

Analyzing Sensitivity Measures Using Moment-Matching Technique

Mahmoud Awad

Industrial Engineering Department, American University of Sharjah
P.O. Box 26666 Sharjah, United Arab Emirates
miawad@aus.edu, +971 6515 2984

ABSTRACT

Sensitivity indices are used to rank the importance of input design variables or components by estimating the degree of uncertainty of output variable influenced by the uncertainty generated from input variables or components. With the advent of highly complex engineering simulation models that describe the relationship between input variables and output response, the need for an efficient and effective sensitivity analysis is more demanding. Traditional importance measures either requires extensive random number generations or unable to measure variables interaction effects. In this article, a generalized approach that can provide efficient and accurate global sensitivity indices is developed. The approach consists of two steps; running an orthogonal array based experiment using moment-matched levels of the input variables followed by a variance contribution analysis. The benefits of the approach are demonstrated through different real life examples.

Keywords:

Sensitivity index
Moment matching
Probabilistic output

1 Introduction

The advancement of computing technologies and numerical approaches drove a tremendous growth of usage of sophisticated system models to assist scientific investigation. As a result, system reliability models are getting more and more complicated with large number of risk events and dependencies between these events. Moreover, engineers or scientists make use of the models to perform various tasks and decision-making by interrogating the model to predict behavior of systems under different input variable settings.

In the context of probabilistic design [1] [2] [3], one is interested in studying the effect of input uncertainty (which are characterized by statistical distribution of input variables) to the variation of output variable. The sensitivity of a model output with respect to input variable is referred to as sensitivity measure of input variable. When the sensitivity measure is used to rank model inputs based on their influence on model output, it is called importance measure [4]. The application of sensitivity analysis is found in product development where sophisticated engineering computer models are eminent [5], nuclear waste [6] [7], safety [8] [9], hurricane losses [10], medical decision [11], marine ecosystem [12], and power plant maintenance decision making [13]

Once effect of input variables is determined, design improvement can be made effectively to mitigate the associated risk due to input variation. The probability distribution function assumed for each parameter then quantifies the uncertainty that is due either to lack of knowledge about the exact value of this parameter or to an actual variability of the value of the parameter [14]. The input variables could be design or process variables such as dimension, material property, or reliability of a component in a complex system. In the latter case, the focus is on system reliability variance which is widely used in

probabilistic safety and risk analysis of industrial plants such as nuclear ones. For example, risk achievement worth (RAW) measure is a common importance measure [15] defined as the ratio of the conditional system unreliability if component i is zero to the actual system unreliability. Other measures have also been introduced such as the ones found in [14] [16] [17].

In some cases, the number of variables may be too large to be managed and design improvements using all design variables are too costly. In this situation, design engineers must make decisions to act on important design variables only and neglect the less significant ones. In such cases, it is crucial to prioritize the input variables importance based on their contribution to overall output variance so that efficient design improvements can be made.

When the input-output relationship is sufficiently captured by low order polynomial models and all inputs are uniformly distributed, one can rank the importance of input variables by simply inspecting the regression coefficients. On the other hand, when the input-output relationship is highly nonlinear or the input variables follow various probability distributions, sensitivity analysis becomes a non-trivial task. Therefore, a general approach where the extension of the idea of traditional analysis of variance (ANOVA) decomposition for model interpretation is needed [18] [19].

One common way used to measure sensitivity is based on measuring the impact of varying one factor on output while the other input factors are fixed at their nominal values using the finite difference in output provoked by individual change of input [20] [1]. Such approach is referred to as one way sensitivity measures such as tornado diagram [21] [22] and spider plots [22]. In both approaches, the sensitivity analysis is evaluated as the maximum deviation of output from its nominal

value such as the mean or median and does not account for all possible output values that stem from output probability distribution. These measures are referred to as local sensitivity measures as opposed to global ones.

Borgonovo and Bischke [4] provided an overview of available local and global sensitivity measures. Examples of local measures are: Tornado diagrams, one way sensitivity functions, differentiation-based measures, and scenario decomposition.

Similarly, research is full of different global sensitivity measures obtained through different methods such as non-parametric methods, variance-based methods, moment-independent methods, and value of information-based sensitivity methods.

Differentiation methods are based on Taylor series expansion of model output where partial derivatives of model output relative to model input is used as the sensitivity measures [7]. However, the partial derivatives can't be used alone if the input variables have different units. Borgonovo and Apostolakis [23] suggested using the product of partial derivative and difference of input variable at a fixed local point to overcome such limitation. Several researchers proposed several ways to extend basic importance measures to account for interaction by including higher order derivatives of Taylor expansion [24]. Similar concept is used in reliability analysis where components importance measures are assessed based on the partial derivative of probability of system failure with respect to component probability failure provided that failures are statistically independent [25].

Local sensitivity quantification suffers from three shortcomings: first it is local since local or nominal values for input variables are used to estimate the derivative or output increase/decrease [20]. Secondly it does not capture the full interaction impact between input variables without an additional computational cost that may hinder its feasibility for models with large number of input variables such as the case in Tornado diagrams [4]. Finally it is deterministic, i.e., the occurrences of input variable values are treated uniformly ignoring shape of distribution where input variables may come from [16]. In some cases, analysts are forced to use local sensitivity measures due to lack of knowledge on model input distributions especially when dealing with new product or process.

When model input distributions are known or can be reasonably approximated, one can refer to global or probabilistic sensitivity measures where a "total effect" index that measures the total effect of input variables including all the possible synergetic terms between that variable and all other variables is considered [26] [17]. Several global sensitivity measures will be presented in section two.

The paper is organized as follows. Section two discusses available techniques such as Sobol indices and correlation ratios. The widely accepted Sobol approach will be used as a benchmark performance. Section three presents our proposal while the performance of the proposed approach is studied in section four using four examples including real engineering

examples such as oil sampling module and automotive engine joint sealing. Finally, conclusions are presented in section five.

2 Global Sensitivity Analysis

Global or probabilistic sensitivity analysis utilizes knowledge gained on model inputs' distributions to estimate its impact on the random model output variable distribution or its moments.

The most common approach of global sensitivity is the sample-based probabilistic sensitivity method using Monte Carlo (MC) approach. Typical steps to rank global importance of input variables for probabilistic model output using sampling-based sensitivity analysis are as follows:

1. Assign a probability density function (PDF) to each input variable
2. Generate samples or input matrix using a certain sampling scheme and PDF's of input variables
3. Evaluate model output to generate output distribution of response variable. When the simulation model is computationally expensive, one typically uses a meta-model to generate the outputs [27].
4. Finally, estimate the influences or relative contributions of each input variable to the output variable.

Although the previous outlined procedure is a general approach followed by almost all sensitivity measures, measures may differ in how samples are generated (step 1) or in influence estimation method (step 4).

In general global sensitivity methods can be categorized into: non-parametric methods [7] [19], variance-based [26] [26] [28] [29] [18] [30], moment-independent [8] [31] [17], and value of information-based sensitivity methods [4]. Borgonovo [1] provided a good comparison between the different uncertainty measures techniques. Non-parametric techniques measures sensitivity through input-output regression models which can be estimated using MC simulation sample of model input variables and estimation of sample output. Once input-output model is well defined, standardized regression coefficient (SCR) [7] or Pearson's product moment correlation coefficient (PEAR) can be used as non-parametric sensitivity measures. The ability of assessing influence of input variability depends on adequacy of regression model measured using model coefficient of determination R^2 which can be low under the presence of model nonlinearity or parameters interaction. Screening methods extends the estimation of partial derivatives around one point to several locations to identify least important model inputs while maintaining limited number of model evaluations [32] [33]. Selected locations are usually determined using design of experiment (DOE) technique such as model input range segmentation [34] and sequential bifurcation introduced by Bettonvil [35] and extended by Bettonvil and Kleijnen [36].

The variance based techniques such as Iman and Hora [37] and Sobol indices [30] [28] measures uncertainty based on input variables contribution to output variance and provides broader interpretation than other techniques and capable of providing insights on model structure.

Finally, moment-independent techniques provide insights on the influence that uncertain inputs have on the output distribution. Borgonovo [1] showed by example that non-informative results can be obtained when a decision maker relies on variance as the sole representative of the output uncertainty. As a result, moment-independent importance measures were proposed [8] [31] [17] where the entire output distribution is considered without specific reference to its moments. These moment independent measures are driven by the area between conditional and unconditional output distribution and hold independently of parameters correlation. For example, Borgonovo [1] proposed moment-independent sensitivity indicator δ_i estimated as shown below :

$$\delta_i = \frac{1}{2} E_{x_i} \left[\int |f_y(y) - f_{y|x_i}(y)| dy \right] \quad (1)$$

where $f_y(y)$ is the PDF of y and $f_{y|x_i}(y)$ is the conditional density function of y given x_i and the expected value represent the area enclosed between the conditional and unconditional model output densities obtained for a particular value of model input x_i . In case of model output sparsity, transformation invariance can be used to improve sensitivity measure estimation accuracy [37] [38].

Since variance based techniques requires large number of model simulations, a meta-model can be used to replace the original model to make computational cost less expensive [39] [33] [32] [40]. For example, Baltman and Sudret [41] proposed the use of sparse polynomial chaos (PC) expansions in order to build up a PC-based meta-model to be used to compute sensitivity indices. More details on variance based techniques in general and Sobol indices in specific are provided in section 2.1.

A second relevant aspect discussed by researchers is role of DOE in general and orthogonal arrays in specific in variance based sensitivity analysis model building. For example, Tissot and Prieur [12] discussed the computational cost of estimating Sobol indices in terms of number of independent input vectors MC simulations. As a result, they proposed a new approach to estimate all the first-and second-order Sobol indices by using only two input vectors. Similarly Morris et al., [42] considered the use of permuted column sampling plans, including replicated Latin hypercube sampling (LHS), to estimate first-order sensitivity coefficients. They used deterministic column permutations characterized by orthogonal array that eliminate coefficients estimation bias.

In the next section, two traditional and widely used sensitivity analysis techniques will be reviewed; Sobol indices and Mckay correlation ratios. The former is a global sensitivity index while the latter is a local sensitivity one. Although both techniques use MC simulation for estimating the indices, Sobol uses simple or quasi random sampling while Mckay uses Latin hypercube sampling (LHS). Both of these techniques will be used as comparators to our proposed technique.

2.1 Sobol Indices:

Suppose that a model $f(\mathbf{x}) = f(x_1, \dots, x_n)$ is an integral function defined in n -dimensional unit cube $K^n = \{x | 0 \leq x_i \leq 1; i = 1 \dots, n\}$, Sobol [18] shows that $f(\mathbf{x})$ can be decomposed into summands of different dimensions in the following functional ANOVA form:

$$f(x_1, \dots, x_n) = f_0 + \sum_i f_{i_1 \dots i_s}(x_{i_1}, \dots, x_{i_s}) \quad (2)$$

where the integral of every summands over any of its independent variables are zero, i.e.,

$$\int_0^1 f_{i_1 \dots i_s}(x_{i_1}, \dots, x_{i_s}) d_{i_k} = 0 \text{ if } 1 \leq k \leq s$$

Decomposition equation (2) has several interesting properties such as orthogonality between different terms as shown in equation (3) below:

$$\int_{K''} f_{i_1 \dots i_s}(x_{i_1}, \dots, x_{i_s}) f_{j_1 \dots j_t}(x_{j_1}, \dots, x_{j_t}) dx = 0 \quad (3)$$

Reader is referred to [26] for more discussion on properties and theorems stemmed from decomposition equation.

The term ANOVA is used here because it provides the same interpretation as an ANOVA model in traditional design of experiments theory. For example $f_i(x_i)$ can be viewed as the main effects, while $f_{ij}(x_i, x_j)$ may be regarded as first order interaction effects, and so on. As a matter of fact, the one-indexed term $f_i(x_i)$ can be obtained by integrating the decomposition equation over all the indices but x_i and using orthogonality property [26] to obtain:

$$\int_0^1 \dots \int_0^1 f(\mathbf{x}) \{d\mathbf{x}/dx_i\} = f_0 + f_i(x_i) \quad (4)$$

where $d\mathbf{x}/dx_i$ indicates integration over all variables except x_i . Equation (4) can be extended to the two-indexed summands $f_{ij}(x_i, x_j)$, i.e.,

$$\begin{aligned} \int_0^1 \dots \int_0^1 f(\mathbf{x}) \{d\mathbf{x}/dx_i dx_j\} \\ = f_0 + f_i(x_i) + f_j(x_j) + f_{ij}(x_i, x_j) \end{aligned} \quad (5)$$

Assume further that $f(\mathbf{x})$ is square integrable. Then, sensitivity estimates S_{i_1, \dots, i_s} can be defined as:

$$S_{i_1, \dots, i_s} = \frac{D_{i_1, \dots, i_s}}{D} \quad (6)$$

where $D = \int f^2(\mathbf{x}) d\mathbf{x} - f_0^2$ and

$$D_{i_1, \dots, i_s} = \int_0^1 \dots \int_0^1 f_{i_1 \dots i_s}^2 d_{x_i} d_{x_j}$$

The constants D and D_{i_1, \dots, i_s} are called variances. Note that:

$$D = \sum D_{i_1, \dots, i_s} \quad (7)$$

If \mathbf{x} is viewed as a random variable vector uniformly distributed over the experimental domain, then D and D_{i_1, \dots, i_s} are the variances of $f(\mathbf{x})$ and $f_{i_1, \dots, i_s}(x_{i_1}, \dots, x_{i_s})$ respectively and the ratios S_{i_1, \dots, i_k} are called global sensitivity indices.

All S_{i_1, \dots, i_s} 's are nonnegative and their sum is

$$\sum S_{i_1, \dots, i_s} = 1 \quad (8)$$

The indices S_{i_1, \dots, i_s} are sometimes called *Sobol indices* as they were proposed by Sobol [30] [18]. For a piecewise continuous function $f(\mathbf{x})$, the equality $S_{i_1, \dots, i_s} = 0$ implies that $f_{i_1, \dots, i_s} = 0$. Thus, the functional structure of $f(\mathbf{x})$ can be studied by estimating the indices.

The integral computation presented in the definition of D_{i_1, \dots, i_s} can be a challenging task. In practice, the simplest way to estimate D_{i_1, \dots, i_s} by using the MC simulation method: generate random vectors x_1, \dots, x_n which are uniformly distributed over the unit cube in K^n , then D_{i_1, \dots, i_s} can be estimated as follows:

$$D_{i_1, \dots, i_s} = \frac{1}{N} \sum_{j=1}^N f_{i_1, \dots, i_s}(x_j) \quad (9)$$

It can be shown that this estimate is consistent and possess a root N convergence rate. For a given randomly generated sample with size N , the sensitivity estimates can be estimated as the following:

$$\hat{f}_0 = \frac{1}{N} \sum_{m=1}^N f(x_m) \quad (10)$$

$$\hat{D} = \frac{1}{N} \sum_{m=1}^N f^2(x_m) - \hat{f}_0^2 \quad (11)$$

$$\hat{D}_i = \frac{1}{N} \sum_{m=1}^N f(\mathbf{u}_m, x_{im}) f(\mathbf{v}_m, x_{im}) - \hat{f}_0^2 \quad (12)$$

where x_m is a sampled point in the space K^n , \mathbf{u} and \mathbf{v} denote projections of \mathbf{x} on $K^n - 1 = K^n$ minus the variable. Hence, the term D_i is generated by summing the products of two function values: one with all sampled variables and the other with all variables re-sampled except the variable x_i .

Homma and Saltelli [26] extended this MC estimation approach to estimate the second order terms D_{ij} as shown below:

$$\hat{D}_{ij} + \hat{D}_i + \hat{D}_j = \frac{1}{N} \sum_{m=1}^N f(\mathbf{r}_m, x_{im}, x_{jm}) f(\mathbf{s}_m, x_{im}, x_{jm}) - \hat{f}_0^2 \quad (13)$$

where $\mathbf{r}_m, \mathbf{s}_m \in K^{n-2}$, the left hand side represent the total variance due to variables x_i and x_j including interaction and it can be estimated by summing the products of two function in which all the variables are re-sampled except the variable x_i and x_j . According to Sobol [18] [30], the total number of random numbers which must be generated to evaluate a complete set of $\hat{D}_{i_1, \dots, i_s}$ is a matrix of size $(N, 2n)$. The random numbers can either be generated using random sampling or some other form such as LHS as suggested by McKay et al., [19].

Note that Sobol decomposition is based on the assumption that $f(x)$ is measurable and its density function equals the product of the density function of each of the parameters x_i [43]. The latter assumption requires that the inputs are statistically independent since correlated inputs deteriorate Sobol decomposition represented by equation 2 [40]. Saltelli et al., [44] proposed variance-based measures under correlation by identifying the smallest number of factors that lead to the highest output variance reduction using importance decomposition [45] shown in equation (16) in the next section. Another shortcoming of Sobol indices is the total reliance on variance to describe uncertainty. Saltelli [28] indicated that focusing on variance as the sole measure of uncertainty is equivalent to describing the output variable distribution using second moment only which could be misleading when considering uncertainty [1]. Huang and Litzenger [46] showed that variance is sufficient to characterize variability only if model output is normal or the utility function of decision-maker is quadratic.

2.2 Correlation Ratios and Local Indices:

Let y be the output response of input vector $\mathbf{x} = \{x_i\}$, the variance of y , $\text{var}(y)$, can be decomposed into:

$$\text{Var}(y) = \text{var}\{E(y|x_i)\} + E\{\text{var}(y|x_i)\} \quad (14)$$

McKay [47] called the term $\text{var}(E(y|x_i))$ variance conditional expectation (VCE) and defined the correlation ratio CR_i as:

$$CR_i = \frac{\text{var}\{E(y|x_i)\}}{\text{Var}(y)} \quad \text{where } i = 1, 2, \dots, k \quad (15)$$

CR_i reflects the proportion of variation of outputs explained by x_i . The correlation ratios play a similar role in nonlinear regression settings to that of the usual correlation coefficient for linear relationships between the output and the input variables. Iman [45] introduced a similar measure I_i represents the expected reduction in output variance that can be achieved if uncertainty of x_i is eliminated [28].

$$I_i = \text{var}[y] - E\{\text{var}[y|x_i]\} = \text{var}\{E(y|x_i)\} \quad (16)$$

Note that unlike Sobol sensitivity indices, correlation ratios only capture the marginal effects of x_i which do not include their interactions. Correlation ratios may be quite effective in identifying prospective variables that require further investigation, but they may not accurately identify the true importance of variables or subset of variables. When first order only indices are of interest, Saltelli [28] pointed that equation 12 and 15 will provide the same rankings where both indices would be measuring the expected percentage of variance remaining if all parameters but the variable of interest were known.

The correlation ratio is similar in principle to the tornado diagrams, one way sensitivity metric proposed by Morton et al. [48], and differential importance measures (DIM) [20]. In all these techniques, one input variable is varied at a time while other variables are fixed at their nominal values. All these one way sensitivity metrics suffer from local sensitivity measures deficiencies mentioned earlier. Namely, inability to capture effect of variables interactions, and sensitivity is evaluated locally around the output nominal value only.

McKay et al., [19] suggested the calculation of correlation ratios using r -LHS where r is the number of LHS samplings. The analysis starts with generating a LHS with size of N for the k input variables. Then a new sample set with the same size N is generated by independently applying new random permutations to each of its k columns. In this way, $N \times k$ LHS is generated first and repeated r times using different permutations.

Despite its accuracy, Sobol indices require a relatively large number of simulations which is dependent on the dimension of the model of interest. McKay correlations ratios, on the other hand, require less number of simulations but lacks accuracy when interactions between input variables are present. In this paper, we propose the use of simpler sampling scheme using orthogonal array DOE followed by a simple variance reduction technique. This approach is presented in the next section and proven to achieve high accuracy with much less sample size and analysis steps.

3 Proposed Technique

The proposed technique consists of two steps: a moment-matching step where a design of experiment (DOE) is planned and conducted followed by a variance contribution analysis to calculate the sensitivity analysis indices. The two steps are discussed in a more detail below.

3.1 Moment-Matching:

In this step, a DOE is used with properly selected levels for each input variable. The theory behind this technique is explained by D'Errico and Zaino [49] for symmetrical distribution and extended by Seo and Kwak [50] to include non-symmetrical distributions as well. For a random variable x , the k th order moment about the origin of model $y = g(x)$ can be approximated using a quadratic formula with m nodes [50] as follows:

$$\begin{aligned} E\{g^k\} &= \int_{-\infty}^{\infty} [g(x)]^k f(x) dx \\ &\cong w_1 [g(\mu + \alpha_1 \sigma)]^k + w_2 [g(\mu + \alpha_2 \sigma)]^k + \dots \\ &\quad + w_m [g(\mu + \alpha_m \sigma)]^k \end{aligned} \quad (17)$$

where $f(x)$ is a density function of a random variable x with mean μ and standard deviation σ , $\{w_1, w_2, \dots, w_m\}$ is weight parameter set, and $\{\alpha_1, \alpha_2, \dots, \alpha_m\}$ is coefficient set.

If we are interested in matching the first four moments only, three nodes in the quadratic formula $m = 3$ is sufficient and parameter set $\{w_1, w_2, w_3, \alpha_1, \alpha_2, \alpha_3\}$ is determined by solving equation (18) below:

$$\begin{aligned} M_k(x) &= \int_{-\infty}^{\infty} (x - \mu)^k f(x) dx = w_1 (\alpha_1 \sigma)^k + w_2 (\alpha_2 \sigma)^k \\ &\quad + w_3 (\alpha_3 \sigma)^k \text{ for } k = 0, 1, 2, 3, 4 \end{aligned} \quad (18)$$

where $M_k(x)$ is the k th order central moment of x

The basic idea here is to replace the random input variable x with three discrete levels (l_1, l_2, l_3) by selecting a mean-distance parameters $\{\alpha_1, \alpha_2, \dots, \alpha_m\}$ where $l_i = \mu + \sigma \alpha_i$ and assigning weights to these three discrete levels such that the first four moments (mean, standard deviation, skewness and kurtosis) of the distribution of x are matched.

Without any loss of generality, let the second level be the mean μ , i.e. $\alpha_2 = 0$. When $k=4$, we end up with four equations from (18) in addition to the constraint that the weights should add to one. The latter constraint is necessary to satisfy the equality in equation (18) when $k = 0$. These five equations are enough to uniquely estimate all unknown parameters as follows:

$$\begin{aligned} w_1 + w_2 + w_3 &= 1 \\ w_1 l_1 + w_2 \mu + w_3 l_3 &= \mu \\ w_1 (l_1 - \mu)^2 + w_2 (0)^2 + w_3 (l_3 - \mu)^2 &= \sigma^2 \\ \frac{w_1 (l_1 - \mu)^3 + w_2 (0)^3 + w_3 (l_3 - \mu)^3}{\sigma^3} &= \sqrt{\beta_1} \\ \frac{w_1 (l_1 - \mu)^4 + w_2 (0)^4 + w_3 (l_3 - \mu)^4}{\sigma^4} &= \beta_2 \end{aligned} \quad (19)$$

where $\sqrt{\beta_1}$ and β_2 are skewness and kurtosis of x PDF and $\mu + \alpha_i \sigma$ is replaced with level l_i .

The above equations can be solved simultaneously to get the three levels with its weights as follows:

$$\begin{aligned} l_1 &= \mu + 0.5\sigma\sqrt{\beta_1} - 0.5\sigma\sqrt{4\beta_2 - 3\beta_1} \\ l_2 &= \mu \\ l_3 &= \mu + 0.5\sigma\sqrt{\beta_1} + 0.5\sigma\sqrt{4\beta_2 - 3\beta_1} \end{aligned}$$

$$\begin{aligned}
w_1 &= \frac{(4\beta_2 - 3\beta_1) + \sqrt{\beta_1(4\beta_2 - 3\beta_1)}}{2(4\beta_2 - 3\beta_1)(\beta_2 - \beta_1)} \\
w_2 &= \frac{\beta_2 - \beta_1 - 1}{\beta_2 - \beta_1} \\
w_3 &= \frac{(4\beta_2 - 3\beta_1) - \sqrt{\beta_1(4\beta_2 - 3\beta_1)}}{2(4\beta_2 - 3\beta_1)(\beta_2 - \beta_1)} \quad (20)
\end{aligned}$$

One obvious limitation for the equations above is $\beta_2 - \beta_1$ has to be greater than one to have a nonnegative w_2

The above process is repeated for all input variables of model response $y = g(x)$. Next, a full factorial DOE with all possible combination for all input variables is generated for which the response variable y_i is evaluated. For each treatment, there is an associated weight, $v_i = \prod_{j=1}^k w_{ij}$ where w_{ij} are the weights associated with the x_{ij} . The moments of the response, y , can be approximated directly from the N data points based on the theory of numerical integration originally discussed by Evans [51] as follows:

$$\widehat{M}_{y1} = \bar{y} = \sum_{i=1}^N v_i y_i \quad (21)$$

and

$$\widehat{M}_{yt} = \sum_{i=1}^N v_i (y_i - \widehat{M}_{y1})^t \text{ for } t \geq 2 \quad (22)$$

where \widehat{M}_{y2} is nothing but the variance of the response y .

Note that a full factorial DOE will provide balance and orthogonality between input variables. These characteristics are necessary to prevent any bias towards any of the input variables and provide equal weight of variance contribution in the next step. Another advantage of the choice of 3^k full factorial DOE is low number of experimental runs which is vital when model does not exist and experimentation is expensive or time consuming. Fractional factorial designs capable of handling variables with three levels and a minimum resolution of V can also be used to further reduce number of experiments. Resolution V guarantees that no main effects or 2-factor interactions are aliased with any other main effect or 2-factor interactions. Reader is referred to Phadke [52] or Montgomery [53] for details on fractional factorial experimental designs.

3.2 Variance Contribution

Let's set one of the input variables x_i to a fixed value using its mean, the total variance will be reduced to a new value, $V_{x \sim i}(y|x_i = \mu_{x_i})$. Similarly, reduced variances can be estimated for the other possible values of x_i and averaged out to estimate the expected reduced variance $E_{x_i}(V_{x \sim i}(y|x_i))$ which is expected to be small if x_i is important. The variance decomposition in (14) can be written as:

$$Var(y) = V_{x_i}\{E_{x \sim i}(y|x_i)\} + E_{x_i}(V_{x \sim i}(y|x_i)) \quad (23)$$

where $x \sim i$ represent all the input variables except x_i . The first term in (23) corresponds to the reduced variance due to fixing x_i over its possible values while the second term is the expected reduction of variance. The second term measures what the variance of all of the input variables other than x_i on average when we vary x_i . When the fixed factor is significant, it is intuitively expected that the reduced variance will be significant while the expected reduced variance will be small. As a result, the main effect contribution of the fixed variable on sensitivity analysis can be approximated using the following:

$$\frac{V_{x \sim i}(E_{x \sim i}(y|x_i))}{Var(y)} \quad (24)$$

The numerator can be estimated using the same way as that of the variance of the response in equation (23) and the total effect term can be estimated using the following:

$$\begin{aligned}
\frac{V_{x \sim i}(E_{x \sim i}(y|x_i))}{Var(y)} &= \frac{Var(y) - E_{x_i}(V_{x \sim i}(y|x_i))}{Var(y)} \\
&= 1 - \frac{E_{x_i}(V_{x \sim i}(y|x_i))}{Var(y)} \quad (25)
\end{aligned}$$

Note that equation (25) is very similar in nature to Sobol indices, where both \widehat{D}_i and $V_{x \sim i}\{E_{x \sim i}(y|x_i)\}$ are estimated as a function of the fractional variance that would be left if all factors but x_i were fixed. The variance estimation as proposed by moment-matching guarantees that the variance is estimated by averaging all the different level values of $x \sim i$.

Since the DOE used in variance estimation is full factorial, i.e. balanced and orthogonal, the sum of squares partitioning can also be used instead of variance decomposition as shown below:

$$SS_{total} = \sum_i SS_{x_i} + \sum_i \sum_{j \neq i} SS_{x_i x_j} \quad (26)$$

where SS_{total} is the total sum of squares and SS_{x_i} is the sum of squares of variable x_i and $SS_{x_i x_j}$ is the sum of square of the interaction between x_i and x_j variables. Reader is referred to [53] for more details on sum of squares calculations. As a result, the sensitivity index of variable x_i can be estimated as follows:

$$\widehat{D}_i = \frac{SS_{x_i}}{SS_{total}} \quad (27)$$

Similarly, the interaction sensitivity index can be estimated as

$$\widehat{D}_{ij} = \frac{SS_{x_i x_j}}{SS_{total}} \quad (28)$$

Keep in mind that the sum of sensitivity indices of all variables and their interaction will not add up to one if the sum of square of error term SS_{error} is not zero. This should not be an issue since SS_{error} should be small relative to other terms if model used is accurate enough.

4 Numerical Examples

In this section, the proposed technique will be illustrated using four examples. The performance is compared to the two current methods discussed in section two.

4.1 Example 1: Beam Deflection

The first example discusses the maximum deflection of a statically indeterminate beam studied by Thoft-Christesen and Baker [54] and re-discussed by Seo et al., [50]. The maximum deflection of a 5 m beam can be written as:

$$g(P, E, I) = EI - 78.125P \quad (29)$$

where P is the concentrated force measured in KN, E is modulus of elasticity measured in KN/m^2 , and I is the moment of inertia measured in m^2 . The description of these input random variables is provided in table 1. Moreover, the levels and weights computed using moment matching in equations 16. Note that the weight w_1, w_2, w_3 for a variable following normal or uniform distribution is $\frac{1}{6}, \frac{4}{6}$, and $\frac{1}{6}$ respectively.

Table 1
Beam deflection design variables

Variable	P	E	I
Distribution	Gumble	Normal	Normal
Mean	4	2.0E+07	1.0E-04
Stdev	1	5.0E+06	2.0E-05
Skewness	1.14	0	0
Kurtosis	2.4	3	3
L_1	2.2679	1.133E+7	0.653E-04
L_2	4	2.0E+07	1.0E-04
L_3	5.7321	2.866E+07	1.3464E-04
w_1	0.1667	0.1667	0.1667
w_2	0.6667	0.6667	0.6667
w_3	0.1667	0.1667	0.1667

Table 2 summarizes the sensitivity indices for the total effect of input variables and conjoint effect (effect of interaction of two variables combined) represented by *. Below is a description of how the analysis was carried out:

- Sobol indices were estimated by generating random samples twice for each input variable to correspond to sample (1) and sample (2) in equation (12). The results show that increasing the random samples generated did slightly improve the accuracy of the sensitivity measures. This is shown by the second-fourth columns in Table 2 where 10K, 100K, and one million random samples were generated to estimate Sobol sensitivity indices⁷.
- The four moment-matching technique employed a 3^3 full factorial DOE with three levels for each input variable and a total of 27 runs. Appendix (A) shows the detailed calculations of the proposed technique. In general, since three levels are selected for each variable a 3^k is needed where k is the number of model variables.

- To be able to compare McKay's correlation ratios with the moment-matching technique, we set the total number of runs equal to what's required by the four moment-matching technique (i.e. 27 runs). Two LHS scenarios each with a total of 27 runs were considered:
 1. Generate a 9-run (9 levels) LHS then augment it with three distinct permutations with 9-runs each ($r = 3, N = 27$).
 2. Generate a 3-run (3 levels) LHS then augment it with nine distinct permutations with 3-runs each ($r = 9, N = 27$).

In both scenarios, the levels are selected by uniformly dividing each variable range into intervals (number of intervals = number of levels for that variable) and select the midpoint of each interval as a level value. Note that the second scenario is nothing but a full factorial DOE, similar to the four moment-matching technique, with different level values. The results in Table 2 are obtained using the second scenario since it provided better results than the former.

Table 2 indicate that four and two moments matching techniques are very close to the Sobol indices but with much lesser runs than Sobol technique requires. The results of McKay's correlation ratios are consistent in terms of the order of the main effects only. Although both McKay's technique and the four moment-matching technique uses a 3^3 full factorial DOE, the four moment technique provided more accurate results due to the two following reasons:

- The four moment-matching technique provided a better selection of the levels used in the DOE.
- McKay's technique stems from traditional ANOVA which treats all variables as uniform while the four moment-matching technique uses the variance contribution step which considers original variables distributions.

Table 2
Sensitivity analysis results of example 1

Variable	Sobol 10K	Sobol 100K	Sobol 1M	CR	Four MM
P	0.015	0.012	0.010	0.004	0.014
E	0.610	0.592	0.591	0.198	0.573
I	0.399	0.362	0.369	0.127	0.367
$P*E$	0.000	0.001	0.002		0.000
$P*I$	0.000	0.006	0.004		0.000
$E*I$	0.019	0.021	0.025		0.046

4.2 Example 2: Oil pump

The flow rate of a hydraulic oil pump FR can be described in terms of three design parameters x_1, x_2 , and x_3 as shown below:

$$FR = \frac{3x_1x_2}{x_3} \quad (30)$$

Based on data provided by suppliers, the three design variables are found to have different skewed distributions as shown in Table 3 below:

Table 3
Oil pump design variables

Variable	x_1	x_2	x_3
Distribution	Beta	Gamma	Exponential
Parameters	shape 1= 0.7 shape 2=0.9	shape=1, scale=2	$\lambda = 3$, threshold=0.1
Mean	0.4375	2.02	3.07
Variance	0.095	4.00	8.77
Skewness	0.225	1.96	1.98
Kurtosis	1.25	5.53	5.67
L_1	0.13	0.713	1.11
L_2	0.436	2.02	3.08
L_3	0.812	7.25	10.93
w_1	0.460	0.469	0.455
w_2	0.166	0.413	0.431
w_3	0.373	0.117	0.114

The last six rows in Table 3 show the results of level and weight selection based on moment matching equations (16). Table 4 show a comparison summary between Sobol sensitivity indices using 10,000 random MC simulations and four moments matching sensitivity indices. The indices are fairly close in general. However there is a slight difference in terms of the importance of x_3 term. This difference is acceptable since the Sobol and moment matching importance ranking of terms is same as shown by third and fifth columns of Table 4 respectively.

Table 4
Sensitivity analysis results of example 2

Variable	Sobol Index	Sobol Importance index	Four Moments Matching	Moment Matching importance index
x_1	0.113	4	0.106	4
x_2	0.165	3	0.156	3
x_3	0.348	1	0.362	1
x_1x_2	0.039	6	0.096	6
x_1x_3	0.106	5	0.085	5
x_2x_3	0.229	2	0.217	2

Although input variables of this example are highly skewed, the moment matching technique was not significantly impacted.

4.3 Example 3: Bottle sampling module

During exploration phase of oil and gas wells, a sample of well reservoir fluid is captured and sent to a chemical lab for analysis. Figure 1 below depicts a concept design for bottle sampling module; one of the core modules of the sampling tool. To fill a sample in the bottle, a pump is used to push the sample downstream where a motorized valve (MV) is opened to direct the sample towards the bottles. The bottle is equipped with a normally open (NO) valve which is closed after filling the bottle. The bottle is typically filled with water behind the piston

which will be pushed out using the pressure of the sample and pass through a check valve (CV). The check valve acts as a one way valve, i.e. prevents any fluid or gas to flow in the opposite direction

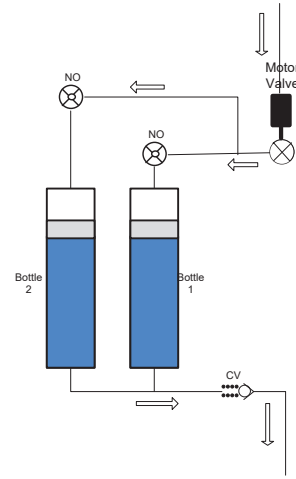


Figure 1 Oil and gas bottle sampling module

The sampling module is a serial system, i.e. any failure in MV, NO, or CV will prohibit filling a sample. In order to increase the likelihood of capturing a sample, another redundant bottle along with NO valve is attached using a parallel configuration as shown in Figure 1. Based on historical field data, all components are found to have a constant failure rate with a uniformly distributed failure rate λ as shown in Table 3. If failure is defined as failure to capture a sample, the mean time to failure can be written as:

$$MTTF = \frac{2}{\lambda_{MV} + \lambda_{CV} + \lambda_{NO}} - \frac{1}{\lambda_{MV} + \lambda_{CV} + 2\lambda_{NO}} \quad (31)$$

Table 5
Bottle sampling module variables

Variable	Min.	Max.	L_1	L_2	L_3
λ_{MV} : Motorized valve failure rate	0.004	0.006	0.004	0.005	0.006
λ_{CV} : Check valve failure rate	0.0005	0.0006	0.0005	0.00055	0.0006
λ_{NO} : Normally open valve failure rate	0.018	0.022	0.018	0.020	0.022

The calculated Sobol indices for the main and total effects are summarized in Table 6 along with the results of both McKay's and moment-matching techniques. Both correlation ratio and moment-matching techniques used 27 total runs similar to the first example.

Once again the four moment-matching technique demonstrated consistent results with respect to Sobol indices.

Table 6
Sensitivity analysis results for bottle sampling module

Variable	Sobol Index	CR	4 MM
λ_{MV}	0.718	0.829	0.726

λ_{CV}	0.0001	0.000	0.0006
λ_{NO}	0.278	0.281	0.279
$\lambda_{MV} * \lambda_{CV}$	0.003		0.004
$\lambda_{MV} * \lambda_{NO}$	0.000		0.000
$\lambda_{CV} * \lambda_{NO}$	0.000		0.000

4.4 Example 4: Engine Block and Head Joint Sealing Assembly

Engine block and head joint sealing assembly is one of the most crucial and fundamental structural designs in the automotive internal combustion engines. The design of engine block and head joint sealing assembly is very complex involving multiple components with complicated geometry to maintain (proper sealing of combustion, high-pressure oil, oil drain, and coolant). The selection of design parameter settings in this assembly (block and head structures, gasket, and fasteners) cannot be analyzed separately because of strong assembly interaction effects. Because design decisions must be made upfront in the product development stage prior to the availability of physical prototypes, the use of computer simulation in the design process is eminent. To best simulate the engine assembly process and its operation, a finite element model (FEM) is used to capture the effects of three-dimensional part geometry, the compliance in the components, non-linear gasket material properties, and contact interface among the block, gasket, head, and fasteners [55]. An example of FEM for this system is shown in Figure 2.

The computer model simulates the assembly process as well as engine operating conditions such as thermal and cylinder pressure cyclical loads due to combustion process. A computer experiment employing a fractional factorial design with 27 runs and six factors described in Table 7 was selected to optimize the design of the head gasket for sealing function. A review of fractional factorial designs and their applications can be found in Phadke [52] and Montgomery [53].

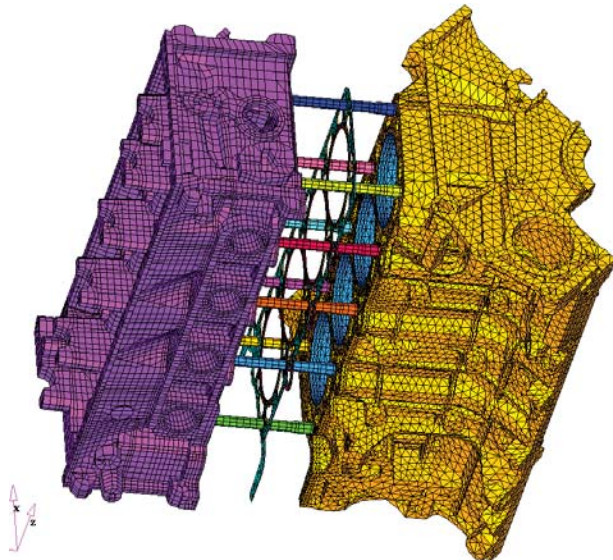


Figure 2 Finite element model of engine assembly

Table 7
Head Assembly Variables

Variable	Mean	STDEV	L_1	L_2	L_3
----------	------	-------	-------	-------	-------

x_1 : Gasket thickness (um)	150	10	132.7	150	167.3
x_2 : Stopper zone-to-zone transition (um/m)	0.5	0.033	0.443	0.5	0.557
x_3 : Coining depth (um)	22.5	2.5	18.17	22.5	25.24
x_4 : Combined cylinder head/block deck face surface flatness	1	0.33	0.428	1	1.57
x_5 : Load/deflection variation	1	0.33	0.428	1	1.57
x_6 : Head bolt force variation (KN)	48	2	44.53	48	51.46

The small number of runs is necessary due to simulation setup complexity and excessive computing requirements. The objective of the design is to optimize the head gasket design factors (x_1 - x_4) to minimize the "gap lift" of the assembly as well as its sensitivity to manufacturing variation (x_5 , x_6). The data of the experiment is shown in Table 8 where the response variable, y , is the gap lift.

The finite element model is computationally intensive, thus a mathematical regression model was developed using the fractional factorial DOE in Table 8 as a surrogate model. For the purposes of sensitivity analysis and comparison with Sobol technique, this cheap-to-compute mathematical meta-model was used.

Table 8
Head Assembly Finite Element Model Results

x_1	x_2	x_3	x_4	x_5	x_6	y
2	3	2	1	2	3	1.53
3	3	3	1	3	1	2.21
1	2	2	1	3	3	1.69
3	2	2	2	3	1	1.92
1	2	3	1	1	2	1.42
1	2	3	3	2	2	5.33
1	1	1	2	3	3	2.00
2	2	1	1	1	1	2.13
3	1	3	2	1	2	1.77
2	1	1	3	1	3	1.89
1	3	3	2	1	3	2.17
3	2	1	2	1	3	2.00
3	1	2	1	2	3	1.66
2	1	3	2	3	1	2.54
1	1	3	1	2	1	1.64
3	3	3	3	2	3	2.14
1	3	1	3	3	2	4.2
3	2	1	3	2	1	1.69
1	1	2	3	1	1	3.74
2	2	3	3	3	3	2.07
2	3	2	3	1	1	1.87
2	2	2	2	2	2	1.19
3	1	2	3	3	2	1.7
2	3	1	1	3	2	1.29
2	1	1	1	2	2	1.82
1	3	1	2	2	1	3.43
3	3	2	2	1	2	1.91

The results for the engine sensitivity analysis using moment-matching technique and Sobol indices are shown in table 9 and 10 respectively. The results of moment-matching method are consistent with Sobol indices except for x_1x_4 term. This is due to the confounding resulted from using L_{27} fractional factorial, which provides clear main effects and confounded second order terms such as x_1x_4 term

Table 9
Head Assembly four moments sensitivity analysis

	x_1	x_2	x_3	x_4	x_5	x_6
x_1	0.304					
x_2	0.000	0.036				
x_3	0.000	0.000	0.003			
x_4	0.456	0.000	0.000	0.165		
x_5	0.000	0.000	0.000	0.000	0.024	
x_6	0.000	0.000	0.000	0.000	0.000	0.013

Table 10
Head assembly Sobol indices

	x_1	x_2	x_3	x_4	x_5	x_6
x_1	0.459					
x_2	0.006	0.032				
x_3	0.006	0.006	0.001			
x_4	0.265	0.006	0.006	0.184		
x_5	0.014	0.006	0.006	0.014	0.017	
x_6	0.006	0.006	0.006	0.006	0.006	0.017

5 Conclusions

As shown by examples, moment matching technique up to the fourth moment provides accurate results for assessing the importance of the input variables with respect to total response variance. Moreover, the technique requires less number of runs while providing comparable results to that of Sobol and McKay techniques.

The main elements behind the moment-matching technique are better selection of the random variable levels, orthogonality in conducting the experiments, and modified variance decomposition. McKay's technique may provide orthogonality as in the examples presented, but moment-matching is superior due to the better selection of the levels of the variables and variance estimations.

The main limitation of the proposed method is the satisfaction of $\beta_2 - \beta_1 > 1$ condition to avoid negative weights. Another limitation indicated by Sobol [18]; variance decomposition reflects the underlying function decomposition in the absence of correlations among the variables. Therefore, a variance based analysis such as the proposed technique is effective only when variables are independent.

Examples 3 and 4 suggests that skewness of input variables and complexity of model output may have an impact on proposed technique accuracy. This impact can be addressed by future research.

References

[1] E. Borgonovo, "Measuring uncertainty importance: investigation and comparison of alternative approaches,"

Risk analysis, vol. 26, no. 5, pp. 1349-61, 2006.

- [2] X. Du and W. Chen, "Sequential optimization and reliability assessment method for efficient probabilistic design," in *ASME Design Automation Conference*, 2002.
- [3] X. Zhuang, R. Pan and X. Du., "Enhancing product robustness in reliability-based design optimization," *Reliability Engineering and System Safety*, vol. 138, pp. 145-153, 2015.
- [4] E. Borgonovo and E. Plischke, "Sensitivity Analysis: A Review of Recent Advances," *Eur. J. Opertion Research*, vol. 3, no. 1, pp. 869-887, 2016.
- [5] G. S. King, R. P. Jones and D. Simner, "A good practice model for implementation of computer-aided engineering analysis in product development," *Journal of Engineering Design*, vol. 14, no. 3, pp. 315-331, 2003.
- [6] A. Saltelli and S. Tarantola, "On the relative importance of input factors in mathematical models: Safety assessment for nuclear waste disposal," *Journal of the American Statistical Association*, vol. 97, no. 459, pp. 702-709, 2002.
- [7] J. C. Helton, "Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal," *Reliability Engineering and System Safety*, vol. 42, pp. 327-367, 1993.
- [8] C. K. Park and K. I. Ahn, "A new approach for measuring uncertainty importance and distributional sensitivity in probabilistic safety assessment," *Reliability Engineering and System Safety*, vol. 46, pp. 253-261, 1994.
- [9] C. H. Frey and S. R. Patil, "Identification and review of sensitivity analysis methods," *Risk Analysis*, vol. 22, no. 3, pp. 553-571, 2002.
- [10] R. L. Iman, M. E. Johnson and C. C. J. Watson, "Sensitivity for computer model projections of hurricane losses," *Risk Analysis*, vol. 25, no. 5, p. 1277-1297, 2005.
- [11] G. Critchfield and K. Willard, "Probabilistic analysis of decision trees using Monte Carlo simulation," *Medical Decision Making*, vol. 6, pp. 85-92, 1986.
- [12] J.-Y. Tissot and C. Prieur, "A randomized Orthogonal Array-based procedure for the estimation of first- and second-order Sobol' indices," *Journal of Statistical Computation and Simulaution*, vol. 85, no. 7, pp. 1358-1381, 2015.
- [13] H. Brewer and K. Canady, "Probabilistic safety assessment support for the maintenance rule at duke power company," *Reliability Engineering & System Safety*, vol. 63, no. 3, pp. 243-249, 1999.
- [14] M. van der Brost and H. Schoonakker, "An overview of PSA importance measures," *Reliability Engineering and System Safety*, vol. 72, pp. 241-245, 2001.
- [15] M. Cheok, G. Parry and S. RR., "Use of importance measures in risk-informed regulatory applications," *Reliability Engineering and System Safety*, vol. 60, pp. 213-226, 1998.
- [16] T. Aven and T. Nøkland, "On the use of uncertainty importance measures in reliability and risk analysis,"

- Reliability Engineering and System Safety*, vol. 95, pp. 127-133, 2010.
- [17] E. Borgonovo, "A new importance measure," *Reliability Engineering and System Safety*, vol. 92, pp. 771-784, 2007.
- [18] I. Sobol, "Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates," *Mathematics and Computers in Simulation*, vol. 55, pp. 271-280, 2001.
- [19] M. D. McKay, J. D. Morrison and S. C. Upton, "Evaluating prediction uncertainty in simulation models," *Computer Physics Communications*, vol. 117, pp. 44-51, 1999.
- [20] Q. Liu and T. Homma, "A New importance measure for sensitivity analysis," *Nuclear Science and Technology*, vol. 1, no. 47, pp. 53-61, 2010.
- [21] R. Howard, "Uncertainty about probability: A decision analysis perspective," *Risk Analysis*, vol. 8, no. 1, pp. 91-98, 1988.
- [22] T. Eschenbach, "Spiderplots versus Tornado diagrams for sensitivity analysis," *Interface*, vol. 22, no. 6, pp. 40-46, 1992.
- [23] E. Borgonovo and G. Apostolakis, "A new importance measure for risk-informed decision making," *Reliability Engineering & System Safety*, vol. 72, no. 2, pp. 193-212, 2001.
- [24] L. Lu and J. Jiang, "Joint failure importance for non coherent fault trees," *IEEE Transactions on Reliability*, vol. 56, no. 3, pp. 435-443, 2007.
- [25] Z. Birnbaum, "On the importance of different components in a multicomponent system," in *Multivariate analysis*, Academic Press, 1969, pp. 581-592.
- [26] T. Homma and A. Saltelli, "Importance measures in global sensitivity analysis of nonlinear models," *Reliability Engineering and System Safety*, vol. 52, pp. 1-17, 1996.
- [27] R. Jin, X. Du and W. Chen, "The use of metamodeling techniques for design under uncertainty," *Structural and Multidisciplinary Optimization*, vol. 25, no. 2, pp. 99-116, 2003.
- [28] A. Saltelli, "Sensitivity analysis for importance assessment," *Risk Analysis*, vol. 22, no. 3, pp. 579-590, 2002.
- [29] A. Saltelli and J. Marivoet, "Non-parametric statistics in sensitivity analysis for model output: A comparison of selected techniques," *Reliability Engineering and System Safety*, vol. 28, pp. 229-253, 1990.
- [30] I. Sobol, "Sensitivity analysis for nonlinear mathematical models," *Mathematical modeling and Computational Experiments*, vol. 1, pp. 407-414, 1993.
- [31] M.-H. Chun, S.-J. Han and N.-I. Tak, "An uncertainty importance measure using a distance metric for the change in a cumulative distribution function," *Reliability Engineering and System Safety*, vol. 70, pp. 313-321, 2002.
- [32] J. Kleijnen, "Kriging metamodeling in simulation: A review.," *European Journal of Operational Research*, vol. 192, no. 3, pp. 707-716, 2009.
- [33] J. Kleijnen, Design and analysis of simulation experiments, Berlin: SpringerVerlag, 2008.
- [34] M. Morris, "Factorial sampling plans for preliminary computational experiments," *Technometrics*, vol. 33, no. 2, pp. 161-174, 1991.
- [35] B. Bettonvil, Detection of important factors by sequential bifurcation., Tilburg: Tilburg University Press, 1990.
- [36] B. Bettonvil and J. Kleijnen, "Searching for important factors in simulation models with many factors: Sequential bifurcation," *European Journal of Operational Research*, vol. 96, no. 1, pp. 180-194, 1997.
- [37] R. Iman and S. Hora, "A robust measure of uncertainty importance for use in fault tree system analysis," *Risk Analysis*, vol. 10, no. 3, pp. 401-406, 1990.
- [38] A. Saltelli and I. Sobol', "About the use of rank transformation in the sensitivity analysis of model output," *Reliability Engineering & System Safety*, vol. 50, no. 3, pp. 225-239, 1995.
- [39] B. Sudret, "Global sensitivity analysis using polynomial chaos expansions," *Reliability Engineering & System Safety*, vol. 93, no. 7, pp. 964-979, 2008.
- [40] J. E. Oakley and A. O'Hagan, "Probabilistic sensitivity analysis of complex models: A Bayesian approach," *Journal of the Royal Statistical Society*, vol. 66, no. 3, pp. 751-769, 2004.
- [41] G. Blatman and B. Sudret, "Efficient computation of global sensitivity indices using sparse polynomial chaos expansions," *Reliability Engineering & System Safety*, vol. 95, no. 11, pp. 1216-1229, 2010.
- [42] M. D. Morris, L. M. Moore and M. D. McKay, "Using orthogonal arrays in the sensitivity analysis of computer models," *Technometrics*, vol. 50, no. 2, pp. 205-215, 2008.
- [43] H. Rabitz and O. F. Alis, "General foundations of high dimensional model representations," *Journal of Mathematical Chemistry*, vol. 25, no. 2-3, pp. 197-233, 1999.
- [44] A. Saltelli, S. Tarantola, F. Campolongo and M. Ratto, Sensitivity Analysis in Practice. A Guide to Assessing Scientific Models, New York: JohnWiley, & Sons publishers, 2004.
- [45] R. Iman, "A matrix-based approach to uncertainty and sensitivity analysis for fault trees," *Risk Analysis*, vol. 7, no. 1, pp. 21-33, 1987.
- [46] C.-F. Huang and R. Litzenberger, Foundations of financial economics., Upper Saddle River, NJ: Prentice Hall, 1988.
- [47] M. McKay, "NUREG/CR-6311: Evaluating Prediction Uncertainty," Los Alamos National Laboratory, NM 87545, 1995.
- [48] D. Morton, B. Leellier, J. Tejada, D. Johnson, Z. Mohagheg, E. Kee, S. Reihani and A. Zolan, "Sensitivity analysis for a high-order simulation used in the STP GS1-191 risk-informed resolution project," in *International Conference on Nuclear Engineering (ICONE22)*, 2014.

- [49] J. R. D'Errico and N. A. Zaino Jr., "Statistical tolerancing using a modification of Taguchi's method," *Technometrics*, vol. 30, pp. 397-405, 1988.
- [50] H. S. Seo and B. M. Kwak, "Efficient statistical tolerance analysis for general distributions using three-point information," *International Journal of Production Research*, vol. 40, pp. 931-944, 2002.
- [51] D. H. Evans, "An application of numerical integration techniques to statistical tolerancing," *Technometrics*, vol. 9, no. 3, pp. 441-456, 1967.
- [52] M. S. Phadke, *Quality Engineering Using Robust Design*, Prentice-Hall, 1989.
- [53] D. C. Montgomery, *Design and Analysis of Experiments*, Hoboken, NJ: John Wiley & Sons, Inc., 2013.
- [54] P. Thoft-Christensen and M. J. Baker, *Structural Reliability Theory and Its Applications*, Berlin: Springer-Verlag Berlin Heidelberg, 1982.
- [55] F. Popielas, C. Chen and S. Obermaier, "CAE approach for multi-layer-steel cylinder head gaskets," in *SAE Technical Paper, SAE 2000-01-1348*, Detroit, USA, 2000.
- [56] K. T. Fang, D. K. J. Lin, P. Winker and Y. Zhang, "Uniform design theory and applications," *Technometrics*, vol. 42, pp. 237-248, 2000.

Appendix A

For the beam deflection example illustrated in section 4.1; the following DOE was conducted:

Table 1A
Beam deflection example full factorial DOE

P	w_p	E $\times 10^7$	w_E	I $\times 10^4$	w_I	w_G	G
2.268	0.1667	1.133	0.1667	0.653	0.1667	0.0046	562.67
2.268	0.1667	1.133	0.1667	1	0.6667	0.0185	955.82
2.268	0.1667	1.133	0.1667	1.3464	0.1667	0.0046	1348.29
2.268	0.1667	2	0.6667	0.653	0.1667	0.0185	1128.82
2.268	0.1667	2	0.6667	1	0.6667	0.0741	1822.82
2.268	0.1667	2	0.6667	1.3464	0.1667	0.0185	2515.62
2.268	0.1667	2.866	0.1667	0.653	0.1667	0.0046	1694.32
2.268	0.1667	2.866	0.1667	1	0.6667	0.0185	2688.82
2.268	0.1667	2.866	0.1667	1.3464	0.1667	0.0046	3681.60
4.000	0.6667	1.133	0.1667	0.653	0.1667	0.0185	427.35
4.000	0.6667	1.133	0.1667	1	0.6667	0.0741	820.50
4.000	0.6667	1.133	0.1667	1.3464	0.1667	0.0185	1212.97
4.000	0.6667	2	0.6667	0.653	0.1667	0.0741	993.50
4.000	0.6667	2	0.6667	1	0.6667	0.2963	1687.50
4.000	0.6667	2	0.6667	1.3464	0.1667	0.0741	2380.30
4.000	0.6667	2.866	0.1667	0.653	0.1667	0.0185	1559.00
4.000	0.6667	2.866	0.1667	1	0.6667	0.0741	2553.50
4.000	0.6667	2.866	0.1667	1.3464	0.1667	0.0185	3546.28
5.732	0.1667	1.133	0.1667	0.653	0.1667	0.0046	292.03
5.732	0.1667	1.133	0.1667	1	0.6667	0.0185	685.18
5.732	0.1667	1.133	0.1667	1.3464	0.1667	0.0046	1077.65
5.732	0.1667	2	0.6667	0.653	0.1667	0.0185	858.18
5.732	0.1667	2	0.6667	1	0.6667	0.0741	1552.18
5.732	0.1667	2	0.6667	1.3464	0.1667	0.0185	2244.98
5.732	0.1667	2.866	0.1667	0.653	0.1667	0.0046	1423.68
5.732	0.1667	2.866	0.1667	1	0.6667	0.0185	2418.18
5.732	0.1667	2.866	0.1667	1.3464	0.1667	0.0046	3410.96

Equations (16) are used to estimate three levels and weights for P , E , and I . A sample of calculations is shown below for E :

$$l_1 = 2.0 \times 10^7 + 0.5(5.0 \times 10^6)(0) - 0.5(5.0 \times 10^6)\sqrt{(4)(3) - 3(0)} = 1.113 \times 10^7$$

$$w_1 = \frac{((4)(3) - 3(0)) + \sqrt{(0)}}{2((4)(3) - 3(0))(3 - 0)} = 0.1667$$

The mean and variance of the response G is estimated using (14) and (15) respectively:

$$\begin{aligned} \mu_G &= \sum_{j=1}^{27} w_{G_j} * G_j = 0.0046 * 562.7 + \dots + 0.0046 * 3,410.9 \\ &= 1,687.133 \\ \sigma_G^2 &= \sum_{j=1}^{27} w_{G_j} * (G_j - \mu_G)^2 = 426,596.9 \end{aligned}$$

Based on the variance reduction technique, the effect of each input variable is estimated. The effect of P will be demonstrated here as an example. First the reduced variance of G is estimated using the nine runs of the full factorial DOE where P equals its mean=4.0 as shown in Table 2A and weight of P is adjusted to be 1. This reduced variance represents $V_{E,I}(G|P=4)$ with $E, I = \sim P$.

$$\begin{aligned} \mu_{G|P=4} &= \sum_{j=1}^9 w_{G_j} * G_j = 0.0278 * 42,735 + \dots + 0.0278 * 3,546 \\ &= 1,687.133 \end{aligned}$$

and

$$V_{E,I}(G|P=4) = \sum_{j=1}^9 w_{G_j} * (G_j - \mu_G)^2 = 420,493$$

The next step is then to estimate the same reduced variance for $P = 2.268$ and $P = 5.732$ and calculate the weighted average to have the contribution of E and I without any interaction with P . This will give you

$$\begin{aligned} &E_P(V_{E,I}(G|P)) \\ &= \frac{V_{E,I}(G|P=4) + V_{E,I}(G|P=2.268) + V_{E,I}(G|P=5.732)}{3} \\ &= 420,493 \end{aligned}$$

Then the total contribution of P without the influence of all interactions between P and E , and P and I can be estimated as shown below:

$$1 - \frac{E_P(V_{E,I}(G|P))}{Var(G)} = 1 - \frac{420,493}{426,596.9} = 0.0143$$

Table 2A
Effect of P in beam deflection example

P	w_p	E $\times 10^7$	w_E	I $\times 10^{-4}$	w_I	w_G	G
4.0	1	1.133	0.1667	0.653	0.1667	0.0278	427.35
4.0	1	1.133	0.1667	1	0.6667	0.1111	820.50
4.0	1	1.133	0.1667	1.3464	0.1667	0.0278	1212.97
4.0	1	2	0.6667	0.653	0.1667	0.1111	993.50
4.0	1	2	0.6667	1	0.6667	0.4444	1687.50
4.0	1	2	0.6667	1.3464	0.1667	0.1111	2380.30
4.0	1	2.866	0.1667	0.653	0.1667	0.0278	1559.00
4.0	1	2.866	0.1667	1	0.6667	0.1111	2553.50
4.0	1	2.866	0.1667	1.3464	0.1667	0.0278	3546.28

Another way to estimate the sensitivity indices is by calculating the sum of squares as shown in Table 3A

Table 3A
Sum of squares of beam deflection example

$Term$	SS	Di
P	329,609	$\frac{329,609}{23,573,581} = 0.0139$
E	13,509,397	0.573
I	8,651,582	0.367
$P * E$	0	0
$P * I$	0	0
$E * I$	1,082,994	0.046
$Total$	23,573,581	