



**HAL**  
open science

## Extension of the RBD-FAST method to the computation of global sensitivity indices

Thierry A. Mara

► **To cite this version:**

Thierry A. Mara. Extension of the RBD-FAST method to the computation of global sensitivity indices. Reliability Engineering and System Safety, 2009, 94, pp.1274-1281. 10.1016/j.ress.2009.01.012 . hal-01093036

**HAL Id: hal-01093036**

**<https://hal.science/hal-01093036>**

Submitted on 10 Dec 2014

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

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

# **Extension of the RBD-FAST method to the computation of global sensitivity indices**

Thierry Alex MARA

University of Reunion Island, LPBS (E.A. 4076), Faculty of Human and Environmental Sciences, 117 Rue General Ailleret, 97430 Tampon, Reunion Island – France.

Tel : +262 262 93 82 12

Fax : +262 262 93 86 65

email : [mara@univ-reunion.fr](mailto:mara@univ-reunion.fr)

## **Abstract**

This paper deals with the sensitivity analysis method named Fourier amplitude sensitivity test (FAST). This method is known to be very robust for the computation of global sensitivity indices but their computational cost remains prohibitive for complex and large dimensional models. Recent developments in the implementation of FAST by use of the random balance designs (RBD) technique have allowed significant reduction of the computational cost. The method is now called RBD-FAST. The drawback of this improvement is that only individual first-order sensitivity indices can be computed. In this article, an extension of RBD is derived for the estimation of any global sensitivity indices of individual factor or group of factors. Several tests are proposed to compare the performances of classical FAST and RBD-FAST.

**Keywords:** Sobol' sensitivity indices, FAST, random balance designs, group of factors

# Table of Contents

<i>Introduction</i> .....	4
<i>1. The Fourier amplitude sensitivity test</i> .....	5
1.1 Computation of the first-order indices.....	5
1.2 Computation of a total sensitivity index.....	7
<i>2. Random balance designs and FAST</i> .....	8
2.1 Illustration on a trivial function.....	8
2.2 Estimation of the first-order sensitivity indices.....	9
2.3 Computation of closed-order sensitivity indices.....	10
2.4 Computation of total sensitivity indices .....	11
<i>3. Numerical tests</i> .....	12
3.1 Computation of the second-order sensitivity indices .....	13
3.2 EFAST versus RBD-FAST .....	14
3.3 Computation of the total sensitivity indices for group of factors .....	16
<i>Conclusion</i> .....	17
<i>References</i> .....	18

## Introduction

Let us denote by  $Y$  a random vector function of a set of  $k$  independent input factors  $\mathbf{X} = \{X_1, X_2, \dots, X_i, \dots, X_k\}$ . Sensitivity analysis of model output is especially an issue when the model at hand is non linear, high dimensional and the uncertain input factors vary generously. Global sensitivity analysis (GSA) is usually employed to compute the sensitivity measures. The latter are generally derived from the analysis of the model response variance (ANOVA). Indeed, it is expected that an influential factor contributes to the variation of the model response. ANOVA-based sensitivity indices (also known as Sobol' indices) can be defined from the following variance decomposition:

$$V[Y] = V[E[Y|\mathbf{X}_i]] + E[V[Y|\mathbf{X}_i]] \quad (1)$$

where  $Y$  is the model response of interest,

$\mathbf{X}_i$  is either an individual factor or a subset of factors,

$V[.]$  is the mathematical variance (the centred second-order moment) and  $V[.|.]$  the conditional variance,

$E[.]$  is the mathematical expectation (the first-order moment) and  $E[.|.]$  the conditional expectation.

$V[E[Y|\mathbf{X}_i]]$  is the partial variance of  $Y$  explained by  $\mathbf{X}_i$  alone. This means that the remaining factors are not involved in this partial variance. In the following,  $\mathbf{X}_{-i}$  denotes the complementary subset so that the entire set of factors (assumed independent) is  $\mathbf{X} = \{\mathbf{X}_i, \mathbf{X}_{-i}\}$ .

For convenience, Eq.(1) is rewritten as follows :

$$1 = S_i + ST_{-i} \quad (2)$$

with  $S_i = V[E[Y|\mathbf{X}_i]] / V[Y]$  is the first-order sensitivity index of  $\mathbf{X}_i$  that represents the amount of the model response variance explained by  $\mathbf{X}_i$  alone.  $ST_{-i} = E[V[Y|\mathbf{X}_i]] / V[Y]$  is the total sensitivity index of  $\mathbf{X}_{-i}$  and provides its total contribution to the model response variance that

is, its marginal as well as its cooperative contributions with  $X_i$ . Eqs. (1-2) are general in the sense that no particular assumption is required so that they hold even for dependent factors. In the literature, many affordable model free computational methods are proposed to estimate the first-order sensitivity indices of individual factor (RBD-FAST [1-2], regression-based methods [2,3], cut-HDMR [4]). There is less as far as group of factors are considered or when higher-order sensitivity indices are investigated (e.g. Sobol' method [5], EFAST [6]). Besides, it has to be noticed that, for models with strictly continuous (or discrete) input/output values, the polynomial chaos expansion is an interesting approach to estimate the Sobol' indices at low computational cost (see [7]).

Originally, RBD-FAST has been proposed to overcome the *curse of dimensionality* inherent to classical FAST (see [1]). The method remains computationally cheap for the estimation of first-order sensitivity indices even for models with many factors. Other authors have extended RBD-FAST to models with correlated factors [8]. In the present paper, exclusively models with independent inputs are discussed. The author proposes different strategies to compute Sobol' indices. Then, some numerical tests are undertaken including comparisons to the so-called extended FAST method (EFAST, see [6]) for the computation of total sensitivity indices.

## **1. The Fourier amplitude sensitivity test**

### **1.1 Computation of the first-order indices**

FAST was originally proposed by Cukier et al. in the 70's to perform sensitivity analysis of a chemical computer model [9]. Later on, the sensitivity measure computed with the classical FAST has been identified as the first-order Sobol' index. FAST exploits Parseval's relationship to decompose the variance of a model response in the frequencies space. For this purpose, the factor values are sampled from a periodic curve with a different frequency  $w_i$

assigned to each factor. Let us set,

$$X_i(s_j) = G_i \left( \sin(w_i s_j + \phi_i) \right), \quad \forall i = 1, 2, \dots, k; \quad s_j = -\pi + \frac{2j\pi}{N}, \quad \forall j = 1, 2, \dots, N, \quad \phi_i \text{ is a random}$$

phase-shift and  $G_i$  is a transformation function chosen in order to ensure that the variable is sampled accordingly with the desired probability density function,  $N$  is the sample size.

The marginal partial variance of an individual factor is obtained from the Fourier coefficients  $A_{nw_i}$  and  $B_{nw_i}$  at the harmonics of  $w_i$  as follows:

$$\hat{D}_i = 2 \sum_{n=1}^M \Lambda_{nw_i} \quad (3)$$

where  $\hat{D}_i$  is an estimate of the marginal partial variance  $V[E[Y|X_i]]$ ,

$M$  is a given integer called interference factor (generally  $< 6$ ),

and  $\Lambda_p = A_p^2 + B_p^2$ .

The Fourier coefficients are numerically estimated by projection,

$$A_p = \frac{1}{N} \sum_{j=1}^N \cos(p s_j) Y(s_j)$$

$$B_p = \frac{1}{N} \sum_{j=1}^N \sin(p s_j) Y(s_j)$$

The estimated total variance is given by Parseval's theorem:

$$\hat{D} = 2 \sum_{n=1}^{N/2} \Lambda_n, \quad (4)$$

Combining Eqs.(3-4) yields to:

$$\hat{S}_i = \hat{D}_i / \hat{D} \quad (5)$$

The drawbacks of this approach are obvious: the choice of the frequency set must be done with great care to avoid aliasing (or interferences that biased the Fourier coefficients estimates) and the highest frequency is proportional to the number of factors (i.e.  $k$ ). As a

consequence, the computational cost  $N$  dramatically increases with  $k$ . Indeed, for a set of frequencies free of interferences to order  $M = 4$ , Cukier et al. in [10] showed that the computational cost is about  $2.6k^{2.5}$ . When the model is additive (i.e.  $\sum S_i = 1$ ), the classical FAST is suited for factors fixing setting (see [11] for details about settings in SA). Otherwise, the computation of higher-order sensitivity indices is necessary.

## 1.2 Computation of a total sensitivity index

In FAST, non linearities generate other frequencies that are linear combinations of the interacting factors frequencies (see [12]). To compute the total sensitivity index of a factor, say  $X_i$ , Saltelli et al. in [6] have proposed to isolate its assigned frequency in the Fourier spectrum ( $w_i$  is high as compared to  $w_{\cdot i}$ ) so that the frequencies generated by its interactions with the other factors remain in the high frequencies of the spectrum (i.e. in the vicinity of  $w_i$ ). Then, the total sensitivity index is computed as follows:

$$\hat{DT}_i = 2 \sum_{n=Mw_{\max}+1}^{N/2} \Lambda_n,$$

where  $\hat{DT}_i$  is an estimate of the total partial variance  $E[V[Y|\mathbf{X}_i]]$ ,

$$w_{\max} = \max(\mathbf{w}_{\cdot i}),$$

the estimate total variance of  $Y$  is still given by Eq.(4) and

$$\hat{ST}_i = \hat{DT}_i / \hat{D} \tag{6}$$

The authors proposed the following steps for the generation of the frequency set:

- fix the interference factor  $M$ ,
- fix a frequency step  $\Delta$ ,
- generate the  $k-1$  first frequencies:  $\mathbf{w}_{\cdot i} = \{1, 1+\Delta, 1+2\Delta, \dots, 1+(k-2)\Delta\}$ ,
- compute the highest frequency :  $w_i = 2M(1+(k-2)\Delta)$



The frequency step is a strictly positive integer that serves to generate the set of low frequencies. The simplest is to choose  $\Delta = 1$  even though it produces interferences. But for fast run models or low dimensional models, it is advised to choose a higher step. The Nyquist limit leads to the expression of the sample size  $N = 2Mw_i + 1 = 4M^2(1+(k-2)\Delta)+1$ . The computational cost can be prohibitive even for the lowest step in the case of models with many factors ( $\Delta = 1, M = 4, \mathbf{w}_i = \{1,2,\dots,k-1\}, w_i = 8(k-1)$  and  $N = 64(k-1)$ ).

## 2. Random balance designs and FAST

### 2.1 Illustration on a trivial function

To reduce the computational cost of the classical FAST, Tarantola et al. [1] employed Satterthwaite's random balance designs technique [13]. In order to introduce the method, let us consider the following simple polynomial function,

$$Y = X_1 + X_2 + 2X_1X_2,$$

where  $X_1$  and  $X_2$  are identically, independently and uniformly distributed over  $[-1,1]$ . The computation of the first-order sensitivity indices with the classical FAST requires the sampling of the variables values along with a periodic curve of different frequencies. Let us assign the frequencies  $\{11,35\}$  to  $X_1$  and  $X_2$  respectively. In order to uniformly sample the input values, the following search curve is employed (see [6] for proof):

$$X_i(s_j) = \frac{2}{\pi} \arcsin\left(\sin(w_i s_j)\right), i = 1,2, s_j = -\pi + \frac{2\pi}{N} j \text{ and } j = 1, \dots, N.$$

In the present numerical application,  $N$  is fixed to 256. Once propagated through the function the resulting spectrum of  $Y$  is plotted in Fig. 1a. As expected, the spectrum contains the marginal contribution of the variables (at frequencies 11 and 35, the harmonics are negligible) as well as their interactions at frequencies 24 and 46 (i.e.  $35 \pm 11$ ). The other contributions are

null. Now, if instead of sampling  $X_1$  periodically one randomly generates its values, then in the spectrum of  $Y$  (see Fig. 1b), its fundamental and interaction frequencies vanish and are replaced by a *white noise* spread all along the spectrum. The level of the noise multiplied by twice the length of the spectrum provides the total partial variance of  $X_1$ .

Let us notice that the random sampling can be achieved by simply permuting the values periodically sampled. This randomization procedure is the basis of the RBD technique. Indeed, let us assume that the two variables are first periodically sampled and then their values independently randomized. The spectrum of  $Y$  will be a white noise. If  $Y^{(1)}$  denotes the response  $Y$  re-ordered so that  $X_1$  describes the original periodic curve then the spectrum of  $Y^{(1)}$  will show some peaks at frequency  $1/l$  and its harmonics (see Fig. 1c). The former provide the marginal partial variance of the variable. The same procedure can be applied to compute the first-order sensitivity index of  $X_2$  and so without extra evaluations of the function.

[Insert Fig. 1 about here]

## 2.2 Estimation of the first-order sensitivity indices

The steps to perform RBD-FAST in order to compute the entire first-order sensitivity indices are the following:

- select  $N$  design points over a periodic curve in the factor space,
- randomly permute the coordinates of the points to generate a set of scrambled points that cover the factor space,
- run the model at each design point,
- for each factor, re-order the model output such that the design points are in increasing

order with respect to  $X_i$  and compute the first-order sensitivity index from Eqs. (3-5)

A sample of  $N$  points is generated using the parametric equation:

$$X_i \left( s^{(i)} \right) = G_i \left( \sin w s^{(i)} \right), \quad \forall i = 1, 2, \dots, k. \quad (7)$$

where  $s^{(i)}$  denotes the  $i$ -th random permutation vector of the  $N$  points  $\{-\pi + 2\pi/N, \dots, -\pi + 2\pi j/N, \dots, \pi\}$ . For each factor  $X_i$ , Eq.(7) provides a different random permutation. Note that, in Eq.(7) one single frequency can be employed and only the estimate of the  $S_i$ 's of individual factor can be computed with one single sample of size  $N$ . The advantage as compared to the classical FAST method is that  $N$  is no longer dependent of the number of inputs and interference factor.

Xu and Gertner generalized RBD-FAST to the case of model with correlated factors defined by a correlation matrix  $\mathbf{C}$  [8]. For this purpose, they use the Iman & Conover's algorithm for generating rank correlated samples (which is in fact a special permutation procedure, see [14]) instead of randomly permuting the design points (in step 2 above).

### 2.3 Computation of closed-order sensitivity indices

In [1], the authors also proposed a hybrid version of RBD-FAST that consists in dividing the factor set into two subsets of equal cardinality. Then, with two frequencies, RBD was applied independently within each group of factors. Let us suppose, for instance, that  $k = 2n$ , then,  $\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2\}$  where  $\mathbf{X}_1 = \{X_1, \dots, X_n\}$  and  $\mathbf{X}_2 = \{X_{n+1}, \dots, X_{2n}\}$ . Let us denote by  $w_1$  and  $w_2$  the two frequencies and  $\{s^{(1)}, s^{(2)}, \dots, s^{(n)}\}$  the set of random paths. Let us denote by  $Y^{(1)}$  the response  $Y$  re-ordered so that  $s^{(1)}$  is sorted in increasing order. Then,  $Y^{(1)}$  provides the first-order sensitivity index of  $X_1$  and  $X_{n+1}$  by considering the Fourier coefficients at frequencies  $Mw_1$  and  $Mw_2$  respectively, and so on for the computation of  $(S_2, S_{n+2}), \dots, (S_n, S_{2n})$ . But actually, the interactions between  $X_j$  and  $X_{n+j}$  also generate frequencies that are linear combinations of  $w_1$  and  $w_2$ . So, to compute an estimate of the closed-sensitivity index of the

couplets, one also has to account for the frequencies  $pw_1 \pm qw_2$  where  $(p,q)$  are integers such that  $p+q \leq M$ . Thus, if the assigned frequencies are  $\{11,35\}$  and  $M = 4$ ,  $X_j$  alone will generate at least the frequency set  $\{11,22,33,44\}$  while  $X_{n+j}$   $\{35,70,105,140\}$  and their interactions up to the fourth-order will generate  $\{46,24,53,13,68,2,81,59,92,48,116,94\}$ . In other words,  $n$  closed second-order sensitivity indices can be computed simultaneously with only one single sample. The approach can be extended to higher-order sensitivity indices. But, by so doing, the number of subsets is increased and one may also need to increase the sample size because of the Nyquist limit. Besides, the determination of the generated frequencies is not a simple task especially when more than two subsets are considered.

## 2.4 Computation of total sensitivity indices

Let us note that, in the RBD-FAST strategy described previously, re-ordering the model response such that the design points are in increasing order with respect to factor  $X_i$  is the same as running the model by randomly generating the values of  $\mathbf{X}_i$  and periodically those of  $X_i$ . To compute, for instance, the total sensitivity index of  $\mathbf{X}_i$  one must generate the  $\mathbf{X}_i$ 's randomly and the  $X_i$ 's periodically with distinct low frequencies. The spectrum of  $Y$  will then show different peaks in the low frequencies (due to  $\mathbf{X}_i$  alone) and a *white noise* in the high frequencies (due to  $\mathbf{X}_i$  alone and its interactions with  $\mathbf{X}_i$ ). Normally, it is expected that the *white noise* spreads all along the spectrum and is stationary that is, its mean remains unchanged all along the spectrum (see Fig. 1b-c). Consequently, by estimating the mean of the spectrum in the high frequencies, one can have an estimate of  $E[V[Y|\mathbf{X}_i]]$ . The steps to compute the total sensitivity indices are:

- fix the interference factor  $M$ ,
- generate  $k$  frequencies : the simplest choice is  $\mathbf{w}_i = \{1,2,3,\dots,k\}$  but if possible avoid frequencies that are harmonics of an already chosen frequency,

- for each factor, generate the  $N$  design points over a periodic curve (with its assigned frequency) in the factor space,  $N = 2(Mk + L)$  where  $L (>100)$  is an arbitrary integer, so that  $[Mk+1, N/2]$  delimits the high frequencies,
- randomly permute the coordinates of the points of the  $\mathbf{X}_i$ 's,
- run the model and, for each output, compute  $\Lambda_p, p = 1, 2, \dots, N/2$ ,
- compute  $\hat{DT}_i = \frac{N}{L} \sum_{n=kM+1}^{N/2} \Lambda_n$ ,  $\hat{D} = 2 \sum_{n=1}^{N/2} \Lambda_n$  and  $\hat{ST}_i = \hat{DT}_i / \hat{D}$ .

Note that, random balance designs can treat group of factors. For this purpose, in step 4 above, all the factors values of the subset shall be subject to the same permutation. Besides, the lowest computational cost of the proposed design ( $N = 2Mk + 200$ ) is significantly cheaper than the one of the extended FAST method,  $N = 4M^2(k-1)$ . But the question is, are the results more accurate? The numerical tests in the next section try to answer this question.

### 3. Numerical tests

The performance of RBD-FAST for  $S_i$ 's estimates of individual factor has already been tested in several articles (see [1,2,15]). So, in the following, only the performance of RBD-FAST for higher order sensitivity indices estimate is tested. The g-function of Sobol' is employed in order to check the accuracy of the methods. Indeed, one interesting features of this function is that analytical sensitivity indices are available (see e.g. [16]).

The g-function is defined by:  $Y = \prod_{i=1}^k g(X_i)$ , where  $g(X_i)$  are given by,  $g(X_i) = \frac{|4X_i - 2| + a_i}{1 + a_i}$ .

The  $X_i$ 's are independent variables defined on the unit hypercube and the  $a_i$ 's are positive integers that condition the importance of the  $X_i$ 's; the smaller  $a_i$ , the higher the effect of  $X_i$  on the function.

To test the robustness of the methods, the bootstrap technique is employed. It consists in

computing the sensitivity indices estimates by resampling the factor space. For this purpose, the factor space is explored with the following search curve:

$$G(s) = \frac{1}{2} + \frac{1}{\pi} \arcsin(\sin(ws + \phi)), \quad (8)$$

Eq.(8) ensures a uniformly distributed sample in the range [0,1]. EFAST requires one single sample path  $s$  and a frequency set  $\mathbf{w}$ . So, when replicating the sampling procedure, a set of phase-shifts  $\phi$  is uniformly sampled over  $(-\pi, \pi)$ . It results that the starting point of the search curve can be anywhere within the factor space. As far as RBD-FAST is concerned, the random permutation of the design points ensures a different random path in the factor space. So, for RBD,  $\phi$  is set to zero.

### 3.1 Computation of the second-order sensitivity indices

For this first application, the g-function of Sobol' is analyzed and six factors are considered,  $\mathbf{a} = \{0, 0.5, 3, 9, 99, 99\}$ . There are  $\binom{6}{2} = 15$  couplets. To compute all the 15 second-order sensitivity indices with RBD-FAST, the factor set is partitioned in two groups of three factors. The frequency set is  $\mathbf{w} = \{11, 35\}$  (already discussed in subsection 2.3) and  $15/3 = 5$  sample sets are necessary. For instance, if the two groups are  $\{X_1, X_2, X_3\}, \{X_4, X_5, X_6\}$  the factors of the couplets  $(X_1, X_4)$ ,  $(X_2, X_5)$  and  $(X_3, X_6)$  are assigned different frequencies (11 and 35 resp.) but the same random design points.

For  $M = 4$ , the maximal frequency expected in the spectrum is 140 (4x35). So a minimum of 280 function evaluations are necessary. The accuracy of the method is tested for  $N = 512, 1024, 2048, 4096$ . One hundred estimates of the  $S_{ij}^c$ 's are computed for each sample size. Finally, the arithmetic mean and the range of the estimated sensitivity indices are computed.

The results are gathered in Fig. 2. As it can be seen, for increasing sample sizes, the estimates

converge to the analytical values. The method is able to compute the closed-sensitivity indices of order 2. Yet, for such a design, the values are biased especially for small sensitivity indices. This result has already been noticed in the case of first-order sensitivity indices estimate [1].

[Insert Fig. 2 about here]

### 3.2 EFAST versus RBD-FAST

In this application, the design described in subsection 2.4 is compared to the Extended FAST as described in subsection 1.2. For this purpose, the same mathematical functions employed by Saltelli et al. [6] are used: the g-function of Sobol' and the Legendre polynomial.

The g-function is employed with a number of factors set to eight. Four test cases are selected: test case A, a very complex *model* as interactions among factors are preponderant ( $a_i = 0$ , for all  $i$ ), test case B, is also a case where all the factors have the same importance but the input/output relationship is additive ( $a_i = 99$ , for all  $i$ ), in case C, the coefficients  $a_i$ 's are set as  $\{0, 1, 4.5, 9, 99, 99, 99, 99\}$ , so that only the three first factors are important and at last, for test case D, the important factors are randomly ranked,  $\mathbf{a} = \{99, 0, 9, 0, 99, 4.5, 1, 99\}$ .

The robustness of the two methods is investigated by repeating the computation of the  $ST_i$ 's 100 times for each of the four cases and for different sample sizes. For each replicate, a total of  $8 \times N$  function evaluations are performed to compute all the  $ST_i$ 's. For EFAST, the frequency set is obtained from the algorithm described in subsection 1.2. As far as RBD-FAST is concerned, the frequency set  $\{1, 2, 3, \dots, 8\}$  is chosen and the design points of one of the factor are randomly permuted (note that  $L = N/2 - 32$ ).

The results of EFAST and RBD-FAST are plotted on Figs. 3 and 4 respectively. Both methods have the same accuracy for  $ST_i$ 's estimates especially at large sample sizes. The lack

of accuracy at low sizes is due to a poor exploration of the factor space. Contrarily to EFAST, RBD-FAST may provide total sensitivity indices greater than 1. Yet, for complex models (case A and B) RBD-FAST seem to give better results. Besides, RBD-FAST is able to compute small  $ST_i$ 's accurately at small sample sizes (see case C and D on Fig. 4). The poor performance of EFAST at small sample sizes is explained by the fact that, for  $N = 128, 256$ , identical frequencies for some of the  $\mathbf{X}_i$ 's are used.

*[Insert Fig. 3 about here]*

*[Insert Fig. 4 about here]*

The second analytical test function is the Legendre polynomials of order  $d$ , denoted by  $L_d(x)$ , also used by Saltelli et al. [6] to compare EFAST and the Sobol' method [5]. Mackay in [17], applied the replicated latin hypercube SA method to compute the first-order sensitivity indices of this function. There are two factors,  $x$  uniformly distributed in  $[-1,+1]$  and  $d = \{1,2,\dots,5\}$  a discrete uniformly distributed variable. The analytical values of the sensitivity indices are:  $S_i = \{0.2,0.0\}$ ,  $ST_i = \{1.0,0.8\}$  for  $\{x,d\}$  respectively. The calculations are performed for  $M = 4$  and  $M = 10$  successively.

The results of EFAST are biased for  $M = 4$  and the estimates converge accurately to the analytical values when  $M = 10$  (see Fig. 5). One can infer that because of the complexity of the function (non linearities) an interference factor of order 4 is not sufficient to capture the factors' total effect. Besides, for  $M = 10$ , EFAST requires a minimum of 400 ( $4M^2$ ) function executions. This explains its poor performance at small sample sizes ( $N = 128, 256$ ). Conversely, RBD-FAST provides lower biased estimates but the variance is higher than EFAST. Besides, it gives sensitivity indices estimates greater than 1 which is not possible.



Nevertheless, it provides satisfactory estimate of total sensitivity indices.

[Insert Fig. 5 about here]

### 3.3 Computation of the total sensitivity indices for group of factors

One interesting feature of the proposed RBD-FAST approach is that group of factors can be dealt with. Let us consider the group of factors  $\mathbf{X}_i = \{X_1, X_2, \dots, X_r\}$ . The first-order sensitivity index of  $\mathbf{X}_i$  is also called the closed effect of order  $r$  of  $\{X_1, X_2, \dots, X_r\}$ . Eq.(2) shows the relationship between a closed effect and the total effect of the complementary subset  $\mathbf{X}_{-i}$ . For instance, for  $r = 3$  we have,

$$S_{123}^c = 1 - ST_{567} = S_1 + S_2 + S_3 + S_{12} + S_{13} + S_{23} + S_{123}$$

To compute the total sensitivity index of a group of factors, one must proceed as follows: select a frequency set (the simplest is  $\{1, 2, \dots, k\}$ ), then sample the factors values accordingly with a periodic curve, randomly permute the values of the factors of the same group (use either a different random permutation per factor or the same one), finally run the model and compute the total sensitivity index of the group (as for an individual factor).

For illustration, let us use again the g-function with six factors where  $\mathbf{a} = \{0, 0.5, 3, 9, 99, 99\}$ . The aim is to estimate all the total second-order sensitivity indices ( $ST_{ij}$ ). The chosen frequency set is  $\{1, 2, 3, \dots, 6\}$  and bootstrapping is again employed to test the robustness of the calculation at different sample sizes. The results are plotted in Fig. 6. One can infer that RBD-FAST provides very good estimates of the sensitivity indices. They converge to the analytical values. In fact, the accuracy is identical to the one obtained for individual total sensitivity indices. Non-significant groups of factors can be detected accurately with the method. This is particularly interesting for model reduction (dimensionality).

[Insert Fig. 6 about here]

In summary, it was shown in the previous tests that RBD-FAST can either provide an estimate of  $V[E[Y|\mathbf{X}_i]]$  or  $E[V[Y|\mathbf{X}_i]]$ . But the computation of the first quantity is less accurate than the second one. This is explained by the fact that the computation of  $V[E[Y|\mathbf{X}_i]]$ , for closed or first-order sensitivity indices estimate, is parametric. Indeed, prior to the calculation of this quantity, one needs to provide a value of the interference factor  $M$ . The latter is directly involved in the calculation of  $V[E[Y|\mathbf{X}_i]]$  since it indicates the frequencies to account for. As a consequence, *aliasing* effects inherent to RBD-FAST slightly bias the calculation. On the other hand, for the computation of  $E[V[Y|\mathbf{X}_i]]$  with RBD-FAST, the interference factor only serves to distinguish the boundary between high and low frequencies.

## Conclusion

Recent previous works have shown that RBD-FAST is efficient to compute the first-order sensitivity indices of computer model factors. In the present paper, RBD-FAST has been extended so that any global sensitivity indices can be evaluated. Numerical tests have demonstrated its efficiency. In particular, compared to the Extended FAST method, RBD-FAST seems to be more efficient at small sample sizes. To sum up, two strategies can be adopted: the first one allows the computation of closed-order sensitivity indices (or first-order sensitivity indices in the case of individual factor) that computes  $n$  (the number of inputs in the group) sensitivity indices with only one single sample. Its drawback is that sensitivity

estimates of non-important inputs are biased at small sample sizes. The second strategy provides accurate estimates of the total sensitivity index of group of factors (or individual factor). The method is efficient and allows the identification of non-important inputs at small sample sizes. Its drawback is that only one sensitivity measure can be evaluated with one single sample. Finally, it has to be noticed that the RBD technique employed here with the Fourier amplitude sensitivity test can also be extended to the sampling-based strategy of Sobol' [5,18] to evaluate any ANOVA-based sensitivity index as already undertaken in [2] for  $S_i$ 's estimates.

## References

- [1] Tarantola, S., Gatelli, D. and Mara, T.A., *Random balance designs for the estimation of first-order global sensitivity indices*, Reliability Engineering and System Safety, Vol. 91, Iss. 6, pp.717-727, 2006.
- [2] Mara T.A., Rakoto Joseph O., *Comparison of some efficient methods to evaluate the main effect of computer model factors*, Journal of Statistical Computation and Simulation, Vol.78, Iss. 2, pp.167-178, 2008.
- [3] Lewandowski D., Cooke R.M., Duintjer Tebbens R.J., *Sample-based estimation of correlation ratio with polynomial approximation*, ACM Transactions on Modeling and Computer Simulation, Vol. V, Iss. N, May 2007, pp.1-16.
- [4] Rabitz H., Alis O., Shorter J. Shim K., *Efficient input-output model representations*, Computer Physics Communication, Vol. 117, pp.11-20, 1999.
- [5] Sobol I.M., *Sensitivity estimates for nonlinear mathematical models*, Mathematical Modelling and Computational Experiments, Vol. 1, Iss. 4, pp.407-414, 1993.

- [6] Saltelli A., Tarantola S. and Chan K., *A quantitative model-independent method for global sensitivity analysis of model output*, Technometrics, Vol. 41, Iss. 1, pp.39-56, 1999.
- [7] Sudret B., *Global sensitivity analysis using polynomial chaos expansions*, Reliability Engineering and System Safety, Vol.93, pp.964-979, 2008.
- [8] Xu C., Gertner G.Z., *A general first-order global sensitivity analysis method*, Reliability Engineering and System Safety, Vol.93, pp.1060-1071, 2008.
- [9] Cukier R.I., Fortuin C. M., Shuler K.E., Petschek A. G. And Schaibly J. H., *Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients: Theory*, Journal of Chemical Physics, Vol. 59, pp. 3873-3878, 1973.
- [10] Cukier R.I., Schaibly J.H., and Shuler K.E., *Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients: Analysis of the approximations*, Journal of Chemical Physics, Vol. 63, pp. 1140-1149, 1975.
- [11] Saltelli A., Tarantola S., *On the relative importance of input factors in mathematical models: Safety assessment for nuclear waste disposal*, Journal of the American Statistical Association, Vol.97, pp.702-709, 2002.
- [12] Mara T.A., Lauret P., *Sampling in the frequency domain to identify a metamodel*, In: proceedings of SAMO 2001, third international symposium on sensitivity analysis of model output, pp. 37-40, Madrid, 2001.
- [13] Satterthwaite F. E., *Random balance experimentation*, Technometrics, Vol.1, pp. 111-137, 1959.
- [14] Iman R.I., Conover W.J., *A distribution-free approach to inducing rank correlation among input variables*, Communications in Statistics: Simulation and Computation, Vol.B11, pp.311-334, 1982.
- [15] Gatelli D., Kucherenko S., Ratto M., Tarantola S., *Calculating first-order sensitivity*

*measures: A benchmark of some recent methodologies*, to appear in Reliability Engineering and System Safety, 2008, doi:10.1016/j.ress.2008.03.028.

[16] Sobol I.M., Tarantola S., Gatelli D., Kucherenko S.S., Mauntz W., *Estimating the approximation error when fixing unessential factors in global sensitivity analysis*, Reliability Engineering and System Safety, Vol.92, pp.957-960, 2007.

[17] McKay M. D., *Evaluating Prediction Uncertainty*, report NUREG/CR-6311, LA-12915-MS, Los Alamos National Laboratory, 1995.

[18] Saltelli A., *Making best use of model evaluations to compute sensitivity indices*, Computer Physics Communication, 145, pp.280-297, 2002.

## Figures Captions

Fig. 1: Illustration of the impact of RBD-FAST on the spectrum of a trivial function. (a) plot of the spectrum when  $x_1$  and  $x_2$  are varied periodically, (b) and (c) are the plots of the spectrum when only  $x_2$  (resp.  $x_1$ ) is varied periodically. In the first plot, the spectrum has peaks at frequencies  $\{11,35,24,46\}$  and is null elsewhere. The frequency set  $\{24,36\}$  is generated by the interaction between the two variables. When one of the two variables is randomized with RBD its interaction and fundamental frequencies are replaced by a noise in the spectrum (b-c). The level of the noise is related to the total contribution of the random variable to the response variance.

Fig. 2: Bootstrap for the closed second-sensitivity indices estimates of the g-function ( $k=6$ ) with RBD-FAST. The error bars represent the ranges of the computed indices for 100 bootstraps estimates and the point (.) is the arithmetic mean. The results are biased especially for small indices.

Fig. 3: Estimation of the total sensitivity indices for the four g-functions with the EFAST method. \* is the analytical values. The error bars represent the ranges of the computed indices for 100 bootstraps estimates and the point (.) is the arithmetic mean. Five sample sizes have been considered:  $N = 128, 256, 512, 1024, 2048$  respectively from the left to the right. The results for case A are less accurate due to the complexity of the model. The method starts to be accurate from  $N = 512$  because of the frequency dependence of EFAST.

Fig. 4: Same as previously except that the RBD-FAST method is employed. The results for case B at small sample sizes are more accurate than EFAST. RBD-FAST is robust and performs as well as EFAST in general.

Fig. 5: Estimation of the total sensitivity indices of the two input factors of the Legendre

polynomial with the RBD-FAST and EFAST. The dashed lines represent the analytical values. The robustness of the method is investigated by use of bootstrapping. The sample sizes employed are :  $N=128, 256, 512, 1024$  and  $2048$ . When the interference factor is underestimated, the methods give biased results (see  $M=4$ ). The impact is particularly important for EFAST.

Fig. 6: Bootstrap for the total sensitivity indices of couple of factors ( $k=6$ ) estimated with RBD-FAST. The results are unbiased and RBD-FAST is especially accurate for small indices, even at small sample sizes.

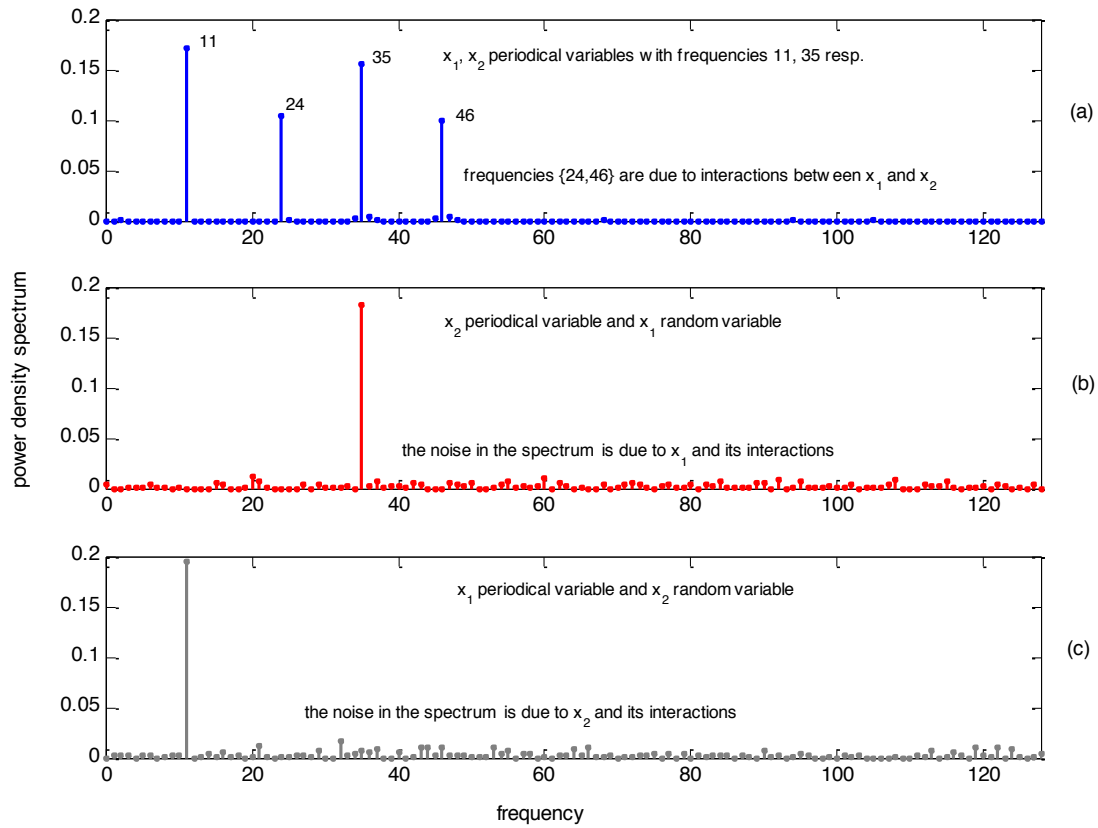


Fig. 1: Illustration of the impact of RBD-FAST on the spectrum of a trivial function. (a) plot of the spectrum when  $x_1$  and  $x_2$  are varied periodically, (b) and (c) are the plots of the spectrum when only  $x_2$  (resp.  $x_1$ ) is varied periodically. In the first plot, the spectrum has peaks at frequencies  $\{11, 35, 24, 46\}$  and is null elsewhere. The frequency set  $\{24, 36\}$  is generated by the interaction between the two variables. When one of the two variables is randomized with RBD its interaction and fundamental frequencies are replaced by a noise in the spectrum (b-c). The level of the noise is related to the total contribution of the random variable to the response variance.



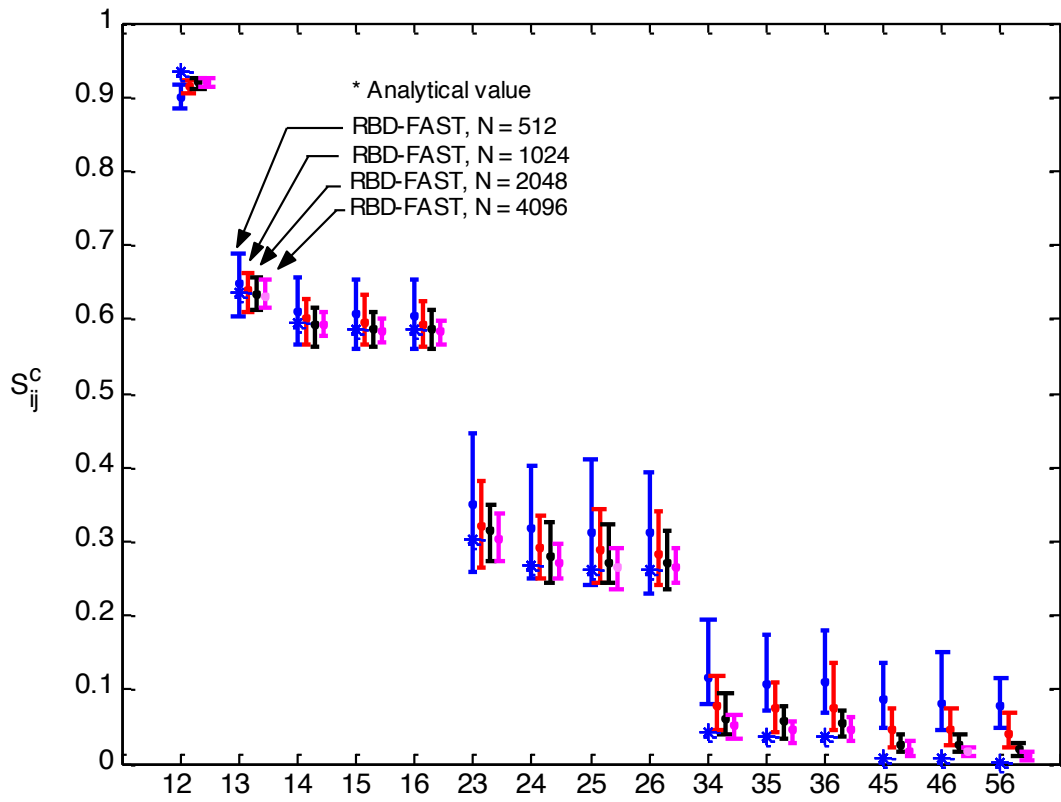


Fig. 2: Bootstrap for the closed second-sensitivity indices estimates of the  $g$ -function ( $k=6$ ) with RBD-FAST. The error bars represent the ranges of the computed indices for 100 bootstraps estimates and the point (.) is the arithmetic mean. The results are biased especially for small indices.

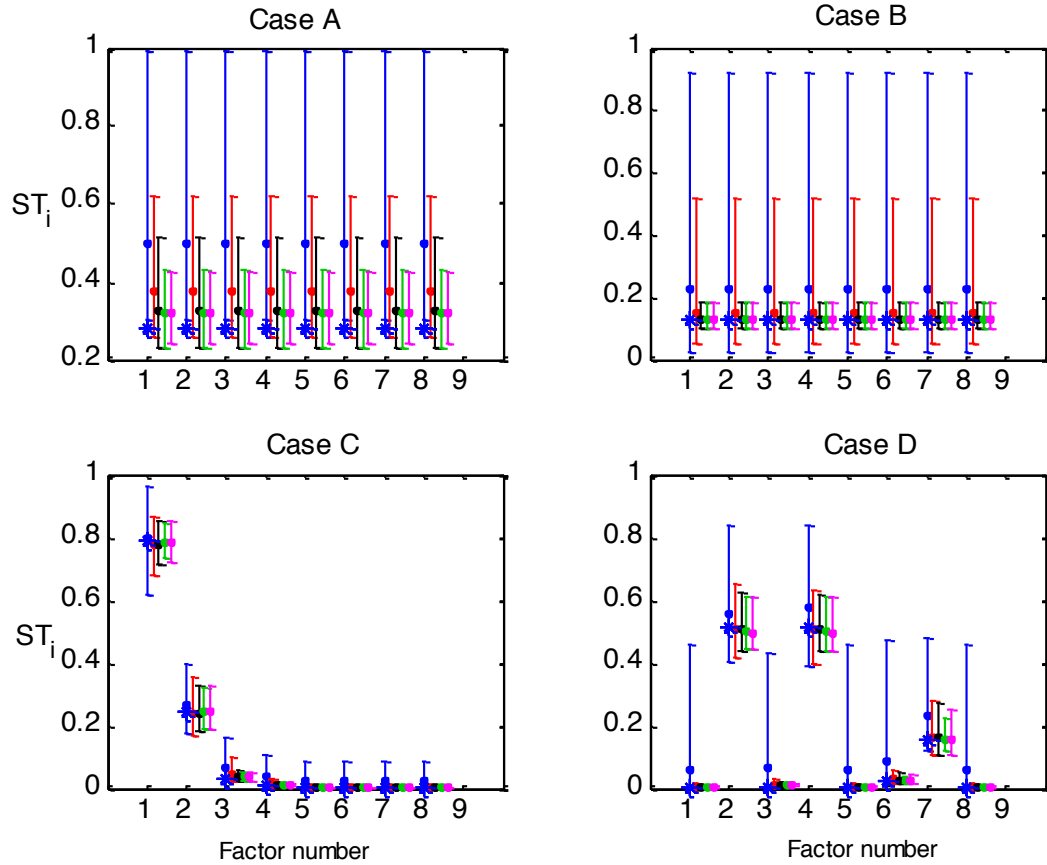


Fig. 3: Estimation of the total sensitivity indices for the four  $g$ -functions with the EFAST method. \* is the analytical values. The error bars represent the ranges of the computed indices for 100 bootstraps estimates and the point (.) is the arithmetic mean. Five sample sizes have been considered:  $N = 128, 256, 512, 1024, 2048$  respectively from the left to the right. The results for case A are less accurate due to the complexity of the model. The method starts to be accurate from  $N = 512$  because of the frequency dependence of EFAST.

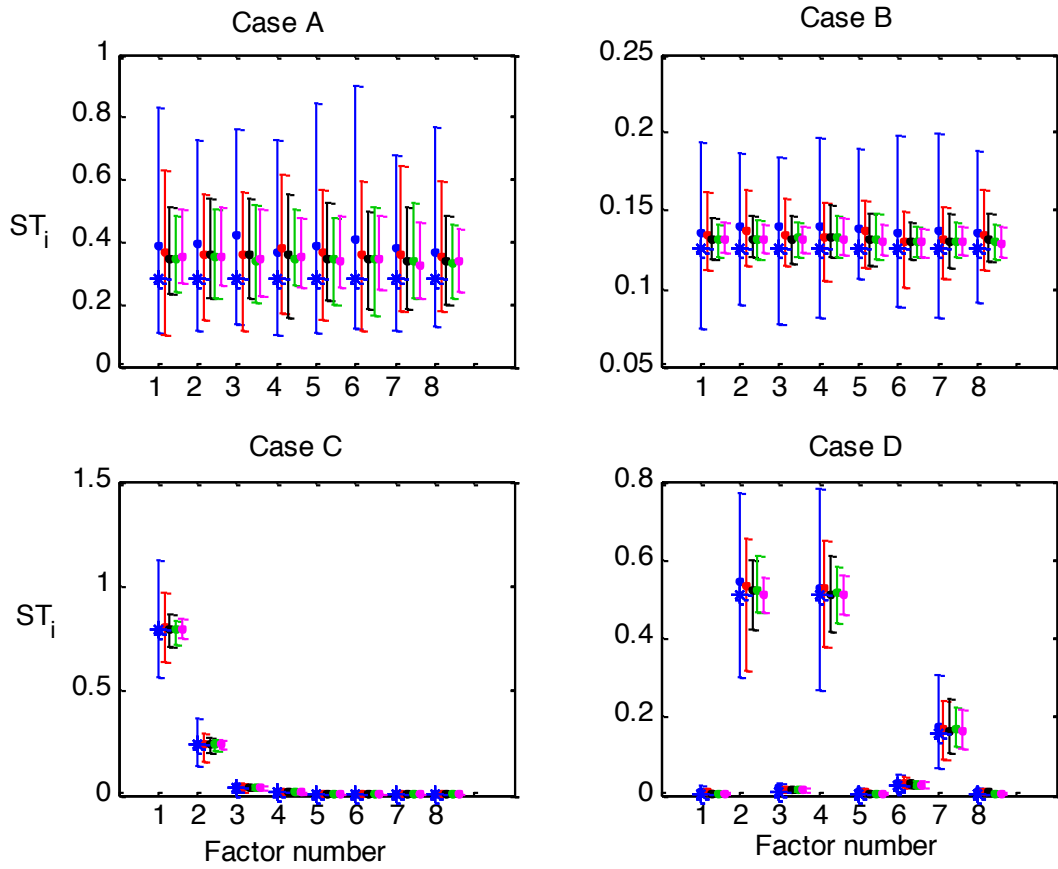


Fig. 4: Same as previously except that the RBD-FAST method is employed. The results for case B at small sample sizes are more accurate than EFAST. RBD-FAST is robust and performs as well as EFAST in general.

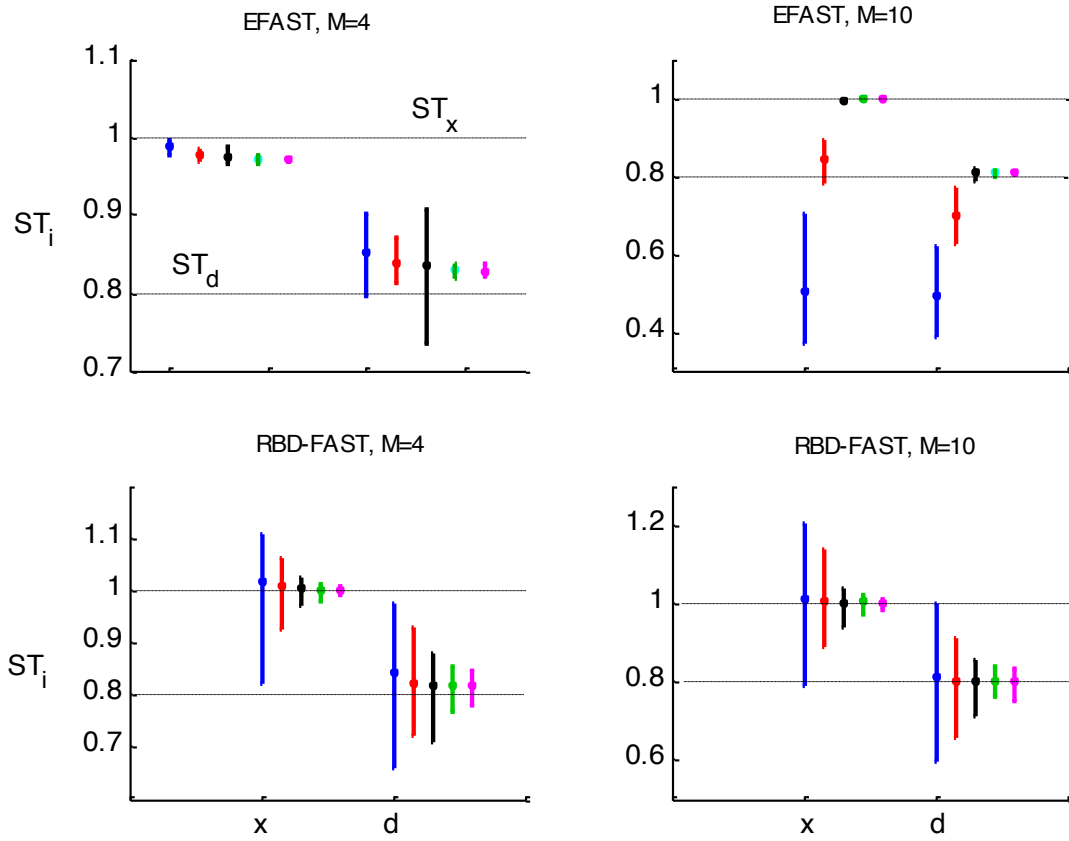


Fig. 5: Estimation of the total sensitivity indices of the two input factors of the Legendre polynomial with the RBD-FAST and EFAST. The dashed lines represent the analytical values. The robustness of the method is investigated by use of bootstrapping. The sample sizes employed are :  $N=128, 256, 512, 1024$  and  $2048$ . When the interference factor is underestimated, the methods give biased results (see  $M=4$ ). The impact is particularly important for EFAST.

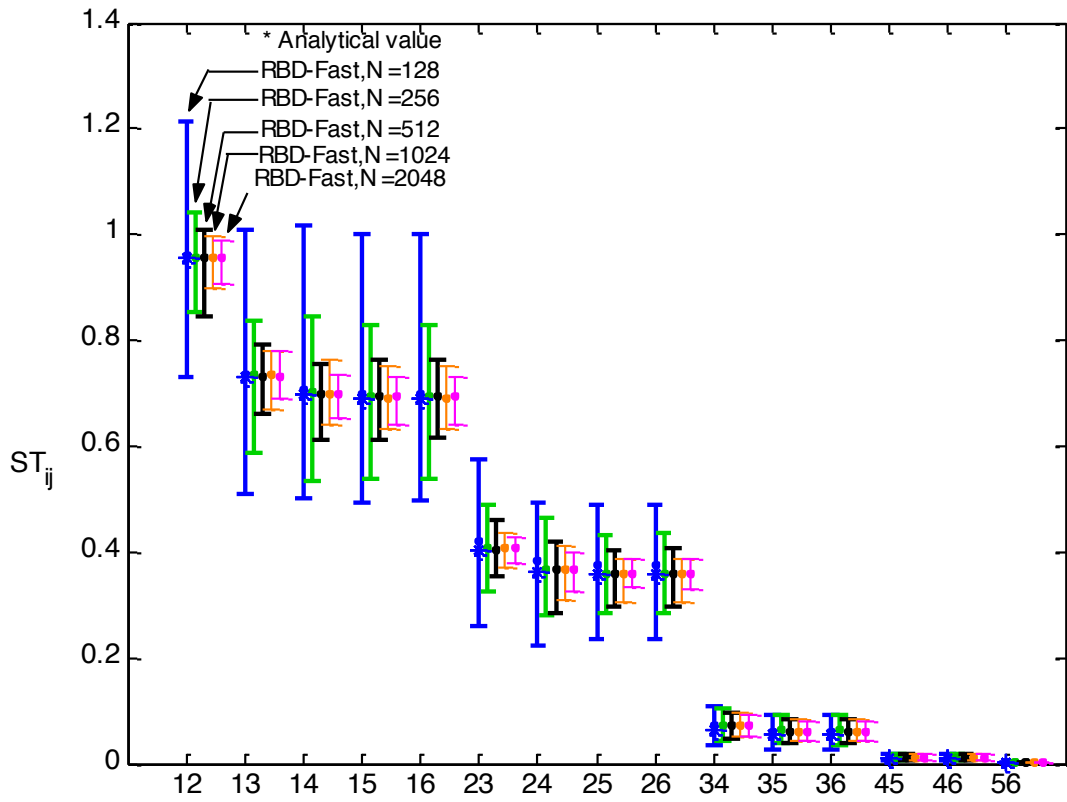


Fig. 6: Bootstrap for the total sensitivity indices of couple of factors ( $k=6$ ) estimated with RBD-FAST. The results are unbiased and RBD-FAST is especially accurate for small indices, even at small sample sizes.