

## **Pick and freeze estimation of sensitivity indices for models with dependent and dynamic input processes**

Mathilde Grandjacques, Alexandre Janon, Benoit Delinchant, Olivier Adrot

► **To cite this version:**

Mathilde Grandjacques, Alexandre Janon, Benoit Delinchant, Olivier Adrot. Pick and freeze estimation of sensitivity indices for models with dependent and dynamic input processes. GCSP. 2014. <hal-00963649v2>

**HAL Id: hal-00963649**

**<https://hal.inria.fr/hal-00963649v2>**

Submitted on 11 Sep 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Pick and freeze estimation of sensitivity indices for models with dependent and dynamic input processes

Mathilde Grandjacques<sup>1</sup>, Alexandre Janon<sup>2</sup>, Benoît Delinchant<sup>1</sup>, Olivier Adrot<sup>3</sup>

## Abstract

This paper addresses sensitivity analysis for dynamic models, linking dependent inputs to observed outputs. The usual method to estimate Sobol indices are based on the independence of input variables. We present a method to overpass this constraint when inputs are Gaussian processes of high dimension in a time related framework. Our proposition leads to a generalization of Sobol indices when inputs are both dependant and dynamic. The method of estimation is a modification of the Pick and Freeze simulation scheme. First we study the general Gaussian cases and secondly we detail the case of stationary models. We then apply the results to an example of heat exchanges inside a building.

## 1 Introduction

To study physical phenomena, it is useful to build mathematic models which translate them. These models can be used for purposes such as managements or forecasts for example. So it is important for the practitioner to assess the fiability of the models used. The sources of uncertainties in a model may be located at two levels :

- on the parameters when they are estimated them for example
- on the inputs of the model (error of measures, variability of the inputs, . . . )

Sensitivity analysis can help to do this work. It aims to quantify uncertainties of each factor on the output of the model. The interest can notably be to :

- reduce variability of the output

---

<sup>1</sup>G2ELab, Université de Grenoble, 38402 St-Martin d'Hères, France

<sup>2</sup>Laboratoire de Mathématiques d'Orsay, Bâtiment 425, Université Paris-Sud, 91405 Orsay, France

<sup>3</sup>GSCOP, 46 avenue Félix Viallet, 38031 Grenoble Cedex 1, France

- prioritize factors : see which factor is the most influent on the output and need more precision on its estimation or its measure
- calibrate the least influent factors.

Among the tools available in global stochastic sensitivity analysis (see for example [1] and references therein), the most used one is Sobol index 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 split down into different partial variances (this is the so-called Hoeffding decomposition, also known as functional ANOVA, see [2]). 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 [3, 4]. 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.

Even when the inputs are not independent, it seems reasonable to consider the same Sobol index but with a quite different interpretation. Several approaches have been proposed in the literature about dependent inputs. In their introduction, Mara et al. [5] cite some of them, which are claimed to be relevant only in the case of a linear model. In that paper, the authors introduce an estimation method for the Sobol index but this method seems computationally intricate. On the other hand, Kucherenko et al. [6] rewrite, as we will do, the Sobol index as a covariance between the output and a copy of the output. Another method ([7]) modifies the Sobol index definition, which leads to indices that are hard to estimate, as well as results that may seem counter-intuitive (for instance, the indices may not be between 0 and 1).

Few works propose to study the sensitivity to 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. It seems necessary to develop a new method to dynamic dependent inputs. In this way, the Sobol index definition is modified.

We set ourselves in a time related framework and we study the following scalar output  $Y$  :

$$Y_t = f_t((\mathbf{U}_s)_{0 \leq s \leq t}), \quad t \in \mathbb{N} \quad (1)$$

The input is a vectorial Gaussian process  $(\mathbf{U}_s)_{s \in \mathbb{N}} \subset \mathbb{R}^p$ . In this context, the sensitivity is defined for  $Y_t$  with respect to the input process  $(\mathbf{U}_s^1)_{0 \leq s \leq t}$  (for example). Thus the sensitivity changes with time  $t$ . The dynamic framework is the most useful in stationary or almost stationary cases. In non stationary cases the problem is no more that a sequence of finite dimensional situations. We focus on two cases with  $\mathbf{U}_t$  a stationary process :

- $Y_t = f(\mathbf{U}_t, \mathbf{U}_{t-1}, \dots, \mathbf{U}_{t-M})$  for some  $M$
- $Y_t = f(\mathbf{U}_t, \mathbf{U}_{t-1}, \dots, \mathbf{U}_0, 0, \dots, 0)$  deduced from a stationary process given by  $Y_t^* = f(\mathbf{U}_t, \mathbf{U}_{t-1}, \dots, \mathbf{U}_0, \mathbf{U}_{-1}, \dots)$ .

This case includes models associated to recurrence equations as Euler schemes of stochastic differential equations.

The method of estimation that seems best suited for functional multidimensional models is the Pick and Freeze scheme (see [4, 8]). It allows flexibility in the form of the inputs and doesn't care of the number of variables by which it is desired to condition the variance, the only constraint being the assumption of independent inputs. In SPF (Scheme Pick and Freeze), a Sobol index is viewed as the correlation 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.

In a first part, after reminding the definition of Sobol index and the Pick and Freeze scheme, we introduce the definition of the index in a dynamic case. We show that under the hypothesis of Gaussian inputs, it is possible to reduce this problem to the case of independent variables and apply the method Pick and Freeze. In the second part we present the properties of our index when inputs are stationary. Finally an application to a physical problem is presented in the last section.

## Notations

Let us give some notations :

- $X, Z$  random variables
- $\mathbf{U}, \mathbf{Z}$  random vectors
- $(\mathbf{U}_t)_{t \in \mathbb{N}}$  a vectorial process,  $\dim(\mathbf{U}_t) = p$
- $\mathbf{U}_{[a,b]} = \{\mathbf{U}_s, a \leq s \leq b\}$ ,  $p \times (b - a + 1)$  matrix, with  $-\infty \leq a \leq b \leq +\infty$
- $\mathbf{U}^*$  or  $\mathbf{U}_t^*$  is the transposed vector of  $\mathbf{U}$  or the vectorial process  $\mathbf{U}_t$

If  $X_t$  and  $\mathbf{Z}_t$  are two stochastic vectorial centered processes with  $\dim(X_t) = 1$ ,  $\dim(\mathbf{Z}_t) = p - 1$ , we define different covariance matrices as following :

**Definition 1.**  $\gamma_{s,v}^{XZ^j} = \mathbf{E}(X_s Z_v^j)$  the covariance between  $X_s$  and  $Z_v^j$  where  $j$  denotes the  $j$ th component of the vector  $\mathbf{Z}_v$

**Definition 2.**  $\boldsymbol{\gamma}_{[0,t],v}^{XZ^j} = \mathbf{E}(X_{[0,t]} Z_v^j)$  a  $(t + 1)$  vector process of generic term  $\gamma_{s,v}^{XZ^j}$ ,  $0 \leq s \leq t$

**Definition 3.**  $\Gamma_{[0,t],v}^{XZ}$  the  $(t + 1) \times (p - 1)$  covariance matrix of generic term  $\gamma_{[0,t],v}^{XZ^j}$  for  $1 \leq j \leq p - 1$ ,  $0 \leq s \leq t$

**Definition 4.**  $\Gamma_{[0,t],[0,u]}^{XZ}$  the  $(t + 1) \times (p - 1)(u + 1)$  matrix using matrix blocks  $\Gamma_{[0,t],v}^{XZ}$  with  $0 \leq v \leq u$

To simplify the exposition we consider an input vector  $\mathbf{U} = (U^1, \dots, U^p)$ . We denote by  $X = U^1$  and  $\mathbf{Z} = (U^2, \dots, U^p)$  when we are in static context and  $X_t = U_t^1$  and  $\mathbf{Z}_t = (U_t^2, \dots, U_t^p)$  in the dynamic case.

## 2 Sobol indices : extended definition and estimation

### 2.1 Definition in a dynamic context

We consider the model given by :  $Y = f(\mathbf{U})$ ,  $\mathbf{U} \in \mathbb{R}^p$  is a random vector with known distributions. We assume that all coordinates of  $\mathbf{U}$  and  $Y$  have a finite non zero variance.

The Sobol index with respect to  $X$  is defined by [3] :

$$S^X = \frac{\mathbf{Var}(\mathbf{E}(Y|X))}{\mathbf{Var}(Y)}. \quad (2)$$

$S^X$  is the Sobol index with respect to  $X$ . More generally,  $S^J$  is the closed Sobol index with respect to the group  $J$  of variables  $\mathbf{U}^J = (U^j, j \in J)$  :

$$S^{U^J} = \frac{\mathbf{Var}(\mathbf{E}(Y|\mathbf{U}^J))}{\mathbf{Var}(Y)}. \quad (3)$$

Total indices and higher-order Sobol indices can also be written by taking the sum or the difference of closed indices. Hence, we can restrict ourselves to the case of two (possibly vector) inputs in the model.

We now introduce the time dimension and we define the input-output relation :

$$Y_t = f_t((\mathbf{U}_s)_{0 \leq s \leq t}) \quad (4)$$

$(\mathbf{U}_s)_{s \in \mathbb{N}} \subset \mathbb{R}^p$  is a vector-valued stochastic process,  $f_t$  being a sequence of functions which will be detailed later. Let  $\mathbf{U}_t = (X_t, \mathbf{Z}_t)$  with  $X_t = U_t^1$  and  $\mathbf{Z}_t = (U_t^2, \dots, U_t^p)$ .

For each  $t \in \mathbb{N}$ , we define a measure of the sensitivity of  $Y_t$  with respect to  $X_{[0,t]} = (X_0, \dots, X_{t-1}, X_t)$  by :

$$S_t^X = \frac{\mathbf{Var}(\mathbf{E}(Y_t|X_{[0,t]}))}{\mathbf{Var}(Y_t)} \quad (5)$$

The index  $t \mapsto S_t^X$  is called the *projection on the past sensitivity index with respect to X*. We notice that, at any given time  $t$ , we consider the sensitivity of  $Y_t$  with respect to all the past  $X_{[0,t]}$  values of the  $X$  process, not just its value  $X_t$ .

Of course, the conditional expectation with respect to  $X_{[0,t]}$  takes into account the dependence of  $\mathbf{Z}_{[0,t]}$  with respect to  $X_{[0,t]}$ .

**Remark 1.** *When the inputs are dependent we keep the property that  $S_t^X \leq 1$  for any  $t$ . The classical term of interaction  $S^{XZ}$  is not defined [9].*

### 2.2 Estimation of $S^X$ , the Pick and Freeze method in the independent case

There exists many methods for estimating  $S^X$ . One of them is the so-called Pick and Freeze scheme [4, 3]. In this case  $f$  plays the role of a black box allowing to simulate the input-output

relationship without any mathematical description. This method is based on the following lemma [4] :

**Lemma 1. Sobol :** *Let  $\mathbf{U} = (X, \mathbf{Z})$ . If  $X$  and  $\mathbf{Z}$  are independent :*

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

with  $Y^X = f(X, \mathbf{Z}')$ ,  $Y = f(X, \mathbf{Z})$  where  $\mathbf{Z}'$  is an independent copy of  $\mathbf{Z}$ .

We can deduce the expression of  $S^X$  when  $X$  and  $\mathbf{Z}$  are independent :

$$S^X = \frac{\mathbf{Cov}(Y, Y^X)}{\mathbf{Var}(Y)}, \quad (6)$$

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

$$\hat{S}^X = \frac{\frac{1}{N} \sum_{i=1}^N Y^{(i)} Y^{X,(i)} - (\frac{1}{N} \sum_{i=1}^N Y^{(i)}) (\frac{1}{N} \sum_{i=1}^N Y^{X,(i)})}{\frac{1}{N} \sum_{i=1}^N (Y^{(i)})^2 - (\frac{1}{N} \sum_{i=1}^N Y^{(i)})^2} \quad (7)$$

If  $X$  and  $\mathbf{Z}$  are finite dimensional random vectors this formula can be justified by asymptotic properties when  $N \rightarrow \infty$ . The speed of convergence of this estimator is in  $O(1/\sqrt{N})$ , see Janon et al. [10]. In practice it can be approximated by  $\frac{C}{\sqrt{N}}$  where  $C$  can be large as we will see later.

In the dependent case the estimation of  $S_t^X$  by a Monte-Carlo method is a challenging task, as one cannot be chosen  $X = (X_{[0,t]})$  and  $\mathbf{Z} = (\mathbf{Z}_{[0,t]})$  in (6) since  $X$  and  $\mathbf{Z}$  are not independent. However, we will see, in the following Section, that, in a particular Gaussian case, whatever the covariance structure is, an efficient Pick and Freeze scheme may be built.

## 2.3 Reduction to independent inputs for Gaussian processes

Suppose that we are able to get another expression of the  $Y_t$  output of the type :

$$Y_t = g_t(X_{[0,t]}, \mathbf{W}_{[0,t]}) \quad (8)$$

where  $\mathbf{W}_{[0,t]}$  is a stochastic process independent of  $X_{[0,t]}$ ,  $\mathbf{W}_{[0,t]}$  being  $(X_s, \mathbf{Z}_s)_{s \leq t}$  measurable.

Then :

$$S_t^X = \frac{\mathbf{Var}(\mathbf{E}(Y_t = g_t(X_{[0,t]}, \mathbf{W}_{[0,t]})|X_{[0,t]}))}{\mathbf{Var}(Y_t)} \quad (9)$$

is defined as in the classical case of independence.

We now prove that if  $Y_t = f_t(\mathbf{U}_t, \mathbf{U}_{t-1}, \dots, \mathbf{U}_0)$  there exists  $g_t$  satisfying (8).

Let :

$$\tilde{\mathbf{X}}_t = \mathbf{E}(\mathbf{Z}_t | X_{[0,t]}) = \Lambda X_{[0,t]} \text{ for a matrix } \Lambda \text{ with } \dim(\Lambda) = (p-1) \times (t+1),$$

$$\text{and } \mathbf{W}_t = \mathbf{Z}_t - \mathbf{E}(\mathbf{Z}_t | X_{[0,t]})$$

Independence of  $X_{[0,t]}$  and  $\mathbf{W}_{[0,t]}$  holds thanks to the Gaussian assumption.

Let :

$$f_t(X_{[0,t]}, \mathbf{Z}_{[0,t]}) = f_t(X_{[0,t]}, \tilde{\mathbf{X}}_{[0,t]} + \mathbf{W}_{[0,t]}) \quad (10)$$

$$= f_t(X_{[0,t]}, \Lambda X_{[0,t]} + \mathbf{W}_{[0,t]}) \quad (11)$$

$$= g_t(X_{[0,t]}, \mathbf{W}_{[0,t]}) \quad (12)$$

Let us now compute  $\tilde{\mathbf{X}}_{[0,t]}$  defined as :

$$\mathbf{E}(\mathbf{Z}_{[0,t]} | X_{[0,t]}) = \{\mathbf{E}(\mathbf{Z}_u | X_{[0,t]}), 0 \leq u \leq t\} = \tilde{\mathbf{X}}_{[0,t]}$$

$(X_t, \mathbf{Z}_t)$  being a Gaussian vector, conditional expectations with respect to  $X_{[0,t]}$  are the projections on the linear space generated by  $X_{[0,t]}$ .

**Assumption 1.** For every  $t$  we suppose that  $\mathbf{U}_{[0,t]}$  is of full rank.

Thus :

$$\mathbf{E}(Z_u^j | X_{[0,t]}) = \boldsymbol{\lambda}_{[0,t],u}^j X_{[0,t]}^* \quad (13)$$

where  $\boldsymbol{\lambda}_{[0,t],u}^j$  is a vector of size  $(t+1)$  given by classical linear regression results :

$$\boldsymbol{\lambda}_{[0,t],u}^j = (\Gamma_{[0,t],[0,t]}^{XX})^{-1} \gamma_{[0,t],u}^{XZ^j} \quad (14)$$

$\Gamma_{[0,t],[0,t]}^{XX}$  is invertible as consequence of assumption 1.

Let  $\Gamma_{[0,t],[0,t]}^{X,Z}$  defined as previously in (3) and

$$\Lambda_{[0,t],[0,t]}^{XZ} = (\Gamma_{[0,t],[0,t]}^{XX})^{-1} \Gamma_{[0,t],[0,t]}^{XZ} \quad (15)$$

then :

$$\tilde{\mathbf{X}}_{[0,t]} = \Lambda_{[0,t],[0,t]}^{XZ} X_{[0,t]} \quad (16)$$

$\tilde{\mathbf{X}}_{[0,t]}$  is a  $(p-1) \times (t+1)$  matrix as is  $\mathbf{W}_{[0,t]} = \mathbf{Z}_{[0,t]} - \tilde{\mathbf{X}}_{[0,t]}$ .

Thanks to (10) we have :

$$S_t^X = \frac{\mathbf{Var}(\mathbf{E}(g_t(X_{[0,t]}, \mathbf{W}_{[0,t]}) | X_{[0,t]}))}{\mathbf{Var}(Y_t)}.$$

Note that the space of all the square integrable functions of the form  $\phi(X_{[0,t]})$  is the same as the space of all the square integrable function of the form  $\psi(X_{[0,t]}, \tilde{\mathbf{X}}_{[0,t]})$ .

Thus  $\mathbf{E}(Y_t | (X_{[0,t]}, \tilde{\mathbf{X}}_{[0,t]})) = \mathbf{E}(Y_t | X_{[0,t]})$ . For  $t$  fixed, we are now exactly in the previous case of two groups of independent inputs  $X_{[0,t]}$  and  $\mathbf{W}_{[0,t]}$  and thus we can apply the Pick and Freeze method (10) and (6) with  $\mathbf{X} = (X_{[0,t]}, \tilde{\mathbf{X}}_{[0,t]})$  and  $\mathbf{Z} = \mathbf{W}_{[0,t]}$ . By copying  $\mathbf{W}'_t$  of  $\mathbf{W}_t$  we mean a stochastic process independent of  $\mathbf{W}_t$  with the same finite dimensional distributions. If  $(X_{[0,t]}^{(i)}, \mathbf{W}_{[0,t]}^{(i)})_{i=1,\dots,N}$  is a sample of  $(X_{[0,t]}, \mathbf{W}_{[0,t]})_{t \in \mathbb{N}}$  we denote  $(X_{[0,t]}^{(i)}, (\mathbf{W}'_{[0,t]}^{(i)})' )_{i=1,\dots,N}$  the sample obtained with  $\mathbf{W}'_{[0,t]}$  a copy of  $\mathbf{W}_{[0,t]}$ .

Let  $Y_t^X = g_t(X_{[0,t]}, \mathbf{W}'_{[0,t]})$ .

We have, for any  $t \in \mathbb{N}$ :

$$S_t^X = \frac{\mathbf{Cov}(Y_t, Y_t^X)}{\mathbf{Var}(Y_t)},$$

where  $Y_t^X = f_t\left(X_{[0,t]}, \tilde{\mathbf{X}}_{[0,t]} + (\mathbf{W}'_{s \leq t})\right)$ , where  $X_{[0,t]}$  and thus  $\tilde{\mathbf{X}}_{[0,t]}$ , which is a function of  $X_{[0,t]}$ , are frozen.

Now to estimate  $S_t^X$ , we need only to get an empirical estimator of the covariance as in (6). Thus we simulate a sample  $(X_{[0,t]}^{(i)}, \mathbf{W}_{[0,t]}^{(i)})$  and  $(X_{[0,t]}^{(i)}, (\mathbf{W}_{[0,t]}^{(i)})')$ ,  $i = 1, \dots, N$ .

To do this, we simulate two independent pairs  $(X_{[0,t]}^{(i)}, \mathbf{Z}_{[0,t]}^{(i)})$  and  $((X_{[0,t]}^{(i)})', (\mathbf{Z}_{[0,t]}^{(i)})')$ . Thanks to these pairs, we build  $\tilde{\mathbf{X}}_{[0,t]}^{(i)}$  and  $(\tilde{\mathbf{X}}_{[0,t]}^{(i)})'$ . We deduce  $(\mathbf{W}_{[0,t]}^{(i)})'$  thanks to  $(\mathbf{W}_{[0,t]}^{(i)})' = (\mathbf{Z}_{[0,t]}^{(i)})' - (\tilde{\mathbf{X}}_{[0,t]}^{(i)})'$ .

In the stationary case, the simulation of the Gaussian vectorial process  $(\mathbf{U}_t)_{t \in \mathbb{N}} = (X_t, \mathbf{Z}_t)_{t \in \mathbb{N}}$  is a classical problem when its covariance is known. In the non stationary case the Cholesky decomposition of the covariance matrix is the most popular method.

Once  $(\mathbf{U}_t)_{t \in \mathbb{N}}$  is simulated, we have to recover  $\tilde{\mathbf{X}}_u$  and  $\mathbf{W}_u = \mathbf{Z}_u - \tilde{\mathbf{X}}_u$ . Formula (16) gives  $\tilde{\mathbf{X}}_{[0,t]} = \Lambda_{[0,t],[0,t]}^{XZ} X_{[0,t]}$ , and directly allows the computation of :

$$\mathbf{W}_{[0,t]} = \mathbf{Z}_{[0,t]} - \Lambda_{[0,t],[0,t]}^{XZ} X_{[0,t]}.$$

## 3 Sensitivity and stationarity

### 3.1 Stationary input-output models

Let  $\mathbf{U}_t = (X_t, \mathbf{Z}_t)_{t \in \mathbb{Z}}$  a stochastic process considered as an input and  $Y_t = f_t(\mathbf{U}_t, \dots, \mathbf{U}_0)$  as the output. It is assumed in the following that  $(\mathbf{U}_t)_{t \in \mathbb{Z}}$  is a stationary process. Remember that a process is stationary if all its multidimensional distributions are translation invariant in time. For a Gaussian process  $(\mathbf{U}_t)_{t \in \mathbb{Z}}$ , stationarity is equivalent to  $\mathbf{E}(\mathbf{U}_t) = m$  and  $\mathbf{E}(\mathbf{U}_t \mathbf{U}_{t+k}^*) = \Gamma^U(k)$  independents of  $t$ .

We consider two cases :

- Case 1 :  $(\mathbf{U}_t)_{t \in \mathbb{Z}}$  is stationary and  $Y_t = f(\mathbf{U}_t, \dots, \mathbf{U}_{t-M})$  is a stationary process  $M$  is fixed as the proper memory of  $Y_t$
- Case 2 :  $(\mathbf{U}_t)_{t \in \mathbb{Z}}$  is stationary and there exists a stationary process  $Y_t^*$  (Bernoulli shift process) such as  $Y_t^* = f(\mathbf{U}_t, \dots, \mathbf{U}_0, \mathbf{U}_{-1}, \dots)$  and  $Y_t = f_t(\mathbf{U}_t, \dots, \mathbf{U}_0) = f(\mathbf{U}_t, \dots, \mathbf{U}_0, 0, \dots)$ .

In the second case,  $Y_t^*$  is a stationary process while  $Y_t$  is not strictly stationary but it is a useful approximation in applications as we will see later.



### 3.2 Sobol indices convergence

We first study the case 1  $Y_t = f(\mathbf{U}_t, \dots, \mathbf{U}_{t-M})$ . We assume  $Y_t$  centered, without loss of generality. The Sobol index is defined as :

$$S_t^X = \frac{\mathbf{Var}(\mathbf{E}(Y_t | X_{[0,t]}))}{\mathbf{Var}(Y_t)}$$

For a fixed  $K$ , let note  $X_{[t-K,t]} = \{X_t, \dots, X_{t-K}\}$ , we have :

$$\forall t \geq K \quad \frac{\mathbf{Var}(\mathbf{E}(Y_t | X_{[0,t]}))}{\mathbf{Var}(Y_t)} \geq \frac{\mathbf{Var}(\mathbf{E}(Y_t | X_{[t-K,t]}))}{\mathbf{Var}(Y_t)} = S_{t,K}^X \quad (17)$$

But the last quantity is constant in  $t$  by translation invariance,  $(Y_t, \mathbf{U}_t)$  being stationary thus :

$$\text{for } t \geq K \quad S_{t,K}^X = S_K^X \quad (18)$$

$S_K^X$  is an increasing sequence in  $K$  and bounded so :

$$S_\infty^X = \sup_K S_K^X \quad (19)$$

thus we see that the sensitivity reaches a limit  $S_\infty^X$  as  $t \rightarrow +\infty$  whatever the stationary system  $(Y_t, \mathbf{U}_t)$ .

**Lemma 2.** *In Case 1 :*

$$\lim_{t \rightarrow +\infty} S_t^X = \lim_{t \rightarrow +\infty} \sup_{K \leq t} S_K^X = S_\infty^X \quad (20)$$

We give in appendix the proof of the same result in the second case when  $Y_t$  has its proper dynamics, but only in particular cases when  $\mathbf{U}_t$  is a linear causal process and  $Y_t$  has a specific form, including the most general linear case.

### 3.3 VAR input case

The simplest input model is the vectorial autoregressive process of order  $p$ , noted  $VAR(p)$ .

The  $VAR(1)$  model is given by :

$$\begin{pmatrix} X_t \\ \mathbf{Z}_t \end{pmatrix} = A \begin{pmatrix} X_{t-1} \\ \mathbf{Z}_{t-1} \end{pmatrix} + \boldsymbol{\omega}_t, \quad t > 0 \quad (21)$$

$A$  is a  $p \times p$  matrix, and  $(\boldsymbol{\omega}_t)_{t \in \mathbb{N}}$  are  $p$ -dimensional iid standard Gaussian variables with covariance  $\Theta$  and  $(X_0, \mathbf{Z}_0)$  given.

We have of course:

$$\begin{pmatrix} X_t \\ \mathbf{Z}_t \end{pmatrix} = \sum_{k=0}^t A^k \boldsymbol{\omega}_{t-k} + A^t \begin{pmatrix} X_0 \\ \mathbf{Z}_0 \end{pmatrix} \quad (22)$$

Stationarity is equivalent to a spectral radius  $|\rho(A)| < 1$ . From now we suppose that this condition is verified. A  $VAR(1)$  process is known to be geometrically ergodic ([11]), thus whatever the distribution of  $(X_0, \mathbf{Z}_0)$  in (22) is, it implies that when  $t \rightarrow +\infty$  the distribution of  $(X_t, \mathbf{Z}_t)$  tends to the stationary distribution  $\mathcal{N}(0, \Gamma_{t,t})$  with  $\Gamma_{t,t} = \mathbf{E}((X_t, \mathbf{Z}_t)(X_t, \mathbf{Z}_t)^*)$  in the stationary case. We use this result in the following way : starting from any  $(X_{t_1}, \mathbf{Z}_{t_1})$  for  $t_1 < 0$  (for instance  $t_1 = -200$ ), the distribution of  $(X_0, \mathbf{Z}_0)$  is the stationary (invariant) distribution except negligible errors and using :

$$\begin{pmatrix} X_0 \\ \mathbf{Z}_0 \end{pmatrix} = \sum_{k=t_1}^0 A^{-k} \boldsymbol{\omega}_k + A^{-t_1} \begin{pmatrix} X_{t_1} \\ \mathbf{Z}_{t_1} \end{pmatrix} \quad (23)$$

In stationary regime we have :

$$\begin{pmatrix} X_t \\ \mathbf{Z}_t \end{pmatrix} = \sum_{k=0}^{\infty} A^k \boldsymbol{\omega}_{t-k} \quad (24)$$

and thus :

$$\Gamma_{t,t} = \sum_{k=0}^{\infty} A^k \Theta(A^k)^* \quad (25)$$

$\Gamma_{[0,t],[0,t]}^{X,X}$  can be easily inverted thanks to its Toeplitz property which allows a faster computation of  $\Lambda_{[0,t],[0,t]}^{XZ}$  (14) used for the simulation of  $\tilde{\mathbf{X}}$  (16).

**Remark 2.** *The  $VAR(p)$  case, given by :*

$$\begin{pmatrix} X_t \\ \mathbf{Z}_t \end{pmatrix} = \sum_{j=1}^p A^j \begin{pmatrix} X_{t-j} \\ \mathbf{Z}_{t-j} \end{pmatrix} + \boldsymbol{\omega}_t, \quad t > 0$$

*can be reduced to the  $VAR(1)$  case by increasing the dimension.*

A pseudo code regarding the reduction of dependent inputs to independent inputs is presented in algorithm 1 for  $VAR(1)$ . The algorithm of the Pick and Freeze method is presented in algorithm 2. We build samples of size  $N$ .

We have to pay attention as explained above to the initial value that we choose for processes  $Simu_1$  and  $Simu_2$  to ensure that they are stationary. To do that we can use a process  $Simu_{init}$ , with some initialisation, that we simulate. We select at a time  $t_1$ ,  $Simu_{init}[t_1]$  that will be the initialisation.

Lines (17) and (18),  $W_1$  and  $W_2$  are what we call previously  $W$  and  $W'$ . The same for  $Y_1$  and  $Y_2$  are what we call previously  $Y$  and  $Y'$

Algorithm 1 returns  $SIMU_1$  and  $SIMU_2$  which is its pick-freezed replication. They will be used to estimate the Sobol index in algorithm 2.

**Remark 3.** *Algorithm 1 and 2 can be applied to any gaussian stationary process. Part 1 of algorithm 1 is a method of simulation of the process.*

---

**Algorithm 1** Reduction to independent inputs

---

**Require:**  $A, \Theta, U, N, init_1, init_2$ 

```
1:  $times \leftarrow \dim(U)[1]$  ,  $input \leftarrow \dim(U)[2]$ 
2:  $Simu_1[times, input, N] \leftarrow 0$  ,  $Simu_2[times, input, N] \leftarrow 0$ 
3: for  $i = 1$  to  $input$  do
4:    $\omega_1[times, input, N] \sim \mathcal{N}(0, \Theta)$ 
5:    $\omega_2[times, input, N] \sim \mathcal{N}(0, \Theta)$ 
6:
7:    $Simu_1[1, , ] \leftarrow init_1$ 
8:    $Simu_2[1, , ] \leftarrow init_2$ 
9:   for  $t = 2$  to  $times$  do
10:     $Simu_1[t, , ] \leftarrow A \cdot Simu_1[t - 1, , ] + \omega_1[t, , ]$ 
11:     $Simu_2[t, , ] \leftarrow A \cdot Simu_2[t - 1, , ] + \omega_2[t, , ]$ 
12:   end for
13:
14:   for  $t = 1$  to  $times$  do
15:     $\Lambda \leftarrow (\mathbf{Cov}(U[1 : t, i], U[1 : t, i]))^{-1} \mathbf{Cov}(U[1 : t, , ], U[1 : t, , ])$ 
16:     $\tilde{X}_1[t, , ] \leftarrow (Simu_1[1 : t, i, , ])^* \cdot \Lambda$ 
17:     $\tilde{X}_2[t, , ] \leftarrow (Simu_2[1 : t, i, , ])^* \cdot \Lambda$ 
18:   end for
19:    $W_1 \leftarrow Simu_1 - \tilde{X}_1$ 
20:    $W_2 \leftarrow Simu_2 - \tilde{X}_2$ 
21:    $SIMU_1 \leftarrow Simu_1$ 
22:    $SIMU_2 \leftarrow \tilde{X}_1 + W_2$ 
23: end for
24: return  $(SIMU_1, SIMU_2)$ 
```

---

---

**Algorithm 2** Pick and Freeze estimation

---

**Require:**  $SIMU_1, SIMU_2, init$ 

```
1:  $times \leftarrow \dim(SMU_1)[1]$  ,  $input \leftarrow \dim(SIMU_1)[2]$ 
2:  $indice \leftarrow 0$ 
3: for  $i = 1$  to  $input$  do
4:   if  $init \neq \text{NULL}$  then
5:      $Y_1 = f(SIMU_1, init)$ 
6:      $Y_2 = f(SIMU_2, init)$ 
7:   else
8:      $Y_1 = f(SIMU_1)$ 
9:      $Y_2 = f(SIMU_2)$ 
10:  end if
11:   $indice[, input] = \frac{\mathbf{E}(Y_1 \otimes Y_2) - \mathbf{E}(Y_1)\mathbf{E}(Y_2)}{\mathbf{Var}(Y_1)}$ 
12: end for
13: return  $indice$ 
```

---

### 3.4 Toy models

We study two stationary toy models one linear and one non linear given by:

$$Y_t = 0.2Y_{t-1} + 0.3X_t + Z_t \quad (26)$$

$$Y_t = X_t Z_t + 0.2 \exp(-Z_t) \quad (27)$$

$\begin{pmatrix} X_t \\ Z_t \end{pmatrix}$  is a  $VAR(1)$  stationary process given by:

$$\begin{pmatrix} X_t \\ Z_t \end{pmatrix} = \begin{pmatrix} 0.8 & 0.4 \\ 0.1 & 0.2 \end{pmatrix} \begin{pmatrix} X_{t-1} \\ Z_{t-1} \end{pmatrix} + \omega_t \quad (28)$$

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

First let's give an example of the result that can be expected with the algorithm 1 when we study the sensitivity of  $Y$  with respect to  $X$  for the model (26).  $\Lambda$  is here a vector  $(\lambda_t, \lambda_{t-1}, \dots, \lambda_{t-k}, \dots, \lambda_0)^*$ . Its values are given in table : 1. After simulation of  $Simu_1$  (table : 2),  $\tilde{X}_1$  is calculated. Using the value of  $\Lambda$  given in table 1 we compute the values corresponding to  $\tilde{X}$  define as previously. The table 2 is the result of the step (17) :  $W_1 = Z - \tilde{X}_1$ . We do the same work with  $Simu_2$  and we obtain  $W_2$ . We can remark that coefficients of  $\Lambda$  decrease. Only the three first past instant are important. We see from table 1 that.

time \ time	0	1	2	3	4
0	$X_t$ 0.12	$X_{t-1}$ 0.38	$X_{t-2}$ 0.07	$X_{t-3}$ -0.01	$X_{t-4}$ -0.01
1		$X_t$ -0.21	$X_{t-1}$ 0.33	$X_{t-2}$ 0.07	$X_{t-3}$ 0.00
2			$X_{t-1}$ -0.21	$X_{t-1}$ 0.33	$X_{t-2}$ 0.07
3				$X_t$ -0.21	$X_{t-1}$ 0.33
4					$X_t$ -0.21

Table 1: Estimated values of  $\Lambda$  (Step (13) of algorithm 1)

time	$X_t$	$Z_t$	$\tilde{X}_t$	$W_t$
0	-0.21	0.40	-0.02	0.42
1	0.11	-0.12	-0.10	-0.02
2	-0.17	-1.02	0.06	-1.08
3	-0.29	-0.79	0.01	-0.80
4	0.24	-0.80	-0.16	-0.64

Table 2: Values of  $X$ ,  $Z$ ,  $\tilde{X}_t$  and  $W$  for  $Simu_1$  (Step (8) of algorithm 1)

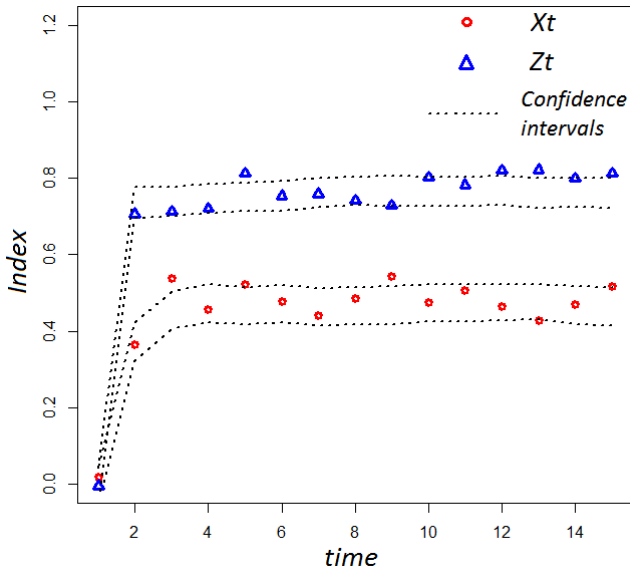


Figure 1: Toy model 26 : Sobol index estimate in function of time. Sample size: 200

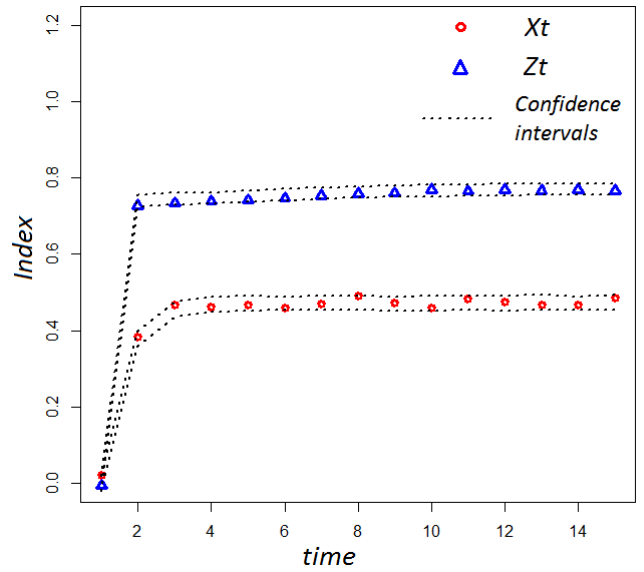


Figure 2: Toy model 26 : Sobol index estimate in function of time. Sample size: 10000

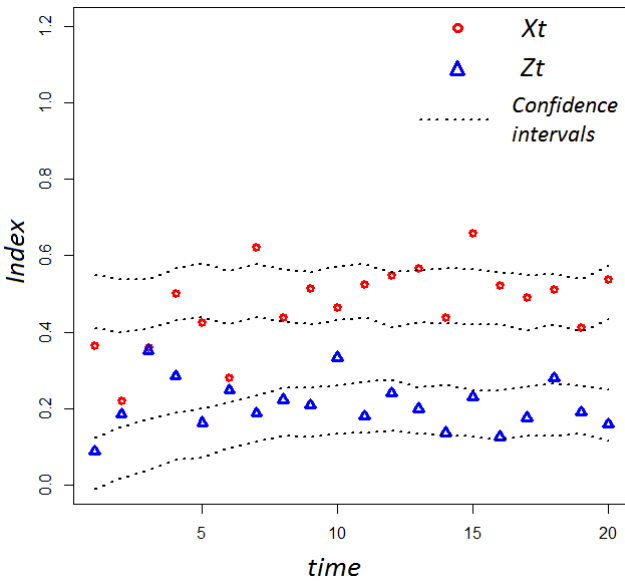


Figure 3: Toy model 27 : Sobol index estimate in function of time. Sample size: 200

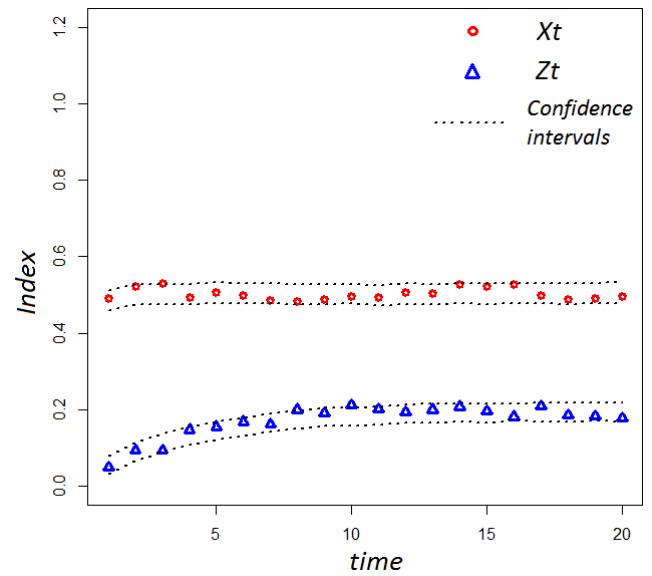


Figure 4: Toy model 27 : Sobol index estimate in function of time. Sample size: 10000

Figures 1, 2, 3, 4, display for each time step,  $S_t^X = \frac{\mathbf{Var}(\mathbf{E}(Y_t|X_{[0,t]}))}{\mathbf{Var}(Y_t)}$ , calculated for models (26) and (27) for different sizes of samples ( $N = 200$  and  $N = 10000$ ). Confidence intervals are stated at the 95% confidence level and plotted on each figures. The following examples exhibit two different types of convergence :

- the convergence of the estimator. At each time step, the Pick and Freeze algorithm estimates  $S_t^X$ . The quality of the estimator  $\widehat{S}_t^X$  depends on the size of the sample. The confidence interval is smaller when  $N = 10000$ . The convergence speed of the estimator is slow ( $O(1/\sqrt{N})$ ) when we use a Monte Carlo sample. We could improve the speed using Quasi Monte Carlo (QMC) sample method [1]. But in our case QMC seems hard to implement.
- the temporal convergence :  $\widehat{S}_t^X$  changes over time at each  $t$ . After just three iterations in time of the estimation algorithm,  $S_t^X$  value reaches a limit (see on figures : 1 and 2). Model (26) is auto-regressive, it means that  $Y_t$  depends on its past  $Y_{t-1}$  and so, on all the past of  $X_t, Z_t$ . We can rewrite it as :

$$Y_t = \sum_{k=0}^{\infty} (0.2)^k (0.3X_{t-k} + Z_{t-k})$$

At time  $t = 0$  for example,  $S_0^X = \frac{\mathbf{Var}(\mathbf{E}(Y_0|X_0))}{\mathbf{Var}(Y_0)}$ .  $Y_0$  is projected on a space of dimension 1 whereas it depends on  $(X_{-\infty}, \dots, X_0, Z_{-\infty}, \dots, Z_0)$ . The projection space is too small. When increasing this space, the index increases and converges to a constant. It's what we call the phenomenon of memory which may refer to the physical concept of inertia. When  $Y_t = \phi(X_t, Z_t)$  the index converges instantaneous (figures : 3, 4).

The time convergence is interesting from a computing point of view. As the index converges to a constant, it useless to compute the index that takes into account the total trajectory of the process considered. It represents a lot of economy on a computational point of view because it requires less iterations for the evaluation of the index.

We can also notice that the method is independent of the expression of the model  $f$ . It is a black-box method requiring just the possibility to simulate a lot of inputs and outputs.

## 4 Application

### 4.1 The physical problem

We now address a model of building constructed as a metamodel [12] estimated thanks to observed data. This building is actually a classroom that welcomes students during the school year. It is therefore sensitive to hours, days but also holidays that occur during the year.

The studied room, whose internal temperature is  $T^{\text{int}}$  is surrounded by a corridor, an adjacent office, an office that is located below and a skylight (a shed). We have equipped these parts with

temperature sensors. Temperatures are measured every hour denoted  $(T^{\text{below}}, T^{\text{above}}, T^{\text{off}}, T^{\text{cor}}, T^{\text{ext}}, T^{\text{int}})$ . We also measure the temperature outside  $T^{\text{ext}}$ . Obviously  $T^{\text{ext}}$  influences all the other temperatures  $T^{\text{below}}, T^{\text{above}}, T^{\text{off}}, T^{\text{cor}}, T^{\text{int}}$ .

We choose to study the summer period when students are not present and when the heating is off.

## 4.2 The input and input-output models

We have to build a statistical input model and also an input-output model. After preprocessing inputs are modelled as a  $VAR(p)$  process. Details are given in appendix B.

The output of the model is the internal temperature  $T^{\text{int}}$ . Our sensitivity analysis aims to determine which among  $T^{\text{below}}, T^{\text{above}}, T^{\text{off}}, T^{\text{cor}}, T^{\text{ext}}$  impact the most the variance of the internal temperature  $T^{\text{int}}$ . We choose to model the input-output system by a linear autoregression of  $T^{\text{int}}$  on variables  $U_t = (T_t^{e'}, T_{t-1}^{e'})_t$ , for  $e' \in \mathcal{E}' = \{\text{below, above, off, cor, ext}\}$ . The used model is given in appendix B. Note that if we have at our disposal a physical model able to simulate the input-output model (for instance an electrical model), we can use the data provided by this model to apply the Pick and Freeze method.

## 4.3 Numerical results

We then compute the different estimators  $\widehat{S}_t^X = \widehat{S}_t^e$ , for  $t = 2, \dots, 48$  and  $e \in \mathcal{E}'$ . The results are gathered in figure : 5. Indices are estimated thanks to a sample of size  $N = 5000$ .

The most influent variables are  $T^{\text{ext}}$  the external temperature and  $T^{\text{above}}$  the temperature of the room above. These two rooms have the most important noise variances : 1.34 and 0.62 (see  $\Theta$  matrix (31)), whereas other noise variances are around 0.02. It seems logical that they are the most important variables in the sensitivity analysis. The number of time iterations is 20. The memory or inertia of the system is long enough.

Despite its small noise variance,  $T^{\text{cor}}$  is important. This variable illustrates the fact that sensitivity in a dependent context is not due only to the variance of the variable. Sensitivity is a function of the input-output relation and of the covariance between the variable of interest and the others.  $T^{\text{cor}}$  is the coldest room, it lowers the temperature  $T^{\text{int}}$ .  $T^{\text{cor}}$  is most likely to cause temperature variations.

$T^{\text{below}}$  is measured in a room located below the studied room  $T^{\text{int}}$  isolated by a slab. It seems logical that it does not affect  $T^{\text{int}}$ .

## Conclusion

In this article, we have proposed a modified definition of Sobol indices, adapted to dependent dynamic inputs. The index is no longer associated with the Hoeffding decomposition and the sum of the index is not equal to 1. Nevertheless, the index is between 0 and 1 and keeps the

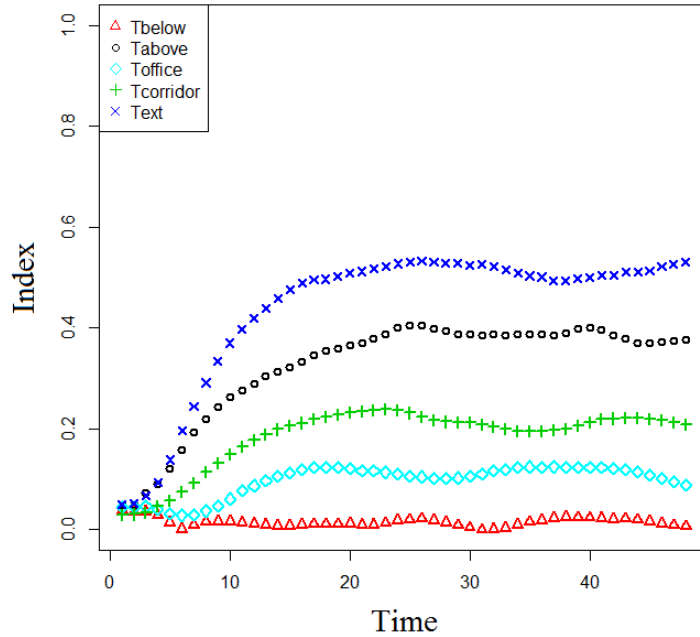


Figure 5: Plots of estimated Sobol indices

same interpretation. In a dynamic context the index varies with time but it is calculated as in a static case, that is to say that the input and the output are frozen at the same  $t$  moment. With our proposition, we set in relief the dynamic aspect of the system by taking into account all the past of the input variables.

To estimate this index in the dependent framework we have chosen a Pick and Freeze method because it is flexible : it works whatever the nature of the input (dynamic or static) and whatever the number of inputs. The only problem is that it requires independent inputs. Yet, we can use this method even when the inputs are dependent and Gaussian because we can separate the inputs into two variables : one corresponding to the variable of interest (the one we need to study the sensitivity of the model) and another one which is totally independent. We can then apply the Pick and Freeze method on these new variables. We propose an algorithm to separate the variables and to calculate the index.

In the case of stationary variables this index approaches a limit quickly. On a computational perspective, this allows us to reduce the computation time, by only simulating the first instants of the process. This method requiring to calculate the inverse covariance matrix of the inputs, *VAR* inputs have several advantages. Their covariance matrix can be calculated analytically and can be inversed easily thanks to their Toeplitz properties. These processes are quickly simulated; this is an advantage when we use the Pick and Freeze method which requires many input samples.



# Appendices

## A Convergence of $S_t^X$ in the case where $Y_t = f_t(\mathbf{U}_t, \dots, \mathbf{U}_0)$

**Assumption 2.**  $(\mathbf{U}_t)_{t \in \mathbb{Z}}$  is a causal regular linear process.

Causality means that  $\mathbf{U}_t$  does not depend on a distant past. Formally if  $H_s^U$  is the Hilbert space (for the covariance scalar product) generated by  $\{\mathbf{U}_v, v \leq s\}$  then  $\bigcap_{s \in \mathbb{Z}} H_s^U = \{0\}$ . Causality and linearity are equivalent to the existence of a representation :

$$\mathbf{U}_t = \sum_{k=0}^{\infty} C_k \boldsymbol{\omega}_{t-k} \quad (29)$$

where  $\{\boldsymbol{\omega}_t\}$  is a vectorial white noise,  $\mathbf{E}(\boldsymbol{\omega}_t \boldsymbol{\omega}_t^*) = \Theta$  and  $C_k$  are  $(p \times p)$  matrices such that  $\sum \|C_k\|^2 < \infty$ .

Basic examples are *VARMA* processes associated to recurrence equations :

$$P(d)(\mathbf{U}_t) = Q(d)(\boldsymbol{\omega}_t)$$

where  $d$  is the backward operator defined as  $d(\mathbf{U}_t) = \mathbf{U}_{t-1}$ ,  $d(\boldsymbol{\omega}_t) = \boldsymbol{\omega}_{t-1}$  and  $P, Q$  are polynomials such that  $|P(z)| \neq 0$  for  $|z| < 1$  for stationarity and  $|Q(z)| \neq 0$  for regularity ( $H_s^\omega = H_s^U$  for every  $-\infty < s < +\infty$ ).

We need the following truncation result. Let  $\eta$  fixed. We can find  $K$  such that :

$$\|\mathbf{U}_t - \mathbf{U}_t^{\tau(K)}\| < \eta$$

where  $\mathbf{U}_t^{\tau(K)} = \sum_{k=t}^{t-K} C_k \boldsymbol{\omega}_{t-k}$

$$\mathbf{E} \left( \left| \sum_{k=t-K+1}^{\infty} C_k \boldsymbol{\omega}_{t-k} \right| \right) \leq \sum_{k=K+1}^{\infty} \|\Theta\| \|C_k\|^2$$

**Assumption 3.** on the output

- $Y_t^*$  has the denominated truncation property as  $t \rightarrow \infty$  :

$$\lim_{t \rightarrow \infty} \mathbf{E}(|Y_t^* - Y_t|^2) = 0$$

and

$$\lim_{t \rightarrow \infty} |E(|Y_t^*|^2) - E(|Y_t|^2)| = 0$$

- $Y_t^*$  has the conditional past truncation property as  $t \rightarrow \infty$  :

$$\lim_{t \rightarrow \infty} \mathbf{E}(\mathbf{E}(Y_t^* | X_{[-\infty, t]})^2) - \mathbf{E}(\mathbf{E}(Y_t^* | X_{[0, t]})^2) = 0$$

From the Jensen inequality :

$$\mathbf{E} \left( \mathbf{E}(Y_t^* | X_{[0,t]}) - \mathbf{E}(Y_t | X_{[0,t]}) \right)^2 \leq \mathbf{E}(Y_t^* - Y_t)^2$$

tends to zero as  $t \rightarrow \infty$

$\mathbf{E} \left( \mathbf{E}(Y_t^* | X_{[-\infty,t]})^2 \right)$  is constant by translation invariance and denoted  $(S_t^X)^* \cdot \mathbf{Var}(Y_t)$ .

From assumption 2  $\mathbf{E}(\mathbf{E}(Y_t^* | X_{[0,t]})^2) \rightarrow (S_t^X)^* \cdot \mathbf{Var}(Y_t)$  as  $t \rightarrow \infty$

By assumption 1 :

We can rewrite :

$$\mathbf{E}(Y_t | X_{[0,t]}) = \mathbf{E}(Y_t^* | X_{[0,t]}) + \mathbf{E}(Y_t - Y_t^* | X_{[0,t]})$$

Then :

$$\mathbf{E}(\mathbf{E}(Y_t | X_{[0,t]})^2) = \mathbf{E}(\mathbf{E}(Y_t^* | X_{[0,t]})^2) + \mathbf{E}(\mathbf{E}(Y_t - Y_t^* | X_{[0,t]})^2) + 2\mathbf{E}(\mathbf{E}(Y_t^* | X_{[0,t]})\mathbf{E}(Y_t - Y_t^* | X_{[0,t]}))$$

When  $t$  tends to  $+\infty$  :

$$\begin{aligned} \lim_{t \rightarrow \infty} \mathbf{E}(\mathbf{E}(Y_t^* | X_{[0,t]})^2) &= (S_t^X)^* \cdot \mathbf{Var}(Y_t) \\ \lim_{t \rightarrow \infty} \mathbf{E}(\mathbf{E}(Y_t - Y_t^* | X_{[0,t]})^2) &= 0 \\ \lim_{t \rightarrow \infty} \mathbf{E}(\mathbf{E}(Y_t^* | X_{[0,t]})\mathbf{E}(Y_t - Y_t^* | X_{[0,t]})) &= 0 \text{ from Schwartz inequality} \end{aligned}$$

So under the hypotheses 1 and 2 :

$$\lim_{t \rightarrow \infty} S_t^X = (S_t^X)^* \tag{30}$$

**Example :** Let us check assumptions 1 and 2 for the example :

$$Y_t^* = \alpha Y_{t-1}^* + BU_t$$

for  $U_t$  causal regular Gaussian process and  $|\alpha| < 1$ .

$Y_t$  can be written as :

$$Y_t^* = \sum_{k=0}^{\infty} \alpha^k BU_{t-k}$$

and

$$Y_t = \sum_{k=0}^t \alpha^k BU_{t-k}$$

thus :

$$\mathbf{E}|Y_t^* - Y_t|^2 \leq C|\alpha|^{2t}$$

thus the assumption 1 is verify. Now we calculate :

$$\mathbf{E}(Y_t^* | X_{[-\infty,t]}) = \sum_{k=0}^{\infty} \alpha^k B \mathbf{E}(U_{t-k} | X_{[-\infty,t]})$$

we know that

$$\mathbf{E}(\mathbf{U}_{t-k} | X_{[-\infty, t]}) = \sum_{j=0}^{\infty} C_j \mathbf{E}(\omega_{t-j} | X_{[-\infty, t]})$$

and if  $\eta$  given it exists  $K$ .

$\|U_{t-k} - U_{t-k}^K\| \leq \eta$  implies  $\|\mathbf{E}(U_{t-k}) - \mathbf{E}(U_{t-k}^K)\| \leq \eta$  for  $K$  large enough from the truncation property of causal processes. But from the regularity of  $\mathbf{U}_t$  we have  $:H_u^X = H_u^\omega$  for all  $u$  thus in Gaussian case :  $\mathbf{E}(\omega_{t-j} | X_{[-\infty, t]}) = \mathbf{E}(\omega_{t-j} | \omega_{[-\infty, t]})$ . But for  $t - j > 0$ ,  $\omega_{t-j}$  is independent of  $\{\omega_s, s < 0\}$ .

Thus

$$\begin{aligned} \mathbf{E}(\omega_{t-j} | X_{[-\infty, t]}) &= \mathbf{E}(\omega_{t-j} | \omega_{[-\infty, t]}) \\ &= \mathbf{E}(\omega_{t-j} | \omega_{[0, t]}) \\ &= \mathbf{E}(\omega_{t-j} | X[0, t]) \end{aligned}$$

and we have proved the result choosing  $t$  large enough to have  $t - K > 0$ .

## B Construction of the statistical metamodel

To build a model easy to use for simulations, we need to preprocess the data in order to be placed in a stationary condition. We built a model for summers.  $T_t^e$  is a scalar time series of input data. Preprocessing means first to transform  $\bar{T}_t^e$  into  $T_t^e$  :

$$\bar{T}_t^e = S(t) + V(t) * T_t^e \quad \forall e \in \mathcal{E}$$

where  $\mathcal{E} = \{\text{int, ext, above, below, cor, off}\}$ ,  $S(t)$  is the mean with period 24 hours,  $V(t)$  the variance periodic function and  $T_t^e$  a stationary or cyclo-stationary process.

Temperatures ( $T^{\text{below}}, T^{\text{above}}, T^{\text{off}}, T^{\text{cor}}, T^{\text{ext}}, T^{\text{int}}$ ) are modelled by a VAR process for working days:

$$\forall e \in \mathcal{E}, \quad T_t^e = \sum_{f \in \mathcal{E}} \gamma_f^e T_{t-1}^f + \omega_t^e,$$

where  $(\omega_t^e)_{t \in \mathbb{N}, e \in \mathcal{E}}$  are Gaussian variables with covariance  $\Theta$ . Let  $\mathbf{T}_t = (T_t^e)_{e \in \mathcal{E}}$ ,  $\boldsymbol{\omega}_t = (\omega_t^e)_{e \in \mathcal{E}}$  thus :

$$\mathbf{T}_t = \sum_{l=1}^p D_l \mathbf{T}_{t-p} + \boldsymbol{\omega}_t$$

For every fixed  $p$  we estimated  $(D_l, l \leq 1, \dots, p)$  and  $\Theta$  the covariance of  $\boldsymbol{\omega}$  by maximum of likelihood under the constraint imposed by the stationarity of  $\mathbf{T}_t$ . Then we choose  $p$  using AIC criteria; obtaining  $p = 2$ .

$D_1$  and  $D_2$  estimated are :

$$D_1 = \begin{pmatrix} 0.88 & 0.01 & 0.06 & 0.06 & 0 \\ -0.23 & 1.21 & 0.01 & 0.49 & 0.3 \\ 0.21 & 0.03 & 1.24 & 0.06 & 0.01 \\ 0.12 & 0.06 & 0.03 & 1 & 0.01 \\ 0.20 & 0.61 & -0.19 & 0.71 & 0.85 \end{pmatrix} \quad D_2 = \begin{pmatrix} 0.04 & -0.01 & -0.07 & -0.04 & 0.01 \\ 0.03 & -0.48 & -0.04 & -0.08 & -0.04 \\ -0.16 & -0.02 & -0.30 & -0.09 & -0.10 \\ -0.03 & -0.04 & -0.05 & -0.15 & -0.01 \\ -0.45 & -0.60 & 0.25 & -0.31 & -0.05 \end{pmatrix}$$

The  $\Theta$  matrix estimated is :

$$\Theta = \begin{matrix} T^{\text{below}} \\ T^{\text{above}} \\ T^{\text{off}} \\ T^{\text{cor}} \\ T^{\text{ext}} \end{matrix} \begin{pmatrix} 0.01 & 0.01 & 0.004 & 0.01 & 0.01 \\ 0.01 & 0.62 & 0.02 & 0.03 & 0.46 \\ 0.004 & 0.02 & 0.03 & 0.002 & 0.04 \\ 0.01 & 0.03 & 0.002 & 0.02 & 0.04 \\ 0.01 & 0.46 & 0.04 & 0.04 & 1.34 \end{pmatrix} \quad (31)$$

The input-output system is modelled by an auto-regression of  $T^{\text{int}}$  on variables  $\mathbf{U}_t = (T^{\text{below}}, T^{\text{above}}, T^{\text{off}}, T^{\text{cor}}, T^{\text{ext}})$  :

$$\forall t \geq 1, T^{\text{int}}(t) = \sum_{e \in \mathcal{E}, k=1}^2 \phi_{e,k} T_{e,t-k} + \sum_{k=1}^2 \phi_{\text{int},k} T_{t-k}^{\text{int}} \quad (32)$$

where  $\mathcal{E} = \{\text{int}, \text{above}, \text{below}, \text{cor}, \text{off}, \text{ext}, \}$

## Acknowledgements

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

## References

- [1] Andrea Saltelli, Karen Chan, E Marian Scott, et al. *Sensitivity analysis*, volume 1. Wiley New York, 2000.
- [2] Ruixue Liu and Art B Owen. Estimating mean dimensionality of analysis of variance decompositions. *Journal of the American Statistical Association*, 101(474):712–721, 2006.
- [3] Il'ya M. Sobol. Sensitivity estimates for nonlinear mathematical models. *Math. Modeling Comput. Experiment*, 1(4):407–414 (1995), 1993.
- [4] Il'ya M. Sobol. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55(1-3):271–280, 2001.
- [5] Thierry A Mara and Stefano Tarantola. Variance-based sensitivity indices for models with dependent inputs. *Reliability Engineering & System Safety*, 2011.
- [6] Sergei Kucherenko, Stefano Tarantola, and Paola Annoni. Estimation of global sensitivity indices for models with dependent variables. *Computer Physics Communications*, 183(4):937–946, 2012.
- [7] Gaëlle Chastaing, Fabrice Gamboa, and Clémentine Prieur. Generalized hoeffding-sobol decomposition for dependent variables-application to sensitivity analysis. *Electronic Journal of Statistics*, 6:2420–2448, 2012.

- [8] Fabrice Gamboa, Alexandre Janon, Thierry Klein, Agnes Lagnoux-Renaudie, and Clémentine Prieur. Statistical inference for sobol pick freeze monte carlo method. *arXiv preprint arXiv:1303.6447*, 2013.
- [9] G. Chastaing, F. Gamboa, and C. Prieur. Generalized hoeffding-sobol decomposition for dependent variables-application to sensitivity analysis. *Arxiv preprint arXiv:1112.1788*, 2011.
- [10] Alexandre Janon, Thierry Klein, Agnès Lagnoux, Maëlle Nodet, and Clémentine Prieur. Asymptotic normality and efficiency of two sobol index estimators. Preprint available at <http://hal.inria.fr/hal-00665048/en>, 2012.
- [11] Peter J. Brockwell and Richard A. Davis. *Time Series: Theory and Methods*. Springer, 2nd ed. 1991. 2nd printing 2009 edition, April 2009.
- [12] Robert Faivre, Bertrand Iooss, Stéphanie Mahévas, David Makowski, and Hervé Monod. *Analyse de sensibilité et exploration de modèles: Application aux sciences de la nature et de l'environnement*. Editions Quae, 2013.