Generalized Sobol sensitivity indices for dependent variables: numerical methods

Gaelle Chastaing, Clémentine Prieur, Fabrice Gamboa

Generalized Sobol sensitivity indices for dependent variables: numerical methods

Gaelle Chastaing[†] Fabrice Gamboa[‡] Clémentine Prieur[†]

Abstract

The hierarchically orthogonal functional decomposition of any measurable function η of a random vector $\mathbf{X} = (X_1, \cdots, X_p)$ consists in decomposing $\eta(\mathbf{X})$ into a sum of increasing dimension functions depending only on a subvector of $\mathbf{X}$. Even when $X_1, \cdots, X_p$ are assumed to be dependent, this decomposition is unique if components are hierarchically orthogonal. That is, two of the components are orthogonal whenever all the variables involved in one of the summands are a subset of the variables involved in the other. Setting $Y = \eta(\mathbf{X})$, this decomposition leads to the definition of generalized sensitivity indices able to quantify the uncertainty of Y with respect to the dependent inputs $\mathbf{X}$ [9]. In this paper, a numerical method is developed to identify the component functions of the decomposition using the hierarchical orthogonality property. Further, the asymptotic properties of the components estimation is studied, as well as the numerical estimation of generalized sensitivity indices though a toy model. In addition, a model coming from real world illustrates the interest of the method.

Key Words: Dependent variables; Extended basis; Greedy algorithm; LARS; Sensitivity analysis; Sobol decomposition.

[†]Université Joseph Fourier, LJK/MOISE BP 53, 38041 Grenoble Cedex, France, **gaelle.chastaing@imag.fr, clementine.prieur@imag.fr**

[‡]Université Paul Sabatier, IMT-EPS, 118, Route de Narbonne, 31062 Toulouse Cedex 9, France, **fabrice.gamboa@univ-tlse.fr**

1. INTRODUCTION

In a nonlinear regression model, input parameters can be subject to many sources of uncertainty. The objective of global sensitivity analysis is to identify and to rank the input variables that drive the uncertainty of the model output. The most popular methods are the variance-based ones [25]. Among them, the Sobol indices are widely used [26]. This last method stands on the assumption that the incomes are independent. Under this assumption, Hoeffding [13] shows that the model output can be uniquely decomposed into a sum of increasing dimension functions, where the integrals of every summand over any of its own variables must be zero. A consequence of these conditions is that all summands of the decomposition are mutually orthogonal. Using this decomposition, Sobol shows that the global variance can also be decomposed as a sum of partial variances. Thus, the so-called Sobol sensitivity index for a group of inputs is the ratio between the partial variance associated to these inputs and the global variance [26]. However, in models with dependent inputs, the use of Sobol indices may lead to a wrong interpretation because the sensitivity induced by the dependence between two factors is implicitly included in their Sobol indices. To handle this problem, a naive solution can consist in computing Sobol sensitivity indices for independent groups of dependent variables. First introduced by Sobol [26], this idea is exploited in practice by Jacques *et al.* [16]. Nevertheless, this technique implies to work with models having several independent groups of inputs. Furthermore, it does not allow to quantify the individual contribution of each input. A different way to deal with this issue has been initiated by Borgonovo *et al.* [2, 3]. These authors define a new measure based on the joint distribution of $(Y, \mathbf{X})$. The new sensitivity indicator of an input X_i measures the shift between the output distribution and the same distribution conditionally to X_i. This moment free index has many properties and has been applied to some real applications [5, 4]. However, the dependence issue remains unsolved as we do not know how the conditional distribution is distorted by the dependence, and so how it impacts the sensitivity index. Another idea is to use the Gram-Schmidt orthogonalization procedure. In an early

work, Bedford [1] suggests to orthogonalize the conditional expectations and then to use the usual variance decomposition on this new orthogonal collection. He then uses the Monte Carlo simulation to compute the indices. In the same spirit, Mara *et al.* [21] use the same Gram-Schmidt tool to decorrelate the inputs, but perform polynomial regressions to approximate the model. In both papers, the decorrelation method depends on the ordering of the variables, making the procedure computationally expensive and difficult to interpret.

Following the construction of Sobol indices previously exposed, Xu *et al.* [31] propose to decompose the partial variance of an input into a correlated and an uncorrelated contribution in the context of linear models. This last work has been later extended by Li *et al.* with the concept of HDMR [19, 20]. In [19], the authors suggest to reconstruct the model function via classical basis (polynomials, splines,...), then to deduce the decomposition of the response variance as a sum of partial variances and covariances. Instead of classical basis, Caniou *et al.* [8] use a polynomial chaos expansion to approximate the initial model as far as the copula theory to model the dependence structure [22]. Thus, in all these papers, the authors choose a type of model reconstruction before proceeding to the splitting of the response variance.

In a previous paper [9], we revisit the Hoeffding decomposition in a different way, bringing a new definition in the case of dependent inputs. Inspired by the pioneering work of Stone [27] and Hooker [14], we show, under a weak assumption on the inputs distribution, that any model function can be decomposed into a sum of hierarchically orthogonal component functions. This means that two of these summands are orthogonal whenever all variables included in one of the components are also involved in the other. The decomposition leads to generalized Sobol sensitivity indices able to quantify the uncertainty brought by dependent inputs on the model.

The goal of this paper is to complete the work done in [9] by providing an efficient numerical method for the estimation of the generalized Sobol sensitivity indices. In our previous paper [9], we have proposed a statistical procedure based on projection operators to identify the components of the hierarchically orthogonal functional decomposition (HOFD). The method consists in projecting the model output onto

constrained spaces to obtain a functional linear system. The numerical resolution of these systems relies on an iterative scheme that requires to estimate conditional expectations at each step. On one hand, this method is well tailored for independent pairs of dependent variables models. On the other hand, it is difficult to apply to more general models because of its computational cost. Hooker [14] has also worked on the estimation of the HOFD components. This author studies the component estimation via a minimization problem under constraints using a sample grid. In general, this procedure is also quite computational demanding. Moreover, it requires to get a prior on the inputs distribution at each evaluation point, or, at least, to be able to estimate them properly. In a recent article, Li *et al.* [18] come back on Hooker's work and also identify the HOFD components by a least-squares method. They propose to approximate these components using their expansions on suitable basis. They bypass some technical problem of degenerate design matrix by using a continuous descent technique [17].

In this paper, we propose an alternative to directly construct a hierarchical orthogonal basis. Inspired by the usual Gram-Schmidt algorithm, the procedure consists in recursively constructing for each component a multidimensional basis that satisfies the hierarchical orthogonal conditions. This procedure will be referred as the Hierarchical Orthogonal Gram-Schmidt (HOGS) procedure. Then, each component of the decomposition can be properly estimated by a linear combination of this basis. The coefficients are then estimated by the usual least-squares method. Thanks to the HOGS procedure, we show that the design matrix has full rank, so the minimization problem admits a unique and explicit solution. Further, we study the asymptotic properties of the estimated components. Nevertheless, the practical estimation of the one-by-one component suffers from the curse of dimensionality when using the ordinary least-squares estimation. To handle this problem, we propose to estimate parameters of the model using variable selection methods. Two usual algorithms are briefly presented, and are adapted to our method. Further, the HOGS procedure coupled with these algorithms is experimented on numerical examples.

The paper is organized as follows. In Section 2, we give and discuss the general

results on the HOFD. We remind Conditions (C.1) and (C.2) under which the HOFD is available. Further, we give the generalized Sobol sensitivity indices, and discuss their interpretation. Section 3 is devoted to the HOGS procedure. We introduce the appropriate notation, and give a detailed procedure. In Section 4, we adapt the least-squares estimation to our problem, and we point out the curse of dimensionality. Further, we discuss about variables selection via a penalized minimization. Section 5 brings asymptotic results on the component estimators, constructed according to the HOGS procedure of Section 3. In Section 6, we present numerical applications. The first example is a toy function, and its objective is to show the efficiency of the HOGS procedure coupled with variable selection methods in the sensitivity estimation. The last example concerns the pressure applied to a tank. In this example, we want to detect the most influent inputs in the model with our procedure.

2. GENERALIZED SOBOL SENSITIVITY INDICES

Functional ANOVA models are specified by a sum of functions depending on an increasing number of variables. A functional ANOVA model is said to be additive if only main effects are included in the model. It is said to be saturated if all interaction terms are included in the model. However, the existence and the uniqueness of such decomposition is ensured by some identifiability constraints. When the inputs are independent, any regular model function is exactly a saturated ANOVA model with pairwise orthogonal components, as reminded in the introduction. It results that the contribution of any group of variables onto the model is measured by the Sobol index, bounded between 0 and 1. Moreover, the Sobol indices are summed to 1 [26]. The use of such an index is not excluded in the dependence context, but the information due to the dependence is considered several times. This could lead to a wrong interpretation of the Sobol indices. In this section, we remind the main results established in Chastaing *et al.* [9] when inputs can be non-independent. In this case, the saturated ANOVA model is established with weaker identifiability constraints than for the independent case. This leads to a generalization of the

Sobol indices well suited to perform global sensitivity analysis when the inputs are not independent.

First, we remind the general context and notation. The last part is dedicated to the generalization of the Hoeffding-Sobol decomposition when inputs are potentially dependent. The definition of the generalized sensitivity indices follows.

2.1. First Settings

We denote by $\subset$ the strict inclusion, that is $A \subset B \Rightarrow A \cap B \neq B$, whereas we use $\subseteq$ when equality is possible.

Consider a measurable function η of a random vector $\mathbf{X} = (X_1, \cdots, X_p) \in \mathbb{R}^p$, $p \geq 1$, and let Y be the real-valued response variable defined as

$$Y : \begin{array}{ccc} (\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p), P_{\mathbf{X}}) & \rightarrow & (\mathbb{R}, \mathcal{B}(\mathbb{R})) \\ \mathbf{X} & \mapsto & \eta(\mathbf{X}) \end{array}$$

where the joint distribution of $\mathbf{X}$ is denoted by $P_{\mathbf{X}}$. For a σ–finite measure ν on $(\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p))$, we assume that $P_{\mathbf{X}} << \nu$ and that $\mathbf{X}$ admits a density $p_{\mathbf{X}}$ with respect to ν, that is $p_{\mathbf{X}} = \dfrac{dP_{\mathbf{X}}}{d\nu}$.

Also, we assume that $\eta \in L^2_{\mathbb{R}}(\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p), P_{\mathbf{X}})$. As usual, we define the inner product $\langle \cdot, \cdot \rangle$ and the norm $\| \cdot \|$ of the Hilbert space $L^2_{\mathbb{R}}(\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p), P_{\mathbf{X}})$ as

$$\langle h_1, h_2 \rangle = \int h_1(\mathbf{x}) h_2(\mathbf{x}) p_{\mathbf{X}} d\nu(\mathbf{x}) = \mathbb{E}(h_1(\mathbf{X}) h_2(\mathbf{X})), \quad h_1, h_2 \in L^2_{\mathbb{R}}(\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p), P_{\mathbf{X}})$$

$$\|h\|^2 = \langle h, h \rangle = \mathbb{E}(h(\mathbf{X})^2), \quad h \in L^2_{\mathbb{R}}(\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p), P_{\mathbf{X}})$$

Here $\mathbb{E}(\cdot)$ denotes the expectation. Further, $V(\cdot) = \mathbb{E}[(\cdot - \mathbb{E}(\cdot))^2]$ denotes the variance, and $\mathrm{Cov}(\cdot, *) = \mathbb{E}[(\cdot - \mathbb{E}(\cdot))(* - \mathbb{E}(*))]$ the covariance.

Let us denote $[1 : k] := \{1, 2, \cdots, k\}$, $\forall\ k \in \mathbb{N}^*$, and let S be the collection of all subsets of $[1 : p]$. As misuse of notation, we will denote the sets $\{i\}$ by i, and $\{ij\}$ by ij. For $u \in S$ with $u = \{u_1, \cdots, u_t\}$, we set the cardinality of u as $|u| = t$ and the random subvector $\mathbf{X_u} := (X_{u_1}, \cdots, X_{u_t})$. Conventionally, if $u = \emptyset$, $|u| = 0$, and $X_\emptyset = 1$. Also, we denote by $\mathbf{X_{-u}}$ the complementary vector of $\mathbf{X_u}$ (that is, $-u$ is the complementary set of u). The marginal density of $\mathbf{X_u}$ (respectively $\mathbf{X_{-u}}$) is

denoted by $p_{\mathbf{X_u}}$ (resp. $p_{\mathbf{X_{-u}}}$).

Further, the mathematical structure of the functional ANOVA models is defined through subspaces $(H_u)_{u\in S}$ and $(H_u^0)_{u\in S}$ of $L^2_{\mathbb{R}}(\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p), P_{\mathbf{X}})$. $H_\emptyset \equiv H_\emptyset^0$ denotes the space of constant functions. For $u \in S\setminus\{\emptyset\}$, H_u is the space of square-integrable functions that depend only on $\mathbf{X_u}$. The space H_u^0 is defined as:

$$H_u^0 = \left\{h_u \in H_u,\ \langle h_u, h_v\rangle = 0, \forall\ v \subset u, \forall\ h_v \in H_v^0\right\} = H_u \cap \left(\sum_{v\subset u} H_v^0\right)^\perp$$

Through the article, we will see that $\eta(\mathbf{X})$ can be written as a functional ANOVA model in terms of low-order components. We define $d := \max_u |u|$ the order of a functional ANOVA model. Thus, if the ANOVA model is additive, $d = 1$. If it is a saturated model, the order is maximal, and $d = p$.

2.2. Generalized Sobol Sensitivity Indices

Let us suppose that

$$\boxed{\begin{array}{ll} & P_{\mathbf{X}} << \nu \\ \textit{where} & \\ & \nu(dx) = \nu_1(dx_1)\otimes\cdots\otimes\nu_p(dx_p) \end{array}} \quad \text{(C.1)}$$

Our main assumption is :

$$\boxed{\exists\ 0 < M \leq 1,\ \forall\ u \subseteq [1:p],\quad p_{\mathbf{X}} \geq M \cdot p_{\mathbf{X_u}} p_{\mathbf{X_{-u}}} \quad \nu\text{-a.e.}} \quad \text{(C.2)}$$

Under these conditions, the following result states a general decomposition of η as a saturated functional ANOVA model, under the specific conditions of the spaces H_u^0 (defined in Section 2.1),

Theorem 1. *Let η be any function in $L^2_{\mathbb{R}}(\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p), P_{\mathbf{X}})$. Then, under (C.1) and (C.2), there exist functions $\eta_0, \eta_1, \cdots, \eta_{\{1,\cdots,p\}} \in H_\emptyset \times H_1^0 \times \cdots H_{\{1,\cdots,p\}}^0$ such that*

the following equality holds :

$$\begin{aligned}\eta(\mathbf{X}) &= \eta_{\emptyset} + \sum_{i=1}^{p} \eta_i(X_i) + \sum_{1\leq i<j\leq p} \eta_{ij}(X_i, X_j) + \cdots + \eta_{\{1,\cdots,p\}}(\mathbf{X}) \\ &= \sum_{u\in S} \eta_u(\mathbf{X_u}) \end{aligned} \tag{1}$$

Moreover, this decomposition is unique.

To get a better understanding of Theorem 1, the reader could refer to its proof and further explanations in [9]. Notice that, unlike the Sobol decomposition with independent inputs, the component functions of (1) are hierarchically orthogonal, and no more mutually orthogonal. Thus, further along the article, the obtained decomposition (1) will be abbreviated HOFD (for Hierarchically Orthogonal Functional Decomposition). Also, as mentioned in [9], the HOFD is said to be a generalized decomposition because it turns out to be the usual functional ANOVA decomposition when incomes are independent.

The general decomposition of the output $Y = \eta(\mathbf{X})$ given in Theorem 1 allows for decomposing the global variance as a simplified sum of covariance terms. Further below, we define the generalized sensitivity indices able to measure the contribution of any group of inputs in the model when inputs can be dependent :

Definition 1. *The sensitivity index S_u of order $|u|$ measuring the contribution of $\mathbf{X_u}$ into the model is given by :*

$$\boxed{S_u = \frac{V(\eta_u(\mathbf{X_u})) + \sum_{u\cap v\neq u,v} Cov(\eta_u(\mathbf{X_u}), \eta_v(\mathbf{X_v}))}{V(Y)}} \tag{2}$$

More specifically, the first order sensitivity index S_i is given by :

$$\boxed{S_i = \frac{V(\eta_i(X_i)) + \sum_{\substack{v\neq\emptyset \\ i\notin v}} Cov(\eta_i(X_i), \eta_v(\mathbf{X_v}))}{V(Y)}} \tag{3}$$

These indices are called generalized Sobol sensitivity indices because if all inputs are independent, it can be shown that $\mathrm{Cov}(\eta_u, \eta_v) = 0$, $\forall\ u \neq v$ [9].

Proposition 1. *Under (C.1) and (C.2), the sensitivity indices S_u previously defined sums to 1, i.e. $\sum_{u\in S\setminus\{\emptyset\}} S_u = 1$.*

Interpretation of the sensitivity indices

As the covariance term $\mathrm{Cov}(\eta_u, \sum_{u\cap v\neq u,v} \eta_v)$ in S_u can be negative, S_u is no more bounded between 0 and 1 as in the independent case. Hence, the interpretation of our indices here is much less obvious. However, we could interprete a sensitivity index S_u given by (2) as follows. The form of a sensitivity index S_u allows for distinguishing two parts: the first part, $V(\eta_u)/V(Y)$ could be identified as the full contribution of $\mathbf{X_u}$, whereas the second part, $\mathrm{Cov}(\eta_u, \sum_{u\cap v\neq u,v} \eta_v)/V(Y)$ could be interpreted as the contribution induced by the dependence with the other terms of the decomposition. Thus, the covariance terms would play the role of compensation. If $S_u > 0$, the covariance contribution strengthens the variance term if it is positive, whereas it weakens the full contribution if not. In the case of a negative sensitivity index, the contribution induced by the dependence dominates in the index, that would show a small full contribution of the variable itself with respect to the total sum of covariances. As an illustration, we consider a model with $p = 2$ inputs. If the component η_1 is strongly negatively correlated with η_2, the variations of η_1 is going to impact on the variations of η_2. Thus, it may result a negative index. This tells us about the importance of the dependent part here. Nevertheless, the covariance contribution is the same in S_1 and S_2. Thus, if S_1 is greater than S_2, the full contribution of S_1 will be bigger than the one in S_2. In this case, we are able to rank the input parameters.

However, these results are not new, as they are developed in [9]. As a continuity of this article, we propose here to estimate the functional ANOVA components in the second part. In Theorem 1, we show that each component η_u belongs to a subspace H_u^0 of $L^2(\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p), P_{\mathbf{X}})$. Thus, to estimate η_u, the most natural approach is to construct a good approximation space of H_u^0. The next section aims at proposing a procedure to construct such a space.

3. THE HIERARCHICAL ORTHOGONAL GRAM-SCHMIDT PROCEDURE

In Section 2, we have seen that the generalized sensitivity indices are defined for any type of reference measures $(\nu_i)_{i\in[1:p]}$. From now and until the end, we will assume that ν_i, $\forall\ i \in [1:p]$, are diffuse measures. Indeed, the non diffuse measures raise additional issues in the results developed further that we will not address in this paper.

In a Hilbert space, it is usual to call in an orthonormal basis to express any of the space element as a linear combination of these basis. Further below, we will define the finite-dimensional spaces $H_u^L \subset H_u$ and $H_u^{0,L} \subset H_u^0$, $\forall\ u \in S$, as linear spans of some orthonormal systems that will be settled later. We take the notation $\text{Span}\,\{B\}$ to define the set of all finite linear combination of elements of B, also called the linear span of B.

Consider, for any $i \in [1:p]$, a truncated orthonormal system $(\psi_{l_i}^i)_{l_i=1}^{L_i}$ of $L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_{X_i})$, with $L_i \geq 1$. Further, let us denote the vector of sizes as $L = {}^t(L_1, \cdots, L_p)$. Also, we denote by $\boldsymbol{l_u} = (l_u^1, \cdots, l_u^i, \cdots)_{i\in u}$ the multi-index associated to the tensor-product of $(\otimes_{i\in u}\psi_{l_u^i}^i)$. Hence, $\boldsymbol{l_u} \in \underset{i\in u}{\times}[1:L_i]$, where $\underset{i\in u}{\times}[1:L_i]$ denotes the Cartesian product of the set $[1:L_i]$, for $i \in u$. To define properly the truncated spaces $H_u^L \subset H_u$, we need first to assume that

$$\boxed{\forall\ u \in S,\ \forall\ l_i \in [1:L_i], \quad \int (\textstyle\prod_{i\in u}\psi_{l_i}^i(x_i))^2 p_{\mathbf{X}} d\nu(\mathbf{x}) < +\infty} \qquad \text{(C.3)}$$

Remark. A sufficient condition for (C.2) is to have $0 < M_1 \leq p_{\mathbf{X}} \leq M_2$ *(see Section 3 of [9]). In this particular case, it is enough to assume that* $\int (\prod_{i\in[1:p]}\psi_{l_i}^i(x_i))^2 d\nu(\mathbf{x}) < +\infty$ *to warrant (C.3).*

Under (C.3), we define, $H_\emptyset^L = \text{Span}\,\{1\}$. Also, we set, $\forall\ i \neq j \in [1:p]$,

$$H_i^L = \text{Span}\left\{1, \psi_1^i, \cdots, \psi_{L_i}^i\right\}$$

$$H_{ij}^L = \text{Span}\left\{1, \psi_1^i, \cdots, \psi_{L_i}^i, \psi_1^j, \cdots, \psi_{L_j}^j, \psi_1^i \otimes \psi_1^j, \cdots, \psi_{L_i}^i \otimes \psi_{L_j}^j\right\}$$

where $\left(\psi^i_{l^i_{ij\}}} \otimes \psi^j_{l^j_{\{ij\}}}\right)$, for $\boldsymbol{l_{\{ij\}}} = (l^i_{\{ij\}}, l^j_{\{ij\}}) \in [1 : L_i] \times [1 : L_j]$ is the tensor product orthonormal system of $H^L_i \otimes H^L_j$. More generally, for any u such that $|u| \geq 1$, we set

$$H^L_u = \text{Span}\left\{1, (\otimes_{i\in v}\psi^i_{l^i_v})_{\boldsymbol{l_v}=(l^i_v)\in \underset{i\in v}{\times}[1:L_i]}, \forall\ v \subseteq u, v \neq \emptyset\right\}$$

Notice that $\dim(H^L_u) = 1 + \sum_{\substack{v\subseteq u\\ v\neq\emptyset}} \prod_{i\in v} L_i$. Now, we define the corresponding spaces up to the hierarchical orthogonality constraints. First, $H^{0,L}_\emptyset = H^L_\emptyset$. For $u \in S \setminus \{\emptyset\}$,

$$H^{0,L}_u = \left\{h_u \in H^L_u,\ \langle h_u, h_v\rangle = 0,\ \forall\ v \subset u,\ \forall\ h_v \in H^{0,L}_v\right\}$$

From now, we denote by L_u the dimension of $H^{0,L}_u$, for $u \in S \setminus \{\emptyset\}$. By definition of $H^{0,L}_u$, we get $L_u = \dim(H^L_u) - [\sum_{\substack{v\subset u\\ v\neq\emptyset}} \prod_{i\in v} L_i + 1] = \prod_{i\in u} L_i$.

Suppose that we observe an independent and identically distributed sample $(y^s, \mathbf{x}^s)_{s=1,\cdots,n}$ of size n from the distribution of $(Y, \mathbf{X})$. We define the *empirical* inner product $\langle\cdot,\cdot\rangle_{\text{n}}$ and norm $\|\cdot\|_{\text{n}}$ as

$$\langle g_1, g_2\rangle_{\text{n}} = \frac{1}{n}\sum_{s=1}^{n} g_1(\mathbf{x}^s)g_2(\mathbf{x}^s), \quad \|g\|^2_{\text{n}} = \langle g, g\rangle_{\text{n}}$$

We define the finite-dimensional linear subspaces $(G^{0,L}_{u,\text{n}})_{u\in S}$ as the approximating spaces of $(H^{0,L}_u)_{u\in S}$, when the scalar product $\langle\cdot,\cdot\rangle$ is replaced by the empirical one. First, $G^{0,L}_{\emptyset,\text{n}} = H^L_\emptyset$, and, for $u \in S \setminus \{\emptyset\}$,

$$G^{0,L}_{u,\text{n}} = \left\{g_u \in H^L_u,\ \langle g_u, g_v\rangle_{\text{n}} = 0, \forall\ v \subset u, \forall\ g_v \in G^{0,L}_{v,\text{n}}\right\}$$

Now we have determined $(H^{0,L}_u)_{u\in S}$ and $(G^{0,L}_{u,\text{n}})_{u\in S}$, we want to build them. In the next procedure, we propose an iterative scheme to construct them, taking into account their specific properties of orthogonality.

HOGS Procedure

1. *Initialization: For any $i \in [1 : p]$, take a truncated orthonormal system $(\psi^i_{l_i})_{l_i=0}^{L_i}$, $L_i \geq 1$, of $L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_{X_i})$ such that $\psi^i_0 = 1$. Set $\phi^i_{l_i} = \psi^i_{l_i}$, $\forall\ l_i \geq 1$ and $H^{0,L}_i = Span\left\{\phi^i_1, \cdots, \phi^i_{L_i}\right\}$. As $(\phi^i_{l_i})_{l_i=1}^{L_i}$ is an orthonormal system, $h(X_i) = \sum_{l_i=1}^{L_i} \beta^i_{l_i}\phi^i_{l_i}(X_i)$ satisfies $\mathbb{E}(h_i(X_i)) = 0$.*

2. *To build a basis* $(\phi^u_{\boldsymbol{l_u}})_{\boldsymbol{l_u}\in \underset{i\in u}{\times}[1:L_i]}$ *of* $H_u^{0,L}$*, with* $|u|=k$*, we proceed recursively on* $|u|$*. Suppose that, for any* $v\in S$ *such that* $1\le |v|\le k-1$*, we get*

$$H_v^{0,L}=Span\left\{\phi^v_{\boldsymbol{l_v}},\boldsymbol{l_v}=(l_v^i)\in \underset{i\in v}{\times}[1:L_i]\right\},\quad L_v:=dim(H_v^{0,L})=\prod_{i\in v}L_i$$

To construct $(\phi^u_{\boldsymbol{l_u}})_{\boldsymbol{l_u}\in \underset{i\in u}{\times}[1:L_i]}$*, with* $u:=\{u_1,\cdots,u_k\}$*, we proceed as follows: for all* $(l_u^{u_1},\cdots,l_u^{u_k})\in[1:L_{u_1}]\times\cdots\times[1:L_{u_k}]$*,*

(a) set

$$\phi^u_{\boldsymbol{l_u}}(\mathbf{X_u})=(\phi^{u_1}_{l_u^{u_1}}\otimes\cdots\otimes\phi^{u_k}_{l_u^{u_k}})(\mathbf{X_u})+\sum_{\substack{v\subset u\\ v\neq\emptyset}}\sum_{\boldsymbol{l_v}\in\underset{i\in v}{\times}[1:L_i]}\lambda^v_{\boldsymbol{l_v},\boldsymbol{l_u}}\phi^v_{\boldsymbol{l_v}}(\mathbf{X_v})+C^u_{\boldsymbol{l_u}}\quad(4)$$

(b) compute the $(1+\sum_{\substack{v\subset u\\ v\neq\emptyset}}L_v)$ *coefficients* $(C,(\lambda^v_{\boldsymbol{l_v},\boldsymbol{l_u}})_{\boldsymbol{l_v}\in\underset{i\in v}{\times}[1:L_i],v\subset u})$ *by solving*

$$\begin{cases}\langle\phi^u_{\boldsymbol{l_u}},\phi^v_{\boldsymbol{l_v}}\rangle=0,\quad\forall v\subset u,\ \forall\ \boldsymbol{l_v}\in\underset{i\in v}{\times}[1:L_i]\\ \langle\phi^u_{\boldsymbol{l_u}},1\rangle=0\end{cases}\quad(5)$$

With (4), the linear system (5) is equivalent to a sparse matrix system of the form $A^u_\phi\lambda^u=D^{\boldsymbol{l_u}}$*, when* $C^u_{\boldsymbol{l_u}}$ *has been removed. The matrix* A^u_ϕ *is a Gramian matrix involving terms* $\mathbb{E}(\Phi_{v_1}(\mathbf{X_{v_1}})^t\Phi_{v_2}(\mathbf{X_{v_2}}))_{v_1,v_2\subset u}$*, with* $(\Phi_{v_i}(\mathbf{X_{v_i}}))_{\boldsymbol{l_{v_i}}}=\phi^{v_i}_{\boldsymbol{l_{v_i}}}(\mathbf{X_{v_i}}),\boldsymbol{l_{v_i}}\in\underset{j\in v_i}{\times}[1:L_j])$*,* $i=1,2$*.* λ^u *involves the terms* $(\lambda^v_{\boldsymbol{l_v},\boldsymbol{l_u}})_{\boldsymbol{l_v}\in\underset{i\in v}{\times}[1:L_i],v\subset u}$*, and* $D^{\boldsymbol{l_u}}$ *involves* $-\mathbb{E}(\otimes_{i=1}^k\phi^{u_i}_{l_u^i}\Phi_{v_i})_{v_i\subset u}$*. In Lemma 1, we show that* A^u_ϕ *is a definite positive matrix, so the system (5) admits a unique solution.*

Set $H_u^{0,L}=Span\left\{\phi^u_{\boldsymbol{l_u}},\ \boldsymbol{l_u}\in\underset{i\in u}{\times}[1:L_i]\right\}$*, with* $L_u:=dim(H_u^{0,L})=\prod_{i=1}^k L_{u_i}$*.*

The construction of $(G^{0,L}_{u,\mathrm{n}})_{u\in S}$ is very similar to the $(H_u^{0,L})_{u\in S}$ one. However, as the spaces $(G^{0,L}_{u,\mathrm{n}})_{u\in S}$ depend on the observed sample, their construction requires to assume that the sample size n is larger than the sizes L_i, $i\in[1:p]$. To build $G^{0,L}_{i,\mathrm{n}}$, $\forall\ i\in[1:p]$, we use the usual Gram-Schmidt procedure on $(\phi^i_{l_i})_{l_i=1}^{L_i}$ to get an orthornormal system $(\varphi^i_{l_i})_{l_i=1}^{L_i}$ with respect to the empirical inner product $\langle\cdot,\cdot\rangle_{\mathrm{n}}$. To build $(G^{0,L}_{u,\mathrm{n}})_{u\in S,|u|\ge2}$, it is enough to replace $\langle\cdot,\cdot\rangle$ by $\langle\cdot,\cdot\rangle_{\mathrm{n}}$ and to use the HOGS

procedure. At the end, we denote $G_{u,\mathrm{n}}^{0,L} = \mathrm{Span}\left\{\varphi_{\boldsymbol{l_u},\mathrm{n}}^{u},\ \boldsymbol{l_u} \in \underset{i\in u}{\times}[1:L_i]\right\},\ \forall\ u \in S\setminus\{\emptyset\}$.

In practice, polynomials or splines basis functions [10] will be considered. Also, for practical reasons and motivation exposed in Section 5, we will approximate the model output by an ANOVA model of order at most $d = 3$. In the next section, we discuss the practical estimation of generalized Sobol sensitivity indices using least-squares minimization. Further, we also discuss the curse of dimensionality, and propose some variable selection methods to handle it.

4. ESTIMATION OF THE GENERALIZED SENSITIVITY INDICES

4.1. Least-Squares Estimation

The effects $(\eta_u)_{u\in S}$ in the HOFD (1) satisfy

$$(\eta_u)_{u\in S} = \underset{\substack{(\tilde{\eta}_u)_{u\in S}\\ \tilde{\eta}_u\in H_u^0}}{\mathrm{Arg\,min}}\ \mathbb{E}[(Y - \sum_{u\in S}\tilde{\eta}_u(\mathbf{X_u}))^2] \tag{6}$$

Notice that η_0, the expected value of Y, has not interest for the sensitivity indices estimation. Thus, $\tilde{Y} := Y - \mathbb{E}(Y)$ replaces Y in (6). Also, the residual term $\eta_{\{1,\cdots,p\}}$ is removed from (6) and it is estimated afterwards. In Section 3, we defined the approximating spaces $G_{u,\mathrm{n}}^{0,L}$ of H_u^0, for $u \in S\setminus\{\emptyset\}$. Thus, the minimization problem (6) may be replaced by its empirical version,

$$\min_{(\beta_{\boldsymbol{l_u}})_{\boldsymbol{l_u},u}} \frac{1}{n}\sum_{s=1}^{n}\left[\tilde{y}^s - \sum_{\substack{u\subset[1:p]\\ u\neq\emptyset}}\ \sum_{\boldsymbol{l_u}\in\underset{i\in u}{\times}[1:L_i]} \beta_{\boldsymbol{l_u}}^{u}\varphi_{\boldsymbol{l_u},\mathrm{n}}^{u}(\mathbf{x_u}^s)\right]^2 \tag{7}$$

where $\tilde{y}^s := y^s - \bar{y}$, $\bar{y} := \frac{1}{n}\sum_{s=1}^{n} y^s$, and where every subspace $G_{u,\mathrm{n}}^{0,L}$ is spanned by basis functions $(\varphi_{\boldsymbol{l_u},\mathrm{n}}^{u})_{\boldsymbol{l_u}\in\underset{i\in u}{\times}[1:L_i]}$ constructed according to the HOGS Procedure of Section 3. The equivalent matrix form of (7) is

$$\min_{\boldsymbol{\beta}} \|\mathbb{Y} - \mathbb{X}_{\varphi}\boldsymbol{\beta}\|_{\mathrm{n}}^2 \tag{8}$$

where $\mathbb{Y}_s = y^s - \bar{y}$, $\mathbb{X}_\varphi = \begin{pmatrix} \boldsymbol{\varphi}_1 & \cdots & \boldsymbol{\varphi}_u & \cdots \end{pmatrix} \in \underset{u\in S}{\times} \mathcal{M}_{n,L_u}(\mathbb{R})$, where, $\underset{u\in S}{\times} \mathcal{M}_{n,L_u}(\mathbb{R})$ denotes the cartesian product of real entries matrices with n rows and L_u columns. For $u \in S$, $(\boldsymbol{\varphi}_u)_{s,\boldsymbol{l_u}} = \varphi^u_{\boldsymbol{l_u},\mathrm{n}}(\mathbf{x_u}^s)$, $\forall\ s \in [1:n]$, $\forall\ \boldsymbol{l_u} \in \underset{i\in u}{\times} [1:L_i]$. $(\boldsymbol{\beta})_{\boldsymbol{l_u},u} = \beta^u_{\boldsymbol{l_u}}$, $\forall\ \boldsymbol{l_u} \in \underset{i\in u}{\times} [1:L_i]$, $\forall\ u \subset [1:p],\ u \neq \emptyset$.
Remind that the dimension of spaces $H_u^{0,L}$ is denoted by $L_u = \prod_{i\in u} L_i$, $\forall\ u \in S \setminus \{\emptyset\}$. Thus, the number of parameters to be estimated in (8) is equal to $m := \sum_{\substack{u\subset[1:p]\\ u\neq\emptyset}} L_u$. Let us remark that it would be numerically very expensive to consider the estimation of all these coefficients. Even for small ANOVA order d, the number of terms blows up with the dimensionality of the problem, and so does the number of model evaluations when using an ordinary least-squares regression scheme. As an illustration, take $d = 3$, $p = 8$ and $L_i = L = 5$, $\forall\ i \in [1:p]$. In this case, $m = 7740$ parameters are to be estimated, which could be a difficult task in practice. To handle this problem, many variable selection methods have been considered in the field of statistics. The next section aims at briefly exposing the variable selection methods via a penalized regression. We particularly focus on the ℓ_0 penalty [28] and on the Lasso regression [29].

4.2. The Variable Selection Methods

For simplicity, we denote by m the number of parameters in (8). The variable selection methods usually deal with the penalized regression

$$\min_{\boldsymbol{\beta}} \|\mathbb{Y} - \mathbb{X}_\varphi \boldsymbol{\beta}\|^2_{\mathrm{n}} + \lambda J(\boldsymbol{\beta}) \tag{9}$$

where $J(\cdot)$ is positive valued for $\boldsymbol{\beta} \neq 0$, and where $\lambda \geq 0$ is a tuning parameter. The most intuitive approach is to consider the ℓ_0-penalty $J(\boldsymbol{\beta}) = \|\boldsymbol{\beta}\|_0$, where $\|\boldsymbol{\beta}\|_0 = \sum_{j=1}^m \mathbb{1}(\beta_j \neq 0)$. Indeed, the ℓ_0 regularization aims at selecting nonzero coefficients, thus at removing the useless parameters from the model. The greedy approximation [28] offers a series of strategies to deal with the ℓ_0-penalty. Nevertheless, the ℓ_0 regularization is a non convex function, and suffers from the statistical instability, as mentioned in [6, 29]. A convex relaxation of the optimization problem can be viewed with the Lasso regression [29]. Indeed, the Lasso regression corre-

sponds to the ℓ_1-penalty, i.e. (9) with $J(\boldsymbol{\beta}) = \|\boldsymbol{\beta}\|_1$, and $\|\boldsymbol{\beta}\|_1 = \sum_{j=1}^{m} |\beta_j|$. The Lasso offers a good compromise between a rough selection of nonzero elements, and a ridge regression ($J(\boldsymbol{\beta}) = \sum_{j=1}^{m} \beta_j^2$) that only shrinks coefficients, but is known to be stable [7, 12]. To offer a good panel to the reader, we will adapt our method to the ℓ_0 and to the ℓ_1 regularization.

The adaptive forward-backward greedy (*FoBa*) algorithm proposed in Zhang [32] is exploited here to deal with the ℓ_0 penalization. From a dictionary $\mathcal{D}$ that can be large and/or redundant, the *FoBa* algorithm is an iterative scheme that sequentially selects and deletes the element of $\mathcal{D}$ that has the least impact on the fit. The aim of the algorithm is to efficiently select a limited number of predictors. The advantage of such approach is that it is very intuitive, and easy to implement. In our problem, the *FoBa* algorithm is applied on the whole set of basis functions. It can then happen that none basis function is retained for the estimation of a HOFD component. In this case, as we want to estimate each component of the HOFD, the coefficient corresponding to this component is set to be zero.

Initiated by Efron *et al.* [11], the modified LARS algorithm is further adapted to our problem to deal with the Lasso regression. The LARS is a general iterative technique that builds up the regression function by successive steps. The adaptation of LARS to Lasso (the modified LARS) is inspired by the homotopy method proposed by Osborne *et al.* [23]. The main advantage of the modified LARS algorithm is that it builds up the whole regularized solution path $\{\hat{\boldsymbol{\beta}}(\lambda), \lambda \in \mathbb{R}\}$, exploiting the property of piecewise linearity of the solutions with respect to λ [7, 24].

In the next part, both *FoBa* algorithm and the modified LARS algorithm are adapted to our problem and they are compared via numerical examples.

4.3. Summary of the Estimation Procedure

Provided an initial good choice of orthonormal systems $(\psi_{l_i}^i)_{l_i=0, i\in[1:p]}^{L_i}$, we first construct the approximating spaces $G_{u,\mathrm{n}}^{0,L}$ of H_u^0 for $|u| \leq d$, for $d < p$, thanks to the HOGS Procedure of Section 3. A HOFD component η_u is then a projection onto $G_{u,\mathrm{n}}^{0,L}$, whose coefficients are defined by least-squares estimation. To bypass the curse

of dimensionality, the *FoBa* algorithm or the modified LARS algorithm is used. Once the HOFD components are estimated, we deduce the empirical estimation of the generalized Sobol sensitivity indices given in Definition 1.

5. ASYMPTOTIC RESULTS

In this section, we then assume that a functional ANOVA of order d, with $d \ll p$, substitutes properly the initial model. Let us denote S^d the collection of all subsets of $[1:p]$ of size at most d. We suppose that

$$\eta(\mathbf{X}) \approx \eta^R(\mathbf{X}) = \sum_{u \in S^d} \eta_u^R(\mathbf{X}), \quad \text{with} \;\; \eta_u^R = \sum_{\boldsymbol{l_u} \in \underset{i \in u}{\times} [1:L_i]} \beta_{\boldsymbol{l_u}}^{u,0} \phi_{\boldsymbol{l_u}}^u(\mathbf{X_u}) \in H_u^{0,L}$$

In the following, we give the convergence properties of the estimator $\hat{\eta}^R$ to η^R, with

$$\hat{\eta}^R(\mathbf{X}) := \sum_{u \in S^d} \hat{\eta}_u^R(\mathbf{X_u}), \quad \text{with} \;\; \hat{\eta}_u^R(\mathbf{X_u}) = \sum_{\boldsymbol{l_u} \in \underset{i \in u}{\times} [1:L_i]} \hat{\beta}_{\boldsymbol{l_u}}^u \varphi_{\boldsymbol{l_u},\mathrm{n}}^u(\mathbf{X_u}) \in G_{u,\mathrm{n}}^{0,L}.$$

where $(\hat{\beta}_{\boldsymbol{l_u}}^u)_{\boldsymbol{l_u} \in \underset{i \in u}{\times} [1:L_i]}$, $u \in S^d$ are estimated by the minimization problem (8). Thus, we are interested in the convergence results when the ANOVA order d and the order of truncation $L_u = \prod_{i \in u} L_i$ $(u \in S^d)$, are fixed.

Proposition 2 gives the convergence result.

Proposition 2. *Assume that*

$$Y = \eta^R(\mathbf{X}) + \varepsilon, \quad \textit{where} \;\; \eta^R(\mathbf{X}) = \sum_{u \in S^d} \sum_{\boldsymbol{l_u} \in \underset{i \in u}{\times} [1:L_i]} \boldsymbol{\beta}_{\boldsymbol{l_u}}^{u,0} \phi_{\boldsymbol{l_u}}^u(\mathbf{X_u}) \in H_u^{0,L}$$

with $\mathbb{E}(\varepsilon) = 0$, $\mathbb{E}(\varepsilon^2) = \sigma_*^2$, $\mathbb{E}(\varepsilon \cdot \phi_{\boldsymbol{l_u}}^u(\mathbf{X_u})) = 0$, $\forall\ \boldsymbol{l_u} \in \underset{i \in u}{\times} [1 : L_i]$, $\forall\ u \in S^d$. *($\boldsymbol{\beta}_0 = (\boldsymbol{\beta}_{\boldsymbol{l_u}}^{u,0})_{\boldsymbol{l_u},u} \in \Theta$ is the true parameter).*

Further, let us consider the least-squares estimation $\hat{\eta}^R$ of η^R using the sample $(y^s, \mathbf{x}^s)_{s \in [1:n]}$ and the functions $(\varphi_{\boldsymbol{l_u}^u})_{\boldsymbol{l_u},u}$, that is

$$\hat{\eta}^R(\mathbf{X}) = \sum_{u \in S^d} \hat{\eta}_u^R(\mathbf{X_u}), \quad \textit{where} \;\; \hat{\eta}_u^R(\mathbf{X_u}) = \sum_{\boldsymbol{l_u} \in \underset{i \in u}{\times} [1:L_i]} \hat{\beta}_{\boldsymbol{l_u}}^u \varphi_{\boldsymbol{l_u},n}^u(\mathbf{X_u}) \in G_{u,n}^{0,L}$$

where $\hat{\boldsymbol{\beta}} = Arg\min_{\boldsymbol{\beta} \in \Theta} \|\mathbb{Y} - \mathbb{X}_\varphi \boldsymbol{\beta}\|_n^2$. If we assume that

(H) The distribution $P_{\mathbf{X}}$ is equivalent to $\otimes_{i=1}^p P_{X_i}$,

then

$$\left\|\hat{\eta}^R - \eta^R\right\| \overset{a.s.}{\to} 0 \quad \textit{when } n \to +\infty. \tag{10}$$

The sketch of proof of Proposition 2 is postponed to Appendix A. For further interest, the detailed proof of Proposition 2 is postponed to Appendix B, as the annex of this document.

Our aim here is to study how the approximating spaces $G_{u,\mathrm{n}}^{0,L}$, constructed with the previous procedure, behave when $n \to +\infty$. However, we assume that the ANOVA order d and the order of truncation L_u ($u \in S^d$) are fixed. By extending the work of Stone [27], Huang [15] explores the convergence properties of functional ANOVA models when d and L_u are not fixed anymore. Nevertheless, the results are obtained for a general model space H_u, and its approximating space G_u, $\forall\ u \in S$. In [15], the author states that if the basis functions are m-smooth and bounded, $\|\hat{\eta} - \eta\|$ converges in probability. For polynomials, Fourier transforms or splines, he specifically shows that $\|\hat{\eta} - \eta\| = O_p(n^{-\frac{2m}{2m+d}})$ (See [15] p. 257). Thus, to get a good rate of convergence, it is in our interest to have a small order d in a model. In the next numerical applications, we use this theoretical result to substitute the initial model to a functional ANOVA of order at most $d = 3$.

6. APPLICATION

In this section, we consider two numerical applications. The first model is a toy function studied in Li *et al.* [18]. It is used to study the numerical properties of the practical method summarized in Section 4.3. The last application is the study of a shell subject to an internal pressure. From a finite elements code, we estimate the generalized sensitivity indices of input parameters implied into the model. The goal is to quantify the sensitivity of each input variable in a context of dependence.

6.1. A Test Case

Remind that the HOGS procedure is used to construct basis functions adapted to the hierarchical orthogonality constraints. Hence, the same method improved by the adaptive greedy algorithm will be abbreviated GHOGS method (G for Greedy). The method improved by the modified LARS algorithm will be called LHOGS (L for LARS).

The estimation procedure suggested in [9], and reminded in the introduction, is compared to the G/LHOGS methods. It will be denoted POM (for Projection Operators Method), relatively to the projection operators it uses.

Let $\mathbf{X} \sim N(0, \Sigma)$ and the model function,

$$\boxed{Y = g_1(X_1, X_2) + g_2(X_2) + g_3(X_3)}$$

with

$$\begin{aligned} g_1(X_1, X_2) &= [a_1 X_1 + a_0][b_1 X_2 + b_0] \\ g_2(X_2) &= c_2 X_2^2 + c_1 X_2 + c_0 \\ g_3(X_3) &= d_3 X_3^3 + d_2 X_3^2 + d_1 X_3 + d_0 \end{aligned}$$

and

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \gamma\sigma_1\sigma_2 & 0 \\ \gamma\sigma_1\sigma_2 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{pmatrix}$$

Condition (C.2) does not allow to use the normal distribution, but rather the mixture Gaussian one [9]. However, the Gaussian distribution allows for computing a HOFD decomposition, as done in [18]. Moreover, if the research of solutions is restricted to the polynomial spaces, the uniqueness of the HOFD components given in [18] is ensured, whatever the type of distribution. Thus, the analytical form of the generalized Sobol indices can be deduced in this case.

We take $a_0 = c_1 = d_0 = 1$, $a_1 = b_0 = c_2 = d_1 = d_2 = 2$ and $b_1 = c_0 = d_3 = 3$. The variations are fixed at $\sigma_1 = \sigma_2 = 0.2$, $\sigma_3 = 0.18$ and $\gamma = 0.6$. To show the interest of the greedy and Lasso application, we proceed to $n = 200$ model

evaluations repeated 50 times. For each component, we choose a Hermite basis of degree 10. Thus, the number of parameters $m = 330 > n = 200$. In view of analytical results given in [18], it is clear that only 8 among 330 basis functions are necessary to restitute entirely the model. Table 1 helps for comparing the POM with the GHOGS/LHOGS procedures on the sensitivity indices estimation and their standard deviation (indicated into brackets). Table 1 also provides the number of nonzero estimated coefficients for the *FoBa* and the modified LARS algorithm. The estimated components $\hat{\eta}_1$, $\hat{\eta}_2$ and $\hat{\eta}_3$ are represented in Figure 1.

	S_1	S_2	S_3	S_{12}	S_{13}	S_{23}	S_{123}	$\sum_j \mathbb{1}(\hat{\beta}_j \neq 0)$
Analytical	0.4429	0.4718	0.0763	0.0091	0	0	0	-
POM	0.4402	0.4718	0.0810	-0.0014	-	-	-	-
	(0.021)	(0.0401)	(0.0012)	(0.001)	-	-	-	
GHOGS	0.4499	0.4647	0.0754	0.0030	0	0	0.0070	5 to 7
	(0.0272)	(0.0328)	(0.0249)	(0.0032)	(0)	(0)	(0.0024)	
LHOGS	0.4534	0.4688	0.0793	0.0060	0.0013	0.0010	-0.0098	22 to 207
	(0.0275)	(0.0314)	(0.0258)	(0.0024)	(0.0012)	(0.0012)	(0.0002)	

Table 1: Sensitivity indices estimation with the POM and the G/LHOGS methods

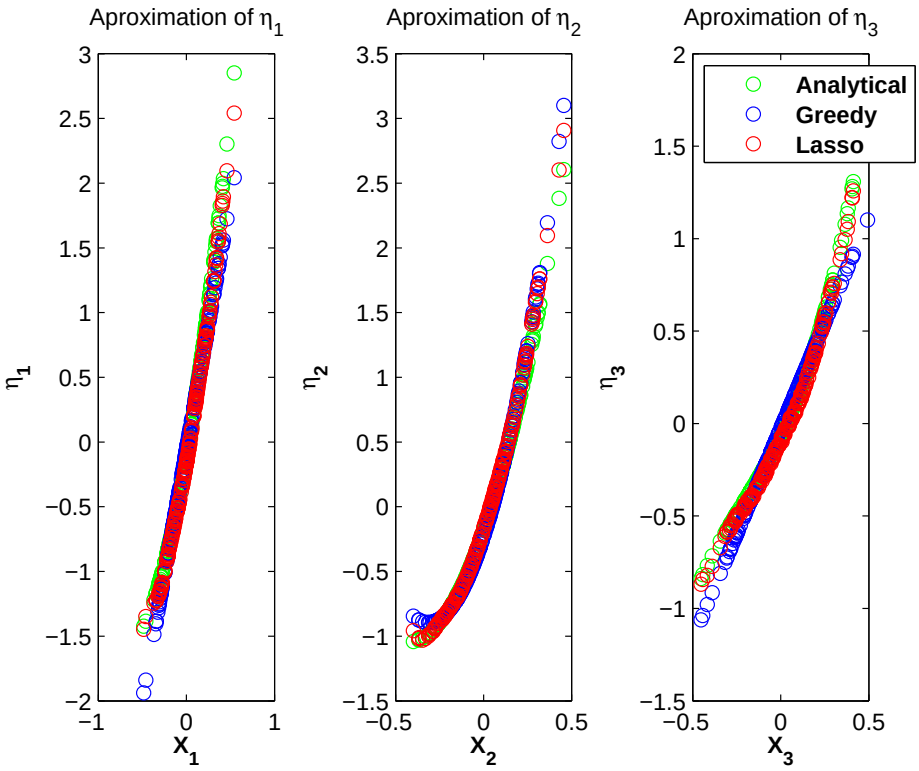

Figure 1: Estimation of the HOFD components by the G/LHOGS methods

The advantage of the GHOGS and of LHOGS methods is to be able to estimate all interaction indices, whereas the POM only estimates interaction indices involved in dependent pairs [9]. Even if many of the nonzero coefficients in LARS are close to zero, this method tends to estimate a large number of nonzero parameters in comparison with the GHOGS procedure. Through this example, the greedy restitutes properly the information with a relevant selection of coefficients.

6.2. The Tank Pressure Model

The case study concerns a shell closed by a cap and subject to an internal pressure. Figure 2 illustrates a simulation of tank distortion. We are interested in the von Mises stress [30] on the point y labeled in Figure 2. The von Mises stress allows for predicting material yielding which occurs when it reaches the material yield strength. The selected point y corresponds to the point for which the von Mises

stress is maximal in the tank. Therefore, we want to prevent the tank from material damage induced by plastic deformations. To offer a large panel of tanks able to resist to the internal pressure, a manufacturer wants to know the most contributive parameters to the von Mises criterion variability. In the model we propose, the von Mises criterion depends on three geometrical parameters: the shell internal radius (R_{int}), the shell thickness (T_{shell}), and the cap thickness (T_{cap}). It also depends on five physical parameters concerning the Young's modulus (E_{shell} and E_{cap}) and the yield strength ($\sigma_{y,shell}$ and $\sigma_{y,cap}$) of the shell and the cap. The last parameter is the internal pressure (P_{int}) applied to the shell. The system is modelized by a 2D finite elements code ASTER.

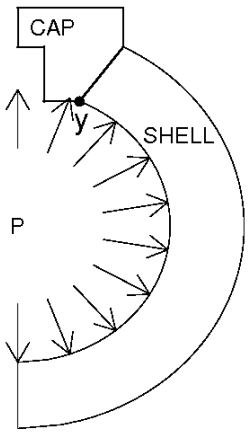

Figure 2: Tank distortion at point y

In table 2, we give the input distributions.

Inputs	Distribution
R_{int}	$\mathcal{U}([1800; 2200]),\ \gamma(R_{int}, T_{shell}) = 0.85$
T_{shell}	$\mathcal{U}([360; 440]),\ \gamma(T_{shell}, T_{cap}) = 0.3$
T_{cap}	$\mathcal{U}([180; 220]),\ \gamma(T_{cap}, R_{int}) = 0.3$
E_{cap}	$\alpha N(\mu, \Sigma) + (1-\alpha)N(\mu, \Omega)$
$\sigma_{y,cap}$	$\alpha = 0.02,\ \mu = \begin{pmatrix} 210 \\ 500 \end{pmatrix},\ \Sigma = \begin{pmatrix} 350 & 0 \\ 0 & 29 \end{pmatrix},\ \Omega = \begin{pmatrix} 175 & 81 \\ 81 & 417 \end{pmatrix}$
E_{shell}	$\alpha N(\mu, \Sigma) + (1-\alpha)N(\mu, \Omega)$
$\sigma_{y,shell}$	$\alpha = 0.02,\ \mu = \begin{pmatrix} 70 \\ 300 \end{pmatrix},\ \Sigma = \begin{pmatrix} 117 & 0 \\ 0 & 500 \end{pmatrix},\ \Omega = \begin{pmatrix} 58 & 37 \\ 37 & 250 \end{pmatrix}$
P_{int}	$N(80, 10)$

Table 2: Description of inputs of the shell model

The geometrical parameters are uniformly distributed because of the large choice left for the tank building. The correlation γ between the geometrical parameters is induced by the constraints of manufacturing processes. The physical inputs are normally distributed and their uncertainty are due to the manufacturing process and the properties of the elementary constituents variabilities. The large variability of P_{int} in the model corresponds to the different internal pressure values which could be applied to the shell by the user.

To measure the contribution of the correlated inputs to the output variability, we estimate the generalized sensitivity indices by the practical method exposed in Section 4. We proceed to $n = 1000$ simulations over 50 runs. We use the 5-spline functions

for the geometrical parameters and the Hermite basis functions of degree 7 for the physical parameters. The first order indices dispersions are displayed in Figure 3 for both Greedy and LARS algorithm.

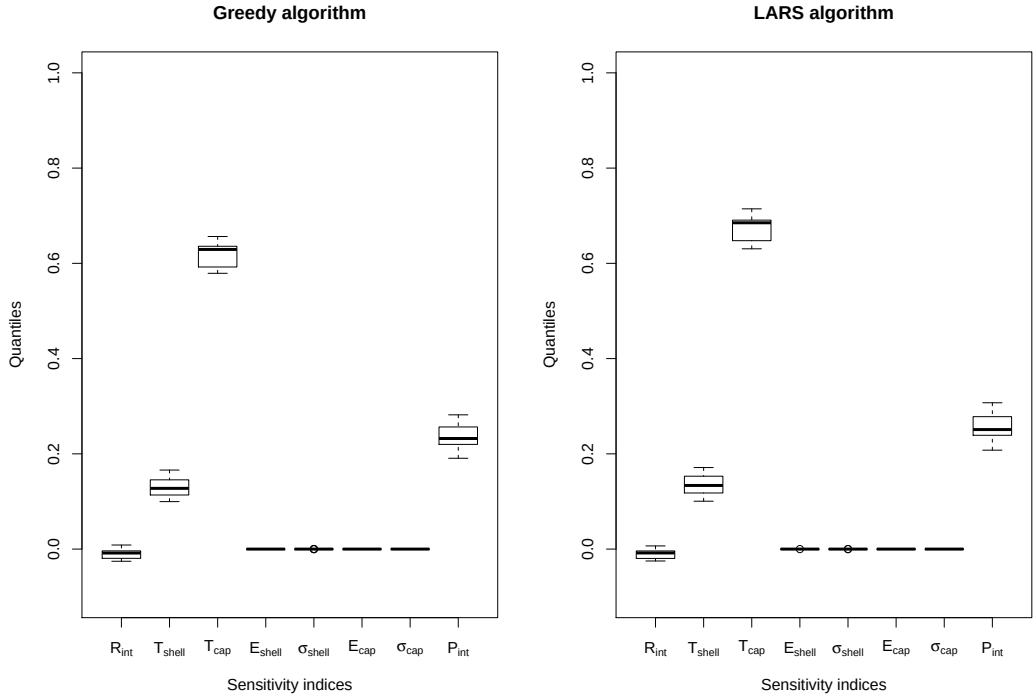

Figure 3: Boxplot representations of the first order sensitivity indices

We observe first that the HOGS procedure applied with the greedy and the LARS techniques give very similar results. Once again, we do not observe a big difference between the two variable selection methods.

The first four physical parameters are independent from the other inputs, and their effects are null, so we can deduce that they do not have any influence in the model. Also, even if the internal pressure plays an important role, the strongest contribution comes from the correlated set of geometrical inputs $(R_{int}, T_{shell}, T_{cap})$. The sensitivity index related to the shell radius (R_{int}) is negative, so the covariance induced by the dependence dominates in the index, showing that either there is a strong negative covariance part or the full contribution of the variable is small. In the first case, it shows that R_{int} is influent through its correlation. In the second one, the input R_{int} is not an influent variable in the model. The sensitivity indices of the shell

thickness (T_{shell}) and the cap thickness (T_{cap}) reveal that these two variables have a strong influence in the model. Thus, to scale down the variability in the model, we should reduce the cap thickness variability first. Because of the strong correlation between the shell radius and its thickness, one should reduce the variability of both parameters.

APPENDIX A : CONVERGENCE RESULTS

For sake of clarity, we first recall and define some notation that will be used further.

Settings

First, as mentioned in Section 4, we assume that Y is centered. Also, we assume that a functional ANOVA of order d, with $d \leq p$, substitutes properly the initial model. Let us denote S^d the collection of all subsets of $[1:p]$ of size at most d, where the empty set has been removed. Thus, $S^d = \{u,\ \emptyset \subset u \subseteq [1:p] \text{ and } |u| \leq d\}$. Recall that, $\forall\ i \in [1:p]$, $L_i := L_{\{i\}}$ is the dimension of the spaces $H_i^{0,L}$ and $G_{i,\mathrm{n}}^{0,L}$. More generally, $\phi_{\{i\}} := \phi_i$. Also, $L_u := \prod_{i\in u} L_i$ is the dimension of the spaces $H_u^{0,L}$ and $G_{u,\mathrm{n}}^{0,L}$.

For $u \in S^d$, $\boldsymbol{l_u} = (l_u^i)_{i\in u}$ is a multi-index of $\underset{i\in u}{\times}[1:L_i]$, where $\underset{i\in u}{\times}[1:L_i]$ is the Cartesian product of the sets $[1:L_i]$, for $i \in u$.

We refer $(\phi^u_{\boldsymbol{l_u}})_{\boldsymbol{l_u}\in \underset{i\in u}{\times}[1:L_i]}$ as the basis of $H_u^{0,L}$ and $(\varphi^u_{\boldsymbol{l_u},\mathrm{n}})_{\boldsymbol{l_u}\in \underset{i\in u}{\times}[1:L_i]}$ as the basis of $G_{u,\mathrm{n}}^{0,L}$ constructed according to HOGS Procedure of Section 3. Thus, these functions all lie in $L^2_{\mathbb{R}}(\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p), P_{\mathbf{X}})$.

$\langle\cdot,\cdot\rangle$ and $\|\cdot\|$ are used as the inner product and norm on $L^2_{\mathbb{R}}(\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p), P_{\mathbf{X}})$,

$$\langle h_1, h_2\rangle = \int h_1(\mathbf{x})h_2(\mathbf{x})p_{\mathbf{X}}d\nu(\mathbf{x}), \quad \|h\|^2 = \langle h, h\rangle$$

while $\langle\cdot,\cdot\rangle_{\mathrm{n}}$ and $\|\cdot\|_{\mathrm{n}}$ denote the empirical inner product and norm, that is

$$\langle g_1, g_2\rangle_{\mathrm{n}} = \frac{1}{n}\sum_{s=1}^{n} g_1(\mathbf{x}^s)g_2(\mathbf{x}^s), \quad \|g\|^2_{\mathrm{n}} = \langle g, g\rangle_{\mathrm{n}}$$

when $(y^s, \mathbf{x}^s)_{s=1,\cdots,n}$ is the n-sample of observations from the distribution of $(Y, \mathbf{X})$.

We set $m := \sum_{u\in S^d} L_u$ the number of parameters in the regression model. Denote, for all $u \in S^d$, $\Phi_u(\mathbf{X_u}) \in (L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_{\mathbf{X}}))^{L_u}$, with $(\Phi_u(\mathbf{X_u}))_{\boldsymbol{l_u}} = \phi^u_{\boldsymbol{l_u}}(\mathbf{X_u})$, and by $\boldsymbol{\beta}$ any vector of $\Theta \subset \mathbb{R}^m$, where $(\boldsymbol{\beta})_{\boldsymbol{l_u},u} = \beta^u_{\boldsymbol{l_u}}$, $\forall\ \boldsymbol{l_u} \in \underset{i\in u}{\times}[1 : L_i]$.

Recall that, for $a, b \in \mathbb{N}^*$, $\mathcal{M}_{a,b}(\mathbb{R})$ denotes the set of all real matrices with a rows and b columns. Set $\mathbb{X}_\varphi = \begin{pmatrix}\boldsymbol{\varphi}_1 & \cdots & \boldsymbol{\varphi}_u & \cdots\end{pmatrix} \in \underset{u\in S^d}{\times}\mathcal{M}_{n,L_u}(\mathbb{R})$, where, for $u \in S^d$, $(\boldsymbol{\varphi}_u)_{s,\boldsymbol{l_u}} = \varphi^u_{\boldsymbol{l_u},\mathrm{n}}(\mathbf{x_u}^s)$, $\forall\ s \in [1 : n]$, $\forall\ \boldsymbol{l_u} \in \underset{i\in u}{\times}[1 : L_i]$. Also, we set $\mathbb{X}_\phi = \begin{pmatrix}\mathfrak{C}_1 & \mathfrak{C}_2 & \cdots\end{pmatrix} \in \underset{u\in S^d}{\times}\mathcal{M}_{n,L_u}(\mathbb{R})$, where, for $u \in S^d$, $(\mathfrak{C}_u)_{s,\boldsymbol{l_u}} = \phi^u_{\boldsymbol{l_u}}(\mathbf{x_u}^s)$, $\forall\ s \in [1 : n]$, $\forall\ \boldsymbol{l_u} \in \underset{i\in u}{\times}[1 : L_i]$.

Denote by A_ϕ be the $m\times m$ Gramian matrix whose block entries are $(\mathbb{E}(\Phi_u(\mathbf{X_u})^t\Phi_v(\mathbf{X_v})))_{u,v\in S^d}$.

At last, we define the functions

$$M_n(\boldsymbol{\beta}) = \|\mathbb{Y} - \mathbb{X}_\varphi\boldsymbol{\beta}\|^2_{\mathrm{n}} \tag{11}$$

where $\mathbb{Y}_s = y^s$, $\forall\ s \in [1 : n]$. At last, the Euclidean norm will be denoted $\|\cdot\|_2$.

Proposition 2. *Assume that*

$$Y = \eta^R(\mathbf{X}) + \varepsilon, \quad \textit{where}\ \ \eta^R(\mathbf{X}) = \sum_{u\in S^d}\sum_{\boldsymbol{l_u}\in\underset{i\in u}{\times}[1:L_i]} \boldsymbol{\beta}^{u,0}_{\boldsymbol{l_u}}\phi^u_{\boldsymbol{l_u}}(\mathbf{X_u}) \in H^{0,L}_u$$

with $\mathbb{E}(\varepsilon) = 0$, $\mathbb{E}(\varepsilon^2) = \sigma^2_*$, $\mathbb{E}(\varepsilon\cdot\phi^u_{\boldsymbol{l_u}}(\mathbf{X_u})) = 0$, $\forall\ \boldsymbol{l_u} \in \underset{i\in u}{\times}[1 : L_i]$, $\forall\ u \in S^d$. *($\boldsymbol{\beta}_0 = (\boldsymbol{\beta}^{u,0}_{\boldsymbol{l_u}})_{\boldsymbol{l_u},u}$ is the true parameter).*

Further, let us consider the least-squares estimation $\hat\eta^R$ of η^R using the sample $(y^s, \mathbf{x}^s)_{s\in[1:n]}$ and the functions $(\varphi_{\boldsymbol{l_u}^u})_{\boldsymbol{l_u},u}$, that is

$$\hat\eta^R(\mathbf{X}) = \sum_{u\in S^d}\hat\eta^R_u(\mathbf{X_u}), \quad \textit{where}\ \ \hat\eta^R_u(\mathbf{X_u}) = \sum_{\boldsymbol{l_u}\in\underset{i\in u}{\times}[1:L_i]} \hat\beta^u_{\boldsymbol{l_u}}\varphi^u_{\boldsymbol{l_u},n}(\mathbf{X_u}) \in G^{0,L}_{u,n}$$

where $\hat{\boldsymbol{\beta}} = Arg\min_{\boldsymbol{\beta}\in\Theta} M_n(\boldsymbol{\beta})$. If we assume that

(H) The distribution $P_{\mathbf{X}}$ is equivalent to $\otimes_{i=1}^p P_{X_i}$,

then

$$\left\|\hat\eta^R - \eta^R\right\| \overset{a.s.}{\to} 0, \quad \textit{when}\ \ n \to +\infty. \tag{12}$$

The proof of Proposition 2 is broken up into Lemmas 1-5. To prove that (12), we introduce $\bar\eta^R$ as the following approximation of η^R,

$$\bar\eta^R = \sum_{u\in S^d}\bar\eta^R_u(\mathbf{X_u}) = \sum_{u\in S^d}\sum_{\boldsymbol{l_u}\in\underset{i\in u}{\times}[1:L_i]} \beta^u_{\boldsymbol{l_u}}\varphi^u_{\boldsymbol{l_u},\mathrm{n}}(\mathbf{X_u}),$$

and we write the triangular inequality,

$$\|\hat{\eta}^R - \eta^R\| = \|\hat{\eta}^R - \bar{\eta}^R + \bar{\eta}^R - \eta^R\| \leq \|\hat{\eta}^R - \bar{\eta}^R\| + \|\bar{\eta}^R - \eta^R\| \tag{13}$$

Thus, it is enough to prove that $\left\|\hat{\eta}^R - \bar{\eta}^R\right\| \overset{a.s.}{\to} 0$, and that $\left\|\bar{\eta}^R - \eta^R\right\| \overset{a.s.}{\to} 0$. Lemmas 4 and 5 deal with convergence results on $\|\bar{\eta}^R - \eta^R\|$ and on $\|\hat{\eta}^R - \bar{\eta}^R\|$, respectively. Lemmas 1, 2, 3 are preliminary results to prove Lemmas 4 and 5. Lemmas 1-5 are enunciated further below.

Preliminary results

Lemma 1. *If (H) holds, then A_ϕ is a non singular matrix.*

Lemma 2. *Let $u, v \in S^d$. Assume that $\|\varphi^u_{\boldsymbol{l_u},n} - \phi^u_{\boldsymbol{l_u}}\| \overset{a.s.}{\to} 0$ and $\|\varphi^v_{\boldsymbol{l_v},n} - \phi^v_{\boldsymbol{l_v}}\| \overset{a.s.}{\to} 0$ for any $\boldsymbol{l_u} \in \underset{i\in u}{\times}[1 : L_i]$, $\boldsymbol{l_v} \in \underset{i\in v}{\times}[1 : L_i]$. Then, the following results hold:*

(i) $\left\|\varphi^u_{\boldsymbol{l_u},n}\right\| \overset{a.s.}{\to} \left\|\phi^u_{\boldsymbol{l_u}}\right\|$ *and* $\left\|\varphi^u_{\boldsymbol{l_u},n}\right\|_n \overset{a.s.}{\to} \left\|\phi^u_{\boldsymbol{l_u}}\right\|$;

(ii) $\langle \phi^u_{\boldsymbol{l_u}}, \varphi^v_{\boldsymbol{l_v},n}\rangle \overset{a.s.}{\to} \langle \phi^u_{\boldsymbol{l_u}}, \phi^v_{\boldsymbol{l_v}}\rangle$ *and* $\langle \phi^u_{\boldsymbol{l_u}}, \varphi^v_{\boldsymbol{l_v},n}\rangle_n \overset{a.s.}{\to} \langle \phi^u_{\boldsymbol{l_u}}, \phi^v_{\boldsymbol{l_v}}\rangle$;

(iii) $\langle \varphi^u_{\boldsymbol{l_u},n}, \varphi^v_{\boldsymbol{l_v},n}\rangle \overset{a.s.}{\to} \langle \phi^u_{\boldsymbol{l_u}}, \phi^v_{\boldsymbol{l_v}}\rangle$ *and* $\langle \varphi^u_{\boldsymbol{l_u},n}, \varphi^v_{\boldsymbol{l_v},n}\rangle_n \overset{a.s.}{\to} \langle \phi^u_{\boldsymbol{l_u}}, \phi^v_{\boldsymbol{l_v}}\rangle$.

Lemma 3. *Let $u \in S^d$. For any $l^i_u \in [1 : L_i]$, assume that $\|\varphi^i_{l^i_u,n} - \phi^i_{l^i_u}\| \overset{a.s.}{\to} 0$. Moreover, we also assume that there exists $v \in S^d$ and $\boldsymbol{l_v} \in \underset{i\in v}{\times}[1 : L_i]$ such that $\|\varphi^v_{\boldsymbol{l_v},n} - \phi^v_{\boldsymbol{l_v}}\| \overset{a.s.}{\to} 0$. Then, we have*

(i) $\left\|\prod_{i\in u}\varphi^i_{l^i_u,n} - \prod_{i\in u}\phi^i_{l^i_u}\right\| \overset{a.s.}{\to} 0$;

(ii) $\langle \prod_{i\in u}\varphi^i_{l^i_u,n}, \varphi^v_{\boldsymbol{l_v},n}\rangle_n \overset{a.s.}{\to} \langle \prod_{i\in u}\phi^i_{l^i_u}, \phi^v_{\boldsymbol{l_v}}\rangle$.

Main convergence results

Lemma 4. *Remind that the true regression function is*

$$\eta^R(\mathbf{X}) = \sum_{u\in S^d} \eta^R_u(\mathbf{X_u}), \text{ where } \eta^R_u(\mathbf{X_u}) = \sum_{\boldsymbol{l_u}\in \underset{i\in u}{\times}[1:L_i]} \beta^{u,0}_{\boldsymbol{l_u}} \phi^u_{\boldsymbol{l_u}}(\mathbf{X_u}) \in H^{0,L}_u$$

Further, let $\bar{\eta}^R$ be the approximation of η^R,

$$\bar{\eta}^R(\mathbf{X}) = \sum_{u\in S^d} \bar{\eta}^R_u(\mathbf{X_u}), \text{ where } \bar{\eta}^R_u(\mathbf{X_u}) = \sum_{\boldsymbol{l_u}\in \underset{i\in u}{\times}[1:L_i]} \beta^{u,0}_{\boldsymbol{l_u}} \varphi^u_{\boldsymbol{l_u},n}(\mathbf{X_u}) \in G^{0,L}_{u,n},$$

Then $\|\bar{\eta}_u^R - \eta_u^R\| \overset{a.s.}{\to} 0 \quad \forall\ u \in S^d$, *and* $\|\bar{\eta}^R - \eta^R\| \overset{a.s.}{\to} 0$.

Lemma 5. *Recall that* $\hat{\boldsymbol{\beta}} = Arg\min_{\boldsymbol{\beta}\in\Theta} M_n(\boldsymbol{\beta})$. *If (H) holds, then*

$$\left\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right\|_2 \overset{a.s.}{\to} 0 \tag{14}$$

Moreover,

$$\left\|\hat{\eta}^R - \eta^R\right\| \overset{a.s.}{\to} 0. \tag{15}$$

ACKNOWLEDGMENTS

The authors would like to thank Loic Le Gratiet and Gilles Defaux for their help in providing and understanding the tank pressure model.
This work has been supported by French National Research Agency (ANR) through COSINUS program (project COSTA-BRAVA number ANR-09-COSI-015).

References

[1] T. Bedford. Sensitivity indices for (tree-)dependent variables. In K. Chan, S. Tarantola, and F. Campolongo, editors, *SAMO98*, pages 17–20. Second International Symposium on Sensitivity Analysis of Model Output, 1998.

[2] E. Borgonovo. A new uncertainty importance measure. *Reliability Engineering and System Safety*, 92:771–784, 2007.

[3] E. Borgonovo, W. Castaings, and S. Tarantola. Moment independent importance measures: New results and analytical test cases. *Risk Analysis*, 31(3):404–428, 2011.

[4] E. Borgonovo, W. Castaings, and S. Tarantola. Model emulation and moment-independent sensitivity analysis: An application to environmental modelling. *Environmental Modelling & Software*, 34:105–115, 2012.

[5] E. Borgonovo and S. Tarantola. Moment-independent and variance-based sensitivity analysis with correlations: an application to the stability of a chemical reactor. *Wiley interscience*, 40:687–698, 2008.

[6] L. Breiman. Better subset regression using the nonnegative garrote. *Technometrics*, 37(4):373–384, 1995.

[7] P. Buhlmann and S. van de Geer. *Statistics for high-dimensional data*. Springer, Berlin, 2011.

[8] Y. Caniou and B. Sudret. Distribution-based global sensitivity analysis using polynomial chaos expansion. In *Procedia social and behavioral sciences 2*, pages 7625–7626. Sixth International Conference on Sensitivity Analysis of Model Output, 2010.

[9] G. Chastaing, F. Gamboa, and C. Prieur. Generalized hoeffding-sobol decomposition for dependent variables -application to sensitivity analysis. *Electronic Journal of Statistics*, 6:2420–2448, 2012.

[10] J.J. Droesbeke and G. Saporta. *Approches non paramétriques en régression*. Technip, Paris, 2006.

[11] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *The Annals of Statistics*, 32(2):407–451, 2004.

[12] T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning*. Springer, New York, 2009.

[13] W. Hoeffding. A class of statistics with asymptotically normal distribution. *The annals of Mathematical Statistics*, 19(3):293–325, 1948.

[14] G. Hooker. Generalized functional anova diagnostics for high-dimensional functions of dependent variables. *Journal of Computational and Graphical Statistics*, 16(3):709–732, 2007.

[15] J.Z. Huang. Projection estimation in multiple regression with application to functional anova models. *The annals of statistics*, 26(1):242–272, 1998.

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

[17] G. Li and H. Rabitz. D-morph regression: application to modeling with unknown parameters more than observation data. *Journal of mathematical chemistry*, 48(4):1010–1035, 2010.

[18] G. Li and H. Rabitz. General formulation of hdmr component functions with independent and correlated variables. *Journal of Mathematical Chemistry*, 50(1):99–130, 2012.

[19] G. Li, H. Rabitz, P.E. Yelvington, O. Oluwole, F. Bacon, Kolb C.E., and J. Schoendorf. Global sensitivity analysis with independent and/or correlated inputs. *Journal of Physical Chemistry A*, 114:6022–6032, 2010.

[20] G. Li, C. Rosenthal, and H. Rabitz. High dimensional model representations. *Journal of Physical Chemistry A*, 105(33):7765–7777, 2001.

[21] T. Mara and S. Tarantola. Variance-based sensitivity analysis of computer models with dependent inputs. *Reliability Engineering and System Safety*, 107:115–121, 2012.

[22] R.B. Nelsen. *An introduction to copulas*. Springer, New York, 2006.

[23] M.R. Osborne. A new approach to variable selection in least squares problems. *Journal of numerical analysis*, 20:389–404, 2000.

[24] M.R. Osborne, B. Presnell, and B.A. Turlach. On the lasso and its dual. *Journal of Computational and Graphical Statistics*, 9(2):319–337, 2000.

[25] A. Saltelli, K. Chan, and E.M. Scott. *Sensitivity Analysis*. Wiley, West Sussex, 2000.

[26] I.M. Sobol. Sensitivity estimates for nonlinear mathematical models. *Mathematical Modeling and Computational Experiment*, 1(4):407–414, 1993.

[27] C.J. Stone. The use of polynomial splines and their tensor products in multivariate function estimation. *The Annals of Statistics*, 22(1):118–171, 1994.

[28] V. Temlyakov. *Greedy approximation*. Cambridge monographs on applied and computational mathematics. Cambridge University Press, Cambridge, 2011.

[29] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*, 58(1):267–288, 1996.

[30] R. von Mises. Mechanik der festen krper im plastisch deformablen zustand. *Gttin. Nachr. Math. Phys.*, 1:582592, 1913.

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

[32] T. Zhang. Adaptive forward-backward algorithm for learning sparse representations. *IEEE transactions on information theory*, 57(7):4689–4708, 2011.