



HAL
open science

Total interaction index: A variance-based sensitivity index for interaction screening

Jana Fruth, Olivier Roustant, Sonja Kuhnt

► **To cite this version:**

Jana Fruth, Olivier Roustant, Sonja Kuhnt. Total interaction index: A variance-based sensitivity index for interaction screening. 2012. hal-00631066v3

HAL Id: hal-00631066

<https://hal.science/hal-00631066v3>

Preprint submitted on 27 Apr 2012 (v3), last revised 24 Jul 2013 (v5)

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.

Total interaction index: A variance-based sensitivity index for interaction screening.

J. Fruth^a, O. Roustant^b, S. Kuhnt^a

^a*Faculty of Statistics, TU Dortmund University, Vogelpothsweg 87, 44221 Dortmund, Germany*

^b*Ecole Nationale Supérieure des Mines, FAYOL-EMSE, LSTI, F-42023 Saint-Etienne, France*

Abstract

We consider the problem of investigating the interaction structure of a multivariate function, possibly containing high order interactions, through variance-based indices. By analogy with the total index, used to detect the most influential variables, a screening of interactions can be done with the so-called total interaction index (TII), defined as the superset importance of a pair of variables. Our aim is to investigate the TII. At the theoretical level, it is connected to total and closed indices, and it is shown that the TII is obtained by averaging the second order interaction of a 2-dimensional function obtained by fixing the original one. We then present several estimation methods and prove the asymptotical efficiency of an estimator from Liu and Owen (2006). Its superiority is also confirmed empirically. Finally, an application is given to recover a block-additive structure of a function, without knowledge about the interaction orders nor about the blocks.

Keywords: Sensitivity analysis, FANOVA decomposition, Sobol indices, High-order interactions, Superset importance, Additive structure

1. Introduction

Global sensitivity analysis has broad applications for screening, interpretation and reliability analysis [1, 2]. A common method is the estimation of sensitivity indices which quantify the importance of an input, or a group

Email addresses: fruth@statistik.tu-dortmund.de (J. Fruth), roustant@emse.fr (O. Roustant), kuhnt@statistik.tu-dortmund.de (S. Kuhnt)

of variables, for the behavior of an output variable. First-order indices and closed indices are then used to quantify the influence of variables or groups of variables [3]. Homma and Saltelli [4] introduced total indices which quantify the influence of a variable and its interactions and can be used for variable screening: If the total index of a variable is zero, this variable can be removed because neither the variable nor its interactions (at any order) have an influence.

In practice, less effort is done to investigate the interaction structure. For instance, the detection of active interactions - or *interaction screening* - is often limited to second order ones. One technical reason is that the interactions of higher orders are defined by recursion, depending on the smaller orders ones. Their computation then faces the *curse of dimensionality* [5]. It is also often advocated that the assumption 'second-order interactions only' is reasonable. However, the investigation of complex phenomena is spreading, and it is desirable to deal with the situation where this assumption is not, or partially, valid.

Among the advantages of screening interactions, is the possibility to visualize them with a graph and to recover block-additive structures from it, as pointed out by Hooker [6], or Muehlenstaedt et al. [7] in the context of computer experiments. This is especially useful for optimization (since it separates into lower dimensional problems) as well as in metamodelling (since the kernel structure can be chosen accordingly).

By analogy with the total index, used to detect the most influential variables, interaction screening can be done with the so-called *total interaction index* (TII), defined as the superset importance (Liu and Owen [8]) of a pair of variables. The TII of a pair of variables $\{X_i, X_j\}$ is thus defined as the variance of an output explained by the two variables X_i and X_j simultaneously. In the usual case of independent inputs, it corresponds to the sum of their (unnormalized) second order interactions and all higher interactions containing both X_i and X_j . If the index is zero, there are no interactions at any order between X_i and X_j , and the pair $\{X_i, X_j\}$ can be removed from the list of interactions, in a similar way as one can remove a variable if the total index is zero. Remark, however, that removing an interaction does not remove the individual variables in the pair.

Once defined, it is worth noting that the TII can be computed in an acceptable numerical cost, that is proportional to the square of the problem dimension, and thus does not face the curse of dimensionality. Our aim is to investigate the total interaction index. At theoretical level, the TII is connected to the total indices of a pair of variables, which is the portion of variance explained by *at least* one variable in a pair. This induces an immediate connection to closed effects of groups of variables as well. Furthermore, we show that the TII is obtained by integrating out the second order interaction of a *2-dimensional* function obtained by fixing the original one. The estimator proposed by Liu and Owen [8] can be interpreted this way. This property also makes the link between TII and the indices introduced in Muehlenstaedt et al. [7]. From that study, we compare several estimation procedures, some of them relying on the new developments of FAST (Cukier et al. [9]) and RBD-FAST (Mara [10]) techniques. The main result is that the estimator proposed by Liu and Owen is asymptotically normal and efficient. An empirical comparison of the estimation methods confirms its superiority, assuming a same number of functions evaluations.

The paper is structured as follows. Section 2 presents the main theoretical results concerning total interaction indices, after giving a quick overview of FANOVA decomposition and Sobol indices. Then in Section 3, several estimation methods are deduced, and compared empirically in Section 4. The asymptotical properties of the Liu and Owen estimator are proved in section 5. Finally in Section 6, the total interaction indices are used to recover the block-additive decomposition of a 6-dimensional function as a sum of two 3-dimensional ones.

2. Theoretical aspects

2.1. A quick overview of FANOVA decomposition and Sobol Indices

Assume that the input factors X_1, \dots, X_d are independent random variables, and let ν denote the probability measure of $\mathbf{X} = (X_1, \dots, X_d)$. Then for any function $f \in L^2(\nu)$, the functional ANOVA decomposition provides a unique decomposition into additive terms

$$f(\mathbf{X}) = \mu_0 + \sum_{i=1}^d \mu_i(X_i) + \sum_{i < j} \mu_{ij}(X_i, X_j) + \dots + \mu_{1, \dots, d}(X_1, \dots, X_d).$$

The terms represent main effects ($\mu_i(X_i)$), second-order interactions ($\mu_{ij}(X_i, X_j)$) and all higher combinations of input variables. For uniqueness two conditions have to hold [11]:

$$E(\mu_I(X_I)) = 0, \quad I \subseteq \{1, \dots, d\} \quad (1)$$

and

$$E(\mu_{ii'}(X_i X_{i'}) | X_i) = E(\mu_{ii'i''}(X_i X_{i'} X_{i''}) | X_i X_{i'}) = \dots = 0. \quad (2)$$

From (1) and (2) it follows that

$$E(\mu_I(X_I)\mu_{I'}(X_{I'})) = 0, \quad I \neq I'. \quad (3)$$

The decomposition can be obtained by recursive integration:

$$\mu_0 = E(f(X)),$$

$$\mu_i(X_i) = E(f(X)|X_i) - \mu_0,$$

$$\mu_{ij}(X_i, X_j) = E(f(X)|X_i, X_j) - \mu_i(X_i) - \mu_j(X_j) - \mu_0$$

and more generally

$$\mu_I(X_I) = E(f(X)|X_I) - \sum_{I' \subsetneq I} \mu_{I'}(X_{I'}). \quad (4)$$

By computing the variance, an ANOVA-like variance decomposition is obtained where each part quantifies the impact of the input variables on the response.

$$D = \text{var}(f(\mathbf{X})) = \text{var}(\mu_0) + \sum_{i=1}^d \text{var}(\mu_i(X_i)) + \sum_{i < j} \text{var}(\mu_{ij}(X_i, X_j)) \\ + \dots + \text{var}(\mu_{1, \dots, d}(X_1, \dots, X_d)).$$

Those variances are widely used as indices for the influence of input variables and their interactions (Sobol indices). In this paper we only look at the variances and ignore the usual normalizing by the overall variance (D) for the sake of simplicity.

$$D_I = \text{var}(\mu_I(X_I)). \quad (5)$$

There are several extensions to the standard Sobol indices as given in (5). The total effect index D_i^T [2] of a single input variable X_i describes the total contribution of the variable including all interactions and is defined by the sum of all indices containing i :

$$D_i^T = \sum_{J \ni \{i\}} D_J.$$

It is straightforward to extend this index to groups of variables X_I , for any $I \subseteq \{1, \dots, d\}$, by the sum of all indices containing at least one of the variables:

$$D_I^T = \sum_{\substack{J \\ J \cap I \neq \emptyset}} D_J. \quad (6)$$

Another way to describe the influence of a group of variables is the closed index D_I^C , see e.g. [10]. In contrast to total indices, interactions with variables not in X_I are not included here, but all effects caused by subsets of it. It is equal to the so-called variance of the conditional expectation (VCE) and for main effects it matches with the standard Sobol index.

$$D_I^C = \text{var}(E[f(\mathbf{X})|X_I]) = \sum_{J \subseteq I} D_J. \quad (7)$$

If we define by $-I$ the complementary subset to I ($-I = \{1, \dots, d\} \setminus I$), we obtain from (6) and (7) the well-known relation (see e.g. [10]):

$$D = D_{-I}^C + D_I^T \quad (8)$$

and in particular, with the formula of total variance, one can deduce that the total index relatively to I is equal to the expectation of the conditional variance (ECV) relatively to the complementary subset $-I$:

$$D_I^T = \text{E}(\text{var}[f(\mathbf{X})|X_{-I}]). \quad (9)$$

2.2. Total interaction indices

Next we aim at an index which measures the portion of variance of an output explained by two input variables simultaneously, which we then call total interaction index.

Definition 1. With the notations and assumptions of section 2.1, the *total interaction* index \mathfrak{D}_{ij} of two variables X_i and X_j is defined by:

$$\mathfrak{D}_{ij} := \text{var} \left(\sum_{I \supseteq \{i,j\}} \mu_I(X_I) \right) = \sum_{I \supseteq \{i,j\}} D_I. \quad (10)$$

The index equals the superset importance, introduced by Liu and Owen [8], for the pair of indices (X_i, X_j) , which aims to give a measure of importance of interactions and their supersets.

It is not difficult to see that the total interaction index is connected to total indices, as well as closed indices:

Proposition 1. *The following relations hold:*

$$\mathfrak{D}_{ij} = D_i^T + D_j^T - D_{i,j}^T \quad (11)$$

$$\mathfrak{D}_{ij} = D + D_{-\{i,j\}}^C - D_{-i}^C - D_{-j}^C \quad (12)$$

Proof. (12) is deduced from (11) using (8). For (11), the results come from the identity:

$$\sum_{I \supseteq \{i\} \text{ or } I \supseteq \{j\}} D_I = \sum_{I \supseteq \{i\}} D_I + \sum_{I \supseteq \{j\}} D_I - \sum_{I \supseteq \{i,j\}} D_I.$$

□

The following proposition shows that it is also possible to compute the total interaction indices by integration of second order interactions index of 2-dimensional functions (fixing method):

Proposition 2. *For any $x_{-\{i,j\}}$, define f_{fixed} as the 2-dimensional function $(x_i, x_j) \rightarrow f(\mathbf{x})$ obtained from f by fixing all variables except x_i and x_j . Let $D_{i,j|x_{-\{i,j\}}}$ denote the second order interaction index of $f_{\text{fixed}}(X_i, X_j)$, which depends on the fixed variables $x_{-\{i,j\}}$. Then the total interaction index of X_i and X_j is obtained by integrating $D_{i,j|x_{-\{i,j\}}}$ with respect to $x_{-\{i,j\}}$:*

$$\mathfrak{D}_{ij} = E \left(D_{i,j|X_{-\{i,j\}}} \right). \quad (13)$$

Proof. A direct connection of $E\left(D_{i,j|X_{-\{i,j\}}}\right)$ to the definition (10) of total interaction indices can be obtained by considering the FANOVA decomposition of f_{fixed} . This approach is detailed in the appendix. However, a simpler approach is to connect the three terms composing the second order interaction of f_{fixed} to the total indices of $f(\mathbf{X})$. Denote respectively by $D_{|x_{-\{i,j\}}}$, $D_{i|x_{-\{i,j\}}}$ and $D_{j|x_{-\{i,j\}}}$ the variance and the main effects of f_{fixed} . Since f_{fixed} is 2-dimensional, there is a unique (second order) interaction, given by

$$D_{i,j|X_{-\{i,j\}}} = D_{|X_{-\{i,j\}}} - D_{i|X_{-\{i,j\}}} - D_{j|X_{-\{i,j\}}} \quad (14)$$

and thus

$$E\left(D_{i,j|X_{-\{i,j\}}}\right) = E\left(D_{|X_{-\{i,j\}}}\right) - E\left(D_{i|X_{-\{i,j\}}}\right) - E\left(D_{j|X_{-\{i,j\}}}\right). \quad (15)$$

Consider each term separately:

- The variance of f_{fixed} , first, is given by

$$D_{|X_{-\{i,j\}}} = \text{var} [f(\mathbf{X})|X_{-\{i,j\}}]$$

which implies with (9) that

$$E\left(D_{|X_{-\{i,j\}}}\right) = E\left(\text{var} [f(\mathbf{X})|X_{-\{i,j\}}]\right) = D_{i,j}^T.$$

- For the main effect of f_{fixed} explained by X_i , using (7) or directly the FANOVA decomposition in (4) the index is equal to:

$$\begin{aligned} D_{i|X_{-\{i,j\}}} &= \text{var} [E[f_{\text{fixed}}(X_i, X_j)|X_i] | X_{-\{i,j\}}] \\ &= \text{var} [E[f(\mathbf{X})|X_{-j}] | X_{-\{i,j\}}]. \end{aligned}$$

Now use the total variance formula, conditional to $X_{-\{i,j\}}$:

$$D_{i|X_{-\{i,j\}}} = \text{var} [f(\mathbf{X})|X_{-\{i,j\}}] - E[\text{var} [f(\mathbf{X})|X_{-j}] | X_{-\{i,j\}}]$$

Thus, by integrating w.r. to $X_{-\{i,j\}}$, and using (9) again, we get

$$\begin{aligned} E\left(D_{i|X_{-\{i,j\}}}\right) &= E\left(\text{var} [f(\mathbf{X})|X_{-\{i,j\}}]\right) - E\left(\text{var} [f(\mathbf{X})|X_{-j}]\right) \\ &= D_{i,j}^T - D_j^T. \end{aligned}$$

- Similarly, we have

$$\mathbb{E} \left(D_{j|X_{-\{i,j\}}} \right) = D_{i,j}^T - D_i^T.$$

Finally, from (15) and (11), we obtain:

$$\begin{aligned} \mathbb{E} \left(D_{i,j|X_{-\{i,j\}}} \right) &= D_{i,j}^T - (D_{i,j}^T - D_i^T) - (D_{i,j}^T - D_j^T) \\ &= D_i^T + D_j^T - D_{i,j}^T = \mathfrak{D}_{ij}. \end{aligned}$$

□

Remark. Proposition 2 shows the equality between the sensitivity index defined in [7] and the total interaction index (10).

As the total interaction index is equal to the superset importance of a pair of indices, another way of computation is given by:

Proposition 3. (*Liu and Owen [8]*)

$$\begin{aligned} \mathfrak{D}_{ij} &= \frac{1}{4} \int [\Delta_{i,j}(z_{i,j}, x_{i,j}, x_{-\{i,j\}})]^2 d\nu_{i,j}(z_{i,j}) d\nu(\mathbf{x}) \\ &= \frac{1}{4} E [\Delta_{i,j}(Z_{i,j}, X_{i,j}, X_{-\{i,j\}})^2] \end{aligned} \quad (16)$$

with $\Delta_{i,j}(z_{i,j}, x_{i,j}, x_{-\{i,j\}}) =$

$$f(x_i, x_j, x_{-\{i,j\}}) - f(x_i, z_j, x_{-\{i,j\}}) - f(z_i, x_j, x_{-\{i,j\}}) + f(z_i, z_j, x_{-\{i,j\}})$$

and where Z_i (resp. Z_j) is an independent copy of X_i (resp. X_j).

Remark. The formula in Proposition 3 can itself be interpreted as a fixing method, in the sense of Proposition 2. Indeed, when fixing $X_{-\{i,j\}}$, the second-order interaction of the 2-dimensional fixed function is equal to its total interaction index, and thus given by Proposition 3 as:

$$\begin{aligned} D_{i,j|X_{-\{i,j\}}} &= \frac{1}{4} E [f_{\text{fixed}}(X_i, X_j) - f_{\text{fixed}}(X_i, Z_j) - f_{\text{fixed}}(Z_i, X_j) + f_{\text{fixed}}(Z_i, Z_j)]^2 \\ &= \frac{1}{4} E [(\Delta_{i,j}(Z_{i,j}, X_{i,j}, X_{-\{i,j\}}))^2 | X_{-\{i,j\}}] \end{aligned}$$

Taking the expectation gives the result.

3. Estimation methods

In this section, we treat different estimation methods for the computation of total interaction indices. The theoretical expressions (11), (12), (13) and (16) suggest different specific estimation methods. The first three ones rely respectively on RBD-FAST, Sobol, and FAST estimation methods. First the underlying FAST method is quickly reviewed. Then these methods are presented together with some remarks on their properties.

3.1. Review of FAST

The Fourier amplitude sensitivity test (FAST) by [9] is a very efficient method to estimate first order Sobol indices. Sample points of X are chosen such that the indices can be interpreted as amplitudes obtained by Fourier analysis of the function. More precisely the design of N points is such that

$$x_i^{(k)} := G_i(\sin(\omega_i s_k)), \quad i = 1, \dots, d, \quad k = 1, \dots, N, \quad s_k = \frac{2\pi(k-1)}{N}$$

with G_i functions to ensure that the sample points follow the distribution of X . The set of integer frequencies $\{\omega_1, \dots, \omega_d\}$ associated to the input variables is chosen as "free of interferences" as possible; *free of interferences up to the order M* means that $\sum_{i=1}^p a_i \omega_i \neq 0$ for $\sum_{i=1}^p |a_i| \leq M + 1$ [12]. In practice, $M = 4$ or 6 .

The Fourier coefficients for each variable can then be numerically estimated by

$$\begin{aligned} A_\omega &= \frac{1}{N} \sum_{j=1}^N f(x(s_j)) \cos(\omega s_j), \\ B_\omega &= \frac{1}{N} \sum_{j=1}^N f(x(s_j)) \sin(\omega s_j), \end{aligned}$$

and the main effects' indices can be estimated by the sum of the corresponding amplitudes up to the order M :

$$\widehat{D}_i = 2 \sum_{p=1}^M (A_{p\omega_i}^2 + B_{p\omega_i}^2).$$

An estimate of the overall variance is given by the sum of all amplitudes

$$\widehat{D} = 2 \sum_{n=1}^{N/2} (A_n^2 + B_n^2). \quad (17)$$

3.2. The four estimators

3.2.1. Estimation with RBD-FAST, via total indices

The computation of a total index of groups of variables is possible with an RBD-FAST method. RBD-FAST is a group of modifications of classical FAST which use random permutations of design points to avoid interferences [10]. To compute the RBD-FAST estimator of the total index of a group of variables \widehat{D}_I^T simple frequencies like $\omega = \{1, \dots, d\}$ are assigned to the variables. Then $N = 2(Md + L)$ design points are generated over a periodic curve where M denotes the fix inference factor (usually 4 or 6) and $L (> 100)$ is a selectable integer number regulating the sample size. The values of the factors in I are then randomly permuted (either differently per factor or identically) and the model is evaluated at the points. The total index is estimated by

$$\widehat{D}_I^T = \frac{N}{L} \sum_{p=dM+1}^{N/2} (A_p^2 + B_p^2).$$

The estimator corresponding to (11) is then given by:

$$\widehat{\mathfrak{D}}_{ij} = \widehat{D}_i^T + \widehat{D}_j^T - \widehat{D}_{\{i,j\}}^T. \quad (18)$$

3.2.2. Estimation with Sobol method, via closed indices

It is also possible to compute closed indices with an RBD-FAST method, which is called *hybrid version* in Mara [10]. But, as in classical FAST, frequencies that are free of interferences are needed. Here to apply (12), the estimation of the closed index D_{-i}^C is necessary which requires a number of $d - 1$ free of interference frequencies. Those frequencies are, especially for high dimensions, not easy to find. Therefore an alternative way to get closed indices, Monte Carlo integration [3], is considered. To obtain the closed index of a group of variables X_I a large number (n_{Sobol}) of random numbers from the distribution of X is sampled and another n_{Sobol} random numbers are sampled for the remaining variables X_{-I} . Denote by $x^{*k} = (x_I^{*k}, x_{-I}^{*k})$ and z_{-I}^{*k} these two samples for $k = 1, \dots, n_{\text{Sobol}}$. The closed index of X_I is then estimated by

$$\widehat{D}_I^C = \frac{1}{n_{\text{Sobol}}} \sum_{k=1}^{n_{\text{Sobol}}} f(x_I^{*k}, x_{-I}^{*k}) f(x_I^{*k}, z_{-I}^{*k}) - \hat{\mu}_0^2 \quad (19)$$

with

$$\hat{\mu}_0 = \sum_{k=1}^{n_{\text{Sobol}}} f(x_I^{*k}, x_{-I}^{*k}).$$

Consequently, with (12), the corresponding estimator for the total interaction index is given by

$$\widehat{\mathfrak{D}}_{ij} = \widehat{D} + \widehat{D}_{-\{i,j\}}^C - \widehat{D}_{-i}^C - \widehat{D}_{-j}^C, \quad (20)$$

where \widehat{D} is the estimation of the variance calculated by the sample variance of (x_I^{*k}, x_{-I}^{*k}) . One may remark that the additional sampling required in the Sobol method is quite economic here, since the complementary subsets $-\{i, j\}$, $-\{i\}$ and $-\{j\}$ have a very small size.

3.2.3. Fixing method using FAST

Following proposition 2, the total interaction index can be computed according to the following scheme:

For $k = 1, \dots, n_{\text{MC}}$, do:

1. Simulate $X_{-\{i,j\}}^{*k}$ from the distribution of $X_{-\{i,j\}}$,
2. Fix all variables except for $\{X_i, X_j\}$ to $X_{-\{i,j\}}^{*k}$, and create the corresponding 2-dimensional function f_{fixed} ,
3. Compute the second order interaction index of f_{fixed} , denoted $\widehat{D}_{i,j|X_{-\{i,j\}}}^k$, by removing the main effects indices from the overall variance as in (14).

Finally, compute the estimator

$$\widehat{\mathfrak{D}}_{ij} = \frac{1}{n_{\text{MC}}} \sum_{k=1}^{n_{\text{MC}}} \widehat{D}_{i,j|X_{-\{i,j\}}}^k. \quad (21)$$

This estimation method seems to be greedy, due to the additional loop to simulate $X_{-\{i,j\}}$. On the other hand, only 2-dimensional functions are considered, utilizing efficient techniques to compute the interaction index in step 3. For that purpose, we suggest the FAST method, since the computation is both quick, and returns a *positive* value provided that the frequency parameters are free of interferences and that the number of FAST evaluations, denoted n_{FAST} , is large enough, as we see in the next section.

3.2.4. Fixing method by Liu and Owen

Liu and Owen suggest the estimation by Monte Carlo integration of the integral in proposition 3, similar to the closed index estimation in (19). Denote by \mathbf{x}^k and \mathbf{z}^k , $k = 1, \dots, n_{\text{LO}}$ two independent samples of length n_{LO} drawn from ν . Then the total interaction index is estimated by

$$\widehat{\mathfrak{D}}_{ij} = \frac{1}{4} \times \frac{1}{n_{\text{LO}}} \sum_{k=1}^{n_{\text{LO}}} [f(x_i^k, x_j^k, x_{-\{i,j\}}^k) - f(x_i^k, z_j^k, x_{-\{i,j\}}^k) - f(z_i^k, x_j^k, x_{-\{i,j\}}^k) + f(z_i^k, z_j^k, x_{-\{i,j\}}^k)]^2. \quad (22)$$

3.3. Some properties of the four estimators

3.3.1. Positivity

With the indices being theoretically non-negative, also the estimates should be non-negative. This applies in any case for the estimator by the *fixing method by Liu and Owen* (22), which is a sum of squares. For the estimator by the *fixing method using FAST* (21) there is a sufficient condition, which results from the following proposition:

Proposition 4. *Let f be a 2-dimensional function, and consider its second order interaction $D_{12} = D - D_1 - D_2$. Denote $\widehat{D}_{12} = \widehat{D} - \widehat{D}_1 - \widehat{D}_2$ its FAST estimate, with the notations of section 3.1. Assume that:*

- (i) ω_1 and ω_2 are free of interference up to order $2M$,
- (ii) $N \geq 2M \times \max(\omega_1, \omega_2)$.

Then $\widehat{D}_{12} \geq 0$.

Proof. Denote the sets $W_{\omega_i, M} = \{p\omega_i, p = 1, \dots, M\}$ for $i = 1, 2$, and $W_N = \{1, \dots, N/2\}$. We have

$$\widehat{D}_{12}/2 = \sum_{n \in W_N} (A_n^2 + B_n^2) - \sum_{n \in W_{\omega_1, M}} (A_n^2 + B_n^2) - \sum_{n \in W_{\omega_2, M}} (A_n^2 + B_n^2).$$

Now, the condition (i) ensures that $W_{\omega_1, M} \cap W_{\omega_2, M} = \emptyset$, while (ii) implies that $W_{\omega_i, M} \subseteq W_N$, for $i = 1, 2$. Hence,

$$\widehat{D}_{12}/2 = \sum_{n \in W_N - (W_{\omega_1, M} \cup W_{\omega_2, M})} (A_n^2 + B_n^2) \geq 0.$$

□

Corollary. *It is a direct consequence of proposition 4 that if (i) and (ii) are satisfied, then (21) returns positive values.*

Remark. In practice, one can use for instance $\omega_1 = 11, \omega_2 = 35$ [10], which are free of interferences up to $2M$ for the usual orders $M = 4, 6$; Then the minimal value of N is $2 \times 6 \times \max\{11, 35\} = 420$.

3.3.2. Bias

The three estimation methods differ in terms of bias.

Sobol method (20). The Sobol method estimator is unbiased since only direct Monte Carlo integrals (mean estimators) are used as estimators for the conditional expectations.

Fixing method by Liu and Owen (22). Here too the estimator is unbiased because of the direct Monte Carlo integration. This is especially remarkable in combination with the positivity. Note that the estimator of an inactive total interaction is identically equal to zero (due to the differences in the squared term).

Fixing method using FAST (21). There are several sources of bias for the FAST estimator of main indices given by [12]: Interference, aliasing and truncation. However, (i) and (ii) in proposition 4 are stronger than the conditions given by [12] to limit the bias due to interferences and aliasing. Furthermore, the bias due to truncation vanishes when n_{FAST} tends to infinity. For that reason one can expect (21) to be only slightly biased.

RBD-FAST (18). [12] also mention a bias for RBD-FAST estimators caused by a random noise in the signal coming from the sampled variables. This bias might be even enhanced here through the use of a combination of RBD-FAST estimators.

4. Comparisons

The performance of the estimators shall be studied empirically here. The parameters for each method are chosen in order to match the number of function evaluations N , since this is supposed to be the most time-consuming part, especially for functions with high complexity. We refer to table 1 for the relation between parameter settings and N within each of the four methods. For the *Sobol method* the first factor in the MC integration (19) is evaluated

only once for all index calculations to keep N low. The $+1$ in the formula is due to that fact (where d is due to the first and $\binom{d}{2}$ to the second order indices). When we fix M (e.g. $M = 6$) then for *RBD-FAST* N is determined only by L and for *Sobol method* only by n_{Sobol} . For the *fixing method using FAST* the number of function evaluation depends on the product of n_{MC} and n_{FAST} . Setting $\omega_1 = 11$ and $\omega_2 = 35$, we chose $n_{FAST} = 500$, which satisfies the condition of proposition 4 and seems sufficient to give reliable estimates, so that only n_{MC} has to be adapted. The *fixing method by Liu and Owen* needs four function evaluations according the four terms in the sum for each for each index for each Monte Carlo sample.

method	number of function evaluations
<i>RBD-FAST</i>	$N = 2(Md + L) \times \left(\binom{d}{2} + d\right)$
<i>Sobol method</i>	$N = \left(\binom{d}{2} + d + 1\right) \times n_{Sobol}$
<i>Fixing method using FAST</i>	$N = \binom{d}{2} \times n_{MC} \times n_{FAST}$
<i>Fixing method by Liu and Owen</i>	$N = 4 \times \binom{d}{2} \times n_{LO}$

Table 1: Number of function evaluations for the three estimators (18), (20) and (21).

4.1. Test functions

In order to study the estimators' performances in different situations we consider three functions with different interaction structures. The first function (*function 1*) is defined by

$$g(X_1, X_2, X_3, X_4) = \sin(X_1+X_2)+0.4 \cos(X_3+X_4), \quad X_k \stackrel{\text{i.i.d.}}{\sim} U[-1, 1], \quad k = 1, 2, 3, 4.$$

Its interactions are visibly not higher than second order, a common situation. As a contrast, the extreme case of a pure third order interaction is applied in *function 2*:

$$g(X_1, X_2, X_3) = X_1X_2X_3, \quad X_k \stackrel{\text{i.i.d.}}{\sim} U[-1, 1], \quad k = 1, 2, 3.$$

As a mixed case the popular g -function [13] is chosen. It is defined by

$$g(X_1, \dots, X_d) = \prod_{k=1}^d \frac{|4X_k - 2| + a_k}{1 + a_k}, \quad a_k \geq 0, \quad X_k \stackrel{\text{i.i.d.}}{\sim} U[0, 1], \quad k = 1, \dots, d.$$

We choose $d = 6$ and $\mathbf{a} = (0, 0, 0, 0.4, 0.4, 5)'$ to have a function that contains high interactions. This is demonstrated by the fact that, analytically, the

overall variance with $D = 3.27522$ is much greater than the sum of first and second-order indices with 2.06419 . For the number of function evaluations we choose around 5 000 evaluations per index (in total: $N = \binom{d}{2} \times 5\,000$) and thus set the parameters L , n_{Sobol} , n_{MC} and n_{LO} according to table 1. We estimate each index 100 times for all three methods. Calculations are conducted using the R package `fanovaGraph` (see section Acknowledgements and supplementary material). The results can be seen in figure 1.

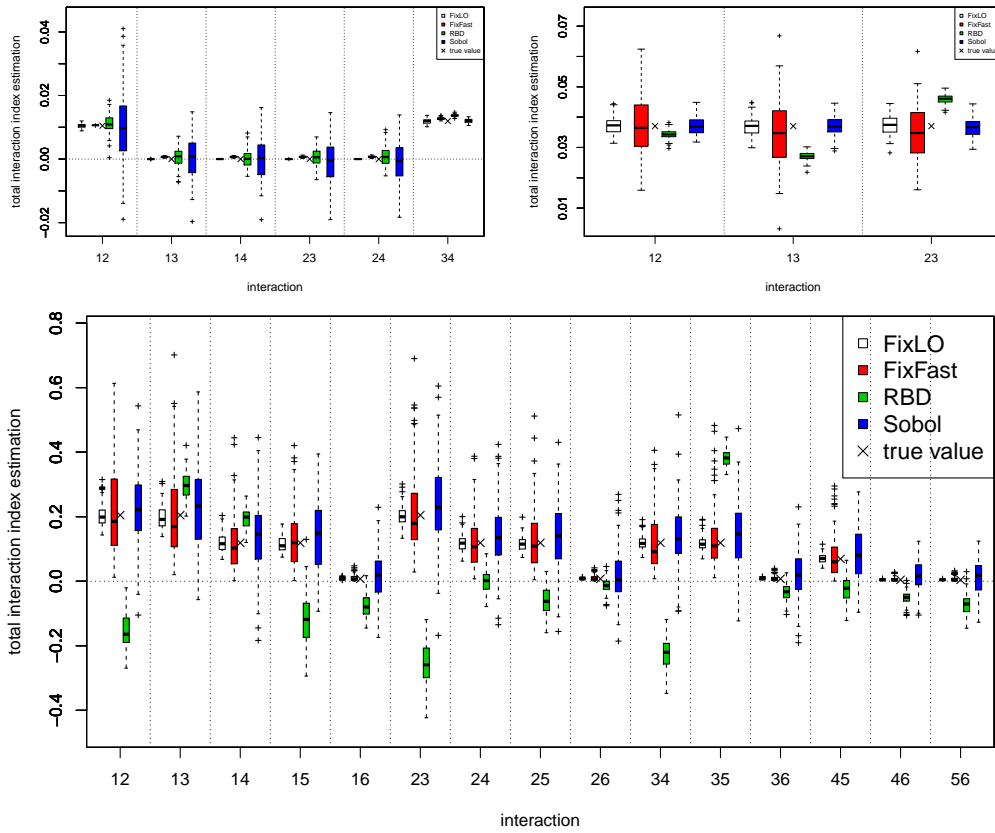


Figure 1: Estimates of the total interaction indices by the four estimators (18), (20), (21) and (22) for test functions 1 (top left), 2 (top right) and g -function (bottom).

4.2. Discussion

As expected in section 3.3.1, negative results can be observed by *RBD-FAST* and the *Sobol method*, but not by the two *fixing methods*. Negative estimates should be treated as zero in applications.

The *RBD-FAST* estimates show a small variance and seem to be unbiased for *function 1*. But with the presence of higher order indices in *function 2* and *g-function*, estimates are severely biased. One reason for this might be the bias for *RBD-FAST* methods described in section 3.3.2.

The estimates by *Sobol method* on the other hand appear to be unbiased, but with a larger variance, resulting from the underlying crude Monte Carlo integration.

The two *fixing methods* both perform well in terms of bias, where the estimates by Liu and Owen seem throughout unbiased and the estimates using *FAST* only slightly biased, as expected in 3.3.2. In terms of variance the Liu and Owen estimates outperform the *FAST* method clearly. The general behaviour of the variance is similar for both estimates. It is very low for the *function 1*, higher for *function 2* and varies for the *g-function*. We observe that the variance is higher when the pair of variables in question $\{X_i, X_j\}$ is part of interactions of a higher order than second order. The reason for this lies in the variance of the estimates $\widehat{D}_{i,j|X_{-\{i,j\}}}^k$ in the fixing method (13). For second order interactions those estimates do not depend on the fixed values $X_{-\{i,j\}}$ and thus vary only slightly, while for higher interactions the estimates should differ with the fixed variables included in the interaction. While this means that the accuracy of the *fixing methods* depends strongly on the interaction situation, they seem to be still the best of the estimation methods. The variances are always smaller than for the *Sobol method* and much less biased than for *RBD-FAST method*. Moreover the *fixing methods*' estimates have the desirable property of having a very low variance for total interaction indices that are close to zero (in the case of Liu and Owen's estimate they are even zero for inactive indices as mentioned in section 3.3.2). That means that the *fixing methods* enable a precise detection of inactive interactions, an important task for interaction screening.

5. Asymptotic properties of Liu and Owen's estimator

The previous section suggests that at least among the four estimation methods, the *fixing method by Liu and Owen* is the most efficient. In fact one can show asymptotical efficiency for this estimator.

Recalling equation (22) we define the estimator for a pair of input variable (X_i, X_j) :

$$T_n = \frac{1}{n} \sum_{k=1}^n \frac{(\Delta_{i,j}^k)^2}{4}$$

with

$$\Delta_{i,j}^k := f(X_i^k, X_j^k, X_{-\{i,j\}}^k) - f(X_i^k, Z_j^k, X_{-\{i,j\}}^k) - f(Z_i^k, X_j^k, X_{-\{i,j\}}^k) + f(Z_i^k, Z_j^k, X_{-\{i,j\}}^k).$$

In the following we assume that \mathbf{X} and \mathbf{Z} are independent random vectors with probability measure ν and that the $(\Delta_{i,j}^k)^2$ are square integrable.

Proposition 5. T_n is consistent for $\mathfrak{D}_{i,j}$

$$T_n \xrightarrow[n \rightarrow \infty]{a.s.} \mathfrak{D}_{i,j}$$

and asymptotically normally distributed

$$\sqrt{n}(T_n - \mathfrak{D}_{i,j}) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}\left(0, \frac{\text{var}[(\Delta_{i,j}^1)^2]}{16}\right)$$

Proof. The results are a direct application of the law of large numbers and the central limit theorem, applied to the variables $(\Delta_{i,j}^k)^2$. \square

Proposition 6. T_n is an asymptotically efficient estimator for $\mathfrak{D}_{i,j}$.

Proof. Denote $\mathcal{X}_k = (X_j^k, Z_j^k, X_{-\{i,j\}}^k)$, $\mathcal{Z}_k = X_i^k$, $\mathcal{Z}'_k = Z_i^k$ and let g be the function defined over $\mathbb{R}^d \times \mathbb{R}$ by:

$$g(\mathbf{a}, b) = f(b, a_1, a_3, \dots, a_d) - f(b, a_2, a_3, \dots, a_d)$$

Then we have

$$\Delta_{i,j}^k = g(\mathcal{X}_k, \mathcal{Z}_k) - g(\mathcal{X}_k, \mathcal{Z}'_k).$$

Therefore

$$T_n = \frac{1}{n} \sum_{k=1}^n \Phi_2(g(\mathcal{X}_k, \mathcal{Z}_k), g(\mathcal{X}_k, \mathcal{Z}'_k))$$

and

$$\mathfrak{D}_{i,j} = \mathbb{E}(\Phi_2(g(\mathcal{X}_1, \mathcal{Z}_1), g(\mathcal{X}_1, \mathcal{Z}'_1)))$$

where Φ_2 is the 2-dimensional function of \mathbb{R}^2 :

$$\Phi_2(u, v) = \frac{(u - v)^2}{4}$$

Remark that \mathcal{Z}_k and \mathcal{Z}'_k are independent copies of each other, both independent of \mathcal{X}_k , and that Φ_2 is a symmetric function. The result then follows from Lemma 2.6 in Janon et al. [14], with the following change of notation

$$i \leftarrow k, X \leftarrow \mathcal{X}, Z \leftarrow \mathcal{Z}, Z' \leftarrow \mathcal{Z}', f \leftarrow g.$$

□

We conclude this section by remarking that the two last propositions extend to the general superset importance, including the case of the total effect of one variable.

Proposition 7. *Let $\Upsilon_I = \sum_{J \supseteq I} D_J$ be the superset importance for a set I . Define*

$$T_{I,n} = \frac{1}{n} \sum_{k=1}^n \frac{(\Delta_I^k)^2}{2^{|I|}}$$

with $\Delta_I^k = \sum_{J \subseteq I} (-1)^{|I-J|} f(Z_J^k, X_{-J}^k)$.

Then $T_{I,n}$ is asymptotically normal and asymptotically efficient for Υ_I .

Proof. Note that $T_{I,n}$ is the sample version of the formula (10) given by Liu and Owen [8] for Υ_I (with suitable change of notations). The proof of asymptotical normality is thus a direct consequence of central limit theorem. For asymptotical efficiency, the proof relies on similar arguments than Proposition 6:

- When $I = \{i\}$ is a single variable, we have:

$$\Delta_I^k = f(Z_i^k, X_{-i}^k) - f(X_i^k, X_{-i}^k),$$

which is of the form $g(\mathcal{X}_k, \mathcal{Z}_k) - g(\mathcal{X}_k, \mathcal{Z}'_k)$ with $\mathcal{Z}_k = Z_i^k$, $\mathcal{Z}'_k = X_i^k$, $\mathcal{X}_k = X_{-i}^k$, and $g(\mathbf{a}, b) = f(b, \mathbf{a})$.

- When $|I| \geq 2$, let choose $i \in I$. Then, by splitting the subsets of I into two parts, depending whether they contain $\{i\}$, we have:

$$\begin{aligned} \Delta_I^k &= \sum_{J \subseteq I - \{i\}} (-1)^{|I-J|} f(Z_J^k, X_{-J}^k) + \sum_{J \subseteq I - \{i\}} (-1)^{|I-(J \cup \{i\})|} f(Z_{J \cup \{i\}}^k, X_{-(J \cup \{i\})}^k) \\ &= \sum_{J \subseteq I - \{i\}} (-1)^{|I-J|} f(Z_J^k, X_i^k, X_{-(J \cup \{i\})}^k) - \sum_{J \subseteq I - \{i\}} (-1)^{|I-J|} f(Z_J^k, Z_i^k, X_{-(J \cup \{i\})}^k) \end{aligned}$$

which is also of the form $g(\mathcal{X}_k, \mathcal{Z}_k) - g(\mathcal{X}_k, \mathcal{Z}'_k)$ with $\mathcal{Z}_k = X_i^k$, $\mathcal{Z}'_k = Z_i^k$, $\mathcal{X}_k = (X_{I-\{i\}}^k, Z_{I-\{i\}}^k, X_{-I}^k)$, and a suitable g , since the second term in the difference is obtained from the first one by exchanging X_i^k and Z_i^k .

The results then derives by applying Lemma 2.6 in Janon et al. [14] to the symmetric function $\Phi_2(u, v) = \frac{(u-v)^2}{2|I|}$, remarking that \mathcal{Z}_k and \mathcal{Z}'_k are independent copies of each other, both independent of \mathcal{X}_k . \square

6. Example of application

In many phenomena, it is not rare, even for complex ones, that some groups of input variables have a separate influence on the output. In that case, the function of interest is decomposed as a sum of lower dimensional terms. In this section, we illustrate how the total interaction indices can be used to recover such decomposition. For instance, let us consider a function, called *function a* from now on:

$$f(X_1, \dots, X_6) = \cos([1, X_1, X_5, X_3] \beta) + \sin([1, X_4, X_2, X_6] \gamma)$$

with $X_k \stackrel{\text{i.i.d.}}{\sim} U[-1, 1]$, $k = 1, \dots, 6$, $\beta = [-0.8, -1.1, 1.1, 1]'$ and $\gamma = [-0.5, 0.9, 1, -1.1]$. Our aim is to recover the decomposition of f into additive parts. First we estimate standard and total indices of the main effects by FAST (section 3.1). The results, divided by the overall variance for comparison purpose, can be seen in table 6. The values for total indices are all very high, so no factor can be removed. The large difference between standard and total indices indicates a strong interaction structure in the function, but the nature of the structure cannot be read from it.

i	1	2	3	4	5	6
\widehat{D}_i/D	0.11639	0.14049	0.09037	0.11290	0.10862	0.19457
\widehat{D}_i^T/D	0.2326	0.20942	0.20770	0.17768	0.23113	0.25874

Table 2: Standard and total indices for the main effects of *function a*

Therefore, in the next step, we estimate total interaction indices. On the basis of the results of section 4 we chose the *fixing method* for estimation. Results, again divided by the overall variance, are given in table 3.

i	1	1	1	1	1	2	2	
j	2	3	4	5	6	3	4	
$\widehat{\mathfrak{D}}_{ij}/D$	0.00000	0.06489	0.00000	0.07757	0.00000	0.00000	0.04130	
i	2	2	3	3	3	4	4	5
j	5	6	4	5	6	5	6	6
$\widehat{\mathfrak{D}}_{ij}/D$	0.00000	0.05076	0.00000	0.06793	0.00000	0.00000	0.04382	0.00000

Table 3: Estimates of total interaction indices for *function a* with $n_{\text{FAST}}=500$ and $n_{\text{MC}} = 50$

A nice way to visualize the estimated interaction structure is the so called FANOVA graph [7]. In the graph, each vertex represents one input factor and an edge between two vertices indicates the presence of second or higher order interactions between the factors. Figure 2 shows the FANOVA graph for *function a*. The thickness of the edges is proportional to the total interaction indices and in addition the thickness of the circles around vertices indicates the standard indices of main effects. The strong connection between factors 1, 3 and 5 is clearly visible as well as the slightly weaker connection between factors 2,4 and 6. We obtain a decomposition into two additive parts.

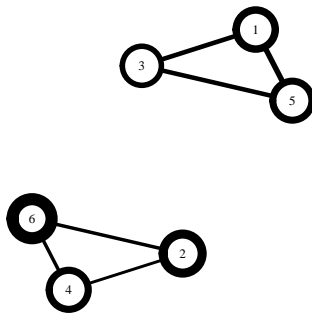


Figure 2: FANOVA-graph for function *a*. The thickness of the circles around the vertices represents main effects, the thickness of the edges represents total interaction effects.

7. Conclusion

We considered the problem of analyzing the interaction structure of a multivariate function, possibly containing high order interactions. For that purpose, we investigated the total interaction index (TII), defined as the superset importance of a pair of variables. The total interaction index generalizes variable screening, usually done with the total index, to interactions.

At theoretical level, we gave several expressions of the TII, including connections to total and closed effects, and an interpretation as an average of the second-order interaction of a 2-dimensional function obtained by fixing the original one (“fixing method”).

Then estimation is considered, and we prove the asymptotical efficiency of Liu and Owen’s estimate. Its superiority is confirmed in empirical comparisons to several other estimation methods. These methods are related to the estimation of closed indices (*Sobol method*), total indices (*RBD-FAST method*), and fixing method where the indices of the fixed 2-dimensional functions are computed by a FAST technique (*fixing-FAST*). The fixing-FAST technique shares the same nice behaviour as Liu and Owen’s for estimating inactive total interactions or second order ones, but its variance is higher. The *RBD-FAST method* could not be trusted, revealing an unpredictable strong bias with some functions, while the *Sobol method* gave unbiased but less accurate (and sometimes negative) results.

Finally we illustrated how the detection of inactive total interactions can be used to recover the decomposition of a complex function by identifying the groups of input variables that have a separate influence on it. Here the indices were also used to graphically visualize the interaction structure of the function.

Among directions for further research, the accuracy of the Liu and Owen estimator could be improved by doing quasi Monte Carlo sampling instead of crude Monte Carlo. Secondly, for real case studies, as interactions may be close to zero but not exactly equal to zero, there is a need to identify a threshold cut below which estimates of total interaction indices are assumed to be close enough to zero. Several techniques, such as tests of significance and decision plots, may be considered.

Supplementary material

The estimation methods as well as FANOVA graphs have been implemented in the R package `fanovaGraph` (version 1.1), published on the official

R website (CRAN). We thank T. Muehlenstaedt for a useful first version of the code, and U. Ligges and O. Mersmann for their relevant advice about programming.

Acknowledgements

We gratefully acknowledge A. Owen for discussions about related works, as well as L. Carraro, M. Roelens, R. Jegou for their comments about the theoretical content.

This paper is based on investigations of the collaborative research centre SFB 708, project C3. The authors would like to thank the Deutsche Forschungsgemeinschaft (DFG) for funding.

Appendix A. Direct proof of proposition 2

The decomposition below, obtained by gathering terms in the FANOVA decomposition of $f(\mathbf{X})$,

$$\begin{aligned} f_{\text{fixed}}(X_i, X_j) &= \sum_{I \subseteq -\{i,j\}} \mu_I(x_I) + \sum_{I \subseteq -\{i,j\}} \mu_{\{i\} \cup I}(X_i, x_I) \\ &+ \sum_{I \subseteq -\{i,j\}} \mu_{\{j\} \cup I}(X_j, x_I) + \sum_{I \subseteq -\{i,j\}} \mu_{\{i,j\} \cup I}(X_i, X_j, x_I) \quad (\text{A.1}) \end{aligned}$$

is the FANOVA decomposition of $f_{\text{fixed}}(X_i, X_j)$. This can be easily shown through conditions (1) and (2). In particular, the second order interaction is given by the last term. Hence, we have by definition:

$$\begin{aligned} D_{i,j|x_{-\{i,j\}}} &= \text{var} \left(\sum_{I \subseteq -\{i,j\}} \mu_{\{i,j\} \cup I}(X_i, X_j, x_I) \right) \\ &= \text{var} \left[\sum_{I \subseteq -\{i,j\}} \mu_{\{i,j\} \cup I}(X_i, X_j, X_I) \middle| X_{-\{i,j\}} = x_{-\{i,j\}} \right] \end{aligned}$$

And, by integrating with respect to $x_{-\{i,j\}}$:

$$\text{E} \left(D_{i,j|X_{-\{i,j\}}} \right) = \text{E} \left(\text{var} \left[\sum_{I \subseteq -\{i,j\}} \mu_{\{i,j\} \cup I}(X_i, X_j, X_I) \middle| X_{-\{i,j\}} \right] \right)$$

Now, one can apply the total variance formula,

$$\begin{aligned} \mathbb{E} \left(D_{i,j|X_{-\{i,j\}}} \right) &= \text{var} \left(\sum_{I \subseteq -\{i,j\}} \mu_{\{i,j\} \cup I}(X_i, X_j, X_I) \right) \\ &\quad - \text{var} \left[\mathbb{E} \left(\sum_{I \subseteq -\{i,j\}} \mu_{\{i,j\} \cup I}(X_i, X_j, X_I) \middle| X_{-\{i,j\}} \right) \right] \end{aligned}$$

and remark that, using (2), $\mathbb{E} \left(\mu_{\{i,j\} \cup I}(X_i, X_j, X_I) \middle| X_{-\{i,j\}} \right) = 0$, for any $i, j \notin I$. Finally:

$$\begin{aligned} \mathbb{E} \left(D_{i,j|X_{-\{i,j\}}} \right) &= \text{var} \left(\sum_{I \subseteq -\{i,j\}} \mu_{\{i,j\} \cup I}(X_i, X_j, X_I) \right) \\ &= \sum_{I \subseteq -\{i,j\}} D_{\{i,j\} \cup I} = \sum_{I \supseteq \{i,j\}} D_I = \mathfrak{D}_{ij}. \end{aligned}$$

□

References

- [1] M. D. Morris, L. M. Moore, M. D. McKay, Sampling plans based on balanced incomplete block designs for evaluating the importance of computer model inputs, *Journal of Statistical Planning and Inference* 136 (2006) 3203–3220.
- [2] A. Saltelli, K. Chan, E. Scott, Sensitivity analysis, Wiley series in probability and statistics, Wiley, Chichester, 2000.
- [3] I. Sobol, Sensitivity estimates for non linear mathematical models, *Mathematical Modelling and Computational Experiments* 1 (1993) 407–414.
- [4] T. Homma, A. Saltelli, Importance measures in global sensitivity analysis of nonlinear models, *Reliability Engineering & System Safety* 52 (1996) 1–17.
- [5] H. Rabitz, O. Alis, J. Shorter, K. Shim, Efficient input–output model representations, *Computer Physics Communications* 117 (1999) 11–20.

- [6] G. Hooker, Discovering additive structure in black box functions, in: Proceedings of KDD 2004, ACM DL, pp. 575–580.
- [7] T. Muehlenstaedt, O. Roustant, L. Carraro, S. Kuhnt, Data-driven Kriging models based on FANOVA-decomposition, To appear in *Statistics & Computing* (2011).
- [8] R. Liu, A. Owen, Estimating mean dimensionality of analysis of variance decompositions, *Journal of the American Statistical Association* 101 (2006) 712–721.
- [9] R. Cukier, H. Levine, K. Shuler, Nonlinear sensitivity analysis of multi-parameter model systems, *Journal of Computational Physics* 26 (1978) 1–42.
- [10] T. A. Mara, Extension of the rbd-fast method to the computation of global sensitivity indices, *Reliability Engineering & System Safety* 94 (2009) 1274–1281.
- [11] B. Efron, C. Stein, The jackknife estimate of variance, *The Annals of Statistics* 9 (1981) 586–596.
- [12] J. Tissot, C. Prieur, A bias correction method for the estimation of sensitivity indices based on random balance designs, 2011. Submitted.
- [13] I. Sobol, Theorems and examples on high dimensional model representation, *Reliability Engineering and System Safety* 79 (2003) 187–193.
- [14] A. Janon, T. Klein, A. Lagnoux-Renaudie, M. Nodet, C. Prieur, Asymptotic normality and efficiency of two sobol index estimators, 2012. HAL: <http://hal.inria.fr/docs/00/66/50/48/PDF/ArtAsymptSobol.pdf>.