Dimensionality Reduction of Optimization Problems Using Variance Based Sensitivity Analysis

Keiichi Ito[1,2], Tom Dhaene[1]

[1]*Department of Information Technology (INTEC), University of Ghent, Gaston Crommenlaan 8 bus 201, 9050 Ledeberg - Ghent, Belgium*
[2]*Noesis Solutions, Gaston Geenslaan 11 B4, 3001 Leuven, Belgium*
{*keiichi.ito, tom.dhaene*}*@intec.ugent.be*

Keywords: Sensitivity Analysis, Sobol Indices, Dimensionality Reduction, Interaction, Independence, Optimization

Abstract: We propose a new interaction index derived from the computation of Sobol indices. In optimization, interaction index can be used to detect lack of interaction among input parameters. First order interaction indices if they return zero, means that those parameters can be optimized independently holding other parameters constant. Likewise, second order interaction indices can tell if a combination of two parameter can be optimized independently of other parameters. In this way, the original optimization problem may be decomposed into a set of lower dimensional problems which may then be solved independently and in parallel. The interaction indices can potentially be useful in robust optimization as well, since it provides importance measure in minimizing output variances.

1 INTRODUCTION

In today's engineering endeavor, it is common to run computer simulations to understand the behavior of complex systems and optimize their parameters to obtain satisfactory designs before actual physical prototypes are built. Very often we encounter optimization problems with many parameters for which optimum values are sought. The simulation codes are usually complex and are expensive to run. Optimizing such high-dimensional problems with expensive objective functions is still a big challenge even at the disposal of powerful computational facilities. Building surrogate models that approximate the mapping of input to output of expensive simulation models is a way to mitigate the difficulty in evaluating the expensive functions many times. However, creating such surrogate models in high-dimensional parameter space is yet another challenge. In this paper, we propose an approach to decompose high-dimensional problem into a set of lower dimensional problem via our new interaction indices. Calculation of these indices is a simple extension to Sobol indices and gives information about particular parameter(s) being independent or interacting with other parameters.

We begin by describing High Dimensional Model Representation and Sobol indices in the next section and in the following section. Then, we describe the use of variance information obtained in the calculation of Sobol indices in optimization problems (section 4), and define the interaction indices (section 5). We compare interaction index with an existing interaction detection method (section 6). Then, we give a simple example (section 7) showing the results of Sobol and interaction indices and their implication are discussed in the subsequent section (section 8).

2 HDMR AND SOBOL' INDICES

Consider a deterministic model $y = f(\mathbf{x})$ where $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is a vector of n input variables and y is the model output. $f(\mathbf{x})$ can be decomposed into a form referred to as high dimensional model representation.

$$
\begin{aligned}
f(\mathbf{x}) = & f_0 + \sum_i f_i(x_i) + \sum_{i<j} f_{ij}(x_i, x_j) \\
& + \sum_{i<j<k} f_{ijk}(x_i, x_j, x_k) + \ldots
\end{aligned} \tag{1}
$$

This decomposition of the function is not unique as the lower order can be selected arbitrarily and the highest order term can be written as the difference between $f(\mathbf{x})$ and the lower order terms. However, if the average of each of the term in the summands of the right hand side of equation (1) is set to zero and

f_0 is set to be a constant, the expression is proven to be unique (Sobol', 1993). The terms are given as the following.

$$f_0 = E(y) \tag{2}$$

$$f_i(x_i) = E(y|x_i) - f_0 \tag{3}$$

$$f_{ij}(x_i,x_j) = E(y|x_i,x_j) - f_i(x_i) \\ -f_j(x_j) - f_0 \tag{4}$$

$$f_{ijk}(x_i,x_j,x_k) = E(y|x_i,x_j,x_k) - f_{ij}(x_i,x_j) \\ -f_{ik}(x_i,x_k) - f_{jk}(x_j,x_k) \\ -f_i(x_i) - f_j(x_j) - f_k(x_k) \\ -f_0. \tag{5}$$

The first order sensitivity index for variable x_i is given by

$$S_i = \frac{V[E(y|x_i)]}{V(y)}, \tag{6}$$

and if we calculated the indices to the highest order, we have

$$\sum_{i=1}^{n} S_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} S_{ij} + \ldots + S_{ij\ldots n} = 1. \tag{7}$$

The numerator in equation (6) means "variance of expected value of y given x_i", the denominator is the "total (unconditional) variance". The S_is are called first order Sobol' indices (European Commission, Joint Research Centre of Ispra, Italy, ; Sobol', 2001; Chan et al., 1997).

Total effect index (Homma and Saltelli, 1996) includes interaction effects in addition to the first order sensitivity indices, and can be defined as

$$S_{Ti} = 1 - S_{-i}, \tag{8}$$

where S_{-i} signifies sum of all the sensitivity indices except those that include variances due to x_i. For example, if $i \in \{1,2,3\}$, the total effect index of x_1 is

$$S_{T1} = S_1 + S_{12} + S_{13} + S_{123} \\ = 1 - S_2 - S_3 - S_{23}. \tag{9}$$

The total effect index defined in equation (8) is useful in variable screening. Variables with $S_{Ti} \simeq 0$ can be held constant at an arbitrary value within its lower and upper bounds since it means that the variable's value does not contribute to the variance in the output. The first order sensitivity indices in equation (6) alone cannot be used for this purpose if there are significant amount of interactions among the variables.

The virtue of this variance based sensitivity analysis is that it takes all the nonlinearities and interactions into account as compared to "one variable at a time" way of sensitivity analysis. Moreover, even if we conduct the sensitivity analysis only to the first order, we will always have an idea of what remains unexplained by calculating $1 - \sum_{i=1}^{n} S_i$.

3 FORMULATION

Following the discussion in the previous section, we now formulate the way to compute the first order sensitivity indices. As before, we express a surrogate model with n-dimensional input space as

$$y = f(x_1, x_2, \ldots, x_n). \tag{10}$$

The total variance is therefore

$$D = \underset{X \in R^n}{V}(f(X)) \\ = \int_{X \in R^n} f^2(x_1, x_2, \ldots, x_n) dx_1 dx_2 \ldots dx_n \\ -f_0^2. \tag{11}$$

The multidimensional integral of equation (11) can be computed using the Monte Carlo integration. Similarly, the following equation gives the $E(y|x_i)$.

$$f_i(x_i) = \int_1 \int_2 \ldots \int_{n-1} f(x_1, x_2, \ldots, x_n) dx_{-i} - f_0, \tag{12}$$

where $dx_{-i} \equiv dx_1 dx_2 \ldots dx_{i-1} dx_{i+1} \ldots dx_n$. We also define $V(y|x_i)$ for later use in our discussion,

$$\underset{x_{-i}}{V}(y|x_i) = \\ \int_1 \int_2 \ldots \int_{n-1} f^2(x_1, x_2, \ldots, x_n) dx_{-i} - f_0^2. \tag{13}$$

This is the variance of y when x_i is fixed at certain value. Again, in equations (12) and (13), the integrations are performed using the Monte Carlo method, but this time x_i is held constant. By fixing x_i at various values, we can conduct the next integration to obtain $V[E(y|x_i)]$.

$$D_i = \underset{x_i}{V}(f_i(x_i)) = \int_{x_i} f_i^2(x_i) dx_i \tag{14}$$

Then,

$$S_i = \frac{D_i}{D}. \tag{15}$$

The computation of $f_i(x_i)$ at different values of x_i to calculate D_i in equation (14) is a brute-force approach. It requires $m \times (n \times l + 1)$ function evaluations, where m is the number of Monte Carlo samples, n is the number of input variables, and l is the number of different x_i values that are used to compute equation (14). There is a more efficient method in which all S_is and S_{Ti}s are calculated in $m \times (n + 2)$ function evaluations (Saltelli, 2002; Saltelli et al., 2008) provided that all input variables x_is' distributions are independent.

4 EXTRACTING INFORMATION FOR OPTIMIZATION

In the process of optimization, for example minimizing y by judicious choice of x_is, one would also be interested in the variance of y given x_i, $V(y|x_i)$ or more generally, the distribution of y given x_i. Let us denote such distribution (or probability density function) as $p(y|x_i)$. These information can easily be obtained during the course of calculating the first order Sobol' Indices. This information can be used in three ways. First, it tells you for what value of x_i one could possibly have the smallest y. Second, it tells you if x_i has any interaction with other parameters. Finally, it tells you what value of x_i would satisfy certain reliability criteria. That is, one could draw a threshold value for y beyond which these variances should not exceed.

For optimization, one would be interested in the minimum of y given x_i. Let us denote this value as $\min(y|x_i)$. Then, one would choose x_i such that

$$x_i^* = \underset{x_i}{\arg\min} \left(\min(y|x_i) \right). \tag{16}$$

If you had control of only one parameter x_i, say x_1, x_1^* would be the point at which $E(y|x_1)$ would be minimum, because no matter what the values for other parameters may be, on average $\underset{x_1}{\arg\min}(E(y|x_1))$ would return minimum y. However, if we know that other parameters are also controllable, global minimum y can be obtained by solving equation (16) because each point in the distribution $p(y|x_1)$ is a deterministic realization of y for certain combination of parameters $x_{i\neq1}$ with x_1 held constant. One can iterate through other parameters determining x_i^* in the same way via equation (16) to obtain the minimum. In the actual calculation of Sobol indices, $p(y|x_i)$ is obtained at finite number of values of x_i. Thus, a good strategy would be to bracket the x_i^* with a new upper and lower bound of x_i, and obtain $p(y|x_i)$ for a new set of finite number of values of x_i in this new interval. One can repeat this procedure until the upper and lower bound of x_i^* becomes narrow enough. Since x_i is fixed at finite set of values in the interval, it is possible that the minimum escapes the search. Therefore, we cannot guarantee if global optimum is found. Note that once the upper and lower bound changes, so will the Sobol indices in equation (6).

If some or all of the x_is contain uncertainties such that their intervals cannot be reduced beyond certain level, the resulting $p(y|x_i)$s will represent the uncertainties in the output due to the uncertainties in these x_is. For reliability purposes, one may also be interested in $\max(y|x_i)$ which is the other end of the distri-

bution tail showing the maximum y that could occur given x_i.

Figure 1 shows an example of representing $p(y|x_i)$ as box plots. The example shows the spread of y i.e. $p(y|x_i)$ in vertical axis with respect to four input parameters x_1, x_2, x_3, and x_4 in horizontal axis. We see by visual inspection that x_2 and x_3 are independent parameters because $V(y|x_2)$ and $V(y|x_3)$ are constant across different values of x_2 and x_3 respectively. On the other hand, x_1 and x_4 have interactions with other parameters because $V(y|x_1)$ and $V(y|x_4)$ are not constant. Since we have only four parameters we can conclude that x_1 and x_4 interact with each other. The red arrows indicate the interval of interest if one is seeking x_i for minimum of y. The green arrows indicate the interval of interest if one is seeking x_i for minimum variance.

5 INTERACTION INDICES

In order to quantify the interaction or independence of input variables, we propose the following interaction index,

$$I_i = \frac{V\left[V(y|x_i)\right]}{V^2(y)}, \tag{17}$$

where we can compute $V(y|x_i)$ from equation (13). We can then set a threshold ε below which we say that the input x_i does not have significant interaction with other input variables and thus can be treated independently. Note that an interaction index I_i is not constant if the upper and lower bound of **x** changes. In the optimization process in which these bounds change every iteration, we need to keep track of I_is. At certain point in the iteration, $I_i \leq \varepsilon$ may hold for x_is that were $I_i > \varepsilon$ at the beginning.

We can extend this concept to detect two and higher dimensional subproblems.

$$I_{ij} = \frac{V\left[V(y|x_i,x_j)\right]}{V^2(y)}, \tag{18}$$

$$I_{ijk} = \frac{V\left[V(y|x_i,x_j,x_k)\right]}{V^2(y)}, \tag{19}$$
$$\cdots$$

6 COMPARISON

It is also possible to detect interaction via the total effect indices and first order Sobol indices, $S_{Ti} - S_i$. However, there are some important differences between the two methods. First, $S_{Ti} - S_i$ does not give the relative importance among input variables driving
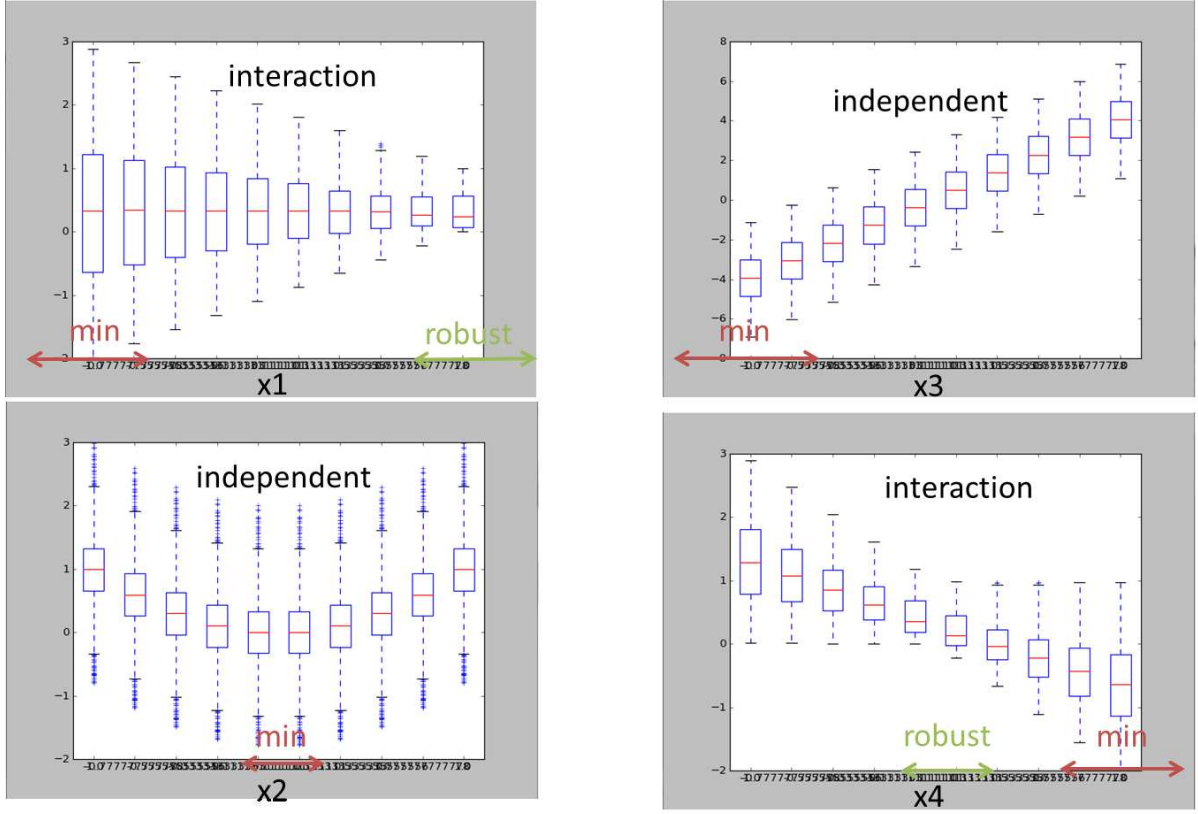
Figure 1: Examples of $p(y|x_i)$

$V[V(y|x_i)]$. For example S_{ij} that is one of the summands of S_{Ti} does not give information about relative importance between x_i and x_j in driving the variance of $f_{ij}(x_i, x_j)$. Thus, x_i could be interacting with many other variables but having minor importance in each of them. In such a case, $S_{Ti} - S_i$ may give a misleading impression of importance in interaction. In the next section, we will see a case in which

$$S_{Ti} - S_i = S_{Tj} - S_j$$

but

$$I_i \neq I_j.$$

Second, the detection of independence $I_i = 0$ is not sensitive to the accuracy of Monte Carlo integration. As long as $V(y|x_i)$ is computed with the same samples in x_{-i}, $V(y|x_i)$, remains constant - although the value of $V(y|x_i)$ may be approximate - throughout different values of x_i. Thus, I_i should show zero to arithmetic precision if x_i does not interact with other variables. On the other hand, $S_{Ti} - S_i$ is subject to the Monte Carlo integration inaccuracy. Therefore, judicious selection of threshold value is needed to detect independence.

Third, as of this writing, computation of $S_{Ti} - S_i$ is less expensive than computation of I_i. The set of

S_{Ti}s and S_is can be obtained from $m \times (n + 2)$ function evaluations, where m is the number of Monte Carlo samples, n is the number of input variables. We currently use the brute-force approach to compute the I_is at the cost of $m \times (n \times l + 1)$ function evaluations, where l is the number of different x_i values that are used to compute equation (14). The computation of I_i does not require any further function evaluation (i.e. computation of response y) beyond what is required for the computation of first order Sobol indices S_i. Computing S_{Ti} in brute-force approach, on the other hand, is often infeasible (requiring computation of up to $n - 1$ order Sobol indices).

We expect that there are shortcuts like in the computation of S_{Ti}s and S_is to economize the computation of I_is. However, the brute-force approach have some advantages over the more efficient ones. It can work even if the input variables are correlated, and it gives distributions at specific values of x_is. This information could further be exploited in the optimization of input variables. To benefit from the shortcuts available for Sobol indices and total effect indices, we will have to have uncorrelated input variables and we will not have distributions for a specific value of x_i.

7 EXAMPLE

Consider the following simple example.

$$y_1 = x_1 + 2x_2 + 4x_3 \quad (20)$$
$$y_2 = x_1^2 - x_2 + x_2 x_3 \quad (21)$$

where $-1 < x_1, x_2, x_3 \le 1$. We obtained the first order Sobol indices of y_1 and y_2 with respect to x_1, x_2, and x_3 as in Table 1. We calculated the results by 1000

Table 1: First order Sobol Indices for the example

	y_1	y_2
x_1	0.05	0.19
x_2	0.19	0.65
x_3	0.77	0.00
Higher Order	-0.01	0.16

uniform random samples. Fifty different values for x_is were used to compute the values for equation (14). The negative value for the second order and higher interaction terms (i.e. the bottom row, second column) is due to numerical error in the Monte Carlo integration and analytically it should be zero. The analytical values of Sobol Indices for y_1 can be easily calculated and they are

$$S_1 = \frac{2^2}{2^2 + 4^2 + 8^2} \simeq 0.04719, \quad (22)$$

$$S_2 = \frac{4^2}{2^2 + 4^2 + 8^2} \simeq 0.19048, \quad (23)$$

$$S_3 = \frac{8^2}{2^2 + 4^2 + 8^2} \simeq 0.76190. \quad (24)$$

The interaction indices for the same example are shown in Table 2. The zero entries in Table 2 indi-

Table 2: First order interaction indices for the example

	y_1	y_2
x_1	0.000	0.000
x_2	0.000	0.036
x_3	0.000	0.573

cate that corresponding variables do not interact with other variables. For y_1, there are no interacting variables, and in optimization, x_1, x_2, and x_3 can be optimized one by one while fixing the remaining variables as constants. For y_2, x_1 is independent but x_2 and x_3 are interacting. Thus, x_1 can be optimized with other parameter fixed but x_2 and x_3 must be optimized jointly while fixing x_1 at a constant value.

Table 3 shows the result of calculating $S_{Ti} - S_i$ using 1000 samples for the Monte Carlo integration. The column for y_1 and the entry for x_1 under the column for y_2 should show zeros. The slightly positive

Table 3: $S_{Ti} - S_i$ for the example

	y_1	y_2
x_1	0.03	0.03
x_2	0.03	0.17
x_3	0.03	0.17

values are again due to inaccuracies in Monte Carlo integration. For the y_2 column, the equation (21) suggests that the entry for x_2 and x_3 should show some values for the interaction. Equations (25) to (27) show the expressions of $S_{Ti} - S_i$ for y_2. The reason that

$$S_{T2} - S_2 = S_{T3} - S_3$$

in Table 3 can be understood from the equations (26) and (27).

$$S_{T1} - S_1 = S_{12} + S_{13} + S_{123} = 0, \quad (25)$$
$$S_{T2} - S_2 = S_{12} + S_{23} + S_{123} = S_{23}, \quad (26)$$
$$S_{T3} - S_3 = S_{13} + S_{23} + S_{123} = S_{23}. \quad (27)$$

The difference between Table 2 and Table 3 illustrates the difference between the two methods of detecting interactions and independences. The reason for $I_2 < I_3$ in Table 2 can be understood by factoring equation (21) as in equation (28),

$$y_2 = x_1^2 + x_2 \cdot (-1 + x_3). \quad (28)$$

For the given upper and lower bounds of x_2 and x_3, we have $0 < -1 + x_3 \le 2$ and $-1 < x_2 \le 1$. Thus, if we sample x_2 and x_3 uniformly between -1 and 1, we have the following. If we let $x_2 = 1$ or -1, then we get the largest $V(y_2|x_2)$ with

$$V[x_2 \cdot (-1 + x_3)|x_2 = \pm 1] = \frac{(2-0)^2}{12} = \frac{1}{3}. \quad (29)$$

On the other hand, if we let $x_3 = -1$, then

$$V[x_2 \cdot (-1 + x_3)|x_3 = -1] = \frac{(2-(-2))^2}{12} = \frac{4}{3}, \quad (30)$$

and $V(y_2|x_3)$ is largest. Furthermore,

$$V[x_2 \cdot (-1 + x_3)|x_2 = 0] = $$
$$V[x_2 \cdot (-1 + x_3)|x_3 = 1] = 0. \quad (31)$$

Thus,

$$\frac{I_3}{I_2} = \frac{V[V(y_2|x_3)]}{V[V(y_2|x_2)]} = \frac{4^2}{1^2} = 16, \quad (32)$$

which confirms Table 2.

8 DISCUSSION

In practical situations in which the calculation of y given an input vector $\mathbf{x}$ is expensive, the computation

of Sobol and interaction indices may be prohibitive due to the number of model evaluations needed to do the Monte Carlo integration. In such cases, fitting surrogate models to the dataset computed by the original model may be useful. Surrogate models are approximation to the original function, and are much cheaper to compute than the original model. It is usually fit on a finite number of input-output data obtained from the original model (usually a complex simulation model). Kriging and Radial Basis Functions are some of the popular surrogate models (Keane and Nair, 2005). Let us denote the output produced by the surrogate model as $\hat{y}$. We can compute the indices based on $\hat{y}$s. It is our experience that if Pearson's correlation coefficient between y and $\hat{y}$ is reasonable, say above 0.8, the Sobol Indices calculated on the surrogate model gives fairly accurate indication of important variables. Whether this robustness holds for interaction indices i.e. whether it reliably predicts non-interacting variables, is yet to be investigated.

An adaptive sampling and optimization method can be conceived. In each iteration, surrogate model can be fit and Sobol indices and interaction indices can be evaluated on the surrogate model. Then, we can determine the next sampling points and bounds of each input parameters. We can also preform problem decomposition to generate a set of lower dimensional problems. The effective way to combine the search, the sampling, and the problem decomposition needs to be investigated. It is expected that the problem decomposition provided by the interaction indices may help in the accuracy of surrogate models since for a given number of samples the lower the input parameter dimension, the more accurate it is in general. For example, one surrogate model could be fit for each of the independent subcomponents. The development of an efficient sampling method based on the evolving information of interaction and sensitivity will be the next step in the research.

The conventional way of detecting interaction obscures each input variable's relative importance in influencing the interaction with other variables. The new interaction indices detects changes (variances) in $V(y|x_i)$ and thus enables one to rank variables in terms of "influence to interaction". This can be a valuable information when conducting robust optimization. It is expected that this will be helpful in devising algorithms to perform problem decomposition to lower dimensional problems.

9   CONCLUSION

We have shown the potential use of Sobol indices and our new interaction indices in decomposing high-dimensional problem into a set of independent lower dimensional problems in optimization. This should effectively simplify the construction of surrogate models and optimization. The interaction index exposes each parameter's importance in influencing the variance in the output through interaction. This is expected to be useful in robust optimization. Application to industrial problems is needed to understand the effectiveness of the proposed index. Also, further research is needed to exploit the concept described in this paper to develop a surrogate model assisted optimization algorithm that is scalable to high-dimensional problems.

ACKNOWLEDGMENTS

REFERENCES

Chan, K., Saltelli, A., and Tarantola, S. (1997). Sensitivity analysis of model output: Variance-based methods make the difference. In *Proceedings of the 1997 Winter Simulation Conference*, pages 261–268.

European Commission, Joint Research Centre of Ispra, Italy. A cookbook for applying sensitivity analysis in practical cases of modeling. http://sensitivity-analysis.jrc.ec.europa.eu/tutorial/index.asp.

Homma, T. and Saltelli, A. (1996). Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering and System Safety*, 52:1 – 17.

Keane, A. J. and Nair, P. B. (2005). *Computational Approaches for Aerospace Design: The Pursuit of Excellence*. John Wiley & Sons.

Saltelli, A. (2002). Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, 145:280 – 297.

Saltelli, A., Rato, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S. (2008). *Global Sensitivity Analysis. The Primer*. John Wiley & Sons.

Sobol', I. M. (1993). Sensitivity estimates for nonlinear mathematical models. *Mathematical Modeling and Computational Experiment*, 1(4):407 – 414.

Sobol', I. M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55:271–280.