Shapley effects for sensitivity analysis with dependent inputs: comparisons with Sobol' indices, numerical estimation and applications

Bertrand Iooss, Clémentine Prieur

Shapley effects for sensitivity analysis with dependent inputs: comparisons with Sobol' indices, numerical estimation and applications

Bertrand Iooss

EDF Lab Chatou, 6 Quai Watier, 78401 Chatou, France

Institut de Mathématiques de Toulouse, 31062, Toulouse, France

Clémentine Prieur

Université Grenoble Alpes, CNRS, LJK, F-38000 Grenoble, France

Inria project/team AIRSEA

Abstract

The global sensitivity analysis of a numerical model aims to quantify, by means of sensitivity indices estimate, the contributions of each uncertain input variable to the model output uncertainty. The so-called Sobol' indices, which are based on the functional variance analysis, present a difficult interpretation in the presence of statistical dependence between inputs. The Shapley effect was recently introduced to overcome this problem as they allocate the mutual contribution (due to correlation and interaction) of a group of inputs to each individual input within the group. In this paper, using several new analytical results, we study the effects of linear correlation between some Gaussian input variables on Shapley effects, and compare these effects to classical first-order and total Sobol' indices. This illustrates the interest, in terms of sensitivity analysis setting and interpretation, of the Shapley effects in the case of dependent inputs. We also investigate the numerical convergence of the estimated Shapley effects. For the practical issue of computationally demanding computer models, we show that the substitution of the original model by a metamodel (here, kriging) makes it possible to estimate these indices with precision at a reasonable computational cost.

Preprint submitted to Reliability Engineering & System Safety *February 22, 2018*

Keywords: Computer experiments, Kriging, Sensitivity Analysis, Shapley effects, Sobol' indices, Uncertainty Quantification

1. Introduction

When constructing and using numerical models simulating physical phenomena, global sensitivity analysis (SA) methods are valuable tools [1]. These methods allow one to determine which model input variables contribute the most to the variability of the model outputs, or on the contrary which are not important and possibly which variables interact with each other. The standard quantitative methods are based on the measure of variance and lead to the so-called Sobol' indices. In the simple framework of d scalar inputs, denoted by $\boldsymbol{X} = (X_1, \dots, X_d)^\top \in \mathbb{R}^d$, and a single scalar output $Y \in \mathbb{R}$, the model response is

$$Y = f(\boldsymbol{X}) \, . \tag{1}$$

In the case of independent inputs, the interpretation of the Sobol' indices is simple because the variance decomposition of Y is unique [2]. For instance, the first-order Sobol' index of X_i, denoted S_i, represents the amount of the output variance solely due to X_i. The second-order Sobol' index (S_{ij}) expresses the contribution of the interactions of the pairs of variables X_i and X_j, and so on for the higher orders. As the sum of all Sobol' indices is equal to one, the indices are interpreted as proportions of explained variance.

However, in many applications, it is common that the input variables have a statistical dependence structure imposed by a probabilistic dependence function [3] (e.g., a copula function) or by physical constraints upon the input or the output space [4, 5, 6]. As shown in previous studies, estimating and interpreting Sobol' indices with correlated inputs are not trivial [7, 8]. Many propositions appear in the literature that are not always easy to interpret. One strategy proposed by [9] is to evaluate the Sobol' indices of subsets of inputs which are correlated within the subsets but not correlated without. However, this approach is not satisfactory because one may need to compute the Sobol' indices of the individual variables.

In [10] the authors proposed to decompose each partial variance V_i contributed by parameter X_i, and defined as $\mathrm{Var}(\mathbb{E}[Y|X_i])$, into partial variance (V_i^{U}) contributed by the uncorrelated variations of parameter X_i and partial variance (V_i^{C}) contributed by the correlated variations of X_i with all other parameters X_j, $j \neq i$. Such an approach allows to exhibit inputs that have an impact on the output only through their strong correlation with other inputs. However, their approach only applies to linear model output with linearly dependent inputs. In the same spirit, the authors in [11] proposed a strategy which support non-linear models and non-linear dependencies. Their methodology is decomposed in two steps: a first step of decorrelation of the inputs and then a second step based on the concept of High Dimensional Model Representation (HDMR). HDMR (see, e.g., [12]) relies on a hierarchy of component functions of increasing dimensions (truncation of ANOVA-Sobol' decomposition in the case of independent variables). The second step in [11] thus consists in performing the HDMR on the decorrelated inputs. At the same time, the authors in [13] proposed a non-parametric procedure to estimate first-order and total indices in the presence of dependencies, not necessarily of linear type. Their methodology requires knowledge of the conditional probability density functions and the ability to draw random samples from those. Later, in [14], the authors established a link between the approaches in [13] and [11], allowing the distinction between the independent contributions of inputs to the response variance and their mutual dependent contributions, via the estimation of four sensitivity indices for each input, namely full and independent first-order indices, and full and independent total indices. They proposed two sampling strategies for dependent continuous inputs. The first one is based on the Rosenblatt transform [15]. The second one is a simpler method that estimates the sensitivity indices without requiring the knowledge of conditional probability density functions, and which can be applied in case the inputs dependence structure is defined by a rank correlation matrix (see, e.g., [16]).

A different approach was introduced in [17]. It is once more based on the HDMR. The component functions are approximated by expansions in terms of some suitable basis functions (e.g., polynomials [18], splines ...). This meta-modeling approach allows a covariance decomposition of the response variance.

It is worth noting that none of these works has given an exact and unambiguous definition of the functional ANOVA for correlated inputs as the one provided by Hoeffding-Sobol' decomposition [19, 2] when inputs are independent. A generalization of the Hoeffding-Sobol' decomposition was proposed in Stone [20] (see also Hooker [21]). Then, in [22], the authors defined a new variable importance measure based on this decomposition. However, such an important measure suffers from two conceptual problems underlined in [23]: the sensitivity indices can be negative and the approach places strong restrictions on the joint probability distribution of the inputs.

As a different approach, [24, 25] initiated the construction of novel generalized moment-free sensitivity indices, called δ-importance measures. Based on some geometrical considerations, these indices measure the shift area between the outcome probability density function and this same density conditioned to a parameter. Thanks to the properties of these new indices, a methodology is available to obtain them analytically through test cases. [6] has shown an application of these sensitivity measures on a gas transmission model with dependent inputs. [26] has also proposed a methodology to evaluate the δ-importance measures by employing the Rosenblatt transform for the decorrelation procedure .

The Shapley values [27], well known in game theory and economics, have been proposed by [28] in the framework of SA of model outputs. In their variance-based form, the Shapley values, also called Shapley effects, have been discussed in [29, 23, 30] to carry out a SA in the case where the model entries exhibit dependencies among them. SA based on Shapley effects does not rely anymore on the Hoeffding-Sobol' decomposition, it rather consists in a direct allocation of a part of the variance of the output at each input. Then, the two main properties and advantages of the Shapley values are that they cannot be negative and they sum-up to the total output variance. The allocation rule is based on an equitable principle. For example, an interaction effect is equally apportioned to each input involved in the interaction. From the conceptual point of view, the major issue is to understand what is the effect of the dependence between inputs on the variance-based Shapley values.

Several test-cases where the variance-based Shapley values can be analytically computed have been presented in [23]. This study has highlighted several properties of these indices (e.g., for $d \geq 2$, if there is a bijection between any two of the X_j, $j = 1, \dots, d$, then those two variables have the same Shapley value). Moreover, [23] gives their general analytical form in the case of Gaussian inputs $\boldsymbol{X}$ and a linear function $f(\boldsymbol{X}) = \beta_0 + \boldsymbol{\beta}^\mathsf{T} \boldsymbol{X}$ (with $\boldsymbol{\beta} \in \mathbb{R}^d$). However, from this analytical result (given in Section 3.1), it is difficult to understand the effects of the dependence between the inputs in the variance-based Shapley values. Therefore, it appears important in this paper to provide a thorough investigation of several particular cases, sufficiently simple to provide some interpretation.

For the sake of practical applications, [29] has proposed two estimation algorithms of the Shapley effects (that we define as the normalized variance-based Shapley values), and illustrated them on two application cases. By using analytical solutions of Shapley effects on several particular test functions, we perform in the present work a numerical convergence study and we compare the theoretical and numerical results of the Shapley effects. As for the Sobol' indices, one important issue in practice is the numerical cost in terms of number of required model evaluations of estimating the Shapley effects. In this case, a classical solution is to use a metamodel which is a mathematical approximation of the numerical model (1) from an initial and limited set of runs [31, 32]. The metamodel solution is a current engineering practice for estimating sensitivity indices [33].

In the following section, we recall the general mathematical formulation of Sobol' indices and Shapley effects when the inputs are dependent. We also provide a discussion of the SA setting [7] that can be addressed with the Shapley effects. In Section 3, we develop the analytical formulas that one can obtain in several particular cases: linear models with Gaussian inputs in various dimensions, block-additive structure, Ishigami model with various dependencies between the three inputs. In particular, we focus on inequalities that can be developed between Sobol' indices and Shapley effects. Numerical algorithms for estimating Shapley effects are studied in Section 4. Section 5 presents an industrial application which requires the use of a metamodel to estimate the sensitivity indices. A conclusion synthesizes the findings and contributions of this work.

2. General formulation of sensitivity indices

2.1. Sobol' indices

Starting from the model (1) $Y = f(\boldsymbol{X})$, the Sobol' index associated with a set of inputs indexed by u ($u \subseteq \{1, \dots, d\}$) is defined by:

$$S_u^{\text{clo}} = \text{Var}(\mathbb{E}[Y|\boldsymbol{X}_u])/\text{Var}(Y) \ . \tag{2}$$

Sobol' indices have been introduced in [2]. Indices defined by (2) are referred as closed Sobol' indices in the literature (see, e.g., [34]) and are always comprised between 0 and 1. For the mathematical developments of the following sections, we denote the numerator of S_u^{clo} as:

$$\underline{\tau}_u^2 = \text{Var}(\mathbb{E}[Y|\boldsymbol{X}_u]) \ . \tag{3}$$

In addition to the Sobol' indices in Eq. (2), total sensitivity indices have also been defined in order to express the "total" sensitivity of the variance of Y to an input variable X_i [35]:

$$S_{T_i} = \frac{\mathbb{E}_{\boldsymbol{X}_{-i}}(\text{Var}_{X_i}[Y|\boldsymbol{X}_{-i}])}{\text{Var}(Y)}, \tag{4}$$

where $\boldsymbol{X}_{-i}$ is the vector $(X_1, \dots, X_d)$ not containing X_i and with the variables over which the conditional operators are applied indicated in subscript.

Recall that this paper is not restricted to independent inputs, so that the knowledge of first-order and total Sobol' indices does not give complete information on the way an input X_i influences the output Y. In [14], the authors propose a strategy based on the estimation of four sensitivity indices per input, namely $S_{(i)}$, $S_{T_{(i)}}$, $S^{\text{ind}}_{(i)}$ and $S^{\text{ind}}_{T_{(i)}}$. $S_{(i)} = S_i$ and $S^{\text{ind}}_{T_{(i)}} = S_{T_i}$ are the classical Sobol' indices, while $S_{T_{(i)}}$ and $S^{\text{ind}}_{(i)}$ are new ones that can be expressed by means of Rosenblatt transform [15]. These indices can also be derived from the law of total variance as discussed in [36].

The indices $S_{(i)}$ and $S_{T_{(i)}}$ include the effects of the dependence of X_i with other inputs, and are referred to as full sensitivity indices in [11]. The indices $S^{\text{ind}}_{(i)}$ and $S^{\text{ind}}_{T_{(i)}}$ measure the effect of an input X_i, that is not due to its dependence with other variables $\boldsymbol{X}_{-i}$. Such indices have also been introduced as uncorrelated effects in [11] and further discussed in [14] which refers to them as the independent Sobol' indices. In [14], the authors propose to estimate the four indices $S_{(i)}$, $S_{T_{(i)}}$, $S^{\text{ind}}_{(i)}$ and $S^{\text{ind}}_{T_{(i)}}$ for the full set of inputs $(i = 1, \ldots, d)$. Note that $S^{\text{ind}}_{(i)} \leq S^{\text{ind}}_{T_{(i)}} = S_{T_i}$ and that $S_i = S_{(i)} \leq S_{T_{(i)}}$, but other inequalities are not known.

In the following, we focus our attention on the full first-order Sobol' indices $S_{(i)}$ and the independent total Sobol' indices $S^{\text{ind}}_{T_{(i)}}$. Indeed, these are the indices which are used in the definition of the Shapley effects (see Section 2.2). Moreover, the full first-order indices $S_{(i)}$ coincide with the associated classical first-order Sobol' indices S_i, and the independent total indices $S^{\text{ind}}_{T_{(i)}}$ coincide with the associated classical total Sobol' indices S_{T_i}. Our preliminary comparative study is completed by the computations of $S_{T_{(i)}}$ and $S^{\text{ind}}_{(i)}$ in [37].

2.2. Shapley effects

As an alternative to the Sobol' indices, the Shapley effects Sh_i are defined as

$$Sh_i = \sum_{u \subseteq -\{i\}} \frac{(d - |u| - 1)!|u|!}{d!} \left[c(u \cup \{i\}) - c(u) \right] \,, \tag{5}$$

where $c(\cdot)$ is a cost function, $-\{i\}$ is the set of indices $\{1, \ldots, d\}$ not containing i and $|u|$ is the cardinality of u. The Shapley values [27], well known in game theory and by economists, have been proposed by [28] in the framework of SA of model outputs. By using, from Eq. (2),

$$c(u) = S_u^{\text{clo}} = \text{Var}(\mathbb{E}[Y|\boldsymbol{X}_u])/\text{Var}(Y) \ , \tag{6}$$

the corresponding Shapley values are new importance measures for SA of model output. These indices are called Shapley effects by the authors in [29] which also prove that it is equivalent to define $c(\cdot)$ as $\mathbb{E}[\text{Var}(Y|\boldsymbol{X}_{-u})]/\text{Var}(Y)$ or as $\text{Var}(\mathbb{E}[Y|\boldsymbol{X}_u])/\text{Var}(Y)$. Note that, in [28] and [29], the cost function is not normalized by the variance of Y while, in the present paper, we consider its normalized version.

The Shapley effect relies on an equitable allocation of part of the variance of the output to each input. Then, it is not any linear combination of Sobol' indices. Indeed, the Shapley effect equitably shares interaction effects of a subset of inputs with each individual input within the subset. Moreover, the Shapley effect associated with input factor X_i ($i \in \{1, \ldots, d\}$) takes into account both interactions and correlations of X_i with X_j, $1 \leq j \leq d$, $j \neq i$. The share allocation has a consequence that Shapley effects are non negative and sum-up to one, allowing an easy interpretation for ranking input factors.

The theoretical formula (5) with $c(u) = S_u^{\text{clo}}$ shows that the Shapley effect of an input is a by-product of its Sobol' indices. Thus, if one can compute the complete set of Sobol' indices, we can compute the Shapley effect of each input. Note that both algorithms proposed in [29] are based on consistent estimators of the Shapley effects. From the exact permutation algorithm, we can extract a consistent estimator of any Sobol' index. Concerning the random permutation algorithm, the sample size N_i related to the inner loop (conditional variance estimation), and the one N_o related to the outer loop (expectation estimation) are fixed to $N_I = 3$ and $N_o = 1$ respectively. Thus it is not possible to extract from that algorithm accurate estimates of Sobol' indices. However, this last algorithm is consistent for the estimation of Shapley effects and is particularly adapted in the case of high-dimensional inputs space (see Section 4.1 for more details).

In case the input factors are independent, first-order (*resp.* total) Sobol' indices provide effectively computable lower- (*resp.* upper-) bounds for the Shapley effects. In the following, as in [29], we will show that these bounds do not hold anymore in case the input factors present some dependencies.

2.3. SA settings

[7] and [38] have defined several objectives, called SA settings, that sensitivity indices can address. These SA settings clarify the objectives of the analysis. They are listed in [38] as follows:

- Factors Prioritization (FP) Setting, to know on which inputs the reduction of uncertainty leads to the largest reduction of the output uncertainty;
- Factors Fixing (FF) Setting, to determine which inputs can be fixed at given values without any loss of information in the model output;
- Variance Cutting (VC) Setting, to know which inputs have to be fixed to obtain a target value on the output variance;
- Factors mapping (FM) Setting, to determine which inputs are most responsible for producing values of the output in a specific region of interest.

In the case of independent inputs, the Sobol' indices directly address the FP, FF and VC Settings (see [38]). In the case of dependent inputs, the classical ANOVA-Sobol' decomposition does not hold anymore and the VC setting cannot be directly obtained with Sobol' indices (see [7]). However, the FP Setting can be achieved using the independent and full first order Sobol' indices. The FF Setting is more difficult to address in the presence of dependencies. Indeed, fixing one or more of the input factors, has an impact on all the input factors that are correlated with them. However, as explained in [14], independent and full total Sobol' indices can help understanding the FF Setting in the dependent framework.

It is of interest now to give some hints about how modelers can use the Shapley effects to address some SA settings. The VC Setting is not achieved, in the independent and dependent inputs cases, because the Shapley effect of an input contains some effects due to other inputs.

In the case of independent inputs, each Shapley effect Sh_i is framed by the corresponding first-order $S_i = S_{(i)}$ and total indices $S_{T_i} = S^{\text{ind}}_{T_{(i)}}$:

$$S_i \leq Sh_i \leq S_{T_i} \ . \tag{7}$$

In addition to the individual effect of the variable X_i, the Shapley effects take into account the effects of interactions by distributing them equally in the index of each input that plays in the interaction [28]. Therefore, the FF Setting is achieved by using the Shapley effects. However, the FP Setting is not precisely achieved because we cannot distinguish the contributions of the main and interaction effects in a Shapley effect.

In the case of dependent inputs, the inequality in Eq. (7) does not hold true anymore. However, due to the equitable principle on which the allocation rule is based, a Shapley effect close to zero means that the input has no significant contribution to the variance of the output, nor by its interactions nor by its dependencies with other inputs. Then, the FF Setting is also achieved.

3. Relations and inequalities between Sobol' indices and Shapley effects

As said before, in the case of dependent inputs, no relation such as the one in Eq. (7) can be directly deduced. The goal of this section is to study particular cases where analytical deduction can be made. The linear model is the first model to be studied in sensitivity analysis because of its simplicity, its interpretability, and the tractable analytical results it provides when the inputs follow Gaussian distributions. Considering two and three inputs allow us to compare our new results to previous studies on Sobol' indices. The Ishigami function is the most used analytical test function in sensitivity analysis. It has been considered in numerous papers dealing with Sobol' indices with dependent inputs.

We focus the analysis on the full first-order indices (corresponding to classical first-order Sobol' indices) $S_i = S_{(i)}$ (called "First-order Sobol" in the figures) and the independent total indices (corresponding to classical total Sobol' indices) $S_{T_i} = S^{\text{ind}}_{T_{(i)}}$ (called "Ind total Sobol" in the figures). Indeed these are the indices mainly studied and discussed in the previous works on this subject, in particular in [13]. The numerical tests of [13] have inspired the ones proposed in this section. Moreover, these indices are easily and directly provided by the Shapley effects estimation algorithms ([29], see Section 4), during the first and last iterations of the algorithm. Finally, the estimation of the independent first-order Sobol' $S_{(i)}$ and of the full total Sobol' indices $S_{T_{(i)}}$ is based on a rather cumbersome process, based on the use of d Rosenblatt transforms (see more details in Section 2.1). Comparisons with these complementary indices are made in [37].

3.1. Gaussian framework and linear model

Let us consider

$$Y = \beta_0 + \boldsymbol{\beta}^\mathsf{T} \boldsymbol{X} \tag{8}$$

with $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{d \times d}$ a positive-definite matrix. We have $\sigma^2 = \text{Var}(Y) = \boldsymbol{\beta}^\mathsf{T}_{\{1,\ldots,d\}} \boldsymbol{\Sigma}_{\{1,\ldots,d\},\{1,\ldots,d\}} \boldsymbol{\beta}_{\{1,\ldots,d\}}$. Note that the subscripts are added on $\boldsymbol{\beta}$ (resp. $\boldsymbol{\Sigma}$) in order to precise which components are contained in the vector (resp. matrix).

We get from [23]:

$$Sh_i = \frac{1}{d} \sum_{u \subseteq -i} \binom{d-1}{|u|}^{-1} \frac{\text{Cov}\big(X_i, \boldsymbol{X}^\mathsf{T}_{-u} \boldsymbol{\beta}_{-u} \mid \boldsymbol{X}_u\big)^2}{\sigma^2 \, \text{Var}(X_i \mid \boldsymbol{X}_u)} \,. \tag{9}$$

Recall now the following classical formula:

$$\text{Var}(\boldsymbol{X} \mid \boldsymbol{X}_{-j}) = \boldsymbol{\Sigma}_{\{1,\ldots,d\},-j} \boldsymbol{\Sigma}^{-1}_{-j,-j} \boldsymbol{\Sigma}_{-j,\{1,\ldots,d\}} \,. \tag{10}$$

From (10) and according to the law of total variance, we easily obtain the Sobol' indices:

$$S_j = \frac{\mathrm{Var}(\mathbb{E}[\beta_0 + \boldsymbol{\beta}^\mathsf{T}\boldsymbol{X}|X_j])}{\sigma^2} = 1 - \frac{\mathbb{E}(\mathrm{Var}[\boldsymbol{\beta}^\mathsf{T}\boldsymbol{X}|X_j])}{\sigma^2}$$
$$= \frac{\boldsymbol{\beta}^\mathsf{T}_{\{1,\dots,d\}}\left(\boldsymbol{\Sigma}_{\{1,\dots,d\},\{1,\dots,d\}} - \boldsymbol{\Sigma}_{\{1,\dots,d\},j}\boldsymbol{\Sigma}^{-1}_{j,j}\boldsymbol{\Sigma}_{j,\{1,\dots,d\}}\right)\boldsymbol{\beta}_{\{1,\dots,d\}}}{\sigma^2}, \quad (11)$$

$$S_{T_j} = \frac{\mathbb{E}(\mathrm{Var}[\beta_0 + \boldsymbol{\beta}^\mathsf{T}\boldsymbol{X}|\boldsymbol{X}_{-j}])}{\sigma^2} = \frac{\mathbb{E}(\mathrm{Var}[\boldsymbol{\beta}^\mathsf{T}\boldsymbol{X}|\boldsymbol{X}_{-j}])}{\sigma^2}$$
$$= \frac{\boldsymbol{\beta}^\mathsf{T}_{\{1,\dots,d\}}\boldsymbol{\Sigma}_{\{1,\dots,d\},-j}\boldsymbol{\Sigma}^{-1}_{-j,-j}\boldsymbol{\Sigma}_{-j,\{1,\dots,d\}}\boldsymbol{\beta}_{\{1,\dots,d\}}}{\sigma^2}. \quad (12)$$

Note that β_0 and $\boldsymbol{\mu}$ do not play any role as translation parameters in variance-based sensitivity analysis. However, no general conclusion can be drawn from this too general example. Therefore, particular cases are studied in the three following sections.

3.2. Gaussian framework, linear model with two inputs

We focus now on the case $d = 2$. We consider $\boldsymbol{X} \sim \mathcal{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $Y = \beta_0 + \boldsymbol{\beta}^\mathsf{T}\boldsymbol{X}$, with

$$\boldsymbol{\mu} = \begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}, \ \boldsymbol{\beta} = \begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix} \text{ and } \boldsymbol{\Sigma} = \begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}, \ -1 \le \rho \le 1, \sigma_1 > 0, \sigma_2 > 0.$$

We have $\sigma^2 = \mathrm{Var}(Y) = \beta_1^2\sigma_1^2 + 2\rho\beta_1\beta_2\sigma_1\sigma_2 + \beta_2^2\sigma_2^2$. From Eq. (3), $\underline{\tau}_u^2 = \mathrm{Var}(\mathbb{E}[Y|\boldsymbol{X}_u])$ ($u \subseteq \{1,\dots,d\}$) and we obtain $\underline{\tau}_\emptyset^2 = 0$, $\underline{\tau}_1^2 = (\beta_1\sigma_1 + \rho\beta_2\sigma_2)^2$, $\underline{\tau}_2^2 = (\beta_2\sigma_2 + \rho\beta_1\sigma_1)^2$ and $\underline{\tau}_{12}^2 = \sigma^2$. For $j = 1, 2$, $d = 2$, the definition of the Shapley effect (Eq. (5)) gives

$$\sigma^2 Sh_j = \frac{1}{d}\sum_{u \subseteq -\{j\}} \binom{d-1}{|u|}^{-1} (\underline{\tau}^2_{u+\{j\}} - \underline{\tau}^2_u). \quad (13)$$

From that, we get

$$\begin{aligned} \sigma^2 Sh_1 &= \beta_1^2\sigma_1^2(1 - \frac{\rho^2}{2}) + \rho\beta_1\beta_2\sigma_1\sigma_2 + \beta_2^2\sigma_2^2\frac{\rho^2}{2}, \\ \sigma^2 Sh_2 &= \beta_2^2\sigma_2^2(1 - \frac{\rho^2}{2}) + \rho\beta_1\beta_2\sigma_1\sigma_2 + \beta_1^2\sigma_1^2\frac{\rho^2}{2}. \end{aligned} \quad (14)$$

For $j = 1, 2$, from Eq. (11) we have

$$\begin{aligned} \sigma^2 S_1 &= \beta_1^2\sigma_1^2 + 2\rho\beta_1\beta_2\sigma_1\sigma_2 + \rho^2\beta_2^2\sigma_2^2 \,, \\ \sigma^2 S_2 &= \beta_2^2\sigma_2^2 + 2\rho\beta_1\beta_2\sigma_1\sigma_2 + \rho^2\beta_1^2\sigma_1^2 \,. \end{aligned} \tag{15}$$

Moreover, for $j = 1, 2$, from the Eq. (12) we have

$$\begin{aligned} \sigma^2 S_{T_1} &= \beta_1^2\sigma_1^2(1-\rho^2) \,, \\ \sigma^2 S_{T_2} &= \beta_2^2\sigma_2^2(1-\rho^2) \,. \end{aligned} \tag{16}$$

From Equations (14), (15) and (16) we can infer that the four following assertions are equivalent

$$\begin{aligned} & Sh_j \leq S_{T_j} \,, \\ & S_j \leq Sh_j \,, \\ & \rho\left(\rho\frac{\beta_1^2\sigma_1^2 + \beta_2^2\sigma_2^2}{2} + \beta_1\beta_2\sigma_1\sigma_2\right) \leq 0 \,, \\ & |\rho| \leq \frac{2|\beta_1\beta_2|\sigma_1\sigma_2}{\beta_1^2\sigma_1^2 + \beta_2^2\sigma_2^2} \,. \end{aligned} \tag{17}$$

The equality of the first three assertions is obtained in the absence of correlation ($\rho = 0$). In this case, the Shapley effects are equal to the first-order and total Sobol' indices in the absence of correlation. In the presence of non-zero correlation, the Shapley effects lie between the full first-order indices and the independent total indices: when $Sh_j \leq S_{T_j}$ then $S_j \leq Sh_j$, and when $Sh_j > S_{T_j}$ then $S_j > Sh_j$. We call this the sandwich effect. We remark that the effects of the dependence that we can see in the independent total indices (*e.g.* $-\rho^2\beta_1^2\sigma_1^2$ for X_1) and in the full first-order indices (*e.g.* $2\rho\beta_1\beta_2\sigma_1\sigma_2 + \rho^2\beta_2^2\sigma_2^2$ for X_1) is allocated in half to the Shapley effect, in addition to the elementary effect (*e.g.* $\beta_1^2\sigma_1^2$ for X_1).

These results are also illustrated in Figure 1. In the left figure, as the standard deviations of each variable are equal, the different sensitivity indices are superimposed and the Shapley effects are constant. In the right figure, X_2 exhibits more variability than X_1, thus the corresponding sensitivity indices of X_2 are logically larger than the ones of X_1. The effect of the dependence between the inputs is clearly shared on each input variable. The dependence between the two inputs lead to a rebalancing of their corresponding Shapley effects, while a full Sobol' index of an input comprises the effect of another input on which it is dependent. We also see on the right figure that the Shapley effects of two perfectly correlated variables are equal. Finally, the sandwich effect is respected for each input: From Eq (17), we can prove that $S_j \leq Sh_j \leq S_{T_j}$ when $\rho \in [-0.8; 0]$ and that $S_j > Sh_j > S_{T_j}$ elsewhere.

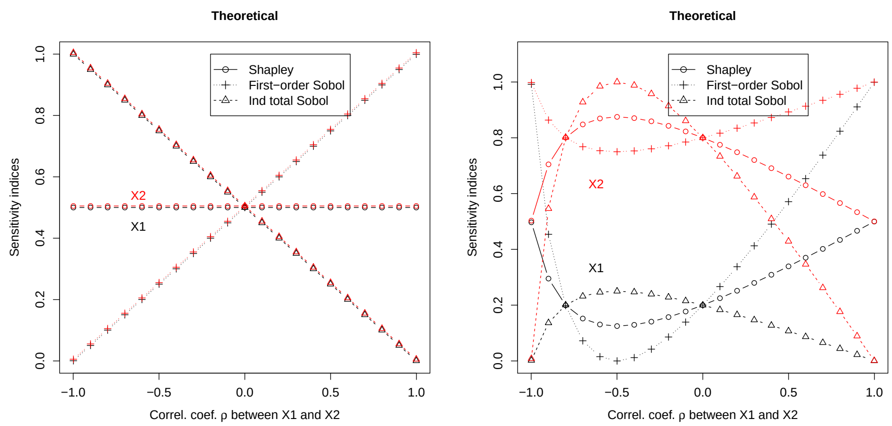

Figure 1: Sensitivity indices on the linear model ($\beta_1 = 1$, $\beta_2 = 1$) with two Gaussian inputs. Left: $(\sigma_1, \sigma_2) = (1, 1)$. Right: $(\sigma_1, \sigma_2) = (1, 2)$.

3.3. Particular case: Effect of a correlated input non included in the model

Consider the model $Y = f(X_1, X_2) = X_1$ with (X_1, X_2) two dependent standard Gaussian variables with a correlation coefficient ρ. It corresponds to the case $\beta_1 = 1$, $\beta_2 = 0$, $\mu_1 = \mu_2 = 0$, $\sigma_1^2 = \sigma_2^2 = 1$ in the model introduced in Section 3.2. We then derive the Shapley effects:

$$Sh_1 = 1 - \frac{\rho^2}{2} \text{ and } Sh_2 = \frac{\rho^2}{2} \, . \tag{18}$$

Eq. (18) leads to the important remark that an input not involved in the numerical model can have a non-zero effect if it is correlated with an influential input of the model. If the two inputs are perfectly correlated, their Shapley effects are equal. This example also illustrates the FF setting that can be achieved with the Shapley effects: if ρ is close to zero, Sh_2 is small and X_2 can be fixed without changing the output variance.

For the Sobol' indices, we have

$$S_1 = 1 \ , \ S_{T_1} = 1 - \rho^2 \text{ and } S_2 = \rho^2 \ , \ S_{T_2} = 0 \ , \tag{19}$$

which indicates that X_2 is only important because of its strong correlation with X_1 (FP setting) and that by only accounting for the uncertainty in X_1, one should be able to evaluate the uncertainty of Y accurately.

3.4. Gaussian framework, linear model with three inputs

We consider a linear model with $\boldsymbol{X} = (X_1, X_2, X_3)^\intercal$ being a Gaussian random vector $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu} = (0,0,0)^\intercal$. We assume that X_1 is independent from both X_2 and X_3, and that X_2 and X_3 may be correlated. The covariance matrix reads:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & \rho\sigma_2\sigma_3 \\ 0 & \rho\sigma_2\sigma_3 & \sigma_3^2 \end{pmatrix} \ , \ -1 \leq \rho \leq 1 \, .$$

We obtained the following analytical results.

$$\begin{aligned} \sigma^2 \quad &= \mathrm{Var}[f(\boldsymbol{X})] = \sum_{j=1}^{3} \beta_j^2 \sigma_j^2 + 2\,\rho\,\beta_2\beta_3\sigma_2\sigma_3 \, , \\ Sh_1 \quad &= (\beta_1^2\sigma_1^2)/\sigma^2 \, , \\ Sh_2 \quad &= [\beta_2^2\sigma_2^2 + \rho\,\beta_2\beta_3\sigma_2\sigma_3 + \frac{\rho^2}{2}(\beta_3^2\sigma_3^2 - \beta_2^2\sigma_2^2)]/\sigma^2 \, , \\ Sh_3 \quad &= [\beta_3^2\sigma_3^2 + \rho\,\beta_2\beta_3\sigma_2\sigma_3 + \frac{\rho^2}{2}(\beta_2^2\sigma_2^2 - \beta_3^2\sigma_3^2)]/\sigma^2 \, , \end{aligned} \tag{20}$$

As expected, we have $\sum_{j=1}^{3} Sh_j = 1$ and we see in Sh_2 and Sh_3 how the correlation effect is distributed in each index. In the case of fully correlated variables, we obtain $Sh_2 = Sh_3 = \left(\beta_2^2\sigma_2^2 + \beta_3^2\sigma_3^2 + 2\rho\,\beta_2\beta_3\sigma_2\sigma_3\right) / \left(2\sigma^2\right)$ with $\rho = \pm 1$.

We study the particular case $\beta_1 = \beta_2 = \beta_3 = 1$, $\sigma_1 = \sigma_2 = 1$ and $\sigma_3 = 2$, for which in [13], the authors provide the formulas of full first-order and independent total Sobol' indices. The analytical indices are depicted on Figure 2 as a function of the correlation coefficient ρ. The Shapley effects are equal to the Sobol' indices in the absence of correlation, and then lie between the associated full first-order and independent total indices in the presence of correlation. The sandwich effect is respected. The effect of an increasing correlation (in absolute value) can be interpreted as an attractive effect both for the full Sobol' indices (here the first-order one) and the Shapley effect. However, for the Shapley effects, the contribution of the correlation is shared with each correlated variable. This leads to the increase of one Shapley effect and the decrease of the other. The Shapley effects allow an easy interpretation. As before, we see that the Shapley effects of two perfectly correlated variables are equal.

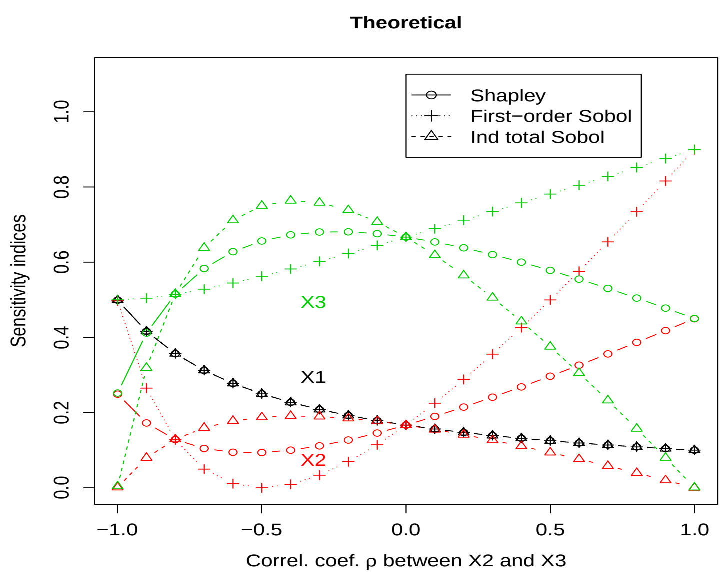

Figure 2: Sensitivity indices on the linear model with three Gaussian inputs.

3.5. Gaussian framework, linear model with an interaction and three inputs

In Sections 3.2 and 3.4, we have presented models for which the Shapley effect takes a value between the associated full first-order and independent total indices. In the present section, we will show that it is not always the case. Let us define the model

$$Y = X_1 + X_2 X_3 \tag{21}$$

with

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim \mathcal{N}_3 \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \boldsymbol{\Sigma} \right) \text{ and } \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & 0 & \rho\sigma_1\sigma_3 \\ 0 & \sigma_2^2 & 0 \\ \rho\sigma_1\sigma_3 & 0 & \sigma_3^2 \end{pmatrix}, \ -1 \leq \rho \leq 1 \,.$$

By simple computations, we get $\sigma^2 S_1 = \sigma_1^2$ and $\sigma^2 S_{T_1} = \sigma^2 - \left(\sigma_2^2\sigma_3^2 + \rho^2\sigma_1^2\right) = (1-\rho^2)\sigma_1^2$. Recall that

$$\sigma^2 Sh_1 = \frac{1}{3}\left(\underline{\tau}_1^2 - \underline{\tau}_\emptyset^2 + \frac{1}{2}\left(\underline{\tau}_{12}^2 - \underline{\tau}_2^2 + \underline{\tau}_{13}^2 - \underline{\tau}_3^2\right) + \sigma^2 - \underline{\tau}_{23}^2 \right) \,.$$

We thus get

$$\sigma^2 Sh_1 = \sigma_1^2(1 - \frac{\rho^2}{2}) + \frac{\sigma_2^2\sigma_3^2}{6}\rho^2 \,. \tag{22}$$

A straightforward computation yields

$$S_{T_1} \leq Sh_1 \leq S_1 \,.$$

We also get $S_2 = 0$, $\sigma^2\, S_{T_2} = \sigma_2^2\sigma_3^2$ and $\sigma^2 Sh_2 = \frac{\sigma_2^2\sigma_3^2}{6}(3+\rho^2)$. Thus

$$S_2 \leq Sh_2 \leq S_{T_2} \,.$$

Concerning the third input variable X_3, one gets $\sigma^2 S_3 = \rho^2\sigma_1^2$, $\sigma^2 S_{T_3} = (1-\rho^2)\sigma_2^2\sigma_3^2$ and $\sigma^2 Sh_3 = \frac{\rho^2\sigma_1^2}{2} + \frac{\sigma_2^2\sigma_3^2}{6}(3-2\rho^2)$. Thus the two following assertions are equivalent:

$$S_3 \leq Sh_3 \leq S_{T_3} \,,$$

$$\rho^2\sigma_1^2 \leq \frac{\sigma_2^2\sigma_3^2}{3}(3-4\rho^2) \,.$$

The two following assertions are also equivalent:

$$S_{T_3} \leq \frac{\phi_3}{\sigma^2} \leq S_3 \,,$$

$$\rho^2\sigma_1^2 \geq \frac{\sigma_2^2\sigma_3^2}{3}(3-2\rho^2) \,.$$

It also happens that Sh_3 is not comprised between S_3 and S_{T_3}. The two following assertions are equivalent:

$$Sh_3 \geq \max\left(S_3, S_{T_3}\right) ,$$

$$\frac{3}{7} \leq \rho^2 \leq \frac{3}{5} .$$

Figure 3 illustrates the previous findings about the Sobol' indices and the Shapley effects for this model. As expected, when the correlation coefficient ρ belongs to two intervals, $[-0.775; -0.655]$ and $[0.655; 0.775]$, the Shapley effects of X_3 are larger than the full first-order Sobol' indices and the independent total Sobol' indices.

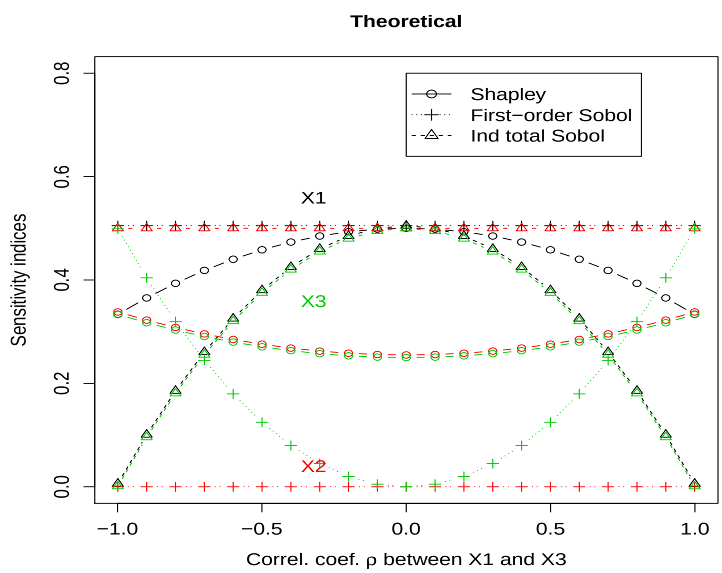

Figure 3: Sensitivity indices on the linear model with three Gaussian inputs and an interaction between X_2 and X_3.

3.6. General model with three inputs and a block-additive structure

We consider the following model:

$$Y = g(X_1, X_2) + h(X_3) , \tag{23}$$

which is called a "block-additive" structure. We consider the general case where the vector $(X_1, X_2, X_3)^\mathsf{T}$ is not restricted to a Gaussian vector. We only assume that the three inputs have finite variances and that X_3 is independent from (X_1, X_2). From the independence properties one has:

$$\sigma^2 = \underline{\tau}_{12}^2 + \underline{\tau}_3^2 , \quad \underline{\tau}_{13}^2 = \underline{\tau}_1^2 + \underline{\tau}_3^2 \text{ and } \underline{\tau}_{23}^2 = \underline{\tau}_2^2 + \underline{\tau}_3^2 . \tag{24}$$

From Equations (2), (4), (5) and (24) we get:

$$S_3 = Sh_3 = S_{T_3} .$$

We also get that, for $j = 1, 2$, the three following assertions are equivalent

$$S_j \leq Sh_j ,$$

$$Sh_j \leq S_{T_j} ,$$

$$\frac{\underline{\tau}_1^2 + \underline{\tau}_2^2}{2} \leq \frac{\underline{\tau}_{12}^2}{2} .$$

We now consider, as in [13], the Ishigami function, a non-linear model involving interaction effects which writes:

$$f(\boldsymbol{X}) = \sin(X_1) + 7\sin(X_2)^2 + 0.1 X_3^4 \sin(X_1) \tag{25}$$

where $X_i \sim \mathcal{U}[-\pi, \pi]$ $\forall i = 1, 2, 3$ with a non-zero correlation ρ between a pair of variables.

Our first study considers a correlation between X_1 and X_3 only. Moreover X_2 is assumed to be independent from X_1 and X_3. This model has a block-additive structure (as in Eq. (23) up to a permutation between X_2 and X_3). The sensitivity measures depicted on Figure 4 were obtained with a numerical procedure explained in the next section. We observe that the sandwich effect is respected for X_1 and X_3. As X_2 is independent from the group (X_1, X_3) and it has no interaction with that group, the Shapley index of X_2 equals both its full first-order and independent total indices. Moreover, the Shapley effects of X_1 and X_3 get closer as the correlation between them increases.

Our second (resp. third) study considers some correlation between X_1 and X_2 (resp. between X_2 and X_3). Let us note that the model structure of Eq.

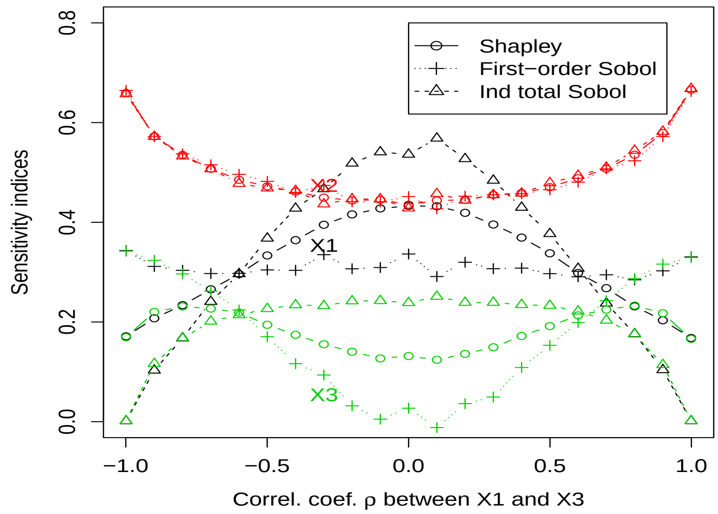

Figure 4: Sensitivity indices on the Ishigami function. Exact permutation method with $N_o = 2 \times 10^4$, $N_i = 3$, $N_v = 10^4$.

(23) with independence between the two blocks is not conserved: The model cannot be decomposed in the sum of two independent terms as in the block-additive model. Figure 5 shows the results of these studies. Surprisingly, the sandwich effect appears still respected. However, it seems difficult to deduce general results on that more general kind of models. In the two studies, we also observe that the correlation effects are smaller on the Shapley effects than in the previous case (correlation between X_1 and X_3), except when the absolute value of the correlation is close to one. It is clear that the strong effects observed in the previous case are due to the conjunction of correlation and interaction between X_1 and X_3.

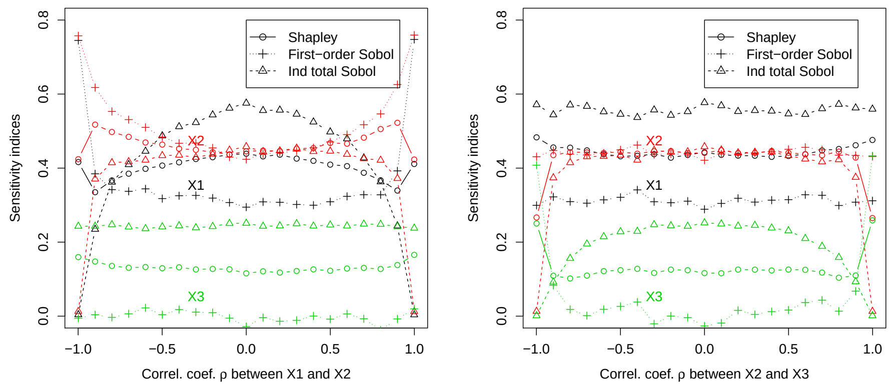

Figure 5: Sensitivity indices on the Ishigami function. Exact permutation method with $N_o = 2 \times 10^4$, $N_i = 3$, $N_v = 10^4$.

4. Numerical estimation issues

4.1. Estimation by direct sampling

For the sake of completeness, the authors in [29] propose two algorithms for estimating the Shapley effects from formula (5) with

$$c(u) = \frac{\mathbb{E}(\mathrm{Var}[Y|\boldsymbol{X}_u])}{\mathrm{Var}(Y)} \tag{26}$$

being the cost function (which has been shown to be more efficient than the variance of the conditional expectation). The first algorithm traverses all possible permutations between the inputs and is called the “Exact permutation method”. The second algorithm consists of randomly sampling some permutations of the inputs and is called the “Random permutation method”. It only makes sense when all permutations is too large to exhaust and then the first algorithm is not applicable.

For each iteration of the inputs’ permutations loop, a conditional variance expectation must be computed. The cost C, in terms of model evaluations, of these algorithms are then the following [29]:

1. Exact permutation method: $C = N_i N_o^{\mathrm{exa}} d!(d-1) + N_v$, with N_i the inner loop size (conditional variance) in (26), N_o^{exa} the outer loop size

(expectation) in (26) and N_v the sample size for the variance computation (denominator in (26));

2. Random permutation method: $C = N_i N_o^{\text{rand}} m(d-1) + N_v$, with m the number of random permutations for discretizing the principal sum in (5), N_i the inner loop size, N_o^{rand} the outer loop size and N_v the sample size for the variance computation.

The $d-1$ term that appears in these costs comes from the fact that $d-1$ Shapley effects are estimated (the last Shapley effect is estimated by using the sum-to-one property). Note that the full first-order Sobol' indices (Eq. (2)) and the independent total Sobol' indices (Eq. (4)) are also estimated by applying these algorithms, each one during only one loop iteration.

From theoretical arguments, the authors in [29] have shown that the near-optimal values of the sizes of the different loops are the following:

- $N_i = 3$ and N_o^{exa} as large as possible for the exact permutation method,
- $N_i = 3$, $N_o^{\text{rand}} = 1$ and m as large as possible for the random permutation method.

We consider these values in all our numerical tests and applications. The value of m which leads to the same numerical cost for the two algorithms is:

$$m = N_o^{\text{exa}} d!\ . \tag{27}$$

In practical applications, this choice is not realistic and the random permutation method is applied with a much smaller value for m (see Section 5). The exact permutation algorithm with fixed N_i is consistent as N_o tends to infinity. The random permutation one with fixed N_i and N_o is consistent as the number of sampled permutations, m, tends to infinity. Indeed, both algorithms are based on unbiased estimators of $\text{Var}(Y)$ and $Sh_i \times \text{Var}(Y)$, whose variance tends to zero (see Appendix A, Equations (18) and (22) in [29] for more details). From Theorem 3 in [29], we know that the variance of the estimator of Sh_i obtained from the random permutation algorithm is bounded by $(\text{Var}(Y))^2/m$. More intuitively, it seems reasonable to think that the difficulty to estimate Sh_i is related on the effective dimension of the model as far as on the complexity of the dependence structure of the inputs.

To illustrate these numerical estimators, we consider the linear model with 3 Gaussian inputs of Section 3.4 with $\beta_1 = \beta_2 = \beta_3 = 1$, $\sigma_1 = \sigma_2 = 1$, $\sigma_3 = 2$. We first set $\rho = 0.9$. On the left-hand side of Figure 6 (resp. right-hand side) are plotted the results of the exact permutation method (resp. random permutation method) versus N_o (resp. m). The error was obtained from the central limit theorem on the permutation loop (Monte Carlo sample of size m) and then by taking two times the standard deviation of the estimates (95% confidence intervals). Similarly, the one in Figure 6 (left) was obtained from the central limit theorem on the outer loop (Monte Carlo sample of size N_o). In the both cases, we observe the convergence of the estimated values toward the exact values as N_o (reps. m) increases. Further analysis about the choice of m, N_o and N_i is nevertheless necessary.

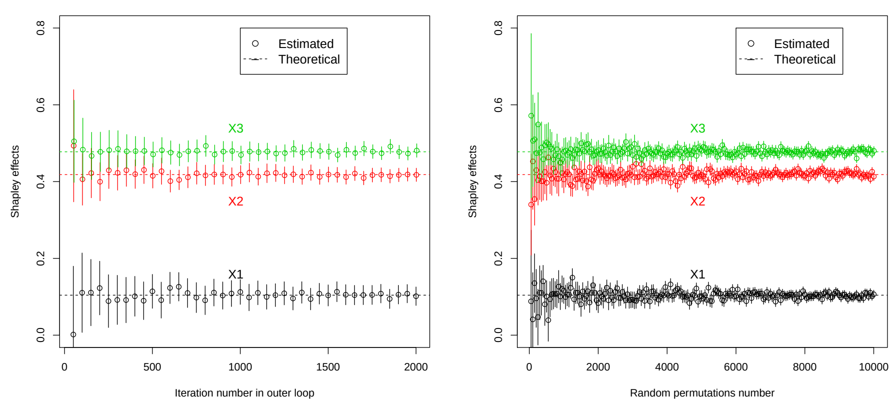

Figure 6: Numerical estimates by the exact permutation method (left) and the random permutation method (right) of the Shapley effects on the linear model.

While varying ρ between -1 and 1, Figure 7 shows the Shapley effects and the Sobol' indices estimated by the two methods (with the same cost by using Eq. (27) with $d = 3$ and $N_v = 10^4$ for both algorithms). We recall that the theoretical results have been obtained in Section 3.4. First, we note that the numerical results and the theoretical values (see Figure 2) are in a good agreement. Second, we can observe that the accuracy of the two methods are equivalent for this low dimensional problem ($d = 3$).

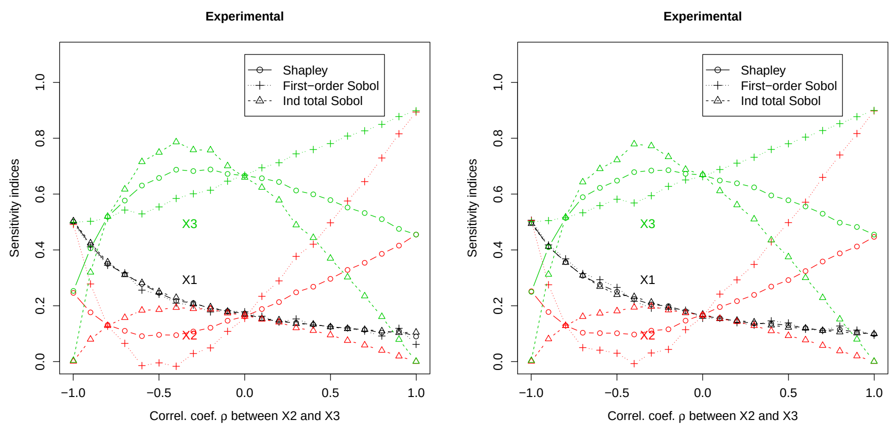

Figure 7: Numerical estimation of sensitivity indices on the linear model. Left: exact permutation method ($N_o = 5 \times 10^3$, $N_i = 3$, $N_v = 10^4$). Right: random permutation method ($m = 3 \times 10^4$, $N_o = 1$, $N_i = 3$, $N_v = 10^4$).

4.2. Metamodel-based estimation

In this section, we consider a relatively common case in industrial applications where the numerical code is expensive in computational time. As a consequence, it cannot be evaluated intensively (e.g. only several hundred calculations are possible). It is therefore not possible to estimate the sensitivity indices with direct use of the model. Indeed, the estimate of Sobol' indices requires for each input several hundreds or thousands of model evaluations [1]). For the Shapley effects, an additional loop is required which increases the computational burden.

In this case, it is recommended to use a metamodel instead of the original numerical model in the estimation procedure. A metamodel is an approximation of the numerical model, built on a learning dataset [31]. The appeal to a metamodel is a current engineering practice for estimating sensitivity indices

[33]. In the present work, we use the Gaussian process metamodel (also called kriging) [39, 40] which has demonstrated in many practical situations to have good predictive capacities (see [41] for example). The Gaussian process model is defined as follows:

$$Y(\boldsymbol{X}) = h(\boldsymbol{X}) + Z(\boldsymbol{X}), \tag{28}$$

where $h(\cdot)$ is a deterministic trend function (typically a multiple linear model) and $Z(\cdot)$ is a centered Gaussian process. The practical implementation details of kriging can be found in [42]. We make the assumption that Z is second-order stationary with a Matérn 5/2 covariance parameterized by the vector of its correlation lengths $\theta \in \mathbb{R}^d$ and variance σ^2. The hyperparameters σ^2 and θ are classically estimated by the maximum likelihood method on a learning sample comprising input/output of a limited number of simulations. Kriging provides an estimator of $Y(\boldsymbol{X})$ which is called the kriging predictor denoted by $\widehat{Y}(\boldsymbol{X})$. To quantify the predictive capability of the metamodel and to validate the predictor, the metamodel predictivity coefficient Q^2 is estimated by cross-validation or on a test sample [41]. More precisely, the Gaussian process model gives the following predictive distribution:

$$\forall \boldsymbol{X}^\star, \quad \left(Y(\boldsymbol{X}^\star) \mid \boldsymbol{y}^N\right) \sim \mathcal{N}\left(\widehat{Y}(\boldsymbol{X}^\star), \sigma_Y^2(\boldsymbol{X}^\star)\right) \tag{29}$$

where $\boldsymbol{X}^\star$ is a point of the input space not contained in the learning sample, $\boldsymbol{y}^N$ is the output vector of the learning sample of size N and $\sigma_Y^2(\boldsymbol{X})$ is the kriging variance that can also be explicitly estimated. In particular, the kriging variance $\sigma_Y^2(\boldsymbol{X})$ quantifies the uncertainty induced by estimating $Y(\boldsymbol{X})$ with $\widehat{Y}(\boldsymbol{X})$.

As an illustration, we study the Ishigami function (Eq. (25)) with a correlation coefficient ρ between X_1 and X_3, on which [13] studied the Sobol' indices (see also Section 3.6 of this paper). When constructing the models, three different sizes N of the learning sample (50, 100 and 200) respectively give three predictive coefficients (Q^2 which is equivalent to the R^2 in prediction) different for the kriging predictor: 0.78, 0.88 and 0.98. Figure 8 shows that with a strong predictive metamodel ($Q^2 = 0.98$ with $N = 200$), the estimations of the Shapley effects by the metamodel are satisfactorily accurate. The precision of the estimated effects deteriorate rapidly with the decrease of the metamodel predictivity.

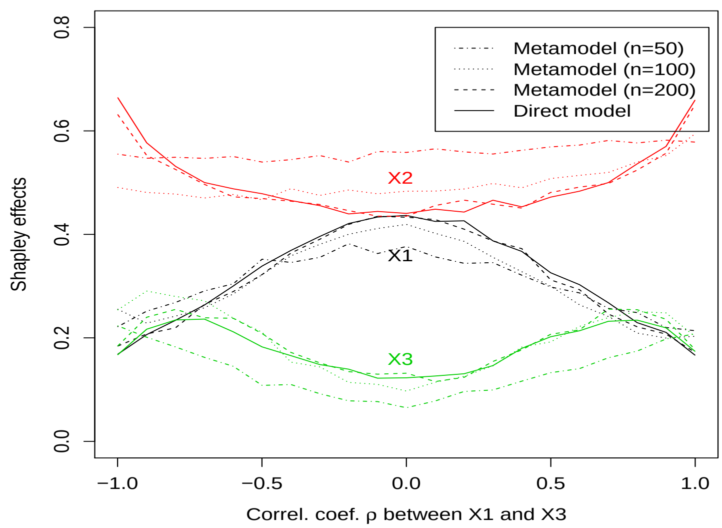

Figure 8: Sensitivity indices on the Ishigami function estimated by 3 metamodels built on 3 different learning databases. The exact permutation method (with $N_o = 2 \times 10^3$, $N_i = 3$, $N_v = 10^4$) is applied on the metamodel predictor.

This example has just illustrated the need to have a sufficiently accurate metamodel in order to have precise estimates of Shapley effects. Controlling the error made on the Shapley effects estimates due to the metamodel approximation is possible thanks to the properties of the Gaussian process model. To provide such confidence bounds on Shapley effects estimates, [37] has developed an algorithm based on conditional Gaussian process simulations (as in [43] for Sobol' indices).

5. Industrial application

This application concerns a probabilistic analysis of an ultrasonic non-destructive control of a weld containing manufacturing defect. Complex phenomena occur in such heterogeneous medium during the ultrasonic wave propagation and a fine analysis to understand the effect of uncertain parameters is important. The simulation of these phenomena is performed via the finite element code ATHENA2D, developed by EDF (Electricité de France). This code is dedicated to the simulation of elastic wave propagation in heterogeneous and anisotropic materials like welds.

A first study [44] has been realized with an inspection configuration aiming to detect a manufactured volumic defect located in a 40 mm thick V grooveweld made of 316L steel (Figure 9). The weld material reveals a heterogeneous and anisotropic structure. It was represented by a simplified model consisting of a partition of 7 equivalent homogeneous regions with a specific grain orientation. Eleven scalar input variables (4 elastic coefficients and 7 orientations of the columnar grains of the weld inspections) have been considered as uncertain and modeled by independent random variables, each one associated to a probability density function. The scalar output variable of the model is the amplitude of the defect echoes resulting from an ultrasonic inspection (maximum value on a so-called Bscan). Uncertainty and sensitivity analysis (based on polynomial chaos expansion [33]) have then been applied from 6000 Monte Carlo simulations of ATHENA2D in [44]. The sensitivity analysis has shown that almost all inputs are influential (only one input has a total Sobol' index smaller than 5%), that the interaction effects are non-negligible (approximately 30% of the total variance) and that the orientations play a major role for explaining the amplitude variability. The analysis confirms that an accurate determination of the micro-structure is essential in these simulation studies. Finally, as a perspective of their work, the authors in [44] explain that the real configuration has been strongly simplified by considering independent input random variables. Indeed, due to the welding physical process, a dependence structure exists between the orientations, in particular between two neighbor domains (see Figure 9 right).

The purpose of the present study is then to perform a sensitivity analysis by using a more realistic probabilistic model for the input random variables. Our SA setting is mainly a FF objective (see Section 2.3): Which parameters can

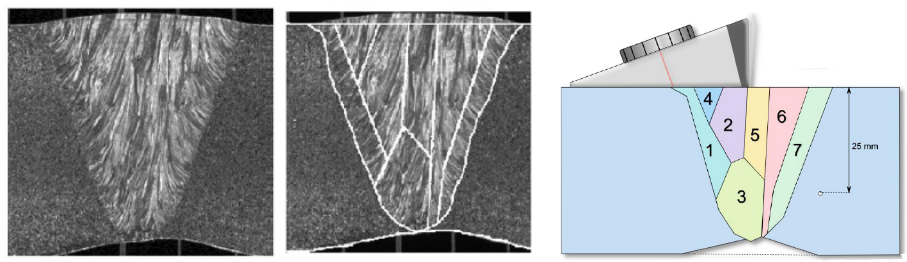

Figure 9: Metallographic picture (left), description of the weld in 7 homogeneous domains (middle) and inspection configuration (right). From [44].

be fixed without impacting the predicted model response uncertainty? Indeed, these SA results are expected to be useful with regards to the qualification process of the non-destructive control technique. As explained in Section 2.3, the Shapley effects are well adapted to FF setting in the case of dependent inputs.

In our study, the probability distributions of all the inputs are considered Gaussian, with the same mean and standard deviation as in [44]. From physical models of welding process and solidification [45], engineers have been able to estimate the following correlation matrix between the 7 orientations $(Or_1, \ldots, Or_7)$ of Figure 9 (right),

$$\mathbf{\Sigma} = \begin{pmatrix} 1 & 0.80 & 0.74 & 0.69 & 0.31 & 0.23 & 0.20 \\ 0.80 & 1 & 0.64 & 0.53 & 0.59 & 0.51 & 0.46 \\ 0.74 & 0.64 & 1 & 0.25 & 0.60 & 0.57 & 0.54 \\ 0.69 & 0.53 & 0.25 & 1 & -0.25 & -0.35 & -0.33 \\ 0.31 & 0.59 & 0.60 & -0.25 & 1 & 0.96 & 0.84 \\ 0.23 & 0.51 & 0.57 & -0.35 & 0.96 & 1 & 0.95 \\ 0.20 & 0.46 & 0.54 & -0.33 & 0.84 & 0.95 & 1 \end{pmatrix}. \quad (30)$$

As only several hundreds of numerical simulations of ATHENA2D can be performed in the schedule time of the present study, our strategy consists of generating a space filling design in order to have a "good" learning sample for a metamodel building process. A Sobol' sequence of $N = 500$ points has then been generated for the $d = 11$ input variables on $[0, 1]^d$. After transformation of this sample to a sample of inputs which follow their physical scales and their joint probability density function, the corresponding 500 runs of ATHENA2D have been computed.

Remark: The 6000 Monte Carlo simulations performed in the previous study [44] were not stored, and thus could not be reused. As already mentioned, the metamodel built in that previous study was based on polynomial chaos expansion, and was not stored as well.

From the resulting N-size learning sample, a Gaussian process metamodel (parameterized as explained in Section 4.2) has then be fitted. We refer to [33] for a comparative study between metamodels based on polynomial chaos expansions and the one based on Gaussian processes. We obtain a predictivity coefficient of $Q^2 = 87\%$. This result is rather satisfactory, especially when it is compared to the predictivity coefficient obtained by a simple linear model ($Q^2 = 25\%$). Moreover, the test on Ishigami function (Section 4.2) has shown that the estimation of Shapley effects with a metamodel of predictivity close to 90% gives results rather close to the exact values.

The Shapley effects are estimated by using the metamodel predictor instead of ATHENA2D (Section 4.2). Due to the input dimension ($d = 11$), the random permutation method is used with $m = 10^4$, $N_i = 3$, $N_o = 1$ and $N_v = 10^4$. The cost is then 3×10^5 in terms of required metamodel evaluations. It would be prohibitive with the "true" computer code ATHENA2D, but it is feasible by using the metamodel predictor. Figure 10 gives the Shapley effects of the elasticity coefficients (C_{11}, C_{13}, C_{55}, C_{13}) and orientations (Or_1, Or_2, Or_3, Or_4, Or_5, Or_6, Or_7). The lengths of the 95%-confidence interval (see Section 4.1) are approximately equal to 4%, which is sufficient to provide a reliable interpretation. Note that negative values of some Shapley effects are caused by the Monte Carlo estimation errors.

By visualizing the Shapley effects, we can propose a discrimination in four groups of inputs according to their degree of influence (note that such discrimination is questionable due to the residual uncertainties on Shapley effects):

- Or_1 and Or_3 whose effects are larger than 20%,
- Or_2 whose effect is 11%,
- C_{11}, Or_4, Or_5, Or_6 and Or_7 whose effects range between 6% and 8%,
- C_{33}, C_{55} and C_{13} whose effects are smaller than 3%. The FF setting could be addressed with the inputs in this group.

To be convinced by this FF setting, the variance of the metamodel output when (C_{33},C_{55},C_{13}) are fixed is compared with the variance of the metamodel output when all the inputs vary. The variance with all inputs is 3.774×10^{-22} and the variance with the three fixed inputs is 3.572×10^{-22}. As expected, the decrease of 5.3% corresponds approximately to the sum of the Shapley effects of C_{33}, C_{55} and C_{13} (approximately 6%).

In the study of [44] which did not take into account the correlation, C_{33} and Or_4 have been identified as influential inputs (effects larger than 9%). This result shows the importance of taking into account the dependence structure between inputs and the usefulness of the Shapley effects for FF setting in this case. If we compare the (normalized) total Sobol' indices of [44] and the Shapley effects of our study, taking into account the correlation has led to:

- an increase in sensitivity indices for Or_1, Or_2 and Or_3,
- a decrease in sensitivity indices for Or_7 ,
- similar sensitivity indices for Or_4, Or_5 and Or_6.

By looking at the input correlation matrix (Eq. (30)), we remark that we can distinguish two groups of inputs as a function of their correlation degrees: (Or_1, Or_2, Or_3, Or_4) and (Or_5, Or_6, Or_7). We observe the homogeneity of the correlation structure effects: the inputs inside the first group correspond to an increase (or a stability) in sensitivity indices and that the inputs inside the second group correspond to a decrease (or a stability) in sensitivity indices.

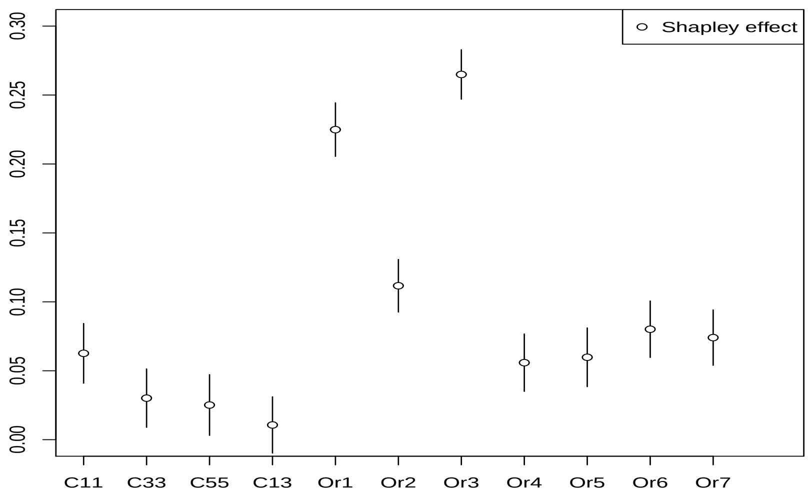

Figure 10: Shapley effects for the ultrasonic non-destructive control application. The vertical bars represent the 95%-confidence intervals of each effect.

6. Conclusion

In many applications of global sensitivity analysis methods, it is common that the input variables have a known statistical dependence structure or that the input space is constrained to a non-rectangular region. In this paper we considered two answers to that issue: the Shapley effects (a normalized version of the variance-based Shapley values proposed in [28] in the framework of sensitivity analysis) and the methodology developed in [14]. The latter suggests the joint analysis of full and independent first-order and total indices to analyze the sensitivity of a model to dependent inputs. In the present paper, we conducted a comparative analysis between Shapley effects on one side and full first-order and independent total indices on the other side. From analytical solutions obtained with linear models and Gaussian variables, we have shown that the dependence between inputs lead to a rebalancing of the corresponding Shapley effects, while a full Sobol' index of an input captures the effect of any input on which it is dependent. Comparisons of Shapley effects with the complementary independent first-order and full total indices are currently under investigation.

We have also illustrated the convergence of two numerical algorithms for estimating Shapley effects. These algorithms depend on various parameters: N_i (conditional variance estimation sample size), N_o (expectation estimation sample size), N_v (output variance estimation sample size) and m (random permutation number). It would be interesting to investigate further the response of the algorithms to these different parameters and to derive empirical and asymptotic confidence intervals for the Shapley effects estimates. Introducing a sequential procedure in the random permutation algorithm, in order to increase m until a sufficient precision on the Shapley effects, seems also promising. Moreover, it would be important in a future work to consider the estimation algorithms capabilities on more complex dependence structures than the pairwise cases exclusively discussed in the present paper.

[46] has started to develop more efficient algorithms in the Gaussian linear case. Finally, we have shown the relevance of using a metamodel (here the Gaussian process predictor) in the industrial situations where the computer model is too time consuming to be evaluated thousands of times for the previous algorithms to be applied. Future work (started in [37]) will consist in developing an algorithm exploiting the complete structure of the Gaussian process allowing to infer the error due to this approximation (see [47], [48] and [43] for the Sobol' indices and [49] for the Derivative-based Global Sensitivity Measures).

Acknowledgments

We are grateful to Thierry Mara and Stefano Tarantola for organizing this SAMO special issue, and to the reviewers for their numerous remarks on the paper. We also thank Géraud Blatman who has performed the computations on ATHENA2D model, Roman Sueur for helpful discussions about Shapley effects and Chu Mai for his remarks and his help for the paper proofreading. Numerical estimation of Shapley effects and Sobol' indices have been realized using the sensitivity package of the R software. Thanks to Eunhye Song, Barry

L. Nelson and Jeremy Staum for providing preliminary versions of the Shapley effects estimation functions.

References

[1] B. Iooss, P. Lemaître, A review on global sensitivity analysis methods, in: C. Meloni, G. Dellino (Eds.), Uncertainty management in Simulation-Optimization of Complex Systems: Algorithms and Applications, Springer, 2015, pp. 101–122.

[2] I. Sobol, Sensitivity estimates for non linear mathematical models, Mathematical Modelling and Computational Experiments 1 (1993) 407–414.

[3] D. Kurowicka, R. Cooke, Uncertainty analysis with high dimensional dependence modelling, Wiley, 2006.

[4] M. Petelet, B. Iooss, O. Asserin, A. Loredo, Latin hypercube sampling with inequality constraints, Advances in Statistical Analysis 94 (2010) 325–339.

[5] S. Kucherenko, O. Klymenko, N. Shah, Sobol' indices for problems defined in non-rectangular domains, Reliability Engineering and System Safety 167 (2017) 218–231.

[6] A. Lopez-Benito, R. Bolado-Lavin, A case study on global sensitivity analysis with dependent inputs: The natural gas transmission model, Reliability Engineering and System Safety 165 (2017) 11–21.

[7] A. Saltelli, S. Tarantola, On the relative importance of input factors in mathematical models: Safety assessment for nuclear waste disposal, Journal of American Statistical Association 97 (2002) 702–709.

[8] S. Da Veiga, F. Wahl, F. Gamboa, Local polynomial estimation for sensitivity analysis on models with correlated inputs, Technometrics 51 (4) (2009) 452–463.

[9] J. Jacques, C. Lavergne, N. Devictor, Sensitivity analysis in presence of model uncertainty and correlated inputs, Reliability Engineering and System Safety 91 (2006) 1126–1134.

[10] C. Xu, G. Gertner, Uncertainty and sensitivity analysis for models with correlated parameters, Reliability Engineering and System Safety 93 (2008) 1563–1573.

[11] T. Mara, S. Tarantola, Variance-based sensitivity indices for models with dependent inputs, Reliability Engineering and System Safety 107 (2012) 115–121.

[12] G. Li, H. Rabitz, J. Hu, Z. Chen, Y. Ju, Regularized random-sampling high dimensional model representation, Journal of Mathematical Chemistry 43 (3) (2008) 6022–6032.

[13] S. Kucherenko, S. Tarantola, P. Annoni, Estimation of global sensitivity indices for models with dependent variables, Computer Physics Communications 183 (2012) 937–946.

[14] T. Mara, S. Tarantola, P. Annoni, Non-parametric methods for global sensitivity analysis of model output with dependent inputs, Environmental Modeling & Software 72 (2015) 173–183.

[15] M. Rosenblatt, Remarks on a multivariate transformation, The Annals of Mathematical Statistics 23 (3) (1952) 470–472.

[16] R. Iman, W. Conover, A distribution-free approach to inducing rank correlation among input variables, Communications in Statistics 11 (3) (1982) 311–334.

[17] G. Li, H. Rabitz, P. Yelvington, O. Oluwole, F. Bacon, C. Kolb, J. Schoendorf, Global sensitivity analysis for systems with independent and/or correlated inputs, Journal of Physical Chemistry 114 (2010) 6022–6032.

[18] B. Sudret, Y. Caniou, Analysis of covariance (ANCOVA) using polynomial chaos expansions, in: G. Deodatis (Ed.), Proceedings of 11th Int. Conf. Struct. Safety and Reliability (ICOSSAR2013), New-York, USA, 2013.

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

[20] C. J. Stone, The use of polynomial splines and their tensor products in multivariate function estimation, The Annals of Statistics 22 (1) (1994) 118–184.

[21] G. Hooker, Generalized functional anova diagnostics for high-dimensional functions of dependent variables, Journal of Computational and Graphical Statistics 16 (2007) 709–732.

[22] G. Chastaing, F. Gamboa, C. Prieur, Generalized Hoeffding-Sobol decomposition for dependent variables - Application to sensitivity analysis, Electronic Journal of Statistics 6 (2012) 2420–2448.

[23] A. Owen, C. Prieur, On Shapley value for measuring importance of dependent inputs, SIAM/ASA Journal on Uncertainty Quantification 5 (2017) 986–1002.

[24] E. Borgonovo, A new uncertainty importance measure, Reliability Engineering and System Safety 92 (2007) 771–784.

[25] E. Borgonovo, W. Castaings, S. Tarantola, Moment independent importance measures: new results and analytical test cases, Risk Analysis 31 (2011) 404–428.

[26] C. Zhou, Z. Lu, L. Zhang, J. Hu, Moment independent sensitivity analysis with correlations, Applied Mathematical Modeling 38 (2014) 4885–4896.

[27] L. Shapley, A value for n-persons game, in: H. Kuhn, A. Tucker (Eds.), Contributions to the theory of games II, Annals of mathematic studies, Princeton University Press, Princeton, NJ, 1953.

[28] A. Owen, Sobol' indices and Shapley value, SIAM/ASA Journal on Uncertainty Quantification 2 (2014) 245–251.

[29] E. Song, B. Nelson, J. Staum, Shapley effects for global sensitivity analysis: Theory and computation, SIAM/ASA Journal on Uncertainty Quantification 4 (2016) 1060–1083.

[30] B. Iooss, C. Prieur, Analyse de sensibilité avec entrées dépendantes : estimation par échantillonnage et par métamodèles des indices de Shapley, in: Proceedings of 49èmes Journées de Statistique, Avignon, France, 2017.

[31] K.-T. Fang, R. Li, A. Sudjianto, Design and modeling for computer experiments, Chapman & Hall/CRC, 2006.

[32] B. Iooss, F. Van Dorpe, N. Devictor, Response surfaces and sensitivity analyses for an environmental model of dose calculations, Reliability Engineering and System Safety 91 (2006) 1241–1251.

[33] L. Le Gratiet, S. Marelli, B. Sudret, Metamodel-based sensitivity analysis: Polynomial chaos expansions and Gaussian processes, in: R. Ghanem, D. Higdon, H. Owhadi (Eds.), Springer Handbook on Uncertainty Quantification, Springer, 2017, pp. 1289–1325.

[34] C. Prieur, S. Tarantola, Variance-based sensitivity analysis: Theory and estimation algorithms, in: R. Ghanem, D. Higdon, H. Owhadi (Eds.), Springer Handbook on Uncertainty Quantification, Springer, 2017, pp. 1217–1239.

[35] T. Homma, A. Saltelli, Importance measures in global sensitivity analysis of non linear models, Reliability Engineering and System Safety 52 (1996) 1–17.

[36] S. Tarantola, T. Mara, Variance-based sensitivity indices of computer models with dependent inputs: The Fourier amplitude sensitivity test, International Journal for Uncertainty Quantification 7 (2018) 511–523.

[37] N. Benoumechiara, K. Elie-Dit-Cosaque, Shapley effects for sensitivity analysis with dependent inputs: bootstrap and kriging-based algorithms, Preprint, arXiv:1801.03300.

[38] A. Saltelli, S. Tarantola, F. Campolongo, M. Ratto, Sensitivity analysis in practice: A guide to assessing scientific models, Wiley, 2004.

[39] J. Sacks, W. Welch, T. Mitchell, H. Wynn, Design and analysis of computer experiments, Statistical Science 4 (1989) 409–435.

[40] T. Santner, B. Williams, W. Notz, The design and analysis of computer experiments, Springer, 2003.

[41] A. Marrel, B. Iooss, F. Van Dorpe, E. Volkova, An efficient methodology for modeling complex computer codes with Gaussian processes, Computational Statistics and Data Analysis 52 (2008) 4731–4744.

[42] O. Roustant, D. Ginsbourger, Y. Deville, DiceKriging, DiceOptim: Two R packages for the analysis of computer experiments by kriging-based metamodeling and optimization, Journal of Statistical Software 21 (2012) 1–55.

[43] L. Le Gratiet, C. Cannamela, B. Iooss, A Bayesian approach for global sensitivity analysis of (multifidelity) computer codes, SIAM/ASA Journal on Uncertainty Quantification 2 (2014) 336–363.

[44] F. Rupin, G. Blatman, S. Lacaze, T. Fouquet, B. Chassignole, Probabilistic approaches to compute uncertainty intervals and sensitivity factors of ultrasonic simulations of a weld inspection, Ultrasonics 54 (2014) 1037–1046.

[45] J. Moysan, A. Apfel, C. Corneloup, B. Chassignole, Modelling the grain orientation of austenitic stainless steel multipass welds to improve ultrasonic assessment of structural integrity, Int. J. Pressure Ves. Piping 80 (2003) 77–85.

[46] B. Broto, F. Bachoc, M. Depecker, J.-M. Martinez, Sensitivity indices for independent groups of variables, Preprint, arXiv:1801.04095.

[47] A. Janon, M. Nodet, C. Prieur, Uncertainties assessment in global sensitivity indices estimation from metamodels, International Journal for Uncertainty Quantification 4 (2014) 21–36.

[48] A. Janon, T. Klein, A. Lagnoux, M. Nodet, C. Prieur, Asymptotic normality and efficiency of two sobol index estimators, ESAIM: Probability and Statistics 18 (2014) 342–364.

[49] M. De Lozzo, A. Marrel, Estimation of the derivative-based global sensitivity measures using a Gaussian process metamodel, SIAM/ASA Journal on Uncertainty Quantification 4 (2016) 708–738.