

Pick and Freeze estimation of sensitivity index for static and dynamic models with dependent inputs

Titre: Estimation d'indice de sensibilité de modèles statiques et dynamiques pour des données dépendantes par la méthode d'échantillonnage *Pick and Freeze*

Mathilde Grandjacques¹, Benoit Delinchant² and Olivier Adrot²

Abstract: This article addresses the estimation of the Sobol index for dependent static and dynamic inputs. We study transformations in the input, whose image is an input with independent components. They have the basic property to give an invariance property for conditional expectation between a subset of inputs and their image that allows to use the Pick and Freeze method.

We first focus on the static case. The general case and the Gaussian case are detailed. In the non Gaussian case we apply the conditional quantile function generally used to simulate random vectors in a new framework. In the Gaussian case the dependent variables are separated into two groups of independent variables.

In the dynamic case the definition of the index has been slightly modified in order to take into account the two dimensions of dependence (temporal and spatial). For Gaussian processes the same method as previously is used. For non Gaussian processes for which in general there is no sufficient information to get a model, we propose to use a copula model to get back to Gaussian inputs. Different cases are studied in order to underline on the weakness, in sensitivity studies, to use the correlations like the measures of dependence.

Résumé : Cet article traite de l'estimation des indices de Sobol pour des entrées dépendantes, statiques et dynamiques. Nous proposons de transformer l'entrée dépendante en une image dont les composantes sont indépendantes. Elles ont la propriété de vérifier l'invariance des espérances conditionnelles sachant le sous espace formé des entrées et celui de leurs images, ce qui nous permet d'appliquer la méthode *Pick and Freeze* au modèle.

Nous traitons tout d'abord l'aspect statique, en détaillant le cas général non Gaussien et le cas Gaussien. Dans le cas non Gaussien, nous appliquons la méthode dite des quantiles conditionnels, généralement utilisée pour simuler des vecteurs aléatoires. Dans le cas Gaussien, les variables dépendantes sont séparées en deux groupes de variables indépendantes.

Concernant l'aspect dynamique, la définition des indices a été légèrement modifiée afin de prendre en compte les deux dimensions de dépendance de l'entrée, la dimension temporelle et la dimension spatiale. Pour les processus Gaussiens, nous utilisons la même méthode que dans le cas statique. Pour les processus non Gaussiens, nous proposons d'utiliser un modèle à base copule pour revenir à des entrées Gaussiennes.

L'étude de différents cas met en évidence le fait que, dans les études de sensibilité, l'utilisation de la corrélation comme mesure de dépendance a ses limites.

Keywords: sensitivity analysis, dependent inputs, time series, copula model, Pick and Freeze estimation

Mots-clés : analyse de sensibilité, entrées dépendantes, séries temporelles, modèle copule, estimation Pick and Freeze

AMS 2000 subject classifications: 35L05, 35L70

¹ G2ELab, Université de Grenoble, 38402 St-Martin d'Hères, France.

E-mail: mathilde.grandjacques@g2elab.grenoble-inp.fr

² GSCOP, 46 avenue Félix Viallet, 38031 Grenoble Cedex 1, France.

1. Introduction

Global sensitivity analysis (GSA) aims to pick out, in a input-output system, the variables that contribute the most to the uncertainty on the output.

GSA is popular for systems such as non linear regression or more complex systems, which are studied mainly by stochastic tools for independent inputs.

Many methods exist in the literature (see for example [Saltelli et al., 2000](#) and references therein). The most used one is the Sobol index which is defined if the variables are assumed to be independent random variables. Their probability distributions account for the practitioner's belief in the input uncertainty. This turns the model output into a random variable, whose total variance can be splitted down into different partial variances (this is the so-called Hoeffding decomposition, also known as functional ANOVA, see [Liu and Owen, 2006](#)). Each partial variance is defined as the variance of the conditional expectation of the output with respect to each input variable. By considering the ratio of each partial variance to the total variance, we obtain the Sobol sensitivity index of the variable [Sobol \(1993, 2001\)](#). This index quantifies the impact of the variability of the factor on the output. Its value is between 0 and 1, allowing to prioritize the variables according to their influence.

Popular methods to calculate Sobol indices are for instance :

- Fourier methods ([Cukier et al., 1973](#); [Mara, 2009](#)) used in a different setting, with the aim to simplify computations
- Orthogonal polynomials for example polynomial chaos [Sudret \(2008\)](#)
- Random balance design [Tarantola et al. \(2006\)](#)

but they require independent input components and (or) a known precise analytical form of f .

A quite different approach, suggested by Sobol (see [Sobol, 2001](#) and [Gamboa et al., 2013](#)) is the Sobol Pick-Freeze (SPF) scheme. It is also based on the independence of components but it is more flexible on the form of the inputs and does not take into account the shape of the input-output model. In SPF, a Sobol index is viewed as the regression coefficient between the output of the model and its pick-frozen replication. This replication is obtained by holding the value of the variable of interest (frozen variable) and by sampling the other variables (picked variables). The sampled replications are then combined to produce an estimator of the Sobol index. There is no requirement about the knowledge of f , except the possibility to simulate the system which is of course a severe constraint. [Janon and al. \(Janon et al., 2013, 2014\)](#) give asymptotic results when the sample size tends to infinity. Estimators are convergent, satisfy a central limit theorem and have robustness properties.

However, in most applications, the parameters or the inputs are dependent due to physical constraints. The interpretation in this case is not easy. If we want to study the sensitivity with respect to a component say X^1 , it is of course not sufficient to study only X^1 as it appears explicitly in the model. It contributes to uncertainty through the other components linked with it. So, conventional methods of sensitivity analysis cannot be used with dependent inputs. The classical orthogonal Hoeffding decomposition must be used with precautions. In particular the notion of interaction between two components, valid for independent cases, is here meaningless.

Several studies have been conducted in the case of dependent parameters. Mara and al. [Mara and Tarantola \(2012\)](#), Xu and al. [Xu and Gertner \(2011\)](#) give approaches that are most often used for linear models and specific forms of dependence. Da Veiga and al. [Da Veiga et al. \(2009\)](#) use a very natural idea : when we have a sample $(X^{(i)}, Y^{(i)})_{i=1, \dots, N}$ of the system, even if f is unknown, we can estimate $\mathbf{E}(Y|X^1)$ and the conditional moments of the output with respect to some factor of interest : X^1 . The use of non parametric statistics for this kind of problem is common in other fields such as econometrics, for instance LOESS method is quite easy in this framework. Kutcherenko and al. [Kucherenko et al. \(2012\)](#) calculate a sensitivity index analogous to Sobol's formula, from a priori knowledge of probability distribution functions. To obtain it they propose to transform the input into a Gaussian copula. Then this copula is used as a model. If we take an input with uniform marginals and given correlation, there are a lot of models with various properties, so the misspecification when one chooses a model can lead to very different sensitivity indices.

A deep work on dependent inputs, starting from ideas of [Stone et al. \(1997\)](#) and [Hooker \(2007\)](#), is the work of Chastaing and al. ([Chastaing et al., 2011](#); [Chastaing and Le Gratiot, 2015](#); [Chastaing, 2013](#)) perhaps limited by computation problems but giving a clear formulation of the dependence role. This work is based on the existence of an Hoeffding representation under light conditions on the density of the input:

$$f(X^1, \dots, X^p) = \sum_{j=1}^p f_j(X^j) + \sum_{\substack{k,j=1 \\ k \neq j}}^p f_{k,j}(X^j, X^k) + \dots + f_{1, \dots, p}(X^1, \dots, X^p) \quad (1)$$

The classical orthogonality due to independence which allows easy computations of the Sobol index is lost but a useful other form of orthogonality extends the classical one. Hierarchical orthogonal decomposition means that a term indexed by k_1, \dots, k_p is orthogonal to any term indexed by a subset of $\{k_1, \dots, k_p\}$ and this property is sufficient to obtain (1).

In a time related framework such that :

$$Y_t = f_t((X_s)_{s=0, \dots, t}) \quad (2)$$

where $f(\cdot) : \mathbb{R}^{p(t+1)} \rightarrow \mathbb{R}$, and the input is (not necessarily independent) a vectorial process $(X_s)_{s \in \mathbb{N}} \subset \mathbb{R}^p$. Few studies propose to study the sensitivity for dynamic inputs. The sensitivity is calculated at each time step t without taking into account the dynamic behaviour of the input. Indeed, the impact of the variability is not always instantaneous. Therefore it seems necessary to develop a new method for dynamic dependent inputs. The Sobol index definition has been modified in order to take into account the dynamic behaviour of the inputs. Each partial variance is defined as the variance of the conditional expectation of the output with respect to a certain time of observation (called memory) : (X_t, \dots, X_{t-k}) of the input vector variable. So the index is defined for each $k \in [0, t]$ and each t . For stationary processes we prove in [Grandjacques et al. \(2015\)](#) that the k -sensitivity is independent of k and converges as $t \rightarrow \infty$ at least for important situations but probably not for all.

Our proposition is to find a transformation that turns dependent inputs into independent inputs and that keeps the sensitivity invariant. It is then possible to apply the Pick and Freeze method to the independent variables to calculate the Sobol indices.

The article is decomposed into two parts. The first regards the static case, when the inputs are vectorial. The second part develops the case where the inputs are dynamic.

In the static case, a method based on the conditional quantile method is proposed. This method is the most general to simulate random vectors. Dependent variables with any distribution are transformed into independent uniform variables and therefore very easily simulated. So, the Pick and Freeze method can be applied.

A remark is done on the easiest case : the case of Gaussian inputs. The transformation is given by a general multivariate regression both for input vectors that input processes. The Pick and Freeze method is easily applied, the independent variables obtained by the transformation are also Gaussian and so are easily generated.

When the information is not sufficient to fix a model, a second method is proposed based on the copula model. This method allows to create a model of dependent variables from the given marginal distributions and correlation matrix [Ghosh and Henderson \(2003\)](#). In the static part we detail a copula type model and to complete [Kucherenko et al. \(2012\)](#) study, we show that the choice of the copula is important, concerning sensitivity analysis. The sensitivities are computed for different simple models and we deal with, specifically for sensitivity values, the difficulties arising from the use of correlation as a dependence measure to build convenient models.

In the dynamic case, after to give the new definition of the Sobol index, some properties are given for the most usual processes : the stationary processes. The calculation of the Sobol index for dependent Gaussian processes is detailed.

When the processes are not Gaussian, [Cario and Nelson \(1996\)](#) propose a method called NORTA (NORmal To Anything) that produces a random process with some desired properties (marginals and different kind of correlations) via a transformation of a multivariate normal random process easily generated them. This model will be used in the dynamic part to calculate sensitivity for non Gaussian processes. We consider a copula type model to get back to Gaussian inputs where sensitivity method of calculation is easily applied. It is chosen starting from each input marginal function and starting from the correlation between X_t and X_{t-1}, X_{t-2}, \dots

2. Sensitivity for dependent inputs : static case

2.1. Sobol index and Pick and Freeze method : independent case

Let an input-output system given by :

$$Y = f(X) \quad (3)$$

with $Y \in \mathbb{R}$, $X = (X^1, \dots, X^p) \in \mathbb{R}^p$.

The closed Sobol index is defined as :

$$S^{X^J} = \frac{\mathbf{Var}(\mathbf{E}(Y|X^J))}{\mathbf{Var}Y}. \quad (4)$$

with $X^J = (X^{j_1}, \dots, X^{j_q})$, $J = \{j_1, \dots, j_q\} \subset \{1, \dots, p\}$.

Lemma 1. *Sobol (2001): Let $X = (X^J, X^{\bar{J}})$ and $Y = f(X^J, X^{\bar{J}})$ with $\bar{J} = \{1, \dots, p\} \setminus J$. If X^J and $X^{\bar{J}}$ are independent :*

$$\mathbf{Var}(\mathbf{E}(Y|X^J)) = \mathbf{Cov}(Y, Y^{X^J})$$

with $Y^{X^J} = f(X^J, (X^{\bar{J}})')$ where $(X^{\bar{J}})'$ is an independent copy of $X^{\bar{J}}$. Copy meaning a random vector independent of $X^{\bar{J}}$ with the same distribution.

We can deduce the expression of the index S^{X^J} when X^J and $X^{\bar{J}}$ are independent :

$$S^{X^J} = \frac{\mathbf{Cov}(Y, Y^{X^J})}{\mathbf{Var}Y} \tag{5}$$

A natural estimator consists in taking the empirical estimators of the covariance and of the variance. Let a N -sample $\{(Y^{(1)}, Y^{X^J, (1)}), \dots, (Y^{(N)}, Y^{X^J, (N)})\}$ a natural estimator of S^{X^J} is :

$$\widehat{S}_N^{X^J} = \frac{\frac{1}{N} \sum_{i=1}^N Y^{(i)} Y^{X^J, (i)} - (\frac{1}{N} \sum_{i=1}^N Y^{(i)}) (\frac{1}{N} \sum_{i=1}^N Y^{X^J, (i)})}{\frac{1}{N} \sum_{i=1}^N (Y^{(i)})^2 - (\frac{1}{N} \sum_{i=1}^N Y^{(i)})^2} \tag{6}$$

Janon et al. (2013) suggest an improvement using a symmetric form :

$$\widetilde{S}_N^{X^J} = \frac{\frac{1}{N} \sum_{i=1}^N Y^{(i)} Y^{X^J, (i)} - (\frac{1}{2N} \sum_{i=1}^N Y^{(i)} + Y^{X^J, (i)})^2}{\frac{1}{N} \sum_{i=1}^N \frac{(Y^{(i)})^2 + (Y^{X^J, (i)})^2}{2} - (\frac{1}{N} \sum_{i=1}^N \frac{Y^{(i)} + Y^{X^J, (i)}}{2})^2} \tag{7}$$

In Janon et al. (2013) these estimators are shown to be consistent, and if $\mathbf{E}(|Y|^4) < \infty$, they satisfy a central limit theorem. Asymptotically, the variance of $\widehat{S}_N^{X^J}$ is the lowest possible one.

2.2. General framework to reduce the dependent case to the independent case

Let $Y = f(X)$ an input-output system and $X = (X^1, \dots, X^p)$. The Pick and Freeze method is based on the Sobol lemma which requires independent inputs. To apply the Pick and Freeze method to dependent inputs we look for an invertible transformation $T : \mathbb{R}^p \rightarrow \mathbb{R}^p$ of the form $(X^J, X^{\bar{J}}) \mapsto (X^J, W)$ such that :

- W independent of X^J
- W is the projection of X on the orthogonal space of $L^2(X^J)$ then $L^2(X) = L^2(X^J, W)$.

Let us consider the random vector $X = (X^J, X^{\bar{J}})$, as previously.

Thus by definition $W \in L^2(X)$, it exists a function ψ such that :

$$W = \psi(X^J, X^{\bar{J}}) = \psi_{X^J}(X^{\bar{J}})$$

and

$$X^{\bar{J}} = \varphi(X^J, W) = \varphi_{X^J}(W)$$

Under this condition the input-output model $Y = f(X^J, X^{\bar{J}})$ can be rewrite as :

$$\begin{aligned} Y &= f(X^J, X^{\bar{J}}) \\ &= f(X^J, \Psi_{X^J}^{-1}(W)) \\ &= h(X^J, W) \end{aligned}$$

Thus to compute the sensitivity S^{X^J} , the Pick and Freeze method is applied to the model h where the inputs (X^J, W) are independent. Indeed :

$$\mathbf{E} \left(f(X^J, X^{\bar{J}}) | X^J \right) = \mathbf{E} (h(X^J, W) | X^J) \quad (8)$$

2.3. General case : the conditional quantile method

We now suppose that X has a density g with respect to the Lebesgue measure.

Let:

$$\begin{aligned} G(x^k | X^1 = x^1, \dots, X^{k-1} = x^{k-1}) &= G_{k|1, \dots, (k-1)}(x^k, x^1, \dots, x^{k-1}) \\ &= \mathbf{P}(X^k < x^k | X^1 = x^1, \dots, X^{k-1} = x^{k-1}) \end{aligned} \quad (9)$$

the conditional distribution of X^k when (X^1, \dots, X^{k-1}) are fixed. It results from the existence of g that all these conditional distributions are well defined.

Lemma 2. *Lévy-Rosenblatt (Rosenblatt, 1956) : Let (U^1, \dots, U^p) the random variables defined for $1 \leq k \leq p$ such that :*

$$U^k = G_{k|1, \dots, (k-1)}(X^k, X^1, \dots, X^{k-1}) \quad (10)$$

then (U^1, \dots, U^p) are uniform and independent random variables.

Proof 1. *The proof is quite obvious*

$$\begin{aligned} \mathbf{P}(U^i \leq u^i, i = 1, \dots, p) &= \int_{\{U, U^i \leq u^i\}} \dots \int G_{p|1, \dots, (p-1)}(x^p, x^1, \dots, x^{p-1}) dx^p \dots \int G_1(x^1) dx^1 \\ &= \int_0^{u^k} \dots \int_0^{u^1} du^k \dots du^1 \end{aligned}$$

by definition of conditional distributions and chain property.

Thus U^k is a function of (X^1, \dots, X^k) .

In order to simplify the definition of inverse functions, we make a (weak) assumption on g .

Let $\mathcal{C} = \text{closure}\{x, g(x) > 0\}$ and assume:

$$g(x) > 0 \text{ if } x \in \text{interior}(\mathcal{C}) \quad (11)$$

From (11), $G_{k|1, \dots, (k-1)}(x^k, x^1, \dots, x^{k-1})$ is a strictly increasing continuous function from \mathbb{R} to $[0, 1]$ for every (x^1, \dots, x^{k-1}) . Thus $G_{k|1, \dots, (k-1)}^{-1}$ is well defined.

So from (11) we get by induction :

$$\begin{aligned} X^k &= G_{k|1,\dots,(k-1)}^{-1}(U^k, X^1, \dots, X^{k-1}) \\ &= G_{k|1,\dots,(k-1)}^{-1}(U^k, X^1(U^1), \dots, X^{k-1}(U^1, \dots, U^{k-1})) \end{aligned}$$

noted for simplicity :

$$X^k = G_{k|1,\dots,(k-1)}^{-1}(U^k, U^1, \dots, U^{k-1}) \tag{12}$$

So for $h \in L^2(X^1, \dots, X^k)$, $k \in [0, p]$:

$$\begin{aligned} h(X^1, \dots, X^k) &= h(X^1(U^1), X^2(U^1, U^2), \dots, X^k(U^1, \dots, U^k)) \\ &= l(U^1, \dots, U^k) \end{aligned}$$

where l is a square integrable function.

Thus by recurrence, we see that for all $k \in [1, p]$:

$$L^2(X^1, \dots, X^k) = L^2(U^1, \dots, U^k) \tag{13}$$

The space $L^2(U^1, \dots, U^k)$ is splitted into two spaces : $L^2(U^1, \dots, U^{k-1}) \oplus L^2(U^k)$, where (U^1, \dots, U^{k-1}) and U^k are independent.

So $L^2(X^1, \dots, X^k)$ is splitted into two spaces : $L^2(X^1, \dots, X^{k-1}) \oplus L^2(U^k)$.

As a consequence, we are in the situation of the part 2.2 :

Lemma 3. *The space $L^2(U^1, \dots, U^k) = L^2(X^1, \dots, X^k)$ for every k and so we have the equality of conditional expectations :*

$$\mathbf{E}(\cdot | X^1, \dots, X^k) = \mathbf{E}(\cdot | U^1, \dots, U^k)$$

Now :

$$Y = f(X^1, \dots, X^p) \tag{14}$$

$$= f(X^1(U^1), \dots, X^p(U^1, \dots, U^p)) \tag{15}$$

$$= \tilde{f}(U^1, \dots, U^p) \tag{16}$$

and by the lemma (3) :

$$S^{U^1} = S^{X^1} \tag{17}$$

where S^{U^1} is given by $\frac{\mathbf{Var}(\mathbf{E}(\tilde{f}(U^1, \dots, U^p) | U^1))}{\mathbf{Var}(Y)}$, \tilde{f} given by (16),

Thus for a specific ordering, say (X^1, \dots, X^p) we can only compute the sensibilities as $(S^{X^1}, S^{X^1 X^2}, \dots, S^{X^1 \dots X^p})$.

Thus if we want to compute all first order sensitivity indices S^{X^k} with $k = 1, \dots, p$, we choose an order among the $(p - 1)!$ possible beginning by X^k .

If we want all second order sensitivity indices we need exactly $\frac{p(p-1)}{2}$ different orders and so on, if we want $S^{X^{j_1}, \dots, X^{j_q}}$ we have to take an order beginning by (j_1, \dots, j_q) .

Example : Let $p = 2$. The input vector is assumed to be : $X = (X^1, X^2)$ defined by a uniform distribution on the triangle:

$$D = \begin{cases} 0 \leq x^1, x^2 \leq 1 \\ x^1 + x^2 \leq 1 \end{cases}$$

and the input-output model is :

$$f(X^1, X^2) = X^1 + X^2$$

We compare the index S^{X^1} calculated directly and the index S^{U^1} using the transformation T^{-1} such that : $(U^1, U^2) \xrightarrow{T^{-1}} (X^1, X^2)$.

To avoid all confusions with the indices, the square of variables is denoted in brackets $(.)^2$.

Let us calculate U^1 and U^2 and deduce X^1 and X^2 :

$$U^1 = G(X^1) = 2X^1 - (X^1)^2 \text{ implies } X^1 = 1 - \sqrt{1 - U^1}$$

$$U^2 = G_{X^2|X^1}(X^2) = \frac{X^2}{1 - X^1} \mathbb{1}_{[0, 1 - X^1]} \text{ so } X^2 = U^2 \sqrt{1 - U^1}$$

The density of $(X^1 + X^2)$ is $2(x^1 + x^2) \mathbb{1}_{0 < x^1 + x^2 < 1}$ and its variance $\frac{1}{18}$.

$\mathbf{E}(X^1 + X^2 | X^1) = \frac{1 + X^1}{2}$ and $\mathbf{Var}\left(\frac{1 + X^1}{2}\right) = \frac{1}{72}$, thus the indices values are :

$$S^{X^1} = S^{X^2} = 1/4$$

Now if we use the function $\tilde{f}(U^1, U^2) = 1 - \sqrt{1 - U^1} + U^2 \sqrt{1 - U^1}$:

$$\mathbf{E}(\tilde{f}(U^1, U^2) | U^1) = 1 - \sqrt{1 - U^1} + \frac{1}{2} \sqrt{1 - U^1} = 1 - \frac{1}{2} \sqrt{1 - U^1}$$

$$\mathbf{Var}\left(1 - \frac{1}{2} \sqrt{1 - U^1}\right) = 1/72$$

So : $S^{U^1} = S^{X^1}$.

But we can notice that $S^{U^2} \neq S^{X^2}$. To calculate S^{X^2} we need to reorder the variables.

The Hoeffding decomposition of \tilde{f} is obtained by centering $\sqrt{1 - U^1}$ and U^2 :

$$\tilde{f}(U^1, U^2) = \frac{1}{3} - \frac{1}{2} \left(\sqrt{1 - U^1} - \frac{2}{3} \right) - \frac{2}{3} \left(U^2 - \frac{1}{2} \right) + \left(U^2 - \frac{1}{2} \right) \left(\sqrt{1 - U^1} - \frac{2}{3} \right)$$

S^{U^1, U^2} can be interpreted as an interaction sensibility between U^1 and U^2 . The sensitivity with respect to X^2 depends on U^1 and U^2 so there is no obvious interpretation in terms of (X^1, X^2) of this interaction.

2.3.1. Pick and Freeze estimation and conditional quantile method

Starting from $Y = f(X^1, \dots, X^p)$ we have built a model $Y = \tilde{f}(U^1, \dots, U^p)$ using a transformation $T : X \mapsto U$ from \mathbb{R}^p to \mathbb{R}^p .

The use of conditional quantile functions is the most general key to simulate any random vector. If $G_{k|1, \dots, k-1}$ is known for every k , it can be possible to simulate $X^{(i)}$ using the simulation of $U^{(i)} = (U^{1,(i)}, \dots, U^{p,(i)})$ then solving this equation recursively :

$$G_{k|1, \dots, k-1}(x^{k,(i)}, X^{1,(i)}, \dots, X^{k-1,(i)}) = U^{k,(i)}$$

when the solution is $X^{1,(i)}$.

So we start with a simulation of U to obtain a simulation for X .

Algorithm 1 Quantile method

Require: $N, G_{i|1, \dots, (i-1)}$ for all $i = 1, \dots, p$

- 1: $U = \text{matrix}(0, \text{ncol} = N, \text{nrow} = p)$; $U' = \text{matrix}(0, \text{ncol} = N, \text{nrow} = p)$
- 2: $X = \text{matrix}(0, \text{ncol} = N, \text{nrow} = p)$; $X' = \text{matrix}(0, \text{ncol} = N, \text{nrow} = p)$
- 3:
- 4: $U \sim \mathcal{U}$ {Simulation of a sample of uniform variables of size N}
- 5: $U' \sim \mathcal{U}$ {Simulation of a second sample of uniform variables of size N}
- 6:
- 7: $X[1,] = \text{Solve}(G_1(X) = U[1,])$
- 8: $X'[1,] = X[1,]$ { $X[1,]$ is frozen}
- 9: **for** $i = 2$ to p **do**
- 10: **for** $j = 1$ to N **do**
- 11: $X[i, j] = \text{Solve}(G_{i|1, \dots, (i-1)}(X) = U[i, j])$
- 12: $X'[i, j] = \text{Solve}(G_{i|1, \dots, (i-1)}(X) = U'[i, j])$
- 13: **end for**
- 14: **end for**
- 15:
- 16: $Y = \eta(X)$ {Sample of size N with the variable $X^1[1,]$ frozen}
- 17: $Y^X = \eta(X')$ {Sample of size N with the variable $X^1[1,]$ frozen}
- 18: **return** Y, Y^X

Thus the algorithm to estimate S^{X^1} is as follows:

1. Simulate $(p - 1)$ -samples $(U^{2,(i)})', \dots, (U^{p,(i)})'$, $i = 1, \dots, N$ of uniform independent variables.
2. **Main step:** Solve equations recursively :

$$G_{k|1, \dots, k-1}((x^{k,(i)})', (x^{1,(i)})', \dots, (x^{k-1,(i)})') = (U^{k,(i)})' \quad i = 1, \dots, N$$

Let $(X^{k,(i)})'$ be the solution, $k = 2, \dots, p$; $i = 1, \dots, N$ and for $k = 1$, $G_1(X^{1,(i)}) = U^{1,(i)}$

The Newton or Quasi Newton method is easy to apply here to solve these one dimensional equations, for $G_{k|1, \dots, k-1}$ are continuous, strictly increasing functions of x^k (Habegger et al., 2010).

3. Compute $Y^{(i)}$ and $Y^{X^j, (i)}$ using as inputs $(X^{1, (i)}, \dots, X^{p, (i)})$ or $(X^{1, (i)}, (X^{2, (i)})', \dots, (X^{p, (i)})')$
4. With these outputs we can estimate \hat{S}^{X^1} using formula (6) or (7).

If we want to compute all first order sensitivity indices S^{X^k} with $k = 1, \dots, p$, we choose an order among the $(p-1)!$ possible beginning by X^k . The computation complexity of this method is higher ($O(p^2)$) than for the independence case ($O(p)$) and so the time of computation increases fastly with the dimension of the input vector.

2.4. The Gaussian case

Let X a Gaussian vector and $Y = f(X)$ a input-output model. The multi-regression of X^J onto X^J is written as :

$$X^{\bar{J}} = \Lambda X^J + W \quad (18)$$

where W is a Gaussian vector independent of X^J and Λ a $(p-q) \times q$ matrix given by :

$$\Lambda = \Gamma^{\bar{J}J} (\Gamma^{JJ})^{-1} \quad (19)$$

with $\Gamma^{\bar{J}J} = \mathbf{E}(X^{\bar{J}}(X^J)^*)$ and $\Gamma^{JJ} = \mathbf{E}(X^J(X^J)^*)$.

The Pick and Freeze method can be applied to the independent variables (X^J, W) of the model h . Indeed :

$$\begin{aligned} f(X^J, X^{\bar{J}}) &= f(X^J, \Lambda X^J + W) \\ &= h(X^J, W) \end{aligned}$$

Levy-Rosenblatt method in the Gaussian case :

For Gaussian distribution, we define sequentially (10) :

$$U^k = G(X^k | X^1, \dots, X^{k-1}) = \Phi \left(\frac{X^k + \sum_{j=1}^{k-1} (C_{kj}/C_{kk})(X^j)}{\sqrt{C/C_{kk}}} \right) \quad (20)$$

where C_{kj}^p is the cofactor of C_{kj} in C^p when C^p is the restriction of the covariance matrix $C = (C_{kj})_{k=1, \dots, p; j=1, \dots, p}$ to $1 \leq j, k \leq p$, and Φ the Gaussian repartition.

$W^k = \Phi^{-1} \left(X^k + \sum_{j=1}^{k-1} (C_{kj}/C_{kk})(X^j) \right) \sqrt{C/C_{kk}}$ is another way to get (18).

2.5. Incomplete information : copula as models and the Pick and Freeze method

When the information on the inputs is not sufficient to build a model, models are chosen in a class of models taking into account this partial information. The choice of a model is supported

by some practical properties. We detail here a class of model which have good properties for sensitivity. These models are the copula one, well adapted to the Pick and Freeze method.

Copulas are used to build models when the information extracted from the data is incomplete and reduced to the p repartitions F^1, \dots, F^p of the p random variables of interest (in our case the input of the system) (Nelsen, 2013; Sklar, 1959). The correlation of the input vector is an information, often easy to obtain and that the metamodel has to incorporate. This led to correlation constrained copulas. Other constraint can be incorporated.

We show that the Pick and Freeze method is adapted to this kind of model and also that sensitivity is not always well associated to correlation.

Suppose that we know F^1, \dots, F^p the distribution functions of $X = (X^1, \dots, X^p)$ and R its correlation.

Definition 1. Let (F^1, \dots, F^p, R) and R a correlation matrix given. We say that $\mathcal{N}_p(0, \Upsilon)$ with Υ a correlation matrix, is a Gaussian copula for (F^1, \dots, F^p, R) if and only if :

$$\begin{cases} X^i = (F^i)^{-1} \circ \Phi(Z^i), \quad i = 1, \dots, p \\ \text{the correlation matrix of } X \text{ is } R \end{cases} \quad (21)$$

where Z^i is a Gaussian variable and ϕ the Gaussian distribution.

Once R is given, Υ can be computed.

By definition the $p \times p$ matrix $R = (R^{i,j})_{\substack{i=1,\dots,p \\ j=1,\dots,p}}$ is :

$$R^{i,j} = \mathbf{Cor}(X^i, X^j) \quad (22)$$

is given and thus we want to compute $\Upsilon = (\Upsilon^{i,j})_{\substack{i=1,\dots,p \\ j=1,\dots,p}}$ (the correlation matrix of the Gaussian vector Z) so that R is the correlation matrix of X by a calculation of the type :

$$\mathbf{E}(X^i X^j) = \frac{1}{2\pi\sqrt{1 - (\Upsilon^{i,j})^2}} \iint ((F^i)^{-1} \circ \Phi)(z^i) ((F^j)^{-1} \circ \Phi)(z^j) \exp^{-\frac{1}{2(1 - (\Upsilon^{i,j})^2)}((z^i)^2 - 2\Upsilon^{i,j}z^i z^j + (z^j)^2)} dz^i dz^j \quad (23)$$

From now, in order to simplify our example we consider uniform variables (X^1, \dots, X^p) .

In this case R and Υ are linked by a simple form Biller and Ghosh (2006):

$$\Upsilon^{i,j} = 2 \sin\left(\frac{\pi R^{i,j}}{6}\right) \quad (24)$$

Note that $|\Upsilon_{i,j}| = 1 \Leftrightarrow |R_{i,j}| = 1$, $\Upsilon_{i,j} = 0 \Leftrightarrow |R_{i,j}| = 0$.

Numerically :

$$\Upsilon^{i,j} = 1.047R^{i,j} - 0.047(R^{i,j})^3 \quad (25)$$

is a good approximation. Thus this correspondence between $R^{i,j}$ and $\Upsilon^{i,j}$ is well defined.

R is given as a correlation matrix but nothing says that (in p dimensions) $\Upsilon = (\Upsilon^{i,j})_{\substack{i=1,\dots,p \\ j=1,\dots,p}}$ is a correlation matrix (positive type matrix). This point is related to the following definition.

Definition 2. Let $(F^i)_{i=1,\dots,p}$ a family of marginal distributions and R a correlation matrix. We say that $((F^i)_{i=1,\dots,p}, R)$ is feasible as a Gaussian copula if and only if there is a Gaussian vector $Z = (Z^i)_{i=1,\dots,p}$ whose correlation matrix is Υ , satisfying :

$$X^i = (F^i)^{-1}(\Phi(Z^i)) \text{ for } i = 1, \dots, p \quad (26)$$

where X has R as correlation.

We don't discuss here the problem of feasibility. If Υ is not positive it is often possible to find correlation matrices "close to" Υ (Ghosh and Henderson, 2006).

Let us now show how to apply Pick and Freeze method for a copula model.

For $Z^i = \Phi^{-1} \circ F^i(X^i)$ is a monotone bijection for all $J \subset \{1, \dots, p\}$ and every output

$$\tilde{f}(Z^1, \dots, Z^p) = f(\Phi^{-1} \circ F^1(X^1), \dots, \Phi^{-1} \circ F^p(X^p)) \quad (27)$$

Then :

$$\mathbf{E}(\tilde{f}(Z^1, \dots, Z^p) | Z^J) = \mathbf{E}(f(X^1, \dots, X^p) | X^J) \quad (28)$$

Thus we can use the Pick and Freeze method on the Gaussian copula to compute all the sensitivities for the model $f(X)$.

In the following example, sensitivities of different copulas chosen as model taking into account the partial information are compared in order to illustrate that for the same constraint of correlation, the sensitivities can be different.

2.5.1. Example on the Ishigami model

Sensitivity is estimated by the Pick and Freeze method in all the cases. We have selected *the model of Ishigami*, a classical toy model in sensitivity and optimisation studies defined as follows :

$$Y = \sin(X^1) + 7 \sin(X^2) + 0.1(X^3)^4 \sin(X^1) \quad (29)$$

(X^1, X^2, X^3) have a uniform distribution with support $[-\pi, \pi]$.

The correlation matrix given is : $\begin{pmatrix} 1 & 0 & \rho \\ 0 & 1 & 0 \\ \rho & 0 & 1 \end{pmatrix}$ and X^2 is supposed (to simplify) to be independent

of the pair (X^1, X^3) . We consider two distributions :

- case 1 : the Gaussian copula
- case 2 : the copula given by this density probability :

$$f_\alpha(x^1, x^2, x^3) = \frac{1}{4\pi^2} \mathbb{1}_{[-\pi, \pi]^2}(x^1, x^3) + \alpha x^1 x^3 \quad (30)$$

f_α is a density probability if $|\alpha| \leq \frac{1}{4\pi^2}$. This condition implies that :

$$\rho = \mathbf{E}(X^1 X^2) = \frac{4\pi^3 \alpha}{9} \text{ thus } |\rho| \leq \frac{\pi}{9}.$$

The sensitivity values are calculated by applying the method of conditional quantile and the results are discussed with respect to different ρ values.

Case 1 : Gaussian copula :

$Z^i, i = 1, 2, 3$ is defined by :

$$X^i = \pi(2\Phi(Z^i) - 1)$$

where X^i is uniform on $[-\pi, \pi]$. The correlation ρ' of Z^1, Z^3 is given by :

$$\rho' = 2 \sin\left(\frac{\pi\rho}{6}\right)$$

Following our previous results in section 2.4 we write the Ishigami model with independent Gaussian variables Z^1, Z^2, W .

Z^1 is defined by : (26) and W by the regression :

$$Z^3 = \rho'Z^1 + \sqrt{1 - (\rho')^2}W \text{ with } W \sim \mathcal{N}(0, 1)$$

Thus the input-output system is now :

$$Y = \sin(\pi(2\Phi(Z^1) - 1)) + 7 \sin(\pi(2\Phi(Z^2) - 1)) + 0.1 \left(\pi(2\Phi(\rho'Z^1 + \sqrt{1 - \rho'^2}W) - 1) \right)^4 \sin(\pi(2\Phi(Z^1) - 1)) \tag{31}$$

As X^1 and X^2 are independent we know that $S^{X^1} = S^{Z^1}$ and $S^{X^2} = S^{Z^2}$.

To compute S^{X^3} the expression (31) of Y .

Results related to ρ are plotted in figure : 1.

$\rho = 0$ corresponds to the case of independent variables.

Case 2 : f_α copula :

The conditional quantile method is used to calculate the indices. First as X^1 and X^2 are independent variables we have :

$$U^1 = \frac{X^1/\pi + 1}{2} \qquad U^2 = \frac{X^2/\pi + 1}{2} \tag{32}$$

U^1 and U^2 are uniform independent variables.

U^3 is defined such that :

$$U^3 = F_{X^3|X^1}(X^3) = \frac{1}{2\pi}(X^3 + \pi) + \pi\alpha X^1 \frac{(X^3)^2 - \pi^2}{2}$$

U^3 is a uniform variable independent from U^1 and U^2 .

Thus :

$$X^3 = \frac{-1 + \sqrt{1 - 8\pi^2\alpha X^1(1 - \pi^2\alpha X^1 - 2U^3)}}{4\alpha\pi X^1} \tag{33}$$

TABLE 1. Sensitivity for Gaussian copula and f_α copula, for different values of ρ

	ρ	S^{X^1}	S^{X^2}	S^{X^3}
copula f_α	0	0.31	0.44	0
copula f_α	$\pi/9$	0.36	0.40	0.46
Gaussian copula	0	0.31	0.44	0
Gaussian copula	$\pi/9$	0.30	0.50	0.08

The canonical formula for the Ishigami and the input given by the α -copula is obtained by the substitution in (29). With this order 1, 2, 3 we can calculate the indices $S^{X^1}, S^{X^2}, S^{X^1X^2}$ (X^1 and X^2 are independent in this case). If we want to compute S^{X^3} we have to resume our work choosing the order (3, 2, 1) for example :

$$U^1 = \frac{X^3/\pi + 1}{2} \quad U^2 = \frac{X^2/\pi + 1}{2} \quad (34)$$

and thus U^3 is defined by :

$$U^3 = \frac{1}{2\pi}(X^1 + \pi) + \pi\alpha X^3 \frac{(X^1)^2 - \pi^2}{2}$$

The results are given in figure 1. We can only compare the results for $0 \leq |\rho| \leq \frac{\pi}{9}$, for instance in table 1

These results show that the practitioner has to be cautious with the use of models with incomplete information when sensitivities are computed. Correlation does not give, in any case, a very good information on dependences when we compute the sensitivity of different inputs. For a same correlation we get different copulas, which gives very different sensitivity results.

3. Sensitivity for vectorial stochastic process inputs

3.1. Sensitivity for vectorial stochastic process inputs and memories

Suppose that we consider an input-output system :

$$Y_t = f_t(X_t, \dots, X_{t-k}, \dots, X_0), \quad t \in \mathbb{N} \quad (35)$$

In the following section the following notations are chosen :

- $(X_t)_{t \in \mathbb{Z}} = (X_t^1, \dots, X_t^p)_{t \in \mathbb{Z}}$ a stochastic vectorial process of dimension p .
- If $J = \{j_1, \dots, j_q\} \subset \{1, \dots, p\}$ and $\bar{J} = \{1, \dots, p\} \setminus J$, $X_t^J = (X_t^{j_1}, \dots, X_t^{j_q})$ is a process of dimension q
- $\mathbb{X}_{t,t-k} = (X_t, X_{t-1}, \dots, X_{t-k})$ a $p \times (k+1)$ matrix.
- $\Gamma_{s,v}^{i,j} = \mathbf{E}(X_s^i X_v^j)$
- $\Gamma_{[s,t],v}^{i,j} = \{\Gamma_{u,v}^{i,j}, s \leq u \leq t\}$ a vector of dimension $(t-s+2)$ whose generic term is $\Gamma_{u,v}^{i,j}$.

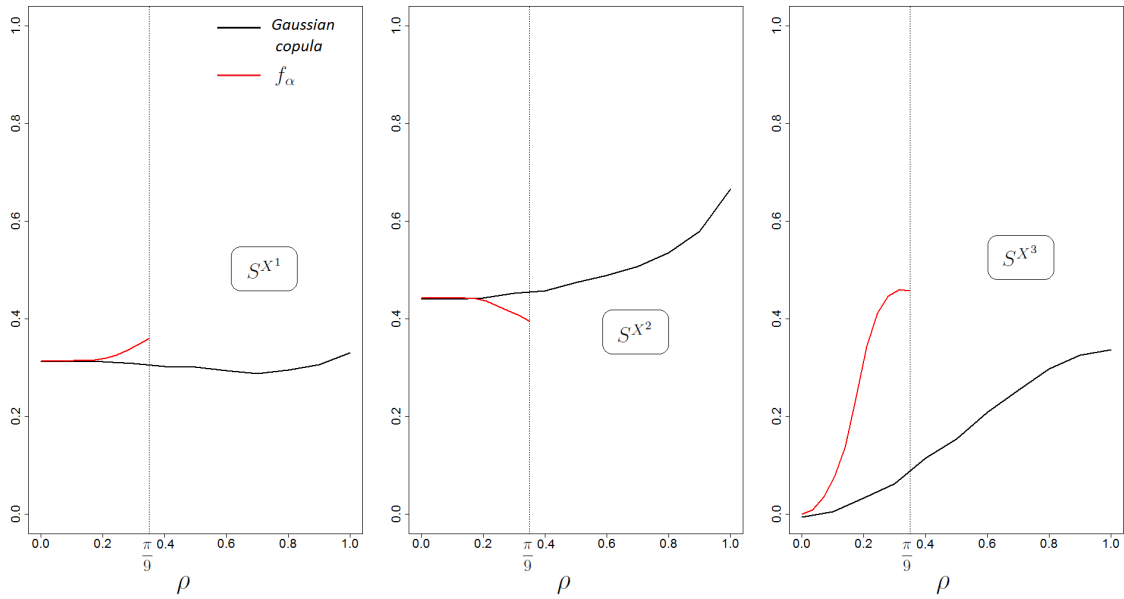


FIGURE 1. Sensitivity indices for different values of ρ^J applied to the Ishigami model for the Gaussian copula and the f_α copula

The output process at time t can depend on its past instants Y_{t-h} and also the past instants of the input process X_{t-k} . Due to this phenomenon of memory, it is not wise to calculate the sensitivity at time t in relation to the input at time t but to calculate the sensitivity with respect to $\mathbb{X}_{t,t-k}^J$.

Definition 3. k -sensitivity

The k -sensitivity is the Sobol index of Y_t with respect to $\mathbb{X}_{t,t-k}^J$ for $0 \leq k < t$.

It is defined by :

$$S_{t,k}^{X^J} = \frac{\text{Var}(\mathbf{E}(Y_t | \mathbb{X}_{t,t-k}^J))}{\text{Var}(Y_t)} \tag{36}$$

The index is measured as the ratio of the conditional expectation of Y_t when $(X_t^J, \dots, X_{t-k}^J)$ is fixed on the total variance Y_t .

$\text{Var}(\mathbf{E}(Y_t | \mathbb{X}_{t,t-k}^J))$ is the square norm of the projection of Y_t onto the space $L^2(\mathbb{X}_{t,t-k}^J)$ defined by $L^2(\mathbb{X}_{t,t-k}^J) = \{\phi(X_t^J, \dots, X_{t-k}^J), \mathbf{E}(\phi^2) \leq \infty\}$.

We can noticed that :

$$L^2(\mathbb{X}_{t,t-k-1}^J) \subset L^2(\mathbb{X}_{t,t-k}^J)$$

So :

$$0 \leq S_{t,k-1}^{X^J} \leq S_{t,k}^{X^J} \leq 1$$

The instantaneous sensitivity corresponds to $k = 0$.

Definition 4. *Total sensitivity :*

The total sensitivity is the sensitivity taking into account the whole past of the input X_t^J . It is thus defined as :

$$S_t^{X^J} = \frac{\mathbf{Var}(\mathbf{E}(Y_t | \mathbb{X}_t^J))}{\mathbf{Var}(Y_t)}. \quad (37)$$

So, we have :

$$S_{t,k}^{X^J} \leq S_t^{X^J} \quad 0 \leq k \leq t$$

$S_{t,k}^{X^J}$ is an increasing function of k . When k tends towards t $S_{t,k}^{X^J}$ converges to $S_t^{X^J}$.

In practice we choose k as the value from which the index $S_{t,k}^{X^J}$ does not increase in a significant manner. This heuristic value k is called *useful memory* in terms of sensitivity. So the definition is :

Definition 5. *Let $\varepsilon > 0$ fixed. The ε -useful memory is defined as :*

$$k_\varepsilon = \inf \left\{ k \geq 0, S_t^{X^J} - S_{t,k}^{X^J} \leq \varepsilon \right\} \quad (38)$$

In applications, ε is of course chosen considering the fit quality of the input and also the statistical errors made when estimating $S_k^{X^J}$.

3.2. Stationary case

The input-output system (X_t, Y_t) defines a stochastic process with values in $\mathbb{R}^p \times \mathbb{R}$. We consider now the case where this process $(X_t, Y_t)_{t \in \mathbb{N}}$ is stationary. This implies that $(Y_t)_{t \in \mathbb{N}}$ and $(X_t)_{t \in \mathbb{N}}$ are stationary stochastic processes.

We consider the stationary case where

$$Y_t = f(X_t, \dots, X_0, X_{-1}, \dots) \quad (39)$$

and the special case :

$$Y_t = f(X_t, \dots, X_{t-h})$$

h is the memory. h is fixed and f non depending on t

The total sensitivity is given by :

$$S_{t,\infty}^{X^J} = \frac{\mathbf{Var}(\mathbf{E}(Y_t | \{X_s, -\infty \leq s \leq t\}))}{\mathbf{Var}(Y_t)} \quad (40)$$

But by shift invariance $S_{t,\infty}^{X^J}$ does not depend on t :

$$S_k^{X^J} \leq S_\infty^{X^J} \quad (41)$$

In [Grandjacques et al. \(2015\)](#), the following intuitive relation is proved for some processes (X_t, Y_t) :

$$\lim_{k \rightarrow \infty} S_k^{X^J} = S_\infty^{X^J} \quad (42)$$

and if the process : $Y_t^* = f_t(X_t, \dots, X_0) = f(X_t, \dots, X_0, 0, \dots)$ associated to the stationary process Y_t :

$$\lim_{t \rightarrow +\infty} \lim_{k \rightarrow t} \frac{\text{Var}(\mathbf{E}(Y_t^* | \{X_s, -\infty \leq s \leq t\}))}{\text{Var}(Y_t^*)} = S_\infty^{*,X^J}$$

Then :

$$S_\infty^{*,X^J} = S_\infty^{X^J} \quad (43)$$

3.3. Stochastic Gaussian processes

We can apply to Gaussian processes the Pick and Freeze method introduced in section 2.4. To compute for instance $S_{t,k}^{X^J}$ we use the decomposition :

$$\mathbb{X}_{t,t-k}^J = \Lambda_{[t-k,t],[t-k,t]} \mathbb{X}_{t,t-k}^J + \mathbb{W}_{t,t-k} \quad (44)$$

with $\Lambda_{t,t}$ given by :

$$\Lambda_{[t-k,t],[t-k,t]} = \left(\Gamma_{[t-k,t],[t-k,t]}^{JJ} \right)^{-1} \Gamma_{[t-k,t],[t-k,t]}^{J\bar{J}} \quad (45)$$

$\Gamma_{[t-k,t],[t-k,t]}^{JJ}$ is invertible.

For each t we apply the Pick and Freeze method to :

$$Y_t = f_t(\mathbb{X}_t^J, \Lambda_{[t-k,t],[t-k,t]} \mathbb{X}_{t,t-k}^J + \mathbb{W}_t) \quad (46)$$

$$= g_t(\mathbb{X}_t^J, \mathbb{W}_t) \quad (47)$$

with \mathbb{X}_t^J and \mathbb{W}_t independent vectors.

3.3.1. Example of toy models for Gaussian inputs

We study two stationary non linear toy models given by:

$$Y_t = 0.5Y_{t-1} + 0.3X_t^1 X_t^2 \quad (48)$$

$$Y_t = X_t^1 X_t^2 - \arctan(X_t^2) \quad (49)$$

X_t^1, X_t^2 is a VAR(1) stationary process given by:

$$\begin{pmatrix} X_t^1 \\ X_t^2 \end{pmatrix} = \begin{pmatrix} 0.1 & 0.4 \\ 0.8 & 0.2 \end{pmatrix} \begin{pmatrix} X_{t-1}^1 \\ X_{t-1}^2 \end{pmatrix} + \omega_t \quad (50)$$

where ω_t is a stationary Gaussian noise of covariance matrix $\Theta = \begin{pmatrix} 0.1 & 0 \\ 0 & 0.1 \end{pmatrix}$.

Indices are estimated with samples of size $N = 10000$. Results are given in figures 2 and 3.

All the indices converge quickly to a constant. The useful memory defines for $\varepsilon = 0.02$ is different according to the model and the variable. It is $k = 4$ for the model (48) and $k = 2$ (variable X^1) or $k = 3$ (variable X^2) for the model (49).

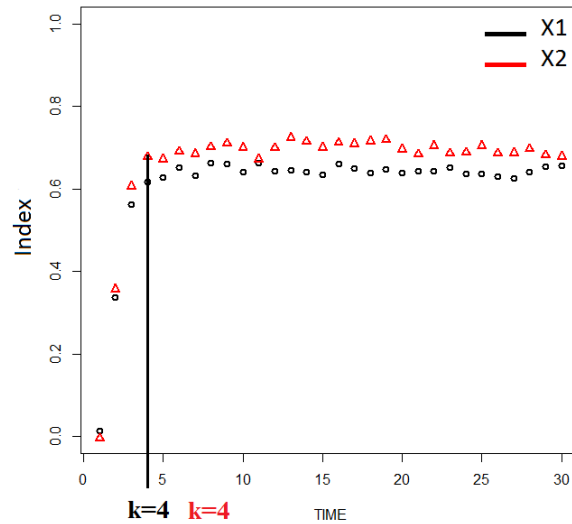


FIGURE 2. Plot of Sobol indices applied to model (48) in function of k .

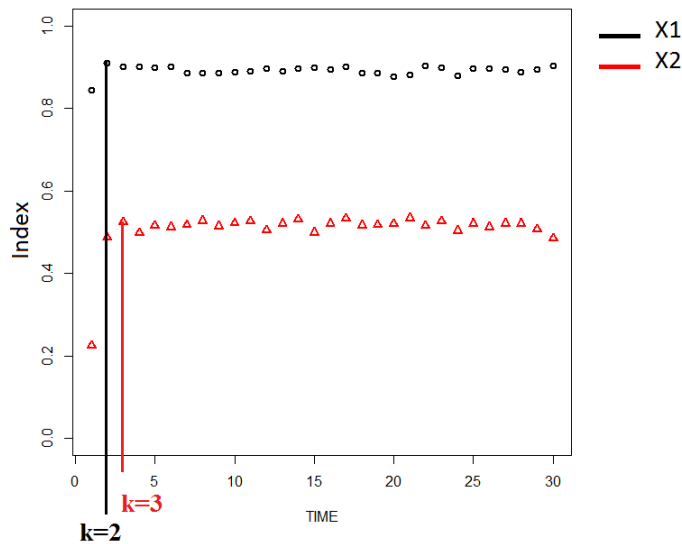


FIGURE 3. Plot of Sobol indices applied to model (49) in function of k .

3.4. Sensitivity and partial information for non Gaussian vectorial process inputs

3.4.1. Non Gaussian input process

A lot of variables have non Gaussian distribution. For example some climatic variables (temperature, wind), heating or energy source variables are usually bounded. If we want to study the impact of extreme cold or even a wave of heat on the indoor temperature, we cannot use Gaussian variables because they have too heavy tails. It is the same for phenomena which present two main values. The density in this case is bimodal.

Starting from data, the construction of a non Gaussian stochastic process is difficult and all the more in a multivariate context. Thus, as often in these situations, the choice of the model is constrained by some criteria. It must take into account some of the information which can be extracted from the data and which seem the most important to the practitioner. These informations can be qualitative or quantitative or mixed. For instance these informations concern :

- the marginal distribution of the inputs and the correlations at a fixed time between the p -components
- the time dependence structure

For marginal distributions, qualitative information is, for instance, the number of modes (important regimes). The semi-qualitative information is for instance the boundedness of the support of the distribution. The quantitative informations can be the mean, the variance, the skewness, the kurtosis. Information on dependence can be translated in terms of some correlation coefficients or in terms of Markovian properties. Once these properties extracted or estimated from the data we have to choose the input model and to be sure concerning our goals that it allows to compute sensitivities with a quite good approximation. This last point is of course an important constraint.

The most classical problem is the following, which can be set in terms of constrained copulas : suppose we want to build a stationary input process X_t with fixed marginals (F^1, \dots, F^p) and some fixed correlations for instance : $\mathbf{Cor}(X_t, X_t)$ and $\mathbf{Cor}(X_t, X_{t-1})$. These correlations are in fact the correlations estimated with the data. The fixed marginals can be estimated using the data in a parametric family, large enough to take into account qualitative and quantitative properties according to the practitioner's experience on sensitivity.

Thus we need to choose a family parametrized for instance by the first four moments of the distribution, flexible enough to allow properties as : boundedness, bimodality, light and heavy tails. This is the case of some families such as Pearson or Johnson one [Johnson \(1949\)](#). Their properties in relation with our work are detailed in the appendix.

Let the correlation matrices R_q defined by :

$$R_q = \mathbf{Cor}(X_t, X_{t-q}) \text{ for } 0 \leq q \leq Q$$

Definition 6. Let $R = \{R_q, 0 \leq q \leq Q\}$ a correlation matrix given. We say that it is a problem (F, R) feasible if there is a stationary stochastic process X_t such that its p marginals are given by $F = (F^1, \dots, F^p)$ and the $Q+1$ first correlations are given by R .

Until to day there are only partial results on this problem [Cario and Nelson \(1996\)](#). The most

usual way is to try to build a model associated to a Gaussian one (modified to be feasible) and which moreover allows to compute sensitivity indices.

The given information on X_t is (F^1, \dots, F^p) the p marginal distributions of the stationary process X_t and the correlation matrix $R = \{R_q, 0 \leq q \leq Q\}$ with $R_q = (R_q^{i,j})_{1 \leq i,j \leq p}$. Let $Z_t \in \mathbb{Z}$ a Gaussian stationary process defined by :

$$Z_t^i = (\Phi^{-1} \circ F^i)(X_t^i) \quad (51)$$

We look for the correlation matrix $\Upsilon = (\Upsilon_0, \dots, \Upsilon_Q)$ of Z_t^i in order that for $0 \leq q \leq Q$:

$$\Upsilon_q^{i,j} = \mathbf{Cor} \left(((F^i)^{-1} \circ \Phi)(Z_t^i), ((F^j)^{-1} \circ \Phi)(Z_{t-q}^j) \right) \quad (52)$$

This can be easily done by computing integrals analogous to (23), taking $\Upsilon_q^{i,j}$ instead of $\Upsilon^{i,j}$.

Thus Υ is now fixed. If Υ is a positive definitive matrix, it gives the first Q correlation of the process Z_t . We discuss later the case when Υ is not positive definite.

The class of stationary Gaussian $\text{VAR}(Q)$ processes can be associated to Υ . This class has the property to allow easy computations of sensitivities by the Pick and Freeze method.

Let

$$Z_t = A_1 Z_{t-1} + \dots + A_Q Z_{t-Q} + \omega_t$$

with $\mathbf{E}(\omega_t \omega_t^*) = \Theta$, ω_t being a Gaussian white noise.

A_1, \dots, A_Q can be quite easily computed from $(\Upsilon_0, \dots, \Upsilon_Q)$ and Θ .

Indeed all $\text{VAR}(Q)$ processes can be rewritten as a $\text{VAR}(1)$ process :

$$V_t = B V_{t-1} + \varepsilon_t$$

$$\text{with } V_t = (Z_t, Z_{t-1}, \dots, Z_{t-Q+1})^* \text{ and } B = \begin{pmatrix} A_1 & A_2 & A_3 & \dots & A_Q \\ I_p & 0 & 0 & \dots & 0 \\ 0 & I_p & \ddots & \ddots & \vdots \\ \vdots & \ddots & I_p & 0 & 0 \\ 0 & \dots & 0 & I_p & 0 \end{pmatrix} \text{ and } \varepsilon = (\omega_t, 0, \dots, 0)^*.$$

Z_t a $\text{VAR}(Q)$ Gaussian process. Suppose to simplify, that $Q = 1$:

$$A_1 = \Upsilon_1 \Upsilon_0^{-1} \quad (53)$$

$$\mathbf{E}(\omega_t \omega_t^*) = \Upsilon_0 - \Upsilon_1 \Upsilon_0^{-1} \Upsilon_1 \quad (54)$$

Thus the process Z_t can be easily simulate.

Z_t defines a $\text{VAR}(Q)$ Gaussian process, which is the $\text{VAR}(Q)$ Gaussian copula associated to X_t by :

$$X_t^i = ((F^i)^{-1} \circ \Phi)(Z_t^i) \quad (55)$$

It may happen that Υ is not a correlation matrix (matrix not positive definite) leading to a stationary process Z_t . There are, until to day, only empirical methods (Biller and Nelson, 2005, 2003) to

overpass this obstacle. The most efficient is to take a smaller Q (in general Q is chosen by an Akaike criterion) and to change R slightly.

Now to compute the sensitivity, we use the following basic facts as previously in the static case : $T^i = (F^i)^{-1} \circ \Phi$ is a monotone function.

Thus we have the equality of the L^2 spaces :

$$L^2(Z_{t_1}^{j_1}, \dots, Z_{t_n}^{j_q}) = L^2(X_{t_1}^{j_1}, \dots, X_{t_n}^{j_q}) \tag{56}$$

for every (t_1, \dots, t_q) and (j_1, \dots, j_q) for every q .

Let Y_t the output of the system :

$$Y_t = f(X_t, X_{t-1}, \dots, X_{t-k})$$

for instance then :

$$\begin{aligned} Y_t &= f(T^{-1}(Z_t), \dots, T^{-1}(Z_{t-k})) \\ &= \tilde{f}(Z_t, \dots, Z_{t-k}) \end{aligned}$$

where $T(Z_t) = ((F^i)^{-1} \circ \Phi(Z_t^i))_{i=1, \dots, p}$ Thus :

$$\mathbf{E} \left(\mathbf{E} (Y_t | X_t^1, \dots, X_{t-s}^1)^2 \right) = \mathbf{E} \left(\mathbf{E} (Y_t | Z_t^1, \dots, Z_{t-s}^1)^2 \right) \tag{57}$$

for every s .

We can compute : $\mathbf{E} (Y_t | X_t^1, \dots, X_{t-s}^1)$ using the Pick and Freeze method already defined for the Gaussian process $(Z_t)_{t \in \mathbb{Z}}$ in section 3.3.

Let us give an example.

3.4.2. Example

We study a non linear stationary model given by:

$$Y_t = 0.5Y_{t-1} - 0.2 \sin(U_t^2) + 0.2U_t^1 \tag{58}$$

where U_t is a stationary process with uniform components. We suppose that the correlation matrices are :

$$\begin{aligned} R_0 &= \begin{pmatrix} R_0^{11} & R_0^{12} \\ R_0^{12} & R_0^{22} \end{pmatrix} = \begin{pmatrix} 0.063 & -0.011 \\ -0.023 & 0.061 \end{pmatrix} \\ R_1 &= \begin{pmatrix} R_1^{11} & R_1^{12} \\ R_1^{12} & R_1^{22} \end{pmatrix} = \begin{pmatrix} -0.003 & 0.023 \\ 0.047 & 0.003 \end{pmatrix} \end{aligned}$$

The correlation matrices of the Gaussian process must verify : $\Upsilon_k^{ij} = 2 \sin(\frac{\pi R_k^{ij}}{6})$, where Υ is the correlation of the process such as : $Z_t^i = \Phi^{-1}(U_t^i)$

So :

$$\Upsilon_0 = \begin{pmatrix} 0.067 & -0.012 \\ -0.024 & 0.064 \end{pmatrix}$$

$$\Upsilon_1 = \begin{pmatrix} -0.029 & 0.024 \\ 0.049 & 0.003 \end{pmatrix}$$

One of the corresponding Gaussian processes might be :

$$Z_t = \begin{pmatrix} 0.1 & 0.40 \\ 0.80 & 0.20 \end{pmatrix} Z_{t-1} + \omega_t$$

where ω_t is a Gaussian noise of covariance matrix $\Theta = \begin{pmatrix} 0.1 & 0 \\ 0 & 0.1 \end{pmatrix}$

So to apply the Pick and Freeze method, the model used is :

$$Y_t = 0.5Y_{t-1} - 0.2 \sin(\Phi(Z_t^2)) + 0.2\Phi(Z_t^1)$$

where Φ is the distribution function and Z a Gaussian process defined as previously. To separate the variables, we use the method developed in section 3.3.

The results are present in figure 4.

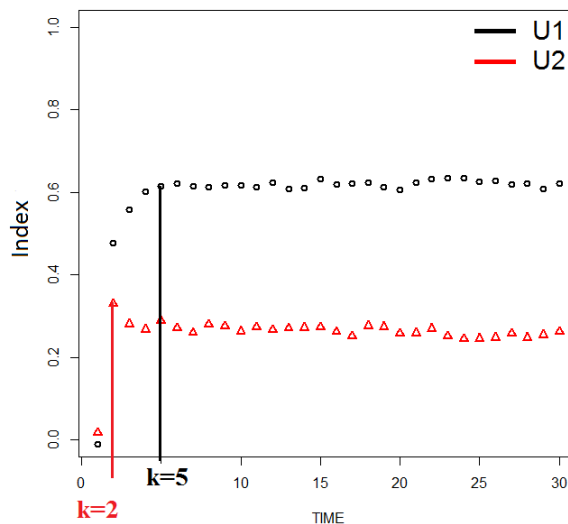


FIGURE 4. Plot of Sobol indices applied to model (58)

4. Conclusion

We give a general framework for computing sensitivities for dependent inputs. In the static case, for dependent inputs the definition of the sensitivity indices remains the same as for independent inputs. But in the dynamic case, in order to take in consideration the temporal dependence, the definition is slightly modified. We take into account the notion of memory related to sensitivity, "memory" being used in a sense linked to physical inertia. The useful memory is the instant when the index does not change significantly.

We study transformations of the input whose image is an input with independent components and for which some conditional expectations are invariant. This property allows to use the Pick and Freeze method to get the sensitivities.

When the inputs are Gaussian we first consider the static case, then the dynamic case and the different ways to solve the problem.

When the inputs are not Gaussian, in the static case we use the conditional quantile functions. They are a nodal point for the sensitivity studies and simulations. This method is the same as the one basic used to simulate random vectors in general. The output Y takes the canonical form $Y = \tilde{f}(U^1, \dots, U^p)$ where (U^1, \dots, U^p) are p uniform (or Gaussian) independent random variables. This canonical form allows to apply the Pick and Freeze method but also all the more or less classical methods to compute sensitivity starting from Hoeffding formula.

We have to take precautions with the order in which we calculate the index. When we want to calculate the index of each variable we have to start with the variable listed first and then reorder the list and so on for the other variables.

In some practical case where the information is not sufficient to allow the choice of a model, we use the copulas and often the Gaussian copulas. A formula links the correlations of the non Gaussian variable and the correlations of the Gaussian variable on which we can apply the Pick and Freeze method.

In practical situations the notion of copula model has to be managed carefully. For the Ishigami example we have shown that the correlation used to represent the dependence between variables can be very weak for sensitivity studies.

In the dynamic case we use the copula model to go back to the Gaussian case. The model chosen is an extension of the Gaussian copula applied to stochastic processes. The correlations used for the model of the non Gaussian process X_t are those between X_t and X_{t-1}, X_{t-2}, \dots . These correlations define the dynamics of the process.

In the case of stochastic process inputs and sensitivity estimation the same caution is required. The specification of inputs becomes in application very important and difficult. Modelisation can be improved using some quantitative and qualitative information, essential for the practical problem and which can be extracted from the input data. For instance, for a practitioner, instead of using the complete probability distribution of the inputs, the information can be summarized by the mean, the variance, the skewness and the kurtosis of all the marginal functions of the inputs. This is the case for the copulas with Johnson (or Pearson) distribution.

Finally, we could apply the conditional quantile function method to a dynamic case but if the processes have a too important memory the computation is heavy.

5. Appendix

Processes having marginal distributions from the Johnson translation system are defined by a cumulative distribution function F_X such as:

$$F_X(x) = \Phi(\gamma + \delta f|(x - \xi)/\lambda|) \quad (59)$$

where γ and δ are shape parameters, ξ location parameter, λ a scale parameter and $f(\cdot)$ is one of the following function :

$$f(y) = \begin{cases} \log(y) & \text{lognormal family} \\ \log(y + \sqrt{y^2 + 1}) & \text{unbounded law} \\ \log\left(\frac{y}{1-y}\right) & \text{bounded law} \\ y & \text{normal family} \end{cases} \quad (60)$$

Φ being the Gaussian repartition. $(\gamma, \delta, \xi, \lambda)$ system is equivalent to the mean, variance, skewness, kurtosis one. The maximal number of modes is 2.

Let X a random variables with distribution F and Z a Gaussian normal variable such that :

$$F(X) = \Phi(Z) \quad (61)$$

equality between uniform variables.

If X is a Johnson distribution thus $Z = \gamma + \delta f\left(\frac{X - \xi}{\lambda}\right)$ or $X = \xi + \lambda f^{-1}\left(\frac{Z - \gamma}{\delta}\right)$ well defined for f is a strictly increasing function. These formula are of course simpler than (61).

Thus the construction of the model is done estimating from the data for every j ; $(f_j, \xi_j, \gamma_j, \lambda_j, \delta_j)$. We have at this stage taken into account the main qualitative features of every F^j , $j = 1, \dots, p$ (maximum likelihood can be the tool for estimation).

Acknowledgements

The authors would like to thank the reviewers for their helpful comments.

References

- Billar, B. and Ghosh, S. (2006). Multivariate input processes. *Handbooks in Operations Research and Management Science*, 13:123–153.
- Billar, B. and Nelson, B. L. (2003). Modeling and generating multivariate time-series input processes using a vector autoregressive technique. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 13(3):211–237.
- Billar, B. and Nelson, B. L. (2005). Fitting time-series input processes for simulation. *Operations Research*, 53(3):549–559.
- Cario, M. C. and Nelson, B. L. (1996). Autoregressive to anything: Time-series input processes for simulation. *Operations Research Letters*, 19(2):51–58.
- Chastaing, G. (2013). *Indices de Sobol généralisés pour variables dépendantes*. PhD thesis, Université de Grenoble.

- Chastaing, G., Gamboa, F., and Prieur, C. (2011). Generalized hoeffding-sobol decomposition for dependent variables-application to sensitivity analysis. *Arxiv preprint arXiv:1112.1788*.
- Chastaing, G. and Le Gratiet, L. (2015). Anova decomposition of conditional gaussian processes for sensitivity analysis with dependent inputs. *Journal of Statistical Computation and Simulation*, 85(11):2164–2186.
- Cukier, R., Fortuin, C., Shuler, K., Petschek, A., and Schaibly, J. (1973). Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. i. theory. *J. Chemical Physics*, 59:3873–3878.
- Da Veiga, S., Wahl, F., and Gamboa, F. (2009). Local polynomial estimation for sensitivity analysis on models with correlated inputs. *Technometrics*, 51(4):452–463.
- Gamboa, F., Janon, A., Klein, T., Lagnoux-Renaudie, A., and Prieur, C. (2013). Statistical inference for sobol pick freeze monte carlo method. *arXiv preprint arXiv:1303.6447*.
- Ghosh, S. and Henderson, S. G. (2003). Behavior of the norta method for correlated random vector generation as the dimension increases. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 13(3):276–294.
- Ghosh, S. and Henderson, S. G. (2006). Corrigendum: Behavior of the norta method for correlated random vector generation as the dimension increases. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 16(1):93–94.
- Grandjacques, M., Janon, A., Delinchant, B., and Adrot, O. (2015). Pick and freeze estimation of sensitivity indices for models with dependent and dynamic input processes. *hal preprint*.
- Habegger, A., Stahel, A., Goette, J., and Jacomet, M. (2010). An efficient hardware implementation for a reciprocal unit. In *Electronic Design, Test and Application, 2010. DELTA'10. Fifth IEEE International Symposium on*, pages 183–187. IEEE.
- Hooker, G. (2007). Generalized functional anova diagnostics for high-dimensional functions of dependent variables. *Journal of Computational and Graphical Statistics*, 16(3).
- Janon, A., Klein, T., Lagnoux, A., Nodet, M., and Prieur, C. (2013). Asymptotic normality and efficiency of two sobol index estimators. *ESAIM: Probability and Statistics*.
- Janon, A., Klein, T., Lagnoux, A., Nodet, M., and Prieur, C. (2014). Asymptotic normality and efficiency of two sobol index estimators. *ESAIM: Probability and Statistics*, 18:342–364.
- Johnson, N. L. (1949). Bivariate distributions based on simple translation systems. *Biometrika*, pages 297–304.
- Kucherenko, S., Tarantola, S., and Annoni, P. (2012). Estimation of global sensitivity indices for models with dependent variables. *Computer Physics Communications*, 183(4):937–946.
- Liu, R. and Owen, A. B. (2006). Estimating mean dimensionality of analysis of variance decompositions. *Journal of the American Statistical Association*, 101(474):712–721.
- Mara, T. A. (2009). Extension of the rbd-fast method to the computation of global sensitivity indices. *Reliability Engineering and System Safety*, 94:1274–1281.
- Mara, T. A. and Tarantola, S. (2012). Variance-based sensitivity indices for models with dependent inputs. *Reliability Engineering & System Safety*, 107:115–121.
- Nelsen, R. B. (2013). *An introduction to copulas*, volume 139. Springer Science & Business Media.
- Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, pages 832–837.
- Saltelli, A., Chan, K., Scott, E. M., et al. (2000). *Sensitivity analysis*, volume 1. Wiley New York.
- Sklar, M. (1959). *Fonctions de répartition à n dimensions et leurs marges*. Université Paris 8.
- Sobol, I. M. (1993). Sensitivity estimates for nonlinear mathematical models. *Math. Modeling Comput. Experiment*, 1(4):407–414 (1995).
- Sobol, I. M. (2001). Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. *Mathematics and computers in simulation*, 55(1-3):271–280.
- Stone, C. J., Hansen, M. H., Kooperberg, C., Truong, Y. K., et al. (1997). Polynomial splines and their tensor products in extended linear modeling: 1994 wald memorial lecture. *The Annals of Statistics*, 25(4):1371–1470.
- Sudret, B. (2008). Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering and System Safety*, 93:964–979.
- Tarantola, S., Gatelli, D., and Mara, T. A. (2006). Random balance designs for the estimation of first order global sensitivity indices. *Reliability Engineering & System Safety*, 91(6):717–727.
- Xu, C. and Gertner, G. Z. (2011). Reliability of global sensitivity indices. *Journal of Statistical Computation and Simulation*, 81(12):1939–1969.