ANOVA decomposition of conditional Gaussian processes for sensitivity analysis with dependent inputs

Gaëlle Chastaing, Loic Le Gratiet

ANOVA decomposition of conditional Gaussian processes for sensitivity analysis with dependent inputs

Gaelle Chastaing
Loic Le Gratiet

October 14, 2013

Abstract

Complex computer codes are widely used in science to model physical systems. Sensitivity analysis aims to measure the contributions of the inputs on the code output variability. An efficient tool to perform such analysis are the variance-based methods which have been recently investigated in the framework of dependent inputs. One of their issue is that they require a large number of runs for the complex simulators. To handle it, a Gaussian process regression model may be used to approximate the complex code. In this work, we propose to decompose a Gaussian process into a high dimensional representation. This leads to the definition of a variance-based sensitivity measure well tailored for non-independent inputs. We give a methodology to estimate these indices and to quantify their uncertainty. Finally, the approach is illustrated on toy functions and on a river flood model.

Keywords: Sensitivity analysis, dependent inputs, Gaussian process regression, functional decomposition, complex computer codes.

1 Introduction

Many physical phenomena are investigated by complex models implemented in computer codes. Often considered as a black box function, a computer code calculates one or several output values which depend on input parameters. However, the code may depend on a very large number of incomes, that can be correlated among them. Moreover, input parameters are also subject to many sources of uncertainty, attributed to errors of measurements or a lack of information. These major flaws undermine the confidence a user have in the model. Indeed, the prediction given by the model may suffer from a large variability, leading to wrong conclusions.
To tackle these issues, the sensitivity analysis offers a series of methods and strategies that has been widely studied over the past decades [Saltelli et al., 2000, Saltelli et al., 2008, Cacuci et al., 2005]. Among the wide range of proposed methods, one could cite the class of global sensitivity analysis. Based on the assumption that the parameters implied in the model are randomly distributed, the global sensitivity analysis aims to identify and to rank the most contributive inputs to the response variability. One of the most popular global measure, the Sobol index, is based on a variance decomposition. Advanced by Hoeffding [Hoeffding, 1948],

the model function can be uniquely decomposed as a sum of mutually orthogonal functional components when input variables are independent. Following this idea, Sobol constructs sensitivity measures by expanding the global variance into partial variances. Then, the Sobol index apportions the individual contribution of a set of inputs by the ratio between the partial variance depending on this set and the global variance [Sobol, 1993].
However, the construction of such measure relies on the assumption that input variables are independent. When incomes are dependent, the use of the Sobol index is not excluded, but it may lead to a wrong interpretation. Indeed, as underlined by Mara *et al.* [Mara and Tarantola, 2012], if inputs are not independent, the amount of the response variance due to a given factor may be influenced by its dependence to other inputs. In other word, as the Sobol index only depends on terms of variance, we ignore how it differentiates the inputs dependence from their interactions. From this perspective, the construction of a sensitivity measure that quantifies the uncertainty brought by dependent inputs becomes clear.
A solution is to use a functional decomposition to build a variance-based sensitivity index. First, Xu *et al.* [Xu and Gertner, 2008] propose to decompose the partial variance of an input into a correlated and an uncorrelated contribution under the hypothesis that the effect of each parameter on the response is linear. The authors learn these contributions by successive linear regressions. To improve this approach, Li *et al.* [Li et al., 2010] propose to approximate the model function by a High Dimension Model Representation (HDMR), that consists of a sum of functional components of low dimensions [Li et al., 2001]. They suggest to reconstruct each term via the usual basis functions (polynomials, splines,...). Then, they deduce the decomposition of the response variance as a sum of partial variances and covariances. Recently, Caniou *et al.* [Caniou, 2012] suggest to build a HDMR by substituting the model function to a truncated polynomial chaos [Wiener, 1938] orthogonal with respect to the product of inputs marginal distribution. This choice is motivated by the fact that, when inputs are independent, the functional decomposition recovers the Hoeffding one, where each (unique) summand is expanded in terms of polynomial chaos [Sudret, 2008].
In a recent paper, Chastaing *et al.* [Chastaing et al., 2012] revisit the Hoeffding decomposition in a different way. To tackle the problem of uniqueness of the components of the decomposition proposed by the previous approaches, the authors give a unique decomposition of the theoretical model. The main strength of the approach is that it is not based on surrogate modeling. Initiated by the pioneering work of Stone [Stone, 1994], they show that any regular function can be uniquely decomposed as a sum of hierarchically orthogonal component functions. This means that two of these summands are orthogonal whenever all variables included in one of the component are also involved in the other. The decomposition leads to the definition of a generalized sensitivity index involving variance and covariance components. Further, the same authors propose a numerical method of estimation [Chastaing et al., 2013].
However, all these approaches suffer from two major flaws for time-consuming computer codes. First, the estimation of these measures is done by a regression method, which requires a very large number of model evaluations to be robust. Secondly, the number of decomposition components exponentially grows with the model dimension. In practice, we assume that only the low-order interaction terms contain the major part of the model behaviour. However, very few theoretical arguments confirm this assumption, and the truncation leads to an error of approximation that can be hardly controlled.
To overcome the first issue, we surrogate the computer code with a Gaussian process regression model. It is a non-parametric approach which considers that our prior knowledge about the code can be modeled by a Gaussian process (GP) [Santner et al., 2003, Rasmussen and

Williams, 2006]. These models are widely used in computer experiments to surrogate a complex computer code from few of its outputs ([Sacks et al., 1989]). Further, the use of a GP model for sensitivity analysis is motivated by arguments given in the literature. For independent inputs, a natural approach is to substitute the model function by the posterior mean of a given GP [Chen et al., 2005]. As this approach does not consider the posterior variance of the GP and thus the uncertainty of the surrogate modeling, Oakley & O'Hagan [Oakley and O'Hagan, 2004] substitute the initial model to a GP in the Sobol index. Then, the sensitivity index is given by the posterior mean of the Sobol index whereas the posterior variance measures its uncertainty. These two approaches are investigated and numerically compared in Marrel *et al.* [Marrel et al., 2009].
To handle with the second issue relative to the decomposition truncation, we propose here to extend the work done by Durrande *et al.* [Durrande et al., 2013] to the case of models with dependent inputs. In particular, we deal with GP specified by a covariance kernel that belongs to a special class of ANOVA kernels studied in [Berlinet and Thomas-Agnan, 2004, Durrande et al., 2013]. But instead of considering the posterior mean, we propose a functional decomposition of a GP distributed with respect to the posterior distribution. Similarly to the work of Caniou *et al.* [Caniou, 2012], the considered GP is decomposed as a sum of processes indexed by increasing dimension input variables. This expansion is such that the summands are mutually orthogonal with respect to the product of the inputs marginal distributions. Thus, as we have accessed to the GP, and as we can deal with each term of its decomposition, we can easily deduce the sum of every other terms, so that a truncation error can not be produced.
Consequently, the GP development leads to the construction of sensitivity measures based on the decomposition of the global variance as a sum of partial variances and covariances. The difference with the use of the polynomial chaos is that GP are not intrinsically linked to the distribution of the input variables, as it is the case for polynomial chaos [Cameron and Martin, 1947]. Also, it should be noticed that our measure is a distribution which takes into account the uncertainty of the surrogate modeling. Further, we propose a numerical method to estimate our new defined measures. The procedure is experimented on several numerical examples. Furthermore, we study the asymptotic properties of the estimated measures.

The paper is organized as follows. In Section 2, we introduce the first definitions and the main features of a GP. We also study a special case of covariance kernels. They will be used all along the article as the referenced kernels because they have good properties for sensitivity analysis. In Section 3, we first remind the ANOVA decomposition proposed by Durrande *et al.* [Durrande et al., 2013]. This expansion is done on the posterior mean of a GP. This introduces the decomposition of the conditional Gaussian processes we develop in this paper. After giving the advantages of such approach, we define in Section 4 a sensitivity index well suited for models with dependent inputs. Furthermore, we describe a numerical procedure based on the Monte Carlo estimate to compute our new sensitivity index. At the end of Section 4, we study the convergence properties of the measure. Section 5 is devoted to numerical applications. The goal is to show the relevance of such a sensitivity index through several test cases. Further, we apply our methodology on a real-world problem.

2 Gaussian process regression for sensitivity analysis

In this section, we introduce the notation that will be used along the document. We remind the basic settings on Gaussian process regression. The purpose is to build a fast approximation — also called meta-model — of the input/output relation of the objective function. Then, we present a particular Gaussian process regression which is relevant to perform sensitivity analysis with dependent inputs.

2.1 First definitions

Let $(\Omega_{\mathbf{X}}, \mathcal{A}_{\mathbf{X}}, \mathbb{P}_{\mathbf{X}})$ be a probability space. Let f be a measurable function of a random vector $\mathbf{X} = (X_1, \cdots, X_p) \in \mathbb{R}^p$, $p \geq 1$, and defined as,

$$\begin{array}{rccccc} & (\Omega_{\mathbf{X}}, \mathcal{A}_{\mathbf{X}}, \mathbb{P}_{\mathbf{X}}) & \to & (\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p), P_{\mathbf{X}}) & \to & (\mathbb{R}, \mathcal{B}(\mathbb{R})) \\ f: & \omega & \mapsto & \mathbf{X}(\omega) & \mapsto & f(\mathbf{X}(\omega)), \end{array}$$

where the joint distribution of $\mathbf{X}$ is denoted by $P_{\mathbf{X}}$. Further, we assume that $P_{\mathbf{X}}$ is absolutely continuous with respect to the Lebesgue measure, and that $\mathbf{X}$ admits a density $p_{\mathbf{X}}$ with respect to the Lebesgue measure, i.e. $p_{\mathbf{X}} d\mathbf{x} = dP_{\mathbf{X}}$.
Also, we assume that $f \in L^2_{\mathbb{R}}(\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p), P_{\mathbf{X}})$. We define the expectation with respect to $P_{\mathbf{X}}$ as follows,

$$\mathbb{E}(h(\mathbf{X})) = \int_{\mathbb{R}^p} h(\mathbf{x}) p_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}, \quad h \in L^2_{\mathbb{R}}(\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p), P_{\mathbf{X}}).$$

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 with respect to the inputs distribution $P_{\mathbf{X}}$.
The collection of all subsets of $\{1, \ldots, p\} \backslash \{\emptyset\}$ is denoted by S. For $u \in S$ with $u = (u_1, \cdots, u_t)$, $|u| = t \geq 1$, the random subvector $\mathbf{X}_u$ of $\mathbf{X}$ is defined as $\mathbf{X}_u := (X_{u_1}, \cdots, X_{u_t})$. The marginal density of $\mathbf{X}_u$ is denoted by $p_{\mathbf{X}_u}$.

2.2 Introduction to Gaussian process regression

For $\mathbf{x} = (x_1, \cdots, x_p) \in \mathbb{R}^p$, we consider that the prior knowledge about $f(\mathbf{x})$ can be modeled by a zero-mean Gaussian Process (GP) $Z(\mathbf{x})$ defined on a probability space $(\Omega_Z, \mathcal{A}_Z, \mathbb{P}_Z)$ plus a known mean $m(\mathbf{x})$,

$$f(\mathbf{x}) = m(\mathbf{x}) + Z(\mathbf{x}).$$

From now, we denote by $\mathbb{E}_Z$, V_Z and Cov_Z the expectation, variance and covariance with respect to $\mathbb{P}_Z$. A GP is completely specify by its mean $\mathbb{E}_Z[Z(\mathbf{x})]$ and its covariance kernel $k(\mathbf{x}, \tilde{\mathbf{x}}) = \mathrm{Cov}_Z(Z(\mathbf{x}), Z(\tilde{\mathbf{x}}))$. Here, we consider a zero-mean GP, that can be written as

$$Z(\mathbf{x}) \sim \mathrm{GP}(0, k(\mathbf{x}, \tilde{\mathbf{x}})). \tag{1}$$

Further, we denote by $\mathbf{D} = \{\mathbf{x}^1, \ldots, \mathbf{x}^n\}$, with $\mathbf{x}^j \in \mathbb{R}^p$ for $j = 1, \cdots, n$, the n-sample of observed inputs. We write the vector of centered outputs $\mathbf{z}_n := {}^t(f(\mathbf{x}^1) - m(\mathbf{x}^1), \cdots, f(\mathbf{x}^n) - m(\mathbf{x}^n))$. Further, we consider the Gaussian random vector $\boldsymbol{Z}_n = (Z(\mathbf{x}^1), \cdots, Z(\mathbf{x}^n))$. Notice that $(\mathbf{x}^j)_{j=1,\ldots,n}$ are generally not sampled from the distribution $P_{\mathbf{X}}$. Indeed, they usually come from a space-filling design procedure [Fang et al., 2006] in order to obtain good prediction accuracy.

In the kriging theory, the aim is to use the known values $\mathbf{z}_n$ of $\boldsymbol{Z}_n$ at points in $\mathbf{D}$ to predict $f(\mathbf{x})$. To perform such prediction, we consider the conditional distribution $[f(\mathbf{x})|\boldsymbol{Z}_n = \mathbf{z}_n]$. Standard results about Gaussian distribution gives that this conditional distribution is given by

$$[f(\mathbf{x})|\boldsymbol{Z}_n = \mathbf{z}_n] = \mathrm{GP}\left(\mu(\mathbf{x}), s^2(\mathbf{x}, \tilde{\mathbf{x}})\right), \tag{2}$$

with

$$\mu(\mathbf{x}) = m(\mathbf{x}) + {}^t\mathbf{k}_n(\mathbf{x})\mathbf{K}_n^{-1}(\mathbf{z}_n - \mathbf{m}_n), \tag{3}$$

and

$$s^2(\mathbf{x}, \tilde{\mathbf{x}}) = k(\mathbf{x}, \tilde{\mathbf{x}}) - {}^t\mathbf{k}_n(\mathbf{x})\mathbf{K}_n^{-1}\mathbf{k}_n(\tilde{\mathbf{x}}), \tag{4}$$

where $\mathbf{k}_n(\mathbf{x}) = [k(\mathbf{x}, \mathbf{x}^j)]_{j=1,\ldots,n}$, $\mathbf{m}_n = {}^t(m(\mathbf{x}^1) \cdots m(\mathbf{x}^n))$ and $\mathbf{K}_n = [k(\mathbf{x}^j, \mathbf{x}^l)]_{j,l=1,\ldots,n}$.

The mean $\mu(\mathbf{x})$ of the predictive distribution $[f(\mathbf{x})|\boldsymbol{Z}_n = \mathbf{z}_n]$ is considered as the meta-model for $f(\mathbf{x})$ and $s^2(\mathbf{x}, \mathbf{x})$ represents its mean squared error. An important property of Gaussian process regression is that the mean $\mu(\mathbf{x})$ interpolates the observations $\mathbf{z}_n$ and the variance $s^2(\mathbf{x}, \mathbf{x})$ equals zero at points in $\mathbf{D}$.

2.3 Covariance kernel for sensitivity analysis

Certainly one of the most important points of Gaussian process regression is the choice of the covariance kernel $k(\mathbf{x}, \tilde{\mathbf{x}})$, for $\mathbf{x} = (x_1, \ldots, x_p)$, $\tilde{\mathbf{x}} = (\tilde{x}_1, \ldots, \tilde{x}_p) \in \mathbb{R}^p$, of the unconditioned Gaussian process $Z(\mathbf{x})$ modeling the residual $f(\mathbf{x}) - m(\mathbf{x})$. We note that $k(\mathbf{x}, \tilde{\mathbf{x}})$ must be positive definite and we consider here that $\sup_{\mathbf{x},\tilde{\mathbf{x}}\in\mathbb{R}^p} k(\mathbf{x}, \tilde{\mathbf{x}}) < \infty$. We choose in this paper a relevant class of kernels for performing sensitivity analysis. They are built from Proposition 1 [Durrande et al., 2013].

Proposition 1 *Let us consider a covariance kernel $\tilde{k}(\mathbf{x}, \tilde{\mathbf{x}})$, $\mathbf{x}, \tilde{\mathbf{x}} \in \mathbb{R}^p$, such that $\tilde{k}_{\mathbf{x}} : \mathbf{x} \mapsto \tilde{k}(\mathbf{x}, \tilde{\mathbf{x}})$ is in $L_1(\mathbb{R}^p)$ for all $\mathbf{x} \in \mathbb{R}^p$ and $\tilde{k} : (\mathbf{x}, \tilde{\mathbf{x}}) \mapsto \tilde{k}(\mathbf{x}, \tilde{\mathbf{x}})$ is in $L_1(\mathbb{R}^p \times \mathbb{R}^p)$. Then the following kernel $k(\mathbf{x}, \tilde{\mathbf{x}})$ is a covariance kernel:*

$$k(\mathbf{x}, \tilde{\mathbf{x}}) = \tilde{k}(\mathbf{x}, \tilde{\mathbf{x}}) - \frac{\int \tilde{k}(\mathbf{x}, \mathbf{w}) p_{\mathbf{X}}(d\mathbf{w}) \int \tilde{k}(\tilde{\mathbf{w}}, \tilde{\mathbf{x}}) p_{\mathbf{X}}(d\tilde{\mathbf{w}})}{\iint \tilde{k}(\mathbf{w}, \tilde{\mathbf{w}}) p_{\mathbf{X}}(d\mathbf{w}) p_{\mathbf{X}}(d\tilde{\mathbf{w}})}. \tag{5}$$

Furthermore, if we consider a Gaussian process $Z(\mathbf{x}) \sim \mathrm{GP}(0, k(\mathbf{x}, \tilde{\mathbf{x}}))$, then we have the following equality almost surely,

$$\int Z(\mathbf{x}) p_{\mathbf{X}}(d\mathbf{x}) = 0.$$

From now and until the end of the article, we are interested by the following covariance kernel:

$$k(\mathbf{x}, \tilde{\mathbf{x}}) = \sigma^2 \prod_{i=1}^{p} \left(1 + k_0^i(x_i, \tilde{x}_i)\right), \tag{6}$$

where, following Proposition 1, for all $i = 1, \ldots, p$, we set:

$$k_0^i(x_i, \tilde{x}_i) = \tilde{k}^i(x_i, \tilde{x}_i) - \frac{\int \tilde{k}^i(x_i, w) p_{X_i}(w) dw \int \tilde{k}^i(v, \tilde{x}_i) p_{X_i}(\tilde{w}) d\tilde{w}}{\iint \tilde{k}^i(w, \tilde{w}) p_{X_i}(w) p_{X_i}(\tilde{w}) dw \; d\tilde{w}}, \tag{7}$$

where $(\tilde{k}^i(x_i, \tilde{x}_i))_{i=1,\dots,p}$ are given covariance kernels such that $\tilde{k}^i_w : \tilde{w} \mapsto \tilde{k}^i(w, \tilde{w})$ is in $L_1(\mathbb{R})$ for all $w \in \mathbb{R}$ and $\tilde{k}^i : (w, \tilde{w}) \mapsto \tilde{k}(w, \tilde{w})$ is in $L_1(\mathbb{R} \times \mathbb{R})$. A nice property of $k^i_0(x_i, \tilde{x}_i)$, $i = 1, \dots, p$, is that it is centered with respect to the marginal probability density function p_{X_i}. This feature is going to be exploited in Section 3.
The choice of $\tilde{k}^i(x, \tilde{x})$, $i = 1, \dots, p$, is of importance since it controls the regularity in the i^{th} direction of the Gaussian process $Z(\mathbf{x})$ and thus the smoothness of the meta-model (see [Stein, 1999] and [Rasmussen and Williams, 2006]). For instance, for $m \in \mathbb{N}$, the partial derivative $\partial^m Z(\mathbf{x})/\partial^m x_i$ exists in mean square sense if and only if the $2m$ derivative of $\tilde{k}^i_0(x_i, \tilde{x}_i)$ exists at point $x_i = \tilde{x}_i$. Examples of such covariance kernels $\tilde{k}^i$ are given in Section 5. For each of them, we will also provide the analytical expression of k^i_0.

Further, we also consider that the objective function can be rewritten as

$$f(\mathbf{x}) = f_0 + Z(\mathbf{x}),$$

where f_0 is the constant mean of $f(\mathbf{x})$ and $Z(\mathbf{x})$ is defined as (1), where the covariance kernel $k(\mathbf{x}, \tilde{\mathbf{x}})$ is given by (6). The choice of $k(\mathbf{x}, \tilde{\mathbf{x}})$ is relevant here to propose a decomposition. Indeed, this definition implies the following properties:

1. If we set $Z_0 \sim \mathcal{N}\left(0, \sigma^2\right)$, and, for all $u \in S$, $Z_u(\mathbf{x}_u) \sim \mathrm{GP}(0, \sigma^2 \prod_{i \in u} k^i_0(x_i, \tilde{x}_i))$ are independent processes, then, if

$$Z(\mathbf{x}) = Z_0 + \sum_{u \in S} Z_u(\mathbf{x}_u), \tag{8}$$

   we have that $Z(x) \sim \mathrm{GP}(0, k(\mathbf{x}, \tilde{\mathbf{x}}))$. Indeed, as done in [Durrande et al., 2013], $k(\mathbf{x}, \tilde{\mathbf{x}})$ can be decomposed as it follows,

$$k(\mathbf{x}, \tilde{\mathbf{x}}) = \sigma^2 + \sigma^2 \sum_{u \in S} \prod_{i \in u} k^i_0(x_i, \tilde{x}_i).$$

2. Let us consider two sets $u, v \in S$ such that $u \neq v$, and $Z_u(\mathbf{x}_u) \sim \mathrm{GP}(0, \sigma^2 \prod_{i \in u} k^i_0(x_i, \tilde{x}_i))$, $Z_v(\mathbf{x}_v) \sim \mathrm{GP}(0, \sigma^2 \prod_{i \in v} k^i_0(x_i, \tilde{x}_i))$. We have the following equalities almost surely:

$$\int Z_u(\mathbf{x}_u) \left(\prod_{i=1}^{p} p_{X_i}(x_i) \right) d\mathbf{x} = 0, \tag{9}$$

   and

$$\int Z_u(\mathbf{x}_u) Z_v(\mathbf{x}_v) \left(\prod_{i=1}^{p} p_{X_i}(x_i) \right) d\mathbf{x} = 0. \tag{10}$$

   Using Proposition 1, we know that the linear transformation $\int Z_u(\mathbf{x}_u) \left(\prod_{i=1}^{p} p_{X_i}(x_i)\right) d\mathbf{x}$ is Gaussian, and

$$\int Z_u(\mathbf{x}_u) \left(\prod_{i=1}^{p} p_{X_i}(x_i) \right) d\mathbf{x} \sim \mathcal{N}\left(0, \sigma^2 \int \prod_{i \in u} k^i_0(x_i, \tilde{x}_i) \prod_{i \in u} p_{X_i}(x_i) d\mathbf{x} \right).$$

By replacing k_0^i by its expression (7), we deduce (9). Now, for $u \neq v$, and $i \in u \setminus v$, we have,

$$\int Z_u(\mathbf{x}_u) Z_v(\mathbf{x}_v) \left(\prod_{i=1}^{p} p_{X_i}(x_i) \right) d\mathbf{x} = \int Z_v(\mathbf{x}_v) \left(\int Z_u(\mathbf{x}_u) p_{X_i}(x_i) dx_i \right) p_{\mathbf{x}_{i^c}} \, d\mathbf{x}_{i^c},$$

where $\mathbf{x}_{i^c}$ is the complementary set of x_i, i.e. $\mathbf{x}_{i^c} = (x_j)_{j \neq i}$. Again with Proposition 1, we conclude that (10) is satisfied.

Discussion about the choice of the covariance kernel when the input parameters are independent. The kernel given in (6) provides a relevant prior for the sensitivity analysis when $P_{\mathbf{X}} = \otimes_{i=1}^{p} P_{X_i}$. In this case, the Sobol index [Sobol, 1993] of $Z(\mathbf{x})$ — modeling our prior knowledge about $f(\mathbf{x})$ — is given by:

$$S_u = \frac{V[Z_u(\mathbf{X}_u)]}{V[Z(\mathbf{X})]}, \quad \forall \ u \in S,$$

where Z_u checks Equalities (9) and (10).
By setting $\sigma_u^2 = \sigma^2 \prod_{i \in u} \left(1 + k_0^i(x_i, x_i)\right)$ and $\sigma_v^2 = \sigma^2 \prod_{i \in v} \left(1 + k_0^i(x_i, x_i)\right)$ where $u \neq v \in S$, we have:

$$\mathbb{E}_Z \left[V(Z_u(\mathbf{X}_u) \right] = \int \sigma_u^2 \prod_{i \in u} p_{x_i}(dx_i)$$

and

$$\mathrm{Cov}_Z \left(V[Z_u(\mathbf{X}_u)], V[Z_v(\mathbf{X}_v)] \right) = 0.$$

Therefore, we notice the sensitivity of $\mathbf{X}_u$ in the model is monitored by σ_u^2, always strictly positive. This means that, through the decomposition (8), we consider *a priori* that every group of input variables is contributive in the model. We also consider *a priori* that the sensitivity indices are uncorrelated.

3 ANOVA decomposition of conditional Gaussian processes

We propose in this section a representation of $f(\mathbf{x})$ as a sum of increasing dimension Gaussian processes. Our main contribution is to consider the complete predictive distribution $[f(\mathbf{x}) | \boldsymbol{Z}_n = \mathbf{z}_n]$, given by (2), and not only the predictive mean $\mu(\mathbf{x})$. This allows us for quantifying the uncertainty due to the meta-modeling on the sensitivity indices estimation.

3.1 ANOVA decomposition of the predictive mean

This paragraph is dedicated to the functional decomposition of the predictive mean $\mu(\mathbf{x})$. Although it has been already developed and studied in [Durrande et al., 2013], we remind it here for the good understanding of the extension proposed in Paragraph 3.2.

Remind that we consider the prior knowledge $Z(\mathbf{x}) \sim \mathrm{GP}(0, k(\mathbf{x}, \tilde{\mathbf{x}}))$, where $k(\mathbf{x}, \tilde{\mathbf{x}})$ is defined by Equations (6)-(7). Following Paragraph 2.2, we consider the predictive distribution

$$[f(\mathbf{x}) | \boldsymbol{Z}_n = \mathbf{z}_n] \sim \mathrm{GP}(\mu(\mathbf{x}), s^2(\mathbf{x}, \tilde{\mathbf{x}})),$$

where, from the definition of $k(\mathbf{x}, \tilde{\mathbf{x}})$, we can decompose $\mu(\mathbf{x})$ as follows,

$$\mu(\mathbf{x}) = \mu_0 + \sum_{u \in S} \mu_u(\mathbf{x}_u), \tag{11}$$

with $\mu_0 = f_0 + {}^t\mathbf{1}_n \mathbf{K}_n^{-1}(\mathbf{z}_n - f_0\mathbf{1}_n)$ and

$$\mu_u(\mathbf{x}_u) = \prod_{i \in u} {}^t\mathbf{k}_{0,n}^i(\mathbf{x}) \mathbf{K}_n^{-1}(\mathbf{z}_n - f_0\mathbf{1}_n), \quad \forall\ u \in S. \tag{12}$$

$\mathbf{1}_n$ the n-vector of 1, $\mathbf{k}_n(\mathbf{x}) = [k(\mathbf{x}, \mathbf{x}^j)]_{j=1,\dots,n}$, $\mathbf{k}_{0,n}^i(x_i) = [k_0^i(x_i, x_i^j)]_{j=1,\dots,n}$ (see Equation (7)) and $\mathbf{K}_n = [k(\mathbf{x}^j, \mathbf{x}^l)]_{j,l=1,\dots,n}$.
Thanks to the property of the kernel k_0^i, we can deduce that, for all $u, v \in S$ with $u \neq v$:

$$\int \mu_u(\mathbf{x}_u) \left(\prod_{i=1}^{p} p_{X_i}(x_i) \right) d\mathbf{x} = 0,$$

and

$$\int \mu_u(\mathbf{x}_u)\mu_v(\mathbf{x}_v) \left(\prod_{i=1}^{p} p_{X_i}(x_i) \right) d\mathbf{x} = 0.$$

From this decomposition, [Durrande et al., 2013] deduce an analytical sensitivity measure, that we call S_u^{D} here, to quantify the contribution of a given $\mathbf{X}_u$ in the model. It is defined as,

$$S_u^{\mathrm{D}} = \frac{V[\mu_u(\mathbf{X}_u)]}{V[\mu(\mathbf{X})]}, \tag{13}$$

with components $(\mu_u)_{u \in S}$ having the same properties as the ones of the Hoeffding expansion [Hoeffding, 1948]. Thus, we can analyse the effect of each group of variables on the global variability, when the initial model f is substituted to the predictive mean μ.

However, when we only consider the predictive mean, we neglect an important part of information contained in the posterior variance. In addition, if the uncertainty of $\mu(\mathbf{x})$ is important, this means that the surrogate model does not ajust properly the objective function, leading to a wrong sensitivity analysis.
In the following part, we take into account the uncertainty of the meta-modeling by defining a functional ANOVA decomposition of a predictive distribution. Inspired by the work of [Durrande et al., 2013], we extend their work to a more general decomposition, and we also extend the sensitivity indices defined by (13) to the definition of sensitivity indices when input variables can be non independent.

3.2 ANOVA decomposition of the conditional Gaussian processes

We saw in the previous paragraph that the considered covariance kernel $k(\mathbf{x}, \tilde{\mathbf{x}})$ (6) leads to an ANOVA decomposition (11) of $\mu(\mathbf{x})$ which is suitable to perform sensitivity analysis. However, as emphasized by [Oakley and O'Hagan, 2004], performing a sensitivity analysis based on a Gaussian process regression using only the predictive mean can be inappropriate. Indeed, in the framework of computer experiments, few observations $\mathbf{z}_n + f_0$ of $f(\mathbf{x})$ are available and

thus the uncertainty on the meta-model $\mu(\mathbf{x})$ can be non-negligible. For this reason, it is worth taking into account the uncertainty on the meta-model and to consider the complete predictive distribution. Nevertheless, it is non trivial to find the analogous of the ANOVA decomposition of the predictive mean to the predictive distribution. The proposition below allows for handling this issue [Chilès and Delfiner, 1999].

Proposition 2 *Let consider the random process $f^n(\mathbf{x})$ defined as*

$$f^n(\mathbf{x}) = \mu(\mathbf{x}) - {}^t\mathbf{k}_n(\mathbf{x})\mathbf{K}_n^{-1}\boldsymbol{Z}_n + Z(\mathbf{x}), \tag{14}$$

where $Z(\mathbf{x}) \sim GP(0, k(\mathbf{x}, \tilde{\mathbf{x}}))$ with $k(\mathbf{x}, \tilde{\mathbf{x}}) = \sigma^2 \prod_{i=1}^p \left(1 + k_0^i(x_i, \tilde{x_i})\right)$, $\mu(\mathbf{x})$ is the predictive mean defined by (11), $\boldsymbol{Z}_n = {}^t(Z(\mathbf{x}^1), \ldots, Z(\mathbf{x}^n))$ is the Gaussian random vector corresponding to the value of $Z(x)$ at points in the experimental design set $\mathbf{D} = \{\mathbf{x}^1, \ldots, \mathbf{x}^n\}$, $\mathbf{k}_n(\mathbf{x}) = [k(\mathbf{x}, \mathbf{x}^j)]_{j=1,\ldots,n}$ and $\mathbf{K}_n = [k(\mathbf{x}^i, \mathbf{x}^j)]_{i,j=1,\ldots,n}$. Then, we have

$$f^n(\mathbf{x}) \sim [f(\mathbf{x}) | \boldsymbol{Z}_n = \mathbf{z}_n]. \tag{15}$$

To get the proof of Proposition 2, the reader could refer to [Chilès and Delfiner, 1999]. Here, this result is of great interest since it allows for defining a Gaussian process $f^n(\mathbf{x})$ distributed with respect to the predictive distribution $[f(\mathbf{x}) | \boldsymbol{Z}_n = \mathbf{z}_n]$. Our goal is now to find a decomposition for $f^n(\mathbf{x})$ which is suitable for performing sensitivity analysis. Following the decompositions of $\mu(\mathbf{x})$ given in (11)-(12) and of $Z(\mathbf{x})$ given in (8), the decomposition of f^n is given in Proposition 3.

Proposition 3 *Let $f^n(\mathbf{x})$ be the random process defined by (14) of Proposition 2. Then,*

$$f^n(\mathbf{x}) = f_0^n + \sum_{u \in S} f_u^n(\mathbf{x}_u), \tag{16}$$

with

$$\begin{cases} f_0^n = \mu_0 - {}^t\mathbf{1}_n\mathbf{K}_n^{-1}\boldsymbol{Z}_n + Z_0, \\ \\ f_u^n(\mathbf{x}_u) = \mu_u(\mathbf{x}_u) - \prod_{i \in u} {}^t\mathbf{k}_{0,n}^i(\mathbf{x})\mathbf{K}_n^{-1}\boldsymbol{Z}_n + Z_u(\mathbf{x}_u), \quad \forall u \in S. \end{cases} \tag{17}$$

Furthermore, since k_0^i defined in Equation (7) is centered with respect to the marginal density p_{X_i}, the following properties holds almost surely for all $u, v \in S$, $u \neq v$:

$$\int f_u^n(\mathbf{x}_u) \left(\prod_{i=1}^p p_{X_i}(x_i) \right) d\mathbf{x} = 0, \tag{18}$$

and

$$\int f_u^n(\mathbf{x}_u) f_v^n(\mathbf{x}_v) \left(\prod_{i=1}^p p_{X_i}(x_i) \right) d\mathbf{x} = 0. \tag{19}$$

The proof of proposition 3 is straightforward, by the decomposition of $\mu(\mathbf{x})$ given by (11)-(12), by the decomposition of $Z(\mathbf{x})$ given by (8), and by the definition of $k(\mathbf{x}, \tilde{\mathbf{x}})$ given by (6). The properties of the summands $(f_u^n)_u$ are also immediate.

Remark: The suggested decomposition (16) allows for taking into account the meta-model uncertainty in the sensitivity index estimates. Furthermore, we highlight that it is easy to sample with respect to the distribution of $f^n(x)$. Indeed, from Proposition 3 we deduce that we can perform it by sampling with respect to the distribution of $Z(x) \sim \mathrm{GP}(0, k(\mathbf{x}, \tilde{\mathbf{x}}))$ and applying the linear transformation presented in (14). Moreover, to obtain a sample of $f_u^n(\mathbf{x}_u)$ we just have to sample $Z_u(\mathbf{x}_u) \sim \mathrm{GP}\left(0, \prod_{i \in u} k_0^i(x_i, \tilde{x}_i)\right)$.

4 Sensitivity measure definition for dependent input variables

Now, we adapt the methodology of [Li et al., 2010] to construct sensitivity indices for models with dependent inputs. First, we define a sensitivity measure based on the result of Proposition 3. Then, we present an estimation procedure of the sensitivity measure in Paragraph 4.2. Finally, we present in Paragraph 4.3 how to take into account the uncertainty of the index estimation.

4.1 Sensitivity measure definition

As presented in Section 3, the model function $f(\mathbf{x})$ is substituted to $f^n(\mathbf{x}) = f_0^n + \sum_{u \in S} f_u^n(\mathbf{x}_u))$. Therefore, the global variance can now be decomposed as

$$V(f^n(\mathbf{X})) = \sum_{u \in S} \left[V(f_u^n(\mathbf{X}_u)) + \mathrm{Cov}(f_u^n(\mathbf{X}_u), f_{u^c}^n(\mathbf{X}_{u^c})) \right] \tag{20}$$

where $f_{u^c}^n(\mathbf{X}) = f^n(\mathbf{X}) - f_u^n(\mathbf{X}_u)$. In this way, the sensitivity index associated to the group of variables $\mathbf{X}_u$ is given by

$$S_u^f = \frac{V[f_u^n(\mathbf{X}_u)] + \mathrm{Cov}[f_u^n(\mathbf{X}_u), f_{u^c}^n(\mathbf{X})]}{V[f^n(\mathbf{X})]} \tag{21}$$

We note that S_u^f is defined on the probability space $(\Omega_Z, \mathcal{A}_Z, \mathbb{P}_Z)$ as V and Cov are the variance and covariance with respect to $P_{\mathbf{X}}$. Therefore, S_u^f integrates the uncertainty related to the meta-model approximation. In practice, the mode or the mean of the distribution S_u^f may be used to get a scalar measure of sensitivity.

It also should be noted that this index is analogous with the Sobol one in the independent case when f is replaced by f^n. Indeed, by (18) and (19), we have

1. For $u \neq v \in S$,
$$\mathrm{Cov}\left(f_u^n(\mathbf{X}_u), f_v^n(\mathbf{X}_v)\right) = 0.$$

2. For a given $u \in S$, by integrating f^n with respect to the distribution of the all inputs except the ones indexed by u, we have that
$$f_u^n(\mathbf{X}_u) = \mathbb{E}\left[f^n(\mathbf{X}) | \mathbf{X}_u\right] + \sum_{\substack{v \subset u \\ v \neq u}} (-1)^{|u|-|v|} \mathbb{E}\left[f^n(\mathbf{X}) | \mathbf{X}_v\right].$$

Hence, when $P_{\mathbf{X}} = P_{X_1} \otimes \cdots \otimes P_{X_p}$,

$$S_u^f = \frac{V(\mathbb{E}\left[f^n(\mathbf{X})|\mathbf{X}_u\right]) + \sum_{\substack{v \subset u \\ v \neq u}} (-1)^{|u|-|v|} V(\mathbb{E}\left[f^n(\mathbf{X})|\mathbf{X}_v\right])}{V[f^n(\mathbf{X})]}.$$

For models with dependent inputs, the parametric functional ANOVA decomposition mentioned in Section 1 [Li et al., 2010, Caniou, 2012, Chastaing et al., 2013] requires to estimate 2^p components, which is hardly achievable in practice when p gets large. Thus, their decomposition must be truncated but in this case we loose a part of model information.
Here, we have accessed to the approximation $f^n(\mathbf{x})$ of $f(\mathbf{x})$ without processing all the terms $f_u^n(\mathbf{x}_u)$, $u \in S$, thanks to the equality $f^n(\mathbf{x}) = \mu(\mathbf{x}) - {}^t\mathbf{k}_n(\mathbf{x})\mathbf{K}_n^{-1}\boldsymbol{Z}_n + Z(\mathbf{x})$ where $Z(\mathbf{x}) \sim \mathrm{GP}(0, k(\mathbf{x}, \tilde{\mathbf{x}}))$, and $\mu(\mathbf{x}) = f_0 + {}^t\mathbf{k}_n(\mathbf{x})\mathbf{K}_n^{-1}(\mathbf{z}_n - f_0\mathbf{1}_n)$. Furthermore, for any $u \in S$, we have an explicit expression of $f_u^n(\mathbf{x}_u)$ given by (17). Thus, as $f^n = f_u^n + f_{u^c}^n$, the complementary summand $f_{u^c}^n$ of f_u^n can be easily deduced, avoiding the truncation error in the estimation.

4.2 Estimation procedure

The aim of this section is to provide an efficient numerical estimation of the sensitivity measure S_u^f, for a given $u \in S$. As already mentioned, S_u^f lies in $(\Omega_Z, \mathcal{A}_Z, \mathbb{P}_Z)$. Further, we describe a numerical procedure for only one realization S_u^f. In practice, this procedure is repeated N_s times to take into account the uncertainty of the meta-model.

We empirically estimate the variance and the covariance involved in (21) with a Monte-Carlo integration. Therefore, we consider the following estimator from m realizations $\mathbf{T} = (\mathbf{t}_1^j, \cdots, \mathbf{t}_p^j)_{j=1,\ldots,m}$ of the random variable $\mathbf{X}$ defined on the probability space $(\Omega_{\mathbf{X}}, \mathcal{A}_{\mathbf{X}}, \mathbb{P}_{\mathbf{X}})$:

$$S_{u,m}^f = \frac{\frac{1}{m}\sum_{j=1}^m f_u^n(\mathbf{t}_u^j)^2 - \left(\bar{f}_u^n\right)^2 + \frac{1}{m}\sum_{j=1}^m f_u^n(\mathbf{t}_u^j) f_{u^c}^n(\mathbf{t}^j) - \bar{f}_u^n \bar{f}_{u^c}^n}{\frac{1}{m}\sum_{j=1}^m f^n(\mathbf{t}^j))^2 - \left(\frac{1}{m}\sum_{j=1}^m f^n(\mathbf{t}^j)\right)^2} \tag{22}$$

where $\bar{f}_u^n = \frac{1}{m}\sum_{j=1}^m f_u^n(\mathbf{t}_u^j)$, $\bar{f}_{u^c}^n = \frac{1}{m}\sum_{j=1}^m f_{u^c}^n(\mathbf{t}^j)$ and $\mathbf{t}_u^j = (\mathbf{t}_i^j)_{i\in u}$, $j = 1, \ldots, m$. We point out that $S_{u,m}^f$ lies in the product probability space $(\Omega_Z, \mathcal{A}_Z, \mathbb{P}_Z)$. Therefore, we generate several realizations of $S_{u,m}^f$ to get an estimate of it. This procedure is described further below.

First, let us denote $\mathbf{M} = \begin{pmatrix} \mathbf{T} \\ \mathbf{D} \end{pmatrix}$, where $\mathbf{D} = (\mathbf{x}^j)_{j=1,\ldots,n}$ is the experimental design set. For $u \in S$, we denote $\mathbf{M}_u = \begin{pmatrix} \mathbf{T}_u \\ \mathbf{D}_u \end{pmatrix}$, where $\mathbf{T}_u = (\mathbf{t}_u^j)_{j=1,\ldots,m}$ and $\mathbf{D}_u = (\mathbf{x}_u^j)_{j=1,\ldots,n}$.

1. To get a realization of $f_u^n(\mathbf{x}_u)$, we generate a sample from the distribution of

$$Z_u(\mathbf{x}_u) \sim \mathrm{GP}\left(0, \prod_{i\in u} k_0^i(x_i, \tilde{x}_i)\right),$$

on $\mathbf{M}$ with the following procedure:

   (a) Compute $\mathbf{K}_{0,m}^u = \bigodot_{i\in u} k_0^i(\mathbf{M}_i, \mathbf{M}_i)$ and the Cholesky decomposition $\mathbf{L}_{0,m}^u$ of $\mathbf{K}_{0,m}^u$ where $\bigodot$ stands for the term-wise matrix product. $\mathbf{K}_{0,m}^u$ is the covariance matrix of $Z_u(\mathbf{x}_u)$ at points in $\mathbf{M}_u$.

(b) Generate one realization $z_u(\mathbf{M}_u)$ of $Z_u(\mathbf{x}_u)$ on $\mathbf{M}_u$ from the Cholesky decomposition of $\mathbf{K}^u_{0,m}$ (see [Rasmussen and Williams, 2006] Appendix A.2) with,

$$z_u(\mathbf{M}_u) = \mathbf{L}^u_{0,m}\boldsymbol{\varepsilon}_u, \tag{23}$$

where $\boldsymbol{\varepsilon}_u$ is a sample generated from the distribution $\mathcal{N}(0, I_{n+m})$ where I is the identity matrix of size $(n+m)\times(n+m)$.

2. To generate a realization of $f_{u^c}(\mathbf{x})$ on $\mathbf{M}$, we generate a sample from the distribution of

$$Z_{u^c}(\mathbf{x}) \sim \mathrm{GP}\left(0, k(\mathbf{x},\tilde{\mathbf{x}}) - \prod_{i\in u} k_0^i(x_i,\tilde{x}_i)\right),$$

on $\mathbf{M}$, with the following procedure:

(a) Compute $\mathbf{K}_m = k(\mathbf{M},\mathbf{M})$ and the Cholesky decomposition $\mathbf{L}^{u^c}_{0,m}$ of $\mathbf{K}_m - \mathbf{K}^u_{0,m}$. $k(\mathbf{M},\mathbf{M})$ is the covariance matrix of $Z(\mathbf{x})$ at points in $\mathbf{M}$.

(b) Generate one realization $z_{u^c}(\mathbf{M})$ of $Z_{u^c}(\mathbf{x})$ on $\mathbf{M}$ with

$$z_{u^c}(\mathbf{M}) = \mathbf{L}^{u^c}_{0,m}\boldsymbol{\varepsilon}_{u^c}, \tag{24}$$

where $\boldsymbol{\varepsilon}_{u^c}$ is sampled from the distribution $\mathcal{N}(0, I_{n+m})$.

3. We deduce a sample $z(\mathbf{M})$ of $Z(\mathbf{x})$ on $\mathbf{M}$ with:

$$z(\mathbf{M}) = z_u(\mathbf{M}_u) + z_{u^c}(\mathbf{M}). \tag{25}$$

Moreover, as $M = \begin{pmatrix}\mathbf{T}\\ \mathbf{D}\end{pmatrix}$ and $M_u = \begin{pmatrix}\mathbf{T}_u\\ \mathbf{D}_u\end{pmatrix}$, $z(\mathbf{M})$, $z_u(\mathbf{M}_u)$ and $z_{u^c}(\mathbf{M})$ can be rewritten in the following forms:

$$z(\mathbf{M}) = \begin{pmatrix}z(\mathbf{T})\\ z(\mathbf{D})\end{pmatrix},\ z_u(\mathbf{M}_u) = \begin{pmatrix}z_u(\mathbf{T}_u)\\ z_u(\mathbf{D}_u)\end{pmatrix} \text{ and } z_{u^c}(\mathbf{M}) = \begin{pmatrix}z_{u^c}(\mathbf{T})\\ z_{u^c}(\mathbf{D})\end{pmatrix}.$$

4. We can deduce the samples $\tilde{f}^n(\mathbf{T})$, $\tilde{f}^n_u(\mathbf{T}_u)$ and $\tilde{f}^n_{u^c}(\mathbf{T})$ of respectively $f^n(\mathbf{x})$, $f^n_u(\mathbf{x}_u)$ and $f^n_{u^c}(\mathbf{x})$ on $\mathbf{T}$ with the following formulas:

$$\begin{cases} \tilde{f}^n(\mathbf{T}) = \mu(\mathbf{T}) - k(\mathbf{T},\mathbf{D})\mathbf{K}_n^{-1}z(\mathbf{D}) + z(\mathbf{T}),\\ \tilde{f}^n_u(\mathbf{T}_u) = \mu_u(\mathbf{T}_u) - \left(\bigodot_{i\in u} k_0^i(\mathbf{T}_i,\mathbf{D}_i)\right)\mathbf{K}_n^{-1}z(\mathbf{D}) + z_u(\mathbf{T}_u),\\ \tilde{f}^n_{u^c}(\mathbf{T}) = \tilde{f}^n(\mathbf{T}) - \tilde{f}^n_u(\mathbf{T}) \end{cases} \tag{26}$$

where $\mathbf{K}_n = k(\mathbf{D},\mathbf{D})$, and

$$\begin{cases} \mu(\mathbf{T}) = {}^t k(\mathbf{T},\mathbf{D})\mathbf{K}_n^{-1}(\mathbf{z}_n - f_0\mathbf{1}_n) + f_0\\ \mu_u(\mathbf{T}_u) = \bigodot_{i\in u} {}^t k_0^i(\mathbf{T}_i,\mathbf{D}_i)\,\mathbf{K}_n^{-1}(\mathbf{z}_n - f_0\mathbf{1}_n) \end{cases}$$

We note that $k(\mathbf{T},\mathbf{D})$ and $\left(\bigodot_{i\in u} k_0^i(\mathbf{T}_i,\mathbf{D}_i)\right)$ have been already computed with $\mathbf{K}_m$ and $\mathbf{K}^u_{0,m}$ (Step 1 and 2) as,

$$\mathbf{K}_m = k(\mathbf{M},\mathbf{M}) = \begin{pmatrix}k(\mathbf{T},\mathbf{T}) & k(\mathbf{T},\mathbf{D})\\ k(\mathbf{D},\mathbf{T}) & k(\mathbf{D},\mathbf{D})\end{pmatrix},$$

and

$$\mathbf{K}^u_{0,m} = \bigodot_{i\in u} k_0^i\left(\mathbf{M}_i, \mathbf{M}_i\right) = \begin{pmatrix} \bigodot_{i\in u} k_0^i\left(\mathbf{T}_i, \mathbf{T}_i\right) & \bigodot_{i\in u} k_0^i\left(\mathbf{T}_i, \mathbf{D}_i\right) \\ \bigodot_{i\in u} k_0^i\left(\mathbf{D}_i, \mathbf{T}_i\right) & \bigodot_{i\in u} k_0^i\left(\mathbf{D}_i, \mathbf{D}_i\right) \end{pmatrix}.$$

5. We deduce that a sample $s^f_{u,m}$ of $S^f_{u,m}$ is given by

$$s^f_{u,m} = \frac{\frac{1}{m}\sum_{j=1}^m \tilde{f}^n_u(\mathbf{t}^j_u)^2 - \left(\bar{\tilde{f}}^n_u\right)^2 + \frac{1}{m}\sum_{j=1}^m \tilde{f}^n_u(\mathbf{t}^j_u)\tilde{f}^n_{u^c}(\mathbf{t}^j) - \bar{\tilde{f}}^n_u\bar{\tilde{f}}^n_{u^c}}{\frac{1}{m}\sum_{j=1}^m \tilde{f}^n(\mathbf{t}^j))^2 - \left(\frac{1}{m}\sum_{j=1}^m \tilde{f}^n(\mathbf{t}^j)\right)^2}, \tag{27}$$

where $\bar{\tilde{f}}^n_u = \frac{1}{m}\sum_{j=1}^m \tilde{f}^n_u(\mathbf{t}^j_u)$ and $\bar{\tilde{f}}^n_{u^c} = \frac{1}{m}\sum_{j=1}^m \tilde{f}^n_{u^c}(\mathbf{t}^j)$.

4.3 Asymptotic normality of the sensitivity index estimator

We have presented in the previous paragraph a procedure to sample $S^f_{u,m}$ defined in Equation (22). However, for given realizations of random processes $f^n_u(\mathbf{X}_u)$ and $f^n_{u^c}(\mathbf{X})$, the estimated sensitivity measure comes from a Monte-Carlo integration and thus integrates a Monte-Carlo error. The purpose of this paragraph is to quantify it. A natural approach is to use an asymptotic normality result as stated in Proposition 4.

Proposition 4 *For $u \in S$, let us consider the respective realizations of $f^n_u(\mathbf{X}_u)$, $f^n_{u^c}(\mathbf{X})$ and $f^n(\mathbf{X})$ denoted by $\tilde{f}^n_u(\mathbf{X}_u)$, $\tilde{f}^n_{u^c}(\mathbf{X})$ and $\tilde{f}^n(\mathbf{X})$ respectively. We denote the theoretical sensitivity measure for $\mathbf{X}_u$ associated to $\tilde{f}^n$ by*

$$s^f_u = \frac{V(\tilde{f}^n_u(\mathbf{X}_u)) + \mathrm{Cov}(\tilde{f}^n_u(\mathbf{X}_u), \tilde{f}^n_{u^c}(\mathbf{X}))}{V(\tilde{f}^n(\mathbf{X}))},$$

where we assume that $V(\tilde{f}^n(\mathbf{X})) \neq 0$. Suppose also that $\mathbb{E}\left[\tilde{f}^n_u(\mathbf{X})^4\right] < \infty$ for all $u \in S$. Then, for any $u \in S$, we have, when $m \to \infty$:

$$\sqrt{m}\left(s^f_{u,m} - s^f_u\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, {}^t(\bigtriangledown \boldsymbol{\phi}(\boldsymbol{\mu}))\boldsymbol{\Gamma} \bigtriangledown \boldsymbol{\phi}(\boldsymbol{\mu})\right) \tag{28}$$

where $\boldsymbol{\mu} = \mathbb{E}(\mathbf{U})$, $\boldsymbol{\Gamma} = V(\mathbf{U})$,

$$\mathbf{U} = \begin{pmatrix} \tilde{f}^n_u(\mathbf{X}_u) & \tilde{f}^n_{u^c}(\mathbf{X}) & \tilde{f}^n(\mathbf{X}) & \tilde{f}^n_u(\mathbf{X}_u)^2 & \tilde{f}^n(\mathbf{X})^2 & \tilde{f}^n_u(\mathbf{X}_u)\tilde{f}^n_{u^c}(\mathbf{X}) \end{pmatrix},$$

and

$$\boldsymbol{\phi}(u_1, u_2, u_3, u_4, u_5, u_6) = \frac{u_4 - u_1^2 + u_6 - u_1u_2}{u_5 - u_3^2}.$$

The proof of Proposition 4 is straightforward by using the Delta method (see [Van Der Vaart, 1998]). We highlight that the terms $\boldsymbol{\mu}$ and $\boldsymbol{\Gamma}$ in Proposition 4 can be estimated from the sample $\mathbf{T} = (\mathbf{t}^j)_{j=1,\ldots,m}$ used in the Monte-Carlo integration (22).

In practice, we use the asymptotic result given in Proposition 4 to estimate the Monte Carlo error. Thus, to take into account both the uncertainty of the surrogate model and the one of the Monte Carlo integration, we proceed as follows,

1. Generate $\tilde{f}^n(\mathbf{T})$, $\tilde{f}^n_u(\mathbf{T}_u)$ and $\tilde{f}^n_{u^c}(\mathbf{T})$ from the sample $\mathbf{T}$ with the estimation procedure given in Paragraph 4.2.

2. Generate a sample of size K from the limit distribution $\mathcal{N}\left(0, {}^t(\bigtriangledown\boldsymbol{\phi}(\hat{\boldsymbol{\mu}}))\hat{\boldsymbol{\Gamma}} \bigtriangledown \boldsymbol{\phi}(\hat{\boldsymbol{\mu}})\right)$ where $\hat{\boldsymbol{\mu}} = \frac{1}{m}\sum_{j=1}^m \mathbf{U}(\mathbf{t}_u^j)$, $\hat{\boldsymbol{\Gamma}} = \frac{1}{m}\sum_{j=1}^m [\mathbf{U}(\mathbf{t}_u^j) - \hat{\boldsymbol{\mu}}]^2$, and

$$\mathbf{U} = \left(\tilde{f}^n_u \quad \tilde{f}^n_{u^c} \quad \tilde{f}^n \quad (\tilde{f}^n_u)^2 \quad (\tilde{f}^n)^2 \quad \tilde{f}^n_u \tilde{f}^n_{u^c}\right).$$

   Thus, a Monte Carlo error is obtained for one realization $s^f_{u,m}$.

3. Repeat Steps 1-2 N_s times to take into account the uncertainty of the surrogate model.

5 Applications

We illustrate in this section our sensitivity measure on academic and industrial examples. First, we present explicit examples of covariance kernels k_0^i, and a procedure to estimate their parameters. Further, we illustrate the estimation procedure of Paragraph 4.2 through several numerical applications.

5.1 Example of covariance kernels

Here, we analytically compute zero mean kernels for two usual kernels associated with uniform distributions. First, let us consider that (see Equation (7)):

$$k_0^i(x,\tilde{x}) = \tilde{k}^i(x,\tilde{x}) - \frac{\int \tilde{k}^i(x,u)p_{X_i}(u)du \int \tilde{k}^i(v,\tilde{x})p_{X_i}(v)dv}{\int\int \tilde{k}^i(u,v)p_{X_i}(u)p_{X_i}(v)dudv}, \quad x,\tilde{x} \in \mathbb{R}.$$

Example 1: We consider an exponential kernel for $\tilde{k}^i(x,\tilde{x})$ with an uniform marginal p_{X_i}, namely,

$$k^i(x,\tilde{x}) = \exp\left(-\frac{1}{2}\frac{|x-\tilde{x}|}{\theta_i}\right), \quad \theta_i > 0,$$

and

$$p_{X_i} \sim \mathcal{U}(a_i, b_i).$$

Then, the corresponding covariance kernel $k_0^i(x,\tilde{x})$ is given by:

$$\begin{aligned} k_0^i(x,\tilde{x}) = \quad & \exp\left(-\tfrac{1}{2}\tfrac{|x-\tilde{x}|}{\theta_i}\right) - \frac{\theta_i}{b_i-a_i+2\theta_i\left(\exp(-\frac{1}{2}\frac{b_i-a_i}{\theta_i})-1\right)} \times \\ & \left(2-\exp(-\tfrac{1}{2}\tfrac{x-a_i}{\theta_i})-\exp(-\tfrac{1}{2}\tfrac{b_i-x}{\theta_i})\right)\cdot\left(2-\exp(-\tfrac{1}{2}\tfrac{\tilde{x}-a_i}{\theta_i})-\exp(-\tfrac{1}{2}\tfrac{b_i-\tilde{x}}{\theta_i})\right) \end{aligned}$$

We note that the exponential kernel is stationary — i.e. it is invariant under translations in the input parameter space — and corresponds to the covariance of an Ornstein-Uhlenbeck process. Furthermore, the corresponding process is continuous in mean square sense and nowhere differentiable. Therefore this kernel is appropriate for rough function $f(\mathbf{x})$.

Example 2: We consider a Gaussian kernel for $\tilde{k}^i(x,\tilde{x})$ with an uniform marginal p_{X_i}, namely,

$$k^i(x,\tilde{x}) = \exp\left(-\frac{1}{2}\frac{(x-\tilde{x})^2}{\theta_i^2}\right), \quad \theta_i > 0,$$

and

$$p_{X_i} \sim \mathcal{U}(a_i, b_i).$$

Then, the corresponding covariance kernel $k_0^i(x,\tilde{x})$ is given by:

$$k_0^i(x,\tilde{x}) = \exp\left(-\tfrac{1}{2}\tfrac{(x-\tilde{x})^2}{\theta_i^2}\right) - A(x)A(\tilde{x})/B,$$

where

$$A(x) = -\frac{\sqrt{\pi}}{\sqrt{2}}\theta_i \mathrm{erf}\left(\frac{a_i - x}{\theta_i\sqrt{2}}\right) + \frac{\sqrt{\pi}}{\sqrt{2}}\theta_i \mathrm{erf}\left(\frac{b_i - x}{\theta_i\sqrt{2}}\right),$$

$$B = -2\theta_i^2 + \theta_i\sqrt{2}\mathrm{erf}\left(\frac{a_i - b_i}{\theta_i\sqrt{2}}\right)\sqrt{\pi}(a_i - b_i) + 2\exp\left(-\frac{1}{2}\frac{(a_i-b_i)^2}{\theta_i^2}\right)\theta_i^2,$$

and the error function is given by

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x \exp(-t^2)dt.$$

We note that the Gaussian kernel corresponds to processes infinitely continuously differentiable in mean square sense. Therefore this kernel is appropriate for very smooth function $f(\mathbf{x})$. Closed form expressions can also be derived for 5/2-Matérn and 3/2-Matérn covariance kernels (see [Stein, 1999]) by straightforward calculations. Due to their complex expression, they are not presented here. Though, we note that these kernels correspond respectively to once and twice continuously differentiable processes in mean square sense. Therefore, they could be a relevant compromise between the exponential and the Gaussian kernels.

5.2 Covariance kernel parameter estimation

We deal in this section with the estimation of the model parameters using a maximum likelihood method. Let us consider the covariance kernel

$$k(\mathbf{x},\tilde{\mathbf{x}}) = \sigma^2 \prod_{i=1}^{p}\left(1 + k_0^i(x_i,\tilde{x}_i)\right),$$

where $k_0^i(x_i,\tilde{x}_i)$ is one of the covariance kernels given in Example 1 or Example 2 of Paragraph 5.1. Therefore, the parameters to be estimated are the variance parameter σ^2, the mean f_0 and the hyper-parameter $\boldsymbol{\theta} = (\theta_i)_{i=1,\ldots,p}$ of $(k_0^i(x_i,\tilde{x}_i))_{i=1,\ldots,p}$.

First, let us consider the maximum likelihood estimate of f_0:

$$\hat{f}_0 = \left({}^t\mathbf{1}_n\mathbf{K}_n^{-1}\mathbf{1}_n\right)^{-1}{}^t\mathbf{1}_n\mathbf{K}_n^{-1}\mathbf{z}_n, \tag{29}$$

where $\mathbf{z}_n = (f(\mathbf{x}^i))_{i=1,\ldots,n}$ and $\mathbf{K}_n = [k(\mathbf{x}^i,\mathbf{x}^j)]_{i,j=1,\ldots,n}$ is the covariance matrix of the observations at points $\mathbf{D} = (\mathbf{x}^i)_{i=1,\ldots,n}$, with $\mathbf{x}^i \in \mathbb{R}$, for all $i = 1,\ldots,n$. Then, we substitute

$\hat{f}_0$ in the likelihood and maximize it with respect to σ^2. We obtain the following maximum likelihood estimate of σ^2:

$$\hat{\sigma}^2 = \frac{{}^t(\mathbf{z}_n - \hat{f}_0\mathbf{1}_n)\mathbf{K}_n^{-1}(\mathbf{z}_n - \hat{f}_0\mathbf{1}_n)}{n}. \tag{30}$$

Finally, we substitute σ^2 with $\hat{\sigma}^2$ in the likelihood to obtain the marginal likelihood:

$$\mathcal{L}(\boldsymbol{\theta}; \mathbf{z}_n) = n\log(\hat{\sigma}^2) + \log(\det \mathbf{K}_n). \tag{31}$$

The estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is obtained by minimizing (31) with respect to $\boldsymbol{\theta}$. In practice, we use an evolutionary algorithm coupled with a BFGS (Broyden-Fletcher-Goldfarb-Shanno) procedure (see [Avriel, 2003]).

5.3 Academic example: the Ishigami function

Let us consider the Ishigami function:

$$z(x_1, x_2, x_3) = \sin(x_1) + 7\sin(x_2)^2 + 0.1x_3^4\sin(x_1) \tag{32}$$

with $(x_1, x_2, x_3) \in [-\pi, \pi]^3$. This function is a classical tabulated function for sensitivity analysis [Saltelli et al., 2000].

5.3.1 Gaussian process regression model building

First of all, let us present the meta-model building. The considered experimental design set is a Latin-Hypercube-Sampling (LHS) [Stein, 1987] of $n = 150$ points optimized with respect to the maximin criterion. This criterion maximizes the minimum distance between the points. We consider the Gaussian covariance kernel presented as Example 2 of Paragraph 5.1. The maximum likelihood estimates of the model parameters are given below (see Paragraph 5.2):

$$\hat{\boldsymbol{\theta}} = \begin{pmatrix} 1.98 & 1.44 & 1.63 \end{pmatrix}, \quad \hat{\sigma}^2 = 16.50, \quad \hat{f}_0 = 3.40.$$

The efficiency of the model is assessed on a test set $\mathbf{X}_{\text{test}}$ of size n_t uniformly spread on $[-\pi, \pi]^3$ with the following coefficient:

$$Q^2 = 1 - \frac{\sum_{\mathbf{x}\in\mathbf{X}_{\text{test}}} \left(\mu(\mathbf{x}) - f(\mathbf{x})\right)^2}{\sum_{\mathbf{x}\in\mathbf{X}_{\text{test}}} \left(\mu(\mathbf{x}) - \bar{f}\right)^2}, \quad \bar{f} = \frac{\sum_{\mathbf{x}\in\mathbf{X}_{\text{test}}} f(\mathbf{x})}{n_t},$$

where $\mu(\mathbf{x})$ is the predictive mean given in (11). The coefficient Q^2 represent the part of the model discrepancy explained by the meta-model. The closer to 1, the more efficient is the meta-model. Here, the estimated efficiency is $Q^2 = 98.2\%$. Then, we have an accurate meta-model.

5.3.2 Ishigami function with independent inputs

We consider the product measure $P_{\mathbf{X}} = P_{X_1} \otimes P_{X_2} \otimes P_{X_3}$ with X_i uniformly distributed on $[-\pi, \pi]$, i.e. $X_i \sim \mathcal{U}(-\pi, \pi)$, for $i = 1, 2, 3$. In this case, the theoretical sensitivity indices coincide to the classical Sobol indices (see [Sobol, 1993]). Their values are indicated in Table 1. The purpose of this paragraph is to study the relevance of the suggested indices in the case of independent inputs.

To perform the Monte-Carlo integration presented in Paragraph 4.2, we generate a sample $\mathbf{T} = (\mathbf{t}^j)_{j=1,\dots,m}$ of $m = 10,000$ points with respect to the product measure $P_{\mathbf{X}}$. Further, we generate $N_s = 200$ realizations (see $s^f_{u,m}$ in Equation (27)) of the estimator $S^f_{u,m}$ of the sensitivity measure S^f_u with $u \in \{\{1\},\{2\},\{3\},\{1,2\},\{1,3\},\{2,3\}\}$.

To get a scalar quantity estimate $\hat{S}^f_u$ of S^f_u, we consider that $\hat{S}^f_u$ is the mode of the probability density estimate of the $N_s = 200$ realizations of $S^f_{u,m}$. We note that the estimate of the probability density is based on a normal kernel function with a window parameter optimal for estimating a normal density [Bowman and Azzalini, 1997].
The estimated indices are given in Table 1. Furthermore, we provide the confidence intervals of each estimators using the procedure given in Paragraph 4.3 that allows to take into account the meta modeling and the Monte Carlo errors. To do that, we consider a sample of size $K = 200$ for each realization.

Index	S_1	S_2	S_3	S_{12}	S_{13}	S_{23}
Analytical	0.314	0.442	0	0	0.244	0
Estimate	0.310	0.447	0.000	0.001	0.238	0.000
2.5%-quantile	0.308	0.430	-0.001	-0.001	0.231	-0.001
97.5%-quantile	0.320	0.452	0.001	0.002	0.253	0.001

Table 1: Sensitivity measure estimates for the Ishigami function with independent input parameters.

We see that the estimated sensitivity measures fit the theoretical ones. This emphasizes the efficiency of the suggested estimation procedure.
To show the relevance of the estimated 95%-confidence intervals, we reiterate the presented procedure with 500 Gaussian process regression models built from different maximin-LHS design sets. For each design set, the parameters $\boldsymbol{\theta}$, σ^2 and f_0 are estimated with a maximum likelihood method. We thus have 500 estimated 95% confidence intervals and we verify whether they include the true index or not. The ratio of intervals including the true indices, also called the coverage rate, has to be close to 95%. The results of this procedure is presented in Table 2. Moreover, these intervals are compared with the ones considering only the meta-modeling error (i.e. without using the procedure presented in Paragraph 4.3 to evaluate the Monte-Carlo integration error).
Furthermore, to show the issue involving the meta-modeling error, we compare the coverage rate of the empirical estimation $\hat{S}^{\mathrm{D}}_u$ of the sensitivity index S^{D}_u proposed by Durrande *et al.*(see Paragraph 3.1) to the one of our sensitivity measure. To evaluate the Monte-Carlo error of $\hat{S}^{\mathrm{D}}_u$, we apply the procedure presented in Paragraph 4.3. The results are presented in Table 2.
We see in Table 2 that the empirical confidence intervals obtained with the suggested procedure are better than those which only take into account the meta-model or the Monte-Carlo error. In particular, the confidence intervals found with the estimator $\hat{S}^{\mathrm{D}}_u$ and considering only the Monte-Carlo error are widely underestimated. However, we see that they are all underestimated for the second order indices. Especially for the indices corresponding to the non-influent interactions, namely, S_{12} and S_{23}. The underestimation could be due to a poor learning of the interactions by the meta-model.

Index	Error	S_1	S_2	S_3	S_{12}	S_{13}	S_{23}
$\hat{S}_u^f$	MC + Meta-model	0.95	0.97	0.91	0.76	0.81	0.74
	Meta-model	0.87	0.94	0.87	0.08	0.29	0.21
$\hat{S}_u^D$	MC + predictive mean	0.67	0.65	0.39	0.00	0.26	0.05

Table 2: Coverage rates of 500 empirical confidence intervals. The theoretical confidence interval is 95%. The coverage rates for the suggested confidence intervals taking into account the uncertainty of both the meta-model approximation and the Monte-Carlo integration are labeled "MC+meta-model" ; the ones taking into account only the meta-model error are labeled "meta-model" ; the ones taking into account only the Monte-Carlo error and using the predictive mean are labeled "MC+predictive mean".

5.3.3 Ishigami function with perfectly correlated inputs

We present here a sensitivity analysis with $P_{X_i} \sim \mathcal{U}(-\pi, \pi)$, $i = 1, 2, 3$, and where we assume that $X_1 = X_2$ and X_1, X_2 independent of X_3.
Therefore, we can either perform a sensitivity analysis considering only two independent variables (Case i) or perform a sensitivity analysis with three input variables where two of them have a perfect positive linear relationship (Case ii). We use the classical Sobol indices for Case i, and we perform our procedure for Case ii.
As the two sensitivity analyses formally correspond to the same underlying function, it should have a connection between them. The purpose of this paragraph is to numerically observe it. Since $X_1 = X_2$, we can consider the following function:

$$z_{(i)}^{\text{sob}}(X_1, X_3) = \sin(X_1) + 7\sin(X_1)^2 + 0.1X_3^4 \sin(X_1),$$

with X_1 independent of X_3. Further, we denote $\hat{S}_u^{\text{sob}}$, for $u \in \{\{1\}, \{3\}, \{13\}\}$ the estimators of the Sobol index. Our indices are denoted $\hat{S}_u^f$, for $u \in \{\{1\}, \{2\}, \{3\}, \{12\}, \{13\}, \{23\}\}$, as we consider the mode of the distribution estimate. Results are given in Table 3.

Index	$\hat{S}_1$	$\hat{S}_2$	$\hat{S}_3$	$\hat{S}_{12}$	$\hat{S}_{13}$	$\hat{S}_{23}$
$\hat{S}_u^f$	0.308	0.439	0.001	0.001	0.238	0.012
$\hat{S}_u^{\text{sob}}$	0.751	-	0.001	-	0.245	-

Table 3: Sensitivity measure estimates for the Ishigami function with $X_1 = X_2$, X_1, X_2 independent of X_3 and $P_{X_i} \sim \mathcal{U}(-\pi, \pi)$, $i = 1, 2, 3$.

We see in Table 3 that we empirically found that $\hat{S}_1^{\text{sob}} \approx \hat{S}_1^f + \hat{S}_2^f + \hat{S}_{12}^f$, $\hat{S}_3^{\text{sob}} \approx \hat{S}_3^f$ and $\hat{S}_{13}^{\text{sob}} \approx \hat{S}_{13}^f + \hat{S}_{23}^f$. Therefore, we numerically observe a direct correspondence between the classical sensitivity analysis for independent inputs and the suggested one for dependent inputs.
This connection strengthen the relevance of the considered index since, in the independent case, the Sobol indices are commonly accepted as a good measure of sensitivity. However, the connection is only established when the considered model can be reduced to an equivalent model which have independent inputs. For general cases, the interpretation will be much more complex (see Paragraph 5.4).

5.3.4 Modeling dependence with copulas

To define the dependence among random variables, it is usual to use the copula functions [Nelsen, 2006]. Indeed, a copula function aims to join the joint distribution of a set of variables to its marginal distributions. If the cumulative distribution function (cdf) of $\mathbf{X}$ is denoted $F_{\mathbf{X}}$, and $F_1, \cdots, F_p$ are the respective marginal cdf of $X_1, \cdots, X_p$, then there exists a copula p-dimensional C such that, for all $\mathbf{x} \in \mathbb{R}^p$,

$$F_{\mathbf{X}}(\mathbf{x}) = C(F_1(x_1), \cdots, F_p(x_p)).$$

Most of copulas belong to a class of copulas, specified by the type of dependence it models. For instance, among the upper tail dependence, one could cite the the family of Gumbel copulas [Nelsen, 2006]. Thus, copulas provide a simple and natural way to measure the dependence. However, copulas are not the most widely used tool in practice. To measure the dependence, it is usual to refer to the Pearson's correlation coefficient, that measures the linear dependence among variables. It is especially appropriated to elliptical distributions [Fang et al., 1990], but it may be misleading for other types of distribution. The Spearman's rho is a good alternative to the Pearson's coefficient because it could be adapted to any distribution. Based on the probability of concordance and discordance of random variables, the Spearman's rho is also well-known as a rank correlation, i.e. the linear Pearson correlation coefficient applied on the rank of observations. The main advantage of this measure is that it does not depend on the marginal distributions, but only on the structure of dependence. Furthermore, it is a copula-based measure of association, i.e. when the dependence between two random variables is modelized by a copula C, the Spearman's rho, denoted ρ^{S}, admits the following expression,

$$\rho^{\mathrm{S}} = 12 \iint_{[0,1]^2} C(u, v) du dv - 3.$$

Through the Ishigami function, we study the influence of this coefficient on the estimation of our sensitivity measures. We fix ρ^{S}, and we model our dependence by two different copulas. The aim is to know if the dependence may be summarized by the Spearman's rho in the sensitivity analysis in presence of dependent incomes.
We assume that each couple of variables (X_i, X_j), $i \neq j$ admits the same Spearman's rho, $\rho^{\mathrm{S}} = 0.7$. We use the Gaussian copula, and the Clayton copula on uniform marginal distributions over $[-\pi, \pi]$. Further, for a given experimental design set of $n = 200$ points, we compare the two dependence structure. Further, the Monte Carlo sample is of size $m = 10,000$, and we made $N_s = 200$ realizations of $S^f_{u,m}$. Figure 1 illustrates the distribution of the indices for both Gaussian and Clayton copulas.

Notice that, depending on the type of dependence, we do not obtain the same conclusion. Especially for S_1 and S_{13}, we notice that the two distributions are disjoint, meaning a significant difference whether we use a Gaussian copula or a Clayton one. This shows that it is not enough to only consider a measure of association to model the dependence in sensitivity analysis.

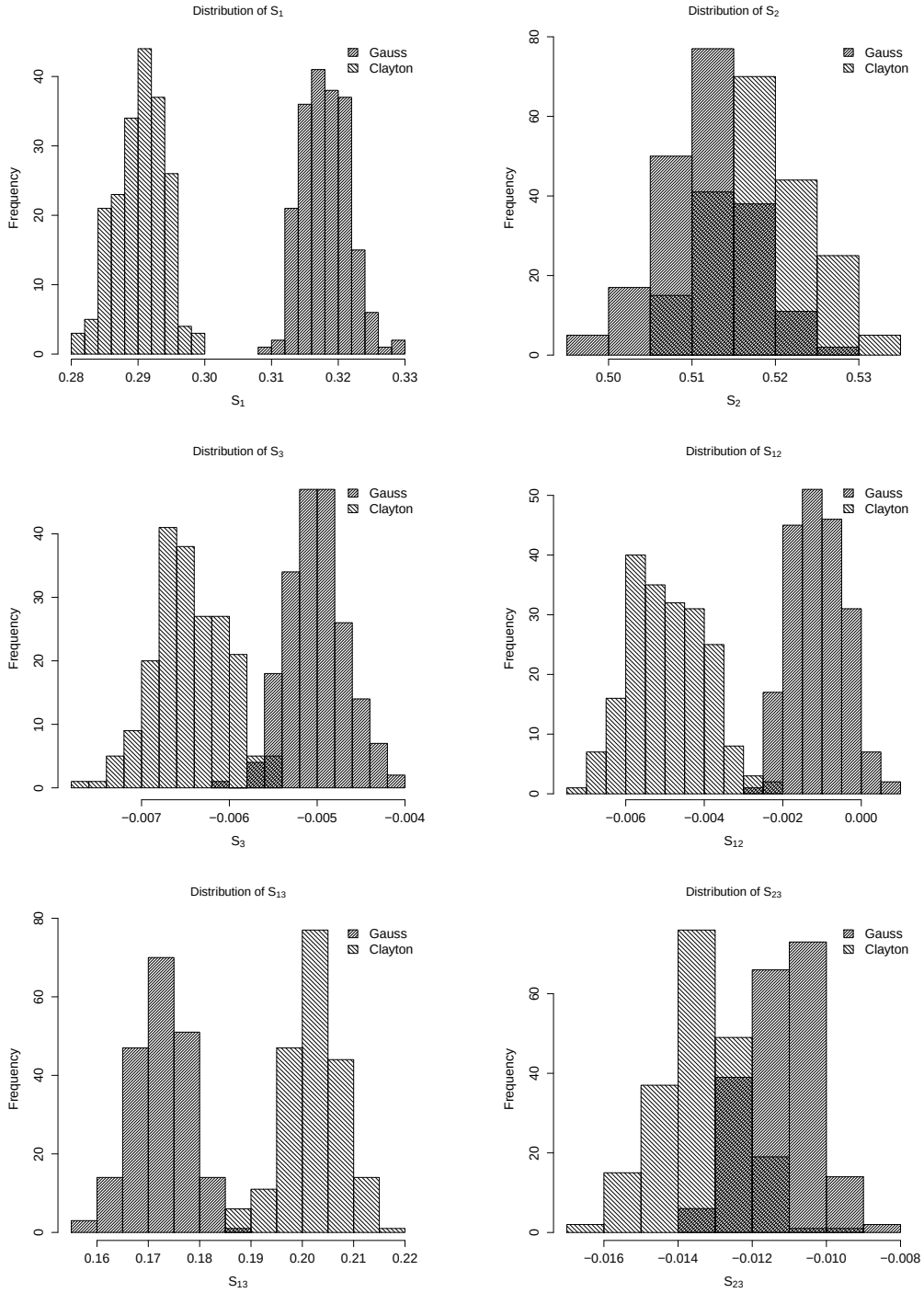


Figure 1: Distribution of the sensitivity measures with Gaussian and Clayton copulas

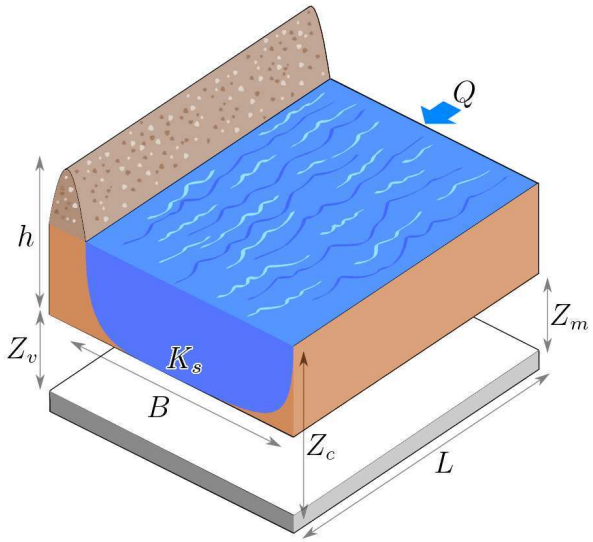


Figure 2: The river flood model

5.4 Industrial example: the river flood inundation

We illustrate our method with the study of a river flood inundation. In this problem, the river flow is compared with the height of a dyke that protects an industrial site [Faivre et al., 2013, De Rocquigny, 2006]. The river flow may lead to inundations that are desirable to avoid. To study this phenomenon, the maximal overflow of the river is modelized by a crude simplification of the 1-D Saint Venant equations, when uniform and constant flow rate is assumed. The model is given by the following expression,

$$S = \underbrace{Z_v + h}_{Z_c} - H_d - C_b, \quad h = \left(\frac{Q}{B K_s \sqrt{\frac{Z_m - Z_v}{L}}} \right)^{0.6},$$

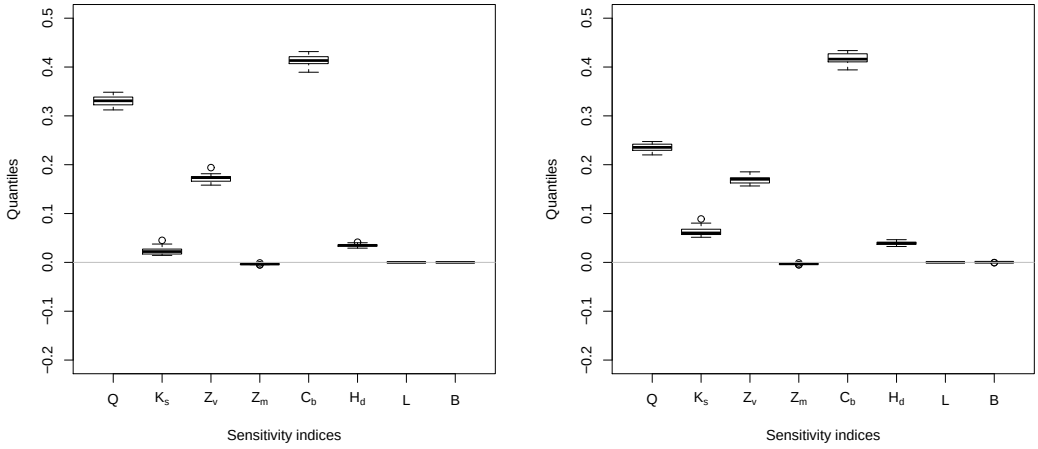
where S is the maximal overflow that depends on eight parameters. The river flow and the parameters implied in the model are represented in Figure 2. These variables are physical and geometrical parameters subject to a spatio-temporal variability or to errors of measurements. Thus, leading a sensitivity analysis in this model has a real interest for this model. The meaning of the incomes and their distribution are given in Table 4.
In this study, we assume that (Q, K_s) is a correlated pair, with correlation coefficient $\rho = 0.5$. This correlation is admitted in real case, as we consider that the friction coefficient increases with the flow rate. Also, (Z_v, Z_m) and (L, B) are assumed to be dependent with the same Pearson coefficient $\rho = 0.3$, because data are supposed to be simultaneously collected by the same measuring device. As for C_b and H_d, they are supposed to be independent.
We take a first sample of $n = 200$ observations, and a Monte Carlo sample of size $m = 5000$. Further, we generate $N_s = 100$ realizations of the first order sensitivity indices. We then consider the mode of the probability density estimate of these realizations. We repeat the procedure 100 times to obtain a Monte Carlo error for each sensitivity index. Further, we compare our result to the generalized sensitivity indices defined in [Chastaing et al., 2013], built from a functional decomposition, called *hierarchical* decomposition. The estimation of these last indices is based on a regression approach, and a recursive procedure [Chastaing

Variables	Meaning	Distribution
h	maximal annual water level	-
Q	maximal annual flow rate	Gumbel $G(1013, 558)$ tr. to $[500; 3000]$
K_s	Strickler coefficient	Normal $N(30, 8)$ tr. to $[15, +\infty[$
Z_v	river downstream level	Triangular $T(49, 50, 51)$
Z_m	river upstream level	Triangular $T(54, 55, 56)$
H_d	dyke height	Uniform $\mathcal{U}([7, 9])$
C_b	bank level	Triangular $T(55, 55.5, 56)$
L	length of the river stretch	Triangular $T(4990, 5000, 5010)$
B	river width	Triangular $T(295, 300, 305)$

Table 4: Description of inputs-output of the river flood model (tr. to=truncated to)

et al., 2013]. Further, as this procedure suffers from the curse of dimensionality, a greedy algorithm is adopted to select a sparse number of informative components. This other strategy will be called the GHOGS (for Greedy Hierarchical Orthogonal Gram-Schmidt) strategy. The comparison with our methodology is given by Figure 3.

Through the result, we observe for both decompositions the same phenomena for the last six inputs (Z_v, Z_m, H_d, C_b, L and B). The width (B) and the length (L) of the river are not influent parameters in the model. Also, for both decompositions, the dyke height is the most contributive variable in the global variability. Moreover, the bank level (C_b) has a negligible impact on the model output and its contribution has the same order of magnitude for both analyses. The main difference between the two procedures is the contribution of the correlated pair (Q, K_s). In the GHOGS procedure, we observe that the flow rate Q is highly contributive with respect to the Strickler coefficient K_s. This contribution is less important for the Gaussian processes approach while the one of K_s is slightly larger. Furthermore, we note that the sum of contributions for the pair (Q, K_s) is similar for the two analyses.
We see in (21) that the sensitivity measure is decomposed into a sum of a variance term $V[f_u^n(\mathbf{X}_u)]/V[f^n(\mathbf{X})]$ and a covariance term $\mathrm{Cov}[f_u^n(\mathbf{X}_u), f_{u^c}^n(\mathbf{X})]/V[f^n(\mathbf{X})]$. The same type of decomposition is present in the index provided by the GHOGS procedure [Chastaing et al., 2013]. The variance term represents the main contribution of the inputs without the dependence part. The covariance term represents the contribution of the dependence to the index. We represent for the two methods the estimated variance and the covariance parts in Figure 4. We observe that the GP and the GHOGS behave differently, as we are not faced to the same decomposition. In the GP approach, the model tends to balance the main contribution, whereas the GHOGS is more discriminant. The covariance contribution is the same in the GHOGS procedure for (Q, K_s), which seems reasonable as it is estimated from a hierarchically orthogonal decomposition [Chastaing et al., 2012]. However, we observe a significant difference between the covariance part of Q and the one of K_s, that may be due to the fact that we measure $\mathrm{Cov}(f_{K_s}^n, f_{K_s^c}^n)$. This last term implies a large sum of terms that may weaken the



(a) GHOGS procedure

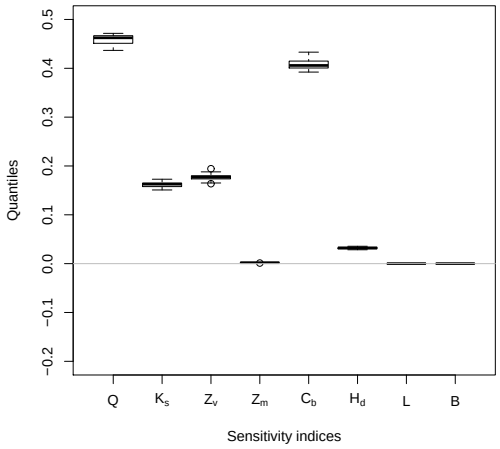
(b) GP modelling

Figure 3: Sensitivity indices estimation with the GHOGS method (a) and the GP modelling (b)

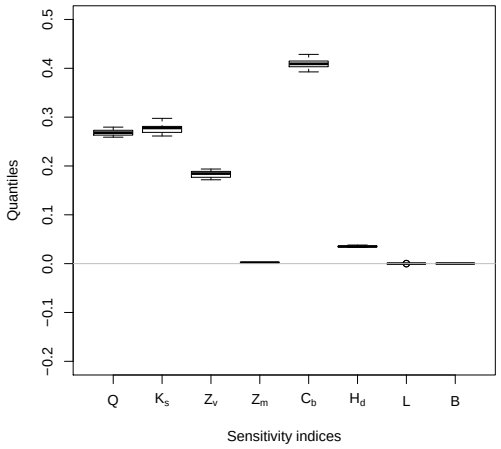
main contribution, and that lead to a negative covariance contribution.

6 Conclusions

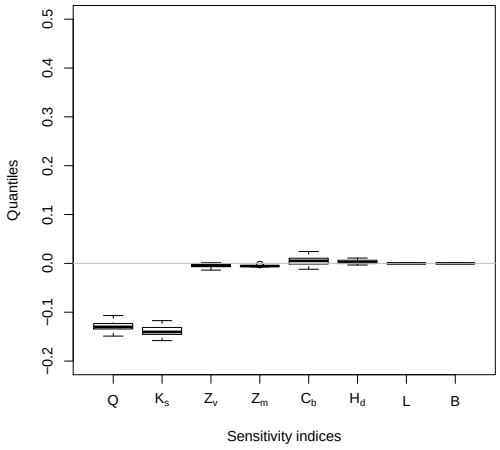
Through this work, we propose a solution for dealing with complex computer codes in presence of dependent input variables in the model. The definition of a variance-based sensitivity index aims at quantifying the contribution of a (group of) variable(s) in the model and can be decomposed as a sum of ratio of variances, interpreted as the main contribution, and a ratio between covariance terms and the global variance, interpreted as the contribution due to the dependence. The attractive side of such methodology is to be able to quantify the uncertainty of the sensitivity measure, and thus to compute confidence intervals for each estimation. The question about the choice of the ANOVA kernel has not been raised in this work, as this choice may have a strong influence on the values of the sensitivity indices. This remains an open problem.



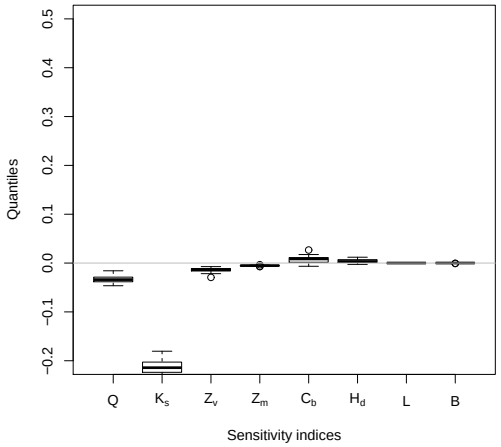
(a) Variance term for the GHOGS procedure



(b) Variance term for the GP modelling



(c) Covariance term for the GHOGS procedure



(d) Covariance term for the GP modelling

Figure 4: Variance (a) & (b) and covariance (c) & (d) terms for the Sensitivity indices estimation with the GHOGS method (a) & (c) and the GP modelling (b) & (d)

References

[Avriel, 2003] Avriel, M. (2003). *Nonlinear programming: analysis and methods*. Dover Publications.

[Berlinet and Thomas-Agnan, 2004] Berlinet, A. and Thomas-Agnan, C. (2004). *Reproducing kernel Hilbert spaces in probability and statistics*. Kluwer Academic Publishers.

[Bowman and Azzalini, 1997] Bowman, A. W. and Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis*. Oxford University Press.

[Cacuci et al., 2005] Cacuci, D., Ionescu-Bujor, M., and Navon, I. (2005). *Sensitivity and Uncertainty Analysis, Volume II: Applications to Large-Scale Systems*, volume 2. Chapman & Hall/CRC.

[Cameron and Martin, 1947] Cameron, R. and Martin, W. (1947). The orthogonal development of non-linear functionals in series of fourier-hermite functionals. *The Annals of Mathematics*, 48(2):385–392.

[Caniou, 2012] Caniou, Y. (2012). *Analyse de sensibilité globale pour les modèles imbriqués et multiéchelles*. PhD thesis, Université Blaise Pascal - Clermont II.

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

[Chastaing et al., 2013] Chastaing, G., Gamboa, F., and Prieur, C. (2013). Generalized sobol sensitivity indices for dependent variables: Numerical methods. Disponible à `http://arxiv.org/abs/1303.4372`.

[Chen et al., 2005] Chen, W., Jin, R., and Sudjianto, A. (2005). Analytical variance-based global sensitivity analysis in simulation-based design under uncertainty. *Journal of Mechanical Design*, 127(5):875–886.

[Chilès and Delfiner, 1999] Chilès, J. and Delfiner, P. (1999). Geostatistics: modeling spatial uncertainty. *Wiley series in probability and statistics (Applied probability and statistics section)*.

[De Rocquigny, 2006] De Rocquigny, E. (2006). La maîtrise des incertitues dans un contexte industriel-1ere partie: une approche méthodologique globale basée sue des exemples. *Journal de la Société Française de Statistique*, 147(2):33–71.

[Durrande et al., 2013] Durrande, N., Ginsbourger, D., Roustant, O., and Carraro, L. (2013). Reproducing kernels for spaces of zero mean functions. Application to sensitivity analysis. *Journal of Multivariate Analysis*, 115:57–67.

[Faivre et al., 2013] Faivre, R., Iooss, B., Mahévas, S., Makowski, D., and Monod, H. (2013). *Analyse de sensibilité et exploration de modèles*. Quae.

[Fang et al., 1990] Fang, K.-T., Kotz, S., and Ng, K.-W. (1990). *Symmetric Multivariate and Related Distributions -Monographs on Statistics and Applied Probability*. Chapman and Hall, London.

[Fang et al., 2006] Fang, K.-T., Li, R., and Sudjianto, A. (2006). *Design and Modeling for Computer Experiments*. Chapman & Hall - Computer Science and Data Analysis Series, London.

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

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

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

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

[Marrel et al., 2009] Marrel, A., Iooss, B., Laurent, B., and Roustant, O. (2009). Calculations of sobol indices for the gaussian process metamodel. *Reliability Engineering & System Safety*, 94(3):742–751.

[Nelsen, 2006] Nelsen, R. (2006). *An introduction to copulas*. Springer, New York.

[Oakley and O'Hagan, 2004] Oakley, J. and O'Hagan, A. (2004). Probabilistic sensitivity analysis of complex models: a bayesian approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(3):751–769.

[Rasmussen and Williams, 2006] Rasmussen, C. and Williams, C. (2006). *Gaussian processes for machine learning*. MIT Press, Cambridge.

[Sacks et al., 1989] Sacks, J., Welch, W., Mitchell, T., and Wynn, H. (1989). Design and analysis of computer experiments. *Statistical science*, 4(4):409–423.

[Saltelli et al., 2000] Saltelli, A., Chan, K., and Scott, E. (2000). *Sensitivity Analysis*. Wiley, West Sussex.

[Saltelli et al., 2008] Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S. (2008). *Global sensitivity analysis: The primer*. Wiley-Interscience, West Sussex.

[Santner et al., 2003] Santner, T., Williams, B., and Notz, W. (2003). *The design and analysis of computer experiments*. Springer Verlag, New York.

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

[Stein, 1987] Stein, M. L. (1987). Large sample properties of simulations using latin hypercube sampling. *Technometrics*, 29:143–151.

[Stein, 1999] Stein, M. L. (1999). *Interpolation of Spatial Data*. Springer Series in Statistics, New York.

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

[Sudret, 2008] Sudret, B. (2008). Global sensitivity analysis using polynomial chaos expansion. *Reliability engineering and system safety*, 93(7):964–979.

[Van Der Vaart, 1998] Van Der Vaart, A. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge.

[Wiener, 1938] Wiener, N. (1938). The homogeneous chaos. *American Journal of Mathematics*, 60(4):897–936.

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