Reliability Engineering and System Safety 107 (2012) 107–114

Contents lists available at SciVerse ScienceDirect

Reliability Engineering and System Safety

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

A variance-based sensitivity index function for factor prioritization

Douglas L. Allaire *, Karen E. Willcox

Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA 02139, United States

ARTICLE INFO

Article history:
Received 15 October 2010
Received in revised form
15 August 2011
Accepted 22 August 2011
Available online 3 September 2011

Keywords:
Global sensitivity analysis
Factor prioritization
Variance reduction
Rejection sampling

ABSTRACT

Among the many uses for sensitivity analysis is factor prioritization—that is, the determination of which factor, once fixed to its true value, on average leads to the greatest reduction in the variance of an output. A key assumption is that a given factor can, through further research, be fixed to some point on its domain. In general, this is an optimistic assumption, which can lead to inappropriate resource allocation. This research develops an original method that apportions output variance as a function of the amount of variance reduction that can be achieved for a particular factor. This variance-based sensitivity index function provides a main effect sensitivity index for a given factor as a function of the amount of variance of that factor that can be reduced. An aggregate measure of which factors would on average cause the greatest reduction in output variance given future research is also defined and assumes the portion of a particular factors variance that can be reduced is a random variable. An average main effect sensitivity index is then calculated by taking the mean of the variance-based sensitivity index function. A key aspect of the method is that the analysis is performed directly on the samples that were generated during a global sensitivity analysis using rejection sampling. The method is demonstrated on the Ishigami function and an additive function, where the rankings for future research are shown to be different than those of a traditional global sensitivity analysis.

1. Introduction

Sensitivity analysis of model output has been defined as the determination of how uncertainty in the output of a model can be apportioned to different sources of uncertainty in the model factors [1]. Sensitivity analysis defined in this manner is often referred to as *global sensitivity analysis*, owing to the fact that entire factor distributions are considered in the apportionment process. Since what is meant by the term "uncertainty" is typically case dependent, several indicators have been developed to apportion different measures of uncertainty among model factors. These indicators are often based on screening methods [2], variance-based methods [3–6], entropy-based methods [6,7], non-parametric methods [6,8], and moment-independent approaches [9–11]. This paper focuses on the development and demonstration of an extension of traditional variance-based global sensitivity analysis that considers the change in output variance caused by a change in factor variance that may arise from researching a factor further. In general, variance-based global sensitivity analysis is the standard practice for determining how each factor contributes to output uncertainty when output variance is considered sufficient to describe output variability [4,5].

Variance-based global sensitivity analysis is a rigorous method for apportioning output variance [3,12]. The method has been applied in a wide variety of applications including hydraulic modeling [13], aviation environmental modeling [14], nuclear waste disposal [4], robust mechanical design practices [15], and many others. The two main metrics computed in variance-based global sensitivity analysis are the main effect sensitivity indices proposed by Sobol' [16] and the total effect sensitivity indices proposed by Homma and Saltelli [5]. One of the primary uses of global sensitivity analysis is in the context of *factor prioritization* [3]. In this setting, the objective is to determine which factor, on average, once fixed to its true value, will lead to the greatest reduction in output variance. It has been established by Saltelli et al. [3] and Oakley and O'Hagan [17] that the main effect sensitivity indices are appropriate measures for ranking factors in this setting, however, as noted in Oakley and O'Hagan [17], it is rarely possible to learn the true value of any uncertain factor, and thus these sensitivity indices only suggest the potential for reducing uncertainty in an output through new research on a factor. Given that it is rarely possible to obtain the true value of any uncertain factor, the assumption that a given factor will be fixed to some point on its domain is a major limitation in the use of main effect sensitivity indices for use in allocating resources aimed at reducing output variance.

To account for the inherent limitations in using global sensitivity analysis results for directing future research, a new method

* Corresponding author.
E-mail address: dallaire@mit.edu (D.L. Allaire).

0951-8320/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ress.2011.08.007

that apportions output variance as a function of the amount of variance reduction that can be achieved for a particular factor has been developed. This function is called the variance-based sensitivity index function. By assuming the portion of a particular factor's variance that can be reduced is a random variable, the mean of this function can be taken to provide average main effect sensitivity indices for ranking purposes. A key aspect of the method is that the analysis is performed directly on the factor and output samples that were generated during a global sensitivity analysis using a rejection sampling technique for Monte Carlo simulation proposed in Beckman and McKay [18]. The derivation of the method is given in Section 2, which is followed by a demonstration of the method on the well-known Ishigami function [19] and a purely additive function in Section 3. Conclusions are presented in Section 4.

2. Methodology

In the following subsections, the method is derived using main effect sensitivity indices from global sensitivity analysis. The notion of a reasonable distribution that could arise through future research on a given factor is considered as well as the rejection sampling technique and how it can be employed to reuse model evaluations from a Monte Carlo simulation.

2.1. Derivation

Consider a generic model, $Y=f(\mathbf{x})$, where $\mathbf{x}=[X_1,\ldots,X_k]^T$, and $X_1,\ldots,X_k$ are random variables on the measurable space $(\mathbb{R},\mathcal{B})$, and $f:\mathbb{R}^k\rightarrow\mathbb{R}$ is $(\mathcal{B}^k,\mathcal{B})$-measurable. Then Y is a random variable, and by definition, the variance of Y can be decomposed according to

$$\mathrm{var}(Y)=\mathbb{E}[\mathrm{var}(Y|X_i)]+\mathrm{var}(\mathbb{E}[Y|X_i]), \tag{1}$$

for any X_i, where $i\in\{1,\ldots,k\}$. According to Saltelli et al. [3], the goal of factor prioritization is the identification of which factor, once fixed at its true value, would reduce the variance of Y the most. Since it is not known *a priori* a given factor's true value, factor prioritization is carried out by identifying the factors which, on average, once fixed, would cause the greatest reduction in the variance of Y. The average amount of variance remaining once a given factor is fixed is just $\mathbb{E}[\mathrm{var}(Y|X_i)]$ for any factor X_i. Thus, according to Eq. (1), the average amount of the variance of Y that could be reduced through fixing factor X_i somewhere on its domain is $\mathrm{var}(\mathbb{E}[Y|X_i])$. Global sensitivity analysis uses this fact for factor prioritization by considering main effect sensitivity indices, which take the form

$$S_i=\frac{\mathrm{var}(\mathbb{E}[Y|X_i])}{\mathrm{var}(Y)}, \tag{2}$$

where S_i is the main effect sensitivity index of factor X_i. The main effect sensitivity index can then be used as a measure of the proportion of the variance of Y that is expected to be reduced once factor X_i is fixed to its true value.

The calculation of main effect sensitivity indices in a global sensitivity analysis is most commonly done using either the Fourier Amplitude Sensitivity Test (FAST) method or the Sobol' method [5,12,20,21]. The FAST method is based on Fourier transforms, while the Sobol' method utilizes Monte Carlo simulation. The Sobol' method is employed in this work.

The Sobol' method is well-developed and in wide use in the sensitivity analysis field. Following Homma and Saltelli [5], the main effect sensitivity indices may be estimated via the Sobol' method for a given factor X_i by first estimating the mean f_0 of the function $Y=f(\mathbf{x})$ as

$$\hat{f}_0=\frac{1}{N}\sum_{m=1}^{N}f(\mathbf{x}^m), \tag{3}$$

where $\mathbf{x}^m$ represents the mth realization of the random vector $[X_1,\ldots,X_k]^T$ and N denotes the total number of realizations of the random vector. Then estimating the variance of the function as

$$\hat{V}=\frac{1}{N}\sum_{m=1}^{N}f(\mathbf{x}^m)^2-\hat{f}_0^2, \tag{4}$$

and the single-factor partial variances as

$$\hat{V}_i=\frac{1}{N}\sum_{m=1}^{N}f([x_1^m,\ldots,x_i^m,\ldots,x_k^m]^T)f([\tilde{x}_1^m,\ldots,x_i^m,\ldots,\tilde{x}_k^m]^T)-\hat{f}_0^2, \tag{5}$$

where x_j^m and $\tilde{x}_j^m$ denote different samples of factor X_j and the partial variances can be computed for $i\in\{1,\ldots,k\}$. The main effect sensitivity index for factor X_i can then be computed according to

$$\hat{S}_i=\frac{\hat{V}_i}{\hat{V}}. \tag{6}$$

Here it should be noted that improvements in estimating main effect indices using sampling-based methods have been developed by Saltelli et al. [22] and using regression or emulator-based methods by Lewandowski et al. [23], Oakley and O'Hagan [17], Tarantola et al. [24], Ratto et al. [25], Storlie and Helton [26]. Our focus in this work is on sampling-based approaches, which are commonly used in situations where both main effect and total effect indices are desired [22]. The methodology developed in this paper can readily be applied to the sampling-based techniques of Saltelli et al. [22], however, the development is more accessible in the context of the traditional Sobol' method. Adapting the work of this paper to regression and emulator-based methods is a topic for future work.

As noted previously, these main effect sensitivity indices may be used for factor prioritization by ranking inputs according to their main effect indices, which give the percentage of how much output variability can be expected to be eliminated by fixing a particular input somewhere on its domain. However, this use of global sensitivity analysis for factor prioritization relies on the assumption that a given factor, through future research, can be fixed to some point on its domain. The key contribution of this work is to relax that assumption by considering the amount of variance that can be reduced for a given factor as a random variable rather than assuming the variance to be completely reducible. More precisely, we assume that for a given amount of variance reduction for a factor X_i, there is a corresponding family of allowable distributions, and we calculate an average change in the variance of the model output over this family.

Let X_i^o be the random variable defined by the original distribution for some factor X_i, and X_i' be the random variable defined by a new distribution for factor X_i after some further research has been done. X_i^o and X_i' have corresponding main effect sensitivity indices S_i^o and S_i' respectively. Then we can define the ratio of the variance of factor X_i that is not reduced and the total variance of the original distribution of factor X_i as $\lambda_i=\mathrm{var}(X_i')/\mathrm{var}(X_i^o)$. Assuming further research reduces the variance of factor X_i,[1] it is clear that $\lambda_i\in[0,1]$. Since it cannot be known in advance how much variance reduction for a given factor is possible through further research, λ_i is cast as a uniform random variable Λ_i on [0,1], which corresponds to a maximum entropy distribution given that all we know is the interval in which λ_i will take a value [27].

[1] It is possible that further research could increase the variance of a factor, however, this would suggest that the original characterization of uncertainty was flawed.

Given that the variance of factor X_i that may be reduced is a random percentage, $100(1-\Lambda_i)\%$, of the total original variance of factor X_i, the variance-based sensitivity index function can be defined as

$$\zeta_i(\lambda_i) = \frac{\text{var}(Y^o)S_i^o - \mathbb{E}[\text{var}(Y')S_i' | \Lambda_i = \lambda_i]}{\text{var}(Y^o)}, \quad (7)$$

where S_i^o is the original main effect sensitivity index of factor X_i, $\text{var}(Y^o)$ is the original output variance, and $\mathbb{E}[\text{var}(Y')S_i' | \Lambda_i = \lambda_i]$ is the expected value of the product of the variance of the output and the main effect global sensitivity index of factor X_i taken over all *reasonable* distributions of factor X_i with $100\lambda_i\%$ of the variance of the original distribution for factor X_i. The reasonable distributions are a pre-specified, parameterized family of distributions. These distributions are discussed further in the following subsection.

The variance-based sensitivity index function given by Eq. (7) provides the main effect sensitivity index for factor X_i if it is known that exactly $100(1-\lambda_i)\%$ of the factor's variance can be reduced. This can be seen by noting that $\text{var}(Y^o)S_i^o$ is the expected value of the variance of Y^o that is due to factor X_i, and $\text{var}(Y')S_i'$ is the expected value of the variance of Y' that is due to factor X_i after $100(1-\lambda_i)\%$ of factor X_i's variance has been reduced. Since there are many ways to reduce the variance of factor X_i by $100(1-\lambda_i)\%$, the expected value of $\text{var}(Y')S_i'$ is taken over all the reasonable distributions for which $100(1-\lambda_i)\%$ has been reduced. Thus, $\text{var}(Y^o)S_i^o - \mathbb{E}[\text{var}(Y')S_i' | \Lambda_i = \lambda_i]$ is the amount of variance in Y^o that cannot be reduced further if factor X_i's variance can only be reduced by $100(1-\lambda_i)\%$.

If it is assumed that all of the variance of a particular factor can be reduced, then $\lambda_i = 0$, and for a given factor X_i, this means that $\mathbb{E}[\text{var}(Y')S_i' | \Lambda_i = 0] = 0$, since once all of the variance of factor X_i has been reduced, factor X_i will simply become a constant, and thus, $S_i' = 0$. Therefore, when $\lambda_i = 0$, $\zeta_i(0) = S_i^o$, and the index reduces to the specific case of global sensitivity analysis. However, as noted previously, since it is not likely known what value λ_i will take prior to further research on a given factor, λ_i is considered to be a uniform random variable, Λ_i, on the interval [0,1].

The expected value of $\zeta_i(\Lambda_i)$ can thus be taken to give an *average main effect sensitivity index* (η), as shown in Eq. (8) for some factor X_i,

$$\eta_i = \mathbb{E}_{\Lambda_i}[\zeta_i(\Lambda_i)]. \quad (8)$$

The average main effect sensitivity index for each factor in a model is then an index that can be used to quantitatively rank factors based on the average amount of output variance that can be reduced when further research is done on a particular factor.

2.2. Defining reasonable distributions

In the discussion of the development of the variance-based sensitivity index function it was noted that reasonable new factor distributions, which represent the result of further research on a factor, be used in the estimation of the function. This is because given some initial distribution for a factor and some λ_i, there will generally not be a unique new distribution with $100\lambda_i\%$ of the variance of the original factor distribution. For example, if a factor has an original distribution that is uniform on the interval [0,1], and $\lambda_i = 0.5$, there are an infinite number of new distributions, such as $U[0,\sqrt{2}/2]$, $U[1-\sqrt{2}/2,1]$, $U[\sqrt{2}/4,1-\sqrt{2}/4]$, etc., that all have variances equal to λ_i times the original variance. The new distributions could also be from a different family of distributions, such as triangular. Therefore, a set of reasonable distributions with $100\lambda_i\%$ of the variance of any given original distribution must be defined. Here we present a procedure for identifying reasonable distributions for factors that are originally uniformly distributed or normally distributed. These distributions tend to be used in the absence of full distributional information and thus are often candidates for future research. In both cases, it is assumed that future research will only impact the parameters of a given distribution, thus the impact of future research that could lead to a change in the underlying distribution family (e.g. from a uniform distribution to a triangular distribution) is not considered here. However, if the distribution family of a given factor were expected to change through further research, then reasonable distributions from the new family, given that the original distribution was from another family, could be defined.

2.2.1. Uniform distributions that may arise through future research

Consider an arbitrary uniform distribution, $U[a,b]$. The variance of this distribution is given as $\text{var}(X) = (b-a)^2/12$. Thus, λ_i for this family of distributions can be written as $\lambda_i = \left[(b'-a')/(b^o-a^o)\right]^2$, where a' and b' are the endpoints of a new distribution and a^o and b^o are the endpoints of the original distribution. In this case, a given λ_i implies all new uniform distributions are intervals of the same width, which is $\lambda_i^{1/2}(b^o-a^o)$. A reasonable method for sampling from the set of intervals on $[a^o,b^o]$ with width $\lambda_i^{1/2}(b^o-a^o)$ is given in Algorithm 1.

Algorithm 1. Sampling uniform distributions.

1: Sample λ_i from a uniform distribution on the interval [0,1].
2: Sample b' from a uniform distribution on the interval $[a^o+\lambda_i^{1/2}(b^o-a^o), b^o]$.
3: Let $a' = b' - \lambda_i^{1/2}(b^o-a^o)$.

This method of sampling ensures the new parameters, a' and b', for a given λ_i, will be such that $a' \sim U[a^o, b^o-\lambda_i^{1/2}(b^o-a^o)]$, and $b' \sim U[a^o+\lambda_i^{1/2}(b^o-a^o), b^o]$. Thus the set of possible uniform distributions with $(1-\lambda_i)$ the variance of the original distribution is sampled uniformly.

2.2.2. Normal distributions that may arise through future research

Consider an arbitrary normal distribution, $\mathcal{N}(\mu_o,\sigma_o^2)$, where μ_o is the mean and σ_o^2 is the variance of the distribution. For the normal family of distributions, λ_i is written as $\lambda_i = \sigma'^2/\sigma_o^2$, where σ'^2 is the variance of a new distribution and σ_o^2 is the original variance. Here a procedure is presented where the mean value of the original distribution is also the mean value of any new distributions after further research has been undertaken. However, if the mean value is expected to change, other procedures can be developed to take that into account. Given that here the mean does not change, a specific λ_i uniquely defines a new distribution $\mathcal{N}(\mu_o,\lambda_i\sigma_o^2)$, where μ_o is the mean of the original distribution. Thus, the proposed procedure for sampling normal distributions is simply selecting a λ_i and using that λ_i to calculate the variance of the new distribution. This procedure is given in Algorithm 2.

Algorithm 2. Sampling normal distributions.

1: Set $\mu' = \mu_o$.
2: Sample λ_i from a uniform distribution on the interval [0,1].
3: Set $\sigma'^2 = \lambda_i\sigma_o^2$.

2.3. Rejection sampling

The evaluation of Eq. (7) and subsequently of Eq. (8) requires consideration of a large number of different distributions for each factor. If a global sensitivity analysis is carried out for each new distribution for each factor, the computational expense would be massive and estimating values of the variance-based sensitivity

index function would likely be too costly to ever carry out. However, if a global sensitivity analysis with the original distributions for each factor is completed, rejection sampling can be used to estimate the values of the variance-based sensitivity index function without any further model evaluations.

Rejection sampling is a method for generating samples from a desired distribution by sampling from a different distribution. The method has been developed for both random samples [28] and quasi-random samples such as low discrepancy sequences [29]. For random samples, following DeGroot and Schervish [28], let $f_Z(z)$ be a probability density function of a desired distribution for some random variable, Z. Let $f_X(x)$ be some other probability density function for a random variable, X, with the property that there exists a constant, k, such that $kf_X(x) \geq f_Z(x)$ for all x, where x is a realization of X. The rejection method can then be used to generate J samples from $f_Z(z)$ as shown in Algorithm 3. For low discrepancy sequences, following Wang [29], a similar procedure for sampling from Z can be developed as shown in Algorithm 4. Both of these algorithms are rigorous techniques for performing rejection sampling on nested uniform distributions. However, in practice, we may use the bounds of the desired new interval to do rejection sampling directly by rejecting all points outside the bounds.

Algorithm 3. Rejection sampling for 1D random samples.

1: Draw a sample, x, from X.
2: Draw a sample, u, from a uniform random variable on [0,1].
3: If $f_Z(x)/f_X(x) \geq ku$
  let $z_j = x$,
  $j = j+1$,
  If $j=J$,
    STOP.
  Else return to 1.
4: Else, discard x and u and return to 1.

Algorithm 4. Rejection sampling for 1D low discrepancy sequences.

1: Generate a low discrepancy sequence of points $(\xi_i, v_i) \in [0,1]^2$, $i = 1,2,\ldots,M$.
2: Map ξ_i to x_i for all i (e.g., using inverse CDF method).
3: Let $u_i = v_i$.
4: For $i=1$ to M
  If $f_Z(x_i)/f_X(x_i) \geq ku_i$,
    let $z_j = x_i$,
    $j = j+1$,
  Else reject x_i.

As an example of how rejection sampling is used in this work, consider a function $Y = f(X_1, X_2)$, where X_1, X_2, and Y are random variables and $X_1, X_2 \sim U[0,1]$. Suppose we care about some quantity $\mathcal{Q}(Y)$ (e.g. an integral, a mean, a sensitivity index, etc.). To evaluate $\mathcal{Q}(Y)$, we first sample from the random variables X_1 and X_2 using a Sobol' quasi-random sequence [30]. The points for a sequence of size 1024 are shown in Fig. 1a.

Following Beckman and McKay [18], if we would now like to evaluate $\mathcal{Q}(Y)$ with a different distribution on say X_1, we can do so by reusing previous samples of X_1 and X_2, and hence samples of Y, by performing the appropriate rejection sampling on X_1. The use of rejection sampling for this task only requires that the support of the new distribution of X_1 be contained within the support of the original distribution of X_1 and the existence of a uniform bound k as in Algorithms 3 and 4. For example, if we would like to evaluate $\mathcal{Q}(Y)$, where $Y = f(X_1, X_2)$ and $X_1 \sim U[1/5, 3/5]$ and $X_2 \sim U[0,1]$, we could do so without reevaluating Y by using the evaluations of Y associated with the samples of X_1 that have been accepted in rejection sampling as samples of the new distribution of X_1. These samples are shown Fig. 1b, which presents the original samples of X_1 and X_2 and the accepted samples after rejection sampling is applied to X_1. It should be noted here that this use of rejection sampling for the case of normal distributions in which the mean may change poses computational difficulties due to the potential small number of previous function evaluations near the new mean of the distribution. This difficulty may also arise if we wish to use rejection sampling on a small subset of the original interval in the case of both uniform and normal distributions.

2.4. Application of rejection sampling to global sensitivity analysis

Rejection sampling can be employed to reuse the results from a global sensitivity analysis to calculate values of the variance-based sensitivity index functions and average main effect sensitivity indices as follows. Consider again a generic model, $Y = f(\mathbf{x})$, where $\mathbf{x} = [X_1, \ldots, X_k]^T$, and $X_1, \ldots, X_k$ and Y are random variables. Suppose we have conducted a global sensitivity analysis to calculate the main effect sensitivity index of factor X_i using the

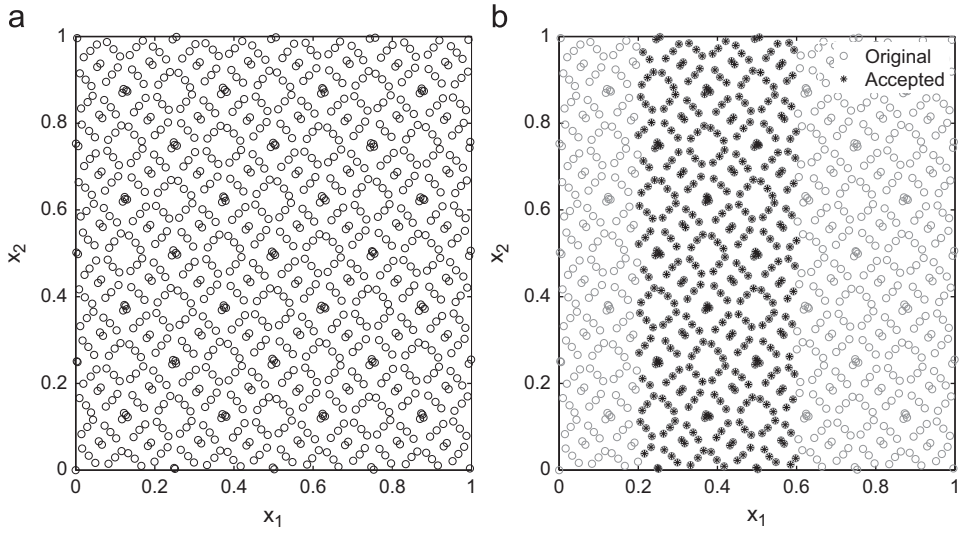

Fig. 1. (a) Original set of 1024 quasi-random Sobol' points for X_1 and X_2, which are used to calculate samples of Y. (b) Accepted samples of X_1 and X_2 after rejection sampling has been performed on X_1. Originally $X_1 \sim U[0,1]$ and is now $X_1 \sim U[1/5, 3/5]$. The original samples are presented for comparison.

Sobol' method [5] and quasi-random Sobol' points. Thus, we have the model evaluations corresponding to $f([x_1^m,\ldots,x_i^m,\ldots,x_k^m]^T)$ and $f([\tilde{x}_1^m,\ldots,x_i^m,\ldots,\tilde{x}_k^m]^T)$, as in Eq. (5), where $m=1,\ldots,N$, where N is the number of samples of each factor. Consider some new distribution, $f_{X_i'}(x')$ for factor X_i, with original distribution $f_{X_i^o}(x_i)$, for which we would like to determine the value of the variance-based sensitivity index function for a particular amount of variance reduction in X_i. To do this, we must estimate the main effect sensitivity index, S_i', and the new variance of the output, var(Y'), for use in the estimation of a variance-based sensitivity index function value for factor X_i given by Eq. (7). The estimation of these quantities using the function evaluations from global sensitivity analysis can be achieved as shown in Algorithm 5.

Algorithm 5. Rejection sampling for the Sobol' method for factor X_i.

1: Use Algorithm 3 or 4 to choose a subsample of the original points.
2: Calculate the new variance of the output var(Y') using Eq. (4) and the output samples associated with the subsample from Step 1.
3: Calculate the single-factor partial variance for X_i using the samples from Step 1 and Eq. (5).
4: Calculate the new Sobol' main effect sensitivity index S_i' using Eq. (6) with the variance from Step 2 and the partial variance from Step 3.

The values of the variance-based sensitivity index function for factor X_i may then be calculated as shown in Algorithm 6.

Algorithm 6. Evaluating $\zeta_i(\lambda_i)$ for factor X_i.

1: Estimate the original output variance var(Y^o).
2: Perform a global sensitivity analysis for factor X_i to estimate the original main effect sensitivity index S_i^o.
3: Use Algorithm 5 to estimate var(Y') and S_i'. Repeat over all reasonable distributions for factor X_i with $100\lambda_i\%$ of the variance of the original distribution for factor X_i.
4: Calculate $\zeta_i(\lambda_i)$ using Eq. (7) with the quantities from Steps 2 and 3.

Given the variance-based sensitivity index function for factor X_i, we may estimate the average main effect sensitivity index for factor X_i as shown in Algorithm 7.

Algorithm 7. Evaluating η_i for factor X_i.

1: Discretize the interval [0,1] and estimate $\zeta_i(\lambda_i)$ at the discretization points (e.g., 0, 0.1, 0.2, …,1.0) using Algorithm 6.
2: Estimate the average main effect sensitivity index η_i using Eq. (8) with the values of ζ_i from Step 1.

3. Test function analysis

To demonstrate the methodology developed in Section 2, the approach is applied here to the Ishigami function and an additive function. The Ishigami function was first introduced by Ishigami and Homma [19]. It is commonly used to test sensitivity and uncertainty analysis techniques. For example, it was used by Ratto et al. [31] to demonstrate the use of state-dependent parameter modeling in the estimation of conditional moments for sensitivity analysis, by Homma and Saltelli [5] to demonstrate the performance of importance measures for sensitivity analysis, by Saltelli et al. [32] to demonstrate the calculation of high dimensional model representation [16] for use in variance-based sensitivity analysis, by Eldred and Swiler [33] to explore refinement approaches for nonintrusive polynomial chaos expansion and stochastic collocation for uncertainty quantification techniques, and by Storlie et al. [34] to investigate the use of metamodels and bootstrap confidence intervals for sensitivity analysis of computationally demanding models. The second example, an additive function, was created for this work to fully demonstrate the benefits of a variance-based sensitivity index function and to provide an example that does not contain strong interactions as the Ishigami function does.

3.1. Ishigami function

The Ishigami function is given as follows:

$$Y=\sin X_1+a\sin^2 X_2+bX_3^4\sin X_1, \tag{9}$$

where the X_i are independent and uniformly distributed on $[-\pi,\pi]$. The constants are set as $a=5$ and $b=0.1$ as in Ratto et al. [31]. A global sensitivity analysis was carried out using the Sobol' method and a Sobol' quasi-random sequence of size 4096. The computed main effect sensitivity indices for each factor are $S_1=0.40$, $S_2=0.28$ and $S_3=0.00$. For factor prioritization purposes then, the conclusion that is drawn from the analysis is to focus future research efforts on factor X_1, since on average once fixed, factor X_1 is expected to reduce the variance of Y by the largest amount.

If the analysis for factor prioritization were to be concluded at this point, a great deal of information regarding the impact of future research on factors X_1 and X_2 is missed and an inappropriate decision regarding how to allocate future resources for reducing the variance of Y may be made. It should be noted here that factor X_3 is not considered further because it does not have a main effect and thus only affects the variance of Y through interactions with the other factors. Thus, to demonstrate the benefits of the variance-based sensitivity index function, an analysis for factors X_1 and X_2 was carried out following the methodology presented in Section 2.

The results of the analysis are presented in Fig. 2a. Indices of each factor for values of λ_i, that is the variance that cannot be reduced, for $\lambda_i=0.0,0.05,0.10,\ldots,1.0$ are provided. Without aggregating the results, the figure shows that the effects of future research on a given factor on the variance of Y is highly nonlinear. Thus, depending on the expected returns of future research on a factor, the main effect global sensitivity indices, which are the rightmost points on the figure, could be misleading if used for determining how to best allocate resources aimed at reducing the variance of Y. As an extreme example of how misleading the information from the main effect indices computed via global sensitivity analysis could be, consider directing future research at factor X_1 (which is supported by factor prioritization analysis) and achieving a 25% reduction in the variance of the factor (thus $1-\lambda_1=0.25$). As can be seen in Fig. 2a, on average, a 25% reduction in the variance of X_1 will actually lead to an *increase* in the variance of Y, and thus not have been an appropriate use of resources. In fact, according to the results of the distributional sensitivity analysis, unless it is believed that future research will lead to a reduction in the variance of X_1 of more than 50%, it does not make sense to direct research at factor X_1 at all if the goal is to reduce the variance of Y. This behavior can be explained by considering the Ishigami function output plotted against input X_1 and X_2 as shown in Fig. 3a and b respectively. In Fig. 3a, the dark points are accepted samples for a $1-\lambda_1=0.25$ case for X_1 and the light points are the rejected points from the full model. It is clear from the figure that the variance of the dark points is greater than the variance of the combination of the dark and light points,

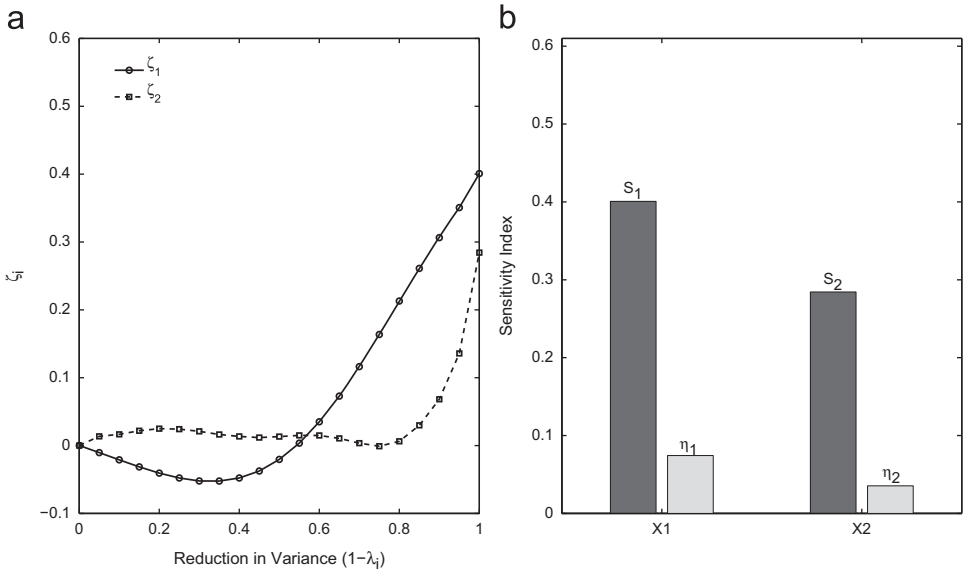

Fig. 2. (a) Variance-based sensitivity index functions of factors 1 and 2 for values of $\lambda_i = 0.0, 0.05, 0.10, \ldots, 1.0$. The function values for each factor for $\lambda_i = 0$ (rightmost points) are the main effect sensitivity indices of each factor as computed via global sensitivity analysis. (b) A comparison of the main effect sensitivity indices and the average main effect sensitivity indices.

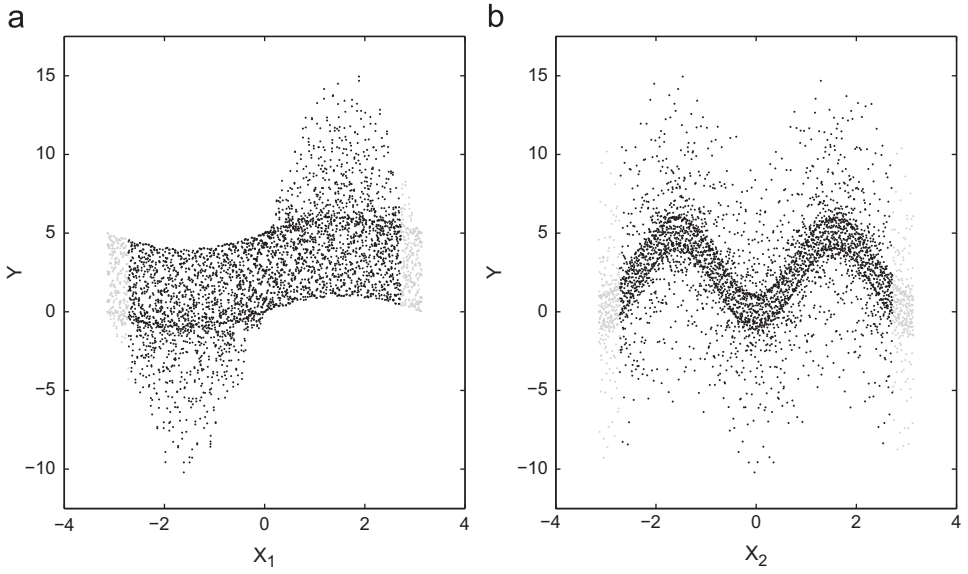

Fig. 3. (a) Ishigami function output versus factor X_1. The dark points (all but the narrow bands on the right and left) are accepted samples for a $1-\lambda_1 = 0.25$ case for X_1 and the light points are the rejected points from the full model. (b) Ishigami function output versus factor X_2. The dark points are accepted samples for a $1-\lambda_2 = 0.25$ case for X_2 and the light points are the rejected points from the full model.

which leads to the behavior of the distributional sensitivity index for factor X_1 shown in Fig. 2. In Fig. 3b, the dark points are accepted samples for a $1-\lambda_2 = 0.25$ case for X_2 and the light points are again the rejected points from the full model. In this case, it is clear from the figure that variance of the dark points should be very similar to the variance of the dark and light points combined, which leads to the nearly constant behavior of the distributional sensitivity index of factor X_2 shown in Fig. 2.

By aggregating the variance-based sensitivity index function values shown in Fig. 2b as discussed in Section 2 to obtain the average main effect sensitivity indices, we are provided with an alternative ranking measure for deciding how to allocate future resources. An important aggregation case to consider is that of having a complete lack of knowledge relating the possible variance reduction for a given factor after future research. As discussed in Section 2, this leads to a uniform distribution on [0,1] for the amount of variance of a given factor that can be reduced. Given this form of aggregation, the average main effect sensitivity indices are $\eta_1 = 0.07$ and $\eta_2 = 0.04$ as shown in Fig. 2b. Here the average main effect sensitivity indices rank factor X_1 above factor X_2 as was the case for global sensitivity analysis. However, the analysis also revealed that neither factor is expected to reduce the variability of Y substantially, and the conclusion of the analysis could be that allotting resources to research further the distributions of X_1 and X_2 may not be worthwhile. This was not the conclusion of the global sensitivity analysis, where it is assumed that all factor uncertainty can be eliminated, leading to a 40% or 28% reduction in the variance Y through researching factor X_1 or factor X_2 respectively. Further, different aggregation procedures can be developed to provide more informative average main

effect sensitivity indices if we have knowledge of the amount of variance that is expected to be reduced given anticipated future research efforts on a given factor.

3.2. Additive function

The additive model is given as follows:

$$Y = 100X_1 + 4\exp(X_2) + 350\sin X_3, \tag{10}$$

where the X_i are independent $\sim\mathcal{N}(0,4)$. A global sensitivity analysis was carried out using the Sobol' method and a Sobol' quasi-random sequence of size 65 536. The computed main effect sensitivity indices for each factor are $S_1 = 0.27$, $S_2 = 0.31$ and $S_3 = 0.42$. For factor prioritization purposes then, according to global sensitivity analysis, the conclusion that is drawn is to focus future research efforts on factor X_3.

The variance-based sensitivity index functions and the average main effect sensitivity indices for the additive model are shown in Fig. 4a and b respectively. As can be seen from the figure, the results are considerably different than those of the global sensitivity analysis. The variance-based sensitivity index functions plot clearly shows that the factors that should be considered for future research depend on the amount of variance that is assumed reducible for each factor. If we have no knowledge of the impact of future research on any of the factors then it is reasonable to assume that the amount of variance of any given factor that can be reduced through future research is a uniform distribution between 0 and 100% of the variance of the factor. Given this, we can aggregate the variance-based sensitivity index function values as discussed in Section 2.1 to obtain the average main effect sensitivity indices for each factor, which are shown in Fig. 4b. The global sensitivity analysis results suggest the ranking for factor prioritization be factor 3, followed by factor 2, and then factor 1, though all of the indices are close to one another. The average main effect sensitivity index results however, suggest that the ranking be factor 2, followed by factor 1, and then factor 3. Thus, in the case of this additive function, assuming the variance of a given factor can be reduced to zero through future research leads to a completely different conclusion regarding which factors should be researched further than the methodology developed here.

4. Conclusions

The sensitivity analysis technique developed in this work provides a method of ranking factors in terms of expected output variance reduction that would occur if future research is done on a given factor. The assumption that a given factor can be fixed somewhere on its domain, a key assumption of using global sensitivity analysis for factor prioritization, has been relaxed by viewing sensitivity indices as functions of the amount of variance of a particular factor can be reduced through research. Using rejection sampling techniques, these variance-based sensitivity index functions are estimated by reusing the information obtained from a traditional global sensitivity analysis. In many cases, this approach does not require additional evaluations of the model being analyzed, though the need for significantly more model evaluations can arise when considering small subsets of original factor intervals and normal distributions in which the mean may change. The variance-based sensitivity index functions reveal a great deal of information regarding expected reduction in output variance given reduction in factor variance, and can be used to determine how to allocate resources for future research among different factors. For any possible forecast a practitioner may make regarding the possible decrease in variance in each of the factors caused by future research, a different possible ranking of the factors may be realized through the variance-based sensitivity index function. For overall ranking purposes, the percentage of a given factor's variance that can be reduced through future research can be considered as a uniform random variable and average main effect sensitivity indices can be computed by taking the mean of the variance-based sensitivity index functions. For the two test functions considered, this approach led to a great deal of useful additional information for consideration in the factor prioritization setting, as well as different conclusions for factor prioritization for the additive test function. Though the results in this work assumed there is no knowledge of anticipated returns in terms of variance reduction for a given factor given some planned future research, if such knowledge does exist, the information can be incorporated into the aggregation of the average main effect sensitivity indices to provide weighted average main effect sensitivity indices for ranking purposes.

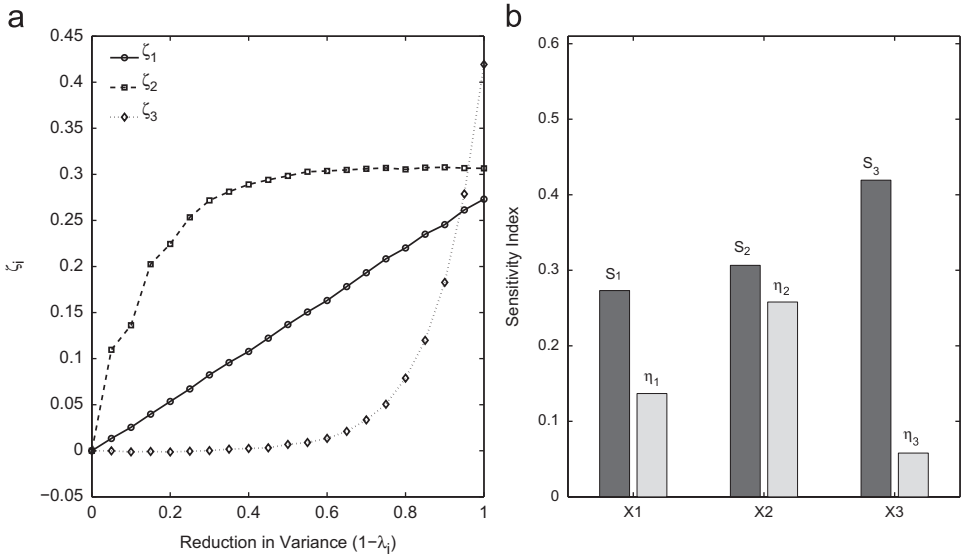

Fig. 4. (a) Variance-based sensitivity index functions of factors 1,2, and 3 for values of $\lambda_i = 0.0, 0.05, 0.10, \ldots, 1.0$. The function values for each factor for $\lambda_i = 0$ (rightmost points) are the main effect sensitivity indices of each factor as computed via global sensitivity analysis. (b) A comparison of the main effect sensitivity indices and the average main effect sensitivity indices.

Acknowledgments

This work was supported by the Office of Environment and Energy, U.S. Federal Aviation Administration (FAA) under Contract no. DTFAWA-05-D-00012, Task Order 0002, managed by Maryalice Locke. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the FAA. The authors acknowledge the support of R. Cointin, FAA Assessment Program Lead.

References

[1] Saltelli A. Sensitivity analysis for importance assessment. Risk Analysis 2002;22(3):579–90.
[2] Morris M. Factorial sampling plans for preliminary computational experiments. Technometrics 1991;33(2):161–74.
[3] Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, Saisana M, Tarantola S. Global sensitivity analysis: the primer. West Sussex, England: John Wiley & Sons, Ltd.; 2008.
[4] Saltelli A, Tarantola S. On the relative importance of input factors in mathematical models: safety assessment for nuclear waste disposal. Journal of the American Statistical Association 2002;97:702–9.
[5] Homma T, Saltelli A. Importance measures in global sensitivity analysis of nonlinear models. Reliability Engineering and System Safety 1996;52:1–17.
[6] Helton J, Johnson J, Sallaberry C, Storlie C. Survey of sampling-based methods for uncertainty and sensitivity analysis. Reliability Engineering and System Safety 2006;91:1175–209.
[7] Liu H, Chen W, Sudjianto A. Relative entropy based method for global and regional sensitivity analysis in probabilistic design. ASME Journal of Mechanical Design 2006;128(2):1–11.
[8] Saltelli A, Marivoet J. Non-parametric statistics in sensitivity analysis for model output: a comparison of selected techniques. Reliability Engineering and System Safety 1990;28:229–53.
[9] Borgonovo E. Measuring uncertainty importance: investigation and comparison of alternative approaches. Risk Analysis 2006;26(5):1349–61.
[10] Borgonovo E. A new uncertainty importance measure. Reliability Engineering and System Safety 2007;92:771–84.
[11] Chun M, Han S, Tak N. An uncertainty importance measure using a distance metric for the change in a cumulative distribution function. Reliability Engineering and System Safety 2000;70:313–21.
[12] Chan K, Saltelli A, Tarantola S. Sensitivity analysis of model output: Variance-based methods make the difference. In: Proceedings of the 1997 winter simulation conference, 1997.
[13] Hall J, Boyce S, Wang Y, Dawson R, Tarantola S, Saltelli A. Sensitivity analysis for hydraulic models. Journal of Hydraulic Engineering 2009;135(11): 959–69.
[14] Allaire D, Willcox K. Surrogate modeling for uncertainty assessment with application to aviation environmental system models. AIAA Journal 2010;48(8): 1791–803.
[15] Chen W, Jin R, Sudjianto A. Analytical global sensitivity analysis and uncertainty propagation for robust design. Journal of Quality Technology 2006;38(4):333–48.
[16] Sobol' I. Sensitivity estimates for nonlinear mathematical models. Mathematical Modeling and Computational Experiment 1993;1:407–14.
[17] Oakley J, O'Hagan A. Probabilistic sensitivity analysis of complex models: A Bayesian approach. Journal of the Royal Statistical Society: Series B 2004;66(3):751–69.
[18] Beckman R, McKay M. Monte Carlo estimation under different distributions using the same simulation. Technometrics 1987;29(2):153–60.
[19] Ishigami T, Homma T. An importance quantification technique in uncertainty analysis for computer models. In: Proceedings of the ISUMA'90, first international symposium on uncertainty modeling and analysis, University of Maryland, December 3–5, 1990.
[20] Sobol' I. Theorems and examples on high dimensional model representation. Reliability Engineering and System Safety 2003;79:187–93.
[21] Saltelli A, Bolado R. An alternative way to compute Fourier amplitude sensitivity test (FAST). Computational Statistics and Data Analysis 1998;26: 445–60.
[22] Saltelli A, Annoni P, Azzini I, Campolongo F, Ratto M, Tarantola S. Variance based sensitivity analysis of model output, design and estimator for the total sensitivity index. Computer Physics Communications 2010;181:259–70.
[23] Lewandowski D, Cooke R, Tebbens R. Sample-based estimation of correlation ratio with polynomial approximation. ACM Transactions on Modeling and Computer Simulation 2007;5:1–17.
[24] Tarantola S, Gatelli D, Mara T. Random balance designs for the estimation of first order global sensitivity indices. Reliability Engineering and System Safety 2006;91(6):717–27.
[25] Ratto M, Pagano A, Young P. State dependent parameter metamodelling and sensitivity analysis. Computer Physics Communications 2007;177:863–76.
[26] Storlie C, Helton J. Multiple predictor smoothing methods for sensitivity analysis: description of techniques. Reliability Engineering and System Safety 2007;93:28–54.
[27] Ayyub B, Klir G. Uncertainty modeling and analysis in engineering and the sciences. Boca Raton, FL: Taylor & Francis Group; 2006.
[28] DeGroot M, Schervish M. Probability and statistics.3rd ed. Boston, MA: Addison-Wesley; 2002.
[29] Wang X. Improving the rejection sampling method in quasi-Monte Carlo methods. Journal of Computational and Applied Mathematics 2000;114: 231–46.
[30] Sobol' IM. Uniformly distributed sequences with an additional uniform property. USSR Computational Mathematics and Mathematical Physics 1977;16:236–42.
[31] Ratto M, Pagano A, Young P. Non-parametric estimation of conditional moments for sensitivity analysis. Reliability Engineering and System Safety 2009;94:237–43.
[32] Saltelli A, Tarantola S, Campolongo F, Ratto M. Sensitivity analysis in practice. West Sussex, England: John Wiley & Sons, Ltd.; 2004.
[33] Eldred M, Swiler L. Towards goal-oriented stochastic design employing adaptive collocation methods. In: Proceedings of the 13th AIAA ATIO/ISSMO conference, No. AIAA-2010-9125, Fort Worth, TX, September 13–15, 2010.
[34] Storlie C, Swiler L, Helton J, Sallaberry C. Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models. Reliability Engineering and System Safety 2009;94:1735–63.