Reliability Engineering and System Safety 107 (2012) 205–213

Contents lists available at SciVerse ScienceDirect

Reliability Engineering and System Safety

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

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

Jean-Yves Tissot *, Clémentine Prieur

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

ARTICLE INFO

Article history:
Received 30 July 2010
Received in revised form
11 June 2012
Accepted 12 June 2012
Available online 19 June 2012

Keywords:
Global sensitivity analysis
Sensitivity indices
Random balance design
RBD-FAST
Bias correction

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 consequently face the main problems inherent to discrete harmonic analysis. We present here a general way to correct a bias which occurs when estimating sensitivity indices (SIs) of any order – except total SI of single factor or group of factors – by the random balance design method (RBD) and its hybrid version, RBD-FAST. In the RBD case, this positive bias has been recently identified in a paper by Xu and Gertner [1]. Following their work, we propose a bias correction method for first-order SIs estimates in RBD. We then extend the correction method to the SIs of any order in RBD-FAST. At last, we suggest an efficient strategy to estimate all the first- and second-order SIs using RBD-FAST.

1. Introduction

Global sensitivity analysis of model output consists in quantifying the respective importance of input factors over their entire range of variation. Many techniques have been developed in this field (see [2,3] for a review), this includes for example screening methods [4], density-based methods [5,6] and also derivative-based methods [7,8]. But the most popular are the variance-based methods that rely on ANOVA decomposition [9–11].

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 the components of $\mathbf{X}$ are independent, the variance V of the model output Y can be decomposed as

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

where

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

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

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

where

$$S_{i_1\ldots i_k}=\frac{V_{i_1\ldots i_k}}{V},\quad 1\le i_1<\cdots<i_k\le p\tag{4}$$

are the so-called kth-order SIs—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 SI 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, the first- and second-order SIs generally provide a good overview of the global variations of a model output.

FAST and its derived methods. Different methods have been developed to estimate variance-based SIs, the FAST method, introduced in the 1970s, is one of the earliest. The three introduction papers [12–14] describe how to compute main effects – i.e. first-order sensitivity indices – exploiting Weyl's ergodic theorem [15]. Then, in a review article [16], the authors precise the underlying theory considering multiple Fourier series, and suggest a decomposition of variance (see Eq. (2.29) in [16]) which allows to consider higher-order SIs. But, in practice, many sources of error occur 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 SIs in

* Corresponding author. Tel.: +33 4 76 63 54 47; fax: +33 4 76 63 12 63.
E-mail addresses: jean-yves.tissot@imag.fr, jeanyvestissot@free.fr (J.-Y. Tissot).

0951-8320/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ress.2012.06.010

small dimension (see the EFAST method due to Saltelli et al. [17] for total SIs).

The RBD and Hybrid FAST-RBD (HFR) methods, proposed in 2006 by Tarantola et al. [18], partially overcome the inherent drawbacks of FAST using a new sampling technique based on Satterthwaite's random balance designs [19]. These methods have been introduced to estimate first-order SIs, and as Mara [20] notices, it is also possible to estimate SIs of any order or closed and total sensitivity indices, using the HFR method (renamed RBD-FAST).

Recently Plischke [21] derived another FAST-like method, named Effective Algorithm for computing global Sensitivity Indices (EASI), which estimates sensitivity indices with any input sample while FAST, RBD and RBD-FAST use specific experimental designs.

In Section 2, we briefly recall the FAST method and discuss the different sources of error that affect the accuracy of SI estimates. In Section 3, we present the specific problem of interferences in RBD which leads to the positive bias of the first-order SIs 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 SI 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 are summarized 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 $(x^k)_{k=1\dots N}$ is such that

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

where the ω_i's are integer frequencies – free of interferences up to a certain order (see Section 2.2) – the G_i's 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 $(s_k)_{k=1\dots N}$ is defined as

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

In particular, to uniformly sample the marginal distributions over [0, 1], one shall use (see for example [17])

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

The Fourier spectrum of the discrete signal $(f(x_1^j,\dots,x_p^j))_{j=1\dots N}$ can be decomposed with respect to the frequencies $\omega_1,\dots,\omega_p$, and the following estimators can be defined

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

$$\widehat{V_i} = \sum_{1 \le |k| \le N_1} |\hat{c}_{k\omega_i}|^2, \tag{9}$$

$$\widehat{V_{ij}} = \sum_{2 \le |k|+|l| \le N_2} |\hat{c}_{k\omega_i + l\omega_j}|^2, \tag{10}$$

and so on; where N_1 is the highest harmonic considered as non-negligible, N_2 is the order over which the linear combinations of ω_i and ω_j are considered as negligible, and

$$\hat{c}_n = \frac{1}{N}\sum_{j=1}^{N} f(x_1^j,\dots,x_p^j)e^{-in2\pi(j-1)/N}, \quad -\frac{N}{2} \le n \le \frac{N}{2} \tag{11}$$

is the n-th complex discrete Fourier coefficient. Finally, dividing (9) (resp. (10)) by (8), we get the estimator of a first-order (resp. second-order) SI:

$$\widehat{S_i} = \frac{\sum_{1 \le |k| \le N_1} |\hat{c}_{k\omega_i}|^2}{\sum_{1 \le |n| \le N/2} |\hat{c}_n|^2}, \tag{12}$$

$$\widehat{S_{ij}} = \frac{\sum_{2 \le |k|+|l| \le N_2} |\hat{c}_{k\omega_i + l\omega_j}|^2}{\sum_{1 \le |n| \le N/2} |\hat{c}_n|^2}. \tag{13}$$

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 the dependence is intricate; in addition to the truncation error, we distinguish two main sources of error.

2.2. Interferences

Whenever a linear combination of the frequencies $\omega_1,\dots,\omega_p$ is equal to zero, some parts of variance could be attributed by error to other ones 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 sometimes cause a bias, and to alleviate their effect, we adopt the criterion proposed by Schaibly and Shuler [13] to choose frequency sets free of interferences up to a certain order M

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

2.3. Aliasing

Only linear combinations such that $-N/2 < \omega = \sum_{i=1}^{p} a_i\omega_i < N/2$ 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 positively biased estimators, it is necessary to satisfy 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:

$$N \ge 2M \max_{1 \le i \le p} \omega_i, \tag{15}$$

where M is defined in the previous paragraph. Practitioners generally set

$$N_1 = N_2 = \cdots = N_d = M, \tag{16}$$

but this constraint is not necessary, and the criterion stated in (14) and (15) can be formulated in a more general way. Indeed, for all $1 \le q \le p$, consider $N_q \in \mathbb{N}^*$, and for all $1 \le i_1 < \cdots < i_q \le p$, define

$$A_{i_1\dots i_q} = \left\{ (a_1,\dots,a_p) \in \mathbb{Z}^p \,|\, \forall i \notin \{i_1,\dots,i_q\}, a_i = 0 \text{ and } \sum_{1 \le m \le q} |a_{i_m}| \le N_q \right\} \tag{17}$$

and

$$A = \bigcup_{q=1}^{p} \bigcup_{1 \le i_1 < \cdots < i_q \le p} A_{i_1\dots i_q}. \tag{18}$$

Hence we propose to replace (14) and (15) by

$$\sum_{i=1}^{p} a_i\omega_i \neq 0 \quad \text{for all } (a_1,\ldots,a_p) \in A \tag{19}$$

and

$$N \geq 2 \max_{(a_1,\ldots,a_p)\in A} \sum_{i=1}^{p} a_i\omega_i, \tag{20}$$

respectively. Note that if (16) is satisfied, (19) and (20) are equivalent to the classic criterion stated in (14) and (15).

3. Random balance design method

As we noted in the previous section, 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

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

$$x_i^k = G_i(\sin(\omega s_{\sigma_i(k)})), \quad \forall i=1,\ldots,p \text{ and } \forall k=1,\ldots,N. \tag{21}$$

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

3.2. Estimator

RBD sampling method can be used to estimate first-order SI. 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

$$\widehat{V_i} = \sum_{1\leq |k| \leq N_1} |\hat{c}_{k\omega}^{\sigma_i}|^2 \tag{22}$$

with

$$\hat{c}_{k\omega}^{\sigma_i} = \frac{1}{N}\sum_{j=1}^{N} f(x_1^{\sigma_i^{-1}(j)},\ldots,x_p^{\sigma_i^{-1}(j)})e^{-ik\omega 2\pi(j-1)/N}, \tag{23}$$

where σ_i^{-1} is the inverse permutation of σ_i. Indeed, considering a fixed i, the design points $(x_1^{\sigma_i^{-1}(j)},\ldots,x_p^{\sigma_i^{-1}(j)})_{j=1\ldots N}$ 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

$$x_k^{\sigma_i^{-1}(j)} = G_k(\sin(\omega s_{\sigma_k(\sigma_i^{-1}(j))})) = \begin{cases} G_k(\sin(\omega s_j)) & \text{if } k=i, \\ G_k(\sin(\omega s_{\sigma_k^i(j)})) & \text{if } k\neq i, \end{cases} \tag{24}$$

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

$$(f(x_1^{\sigma_i^{-1}(j)},\ldots,x_p^{\sigma_i^{-1}(j)}))_{j=1\ldots N}, \tag{25}$$

the harmonics of ω are attributed to the partial variance of X_i. Thus, using FAST estimator, we get Eqs. (22) and (23).

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 in the signal $(f(x_1^{\sigma_i^{-1}(j)},\ldots,x_p^{\sigma_i^{-1}(j)}))_{j=1\ldots N}$ are supposed to be only related to the part of variance V_i due to X_i. But it is essential to notice that, since the factors $(X_k)_{k\neq i}$ are randomly sampled, the remaining part of variance – denoted V_{-i} – appears in the signal $(f(x_1^{\sigma_i^{-1}(j)},\ldots,x_p^{\sigma_i^{-1}(j)}))_{j=1\ldots N}$ 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 this interference between the harmonics of ω and the random noise, showing that for any $\hat{c}_{k\omega}^{\sigma_i}$ we have

$$E(|\hat{c}_{k\omega}^{\sigma_i}|^2) = |c_{k\omega}^{\sigma_i}|^2 + \frac{V_{-i}}{N}, \tag{26}$$

where $c_{k\omega}^{\sigma_i}$ denotes the theoretical unbiased kth harmonic of ω. Thus, following Eq. (22), we define the bias-corrected estimator of V_i as

$$\widehat{V_i^c} = \widehat{V_i} - \frac{2N_1}{N}\widehat{V_{-i}}, \tag{27}$$

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

$$\widehat{V_{-i}} = \widehat{V} - \widehat{V_i^c}. \tag{28}$$

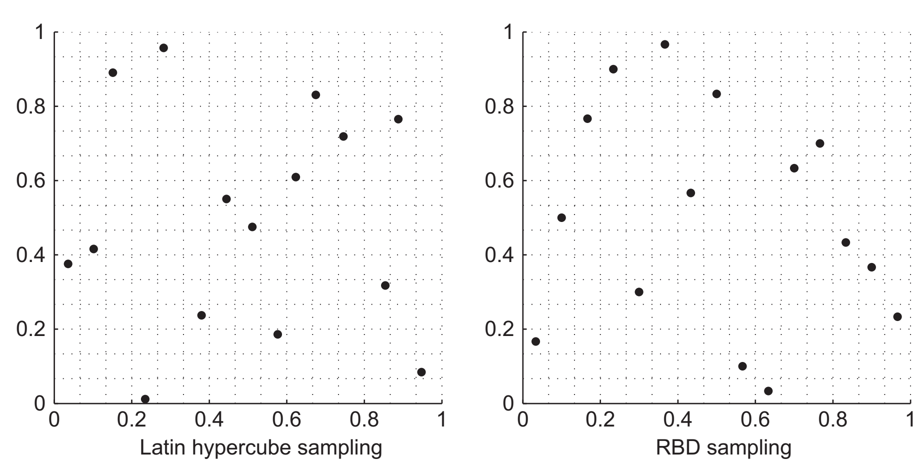

Fig. 1. Comparison between Latin hypercube and RBD samples in two-dimensional unit hypercube with sample size 15.

Hence

$$\widehat{V_i^c} = \widehat{V_i} - \frac{2N_1}{N}(\widehat{V} - \widehat{V_i^c}); \tag{29}$$

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

$$\widehat{S_i^c} = \widehat{S_i} - \frac{2N_1}{N}(1 - \widehat{S_i^c}), \tag{30}$$

where $\widehat{S_i}$ and $\widehat{S_i^c}$ are the RBD estimator of the first-order sensitivity index and the corrected one, respectively. Finally, setting $\lambda = 2N_1/N$, we get the explicit formula

$$\widehat{S_i^c} = \widehat{S_i} - \frac{\lambda}{1-\lambda}(1 - \widehat{S_i}). \tag{31}$$

Remark 2. It is important to observe that the larger N and S_i, the lower the bias.

Remark 3. In his paper, Plischke [21] suggests to apply exactly the same bias correction to the EASI estimates (see Eq. (7) in [21]). His approach is based on a bias correction method for correlation ratios due to Kelley [23].

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 interest of the hybrid approach is that estimation of higher-order SI is possible.

4.1. Sampling method

For this purpose, the p input variables are divided into groups of approximatively equal cardinal. Then each group is assigned a distinct random permutation and each factor of the group a distinct frequency chosen from a frequency set assumed free of interferences up to a given order (see Section 2.2). For example, we can have the following configurations:

$$\text{6 factors}: \quad X_1\ X_2\ X_3\ X_4\ X_5\ X_6 \quad \underbrace{\omega_1\omega_2\omega_3}_{\sigma_1}\ \underbrace{\omega_1\omega_2\omega_3}_{\sigma_2},$$

$$\text{6 factors}: \quad X_1\ X_2\ X_3\ X_4\ X_5\ X_6 \quad \underbrace{\omega_1\omega_2}_{\sigma_1}\ \underbrace{\omega_1\omega_2}_{\sigma_2}\ \underbrace{\omega_1\omega_2}_{\sigma_3},$$

$$\text{7 factors}: \quad X_1\ X_2\ X_3\ X_4\ X_5\ X_6\ X_7 \quad \underbrace{\omega_1\omega_2\omega_3}_{\sigma_1}\ \underbrace{\omega_1\omega_2}_{\sigma_2}\ \underbrace{\omega_1\omega_2}_{\sigma_3}.$$

Remark 4. Tarantola et al. [18] and Mara [20] 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; these just are two different points of view.

4.2. Estimators

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

$$\widehat{V_{ij}} = \sum_{2 \le |k|+|l| \le N_2} \left| c^{\sigma_m}_{k\omega_i + l\omega_j} \right|^2, \tag{32}$$

where N_2 is the value over which the linear combinations of ω_i and ω_j are considered as negligible and where

$$c^{\sigma_m}_{k\omega_i + l\omega_j} = \frac{1}{N}\sum_{n=1}^{N} f(x_1^{\sigma_m^{-1}(n)}, \ldots, x_p^{\sigma_m^{-1}(n)}) e^{-i(k\omega_i + l\omega_j)2\pi(n-1)/N}. \tag{33}$$

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

$$\widehat{V_i} = \sum_{1 \le |k| \le N_1} \left| c^{\sigma_m}_{k\omega_i} \right|^2, \tag{34}$$

where N_1 is the highest harmonic considered as non-negligible and with

$$c^{\sigma_m}_{k\omega_i} = \frac{1}{N}\sum_{n=1}^{N} f(x_1^{\sigma_m^{-1}(n)}, \ldots, x_p^{\sigma_m^{-1}(n)}) e^{-ik\omega_i 2\pi(n-1)/N}. \tag{35}$$

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

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

$$x_k^{\sigma_m^{-1}(j)} = G_k(\sin(\omega_i s_{\sigma_m(\sigma_m^{-1}(j))})) = G_k(\sin(\omega_i s_j)), \tag{36}$$

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

$$x_k^{\sigma_m^{-1}(j)} = G_k(\sin(\omega_i s_{\sigma_n(\sigma_m^{-1}(j))})), \tag{37}$$

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. (32)–(35) 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. (32) and (34) for $card(P) = 1$ and 2. Following RBD bias correction, we first define the estimator of the positive bias B_P as

$$\widehat{B_P} = \frac{n(P)}{N}\widehat{V_{-P}} \tag{38}$$

and the corrected estimator of V_P as

$$\widehat{V_P^c} = \widehat{V_P} - \widehat{B_P}, \tag{39}$$

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

$$\widehat{V_{-P}} = \widehat{V} - \sum_{\substack{Q \subseteq \{m_1,\ldots,m_d\} \\ Q \ne \emptyset}} \widehat{V_Q^c}. \tag{40}$$

Hence

$$\widehat{V_P^c} = \widehat{V_P} - \frac{n(P)}{N}\left(\widehat{V} - \sum_{\substack{Q \subseteq \{m_1,\ldots,m_d\} \\ Q \ne \emptyset}} \widehat{V_Q^c}\right), \tag{41}$$

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

$$\widehat{S_P^c} = \widehat{S_P} - \frac{n(P)}{N}\left(1 - \sum_{\substack{Q \subseteq \{m_1,\ldots,m_d\} \\ Q \neq \emptyset}} \widehat{S_Q^c}\right), \tag{42}$$

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

$$\lambda_Q = \frac{n(Q)}{N} \text{ for any nonempty subset } Q \in \{m_1,\ldots,m_d\} \tag{43}$$

and

$$\overline{\lambda} = \sum_{\substack{Q \subseteq \{m_1,\ldots,m_d\} \\ Q \neq \emptyset}} \lambda_Q, \tag{44}$$

we conclude with the explicit formula

$$\widehat{S_P^c} = \widehat{S_P} - \frac{\lambda_P}{1-\overline{\lambda}}\left(1 - \sum_{\substack{Q \subseteq \{m_1,\ldots,m_d\} \\ Q \neq \emptyset}} \widehat{S_Q}\right). \tag{45}$$

(see details in Appendix A).

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 SI over a certain degree δ and to consider the following bias correction:

$$\widehat{S_P^c} = \widehat{S_P} - \frac{\lambda_P}{1-\overline{\lambda}}\left(1 - \sum_{\substack{Q \subseteq \{m_1,\ldots,m_d\} \\ Q \neq \emptyset, card(Q) \leq \delta}} \widehat{S_Q}\right), \tag{46}$$

where

$$\overline{\lambda} = \sum_{\substack{Q \subseteq \{m_1,\ldots,m_d\} \\ Q \neq \emptyset, card(Q) \leq \delta}} \lambda_Q. \tag{47}$$

Remark 6. An analogous formula for closed SI can be deduced from (45). Keeping the same notations as previously, such indices are defined as

$$S_P^{closed} = \sum_{Q \subseteq P, Q \neq \emptyset} S_Q \tag{48}$$

and we have

$$\widehat{S}_P^{closed,c} = \widehat{S}_P^{closed} - \frac{\lambda_P^{closed}}{1-\overline{\lambda}}\left(1 - \sum_{\substack{Q \subseteq \{m_1,\ldots,m_d\} \\ Q \neq \emptyset}} \widehat{S_Q}\right), \tag{49}$$

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

$$\lambda_P^{closed} = \sum_{Q \subseteq P, Q \neq \emptyset} \lambda_Q. \tag{50}$$

5. An efficient strategy to estimate both 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 SI of a model in which we assume that the SI over a certain order δ are negligible. In this case, we can get the first-order and second-order indices by applying Eqs. (46) and (47).

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 [20] observes that five sample sets are necessary to estimate all the 15 second-order SI – and naturally the first-order ones – of a six-dimensional model. In fact, in the case of six 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

$$1 + \min_{\substack{\sqrt{p} \leq q \\ q \text{ prime}}} q \quad \text{for } p \geq 4,$$

$$1 \quad \text{for } p \leq 3, \tag{51}$$

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 [20], input variables are divided into groups of two factors, while our configurations can contain subgroups of more than two 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 SI 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 SI 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

$$\text{configuration } 0: \quad \underbrace{X_4 X_1 X_5}_{G_1^0} \; \underbrace{X_7 X_9 X_2}_{G_2^0} \; \underbrace{X_3 X_8 X_6}_{G_3^0}. \tag{52}$$

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:

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

(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,

Table 1
Estimation of the first- and second-order SI 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 SI, the empirical means and variances of a sample of 200 estimator replicates.

	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

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:

$$\text{configuration 1}: \quad \underbrace{X_9X_4X_8}_{G_1^1}\ \underbrace{X_7X_5X_6}_{G_2^1}\ \underbrace{X_3X_2X_1}_{G_3^1},$$

$$\text{configuration 2}: \quad \underbrace{X_6X_1X_9}_{G_1^2}\ \underbrace{X_3X_7X_4}_{G_2^2}\ \underbrace{X_2X_8X_5}_{G_3^2},$$

$$\text{configuration 3}: \quad \underbrace{X_7X_1X_8}_{G_1^3}\ \underbrace{X_5X_3X_9}_{G_2^3}\ \underbrace{X_6X_2X_4}_{G_3^3}, \tag{53}$$

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 SI and four estimates of all the first-order terms.

More generally, we have the following proposition:

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 SI 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

$$q^* = \min_{\substack{\sqrt{p} \leq q \\ q \text{ prime}}} q \tag{54}$$

and

$$p^* = (q^*)^2. \tag{55}$$

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 eight-dimensional model, we get $q^*=3$ and $p^*=9$, and we can use the designs of experiments presented in Eqs. (52) and (53), and deleting the factor X_9, we get

$$\text{configuration 0}: \quad \underbrace{X_4X_1X_5}_{G_1^0}\ \underbrace{X_7X_2}_{G_2^0}\ \underbrace{X_3X_8X_6}_{G_3^0},$$

$$\text{configuration 1}: \quad \underbrace{X_4X_8}_{G_1^1}\ \underbrace{X_7X_5X_6}_{G_2^1}\ \underbrace{X_3X_2X_1}_{G_3^1},$$

$$\text{configuration 2}: \quad \underbrace{X_6X_1}_{G_1^2}\ \underbrace{X_3X_7X_4}_{G_2^2}\ \underbrace{X_2X_8X_5}_{G_3^2},$$

$$\text{configuration 3}: \quad \underbrace{X_7X_1X_8}_{G_1^3}\ \underbrace{X_5X_3}_{G_2^3}\ \underbrace{X_6X_2X_4}_{G_3^3}. \tag{56}$$

Hence, for any p, we have an economical strategy for which the number of experimental designs satisfies Eq. (51).

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 [24] who provides an economical way to estimate all the first-order, second-order and total SI 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. [25]). Considering uniformly distributed independent input variables $(X_i)_{i=1\ldots,p}$ on the unit hypercube, this function is defined as

$$f(X_1,\ldots,X_p) = \prod_{i=1}^{p} g_i(X_i), \tag{57}$$

where $g_i(X_i)$ is given by

$$g_i(X_i) = \frac{|4X_i - 2| + a_i}{1 + a_i}. \tag{58}$$

We consider a six-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 do not play any role in the model.

The bias correction method and the efficient strategy are tested on this nine-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 SI 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 appear as a non-negligible one using the basic RBD method (see B_7, B_8 and B_9 in Fig. 2).

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 SI using only four experimental designs – those presented in Eqs. (52) and (53) – with sample size

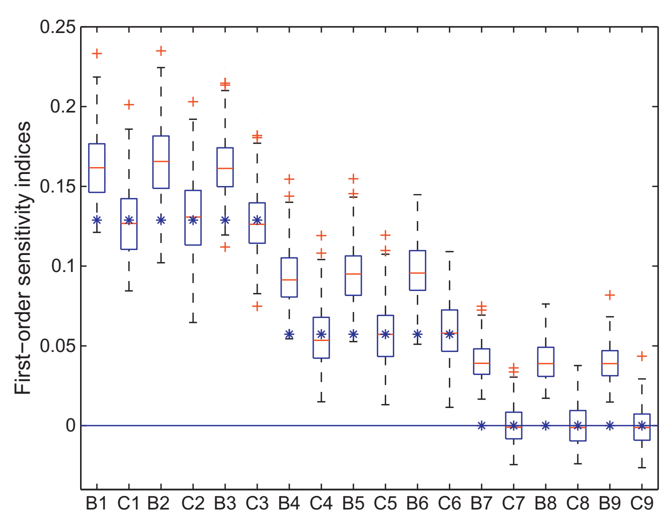

Fig. 2. Estimation of the first-order SI using RBD. We compare, for a fixed sample size $N=501$, the basic estimator (B1–B9) with the bias-corrected one (C1–C9). In each column, we mark the theoretical SI with a blue asterisk and plot several summaries of a sample of 200 estimator replicates: the red 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. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

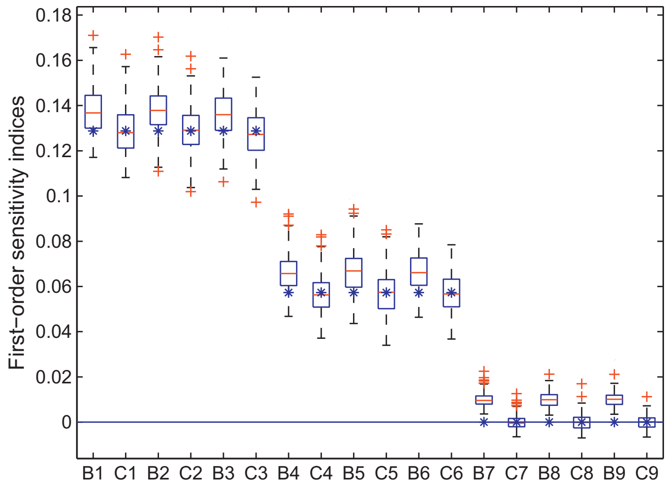

Fig. 3. Estimation of the first-order SI using RBD. We compare, for a fixed sample size $N=2001$, the basic estimator (B1–B9) with the bias-corrected one (C1–C9). In each column, we mark the theoretical SI with a blue asterisk and plot several summaries of a sample of 200 estimator replicates: the red 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. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

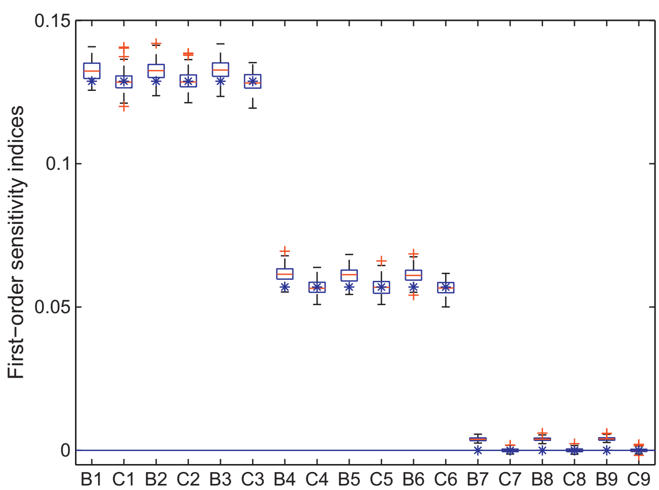

Fig. 4. Estimation of the first-order SI using RBD-FAST. We compare, for a fixed sample size $N=4001$, the basic estimator (B1–B9) with the bias-corrected one (C1–C9). In each column, we mark the theoretical SI with a blue asterisk and plot several summaries of a sample of 200 estimator replicates: the 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. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

4001. Following Remark 5, we neglect the third-order effects – their contribution in the variance is theoretically lower than 10% – so we apply Eqs. (46) and (47) with $\delta=2$.

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 in Figs. 4–6 boxplots of 200 replicates; all first-order SI are shown in Fig. 4, and a representative subset of the second-order SI is shown in 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 the absence of correction. Two main reasons justify the difference between the variances. Firstly the first-order terms are evaluated thanks to four estimates per indices while the second-order ones are computed with only one estimate, and secondly the complexity of SI 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. (46), we can notice that the amplitude of the bias

$$\frac{\lambda_P}{1-\bar{\lambda}}\left(1-\sum_{\substack{Q\subseteq\mathcal{G}(P),Q\neq\emptyset\\ card(Q)\leq\delta}}\widehat{S_Q}\right) \tag{59}$$

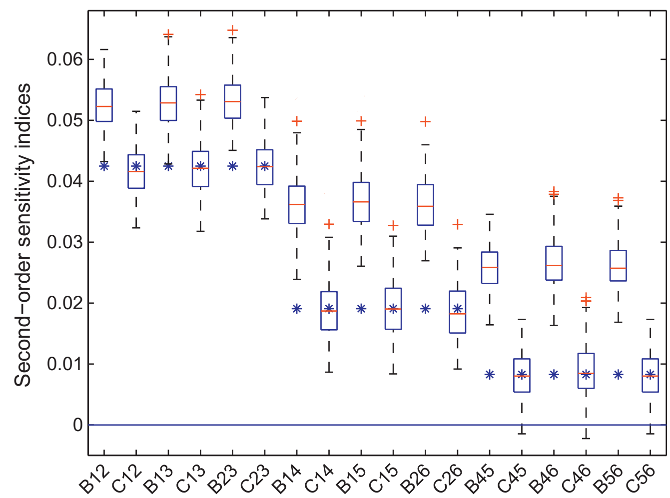

Fig. 5. Estimation of the second-order SI using RBD-FAST. We compare, for a fixed sample size $N=4001$, the basic estimator (Bij) with the bias-corrected one (Cij). In each column, we mark the theoretical SI with a blue asterisk and plot several summaries of a sample of 200 estimator replicates: the 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. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

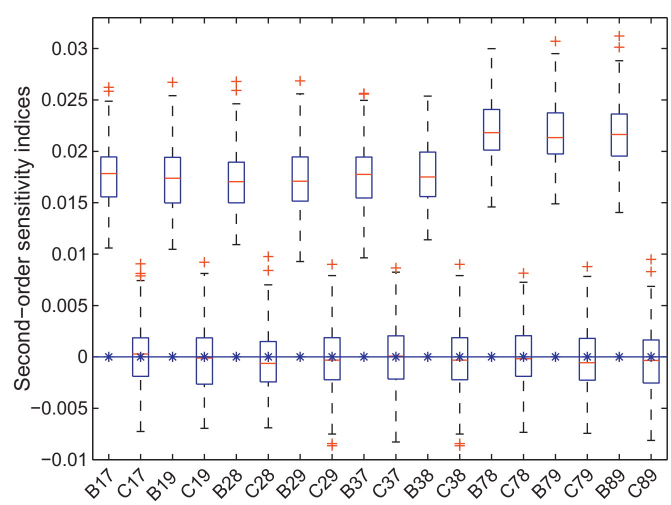

Fig. 6. Estimation of the second-order SI using RBD-FAST. We compare, for a fixed sample size $N=4001$, the basic estimator (Bij) with the bias-corrected one (Cij). In each column, we mark the theoretical SI with a blue asterisk and plot several summaries of a sample of 200 estimator replicates: the 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. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

is proportional to $\lambda_P = n(P)/N$. In this test, we have $n(P) = 2N_1 = 2\times 10 = 20$ for the first-order SI, and $n(P) = 2N_2(N_2-1) = 2\times 7\times(7-1) = 84$ for the second-order SI. Note that the frequency set {177,186,193} satisfies the criterion stated in Eqs. (19) and (20) with parameters $N_1 = 10$, $N_2 = 7$ and $N_3 = 0$.

6.2.2. Comparison with Mara's approach

We now estimate all the first- and second-order SI using the strategy described in Mara [20]. With such an approach, input variables are divided into four groups of two factors and one single term. Hence, nine 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 frequencies $\{\omega_1,\omega_2\} = \{79,83\}$. Note that this frequency set satisfies the criterion stated in Eqs. (19) and (20) with parameters $N_1 = 10$ and $N_2 = 7$. 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 SI 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 SI 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.

Acknowledgments

This work has been partially supported by French National Research Agency (ANR) through COSINUS program (project COSTA-BRAVA no ANR-09-COSI-015). The authors are grateful to the referees for their helpful comments and suggestions.

Appendix A. Details on formula (45)

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. (42) to each of the P_i, we get the linear system

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

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

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

Hence, with the definition in Eq. (43), 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

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

and we conclude that Eq. (45) 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 SI 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

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

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

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

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 $(G_j^i)_{j=1\ldots q}$ is filled with one factor of each original group $(G_k^0)_{k=1\ldots q}$; but (R3) is not always verified. However we can observe that, letting c be a cyclic permutation of order q, the permutations

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

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 us 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.

References

[1] Xu C, Gertner G. Understanding and comparisons of different sampling approaches for the Fourier amplitude sensitivity test (FAST). Computational & Statistics Data Analysis 2011;55(1):184–98.
[2] Saltelli A, Chan KPS, Scott EM. Sensitivity Analysis. John Wiley & Sons; 2000.
[3] Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, et al. Global Sensitivity Analysis: The Primer. John Wiley & Sons; 2008.
[4] Morris MD. Factorial sampling plans for preliminary computational experiments. Technometrics 1991;33(2):161–74.
[5] Borgonovo E. A new uncertainty importance measure. Reliability Engineering and System Safety 2007;92(6):771–84.
[6] Borgonovo E, Castaings W, Tarantola S. Moment independent importance measures: new results and analytical test cases. Risk Analysis 2011;31(3):404–28.
[7] Sobol' IM, Kucherenko S. Derivative based global sensitivity measures and their link with global sensitivity indices. Mathematics and Computers in Simulation 2009;79(10):3009–17.
[8] Sobol' IM, Kucherenko S. A new derivative based importance criterion for groups of variables and its link with the global sensitivity indices. Computer Physics Communications 2010;181(7):1212–7.
[9] Sobol' IM. Sensitivity analysis for nonlinear mathematical models. Mathematical Modeling and Computational Experiment 1993;1:407–14.
[10] Hoeffding WF. A class of statistics with asymptotically normal distributions. Annals of Mathematical Statistics 1948;19:293–325.
[11] Efron B, Stein C. The jackknife estimate of variance. The Annals of Statistics 1981;9(3):586–96.
[12] Cukier RI, Fortuin CM, Shuler KE, Petschek AG, Schaibly JH. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients: theory. Journal of Chemical Physics 1973;59:3873–8.
[13] Schaibly JH, Shuler KE. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients: applications. Journal of Chemical Physics 1973;59:3879–88.
[14] Cukier RI, Schaibly JH, Shuler KE. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients: analysis of the approximations. Journal of Chemical Physics 1975;63:1140–9.
[15] Weyl H. Mean motion. American Journal of Mathematics 1938;60:889–96.
[16] Cukier RI, Levine HB, Shuler KE. Nonlinear sensitivity analysis of multiparameter model systems. Journal of Computational Physics 1978;26:1–42.
[17] Saltelli A, Tarantola S, Chan KPS. A quantitative model-independent method for global sensitivity analysis of model output. Technometrics 1999;41:39–56.
[18] Tarantola S, Gatelli D, Mara TA. Random balance designs for the estimation of first-order global sensitivity indices. Reliability Engineering and System Safety 2006;91:717–27.
[19] Satterthwaite FE. Random balance experimentation. Technometrics 1959;1(2):111–37.
[20] Mara TA. Extension of the RBD-FAST method to the computation of global sensitivity indices. Reliability Engineering and System Safety 2009;94:1274–81.
[21] Plischke E. An effective algorithm for computing global sensitivity indices (EASI). Reliability Engineering and System Safety 2010;95(4):354–60.
[22] McKay MD, Conover WJ, Beckman RJ. 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 1979;21(2):239–45.
[23] Kelley TL. An unbiased correlation ratio measure. Proceedings of the National Academy of Sciences of the United States of America 1935;21(9):554–9.
[24] Saltelli A. Making best use of model evaluations to compute sensitivity indices. Computer Physics Communications 2002;145:208–97.
[25] Saltelli A, Sobol' IM. About the use of rank transformation in sensitivity analysis model. Reliability Engineering and System Safety 1995;50:225–39.