



HAL
open science

A bias correction method for the estimation of sensitivity indices based on random balance designs

Jean-Yves Tissot, Clémentine Prieur

► **To cite this version:**

Jean-Yves Tissot, Clémentine Prieur. A bias correction method for the estimation of sensitivity indices based on random balance designs. 2010. hal-00507526v1

HAL Id: hal-00507526

<https://hal.science/hal-00507526v1>

Preprint submitted on 30 Jul 2010 (v1), last revised 1 Nov 2011 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A bias correction method for the estimation of sensitivity indices based on random balance designs

Jean-Yves Tissot^{a,*}, Clémentine Prieur^a

^a*Joseph Fourier University, LJK/MOISE, BP 53, 38041 Grenoble cedex, France*

Abstract

In this paper, we present a general way to correct a positive bias which occurs in all the estimators in random balance design method (RBD) and in its hybrid version, RBD-FAST. Both these techniques derive from Fourier amplitude sensitivity test (FAST) and, as a consequence, are faced with most of its inherent issues. And up to now, one of these, the well-known problem of interferences, has always been ignored in RBD. After presenting in which way interferences lead to a positive bias in the estimator of first-order sensitivity indices in RBD, we explain how to overcome this issue. We then extend the bias correction method to the estimation of sensitivity indices of any order in RBD-FAST. We also give an economical strategy to estimate all the first-order and second-order sensitivity indices using RBD-FAST.

Keywords: global sensitivity analysis, sensitivity indices, random balance design, RBD-FAST, bias correction

1. Introduction

Global sensitivity analysis of a model output consists in quantifying the respective importance of input factors over their entire range of values. Many techniques have been developed in this field (see [1] for a review), and one of

*Corresponding author. Email: jean-yves.tissot@imag.fr, phone: +33 (0)4 76 63 54 47, fax: +33 (0)4 76 63 12 63.

the most robust is built on the general decomposition of variance¹ proposed by Sobol' in the early nineties² [5].

Decomposition of variance and sensitivity indices. Given a mathematical explicit model, $Y = f(X_1, \dots, X_p)$ where the input factors X_i are independent scalar random variables and f is any square-integrable scalar function — e.g. nonlinear or nonmonotonic —, we can define an exhaustive group of indices measuring the relative effect of input variables or combinations thereof. Indeed, the total variance V of the model output Y can be decomposed as:

$$V = \sum_{k=1}^p \sum_{1 \leq i_1 < \dots < i_k \leq p} V_{i_1 \dots i_k} \quad (1)$$

where

$$V_i \triangleq \text{Var}(E(Y|X_i)) , \quad (2)$$

$$V_{ij} \triangleq \text{Var}(E(Y|X_i, X_j)) - V_i - V_j \quad (3)$$

and so on. Thus, dividing both sides of (1) by V , we obtain a positive and normalized decomposition of the global variations of the function f ,

$$1 = \sum_{k=1}^p \sum_{1 \leq i_1 < \dots < i_k \leq p} S_{i_1 \dots i_k} \quad (4)$$

where the $S_{i_1 \dots i_k}$ are the so-called sensitivity indices — or Sobol' indices —; in particular, the first-order sensitivity indices,

$$S_i \triangleq \frac{V_i}{V} , \quad 1 \leq i \leq p , \quad (5)$$

describe the main effect of each input factor, and the second-order sensitivity indices,

$$S_{ij} \triangleq \frac{V_{ij}}{V} , \quad 1 \leq i < j \leq p , \quad (6)$$

¹Variance-based methods constitute an important component of global sensitivity analysis. Alternative approaches are nevertheless possible; one can cite, in particular, a recent work due to Sobol' and Kucherenko [2] providing a new global sensitivity measure based on derivatives.

²This decomposition has a long history; one of the earliest work related to this topic is due to Hoeffding [3] (see [4] for further historical details).

measure the part of variance due to both input variables X_i and X_j .

In the case of an additive model — i.e. $f(X_1, \dots, X_p) = \sum_{k=1}^p f_k(X_k)$ — all terms but the first-order sensitivity indices are null and we then obtain a full decomposition with only S_1, \dots, S_p . On the contrary, if f is a non-additive function, it is necessary to evaluate higher-order terms to point out which interactions are significant. However, in practice, numerical models can't reach a high degree of complexity — especially in high dimension — and most of the time, it is sufficient to compute only the second-order indices to get a good overview of interactions.

The FAST method and its developments. Different methods have been proposed to estimate this kind of indices based on variance, and FAST, introduced in the seventies, was one of the earliest. The three introduction papers [6–8] describe how to compute main effects — i.e. first-order sensitivity indices³ — exploiting Weyl's ergodic theorem⁴ [12]. Then, in the review article [13], the authors precise the underlying theory considering multiple Fourier series; it allows, in particular, to define higher-order sensitivity indices estimators and to suggest a decomposition of variance (see formula (2.29) in [13]). But, in practice, because of various sources of error — especially the one due to interferences — FAST has only been applied to the estimation of first-order and total⁵ sensitivity indices in small dimension.

The RBD and Hybrid FAST-RBD (HFR) methods, introduced in 2006 by Tarantola *et al.* [15], partially overcome the inherent drawbacks of FAST using a new kind of sampling technique based on Satterthwaite's random balance de-

³In 1998, Saltelli and Bolado [9] discussed the equivalence between first-order sensitivity indices estimated with Sobol' method — i.e. Monte Carlo method — and main effects computed with FAST, and proved the identity of their prediction on two test cases. Even if there is no theoretical result on this issue, it is commonly admitted that these different techniques estimate the same statistical quantity.

⁴From a discrete point of view, it would probably be better to use the famous Weyl's equidistribution criterion ([10], in german). For an english reference, see for example [11].

⁵See the EFAST method due to Saltelli *et al.* [14].

signs [16]. However, as the authors state, both these new methods only allow to estimate first-order terms, and it remains a specific issue due to the interferences leading to an over-evaluation of the sensitivity indices, especially at low sample size. This last problem has never been solved. Nevertheless, in 2009, Mara [17] notices that interaction terms estimators can be defined using the hybrid approach. More precisely, he explains how to compute the sensitivity index of any group of input variables using the HFR method (renamed RBD-FAST).

In section 2, we briefly recall the FAST method and discuss the different sources of error of its estimators. We then present, in section 3, the specific problem of interferences in RBD which leads to the positive bias of the first-order sensitivity indices and we propose a correction method to overcome this issue. In section 4, we extend this technique to the estimators of sensitivity indices of any order in RBD-FAST, and in section 5, we give an economical strategy to estimate, under a rather general assumption, all the first-order and second-order sensitivity indices using RBD-FAST. Numerical examples are presented in section 6 to illustrate the accuracy of the proposed bias correction method and, conclusions and ideas for a future work are summarised in section 7.

2. Sources of error in the FAST method

The FAST method is based on a sequence of approximations which generates various error terms. Let us detail the four main sources of error which affect the estimation of sensitivity indices by FAST.

2.1. Description of the FAST method

The FAST method is based on a specific experimental design — the so-called search curve — which allows to make use of techniques relative to the Fourier series. The sample points $(x^k)_{k=1\dots N}$ are defined as

$$\forall i = 1, \dots, p, \forall k = 1, \dots, N, \quad x_i^k = G_i(\sin(\omega_i s_k + \varphi_i)) , \quad (7)$$

where G_i are functions to be settled to impose probability density functions on the input variables X_i , ω_i are integer frequencies — free of interferences up to a certain order —, φ_i are random phase-shifts and $(s_k)_{k=1\dots N}$ is the discrete parametrical variable defined as

$$\forall k = 1, \dots, N, \quad s_k = \frac{2\pi(k-1)}{N}. \quad (8)$$

In particular, to get uniform in $[0, 1]$ marginal distributions, one shall use (see for example [14]),

$$G_i = \frac{1}{\pi} \arcsin + \frac{1}{2}. \quad (9)$$

[Figure 1 about here.]

The set of image points $(f(x_1^j, \dots, x_p^j))_{j=1\dots N}$, obtained from the experimental design defined by (7) is then considered as a uniformly sampled signal. Hence, decomposing the finite spectrum of this discrete signal with respect to the frequencies $\omega_1, \dots, \omega_p$, it is possible to define the estimators of the different parts of variance as

$$\widehat{V} = \sum_{1 \leq |n| \leq N/2} |c_n|^2, \quad (10)$$

$$\widehat{V}_i = \sum_{1 \leq |k| \leq N_h} |c_{k\omega_i}|^2, \quad (11)$$

$$\widehat{V}_{ij} = \sum_{2 \leq |k|+|l| \leq N_{lc2}} |c_{k\omega_i+l\omega_j}|^2, \quad (12)$$

and so on; where N_h is the highest harmonic considered as non-negligible and N_{lc2} is the value over which the linear combinations of ω_i and ω_j are considered as negligible, and where

$$\text{for all } -\frac{N}{2} \leq n \leq \frac{N}{2}, \quad c_n = \frac{1}{N} \sum_{j=1}^N f(x_1^j, \dots, x_p^j) e^{-in \frac{2\pi(j-1)}{N}} \quad (13)$$

are the discrete Fourier coefficients. Finally, dividing (11) (resp. (12)) by (10), we get the estimator of a first-order (resp. second-order) sensitivity index:

$$\widehat{S}_i = \frac{\sum_{1 \leq |k| \leq N_h} |c_{k\omega_i}|^2}{\sum_{1 \leq |n| \leq N/2} |c_n|^2}, \quad (14)$$

$$\widehat{S}_{ij} = \frac{\sum_{2 \leq |k|+|l| \leq N_{lc2}} |c_{k\omega_i+l\omega_j}|^2}{\sum_{1 \leq |n| \leq N/2} |c_n|^2}. \quad (15)$$

2.2. Sources of error

The accuracy of these estimators is naturally function of the sample size and we can observe an empirical convergence to the theoretical values as N tends to $+\infty$. But, this dependence with respect to the sample size is intricate; we distinguish four main sources of problems generating an error which vanishes as N tends to $+\infty$.

The experimental design. We can notice that, even if it is possible to impose any probability density functions to the marginal distributions using the functions G_i , the search curve has bad space-filling properties and doesn't respect the joint probability density function (see figure 1). As a consequence, this not optimal experimental design leads to some error in terms of bias and variance.

The interferences. In the decomposition of the finite spectrum, if there exists a linear combination of the frequencies $\omega_1, \dots, \omega_p$ which is equal to zero, some parts of variance could be attributed by error to any estimator. For example, if $-2\omega_1 + \omega_2 = 0$, the discrete Fourier coefficient $c_{2\omega_1} = c_{\omega_2}$ contains information of both X_1 and X_2 , and should not be totally attributed to both \widehat{S}_1 and \widehat{S}_2 . These interferences can lead to an over-evaluation of all the sensitivity indices, and in order to keep low these positive biases we adopt the criterion proposed by Schaibly and Shuler [7] to get a set of frequencies free of interferences up to

a certain order M ,

$$\sum_{i=1}^p a_i \omega_i \neq 0 \text{ for } \sum_{i=1}^p |a_i| \leq M + 1 . \quad (16)$$

The aliasing. Only the linear combinations $\omega = \sum_{i=1}^p a_i \omega_i$ that lie inside the finite spectrum — i.e. between $-N/2$ and $N/2$ — are unambiguously represented by the discrete sampled signal. If ω lies out of this range, its spectral component is falsely attributed to another frequency inside the finite spectrum. To avoid, in part, this aliasing phenomenon, which can lead to positive biased estimators, it is necessary to follow the Nyquist-Shannon theorem, i.e. to impose that the sampling rate is large enough. As a consequence, the sample size is bounded from below as follows,

$$N \geq 2N_h \max_{1 \leq i \leq p} \omega_i . \quad (17)$$

The truncation. The different finite sums defining the estimators of the total variance and the partial variances in formulas (10), (11) and (12) should be in fact infinite series. But, as the spectrum only contains a finite number of frequencies, it is necessary to consider truncated sums. As a consequence, all the estimators being sums of squares, the truncations lead to an under-evaluation of all these quantities. These negative biases naturally vanish as the spectrum becomes larger, i.e. as N tends to $+\infty$.

3. The random balance design method

Keeping the underlying theory of FAST, but using a new sampling method, RBD allows to get round most of the issues inherent to FAST. However, the problem of interferences remains in a sense clarified further, and it is necessary to apply a specific correction method to eliminate the induced positive biases.

3.1. The sampling method

Contrarily to FAST, all the ω_i , in RBD, are equal to a unique frequency ω and the way to differentiate input variables consists in taking random permutations of the coordinates of the sample points. Let π_1, \dots, π_p be random

permutations on the set $\{1, \dots, N\}$, x^k are defined as

$$x_i^k = G_i(\sin(\omega s_{\pi_i(k)})), \quad \forall i = 1, \dots, p \quad \text{and} \quad \forall k = 1, \dots, N. \quad (18)$$

Because of symmetries in sine, one shall choose an odd integer N to get a good space-filling design. In this case, RBD technique is very close to Latin hypercube sampling introduced in 1979 (see [18]); the only difference is that, in RBD, sample points are located in the center of the cells (see figure 2).

[Figure 2 about here.]

3.2. The estimator

Proceeding with RBD sampling method, we can estimate the first-order sensitivity indices; the estimator of the total variance is defined as in FAST and the part of variance due to one factor is estimated by

$$\widehat{V}_i = \sum_{1 \leq |k| \leq N_h} |c_{k\omega}^{\pi_i}|^2 \quad (19)$$

with

$$c_{k\omega}^{\pi_i} = \frac{1}{N} \sum_{j=1}^N f(x_1^{\pi_i^{-1}(j)}, \dots, x_p^{\pi_i^{-1}(j)}) e^{-ik\omega \frac{2\pi(j-1)}{N}}. \quad (20)$$

Indeed, considering a fixed i , the sample points $(x_1^{\pi_i^{-1}(j)}, \dots, x_p^{\pi_i^{-1}(j)})_{j=1 \dots N}$ are such that the i^{th} coordinate is sampled with respect to the frequency ω and the other ones are sampled in a chaotic way because

$$\begin{aligned} x_k^{\pi_i^{-1}(j)} &= G_k \left(\sin(\omega s_{\pi_k(\pi_i^{-1}(j))}) \right) \\ &= \begin{cases} G_k(\sin(\omega s_j)) & \text{if } k = i \\ G_k(\sin(\omega s_{\pi_k^i(j)})) & \text{if } k \neq i, \end{cases} \end{aligned} \quad (21)$$

where $\pi_k^i = \pi_k \circ \pi_i^{-1}$ is almost surely a non-trivial permutation. As a consequence, in the finite spectrum — contained in $[-\frac{N}{2}, \frac{N}{2}]$ — of the signal

$$\left(f(x_1^{\pi_i^{-1}(j)}, \dots, x_p^{\pi_i^{-1}(j)}) \right)_{j=1 \dots N}, \quad (22)$$

the harmonic of ω are only related to X_i . Thus, using FAST's estimator, we get formulas (19) and (20).

Remark 1. *The choice of the frequency ω seems to be of secondary importance; but, to avoid aliasing, the most efficient value is the smallest one, typically $\omega = 1$. In this case, the aliasing phenomenon is negligible and consequently, there is no more restriction on the sample size like in formula (17).*

3.3. The bias

In order to understand the underlying phenomenon which leads to the positive bias, it is necessary to precise the composition of the finite spectrum of the signal defined in (22). As we explained in the last section, the part of variance due to X_i weights the harmonics of ω , but it is also essential to notice that, because of chaotic sampling of the $(X_k)_{k \neq i}$, the rest of variance — denoted V_{-i} — spreads on the whole finite spectrum. In particular, the rest of variance weights the harmonic of ω too, and these interferences lead to an over-evaluation of V_i .

The positive bias is denoted B_i and is defined as the part of V_{-i} located on the harmonics,

$$-N_h\omega, -(N_h - 1)\omega, \dots, -\omega, \omega, \dots, (N_h - 1)\omega, N_h\omega. \quad (23)$$

It obviously depends on the sample points but chaotic sampling implies that no frequency is favoured; so in mean⁶, V_{-i} is equally shared on the whole finite spectrum. Thus, for an odd integer N , we define the estimator of B_i as

$$\widehat{B}_i = \frac{2N_h}{N} \widehat{V}_{-i}, \quad (24)$$

and the corrected estimator of V_i as

$$\widehat{V}_i^c = \widehat{V}_i - \widehat{B}_i. \quad (25)$$

Thus, to get an explicit solution, it is necessary to define the estimator \widehat{V}_{-i} in (24). An efficient and economical way to get it consists in noticing that V_{-i} is

⁶i.e. when the experience is repeated, with a fixed sample size N , using different permutations π_i which are assumed to be the result of a uniform sampling inside the set of permutations on $\{1, 2, \dots, N\}$.

the complementary variance of V_i , i.e.

$$\widehat{V}_{-i} = V - \widehat{V}_i ; \quad (26)$$

so, assuming the bias correction, it comes

$$\widehat{V}_{-i}^c = \widehat{V} - \widehat{V}_i^c . \quad (27)$$

As a consequence, the corrected estimator of the part of variance is such that

$$\widehat{V}_i^c = \widehat{V}_i - \frac{2N_h}{N} (\widehat{V} - \widehat{V}_i^c) ; \quad (28)$$

and dividing both sides of the equality by \widehat{V} , we obtain

$$\widehat{S}_i^c = \widehat{S}_i - \frac{2N_h}{N} (1 - \widehat{S}_i^c) , \quad (29)$$

where \widehat{S}_i et \widehat{S}_i^c are respectively the RBD estimator of the first-order sensitivity index and the corrected one. Finally, setting $\lambda = \frac{2N_h}{N}$, we get the explicit formula

$$\widehat{S}_i^c = \widehat{S}_i - \frac{\lambda}{1 - \lambda} (1 - \widehat{S}_i) . \quad (30)$$

Remark 2. *It is important to notice that,*

- (i) *the more N is large, the less the bias is significant,*
- (ii) *the more S_i is large, the less the bias is significant.*

4. The hybrid approach: RBD-FAST

The underlying idea in RBD-FAST is to combine both the RBD and FAST sampling approaches. Therefore, this new method is naturally faced with the classical issues of FAST, but in a lesser extent. The main advantage in the hybrid approach is that the estimation of interaction terms is possible, but because of the same problem of interferences as in RBD, it is necessary to perform a bias correction method.

4.1. The sampling method

First, the p input variables are divided into groups of approximatively equal cardinality. Then free of interferences frequencies are allocated within each group of factors and random permutations are applied on each group. For example, we can have the following configurations,

$$\begin{aligned}
\text{6 factors: } & X_1 X_2 X_3 X_4 X_5 X_6 \\
& \underbrace{\omega_1 \ \omega_2 \ \omega_3}_{\pi_1} \quad \underbrace{\omega_1 \ \omega_2 \ \omega_3}_{\pi_2} \\
\text{6 factors: } & X_1 X_2 X_3 X_4 X_5 X_6 \\
& \underbrace{\omega_1 \ \omega_2}_{\pi_1} \quad \underbrace{\omega_1 \ \omega_2}_{\pi_2} \quad \underbrace{\omega_1 \ \omega_2}_{\pi_3} \\
\text{7 factors: } & X_1 X_2 X_3 X_4 X_5 X_6 X_7 \\
& \underbrace{\omega_1 \ \omega_2 \ \omega_3}_{\pi_1} \quad \underbrace{\omega_1 \ \omega_2}_{\pi_2} \quad \underbrace{\omega_1 \ \omega_2}_{\pi_3} .
\end{aligned}$$

Remark 3. *Tarantola et al. [15] and Mara [17] present RBD-FAST (or HFR) in another way: the p input variables are partitioned in the same way but the permutations are applied within the groups and a different frequency is associated to each group. Actually, the methods are strictly equivalent; it is just two different points of view.*

4.2. The estimators

As Mara explained in [17], this hybrid sampling method allows to define the estimator of sensitivity indices of any group of input variables. In particular, considering two factors inside the m^{th} group and respectively associated with the frequencies ω_i and ω_j , we can define the part of variance of their interaction as,

$$\widehat{V}_{ij} = \sum_{2 \leq |k|+|l| \leq N_{lc2}} |c_{k\omega_i+l\omega_j}^{\pi_m}|^2 \quad (31)$$

where N_{lc2} is the value over which the linear combinations of ω_i and ω_j are considered as negligible and with

$$c_{k\omega_i+l\omega_j}^{\pi_m} = \frac{1}{N} \sum_{n=1}^N f(x_1^{\pi_m^{-1}(n)}, \dots, x_p^{\pi_m^{-1}(n)}) e^{-i(k\omega_i+l\omega_j) \frac{2\pi(n-1)}{N}} . \quad (32)$$

In the same way, considering a factor inside the m^{th} group and associated with the frequency ω_i , we can define its part of variance as,

$$\widehat{V}_i = \sum_{1 \leq |k| \leq N_h} |c_{k\omega_i}^{\pi_m}|^2 \quad (33)$$

where N_h is the highest harmonic considered as non-negligible and with

$$c_{k\omega_i}^{\pi_m} = \frac{1}{N} \sum_{n=1}^N f(x_1^{\pi_m^{-1}(n)}, \dots, x_p^{\pi_m^{-1}(n)}) e^{-ik\omega_i \frac{2\pi(n-1)}{N}}. \quad (34)$$

Indeed, considering the sample points $(x_1^{\pi_m^{-1}(j)}, \dots, x_p^{\pi_m^{-1}(j)})_{j=1 \dots N}$ where m is fixed, for $1 \leq k \leq p$, we have,

(i) if X_k is associated with the couple (ω_i, π_m) then

$$x_k^{\pi_m^{-1}(j)} = G_k \left(\sin(\omega_i s_{\pi_m(\pi_m^{-1}(j))}) \right) = G_k(\sin(\omega_i s_j)) \quad (35)$$

(ii) if X_k is associated with a couple (ω_i, π_n) , ($n \neq m$) then

$$x_k^{\pi_m^{-1}(j)} = G_k \left(\sin(\omega_i s_{\pi_n(\pi_m^{-1}(j))}) \right) \quad (36)$$

where $\pi_n \circ \pi_m^{-1}$ is almost surely a non-trivial permutation. As a consequence, all the input variables outside the group associated with π_m are sampled in a chaotic way, and the other ones are sampled with respect to their frequencies. Applying FAST's estimator, formulas (31)–(34) follow.

4.3. The bias

The phenomenon leading to positive biases described for the RBD method occurs in the same way for RBD-FAST. As a consequence, parts of variance can be corrected with an analogous technique.

Let X_{m_1}, \dots, X_{m_d} be the d input factors inside the m^{th} group, and P be a nonempty subset of $\{m_1, \dots, m_d\}$. We denote V_P the part of variance due to all the input variables $(X_i)_{i \in P}$, i.e. the term due to the interaction of all the $(X_i)_{i \in P}$, if $\text{card}(P) > 1$, or the term due to the only effect of X_{i_0} , if $P = \{i_0\}$. Let \widehat{V}_P be the RBD-FAST classical estimator of V_P , previously detailed

in formulas (31) and (33) for $\text{card}(P) = 1$ or 2 . We first define the estimator of the positive bias B_P as

$$\widehat{B}_P = \frac{\#P}{N} \widehat{V}_{-\mathcal{G}(P)} \quad (37)$$

and the corrected estimator of V_P as

$$\widehat{V}_P^c = \widehat{V}_P - \widehat{B}_P \quad (38)$$

where $\#P$ is the number of frequencies taken into account to estimate V_P — for example, $\#P = 2N_h$ if $\text{card}(P) = 1$ and $\#P = 2N_{lc2}(N_{lc2} - 1)$ if $\text{card}(P) = 2$ — and $\widehat{V}_{-\mathcal{G}(P)}$ is the part of variance not due to any subset of factors contained in the group P belongs to, $\mathcal{G}(P) = \{m_1, \dots, m_d\}$. Assuming the bias correction, it comes

$$\widehat{V}_{-\mathcal{G}(P)} = \widehat{V} - \sum_{\substack{Q \subset \mathcal{G}(P) \\ Q \neq \emptyset}} \widehat{V}_Q^c. \quad (39)$$

As a consequence, we have

$$\widehat{V}_P^c = \widehat{V}_P - \frac{\#P}{N} \left(\widehat{V} - \sum_{\substack{Q \subset \mathcal{G}(P) \\ Q \neq \emptyset}} \widehat{V}_Q^c \right), \quad (40)$$

and dividing both sides of the equality by \widehat{V} , we get

$$\widehat{S}_P^c = \widehat{S}_P - \frac{\#P}{N} \left(1 - \sum_{\substack{Q \subset \mathcal{G}(P) \\ Q \neq \emptyset}} \widehat{S}_Q^c \right), \quad (41)$$

where \widehat{S}_P and \widehat{S}_P^c are respectively the RBD-FAST estimator of the sensitivity index S_P and the corrected one. Then, setting

$$\lambda_P = \frac{\#P}{N}, \quad \text{for any nonempty subset } P \in \{m_1, \dots, m_d\} \quad (42)$$

and

$$\bar{\lambda} = \sum_{\substack{Q \subset \mathcal{G}(P) \\ Q \neq \emptyset}} \lambda_Q, \quad (43)$$

we conclude with the explicit formula

$$\widehat{S}_P^c = \widehat{S}_P - \frac{\lambda_P}{1 - \bar{\lambda}} \left(1 - \sum_{\substack{Q \subset \mathcal{G}(P) \\ Q \neq \emptyset}} \widehat{S}_Q^c \right) \quad (\text{see details in Appendix A}). \quad (44)$$

Remark 4. *This bias correction formula requires the knowledge of the biased estimators \widehat{S}_Q of any order relative to the input factors $(X_i)_{i \in P}$. Unfortunately, the estimation of the terms over a certain order is quite difficult; so in practice, it is necessary to neglect sensitivity indices over a certain degree δ and to consider the following bias correction,*

$$\widehat{S}_P^c = \widehat{S}_P - \frac{\lambda_P}{1 - \bar{\lambda}} \left(1 - \sum_{\substack{Q \subset \mathcal{G}(P), Q \neq \emptyset \\ \text{card}(Q) \leq \delta}} \widehat{S}_Q \right) \quad (45)$$

where

$$\bar{\lambda} = \sum_{\substack{Q \subset \mathcal{G}(P), Q \neq \emptyset \\ \text{card}(Q) \leq \delta}} \lambda_Q . \quad (46)$$

5. An economical strategy to estimate all the first-order and second order sensitivity indices

Through this section, we develop a strategy using RBD-FAST to get all the bias-corrected estimates of the first-order and second-order sensitivity indices of a model in which we assume that the sensitivity indices over a certain order δ are negligible. In this case, we can get the first-order and second-order indices applying formulas (45) and (46).

However, contrarily to the RBD method in which all the main effects of any model can be estimated using only one experimental design, the computation of all the first-order and second-order indices using RBD-FAST requires a number of sample sets increasing with the number of factors p . Through an example, Mara [17] observes that 5 sample sets are necessary to estimate all the 15 second-order sensitivity indices — and naturally the first-order ones — of a 6-dimensional model. In fact, in the case of 6 input factors, the number of experimental designs can be restricted to 4. More generally, we establish that

the required number of experimental designs is equal to:

$$\begin{aligned}
1 + \min_{\substack{\sqrt{p} \leq q \\ q \text{ prime}}} q & \quad \text{for } p \geq 6 \\
3 & \quad \text{for } p = 4 \text{ or } 5 \\
1 & \quad \text{for } p \leq 3,
\end{aligned} \tag{47}$$

where p is the number of input factors. Low-dimensional models — $p \leq 3$ — can be treated using FAST method with only one design of experiments; in the other cases we implement a strategy based on elementary combinatory considerations.

5.1. Designs of experiments in the case $p = q^2$ with q prime

In this particular case, the different configurations of the designs of experiments required to estimate all the first-order and second-order sensitivity indices are quite natural. First, we divide the set of input variables $\{X_1, \dots, X_p\}$ into q groups of q factors; for example, in the case $p = 9$, we can have,

$$\text{configuration 0 : } \underbrace{X_4 X_1 X_5}_{G_1^0} \quad \underbrace{X_7 X_9 X_2}_{G_2^0} \quad \underbrace{X_3 X_8 X_6}_{G_3^0} . \tag{48}$$

Following RBD-FAST approach, allocating free of interferences frequencies within each group and applying a random permutation on each group, this configuration allows to estimate the second-order indices $S_{14}, S_{15}, S_{45}, S_{27}, S_{29}, S_{79}, S_{36}, S_{38}$ and S_{68} , and all the first-order terms.

We then obtain the other configurations applying the following rules:

- (R1) each of the new configurations is a partition of the input variables into q groups of q factors,
- (R2) each group in the new configurations is filled with one factor of each original group $(G_i^0)_{i=1\dots q}$,
- (R3) if a set of two distinct variables $\{X_i, X_j\}$ is already contained in a group of a certain configuration, G_n^k , then we are not allowed to define a group G_m^l in a next configuration containing both X_i and X_j .

As a consequence, in the case $p = 9$, it is only possible to create three new configurations, for example,

$$\begin{aligned}
\text{configuration 1 : } & \underbrace{X_9 X_4 X_8}_{G_1^1} \quad \underbrace{X_7 X_5 X_6}_{G_2^1} \quad \underbrace{X_3 X_2 X_1}_{G_3^1} \\
\text{configuration 2 : } & \underbrace{X_6 X_1 X_9}_{G_1^2} \quad \underbrace{X_3 X_7 X_4}_{G_2^2} \quad \underbrace{X_2 X_8 X_5}_{G_3^2} \\
\text{configuration 3 : } & \underbrace{X_7 X_1 X_8}_{G_1^3} \quad \underbrace{X_5 X_3 X_9}_{G_2^3} \quad \underbrace{X_6 X_2 X_4}_{G_3^3}
\end{aligned} \tag{49}$$

Then, it is easy to notice that these four configurations 0, 1, 2 and 3 allow to compute one estimate of all the second-order sensitivity indices and four estimates of all the first-order terms.

More generally, we have the proposition below.

Proposition 1. *In the case $p = q^2$ with q prime, it exists an economical strategy which, using $q + 1$ designs of experiments, allows to compute $q + 1$ estimates of all the first-order sensitivity indices and one estimate of all the second-order terms.*

Proof. See Appendix B.

5.2. Experimental designs for any p

In the general case, we first define

$$q^* = \min_{\substack{\sqrt{p} \leq q \\ q \text{ prime}}} q, \tag{50}$$

and

$$p^* = (q^*)^2. \tag{51}$$

Following the strategy detailed in the previous section, we can create $q + 1$ designs of experiments with p^* factors, $X_1, \dots, X_p, \dots, X_{p^*}$. We then delete the variables X_{p+1}, \dots, X_{p^*} in all the configurations. For example, considering an 8-dimensional model, we get $q^* = 3$ and $p^* = 9$. As a consequence, we can use the designs of experiments presented in formulas (48) and (49), and delete the

factor X_9 , it comes

$$\begin{aligned}
\text{configuration 0 : } & \underbrace{X_4 X_1 X_5}_{G_1^0} \quad \underbrace{X_7 X_2}_{G_2^0} \quad \underbrace{X_3 X_8 X_6}_{G_3^0} \\
\text{configuration 1 : } & \underbrace{X_4 X_8}_{G_1^1} \quad \underbrace{X_7 X_5 X_6}_{G_2^1} \quad \underbrace{X_3 X_2 X_1}_{G_3^1} \\
\text{configuration 2 : } & \underbrace{X_6 X_1}_{G_1^2} \quad \underbrace{X_3 X_7 X_4}_{G_2^2} \quad \underbrace{X_2 X_8 X_5}_{G_3^2} \\
\text{configuration 3 : } & \underbrace{X_7 X_1 X_8}_{G_1^3} \quad \underbrace{X_5 X_3}_{G_2^3} \quad \underbrace{X_6 X_2 X_4}_{G_3^3}
\end{aligned} \tag{52}$$

With $p = 5$, this technique leads to use four designs of experiments; but we had better consider the three following configurations

$$\begin{aligned}
\text{configuration 1 : } & \underbrace{X_1 X_2 X_3}_{G_1^1} \quad \underbrace{X_4 X_5}_{G_2^1} \\
\text{configuration 2 : } & \underbrace{X_1 X_2 X_4}_{G_1^2} \quad \underbrace{X_3 X_5}_{G_2^2} \\
\text{configuration 3 : } & \underbrace{X_1 X_2 X_5}_{G_1^3} \quad \underbrace{X_3 X_4}_{G_2^3} .
\end{aligned} \tag{53}$$

Finally, we have defined, for any p , an economical strategy in which the number of experimental designs satisfies formula (47).

Remark 5. *Elaborating economical strategies is also of major importance for the Sobol' method in which the curse of dimensionality is clearly problematic. In particular, one can cite the work of Saltelli [19] who provides an economical way to estimate all the first-order, second-order and total sensitivity indices using the Sobol' method.*

6. Numerical tests

The accuracy of the proposed bias correction method is tested on the g -function introduced by Sobol' (see e.g. [20]). Considering uniformly distributed independent input variables $(X_i)_{i=1,\dots,p}$ on the unit hypercube, the function is defined as

$$f(X_1, \dots, X_p) = \prod_{i=1}^p g_i(X_i) \tag{54}$$

where $g_i(X_i)$ is given by

$$g_i(X_i) = \frac{|4X_i - 2| + a_i}{1 + a_i}. \quad (55)$$

We consider an 8-dimensional model and we fix the parameters a_i as $\{0, 0, 0, 0.5, 0.5, 9, 9, 9\}$, so that the three first parameters are important, the two next ones are less important, the three last ones are negligible and interactions are quite important.

After testing the correction method on the estimation of the first-order sensitivity indices by RBD, we focus on the estimation of both first-order and second-order indices using the economical approach of RBD-FAST developed in the previous section.

6.1. Test on RBD

The correction method is tested at increasing sample sizes, $N = 501$ and $N = 2001$ (see figure 4). In both configurations, we estimate all the first-order sensitivity indices with the basic RBD method and with the corrected one. The experience is replicated 150 times on different random samples and boxplots are presented.

We can observe that the corrected boxplots are centered on the analytical values whatever the sample size. On the contrary, in the absence of correction method, the estimates are considerably biased — even at large sample size —, and we can notice that negligible input factors are not accurately identified, especially at low sample size (see X_6 , X_7 and X_8 in the case $N = 501$). More generally, defining the total quadratic error,

$$\varepsilon_{\text{TQE}} = \sqrt{\sum_{i=1}^p (\widehat{S}_i - S_i)^2}, \quad (56)$$

we can observe the nonnegligible gain of accuracy due to the bias correction method (see figure 5).

[Figures 3, 4 and 5 about here.]

6.2. Test on RBD-FAST

We now test the bias correction method on the estimation of the interaction terms. Applying the economical approach of RBD-FAST detailed in the previous section, we estimate all the first-order and second-order sensitivity indices using only 4 experimental designs — those presented in formula (52) — with a sample size equal to 4001. Following Remark 4, we neglect the importance of the third-order sensitivity indices — their part of variance is theoretically lower than 10% —, so we apply formulas (45) and (46) with $\delta = 2$.

We present the boxplots of 150 replicates (see figures 6, 7 and 9); the result show that, as in the previous test, the corrected indices are centered on their respective theoretical value; but some differences exist between main effects and interactions estimation. On the one hand, the first-order terms are accurately evaluated, and the bias, in the absence of correction, is rather low; on the other hand, the interactions estimates suffer from a more important variance and a larger bias in the absence of correction. Two main reasons justify the difference between the variance; firstly, the first-order terms are evaluated thanks to 4 estimates per indices though the second-order ones are computed with only one estimate, and secondly, the complexity of the sensitivity indices grows with the order. In terms of bias, the lower performance of the interactions estimation without correction is essentially due to the larger number of frequencies taken into account to evaluate the second-order indices. Indeed, considering formula (45), we can notice that the amplitude of the bias

$$\frac{\lambda_P}{1 - \bar{\lambda}} \left(1 - \sum_{\substack{Q \subset \mathcal{G}(P), Q \neq \emptyset \\ \text{card}(Q) \leq \delta}} \widehat{S}_Q \right) \quad (57)$$

is in proportion to $\lambda_P = \#P/N$; and, in this test, we have $\#P = 2N_h = 2 * 10 = 20$ for the first-order sensitivity indices, and $\#P = 2N_{lc2}(N_{lc2} - 1) = 2 * 9 * (9 - 1) = 144$ for the second-order sensitivity indices.

[Figures 6, 7, 8 and 9 about here.]

7. Conclusion

In this paper, we presented a bias correction method for the estimation of sensitivity indices of any order by both RBD and RBD-FAST. In particular, as we can notice through the numerical tests, this technique successfully overcomes the over-estimation of the first-order and second-order indices whatever the sample size.

We also introduced an economical strategy which, combined with the bias correction method, provides an efficient way to estimate all the first-order and second-order indices using RBD-FAST. In particular, this kind of approach allows to get a good overview of the sensitivity of a model at a low cost.

Finally, the source of error that remains is essentially due to the experimental design, and a further work would be to improve RBD and RBD-FAST sampling methods. In particular, optimization algorithms are commonly used for Latin hypercube sampling, and an idea would be to transpose these techniques to RBD experimental designs which are, as we have noticed in Section 3, very close to Latin hypercube designs. But, this is beyond the scope of the present paper.

Appendix A. Details on formula (44)

We denote $(P_i)_{i=1\dots n}$ the nonempty subsets of $\{m_1, \dots, m_d\}$ where n is such that

$$n = \sum_{k=1}^d \binom{d}{k} = 2^d - 1, \quad (\text{A.1})$$

and to simplify the notations, we denote λ_i the coefficients λ_{P_i} . Applying formula (41) to each of the P_i , we get the linear system,

$$\begin{pmatrix} \widehat{S_{P_1}} \\ \widehat{S_{P_2}} \\ \vdots \\ \widehat{S_{P_{n-1}}} \\ \widehat{S_{P_n}} \end{pmatrix} = \begin{pmatrix} 1 - \lambda_1 & \lambda_1 & \cdots & \cdots & -\lambda_1 \\ -\lambda_2 & 1 - \lambda_2 & -\lambda_2 & \cdots & -\lambda_2 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ -\lambda_{n-1} & \cdots & \cdots & 1 - \lambda_{n-1} & -\lambda_{n-1} \\ -\lambda_n & \cdots & \cdots & -\lambda_n & 1 - \lambda_n \end{pmatrix} \begin{pmatrix} \widehat{S_{P_1}^c} \\ \widehat{S_{P_2}^c} \\ \vdots \\ \widehat{S_{P_{n-1}}^c} \\ \widehat{S_{P_n}^c} \end{pmatrix} + \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_{n-1} \\ \lambda_n \end{pmatrix} \quad (\text{A.2})$$

The determinant Δ of the matrix of the system — denoted A — is easy to get. Subtracting the first column to all other ones, it comes

$$\Delta = \begin{vmatrix} 1 - \lambda_1 & -1 & \cdots & \cdots & -1 \\ -\lambda_2 & 1 & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ -\lambda_n & 0 & \cdots & 0 & 1 \end{vmatrix} \quad (\text{A.3})$$

and, using Laplace expansion, we get

$$\Delta = 1 - \lambda_1 - \lambda_2 \cdots - \lambda_n . \quad (\text{A.4})$$

In practice, we fix N such that

$$\sum_{i=1}^n \#P_i < N \quad (\text{A.5})$$

so, with the definition in formula (42), it comes

$$\sum_{i=1}^n \lambda_i < 1 . \quad (\text{A.6})$$

It implies that Δ is positive and, as a consequence, the matrix A is invertible.

We get A^{-1} using the formula based on the adjugate matrix,

$$A^{-1} = \frac{{}^t\text{adj}(A)}{\Delta}. \quad (\text{A.7})$$

We easily obtain,

$$\text{adj}(A) = \begin{pmatrix} \Delta + \lambda_1 & \lambda_2 & \cdots & \lambda_{n-1} & \lambda_n \\ \lambda_1 & \Delta + \lambda_2 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \Delta + \lambda_{n-1} & \lambda_n \\ \lambda_1 & \lambda_2 & \cdots & \lambda_{n-1} & \Delta + \lambda_n \end{pmatrix}. \quad (\text{A.8})$$

Finally, we invert the linear system (A.2), it comes

$$\begin{pmatrix} \widehat{S_{P_1}^c} \\ \widehat{S_{P_2}^c} \\ \vdots \\ \widehat{S_{P_{n-1}}^c} \\ \widehat{S_{P_n}^c} \end{pmatrix} = \begin{pmatrix} 1 + \frac{\lambda_1}{\Delta} & \frac{\lambda_1}{\Delta} & \cdots & \cdots & \frac{\lambda_1}{\Delta} \\ \frac{\lambda_2}{\Delta} & 1 + \frac{\lambda_2}{\Delta} & \frac{\lambda_2}{\Delta} & \cdots & \frac{\lambda_2}{\Delta} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \frac{\lambda_n}{\Delta} & \cdots & \cdots & \frac{\lambda_n}{\Delta} & 1 + \frac{\lambda_n}{\Delta} \end{pmatrix} \begin{pmatrix} \widehat{S_{P_1}} \\ \widehat{S_{P_2}} \\ \vdots \\ \widehat{S_{P_{n-1}}} \\ \widehat{S_{P_n}} \end{pmatrix} - \begin{pmatrix} \frac{\lambda_1}{\Delta} \\ \frac{\lambda_2}{\Delta} \\ \vdots \\ \frac{\lambda_{n-1}}{\Delta} \\ \frac{\lambda_n}{\Delta} \end{pmatrix} \quad (\text{A.9})$$

and it allows to conclude to formula (44).

Appendix B. Proof of Proposition 1

Let $p = q^2$ with q prime; it is obvious that if there exists $q + 1$ designs of experiments satisfying the rules established in section 5.1, then these configurations allow to compute $q + 1$ estimates of all the first-order sensitivity indices and one estimate of all the second-order terms. So, to show that the economical strategy exists, it is sufficient to prove the existence of such configurations under the rules (R1), (R2) and (R3) of section 5.1; let us give a constructive proof.

First, renaming the factors $(X_i)_{i=1\dots p}$, we define an initial configuration,

$$\text{configuration 0 : } \underbrace{X_1^1 \dots X_1^q}_{G_1^0} \quad \underbrace{X_2^1 \dots X_2^q}_{G_2^0} \quad \dots \quad \underbrace{X_q^1 \dots X_q^q}_{G_q^0} . \quad (\text{B.1})$$

We then obtain the q other experimental designs considering for i from 1 to q ,

$$\text{configuration } i : \quad \underbrace{X_1^{\sigma_i^1(1)} \dots X_q^{\sigma_i^q(1)}}_{G_1^i} \quad \underbrace{X_1^{\sigma_i^1(2)} \dots X_q^{\sigma_i^q(2)}}_{G_2^i} \quad \dots \quad \underbrace{X_1^{\sigma_i^1(q)} \dots X_q^{\sigma_i^q(q)}}_{G_q^i} , \quad (\text{B.2})$$

where for all $1 \leq i, j \leq q$, σ_i^j is a permutation on the set $\{1, \dots, q\}$. These configurations naturally satisfy rules (R1) and (R2); but (R3) is not always completed. However, we can observe that, letting c a cyclic permutation of order q , the permutations

$$\sigma_i^j = c^{i*j} \quad (\text{B.3})$$

allow to satisfy rule (R3). Indeed, following the formalism of formula (B.2), rule (R3) reads as

$$\forall 1 \leq i, i', k, k', j_1, j_2 \leq q, i \neq i', j_1 \neq j_2, \left\{ \begin{array}{l} \sigma_i^{j_1}(k) \neq \sigma_{i'}^{j_1}(k') \\ \text{or} \\ \sigma_i^{j_2}(k) \neq \sigma_{i'}^{j_2}(k') . \end{array} \right. \quad (\text{B.4})$$

So, assuming formula (B.3), let us prove

$$\forall 1 \leq i, i', k, k', j_1, j_2 \leq q, i \neq i', j_1 \neq j_2, \left\{ \begin{array}{l} c^{i*j_1}(k) \neq c^{i'*j_1}(k') \\ \text{or} \\ c^{i*j_2}(k) \neq c^{i'*j_2}(k') . \end{array} \right. \quad (\text{B.5})$$

Let us suppose, by contradiction, that

$$c^{i*j_1}(k) = c^{i'*j_1}(k') \quad \text{and} \quad c^{i*j_2}(k) = c^{i'*j_2}(k') \quad (\text{B.6})$$

for a certain 6-uplet (i, i', k, k', j_1, j_2) with $i \neq i'$ and $j_1 \neq j_2$. It implies that

$$c^{(i-i')*(j_1-j_2)}(k) = k . \quad (\text{B.7})$$

Then c being a cyclic permutation of order q with q prime and i being different from i' , we deduce that $c^{(i-i')}$ is a cyclic permutation of order q . As a consequence, it follows that $j_1 - j_2 = q * r$ for a certain integer r . But, assuming

$1 \leq j_1, j_2 \leq q$, we conclude that $r = 0$ and $j_1 = j_2$, which is in contradiction with our assumption $j_1 \neq j_2$. The conclusion follows.

Acknowledgments

This work has been partially supported by French National Research Agency (ANR) through COSINUS program (project COSTA-BRAVA n° ANR-09-COSI-015).

References

- [1] A. Saltelli, K. P. S. Chan, E. M. Scott, Sensitivity Analysis, John Wiley & Sons, 2000.
- [2] I. M. Sobol', S. Kucherenko, Derivative based global sensitivity measures and their link with global sensitivity indices, *Mathematics and Computers in Simulation* 79(10) (2009) 3009–3017.
- [3] W. F. Hoeffding, A class of statistics with asymptotically normal distributions, *Annals of Mathematical Statistics* 19 (1948) 293–325.
- [4] G. E. B. Archer, A. Saltelli, I. M. Sobol', Sensitivity measures, anova-like techniques and the use of bootstrap, *Journal of Statistical Computation and Simulation* 58 (1997) 99–120.
- [5] I. M. Sobol', Sensitivity analysis for nonlinear mathematical models, *Mathematical Modeling and Computational Experiment* 1 (1993) 407–414.
- [6] R. I. Cukier, C. M. Fortuin, K. E. Shuler, A. G. Petschek, J. H. Schaibly, Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients: Theory, *Journal of Chemical Physics* 59 (1973) 3873–3878.
- [7] J. H. Schaibly, K. E. Shuler, Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients: Applications, *Journal of Chemical Physics* 59 (1973) 3879–3888.

- [8] R. I. Cukier, J. H. Schaibly, K. E. Shuler, Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients: Analysis of the approximations, *Journal of Chemical Physics* 63 (1975) 1140–1149.
- [9] A. Saltelli, R. Bolado, An alternative way to compute fourier amplitude sensitivity test (FAST), *Computational Statistics & Data Analysis* 26 (1998) 445–460.
- [10] H. Weyl, Über ein problem aus dem gebiete der diophantischen approximationen, *Nachrichten der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* (1914) 234–244.
- [11] K. Chandrasekharan, *Introduction to Analytic Number Theory*, Springer-Verlag, 1968.
- [12] H. Weyl, Mean motion, *American Journal of Mathematics* 60 (1938) 889–896.
- [13] R. I. Cukier, H. B. Levine, K. E. Shuler, Nonlinear sensitivity analysis of multiparameter model systems, *Journal of Computational Physics* 26 (1978) 1–42.
- [14] A. Saltelli, S. Tarantola, K. P. S. Chan, A quantitative model-independent method for global sensitivity analysis of model output, *Technometrics* 41 (1999) 39–56.
- [15] S. Tarantola, D. Gatelli, T. A. Mara, Random balance designs for the estimation of first-order global sensitivity indices, *Reliability Engineering and System Safety* 91 (2006) 717–727.
- [16] F. E. Satterthwaite, Random balance experimentation, *Technometrics* 1(2) (1959) 111–137.
- [17] T. A. Mara, Extension of the rbd-fast method to the computation of global sensitivity indices, *Reliability Engineering and System Safety* 94 (2009) 1274–1281.

[18] M. D. McKay, W. J. Conover, R. J. Beckman, A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Reliability Engineering and System Safety* 21(2) (1979) 239–245.

[19] A. Saltelli, Making best use of model evaluations to compute sensitivity indices, *Computer Physics Communications* 145 (2002) 208–297.

[20] A. Saltelli, I. M. Sobol', About the use of rank transformation in sensitivity analysis model, *Reliability Engineering and System Safety* 50 (1995) 225–239.

List of Figures

1	Plot of a search curve with uniform marginals in a three-dimensional unit hypercube with a sample size of 301.	28
2	Plots of uniform sampling in a two-dimensional unit hypercube with a sample size of 15.	29
3	Boxplots of the first-order sensitivity indices by the RBD method with a sample size of 501. the analytical values are $S_1 = S_2 = S_3 = 0.1546$, $S_4 = S_5 = 0.0687$ and $S_6 = S_7 = S_8 = 0.0015$	30
4	Boxplots of the first-order sensitivity indices by the RBD method with a sample size of 2001. The analytical values are $S_1 = S_2 = S_3 = 0.1546$, $S_4 = S_5 = 0.0687$ and $S_6 = S_7 = S_8 = 0.0015$	31
5	Plots and boxplots (150 replicates) of total quadratic error against the sample size N.	32
6	Boxplots of the first-order sensitivity indices by the RBD-FAST method with a sample size of 4001. The analytical values are $S_1 = S_2 = S_3 = 0.1546$, $S_4 = S_5 = 0.0687$ and $S_6 = S_7 = S_8 = 0.0015$	33

7	Boxplots of the second-order sensitivity indices by the RBD-FAST method with a sample size of 4001. The analytical values are $S_{12} = S_{13} = S_{23} = 0.0515$, $S_{14} = S_{15} = S_{24} = S_{25} = S_{34} = S_{35} = 0.0229$	34
8	Boxplots of the second-order sensitivity indices by the RBD-FAST method with a sample size of 4001. The analytical values are $S_{45} = 0.0102$, $S_{16} = S_{17} = S_{18} = S_{26} = S_{27} = S_{28} = S_{36} = S_{37} = S_{38} = 0.0005$	35
9	Boxplots of the second-order sensitivity indices by the RBD-FAST method with a sample size of 4001. The analytical values are $S_{46} = S_{47} = S_{48} = S_{56} = S_{57} = S_{58} = 0.0002$, $S_{67} = S_{68} = S_{78} = 5.10^{-6}$	36

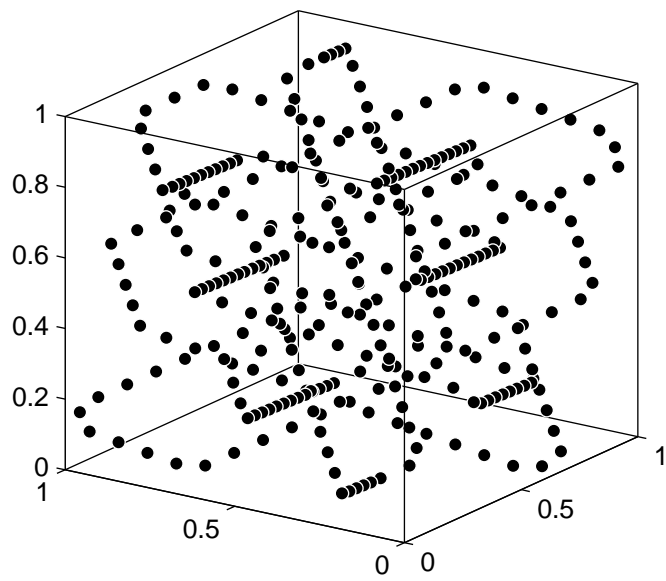


Figure 1: Plot of a search curve with uniform marginals in a three-dimensional unit hypercube with a sample size of 301.

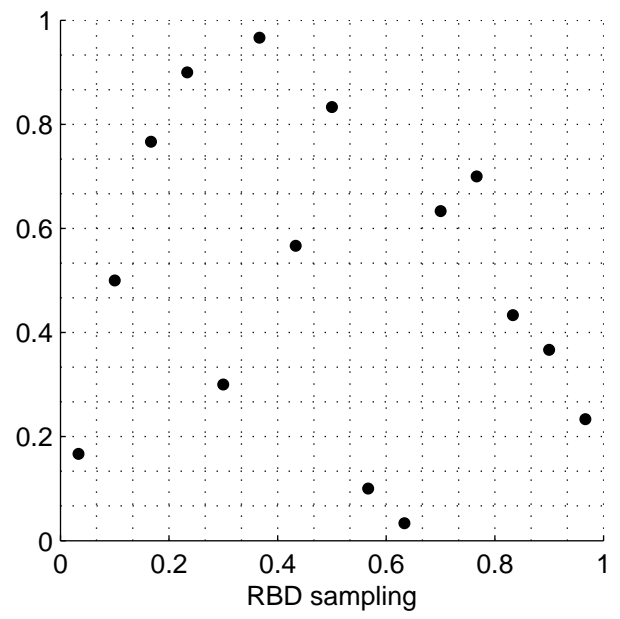
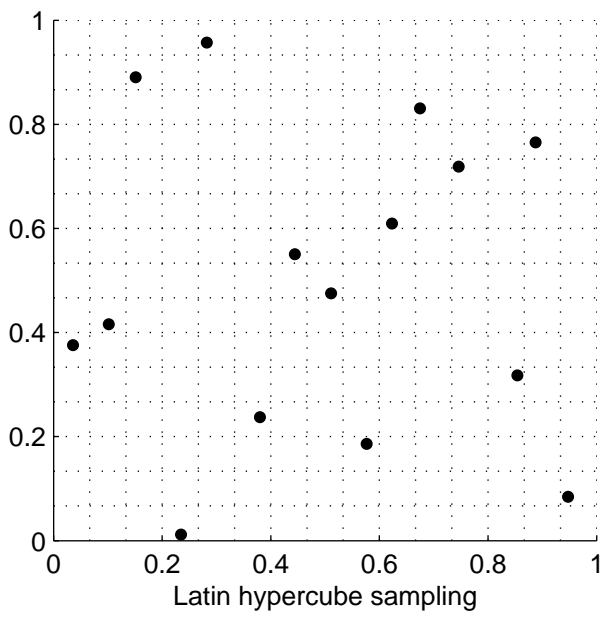


Figure 2: Plots of uniform sampling in a two-dimensional unit hypercube with a sample size of 15.

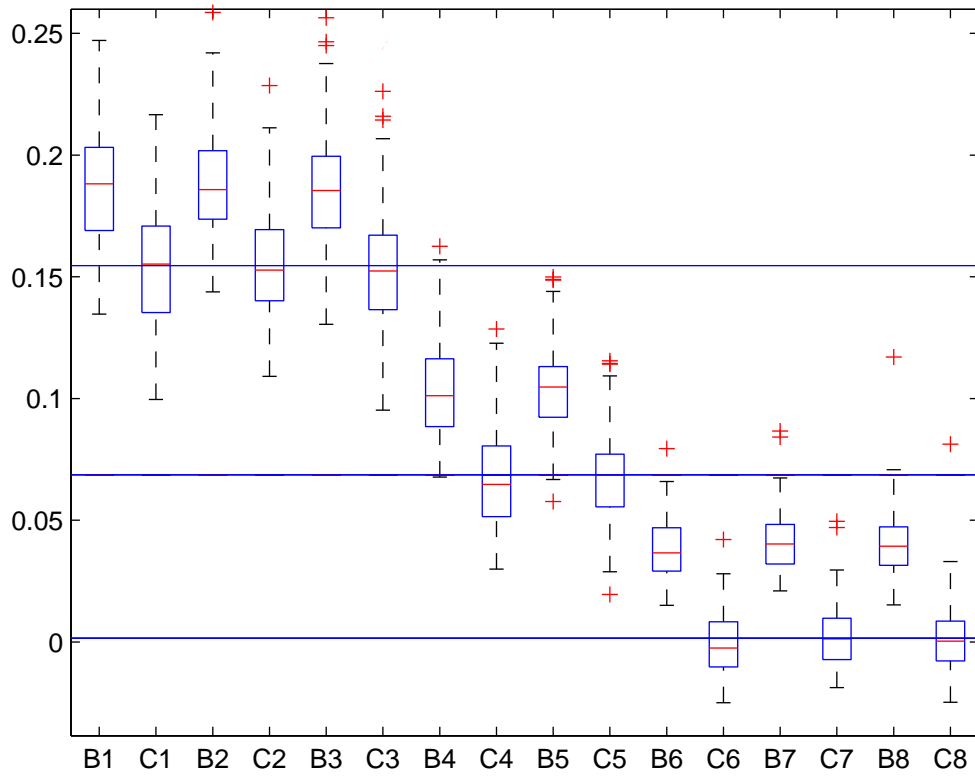


Figure 3: Boxplots of the first-order sensitivity indices by the RBD method with a sample size of 501. the analytical values are $S_1 = S_2 = S_3 = 0.1546$, $S_4 = S_5 = 0.0687$ and $S_6 = S_7 = S_8 = 0.0015$.

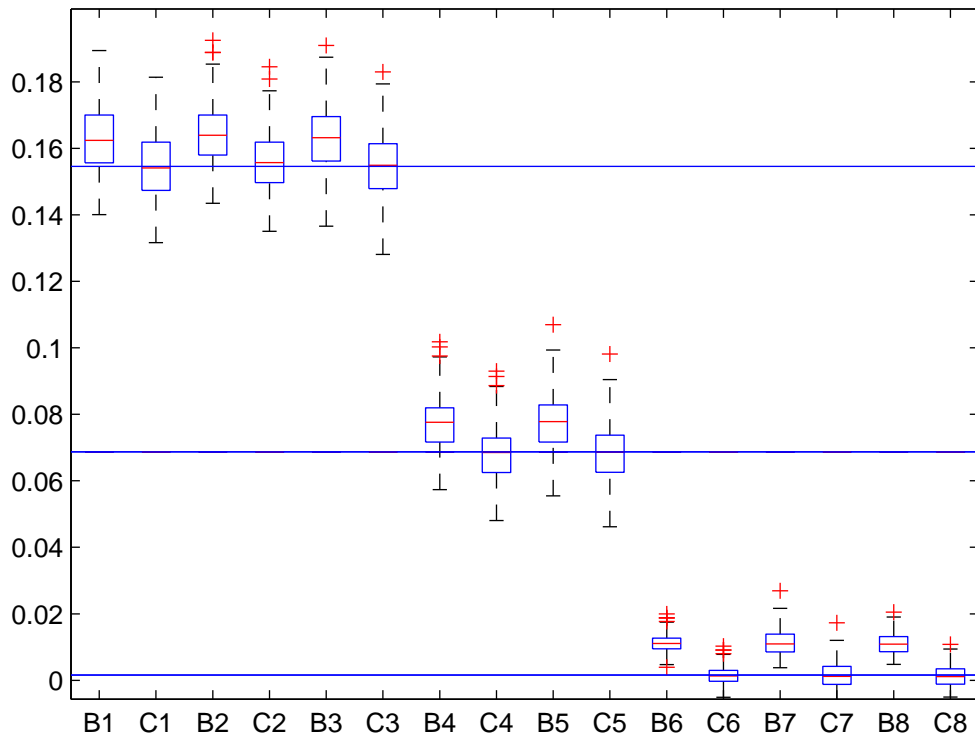


Figure 4: Boxplots of the first-order sensitivity indices by the RBD method with a sample size of 2001. The analytical values are $S_1 = S_2 = S_3 = 0.1546$, $S_4 = S_5 = 0.0687$ and $S_6 = S_7 = S_8 = 0.0015$.

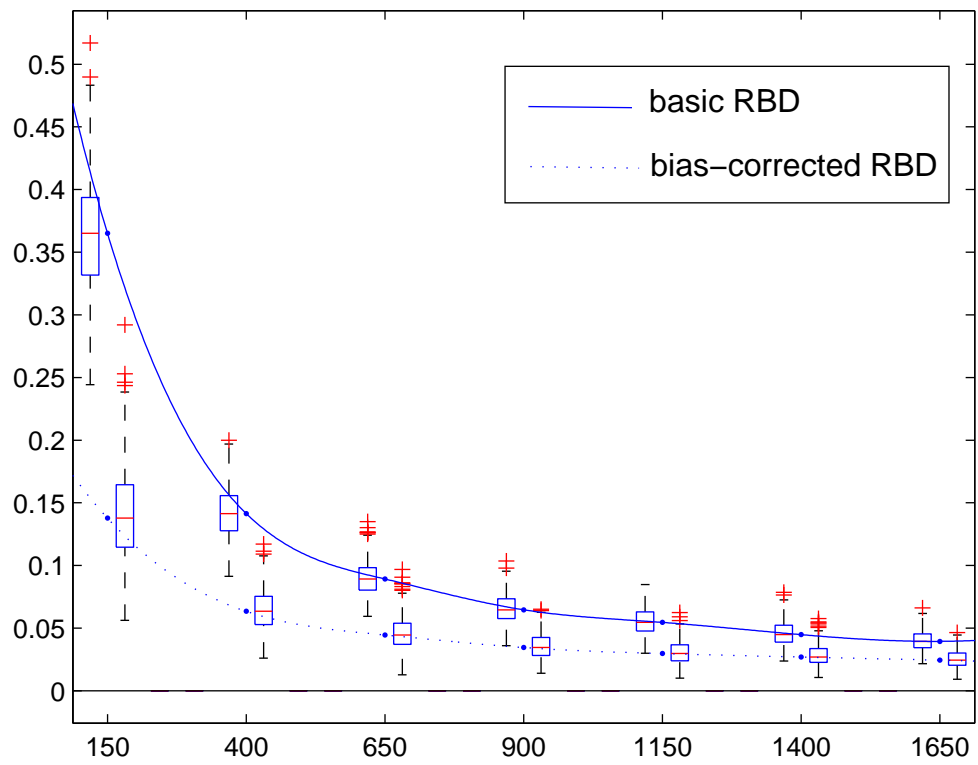


Figure 5: Plots and boxplots (150 replicates) of total quadratic error against the sample size N .

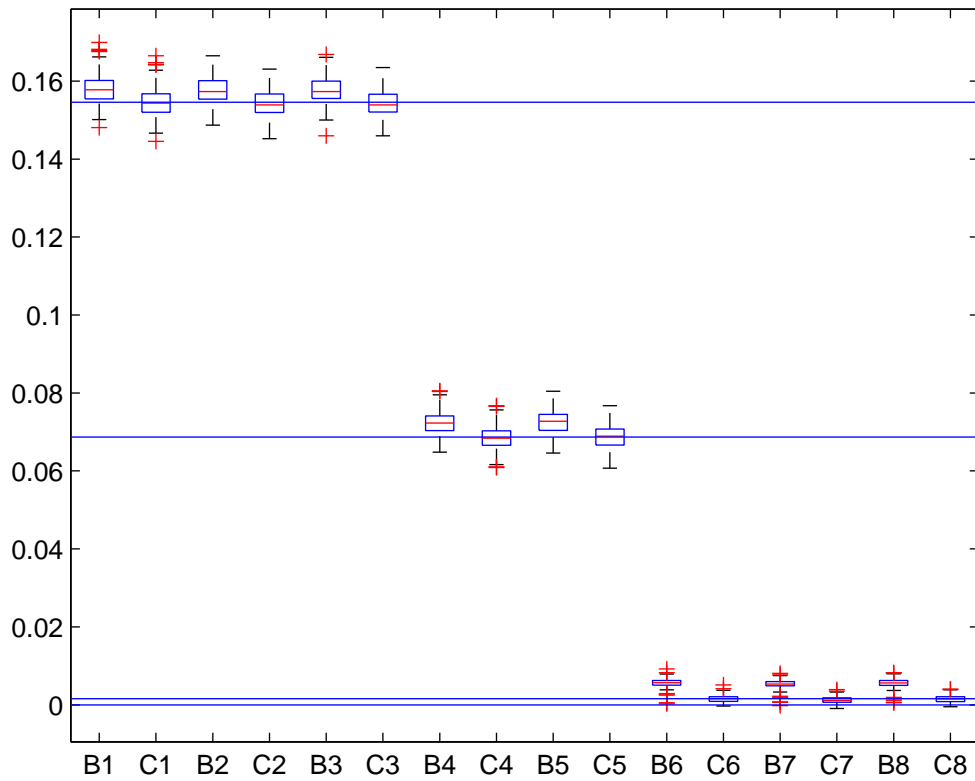


Figure 6: Boxplots of the first-order sensitivity indices by the RBD-FAST method with a sample size of 4001. The analytical values are $S_1 = S_2 = S_3 = 0.1546$, $S_4 = S_5 = 0.0687$ and $S_6 = S_7 = S_8 = 0.0015$.

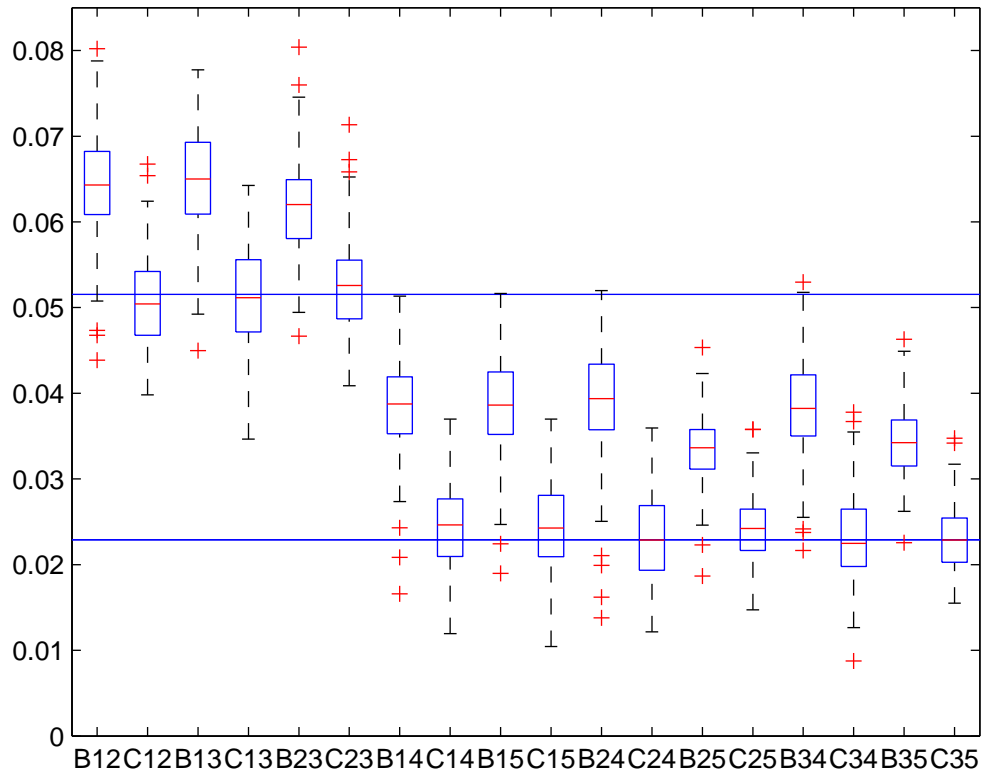


Figure 7: Boxplots of the second-order sensitivity indices by the RBD-FAST method with a sample size of 4001. The analytical values are $S_{12} = S_{13} = S_{23} = 0.0515$, $S_{14} = S_{15} = S_{24} = S_{25} = S_{34} = S_{35} = 0.0229$.

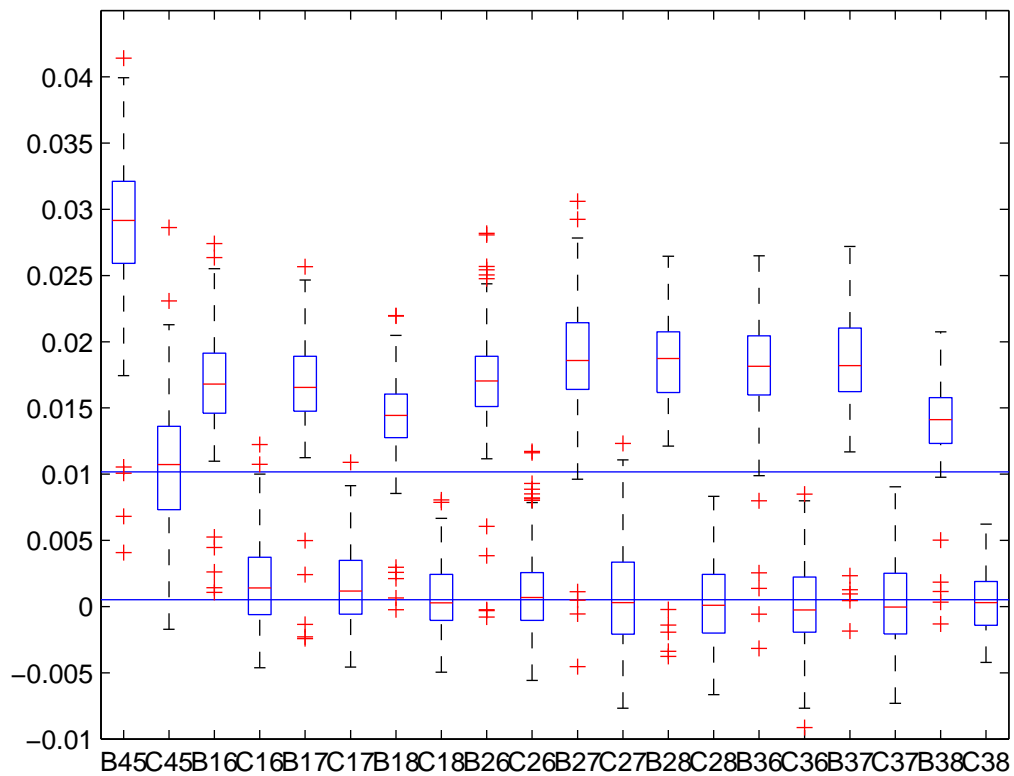


Figure 8: Boxplots of the second-order sensitivity indices by the RBD-FAST method with a sample size of 4001. The analytical values are $S_{45} = 0.0102$, $S_{16} = S_{17} = S_{18} = S_{26} = S_{27} = S_{28} = S_{36} = S_{37} = S_{38} = 0.0005$.

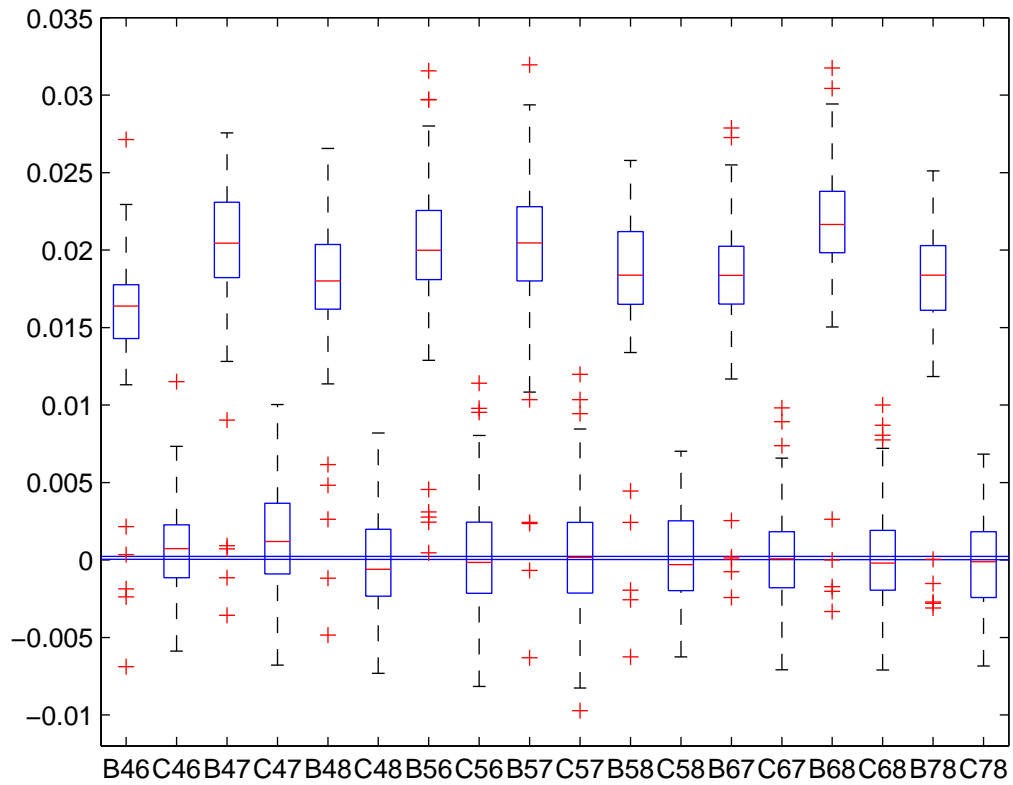


Figure 9: Boxplots of the second-order sensitivity indices by the RBD-FAST method with a sample size of 4001. The analytical values are $S_{46} = S_{47} = S_{48} = S_{56} = S_{57} = S_{58} = 0.0002$, $S_{67} = S_{68} = S_{78} = 5.10^{-6}$.