]. Some parts of the integral were numerically solved using vectorized adaptive quadrature [55] and Simpson quadrature [56] (implemented as the "integral" and "quad" functions in MATLAB). In T1-3, T2-4, and T2-5, the theoretical values were difficult to obtain due to the distributions being used, therefore values from the MCS are used as references. The theoretical results of the adopted test-cases are obtained from relevant papers referred to above.

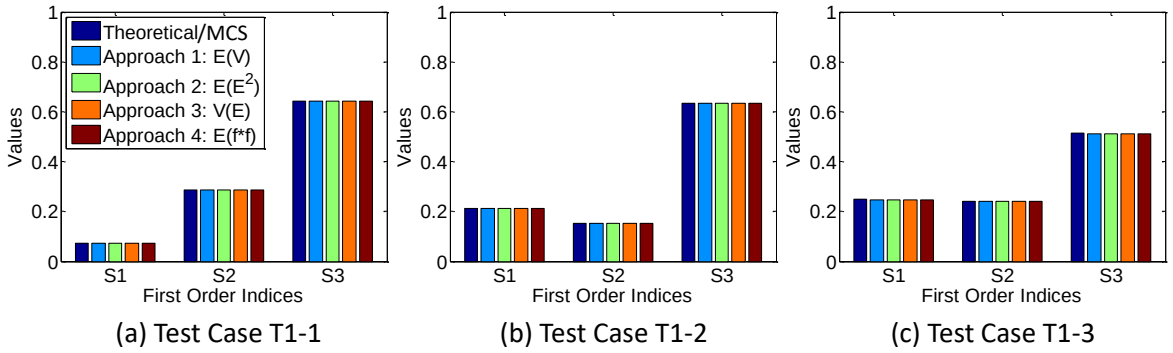

Figure 3. Results of the linear model under different settings

It can be seen from Figure 3 that the proposed method worked well on the linear model, regardless of the distributions used (including multi-modal distributions). The results from the four options are almost the same and are very close to the theoretical/reference values.

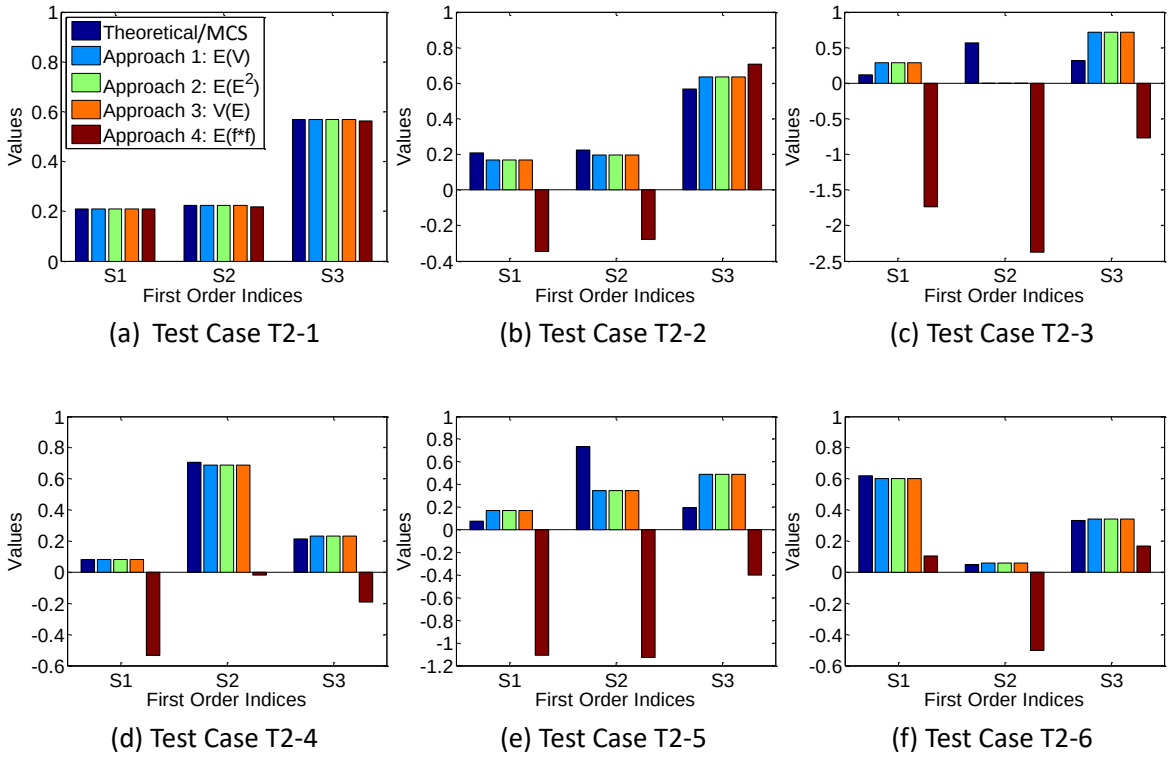

Figure 4. Results of the non-linear model 1 under different settings

Regarding the first non-linear model, the proposed method was able to deliver accurate results when all the inputs are inside the monotonic region of the function, as shown by Figure 4(a). In test-case T2-2 (Figure 4(b)), as the input distributions are expanded to the non-monotonic regions, the errors increase. It could be seen that, option 4 is the most affected one, resulting in negative values for S_1 and S_2. The results of the other three options are still relatively close to the theoretical/reference values. Figure 4 (c) shows an extreme case where S_2 is not detectable by the proposed method. Figure 4 (d) and (e) represent the results of test-case T2-4 and T2-5. It can be seen that the accuracy of the former is considerably higher, especially regarding S_2. This difference was initially unexpected, because in these two cases, the same model is used and the distribution settings are very close. The only slight difference is the distribution of x_2, which still covers roughly the same region (from -4 to 4) in both cases. A further investigation of this effect will be presented in Section 4.1.3. Figure 4 (f) shows the results of test-case T2-6, in which a number of different distributions are used. The results are close and relatively accurate, regarding the first three options.

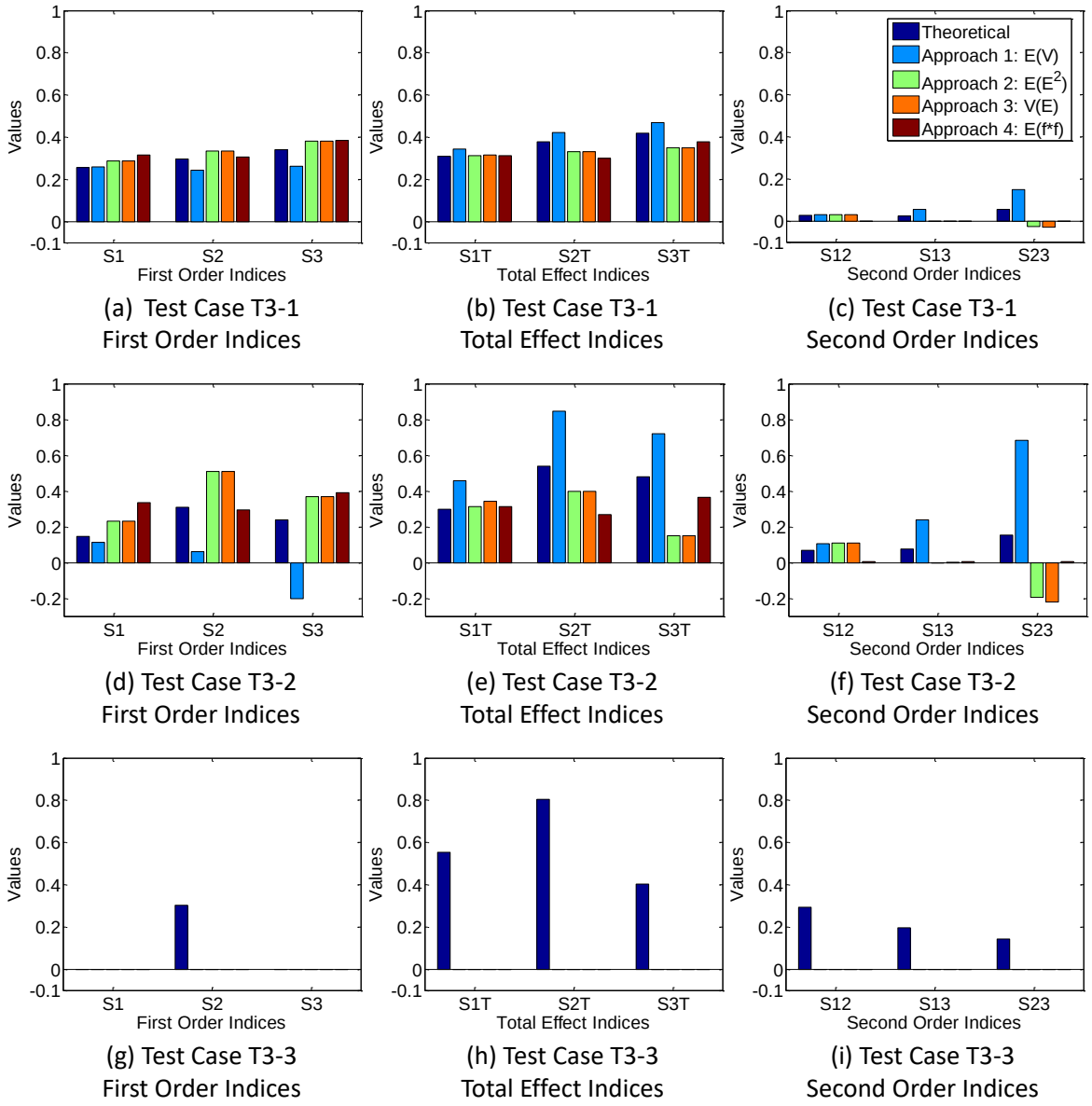

Figure 5. Results of test cases with interaction effects

The results of T3-1, T3-2 and T3-3 are shown in Figure 5. It can be seen that, the predicted first order and total effect indices are relatively accurate in test case T3-1, as plotted in Figure 5, (a) and (b). In T3-2, when the region of input is expanded to $(-0.5,1.5)$, the accuracy has been reduced (Figure 5, (d) and (e)). In both T3-1 and T3-2, the second order indices do not come very close to the benchmark (Figure 5, (c) and (f)). T3-3 is an extreme case, when the proposed method fails altogether. It should be noted that in Figure 5 (g), the reference values of S_1 and S_3 are also zeros.

The results of test-cases T4 and T5 are shown in Figure 6 and Figure 7, respectively. The quality of results is dependent on the selected coefficients in equations (4.4) and (4.5). Regarding the Sobol' G-Function, when the a_i's are all zeros, the proposed method could not produce any results as shown in Figure 6 (a), (b), and (c). As the values of a_i's are increased, the first order and total order indices

from option 1 and option 3 become more accurate (Figure 6 (d), (e), (g), and (h)). Regarding the Ishigami Function, a smaller value of *b* will lead to better predictions (Figure 7). For all the cases, the second order indices are still quite irrelevant.

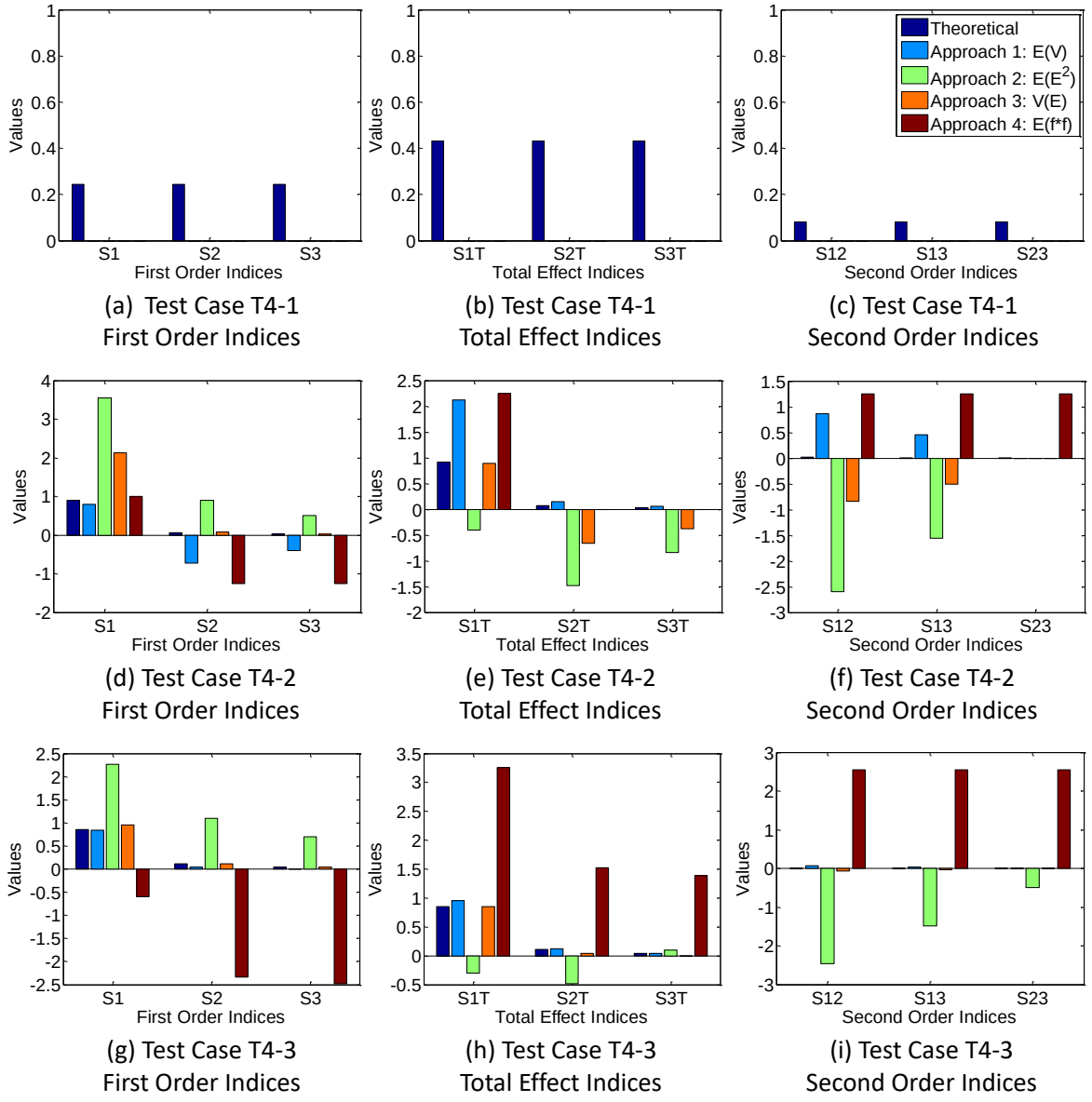

Figure 6. Results for test on the Sobol' G-function

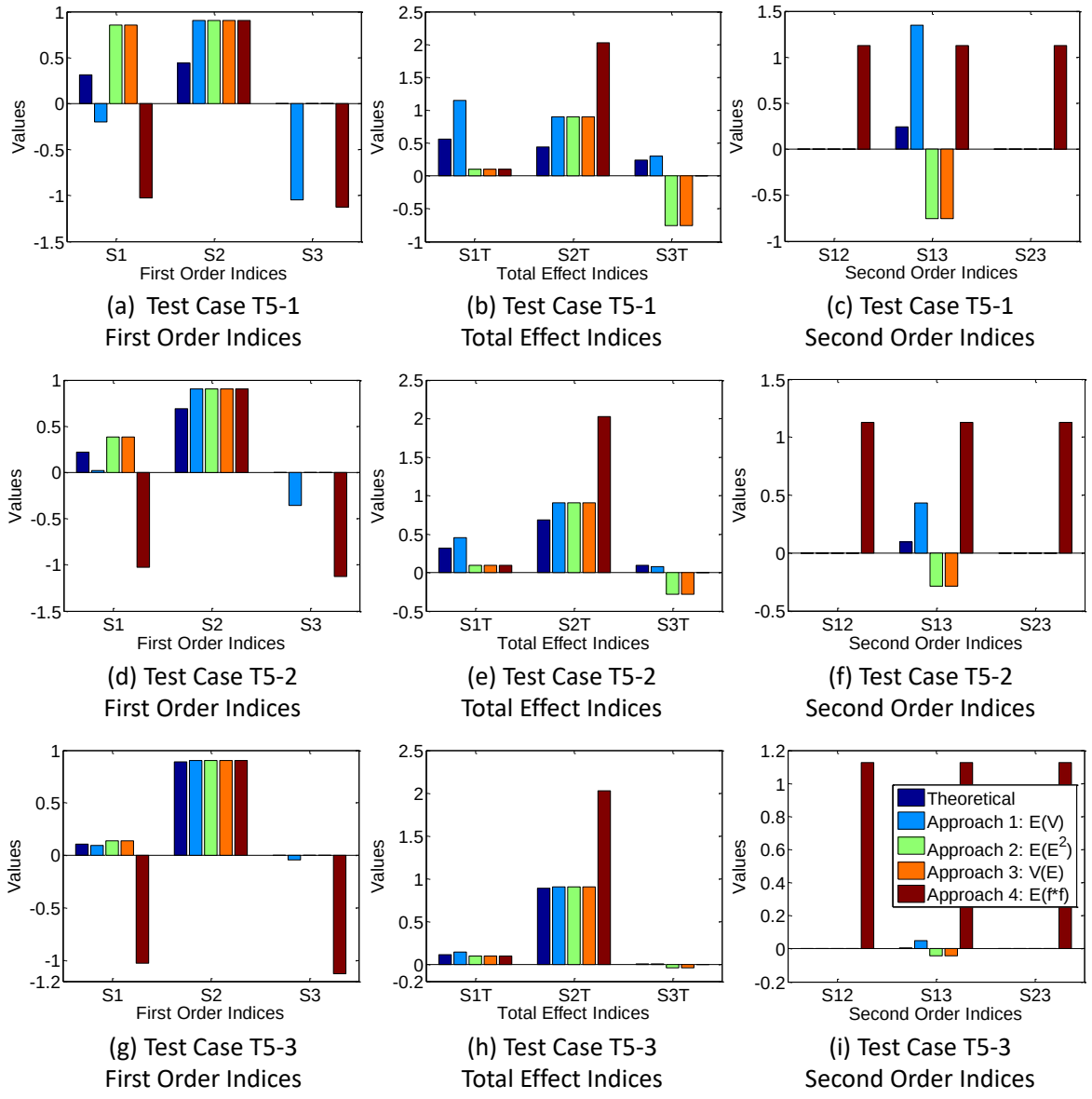

Figure 7. Results for tests on Ishigami Function

4.1.3 Discussion

The evaluation confirmed at large the effectiveness of the method, but also revealed some limitations. It was discovered that these are largely limitations of the URQ technique itself. The first limitation is due to aliasing, which could explain the low accuracy in T2-3 and T2-5.

Considering T2-3, due to the selected distribution for $x_2 \sim U(-2.5, 2.5)$, URQ will sample x_2 at the points of -1.9365, 0, and 1.9365 (corresponding to $\mu_{x_2} + h_2^- \sigma_{x_2}$, μ_{x_2}, and $\mu_{x_2} + h_2^+ \sigma_{x_2}$ respectively). At these three points, the variation of $sin\left(\frac{\pi}{2} x_2\right)$ are grossly underestimated, as illustrated in Figure 8. This directly leads to the disregard of S_2 and a reduction in the calculated total variance. The latter will also cause over-estimation of S_1 and S_3 (Figure 4 (c)).

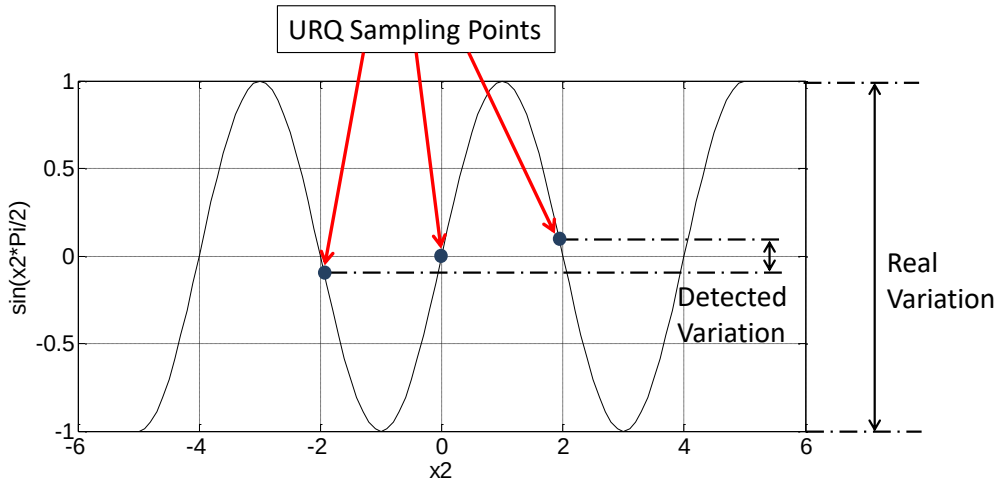

Figure 8. Under-estimated variation for $sin\left(\frac{\pi}{2}x_2\right)$

For T2-4 and T2-5, although in both cases, the distributions of x_2 have covered roughly the same region from -4 to 4, the subtle difference between the two PDFs has shifted the sampling points accordingly, as illustrated in Figure 9. Due to this shift, the accuracy of results in T2-4 (Figure 4(d)) is higher than that of T2-5 (Figure 4(e)).

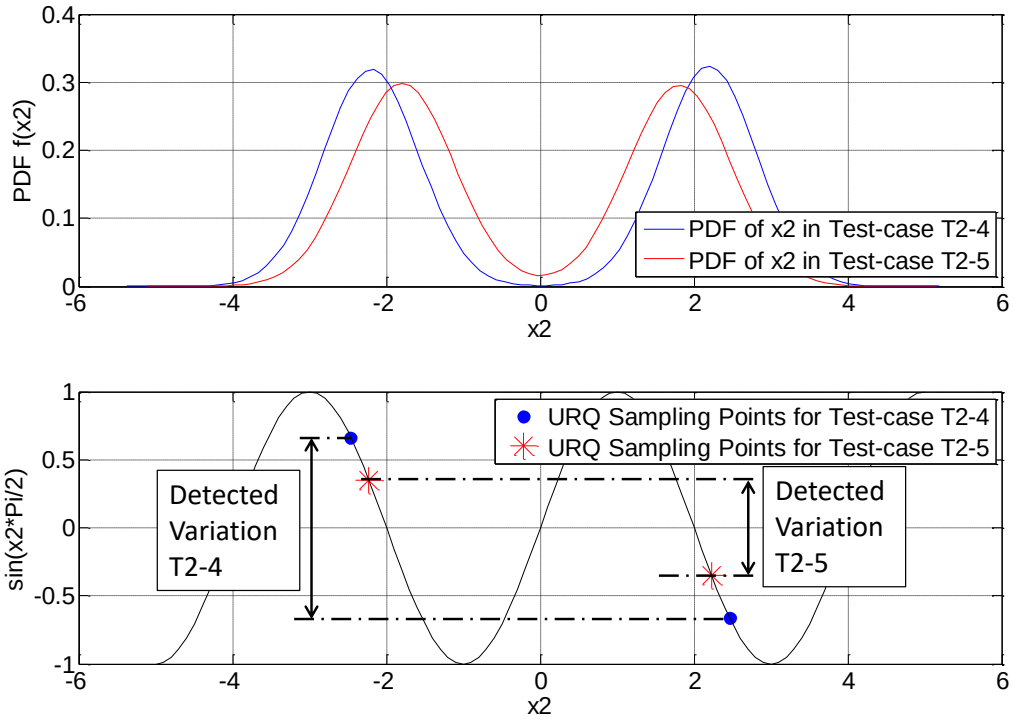

Figure 9. Comparison of the two PDFs and the relevant sampling points

The second limitation is illustrated in Figure 10, using a 2-dimensional example of $y = sin(x_1) * sin(x_2)$. The model output will be zero, on the x_1 or x_2 axis. However URQ would perturb only one variable at a time and keep the others at their mean values. If the means of the distributions happen to be zero, model evaluations at all the URQ sampling points (as indicated by the red dots in Figure 10)

will become zero. Therefore the variance will be totally disregarded, and cannot be used as denominator to calculate the indices (the estimated mean is zero as well). In test case T3-3 and T4-1 the models have similar (multiplicative) features and the distributions of inputs are symmetric to the zero point. This is the reason why no results could be given by this method for these two cases. In the Ishigami function (equation (4.5)), this effect is caused by the component $b \cdot x_3^4 sin(x_1)$, therefore by using smaller values of coefficient b, the effect will weaken, which gives more accurate predictions (as shown by T5-1, T5-2, and T5-3) .

This limitation also leads to the discrepancies regarding option 2 and especially option 4. In option 4, a new function $f_*(\boldsymbol{x}_{*\boldsymbol{i}})$ is defined by multiplying the original function $f(x)$ with itself, as shown in equation (2.6). URQ is then used to estimate the mean of $f_*(\boldsymbol{x}_{*\boldsymbol{i}})$. This new function has a multiplicative feature which URQ cannot handle (similar to the case shown in Figure 10), therefore the estimation is less accurate. For instance, the negative values in Figures 4, 6, and 7 are caused by the under-estimation of $E\left(f_*(\boldsymbol{x}_{*\boldsymbol{i}})\right)$, which leads to negative $E\left(f_*(\boldsymbol{x}_{*\boldsymbol{i}})\right) - E^2(y)$ in equation (2.5). In option 2, a similar multiplicative structure exists in the term $E_{X_i}(E^2_{x_{\sim i}}(\,y|_{x_i=X_i}))$, as defined in equation (2.3), where $E_{x_{\sim i}}(\,y|_{x_i=X_i})$ is regarded as a function of x_i, and URQ is used to predict its mean. In this case, the discrepancy is less influential when estimating $E_{X_i}(E^2_{x_{\sim i}}(\,y|_{x_i=X_i}))$, because x_i is the only variable, while in $E\left(f_*(\boldsymbol{x}_{*\boldsymbol{i}})\right)$, there are $2n-1$ variables (refer also to equation (2.7)).

As the aforementioned limitations are largely due to the URQ technique itself, it is expected that the proposed approach could be improved by adopting alternative propagation techniques. For instance, by using the propagation technique proposed by [57], additional sampling points will be used as indicated by the green dots in Figure 10. This will capture more information of the model response surface, and avoid disregarding the output variance. Another potential solution is to introduce coordinate transformations in future developments.

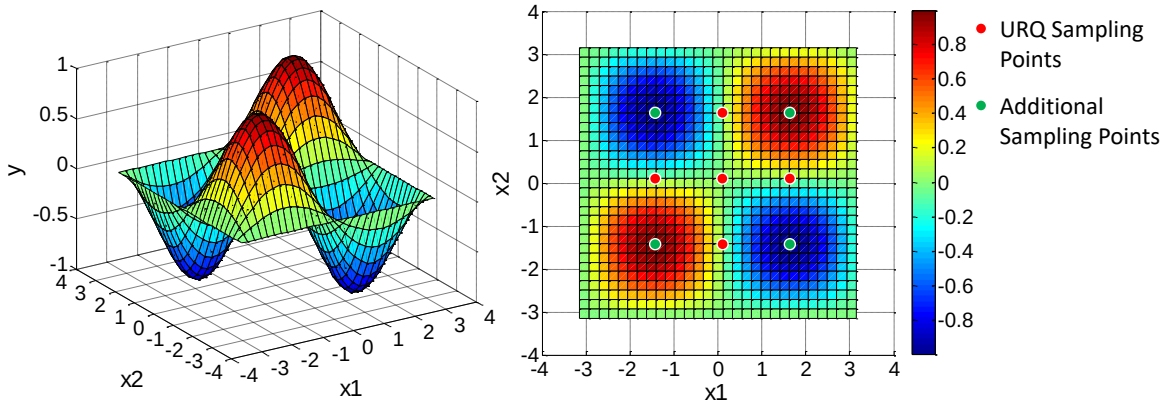

Figure 10. Example of a 2D case

4.2 Practical Test Case

4.2.1 Problem Statement

The proposed method was applied on an example of an aircraft environmental control system (ECS), adopted from Pérez-Grande and Leo [58]. In this example, thermal characteristics of an ECS, such as pressures, temperatures at different locations, and the overall entropy generation rate are calculated, given the designer-specified geometry properties. The latter include the heat exchanger dimensions, fin characteristics, and efficiencies of turbines, compressors, nozzles, and diffusers. There are 119 models in total, adopted from [58], including 59 independent (design variables and parameters) and 120 dependent variables (intermediate variables and outputs), respectively. The details of the inputs and outputs of interest are given in the appendix (Table 6 and Table 7, respectively).

While the original work of Pérez-Grande and Leo [58] was focused on ECS optimization, the current study is concerned with sensitivity analysis of their optimal solution. Uncertainties have been considered for all the 42 input variables, listed in Table 6. In the original paper [58], some of these inputs were considered as fixed parameters, while in the current test case, their variations are also taken into account. All the nominal values are adopted from the original paper and the uncertainty distributions are arbitrarily assigned, with combinations of uniform, Gaussian, triangular, and mixture-Gaussian (multi-modal) distributions. It should be emphasised that the purpose of this study is to validate the proposed method with a practical application of realistic size and complexity, rather than conduct an investigation into the merits of the particular ECS design. In a real design scenario, the distributions could be based on historical data or expert elicitation [59–61].

4.2.2 Results

Given the complexity of the test case, it was deemed infeasible to calculate the theoretical Sobol' indices. Thus the results of the proposed method are compared with reference values obtained from the traditional (benchmark) Monte-Carlo approach. As it is impractical to show the different sets of Sobol' indices for all the 19 output variables, only N_s is chosen for illustration purposes. In the original paper [58], this output was used as the objective function (to be minimised) in the design optimization.

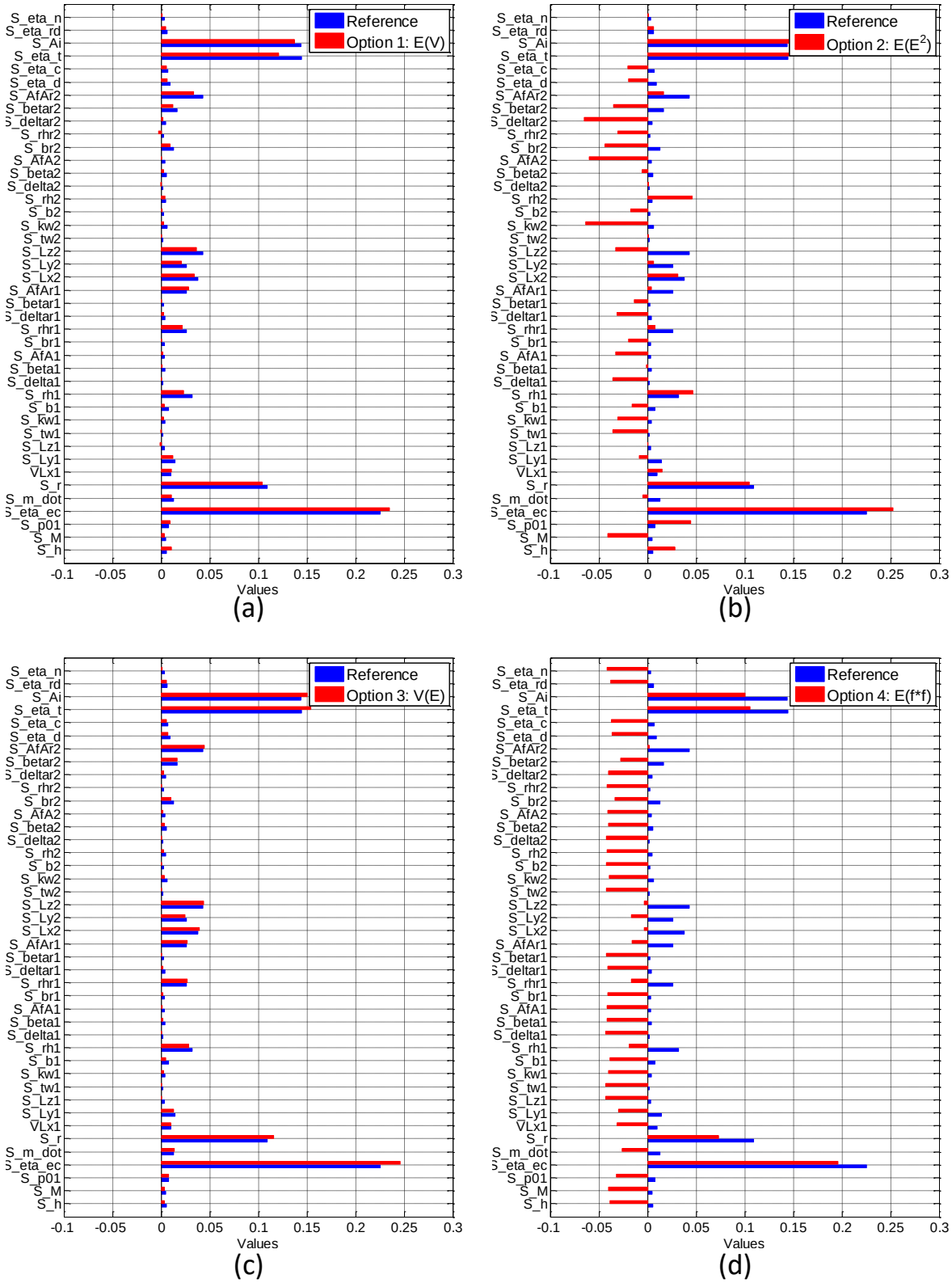

Figure 11. First order indices calculated from the four options, compared with the reference values

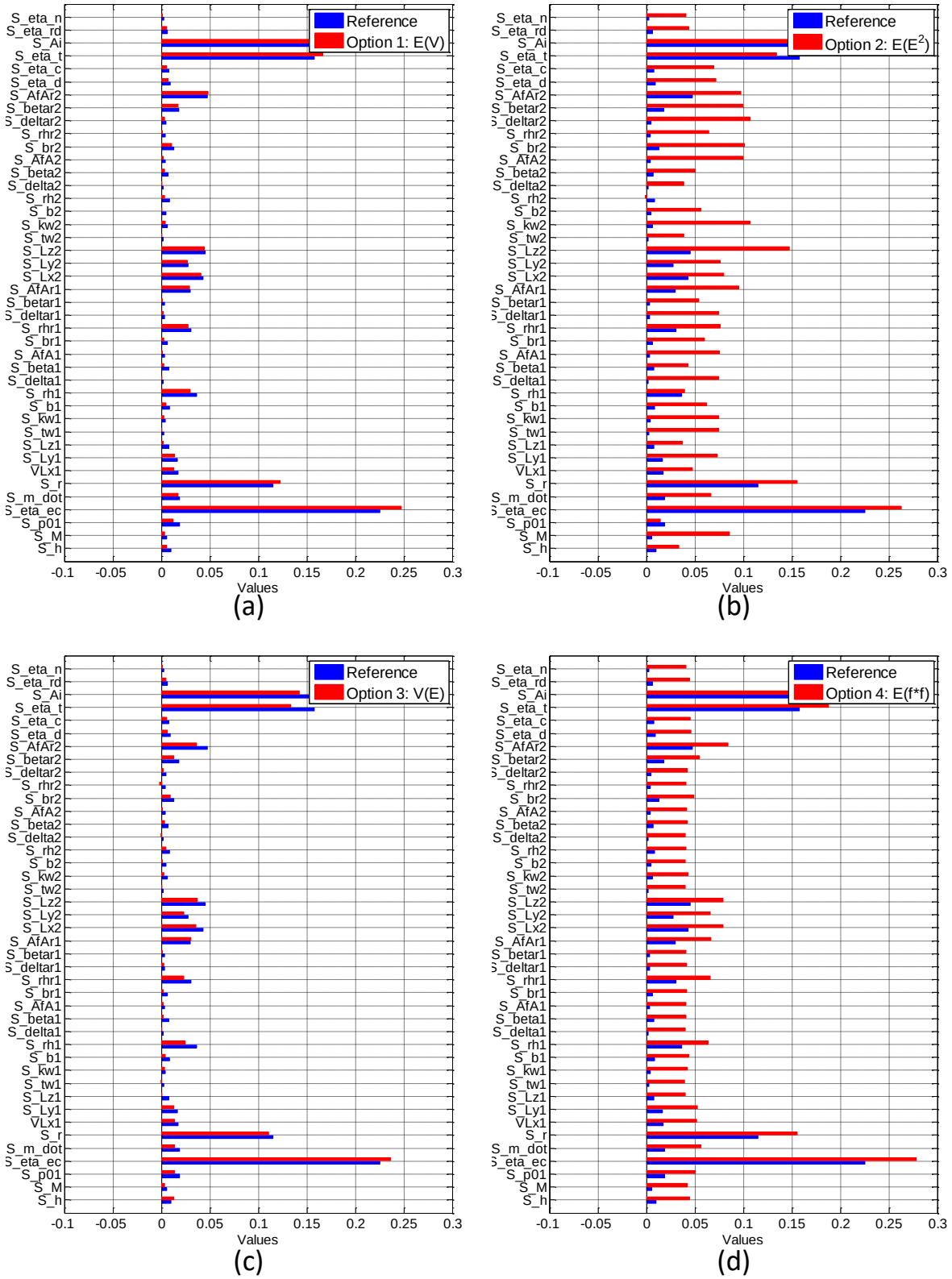

Figure 12. Total indices calculated from the four options, compared with the reference values

Shown in Figure 11 and Figure 12 are the first order and the total effect indices calculated as per the four options from Table 2 (marked as red in each subplot), compared with the reference values from

MCS (marked as blue). The interaction effects in this model turned out to be insignificant. Comparison of the results are not shown as all the values (both the calculated and reference ones) are very close to zero.

It can be seen that option 1 and 3 produced relatively good matches with the reference values, for both first order (Figure 11 (a) & (c)) and total effect indices (Figure 12 (a) & (c)). While option 2 and 4 would sometimes give negative values for S_i (Figure 11 (b) & (d)) and over-estimated values for S_i^T (Figure 12 (b) & (d))). The discrepancies are related mostly to the estimation of the insignificant factors, while the most influential ones are captured well.

4.2.3 Discussion

In Figures 11-12 (c) and (d), the discrepancies regarding option 2 and 4 are mainly due to the second limitation as discussed in Section 4.1.3 and illustrated in Figure 10.

Figures 13 and 14 illustrate the efficiency and the effectiveness of the proposed method with regard to the practical test case. The solid blue lines represent the first order and total effect indices for cooling ratio (S_r and S_r^T, respectively) calculated from the traditional (benchmark) MCS approach, at different model evaluations, $N_{total} = m(n+1)$ for S_r and $N_{total} = m(n+2)$ for S_r^T, respectively, where $m = 500,1000,2500,5000,7500,10000,12500,15000,17500,20000,30000,50000$. The red error bars in Figure 13 and 14 indicate 95% confidence intervals of the estimated values, obtained by using the boot-strap method from [62]. Regarding the first order index, it can be seen that the value start to converge after 645000 points. The straight black dash line is the value calculated by the proposed method (Option 1) at a computational cost of 3529 model evaluations. This is 183 times faster (compared with 645000 points), at a similar accuracy.

A convergence plot for the second order index $S_{r\eta_t}$ (interaction between cooling ratio and turbine efficiency), is shown in Figure 15. We have already shown in the analytical test-cases that the proposed method is not suitable for computing second order indices. In Figure 15, the result from the proposed method matches the one from MCS, because the interaction effects are very small and all the second indices are close to zero. Therefore we only claim effectiveness for the calculation of the first order indices.

Also it should be noted that, there are $\binom{n}{2} = \frac{n(n-1)}{2}$ second order indices (n in this case is 42, which leads to 861 indices). For demonstration, we only conducted a convergence study for $S_{r\eta_t}$ (interaction between cooling ratio and turbine efficiency). This leads to $N_{total} = m(n+2)$ model evaluations. Equation (4.9) should be used to obtain the computational cost of all the first order, second order, and total effect indices.

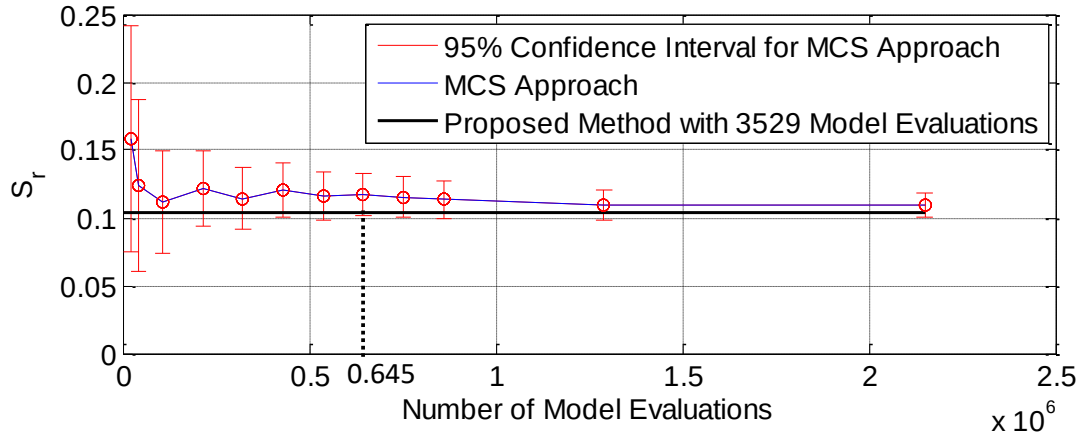

Figure 13. Convergence of the first order index of cooling ratio

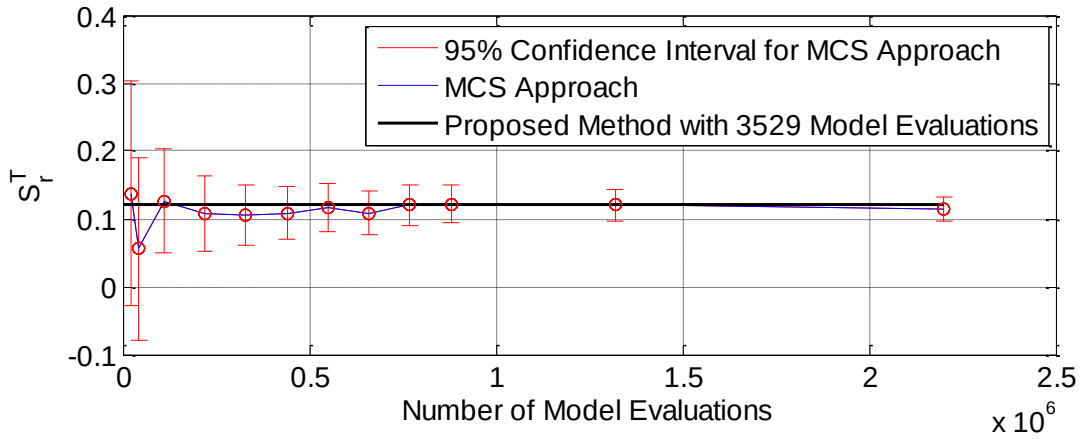

Figure 14. Convergence of the total effect index of cooling ratio

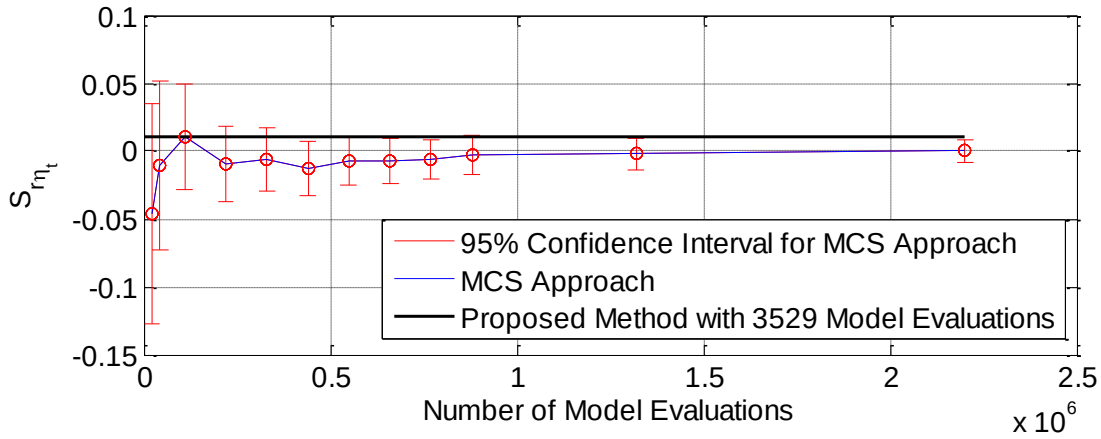

Figure 15. Convergence of the second order index of cooling ratio and turbine efficiency

4.3 Comparison of Computational Cost

Numerical comparison with other available techniques has not been conducted yet. However, the costs can be predicted analytically using the equations (4.7)-(4.18) in Table **5** and Figure 16, which demonstrated the efficiency of the proposed method.

Regarding the traditional MCS approach, equations (4.7), (4.8), and (4.9) are adopted from [29]. The cost of the improved MCS with Random Balance Design [27] is given by equation (4.10). In these equations, m is the number of samples, which will be used for a general MCS. It is chosen by the user, based on the required accuracy (the error is inversely proportional to $\sqrt{m}$) [2]. If a purely random sampling strategy is used, a representative value of m could be 1000. Some of the modified estimators [23,28] or quasi-random sampling methods can speed up the convergence, therefore smaller m could be used. In Figure 16, the computational costs against the number of uncertain input variables for all the MCS approaches are plotted as red lines, where the different settings (m values and indices included) are indicated as different markers. For instance, assume that the modified estimator from [23,28] could increase the efficiency by 50% (which is very optimistic), the cost of computing the first order indices (with $m = 500$) is indicated by the red line with asterisk markers.

The computational costs of PCE approaches are dependent on the detailed techniques for estimating the polynomial coefficients. For intrusive PCE, equation (4.11) can be used, where p is the order of truncation. This approach requires modification of the original model, which becomes prohibitive if the models are complex or black-boxes [63]. The non-intrusive PCE can be further classified as full tensor approach, sparse grid approach, and Least Square Approximation (LSA). The cost of the full tensor approach grows exponentially as shown in (4.12), while the cost of sparse grid approach can be estimated with equation (4.13), where k is the level of the grid. The LSA approach is relatively more flexible with regard to computational cost. In [64] and [34], the authors suggested that the number of samples should be more than twice or $(n - 1)$ times that of the terms in the truncated polynomial expansion. Research on the optimal sampling strategy has been conducted in [38–40,65,66], which reduces the number of samples down to the number of polynomial terms. A technique named Low-Rank tensor Approximation (LRA) was recently proposed in [37], which was reported to be more efficient than the conventional LSA. The computational cost is not included in Table **5**, as an iterative algorithm is used for sampling. The stopping criteria of this algorithm include a pre-defined tolerance for error and a number of maximum iterations. In Figure 16, the PCE related approaches are plotted as blue lines, with the markers indicating different settings. In the above equations, p is chosen to be 2 and 4, and k is chosen to be 2 (as representative values). The cost of sparse grid almost overlaps with the proposed method.

The computational cost of Lamboni method for total effect indices is given by equation (4.17), where m is the sampling of a Latin Hyper Cube and q is the order of quadrature. In the plot, m is assumed to

be 250 and q is assumed to be 2, indicated by the green line. The cost of the grid quadrature method used by Kucherenko et al [32] is not compared here, because the quantities being calculated are not equivalent. However a higher cost can be expected as the cost of grid quadrature grows exponentially.

The cost of FAST is given by (4.18), where M is the order of interference (the value is set as 4) and ω_{max} is the maximum frequency in the Fourier Transformation, the set of frequencies are adopted from [67]. In Figure 16, the cost is indicated by the purple line. The cost of extended-FAST (which is not plotted here) will be higher due to re-samplings.

The Bayesian approach is not included in the table, as no explicit estimation of the computational cost has been provided in the original paper [33]. However an example was given, in which a 40-dimenssional problem was solved, using 101 model evaluations.

In general, the computational cost of the proposed method is comparable to the RBD [27], Bayesian approach [33], and PCE using LSA [40,65], low-rank tensor [37] and sparse grid [68]. As the exact computational costs of these methods are also dependent on the required accuracy, a further comparison is planned for future work.

Table 5. Computational cost of various method

Method		Computational Cost	Equation
Proposed Method		$N_{total} = \begin{cases} 2n^2+1 \text{ (Option 1,2 \& 3)} \\ 4n-1 \text{ (Option 4)} \end{cases}$	(4.6)
MCS for First Order Indices Only [29]		$N_{total} = m(n+1)$	(4.7)
MCS for First Order and Total Effect Indices [29]		$N_{total} = m(n+2)$	(4.8)
MCS for First Order, Second Order, and Total Effect Indices [29]		$N_{total} = m(n+2+\binom{n}{2})$	(4.9)
MCS using Random Balance Design [27]		$N_{total} = 2m$	(4.10)
Intrusive PCE for All the Indices [69]		$N_{total} = \frac{(n+p)!}{n!\,p!}$	(4.11)
Non- Intrusive PCE (using full tensor quadrature) for All the Indices [64]		$N_{total} = (p+1)^n$	(4.12)
Non- Intrusive PCE (using sparse grid) for All the Indices [68]		$N_{total} \sim 2^k n^k / k!$	(4.13)
Non- Intrusive PCE (using least square approximation) for All the Indices	With optimal sampling proposed in [40,65]	$N_{total} = \frac{(n+p)!}{n!\,p!}$	(4.14)
	Recommended in [64] with random sampling	$N_{total} = 2\frac{(n+p)!}{n!\,p!}$	(4.15)
	Recommended in [34] with random sampling	$N_{total} = (n-1)\frac{(n+p)!}{n!\,p!}$	(4.16)
Lamboni Method [31]		$N_{total} = m(nq+1)$	(4.17)
Classic FAST [21]		$N_{total} = 2M\omega_{max}(n)+1$	(4.18)

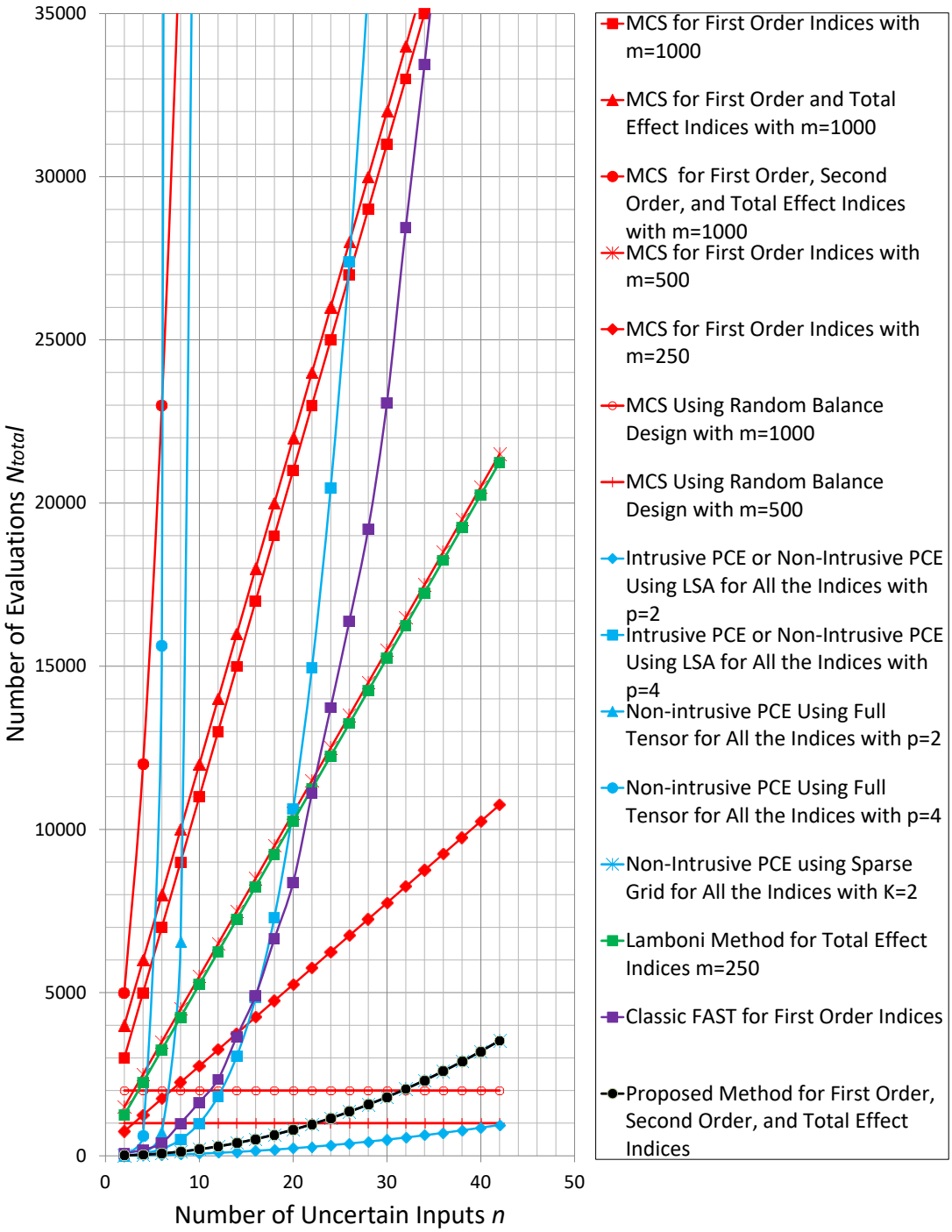

Figure 16. Number of model evaluations required

5 Conclusions

Presented in this paper is a method for efficient variance-based sensitivity analysis. The main contribution can be seen in two aspects. The first aspect includes specific formulations allowing to integrate URQ with variance-based sensitivity analysis, which provides an alternative way of calculating Sobol' indices, with considerable efficiency and effectiveness. The second aspect is

regarding the general approach to transforming a sensitivity analysis problem into an uncertainty propagation one. As extensive research has been conducted for the latter, existing propagation techniques (not only the URQ) can be exploited to compute the Sobol' indices for further reduction of the computational cost.

As demonstrated through the numerical experiments, the proposed method is very accurate in all linear or monotonic cases, under various probability distributions. For most non-linear and non-monotonic conditions, this method still provides fairly accurate estimations of the first order and total effect indices. In general, option 1 has the highest accuracy and robustness, followed by option 3 and option 2. Option 4 is the least robust to non-linearity.

In general, the particular strength is the low computational cost for large scale problems, compared with conventional techniques (such as MCS or FAST). In practical applications, the proposed method can be used as a first assessment to identify the most influential factors, which will reduce the dimensionality of the original problem. If needed, further sensitivity analysis with more accurate, but computationally expensive methods (such as MCS or FAST) can be conducted, within the reduced set of the most influential factors. Several other competitive approaches exist, including the RBD-based MCS [27], Bayesian approach [33], and PCE method (using sparse grid [68], LSA [40,65], and LRA [37]). Considering the cases reported in literature [33] and the comparison in Figure 16, it can be concluded that the Bayesian approach and PCE with LSA are currently the methods with the lowest computational cost.

The current limitations of the proposed method with regard to non-linearity are mainly due to the limitations of the URQ method used. Interaction effects in the computational models will cause a decrease in the accuracy. The higher order indices are not yet usable and their computation needs further improvement. Certain combinations of non-linearity, non- monotonicity, and shapes of the distributions will restrict the application of the current method. In particular:

- Periodicity of the model plus input distributions with spans larger than half of the model period may cause aliasing.
- Multiplicative models with constant output values along the axes, plus the sampling points corresponding to the input distributions happen to be on the axes, will lead to disregard of the variance.

Future work will focus on three tasks: the first is to further develop the specific method implemented with URQ. This will involve further investigation of the numerical errors and more rigorous comparisons with the aforementioned competitive approaches. The second is to implement the general approach with other approximation techniques. Some of the candidates may include: the unscented transformation [70], Gaussian Quadrature [71], and generalised Taguchi method [57]. These implementations will also be compared with current techniques for sensitivity analysis. The third task

is to organise these specific techniques as a comprehensive system to perform sensitivity analysis, adapting to different mathematical characteristics of the computational model under study.

References

[1] Saltelli, A., Tarantola, S., Campolongo, F., and Ratto, M., *Sensitivity analysis in practice : a guide to assessing scientific models*, Chichester: John Wiley & Sons, 2004.

[2] Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S., *Global Sensitivity Analysis: The Primer*, Chichester: John Wiley & Sons, 2008.

[3] Helton, J. C., Johnson, J. D., Sallaberry, C. J., and Storlie, C. B., "Survey of sampling-based methods for uncertainty and sensitivity analysis," *Reliability Engineering & System Safety*, vol. 91, Oct. 2006, pp. 1175–1209.

[4] Iooss, B., and Lemaître, P., "A Review on Global Sensitivity Analysis Methods," Springer US, 2015, pp. 101–122.

[5] Cukier, R. I., "Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory," *The Journal of Chemical Physics*, vol. 59, Sep. 1973, p. 3873.

[6] Cukier, R. I., Schaibly, J. H., and Shuler, K. E., "Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. III. Analysis of the approximations," *The Journal of Chemical Physics*, vol. 63, Aug. 1975, pp. 1140–1149.

[7] Cukier, R. I., Levine, H. B., and Shuler, K. E., "Nonlinear sensitivity analysis of multiparameter model systems," *Journal of Computational Physics*, vol. 26, Jan. 1978, pp. 1–42.

[8] Schaibly, J. H., and Shuler, K. E., "Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. II Applications," *The Journal of Chemical Physics*, vol. 59, Oct. 1973, pp. 3879–3888.

[9] Hora, S. C., and Iman, R. L., "Comparison of Maximus/Bounding and Bayes/Monte Carlo for fault tree uncertainty analysis (SAND-85-2839)," *Sandia National Laboratories Report*, 1986.

[10] Iman, R. L., and Hora, S. C., "A Robust Measure of Uncertainty Importance for Use in Fault Tree System Analysis," *Risk Analysis*, vol. 10, Sep. 1990, pp. 401–406.

[11] Ishigami, T., and Homma, T., "An importance quantification technique in uncertainty analysis for computer models," *Proceedings. First International Symposium on Uncertainty Modeling and Analysis*, IEEE Comput. Soc. Press, 1990, pp. 398–403.

[12] Saltelli, A., and Homma, T., "Sensitivity analysis for model output," *Computational Statistics & Data Analysis*, vol. 13, Jan. 1992, pp. 73–94.

[13] Saltelli, A., Andres, T. H., and Homma, T., "Sensitivity analysis of model output: An investigation of new techniques," *Computational Statistics & Data Analysis*, vol. 15, 1993, pp. 211–238.

[14] Jansen, M. J., Rossing, W. A., and Daamen, R. A., "Monte carlo estimation of uncertainty contributions from several independent multivariate sources," *Predictability and Nonlinear Modelling*, J. Grasman and G. van Straten, eds., Dordrecht: Springer Netherlands, 1994, pp. 334–343.

[15] Sobol', I. M., "On sensitivity estimation for nonlinear mathematical models (in Russian)," *Matematicheskoe Modelirovanie*, vol. 2, 1990, pp. 112–118.

[16] Sobol', I. M., "Sensitivity analysis for nonlinear mathematical models," *Mathematical Modeling and Computational Experiment*, vol. 1, 1993, pp. 407–414.

[17] Sobol', I. M., and Levitan, Y. L., "On the use of variance reducing multipliers in Monte Carlo computations of a global sensitivity index," *Computer Physics Communications*, vol. 117, Mar. 1999, pp. 52–61.

[18] Sobol′, I. M., “Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates,” *Mathematics and Computers in Simulation*, vol. 55, Feb. 2001, pp. 271–280.

[19] Sobol’, I. M., “Theorems and examples on high dimensional model representation,” *Reliability Engineering & System Safety*, vol. 79, 2003, pp. 187–193.

[20] Gamboa, F., Janon, A., Klein, T., Lagnoux, A., and Prieur, C., “Statistical inference for Sobol pick-freeze Monte Carlo method,” *Statistics*, vol. 50, Jul. 2016, pp. 881–902.

[21] Saltelli, A., Tarantola, S., and Chan, K. P.-S., “A Quantitative Model-Independent Method for Global Sensitivity Analysis of Model Output,” *Technometrics*, vol. 41, Feb. 1999, pp. 39–56.

[22] Homma, T., and Saltelli, A., “Importance measures in global sensitivity analysis of nonlinear models,” *Reliability Engineering & System Safety*, vol. 52, Apr. 1996, pp. 1–17.

[23] Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., and Tarantola, S., “Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index,” *Computer Physics Communications*, vol. 181, Feb. 2010, pp. 259–270.

[24] Helton, J. C., and Davis, F. J., “Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems,” *Reliability Engineering & System Safety*, vol. 81, Jul. 2003, pp. 23–69.

[25] Satterthwaite, F. E., “Random Balance Experimentation,” *Technometrics*, vol. 1, May 1959, pp. 111–137.

[26] Tarantola, S., Gatelli, D., and Mara, T. A., “Random balance designs for the estimation of first order global sensitivity indices,” *Reliability Engineering & System Safety*, vol. 91, Jun. 2006, pp. 717–727.

[27] Mara, T. A., and Rakoto Joseph, O., “Comparison of some efficient methods to evaluate the main effect of computer model factors,” *Journal of Statistical Computation and Simulation*, vol. 78, Feb. 2008, pp. 167–178.

[28] Jansen, M. J. W., “Analysis of variance designs for model output,” *Computer Physics Communications*, vol. 117, Mar. 1999, pp. 35–43.

[29] Saltelli, A., “Making best use of model evaluations to compute sensitivity indices,” *Computer Physics Communications*, vol. 145, 2002, pp. 280–297.

[30] Sobol’, I. M., Tarantola, S., Gatelli, D., Kucherenko, S. S., and Mauntz, W., “Estimating the approximation error when fixing unessential factors in global sensitivity analysis,” *Reliability Engineering & System Safety*, vol. 92, Jul. 2007, pp. 957–960.

[31] Lamboni, M., “Global sensitivity analysis: an efficient numerical method for approximating the total sensitivity index,” *International Journal for Uncertainty Quantification*, vol. 6, 2016, pp. 1–17.

[32] Kucherenko, S., Klymenko, O. V., and Shah, N., “Sobol’ indices for problems defined in non-rectangular domains,” *Reliability Engineering & System Safety*, vol. 167, 2017, pp. 218–231.

[33] Oakley, J. E., and O’Hagan, A., “Probabilistic sensitivity analysis of complex models: a Bayesian approach,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 66, Aug. 2004, pp. 751–769.

[34] Sudret, B., “Global sensitivity analysis using polynomial chaos expansions,” *Reliability Engineering & System Safety*, vol. 93, Jul. 2008, pp. 964–979.

[35] Crestaux, T., Le Maître, O., and Martinez, J. M., “Polynomial chaos expansion for sensitivity analysis,” *Reliability Engineering and System Safety*, vol. 94, Jul. 2009, pp. 1161–1172.

[36] Blatman, G., and Sudret, B., "Efficient computation of global sensitivity indices using sparse polynomial chaos expansions," *Reliability Engineering & System Safety*, vol. 95, Nov. 2010, pp. 1216–1229.

[37] Konakli, K., and Sudret, B., "Global sensitivity analysis using low-rank tensor approximations," *Reliability Engineering & System Safety*, vol. 156, Dec. 2016, pp. 64–83.

[38] Ghisu, T., and Shahpar, S., "Toward Affordable Uncertainty Quantification for Industrial Problems: Part II — Turbomachinery Application," *Volume 2C: Turbomachinery*, ASME, 2017, p. V02CT47A021.

[39] Ghisu, T., and Shahpar, S., "Toward Affordable Uncertainty Quantification for Industrial Problems: Part I — Theory and Validation," *Volume 2C: Turbomachinery*, ASME, 2017, p. V02CT47A019.

[40] Ghisu, T., and Shahpar, S., "Affordable uncertainty quantification for industrial problems: application to aero-engine fans," *Journal of Turbomachinery*, Jan. 2018.

[41] Padulo, M., Campobasso, M. S., and Guenov, M. D., "Novel Uncertainty Propagation Method for Robust Aerodynamic Design," *AIAA Journal*, vol. 49, 2011, pp. 530–543.

[42] Campbell, K., McKay, M. D., and Williams, B. J., "Sensitivity analysis when model outputs are functions," *Reliability Engineering & System Safety*, vol. 91, Oct. 2006, pp. 1468–1472.

[43] Lamboni, M., Makowski, D., Lehuger, S., Gabrielle, B., and Monod, H., "Multivariate global sensitivity analysis for dynamic crop models," *Field Crops Research*, vol. 113, Sep. 2009, pp. 312–320.

[44] Lamboni, M., Monod, H., and Makowski, D., "Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models," *Reliability Engineering & System Safety*, vol. 96, Apr. 2011, pp. 450–459.

[45] Gamboa, F., Janon, A., Klein, T., and Lagnoux, A., "Sensitivity indices for multivariate outputs," *Comptes Rendus Mathematique*, vol. 351, Apr. 2013, pp. 307–310.

[46] Garcia-Cabrejo, O., and Valocchi, A., "Global Sensitivity Analysis for multivariate output using Polynomial Chaos Expansion," *Reliability Engineering & System Safety*, vol. 126, Jun. 2014, pp. 25–36.

[47] Xiao, S., Lu, Z., and Xu, L., "Multivariate sensitivity analysis based on the direction of eigen space through principal component analysis," *Reliability Engineering & System Safety*, vol. 165, Sep. 2017, pp. 1–10.

[48] Xiao, S., Lu, Z., and Wang, P., "Multivariate global sensitivity analysis for dynamic models based on wavelet analysis," *Reliability Engineering & System Safety*, vol. 170, Feb. 2018, pp. 20–30.

[49] Li, L., Lu, Z., and Wu, D., "A new kind of sensitivity index for multivariate output," *Reliability Engineering & System Safety*, vol. 147, Mar. 2016, pp. 123–131.

[50] Mara, T. A., and Tarantola, S., "Variance-based sensitivity indices for models with dependent inputs," *Reliability Engineering & System Safety*, vol. 107, Nov. 2012, pp. 115–121.

[51] Xu, C., "Decoupling correlated and uncorrelated parametric uncertainty contributions for nonlinear models," *Applied Mathematical Modelling*, vol. 37, Dec. 2013, pp. 9950–9969.

[52] Kucherenko, S., Tarantola, S., and Annoni, P., "Estimation of global sensitivity indices for models with dependent variables," *Computer Physics Communications*, vol. 183, Apr. 2012, pp. 937–946.

[53] McLachlan, G. J., and Peel, D., *Finite mixture models*, New York: John Wiley & Sons, 2000.

[54] Saltelli, A., and Sobol', I. M., "Sensitivity analysis for nonlinear mathematical models: numerical experience," *Matematicheskoe Modelirovanie*, vol. 7, 1995, pp. 16–28.

[55] Shampine, L. F., "Vectorized adaptive quadrature in MATLAB," *Journal of Computational and Applied Mathematics*, vol. 211, Feb. 2008, pp. 131–140.

[56] Gander, W., and Gautschi, W., "Adaptive Quadrature—Revisited," *Bit Numerical Mathematics*, vol. 40, 2000, pp. 84–101.

[57] Seo, H. S., and Kwak, B. M., "Efficient statistical tolerance analysis for general distributions using three-point information," *International Journal of Production Research*, vol. 40, Jan. 2002, pp. 931–944.

[58] Pérez-Grande, I., and Leo, T. J., "Optimization of a commercial aircraft environmental control system," *Applied Thermal Engineering*, vol. 22, 2002, pp. 1885–1904.

[59] O'Hagan, A., and Wiley InterScience (Online service), *Uncertain judgements : eliciting experts' probabilities*, Wiley, 2006.

[60] Cooke, R. M., *Experts in Uncertainty : Opinion and Subjective Probability in Science: Opinion and Subjective Probability in Science*, New York: Oxford University Press, 1991.

[61] Ayyub, B. M., *Elicitation of expert opinions for uncertainty and risks*, Boca Raton: CRC Press, 2001.

[62] Archer, G. E. B., Saltelli, A., and Sobol, I. M., "Sensitivity measures, anova-like techniques and the use of bootstrap," *Journal of Statistical Computation and Simulation*, vol. 58, May 1997, pp. 99–120.

[63] Hosder, S., Walters, R. W., and Perez, R., "A Non-Intrusive Polynomial Chaos Method For Uncertainty Propagation in CFD Simulations," *44th AIAA Aerospace Sciences Meeting and Exhibit, Aerospace Sciences Meetings*, Reno, Nevada: American Institute of Aeronautics and Astronautics, 2006.

[64] Eldred, M., and Burkardt, J., "Comparison of Non-Intrusive Polynomial Chaos and Stochastic Collocation Methods for Uncertainty Quantification," *47th AIAA Aerospace Sciences Meeting including The New Horizons Forum and Aerospace Exposition*, Reston, Virigina: American Institute of Aeronautics and Astronautics, 2009.

[65] Seshadri, P., Narayan, A., and Mahadevan, S., "Effectively Subsampled Quadratures for Least Squares Polynomial Approximations," *SIAM/ASA Journal on Uncertainty Quantification*, vol. 5, Jan. 2017, pp. 1003–1023.

[66] Cuneo, A., Traverso, A., and Shahpar, S., "Comparative Analysis of Methodologies for Uncertainty Propagation and Quantification," *ASME Turbo Expo 2017: Turbomachinery Technical Conference and Exposition*, Charlotte: American Society of Mechanical Engineers, 2017.

[67] McRae, G. J., Tilden, J. W., and Seinfeld, J. H., "Global sensitivity analysis—a computational implementation of the Fourier Amplitude Sensitivity Test (FAST)," *Computers & Chemical Engineering*, vol. 6, Jan. 1982, pp. 15–25.

[68] Xiu, D., and Hesthaven, J. S., "High-Order Collocation Methods for Differential Equations with Random Inputs," *SIAM Journal on Scientific Computing*, vol. 27, Jan. 2005, pp. 1118–1139.

[69] Xiu, D., and Karniadakis, G. E., "The Wiener--Askey Polynomial Chaos for Stochastic Differential Equations," *SIAM Journal on Scientific Computing*, vol. 24, Jan. 2002, pp. 619–644.

[70] Steiner, G., Watzenig, D., Magele, C., and Baumgartner, U., "Statistical robust design using

the unscented transformation," *COMPEL - The international journal for computation and mathematics in electrical and electronic engineering*, vol. 24, Jun. 2005, pp. 606–619.

[71] Evans, D. H., "An Application of Numerical Integration Techniques to Statistical Tolerancing, III—General Distributions," *Technometrics*, vol. 14, Feb. 1972, pp. 23–35.

Appendix

Table 6. Input Variables and Associated Uncertainties

	Inputs	Symbol	Probability Distribution
Flight Condition	Altitude (m)	h	$N(11000, 550)$
	Mach Number	$Mach$	$U(0.78, 0.82)$
Engine	Bleed Pressure (kPa)	P_1	$U(225, 275)$
	Engine Fan/Compressor Efficiency	η_{ec}	$Tri(0.8, 0.98, 0.9)$
Overall Parameters	Conditioning Mass Flow Rate (kg/s)	$\dot{m}$	$U(0.65, 0.75)$
	Cooling Ratio	r	$U(0.315, 0.385)$
Pre-cooler	Length (m)	L_{x_1}	$U(0.09, 0.11)$
	Width (m)	L_{y_1}	$U(0.27, 0.33)$
	Height (m)	L_{z_1}	$U(0.3852, 0.4708)$
	Wall Thickness (m)	t_{W1}	$U(5.4e-4, 6.6e-4)$
	Sheet Fin Thermal Conductivity (W/(m·K))	k_{W1}	$MG(20,22,0.5,0.5,0.6,0.4)$
Pre-cooler Heat Transfer Surface (Main Stream side)	Plate spacing (m)	b_1	$U(4.69e-3, 5.73e-3)$
	Hydraulic Diameter (m)	$4r_{h1}$	$U(1.38e-3, 1.69e-3)$
	Fin Thickness (m)	δ_1	$U(0.92e-4, 1.12e-4)$
	Heat Transfer Area/Volume Between Plates (m^2/m^3)	β_1	$N(2231, 111.55)$
	Fin Area/Total Area	$(A_f/A)_1$	$MG(0.8,0.9,0.03,0.03,0.59,0.41)$
Pre-cooler Heat Transfer Surface (Ram Air side)	Plate spacing (m)	b_{1_r}	$U(11.07e-3, 13.53e-3)$
	Hydraulic Diameter (m)	$4r_{h1_r}$	$U(3.07e-3, 3.75e-3)$
	Fin Thickness (m)	δ_{1_r}	$U(0.92e-4, 1.12e-4)$
	Heat Transfer Area/Volume Between Plates (m^2/m^3)	β_{1_r}	$N(1115, 55.75)$
	Fin Area/Total Area	$(A_f/A)_{1_r}$	$MG(0.8,0.9,0.03,0.03,0.38,0.62)$
Main Heat Exchanger	Length (m)	L_{x_2}	$U(0.09,0.11)$
	Width (m)	L_{y_2}	$U(0.315, 0.385)$
	Height (m)	L_{z_2}	$U(0.2781, 0.3399)$
	Wall Thickness (m)	t_{W2}	$U(5.4e-4, 6.6e-4)$
	Sheet Fin Thermal Conductivity (W/(m·K))	k_{W2}	$MG(20,22,0.5,0.5,0.6,0.4)$
Main Heat Exchanger Heat Transfer Surface (Main Stream side)	Plate spacing (m)	b_2	$U(4.69e-3, 5.73e-3)$
	Hydraulic Diameter (m)	$4r_{h2}$	$U(1.38e-3, 1.69e-3)$
	Fin Thickness (m)	δ_2	$U(0.92e-4, 1.12e-4)$
	Heat Transfer Area/Volume Between Plates (m^2/m^3)	β_2	$N(2231, 111.55)$
	Fin Area/Total Area	$(A_f/A)_2$	$MG(0.8,0.9,0.03,0.03,0.59,0.41)$
Main Heat Exchanger Heat Transfer Surface (Ram Air side)	Plate spacing (m)	b_{2_r}	$U(11.07e-3, 13.53e-3)$
	Hydraulic Diameter (m)	$4r_{h2_r}$	$U(3.07e-3, 3.75e-3)$
	Fin Thickness (m)	δ_{2_r}	$U(0.92e-4, 1.12e-4)$
	Heat Transfer Area/Volume Between Plates (m^2/m^3)	β_{2_r}	$N(1115, 55.75)$
	Fin Area/Total Area	$(A_f/A)_{2_r}$	$MG(0.8,0.9,0.03,0.03,0.38,0.62)$
Miscellaneous	Main Stream Diffuser Efficiency	η_d	$Tri(0.85, 0.97, 0.95)$
	Air Cycle Machine Compressor Efficiency	η_c	$Tri(0.7, 0.8, 0.75)$
	Air Cycle Machine Turbine Efficiency	η_t	$Tri(0.7, 0.9, 0.8)$
	Ram Air Section (m^2)	A_i	$U(1.08e-2, 1.32e-2)$
	Ram Air Diffuser Efficiency	η_{rd}	$Tri(0.85, 0.97, 0.95)$
	Ram Air Nozzle Efficiency	η_n	$Tri(0.85, 0.97, 0.95)$

Table 7. Part of output variables

	Output	Symbol
Main stream	Bleed Temperature (K)	T_1
	Temperature at the Exit of Pre-cooler (K)	T_2
	Temperature after the Compressor (K)	T_3
	Temperature at the Exit of Main Heat Exchanger (K)	T_4
	Temperature after the Turbine (to the Cabin) (K)	T_5
	Pressure at the Exit of Pre-cooler (kPa)	P_2
	Pressure after the Compressor (kPa)	P_3
	Pressure at the Exit of Main Heat Exchanger (kPa)	P_4
	Pressure after the Turbine (to the Cabin) (kPa)	P_5
Ram Air	Temperature at the Exit of Ram Air Diffuser (K)	T_{1r}
	Temperature at the Exit of Main Heat Exchanger (K)	T_{2r}
	Temperature at the Exit of Pre-cooler (K)	T_{3r}
	Temperature at the Exit of Ram Air Nozzle (K)	T_{4r}
	Pressure at the Exit of Ram Air Diffuser (kPa)	P_{1r}
	Pressure at the Exit of Main Heat Exchanger (kPa)	P_{2r}
	Pressure at the Exit of Pre-cooler (kPa)	P_{3r}
	Pressure at the Exit of Ram Air Nozzle (kPa)	P_{4r}
Overall Performance	Entropy Generation Rate	N_S
	Total Volume of Heat Exchangers (m^3)	V_T