Bias correction for the estimation of sensitivity indices based on random balance designs

Jean-Yves Tissot, Clémentine Prieur

▶ To cite this version:

HAL Id: hal-00507526
https://hal.science/hal-00507526v2

Bias correction 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

This paper deals with the random balance design method (RBD) and its hybrid approach, RBD-FAST. Both these global sensitivity analysis methods originate from Fourier amplitude sensitivity test (FAST) and are consequently faced with the main problems inherent to discrete harmonic analysis. As some authors pointed out in these methods, certain estimates of sensitivity indices are systematically over-evaluated. This positive bias has recently been identified in the RBD method by Xu and Gertner [1]. Following their work, we propose a bias correction method for first-order sensitivity indices estimates in RBD. We then extend the correction method to the sensitivity indices of any order in RBD-FAST. At last, we suggest an efficient strategy to estimate all the first 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 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 [2, 3] for a review, see also [4, 5] for recent developments on derivative based methods), the most well-

---

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

known of which are methods of computing variance-based sensitivity indices [6] using ANOVA decomposition [7, 8].

ANOVA decomposition and sensitivity indices. Let $\mathbf{X} = (X_1, \ldots, X_p)$ be a random vector and $Y = f(\mathbf{X}) \in \mathbb{R}$, where f is a square-integrable function. Under the assumption that $\mathbf{X}$ has independent components, the variance V of the model output Y can be decomposed as:

$$V = \sum_{k=1}^{p} \sum_{1 \leq i_1 < \cdots < i_k \leq p} V_{i_1 \ldots i_k} \tag{1}$$

where

$$V_{i_1 \ldots i_k} \triangleq \sum_{J \subset \{i_1, \ldots, i_k\}} (-1)^{k - \#J} \mathrm{Var}\big(\mathbb{E}(Y | X_j, j \in J)\big) \tag{2}$$

where $\mathrm{Var}(\cdot)$ and $\mathbb{E}(\cdot|\cdot)$ denote variance and conditional expectation, respectively. Thus, if $V \neq 0$ (i.e. Y is not almost surely constant), dividing both sides of (1) by V, yields a positive and normalized decomposition,

$$1 = \sum_{k=1}^{p} \sum_{1 \leq i_1 < \cdots < i_k \leq p} S_{i_1 \ldots i_k} \tag{3}$$

where for each k,

$$S_{i_1 \ldots i_k} \triangleq \frac{V_{i_1 \ldots i_k}}{V} , \ 1 \leq i_1 < \cdots < i_k \leq p \tag{4}$$

are the so-called k^{th}-order sensitivity indices — or Sobol' indices —.

In the case of an additive model — i.e. $f(X_1, \ldots, X_p) = \sum_{k=1}^{p} f_k(X_k)$ — all terms but the first-order sensitivity indices are zero and we obtain a full decomposition with only $S_1, \ldots, S_p$. 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. In practice, it is sufficient to compute only the first and second-order sensitivity indices to get a good overview of the global variations of a model output.

FAST and its derived methods. Different methods have been developed to estimate sensitivity indices, and FAST, introduced in the 1970's, is one of the

earliest. The three introduction papers [9–11] describe how to compute main effects — i.e. first-order sensitivity indices — exploiting Weyl's ergodic theorem [12]. Then, in a review article [13], the authors precise the underlying theory considering multiple Fourier series, and suggest a decomposition of variance (see Eq. (2.29) in [13]) which allows to consider higher-order sensitivity indices. But, in practice, many sources of error occurs and it is generally impossible to get accurate estimates at low computational cost. As a consequence, FAST has only been applied to estimate first-order and total (see the EFAST method due to Saltelli *et al.* [14]) sensitivity indices in small dimension.

The RBD and Hybrid FAST-RBD (HFR) methods, proposed in 2006 by Tarantola *et al.* [15], partially overcome the inherent drawbacks of FAST using a new sampling technique based on Satterthwaite's random balance designs [16]. These methods have been introduced to estimate first-order sensitivity indices, and as Mara [17] notices, it is also possible to estimate sensitivity indices of any order or closed and total sensitivity indices, using the HFR method (renamed RBD-FAST).

At last, Plischke [18] recently derived from FAST a method named Effective Algorithm for computing global Sensitivity Indices (EASI) which estimates sensitivity indices. The main advantage of this method is that, while FAST, RBD and RBD-FAST use specific experimental designs, EASI can be applied to any input sample.

In Section 2, we briefly recall the FAST method and discuss the different sources of error that affect the accuracy of sensitivity indices estimates. In Section 3, we present, the specific problem of interferences in RBD which leads to the positive bias of the first-order sensitivity indices and we propose a bias-correction method. In Section 4, we extend this technique to the sensitivity indices of any order in RBD-FAST, and in Section 5, we describe an efficient strategy to estimate all the first 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. Conclusions and ideas for a future work

3

are summarised in Section 7.

2. Sources of error in the FAST method

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 use discrete Fourier transform. The experimental design $(x^k)_{k=1...N}$ is such that

$$x_i^k = G_i\big(\sin(\omega_i s_k + \varphi_i)\big) , \quad i = 1, \ldots, p, \ k = 1, \ldots, N \tag{5}$$

where ω_i are integer frequencies — free of interferences up to a certain order (see Section 2.2.1) —, G_i are functions to be settled so as to impose probability density functions on the input variables X_i, φ_i are random phase-shifts and $(s_k)_{k=1...N}$ is defined as

$$s_k = \frac{2\pi(k-1)}{N} . \tag{6}$$

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

$$G_i(\cdot) = \frac{1}{\pi} \arcsin(\cdot) + \frac{1}{2} . \tag{7}$$

The finite sequence $\Big(f\big(x_1^j, \ldots, x_p^j\big)\Big)_{j=1...N}$ obtained from the experimental design defined by (5) is then considered as a uniformly sampled signal. Hence, the Fourier spectrum of this discrete signal can be decomposed with respect to the frequencies ω_1, $\ldots$, ω_p, and we can define the estimators of the different parts of variance,

$$\widehat{V} = \sum_{1 \le |n| \le N/2} |\hat{c}_n|^2 , \tag{8}$$

$$\widehat{V_i} = \sum_{1 \le |k| \le N_h} |\hat{c}_{k\omega_i}|^2 , \tag{9}$$

$$\widehat{V_{ij}} = \sum_{2 \le |k|+|l| \le N_{lc2}} |\hat{c}_{k\omega_i + l\omega_j}|^2 , \tag{10}$$

4

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

$$\hat{c}_n = \frac{1}{N} \sum_{j=1}^{N} f\left(x_1^j, \ldots, x_p^j\right) e^{-in\frac{2\pi(j-1)}{N}} \ , \quad -\frac{N}{2} \leq n \leq \frac{N}{2} \tag{11}$$

are discrete Fourier coefficients. Finally, dividing (9) (resp. (10)) by (8), we get the estimator of a first-order (resp. second-order) sensitivity index:

$$\widehat{S_i} = \frac{\displaystyle\sum_{1 \leq |k| \leq N_h} |\hat{c}_{k\omega_i}|^2}{\displaystyle\sum_{1 \leq |n| \leq N/2} |\hat{c}_n|^2} \ , \tag{12}$$

$$\widehat{S_{ij}} = \frac{\displaystyle\sum_{2 \leq |k|+|l| \leq N_{lc2}} |\hat{c}_{k\omega_i + l\omega_j}|^2}{\displaystyle\sum_{1 \leq |n| \leq N/2} |\hat{c}_n|^2} \ . \tag{13}$$

2.2. Sources of error

The accuracy of these estimates naturally depends on 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 three main sources of error.

2.2.1. Interferences

If there exists a linear combination of the frequencies $\omega_1, \ldots, \omega_p$ equal to zero, some parts of variance could be attributed by error to another one in the decomposition of the Fourier spectrum. For example, if $-2\omega_1 + \omega_2 = 0$, the discrete Fourier coefficient $\hat{c}_{2\omega_1} = \hat{c}_{\omega_2}$ contains information from both X_1 and X_2, and should not be totally attributed to $\widehat{S_1}$ and $\widehat{S_2}$. These interferences can lead to an over-evaluation of sensitivity indices estimates, and in order to keep these positive biases low, we adopt the criterion proposed by Schaibly and Shuler [10] to get a set of frequencies free of interferences up to a certain order M,

$$\sum_{i=1}^{p} a_i \omega_i \neq 0 \ \ \text{for} \ \ \sum_{i=1}^{p} |a_i| \leq M + 1 \ . \tag{14}$$

2.2.2. Aliasing

Only the linear combinations $\omega = \sum_{i=1}^{p} a_i \omega_i$ lying inside the Fourier spectrum — i.e. between $-N/2$ and $N/2$ — are unambiguously represented by the discrete sampled signal. If ω is out of this range, its spectral component is falsely attributed to another frequency inside the Fourier spectrum. To avoid 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:

$$N \geq 2M \max_{1 \leq i \leq p} \omega_i \ , \tag{15}$$

where M is defined in the previous paragraph.

2.2.3. Truncation

The different finite sums defining the estimators of the total variance and the partial variances in Eqs. (8), (9) and (10) should be 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. Random balance design method

As we noted in the previous part, using a distinct frequency per input factor in the FAST method imposes restrictive constraints on the sample size. To overcome this drawback, an alternative sampling method is employed in RBD.

3.1. Sampling method

Contrarily to FAST, in RBD, all the ω_i are equal to a unique frequency ω and input variables are distinguished by taking random permutations of the

6

coordinates of the sample points. Let $\sigma_1, \ldots, \sigma_p$ denote random permutations on the set $\{1, \ldots, N\}$, the experimental design $(x^k)_{k=1\ldots N}$ is such that

$$x_i^k = G_i\big(\sin(\omega s_{\sigma_i(k)})\big), \quad \forall i = 1, \ldots, p \ \text{ and } \ \forall k = 1, \ldots, N \ . \qquad (16)$$

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 [19]); the only difference is that, the RBD design points are located in the center of the cells (see Fig. 1).

[Fig. 1 about here.]

3.2. Estimator

RBD sampling method can be used to estimate first-order sensitivity indices. The estimator of the total variance is defined as in FAST and the part of variance due to the factor X_i is estimated by

$$\widehat{V_i} = \sum_{1 \leq |k| \leq N_h} |\hat{c}_{k\omega}^{\sigma_i}|^2 \qquad (17)$$

with

$$\hat{c}_{k\omega}^{\sigma_i} = \frac{1}{N} \sum_{j=1}^{N} f\big(x_1^{\sigma_i^{-1}(j)}, \ldots, x_p^{\sigma_i^{-1}(j)}\big) e^{-ik\omega \frac{2\pi(j-1)}{N}} \ , \qquad (18)$$

where σ_i^{-1} is the inverse permutation of σ_i. Indeed, considering a fixed i, the design points $\big(x_1^{\sigma_i^{-1}(j)}, \ldots, x_p^{\sigma_i^{-1}(j)}\big)_{j=1\ldots N}$ are such that the ith coordinate is sampled with respect to the frequency ω and the other ones are sampled in a random way because

$$
\begin{aligned}
x_k^{\sigma_i^{-1}(j)} \ &= \ G_k\bigg(\sin\big(\omega s_{\sigma_k(\sigma_i^{-1}(j))}\big)\bigg) \\[2mm]
&= \ \begin{cases} G_k\big(\sin(\omega s_j)\big) & \text{if } k = i \\[2mm] G_k\big(\sin\big(\omega s_{\sigma_k^i(j)}\big)\big) & \text{if } k \neq i \ , \end{cases}
\end{aligned}
\qquad (19)
$$

where $\sigma_k^i = \sigma_k \circ \sigma_i^{-1}$ is almost surely a non-trivial permutation. Therefore, in the Fourier spectrum of the signal

$$\Big(f\big(x_1^{\sigma_i^{-1}(j)}, \ldots, x_p^{\sigma_i^{-1}(j)}\big)\Big)_{j=1\ldots N} \ , \qquad (20)$$

7

the harmonics of ω are attributed to the part of variance of X_i. Thus, using FAST estimator, we get Eqs. (17) and (18).

Remark 1. *The choice of the frequency ω seems to be of secondary importance. However, 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 as in Eq. (15).*

3.3. Bias

As we explained in the last section, the RBD estimator is so defined because the harmonics of ω of the signal $\left(f\left(x_1^{\sigma_i^{-1}(j)}, \ldots, x_p^{\sigma_i^{-1}(j)}\right)\right)_{j=1\ldots N}$ are supposed to be only related to the part of variance V_i due to X_i. But it is essential to notice that, as the factors $(X_k)_{k \neq i}$ are randomly sampled, the remaining part of variance — denoted V_{-i} — appears in the signal $\left(f\left(x_1^{\sigma_i^{-1}(j)}, \ldots, x_p^{\sigma_i^{-1}(j)}\right)\right)_{j=1\ldots N}$ as a random noise. Therefore, a random fraction of each harmonic of ω is related to V_{-i} and is falsely attributed to V_i. Xu and Gertner [1] quantified, in mean, this interference between the harmonics of ω, and the random noise, showing that for any harmonic of ω, $\hat{c}_{k\omega}^{\sigma_i}$, we have

$$E\left(|\hat{c}_{k\omega}^{\sigma_i}|^2\right) = |c_{k\omega}^{\sigma_i}|^2 + \frac{V_{-i}}{N} \tag{21}$$

where $c_{k\omega}^{\sigma_i}$ denotes the theoretical unbiased k^{th} harmonic of ω. Thus, following Eq. (17), we define the bias-corrected estimator of V_i as

$$\widehat{V_i^c} = \widehat{V_i} - \frac{2N_h}{N}\widehat{V_{-i}} \ , \tag{22}$$

where $\widehat{V_{-i}}$ is an estimator of V_{-i} defined, assuming the bias correction, as

$$\widehat{V_{-i}} = \widehat{V} - \widehat{V_i^c} \ . \tag{23}$$

Hence,

$$\widehat{V_i^c} = \widehat{V_i} - \frac{2N_h}{N}\left(\widehat{V} - \widehat{V_i^c}\right) \ ; \tag{24}$$

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

$$\widehat{S_i^c} = \widehat{S_i} - \frac{2N_h}{N}\left(1 - \widehat{S_i^c}\right) \ , \tag{25}$$

8

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

$$\widehat{S_i^c} = \widehat{S_i} - \frac{\lambda}{1 - \lambda}\left(1 - \widehat{S_i}\right) . \tag{26}$$

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

(i) the larger N is, the less significant the bias is,

(ii) the larger S_i is, the less significant the bias is.

Remark 3. *In his paper, Plischke [18] suggests to apply exactly the same bias correction to the EASI estimates (see Eq. (7) in [18]). His approach is based on a bias correction method for correlation ratios due to Kelley [20].*

4. Hybrid approach: RBD-FAST

The underlying idea in RBD-FAST is to combine both RBD and FAST sampling approaches. Therefore, this new method is naturally faced with the classical drawbacks of FAST, but in a lesser extent. The main advantage of the hybrid approach is that estimation of higher-order sensitivity indices is possible.

4.1. Sampling method

First, the p input variables are divided into groups of approximatively equal size. 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:

6 factors: $X_1 \ X_2 \ X_3 \ X_4 \ X_5 \ X_6$

$$\underbrace{\omega_1 \ \omega_2 \ \omega_3}_{\sigma_1} \ \underbrace{\omega_1 \ \omega_2 \ \omega_3}_{\sigma_2}$$

6 factors: $X_1 \ X_2 \ X_3 \ X_4 \ X_5 \ X_6$

$$\underbrace{\omega_1 \ \omega_2}_{\sigma_1} \ \underbrace{\omega_1 \ \omega_2}_{\sigma_2} \ \underbrace{\omega_1 \ \omega_2}_{\sigma_3}$$

7 factors: $X_1 \ X_2 \ X_3 \ X_4 \ X_5 \ X_6 \ X_7$

$$\underbrace{\omega_1 \ \omega_2 \ \omega_3}_{\sigma_1} \ \underbrace{\omega_1 \ \omega_2}_{\sigma_2} \ \underbrace{\omega_1 \ \omega_2}_{\sigma_3} \ .$$

Remark 4. *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. Estimators

This hybrid sampling method allows to define the estimator of sensitivity indices of any order. In particular, considering two factors inside the mth group respectively associated with the frequencies ω_i and ω_j, we can define the part of variance of their interaction as,

$$\widehat{V_{ij}} = \sum_{2 \leq |k|+|l| \leq N_{lc2}} |c^{\sigma_m}_{k\omega_i + l\omega_j}|^2 \tag{27}$$

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

$$c^{\sigma_m}_{k\omega_i + l\omega_j} = \frac{1}{N} \sum_{n=1}^{N} f\left(x_1^{\sigma_m^{-1}(n)}, \ldots, x_p^{\sigma_m^{-1}(n)}\right) e^{-i(k\omega_i + l\omega_j)\frac{2\pi(n-1)}{N}} \ . \tag{28}$$

In the same way, considering a factor inside the mth group associated with the frequency ω_i, we can define its part of variance as,

$$\widehat{V_i} = \sum_{1 \leq |k| \leq N_h} |c^{\sigma_m}_{k\omega_i}|^2 \tag{29}$$

10

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

$$c^{\sigma_m}_{k\omega_i} = \frac{1}{N} \sum_{n=1}^{N} f\left(x_1^{\sigma_m^{-1}(n)}, \ldots, x_p^{\sigma_m^{-1}(n)}\right) e^{-ik\omega_i \frac{2\pi(n-1)}{N}} . \tag{30}$$

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

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

$$x_k^{\sigma_m^{-1}(j)} = G_k\left(\sin\left(\omega_i s_{\sigma_m(\sigma_m^{-1}(j))}\right)\right) = G_k\left(\sin\left(\omega_i s_j\right)\right) \tag{31}$$

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

$$x_k^{\sigma_m^{-1}(j)} = G_k\left(\sin\left(\omega_i s_{\sigma_n(\sigma_m^{-1}(j))}\right)\right) \tag{32}$$

where $\sigma_n \circ \sigma_m^{-1}$ is almost surely a non-trivial permutation. Therefore, all input variables outside the group associated with σ_m are randomly sampled, and the other ones are sampled with respect to their frequencies. Applying FAST's estimator, Eqs. (27)–(30) follow.

4.3. Bias

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

Let $X_{m_1}, \ldots, X_{m_d}$ be the d input factors inside the mth group, and P be a nonempty subset of $\{m_1, \ldots, m_d\}$. We denote V_P the part of variance due to the interaction between the input variables $(X_i)_{i \in P}$ (e.g. if $P = \{i\}$, V_P is simply V_i, and if $P = \{i, j\}$, V_P is V_{ij}). Let $\widehat{V_P}$ be the RBD-FAST classical estimator of V_P, previously described in Eqs. (27) and (29) for $\#P = 1$ and 2. Following RBD bias correction, we first define the estimator of the positive bias B_P as

$$\widehat{B_P} = \frac{n(P)}{N} \widehat{V_{-P}} \tag{33}$$

and the corrected estimator of V_P as

$$\widehat{V_P^c} = \widehat{V_P} - \widehat{B_P} \tag{34}$$

11

where $n(P)$ is the number of Fourier coefficients taken into account to estimate V_P and $\widehat{V_{-P}}$ is an estimate of the part of variance not due to any subset of factors contained in $\{m_1, \ldots, m_d\}$ defined, assuming the bias correction, as

$$\widehat{V_{-P}} = \widehat{V} - \sum_{\substack{Q \subset \{m_1, \ldots, m_d\} \\ Q \neq \emptyset}} \widehat{V_Q^c} \ . \tag{35}$$

Hence,

$$\widehat{V_P^c} = \widehat{V_P} - \frac{n(P)}{N} \left(\widehat{V} - \sum_{\substack{Q \subset \{m_1, \ldots, m_d\} \\ Q \neq \emptyset}} \widehat{V_Q^c} \right), \tag{36}$$

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

$$\widehat{S_P^c} = \widehat{S_P} - \frac{n(P)}{N} \left(1 - \sum_{\substack{Q \subset \{m_1, \ldots, m_d\} \\ Q \neq \emptyset}} \widehat{S_Q^c} \right), \tag{37}$$

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

$$\lambda_Q = \frac{n(Q)}{N} \ , \quad \text{for any nonempty subset } Q \in \{m_1, \ldots, m_d\} \tag{38}$$

and

$$\overline{\lambda} = \sum_{\substack{Q \subset \{m_1, \ldots, m_d\} \\ Q \neq \emptyset}} \lambda_Q \ , \tag{39}$$

we conclude with the explicit formula

$$\widehat{S_P^c} = \widehat{S_P} - \frac{\lambda_P}{1 - \overline{\lambda}} \left(1 - \sum_{\substack{Q \subset \{m_1, \ldots, m_d\} \\ Q \neq \emptyset}} \widehat{S_Q} \right) \quad \text{(see details in Appendix A).} \tag{40}$$

Remark 5. *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,*

$$\widehat{S_P^c} = \widehat{S_P} - \frac{\lambda_P}{1 - \overline{\lambda}} \left(1 - \sum_{\substack{Q \subset \{m_1, \ldots, m_d\} \\ Q \neq \emptyset, \ \#Q \leq \delta}} \widehat{S_Q} \right) \tag{41}$$

12

where

$$\overline{\lambda} = \sum_{\substack{Q \subset \{m_1,\ldots,m_d\} \\ Q \neq \emptyset,\ \#Q \leq \delta}} \lambda_Q \ . \tag{42}$$

Remark 6. *An analagous formula for closed sensitivity indices can be deduced from* (40). *Keeping the same notations as previously, such indices are defined as,*

$$S_P^{closed} \triangleq \sum_{Q \subset P,\ Q \neq \emptyset} S_Q \tag{43}$$

and we have

$$\widehat{S}_P^{closed,c} = \widehat{S}_P^{closed} - \frac{\lambda_P^{closed}}{1 - \overline{\lambda}} \left(1 - \sum_{\substack{Q \subset \{m_1,\ldots,m_d\} \\ Q \neq \emptyset}} \widehat{S_Q} \right) \tag{44}$$

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

$$\lambda_P^{closed} = \sum_{Q \subset P,\ Q \neq \emptyset} \lambda_Q \ . \tag{45}$$

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

Throughout this section, we develop a strategy using RBD-FAST to get all the bias-corrected estimates of the first 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 by applying Eqs. (41) and (42).

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

13

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:

$$1 + \min_{\substack{\sqrt{p} \leq q \\ q \text{ prime}}} q \qquad \text{for } p \geq 4 \tag{46}$$
$$1 \qquad \text{for } p \leq 3 \; ,$$

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.

It has to be noted that, in Mara's paper [17], input variables are divided into groups of 2 factors, while our configurations can contain subgroups of more than 2 factors. Thus, the constraints on the sample size that arise from FAST — see Eqs. (14) and (15) — are more restrictive in our approach. Nevertheless, as we can observe in Table 1, at the same computational cost, our strategy provides second-order sensitivity indices estimates with smaller variance.

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 $\{X_1, \ldots, X_p\}$ into q groups of q factors; for example, in the case $p = 9$, we can have,

$$\text{configuration 0}: \quad \underbrace{X_4 X_1 X_5}_{G_1^0} \quad \underbrace{X_7 X_9 X_2}_{G_2^0} \quad \underbrace{X_3 X_8 X_6}_{G_3^0} \; . \tag{47}$$

Following RBD-FAST approach, each group receives a set of free of interferences frequencies and is randomly permuted. This 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:

14

(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 $(G_i^0)_{i=1...q}$,

(R3) if a set of two distinct variables $\{X_i, X_j\}$ is already contained in a group G_n^k, then we are not allowed to define a group G_m^l, with $l \neq k$ and $m \neq n$, in a next configuration containing both X_i and X_j.

For instance, in the case $p = 9$, it is only possible to create three new configurations,

$$
\begin{array}{llll}
\text{configuration 1:} & \underbrace{X_9 X_4 X_8}_{G_1^1} & \underbrace{X_7 X_5 X_6}_{G_2^1} & \underbrace{X_3 X_2 X_1}_{G_3^1} \\[2mm]
\text{configuration 2:} & \underbrace{X_6 X_1 X_9}_{G_1^2} & \underbrace{X_3 X_7 X_4}_{G_2^2} & \underbrace{X_2 X_8 X_5}_{G_3^2} \\[2mm]
\text{configuration 3:} & \underbrace{X_7 X_1 X_8}_{G_1^3} & \underbrace{X_5 X_3 X_9}_{G_2^3} & \underbrace{X_6 X_2 X_4}_{G_3^3}
\end{array}
\tag{48}
$$

Here, 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, there exists an efficient strategy using $q + 1$ designs of experiments and allowing 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

$$
q^* = \min_{\substack{\sqrt{p} \leq q \\ q \text{ prime}}} q \ ,
\tag{49}
$$

15

and

$$p^* = \left(q^*\right)^2 . \tag{50}$$

Following the strategy presented in the previous section, we can create $q + 1$ designs of experiments with p^* factors, $X_1, \ldots, X_p, \ldots, X_{p^*}$. We then delete variables $X_{p+1}, \ldots, X_{p^*}$ in all configurations. For example, considering an 8-dimensional model, we get $q^* = 3$ and $p^* = 9$, and we can use the designs of experiments presented in Eqs. (47) and (48), and delete the factor X_9, we get

$$
\begin{array}{llll}
\text{configuration 0}: & \underbrace{X_4 X_1 X_5}_{G_1^0} & \underbrace{X_7 X_2}_{G_2^0} & \underbrace{X_3 X_8 X_6}_{G_3^0} \\[2mm]
\text{configuration 1}: & \underbrace{X_4 X_8}_{G_1^1} & \underbrace{X_7 X_5 X_6}_{G_2^1} & \underbrace{X_3 X_2 X_1}_{G_3^1} \\[2mm]
\text{configuration 2}: & \underbrace{X_6 X_1}_{G_1^2} & \underbrace{X_3 X_7 X_4}_{G_2^2} & \underbrace{X_2 X_8 X_5}_{G_3^2} \\[2mm]
\text{configuration 3}: & \underbrace{X_7 X_1 X_8}_{G_1^3} & \underbrace{X_5 X_3}_{G_2^3} & \underbrace{X_6 X_2 X_4}_{G_3^3}
\end{array}
\tag{51}
$$

Hence, for any p, we have an economical strategy for which the number of experimental designs satisfies Eq. (46).

Remark 7. *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 [21] 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. [22]). Considering uniformly distributed independent input variables $(X_i)_{i=1\ldots,p}$ on the unit hypercube, this function is defined as

$$f(X_1, \ldots, X_p) = \prod_{i=1}^{p} g_i(X_i) \tag{52}$$

where $g_i(X_i)$ is given by

$$g_i(X_i) = \frac{|4X_i - 2| + a_i}{1 + a_i} \ .$$ (53)

We consider a 6-dimensional g-function where $(a_i) = (0, 0, 0, 0.5, 0.5, 0, 5)$, so that the three first parameters are important, the others are less important and interactions are quite important. We then add three dummy factors X_7, X_8 and X_9 that don't play any role in the model.

The bias correction method and the efficient strategy are tested on this 9-dimensional model.

6.1. Test on RBD

The correction method is tested using increasing sample sizes, $N = 501$ and $N = 2001$ (see Figs. 2 and 3). In both cases, we estimate all the first-order sensitivity indices with the basic RBD method and with the corrected one. The experiment is replicated 200 times using different random permutations.

We 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 for a large sample size. For a low sample size, we can notice that the bias correction is of great importance because a factor without any effect on the output can appears as a nonnegligible one using the basic RBD method (see B_7, B_8 and B_9 in Fig. 2).

[Figs. 2 and 3 about here.]

6.2. Tests on RBD-FAST

6.2.1. Computations using the efficient strategy

In this section, we test the bias correction method on RBD-FAST. Applying the efficient strategy using RBD-FAST, we estimate all the first and second-order sensitivity indices using only 4 experimental designs — those presented in Eqs. (47) and (48) — with sample size 4001. Following Remark 5, we neglect the third-order effects — their contribution in the variance is theoretically lower than 10% —, so we apply Eqs. (41) and (42) with $\delta = 2$.

17

Here, designs are constructed using different random permutations and the set of frequencies free of interferences is $\{\omega_1, \omega_2, \omega_3\} = \{177, 186, 193\}$. We show on Figs. 4–6 boxplots of 200 replicates; all first-order sensitivity indices are shown in Fig. 4, and a representative subset of the second-order sensitivity indices is shown on Figs. 5 and 6. As in the previous test, the corrected indices are centered on their respective theoretical value; but some differences exist between main effects and interaction estimations. On the one hand, first-order terms are accurately evaluated, and their bias, in the absence of correction, are rather low; on the other hand, interaction estimates suffer from a more important variance and a larger bias in absence of correction. Two main reasons justify the difference between the variances. Firstly, the first-order terms are evaluated thanks to 4 estimates per indices while the second-order ones are computed with only one estimate, and secondly, the complexity of sensitivity indices grows with the order. In terms of bias, the lower performance of the interaction estimations without correction is essentially due to the larger number of frequencies taken into account to evaluate the second-order indices. Indeed, considering Eq. (41), we can notice that the amplitude of the bias:

$$\frac{\lambda_P}{1 - \overline{\lambda}} \left(1 - \sum_{\substack{Q \subset \mathcal{G}(P), Q \neq \emptyset \\ \#Q \leq \delta}} \widehat{S_Q} \right) \tag{54}$$

is proportional to $\lambda_P = n(P)/N$. In this test, we have $n(P) = 2N_h = 2 \times 10 = 20$ for the first-order sensitivity indices, and $n(P) = 2N_{lc2}(N_{lc2} - 1) = 2 \times 7 \times (7 - 1) = 84$ for the second-order sensitivity indices.

[Figs. 4, 5 and 6 about here.]

6.2.2. Comparison with Mara's approach

We now estimate all the first and second-order sensitivity indices using the strategy described in Mara [17]. With such an approach, input variables are divided into 4 groups of 2 factors and 1 single term. Hence, 9 experimental designs have to be employed. To keep the same computational cost as for the previous experiment in Section 6.2.1, sample size is 1791 and we use the set of

18

	S_1	S_4	S_7	S_{14}	S_{17}	S_{47}	S_{12}	S_{45}	S_{78}
Theoretical value	0.1288	0.0573	0	0.0191	0	0	0.0429	0.0085	0
Mean ES	0.1286	0.0573	0.0000	0.0187	-0.0002	-0.0001	0.0423	0.0083	-0.0001
Variance ES($\times 10^{-5}$)	1.1	0.8	0.1	2.5	0.9	1.0	1.9	1.9	1.1
Mean MS	0.1289	0.0568	0.0000	0.0189	-0.0004	-0.0002	0.0423	0.0078	0.0001
Variance MS($\times 10^{-5}$)	1.3	0.8	0.1	10.0	6.4	6.0	9.9	9	6

Table 1: Estimation of the first and second-order sensitivity indices using the RBD-FAST method with sample size 4001 with Mara's strategy (MS) and the proposed efficient strategy (ES). We give, together with the theoretical value of the sensitivity index, the empirical means and variances of a sample of 200 estimator replicates.

frequencies $\{\omega_1, \omega_2\} = \{79, 83\}$. The experiment is replicated 200 times using different random permutations, and results (empirical mean and variance for each strategy) are reported in Table 1. On the one hand, the accuracy of first-order sensitivity indices estimates is the same, and on the other hand, we observe that the efficient strategy provides second-order indices with lower variance. We conclude that the choice of strategy seems to be important in terms of variance reduction.

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 avoids the over-estimation of the first-order and second-order indices, for any sample size.

We also introduced a 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 output at a low cost.

Finally, this efficient strategy introduces the question of variance reduction techniques (see Section 6.2.2), and a further work is to improve RBD and RBD-FAST sampling methods. In particular, optimization algorithms commonly used

for Latin hypercube sampling, could be adapted for RBD experimental designs which are, as we have noticed in Section 3, very close to Latin hypercube designs.

Appendix A. Details on formula (40)

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

$$n = \sum_{k=1}^{d} \binom{d}{k} = 2^d - 1 \ , \tag{A.1}$$

and, to simplify the notations, we denote by λ_i the coefficients λ_{P_i}. Applying Eq. (37) to each of the P_i, we get the linear system,

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

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

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

and, using Laplace expansion,

$$\Delta = 1 - \lambda_1 - \lambda_2 \cdots - \lambda_n \ . \tag{A.4}$$

In practice, we fix N so that

$$\sum_{i=1}^{n} \#P_i < N \tag{A.5}$$

Hence, with the definition in Eq. (38), we have

$$\sum_{i=1}^{n} \lambda_i < 1 . \tag{A.6}$$

This implies that Δ is positive, in particular A is invertible.

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

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

We easily obtain,

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

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

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

and we conclude that Eq. (40) holds.

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 first-order sensitivity indices and one estimate of all second-order terms. So, to show that an efficient strategy exists, it is sufficient to prove the existence of such configurations under the rules (R1), (R2) and (R3) of Section 5.1. We give a constructive proof.

We begin by renaming the factors $(X_i)_{i=1\ldots p}$, and defining an initial configuration,

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

where $X_i^j = X_{(i-1)q+j}$. We then obtain the q other experimental designs by considering for $i = 1, \ldots, q$

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

where for all i and j between 1 and q, σ_i^j is a permutation on the set $\{1, \ldots, q\}$. These configurations obviously satisfy rules (R1) and (R2) since each group $(G_j^i)_{j=1..q}$ is filled with one factor of each original group $(G_k^0)_{k=1..q}$; but (R3) is not always verified. However, we can observe that, letting c be a cyclic permutation of order q, the permutations

$$\sigma_i^j = c^{ij} = \underbrace{c \circ c \circ \cdots \circ c}_{ij \text{ times}} \qquad \text{(B.3)}$$

allow to satisfy rule (R3). Indeed, following the formalism of Eq. (B.2), rule (R3) reads as: for all i, i', k, k', j_1 and j_2 between 1 and q, with $i < i'$ and $j_1 \neq j_2$, either the factor from $G_{j_1}^0$ inside G_k^i — i.e. $X_{j_1}^{\sigma_i^{j_1}(k)}$ — is different from the factor from $G_{j_1}^0$ inside $G_{k'}^{i'}$ — i.e. $X_{j_1}^{\sigma_{i'}^{j_1}(k')}$ — or the factor from $G_{j_2}^0$ inside G_k^i — i.e. $X_{j_2}^{\sigma_i^{j_2}(k)}$ — is different from the factor from $G_{j_2}^0$ inside $G_{k'}^{i'}$ — i.e.

$X_{j_2}^{\sigma_{i'}^{j_2}(k')}$ —. That is to say

$$\forall\ 1 \le i, i', k, k', j_1, j_2 \le q,\ i < i',\ j_1 \neq j_2, \quad \begin{cases} \sigma_i^{j_1}(k) \neq \sigma_{i'}^{j_1}(k') \\ \text{or} \\ \sigma_i^{j_2}(k) \neq \sigma_{i'}^{j_2}(k')\ . \end{cases} \tag{B.4}$$

So, assuming Eq. (B.3), let's prove that

$$\forall\ 1 \le i, i', k, k', j_1, j_2 \le q,\ i < i',\ j_1 \neq j_2, \quad \begin{cases} c^{ij_1}(k) \neq c^{i'j_1}(k') \\ \text{or} \\ c^{ij_2}(k) \neq c^{i'j_2}(k')\ . \end{cases} \tag{B.5}$$

Suppose, by contradiction, that

$$c^{ij_1}(k) = c^{i'j_1}(k') \quad \text{and} \quad c^{ij_2}(k) = c^{i'j_2}(k') \tag{B.6}$$

for some (i, i', k, k', j_1, j_2) with $i \neq i'$ and $j_1 \neq j_2$. It follows that

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

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. Hence, $j_1 - j_2 = qr$ for a certain integer r. But, assuming $1 \le j_1, j_2 \le q$, we conclude that $r = 0$ and $j_1 = j_2$, a contradiction to our assumption $j_1 \neq j_2$. The conclusion follows.

Acknowledgments

References

[1] C. Xu, G. Gertner, Understanding and comparisons of different sampling approaches for the fourier amplitudes sensitivity test (fast), Computational Statistics & Data Analysis 55(1) (2011) 184–198.

[2] A. Saltelli, K. P. S. Chan, E. M. Scott, Sensitivity Analysis, John Wiley & Sons, 2000.

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

[4] 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.

[5] I. M. Sobol', S. Kucherenko, A new derivative based importance criterion for groups of variables and its link with the global sensitivity indices, Computer Physics Communications 181(7) (2010) 1212–1217.

[6] I. M. Sobol', Sensitivity analysis for nonlinear mathematical models, Mathematical Modeling and Computational Experiment 1 (1993) 407–414.

[7] W. F. Hoeffding, A class of statistics with asymptotically normal distributions, Annals of Mathematical Statistics 19 (1948) 293–325.

[8] B. Efron, C. Stein, The jackknife estimate of variance, The Annals of Statistics 9(3) (1981) 586–596.

[9] 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.

[10] 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.

[11] 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.

[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] E. Plischke, An effective algorithm for computing global sensitivity indices (easi), Reliability Engineering and System Safety 95(4) (2010) 354–360.

[19] 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.

[20] T. L. Kelley, An unbiased correlation ratio measure, Proceedings of the National Academy of Sciences of the United States of America 21(9) (1935) 554–559.

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

[22] 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

Figure 1: Comparison between Latin hypercube and RBD samples in two-dimensional unit hypercube with sample size 15.

Figure 2: Estimation of the first-order sensitivity indices using RBD. We compare, for a fixed sample size $N = 501$, the basic estimator (B1 to B9) with the bias-corrected one (C1 to C9). In each column, we mark the theoretical sensitivity index with a blue asterisk and plot several summaries of a sample of 200 estimator replicates: the red central mark is the median; the box has its lower and upper edges at the 25th percentile q and the 75th percentile Q, respectively; the whiskers extend between $q - 1.5(Q - q)$ and $Q + 1.5(Q - q)$; the red crosses are outliers.

Figure 3: Estimation of the first-order sensitivity indices using RBD. We compare, for a fixed sample size $N = 2001$, the basic estimator (B1 to B9) with the bias-corrected one (C1 to C9). In each column, we mark the theoretical sensitivity index with a blue asterisk and plot several summaries of a sample of 200 estimator replicates: the red central mark is the median; the box has its lower and upper edges at the 25th percentile q and the 75th percentile Q, respectively; the whiskers extend between $q - 1.5(Q - q)$ and $Q + 1.5(Q - q)$; the red crosses are outliers.

Figure 4: Estimation of the first-order sensitivity indices using RBD-FAST. We compare, for a fixed sample size $N = 4001$, the basic estimator (B1 to B9) with the bias-corrected one (C1 to C9). In each column, we mark the theoretical sensitivity index with a blue asterisk and plot several summaries of a sample of 200 estimator replicates: the red central mark is the median; the box has its lower and upper edges at the 25th percentile q and the 75th percentile Q, respectively; the whiskers extend between $q - 1.5(Q - q)$ and $Q + 1.5(Q - q)$; the red crosses are outliers.

Figure 5: Estimation of the second-order sensitivity indices using RBD-FAST. We compare, for a fixed sample size $N = 4001$, the basic estimator (Bij) with the bias-corrected one (Cij). In each column, we mark the theoretical sensitivity index with a blue asterisk and plot several summaries of a sample of 200 estimator replicates: the red central mark is the median; the box has its lower and upper edges at the 25th percentile q and the 75th percentile Q, respectively; the whiskers extend between $q - 1.5(Q - q)$ and $Q + 1.5(Q - q)$; the red crosses are outliers.

28

Figure 6: Estimation of the second-order sensitivity indices using RBD-FAST. We compare, for a fixed sample size $N = 4001$, the basic estimator (Bij) with the bias-corrected one (Cij). In each column, we mark the theoretical sensitivity index with a blue asterisk and plot several summaries of a sample of 200 estimator replicates: the red central mark is the median; the box has its lower and upper edges at the 25th percentile q and the 75th percentile Q, respectively; the whiskers extend between $q - 1.5(Q - q)$ and $Q + 1.5(Q - q)$; the red crosses are outliers.