Variance-Based Sensitivity Indices For Models With Dependent Inputs

Thierry A. Mara[a,*], Stefano Tarantola[b]

[a] *PIMENT/MASC, University of Reunion Island, 15 Avenue Rene Cassin 97715 Saint-Denis, Reunion*
[b] *Institute for the Protection and Security of the Citizen, Joint Research Centre, European Commission, Ispra (VA), Italy*

Abstract

Computational models are intensively used in engineering for risk analysis or prediction of future outcomes. Uncertainty and sensitivity analyses are of great help in these purposes. Although several methods exist to perform variance-based sensitivity analysis of model output with independent inputs only a few are proposed in the literature in the case of dependent inputs. This is explained by the fact that the theoretical framework for the independent case is set and an univocal set of variance-based sensitivity indices is defined. In the present work, we propose a set of variance-based sensitivity indices to perform sensitivity analysis of models with dependent inputs. These measures allow us to distinguish between the mutual dependent contribution and the independent contribution of an input to the model response variance. Their definition relies on a specific orthogonalisation of the inputs and ANOVA-representations of the model output. In the applications, we show the interest of the new sensitivity indices for model simplification setting.

Keywords: ANOVA-HDMR, dependent inputs, variance-based sensitivity indices, dependent contributions, independent contributions

*Corresponding author: mara@univ-reunion.fr

Contents

1 Introduction **3**

2 Sensitivity analysis of model output with independent inputs **5**

2.1 The ANOVA-representation 5

2.2 The variance-based sensitivity indices 6

3 Sensitivity analysis of model output with dependent inputs **7**

3.1 Dependences among random variables 7

3.2 Orthogonalisation of the dependent inputs 8

3.3 Interpretation of the sensitivity indices 9

3.4 Computational issues 11

4 Numerical test cases **11**

4.1 Analytical test: a linear model 12

4.2 Computational test : a non-linear model 14

5 Conclusion **17**

6 Appendix **18**

7 References **20**

8 Tables **24**

9 Figures **27**

1. Introduction

Simulation models are becoming more and more popular nowadays for forecasting, process control or systems characterization. Whatever the issue the computer model is required to answer, one needs to feed some input values into it. Such inputs are necessarily known at some degree of uncertainty, and they are treated as random variables. It is then an issue to investigate the inputs mainly responsible for the response variance (here assumed as representative of the uncertainty). Affordable model-free computational methods for evaluating the relative importance of independent inputs on the output variance are now available (Sudret 2008, Saltelli *et al.* 1999, Rabitz *et al.* 1999, Morris 1991, Saltelli 2002, Mara 2009, Tarantola *et al.* 2006). Yet, only a few methods and applications are proposed in the literature to tackle models with dependent inputs (for instance, McKay 1996, Saltelli & Tarantola 2002, Xu & Gertner 2008a-2008b, Borgonovo & Tarantola 2008, Bedford 1998, Li *et al.* 2010). In McKay (1996), the author derived the so-called replicated latin hypercube sampling technique (r-LHS) to compute marginal contribution of input factors to the output variance, that is,

$$S_i = V[E[y|x_i]]/V_y, \tag{1}$$

where y denotes the model response and $\mathbf{x} = \{x_1, \ldots, x_n\}$ is a set of dependent inputs characterized by a joint probability density function $\mathbf{p}(\mathbf{x})$, V_y is the total variance of y, S_i is the marginal sensitivity index of the input x_i, $V[\cdot]$ stands for variance operator and $E[\cdot|\cdot]$ for the conditional expectation operator. The sensitivity index defined by equation (1) represents the amount of the response variance due to the factor x_i. If x_i and the other inputs are dependent, then the value of S_i can be influenced by this dependence. Saltelli & Tarantola (2002) applied r-LHS sampling with McKay's method

to the Level E model with dependent inputs in a variance reduction setting (the readers are referred to the cited paper for more details about settings in sensitivity analysis).

In Xu & Gertner (2008a), the authors extended the original procedure of Iman & Conover (1982), for generating correlated LHS samples to the Random Balance Design technique proposed by Tarantola *et al.* (2006) in order to compute all the S_i's. They showed that the method outperformed r-LHS in terms of computational cost since it only requires one single sample set. Then, in a second article, the same authors Xu & Gertner (2008b) proposed a computational approach to estimate the correlated S_i^c and the uncorrelated S_i^u contribution of x_i on the model response y, namely,

$$S_i = S_i^u + S_i^c.$$

This allows the investigation of spurious inputs that have an impact on the model response only because of their strong correlations with the other significant ones. Indeed, a large S_i with a low S_i^u indicates that the contribution of x_i to the response variance is only due to its correlation with the other inputs. The drawback of their approach is that only linear models with correlations (linear dependences) are supported.

In Li *et al.* (2010), the authors propose to build a random sampling high dimensional model representation (RS-HDMR) of the original model in order to be able to evaluate the correlative and the structural contribution of an input to the response variance. This allows to investigate whether the model is structurally a function of all the inputs or only of some of them. They defined new sensitivity indices that are different of those of Xu & Gertner (2008a). According to their results, it seems that the method performs well for linear models.

Our objective in this work is to derive a set of variance-based sensitivity

indices that can support non-linear models and non-linear dependences. The new sensitivity indices are more related to those of Xu & Gertner (2008a) than those of Li *et al.* (2010). This is achieved by performing the ANOVA-HDMR after *decorrelating* the inputs. To this aim, we preliminarily decorrelate the inputs with the Gram-Schmidt procedure . The computation of the sensitivity indices of the independent variables is then straightforward. We show that the variance-based sensitivity indices, so obtained, can be interpreted as the fully, partially correlated and independent contributions of the inputs to the ouput variance.

The rest of the paper is organized as follows; in § 2 we recall the basic concept of the ANOVA decomposition and the definition of the variance-based sensitivity indices. In § 3, we describe an algorithm to generate independent variables from a set of dependent inputs. We also discuss the link between the sensitivity indices of the new variables and those of the original inputs. In § 4, two numerical examples are treated in which the sensitivity indices are computed analytically and numerically before concluding (§ 5).

2. Sensitivity analysis of model output with independent inputs

2.1. The ANOVA-representation

Let $y = f(\mathbf{x})$ be a square-integrable function of n independent random variables $\mathbf{x} = \{x_1, \ldots, x_n\}$. Variance-based sensitivity indices result from the ANOVA decomposition of Sobol (1993). This decomposition can be summed up as follows: $f(\mathbf{x})$ can always be expanded into summands of different dimensions, that is,

$$f(\mathbf{x}) = f_0 + \sum_{i=1}^{n} f_i(x_i) + \sum_{j>i}^{n} f_{ij}(x_i, x_j) + \ldots + f_{1\ldots n}(x_1, \ldots, x_n), \qquad (2)$$

with $E[f_{i_1\ldots i_s}(x_{i_1},\ldots,x_{i_s})]=0$, $i_q=i_1,\ldots,i_s$ and $\mathbf{x_{-i_q}}=\mathbf{x}\backslash\{x_{i_q}\}$. Hence, the functions are pairwise orthogonal, that is,

$$E\Big[f_{i_r\ldots i_s}(x_{i_r},\ldots,x_{i_s})\times f_{i_q\ldots i_t}(x_{i_q},\ldots,x_{i_t})\Big]=0,\ \forall(i_r,\ldots,i_s)\neq(i_q,\ldots,i_t). \tag{3}$$

Expansion (2) is unique only if the inputs are independent. Then, it can be easily deduced that,

$$V_y=\sum_{i=1}^{n}V_i+\sum_{j>i}^{n}V_{ij}+\ldots+V_{1\ldots n}, \tag{4}$$

where V_y is the total variance of $f(\mathbf{x})$, V_i is the marginal variance of x_i and $V_{i_1\ldots i_s}$ is the cooperative fractional variance of $\{x_{i_1},\ldots,x_{i_s}\}$ which quantifies the interactions between $\{x_{i_1},\ldots,x_{i_s}\}$.

2.2. The variance-based sensitivity indices

Eq.(4) leads to the definition of variance-based sensitivity indices,

$$S_{i_1\ldots i_s}=V_{i_1\ldots i_s}/V_y, \tag{5}$$

where $S_i=V_i/V_y$ is the first-order sensitivity index, which measures the amount of the variance of y explained by x_i alone. $S_{ij}=V_{ij}/V_y$ is the second-order sensitivity index that measures the amount of the variance explained by the interaction between x_i and x_j and so on. In practice, two variance-based sensitivity indices are investigated, the first-order sensitivity index S_i defined earlier and the total sensitivity index defined as,

$$ST_i=S_i+\sum_{j\neq i}^{n}S_{ij}+\ldots+S_{1\ldots i\ldots n}. \tag{6}$$

Indeed, while S_i only measures the marginal contribution of x_i to the variance of the outcome, ST_i also includes its cooperative contribution with all the other inputs. These sensitivity indices allow to classify $f(\mathbf{x})$ as additive

($\sum_1^n S_i = 1$) or non-additive with interactions ($\sum_1^n S_i \ll 1$). Consequently, in the case of independent inputs, the ANOVA–HDMR (2) reflects the model structure. In the case of dependent variables, an expansion such as equation (2) does not reflect the model structure anymore (Oakley & O'Hagan 2004).

3. Sensitivity analysis of model output with dependent inputs

3.1. Dependences among random variables

Pearson and Spearman correlation coefficients are often employed to characterize the linear relationship between two random variables. In particular, the correlation matrix, a symmetric positive-definite matrix composed of Pearson correlation coefficients, uniquely defines the dependence of normally distributed variables. Moreover, if $\mathbf{x}$ is a set of normal correlated variables, $E[x_i|x_j]$ defines the marginal relationship between the two random variables. Hence, if x_i and x_j are independent, $E[x_i|x_j] = E[x_i]$. In the same manner, $E[x_i|x_j, x_k]$ characterizes the dependence of x_i to both x_j and x_k. For non-normal distributions, higher conditional moments are required to characterize the relationship between the random variables. In the sequel, we assume that only the first-order conditional moment characterizes dependencies between the random variables. In this case, pairwise dependencies are monotonic and homoscedastic. These assumptions are satisfied, for instance, if the dependence structure is defined by a correlation matrix or a rank correlation matrix.

Let us recall the following relationship for three dependent inputs,

$$p(x_1, x_2, x_3) = p(x_1)p(x_2|x_1)p(x_3|x_1, x_2), \tag{7}$$

where $p(x_1, x_2, x_3)$ is the joint probability density function (pdf), $p(x_1)$ is the marginal distribution of x_1, $p(x_2|x_1)$ the marginal pdf of x_2 conditioned on

x_1 and $p(x_3|x_1, x_2)$ the marginal pdf of x_3 conditioned on $\{x_1, x_2\}$. Under the previous assumption, x_2 is dependent of x_1 because $E[x_2|x_1] \neq E[x_2]$. Similarly, x_3 depends on $\{x_1, x_2\}$ if $E[x_3|x_1, x_2] \neq E[x_3]$. As a consequence, if we set,

$$x_{2-1} = x_2 - E[x_2|x_1], x_{3-12} = x_3 - E[x_3|x_1, x_2],$$

then,

$$p(x_1, x_{2-1}, x_{3-12}) = p(x_1)p(x_{2-1})p(x_{3-12}). \tag{8}$$

This transformation is nothing else but the one of Rosenblatt (1952) for normal variables (as defined in Lebrun & Dutfoy 2009, regardless the standardisation) and $\{x_1, x_{2-1}, x_{3-12}\}$ is a set of independent variables.

3.2. Orthogonalisation of the dependent inputs

We discuss in this subsection how to generate an independent sample set from a dependent one. Let us assume, without loss of generality, a set of standardised dependent random variables $\mathbf{x}$. A procedure to derive a set of orthogonal variables is the following,

$$\bar{x}_1 = x_1, \tag{9}$$

$$\bar{x}_i = x_i - E[x_i|\bar{x}_1, \ldots, \bar{x}_{i-1}], \forall i = 2, \ldots, n. \tag{10}$$

Under the assumption that only the first-order conditional moment characterizes the dependences between the inputs, the new variables are orthogonal and independent. A particular case, although common, is when the dependent sample has been generated from a correlation matrix or rank correlation matrix (*e.g.* with the technique of Iman & Conover 1982). Then, the correlated variables are pairwise dependent, that is,

$$E[x_i|\mathbf{x}_{-\mathbf{i}}] = \sum_{j \neq i}^{n} E[x_i|x_j]. \tag{11}$$

As a consequence, $E[x_i|\bar{x}_1,\ldots,\bar{x}_{i-1}]$ can be approximated by one-dimensional regressions (Lewandowski *et al.* 2007, Storlie & Helton 2008, Mara & Rakoto Joseph 2008). Figure 1 depicts the results of the decorrelation procedure for several dependence structures between two random variables. If equation (11) is not satisfied, then $E[x_i|\bar{x}_1,\ldots,\bar{x}_{i-1}]$ shall be approximated by a complex multidimensional function similar to equation (2).

Finally, one can perform the ANOVA-HDMR of the model and define the sensitivity indices of the new variables $\{\bar{S}_i, \bar{S}_{ij}, \cdots, \bar{S}_{1\cdots n}\}$. The orthogonal set so obtained is not unique as it depends on the order of the inputs in the original set. Actually, $n!$ orthogonal sets can be generated in this manner and a huge number of sensitivity indices can be defined.

Applying the procedure given in (Equations 9-10) does not in general provide a set of independent variables, especially when the inputs are not normally distributed. An alternative is then to use Nataf isoprobabilistic transformation which consists in generating orthogonal normal random variables from any set of dependent variables (Nataf 1962). In Lebrun & Dutfoy 2009, it is proved that Rosenblatt and Nataf transformations are equivalent for joint normal distributions (*i.e.* to the procedure given in Equations 9-10). However, Nataf transformation could make the relationship between the model response and the new variables more complex than using equations (9-10).

3.3. Interpretation of the sensitivity indices

The new sensitivity indices are interpreted as follows,

- $\bar{S}_1 = \dfrac{V\left[E\left[f(\mathbf{x})|\bar{x}_1\right]\right]}{V\left[f(\mathbf{x})\right]}$, is the well-known main effect of $\bar{x}_1$. But since $x_1 = \bar{x}_1$, $\bar{S}_1 = S_1$ is the *full* marginal contribution of x_1 to the variance of $f(\mathbf{x})$.

- $\bar{S}_2 = \frac{V\Big[E\big[f(\mathbf{x})|\bar{x}_2\big]\Big]}{V\big[f(\mathbf{x})\big]} = S_{2-1}$, is the marginal contribution of x_2 to the variance of $f(\mathbf{x})$ without its *correlative* contribution with x_1, since $\bar{x}_2$ is not correlated to x_1,
- $\bar{S}_3 = \frac{V\Big[E\big[f(\mathbf{x})|\bar{x}_3\big]\Big]}{V\big[f(\mathbf{x})\big]} = S_{3-12}$, is the marginal contribution of x_3 to the variance of $f(\mathbf{x})$ without its *correlative* contribution with $\{x_1, x_2\}$,
- $\cdots$,
- $\bar{S}_n = \frac{V\Big[E\big[f(\mathbf{x})|\bar{x}_n\big]\Big]}{V\big[f(\mathbf{x})\big]} = S_n^u$, is the uncorrelated marginal contribution of x_n to the variance of $f(\mathbf{x})$.

Indeed, $\bar{x}_1$ keeps all information concerning x_1 including its common part with the other variables. While $\bar{x}_n$, the last factor in the iterative procedure, only contains the proper information of x_n excluding its dependent part. For convenience, in the sequel, we also note $\bar{x}_2 = x_{2-1}$, $\bar{x}_3 = x_{3-12}$ and so on.

It can also be infered that,

- $\bar{S}_{12}^c = \frac{V\Big[E\big[f(\mathbf{x})|\bar{x}_1,\bar{x}_2\big]\Big]}{V\big[f(\mathbf{x})\big]} = \frac{V\Big[E\big[f(\mathbf{x})|x_1,x_2\big]\Big]}{V\big[f(\mathbf{x})\big]} = S_{12}^c$, since to derive $\{\bar{x}_1, \bar{x}_2\}$ only $\{x_1, x_2\}$ are required.
- $\bar{S}_{123}^c = \frac{V\Big[E\big[f(\mathbf{x})|\bar{x}_1,\bar{x}_2,\bar{x}_3\big]\Big]}{V\big[f(\mathbf{x})\big]} = S_{123}^c$, since to derive $\{\bar{x}_1, \bar{x}_2, \bar{x}_3\}$ only $\{x_1, x_2, x_3\}$ are necessary, and so on,
- $\cdots$,
- $ST_n^u = 1 - \frac{V\Big[E\big[f(\mathbf{x})|\bar{\mathbf{x}}_{-n}\big]\Big]}{V\big[f(\mathbf{x})\big]}$, is the uncorrelated total contribution of x_n to the variance of $f(\mathbf{x})$.

Besides, according to the law of total variance, that is,

$$V[f(\mathbf{x})] = V[E[f(\mathbf{x})|\mathbf{x}_{-n}]] + E[V[f(\mathbf{x})|\mathbf{x}_{-n}]], \tag{12}$$

[Insert Figure 1 about here]

we can also infer that,

$$ST_n^u = \frac{E[V[f(\mathbf{x})|\bar{\mathbf{x}}_{-n}]]}{V[f(\mathbf{x})]} = \frac{E[V[f(\mathbf{x})|\mathbf{x}_{-n}]]}{V[f(\mathbf{x})]}. \tag{13}$$

3.4. Computational issues

In order to numerically compute the new sensitivity indices, one has to: (i) generate a sample set of correlated inputs, (ii) deduce from the original dependent sample, a set of independent samples accordingly with equations (9-10), (iii) compute the variance-based sensitivity indices of interest. We achieve the first step using the procedure of Iman & Conover (1982), which allows us to generate LHS correlated samples by imposing a given rank correlation matrix. In alternative, the theory of copulas provides computational methods to generate dependent samples (see Nelsen 2006). Once a correlated sample has been obtained, step (ii) is achieved by approximating $E[x_i|\bar{x}_1, \ldots, \bar{x}_{i-1}]$. This is easily achieved if the inputs are pairwise dependent as discussed previously. In step (ii), the estimation of the new sensitivity indices can be made using RS-HDMR (Li *et al.* 2008) which consists in performing an ANOVA decomposition of the model of interest. Polynomial regressions are other possible approximations of the model (these approximations are called surrogate models) and in the case of orthogonal polynomials (e.g. Legendre, Hermite polynomials) the surrogate modelling approach is called polynomial chaos expansion (Sudret 2008).

4. Numerical test cases

In the following two examples we are interested in the reduction of the dimensionality of the input space for the analysed model for the subsequent

development of a surrogate model with the smallest possible number of inputs. To identify the inputs that could be fixed we look at the unconditional total indices $\widehat{ST}_i^u$, which have to be as small as possible. In the case of uncorrelated inputs, this procedure coincides with the *factors fixing setting* of Saltelli et al. 2006. However, in the presence of correlation between inputs, we have to test n different orderings of the inputs and the $\widehat{ST}_i^u$ can be estimated for each of these orders (see Table 3 for a non-linear model). More simply, the first case study is a linear model, therefore the same investigation can be done using only first order sensitivity indices (see Table 1).

4.1. Analytical test: a linear model

In the first case, we consider the linear simple model $y = x_1 + x_2 + x_3$, where the x_i's are standard normal random variables with the following correlation matrix,

$$\mathbf{C} = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix}.$$

In this case, the new independent random variables $\bar{\mathbf{x}}$ are also normally distributed and it can be proved that (see in appendix):

$$\begin{aligned} V_y &= 3 + 2(\rho_{12} + \rho_{13} + \rho_{23}), \\ V_i &= V_{x_i}(E(y|x_i)) = \Big(1 + \rho_{ij} + \rho_{ik}\Big)^2, \\ V_i^u &= E(V(y|x_j, x_k)) = 1 + \Big(2\rho_{12}\rho_{13}\rho_{23} - \rho_{ij}^2 - \rho_{ik}^2\Big)\Big/(1 - \rho_{jk}^2), \end{aligned}$$

with $i \neq j \neq k$ and $S_i = V_i/V_y$, $S_i^u = V_i^u/V_y$. We investigate the behaviour of such estimates for some different correlation scenarios. Since the model is linear, we only focus on the marginal sensitivity indices S_i and the uncorre-

lated part S_i^u for different sets of correlation coefficient $(\rho_{12}, \rho_{13}, \rho_{23})$ (indeed, $S_i = ST_i$, $S_i^u = ST_i^u$). The results are gathered in Table 1.

For the case $(\rho_{12}, \rho_{13}, \rho_{23}) = (0.5, 0.8, 0)$, we find that the original three-dimensional function can be reduced to a one-dimensional model that only depends on x_1. x_1 explains 94% of the output variance, while the remaining contribution due to $\{x_{2-1}, x_{3-12}\}$ is about 6% (see Table 1). Indeed, six (*i.e.* 3!) equivalent polynomial expansions can be derived with the proposed method. In particular, with the set $\{x_1, x_2, x_3\}$, the following surrogate model is obtained (see in appendix for details):

$$y = \frac{23}{10}x_1 + \frac{7}{15}x_{2-1} + x_{3-12} \simeq \frac{23}{10}x_1,$$

where $x_1 \sim \mathcal{N}(0,1)$, $x_{2-1} \sim \mathcal{N}(0, \frac{3}{4})$, $x_{3-12} \sim \mathcal{N}(0, \frac{11}{75})$.

For the set $\{x_2, x_3, x_1\}$ ($\{x_3, x_1, x_2\}$ resp.) a two dimensional first-order polynomial function of $\{x_2, x_{3-2}\}$ ($\{x_3, x_{1-3}\}$ resp.) is necessary. The expressions of the surrogate models are respectively,

$$y = \frac{3}{2}x_2 + \frac{9}{5}x_{3-2} + x_{1-23} \simeq \frac{3}{2}x_2 + \frac{9}{5}x_{3-2},$$
$$\text{with } x_2 \sim \mathcal{N}(0,1), x_{3-2} \sim \mathcal{N}(0,1), x_{1-23} \sim \mathcal{N}(0, \frac{11}{100}),$$
$$y = \frac{9}{5}x_3 + \frac{43}{18}x_{1-3} + x_{2-13} \simeq \frac{9}{5}x_3 + \frac{43}{18}x_{1-3},$$
$$\text{with } x_3 \sim \mathcal{N}(0,1), x_{1-3} \sim \mathcal{N}(0, \frac{9}{25}), x_{2-13} \sim \mathcal{N}(0, \frac{11}{36}).$$

These three approximations are possible ANOVA-representations of the original model. But the first one has the lowest dimension. This means that y only depends on x_1 in the form given above (*i.e.* $y \simeq \frac{23}{10}x_1$) instead of the original one. This means that, in order to improve the knowledge on y, one should reduce the uncertainty of x_1. But, if the analyst is only able to improve its knowledge on x_2 and x_3 then the second polynomial is to be employed.

[Insert Table 1 about here]

In the next case we investigate negative correlations: $(\rho_{12}, \rho_{13}, \rho_{23}) = (-0.5, 0.2, -0.7)$. As expected (also shown in Xu & Gertner 2008b), the particularity of such a correlation structure is that the uncorrelated marginal contributions are larger than the *full* marginal contribution. Besides, $S_1 + S_{3-1} = S_3 + S_{1-3} = 0.97$ means that the variance of y is controlled by $\{x_1, x_3\}$ and their correlation.

Finally, let us examine a symmetric case $(\rho_{12}, \rho_{13}, \rho_{23}) = (-0.49, -0.49, -0.49)$ (we do not use $\rho_{ij} = -0.5$ because it would correspond to a non-positive definite matrix, which cannot represent a true correlation matrix). Here the output uncertainty is very low since the uncertainties of the inputs compensate each other in the model ($V_y = 0.06$). Indeed, the full sensitivity index is zero ($S_i = ST_i \simeq 0$) while the uncorrelated contribution of an input almost explains the whole response variance ($S_i^u = ST_i^u \simeq 0.98$). A possible ANOVA-representation of the original model is,

$$y = 0.02x_1 + 0.04x_{2-1} + x_{3-12} \simeq x_{3-12},$$

with $x_1 \sim \mathcal{N}(0, 1)$, $x_{2-1} \sim \mathcal{N}(0, 0.76)$, $x_{3-12} \sim \mathcal{N}(0, 0.058)$. Obviously, all the inputs are equally relevant for the model and it is not possible to further reduce the input dimension. Although y mainly depends on x_{3-12} in the previous equation, the determination of x_{3-12} requires the knowledge of x_1 and x_2.

4.2. Computational test : a non-linear model

The function we analyse in this example was introduced in Jacques *et al.* (2006), $y = x_1x_2 + x_3x_4 + x_5x_6$ where the six variables are standard normal and identically distributed with the correlation matrix,

$$\mathbf{C} = \begin{bmatrix} 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.0 & 0.3 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.3 & 1.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.8 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.8 & 1.0 \end{bmatrix}$$

$\{x_1, x_2\}$ is a set of independent inputs, $\{x_3, x_4\}$ are slightly correlated and $\{x_5, x_6\}$ are strongly correlated. In their paper, the authors proposed to treat this problem with the classical Sobol' method by computing the closed second-order sensitivity indices, S^c_{12}, S^c_{34}, S^c_{56} given that here $S^c_{12}+S^c_{34}+S^c_{56} = 1$. With our proposed methodology one can investigate the correlated and uncorrelated variance-based sensitivity indices as well as the closed-order sensitivity indices. Contrarily to the previous case, the sensitivity indices are investigated numerically. Indeed, after generating a correlated LHS sample of size $N = 4096$ and evaluating the function, the decorrelated samples are deduced thanks to the set of equations (9-10) by a least-square regression method (Mara & Rakoto Joseph 2008). Then, a polynomial chaos expansion is built from the decorrelated sample. For this purpose, we use the multi-dimensional Hermite polynomials since the inputs are normally distributed. The estimated variance-based sensitivity indices are then computed. The procedure is repeated six times by circularly permuting the original input set in order to estimate all the desired sensitivity indices.

The marginal sensitivity indices are gathered in Table 2. Most of the marginal contributions are null due to the preponderance of interactions within the function. Only the marginal contributions of x_5 and x_6 are non negligible: this is due to the combined effect of non-linearities and strong

correlation between these two inputs. Moreover, we can note that, as long as x_5 precedes x_6 in the input set, x_5 contains the full marginal effect while the uncorrelated marginal contribution of x_6 is null and conversely when x_6 is positioned before x_5 (last row in the table).

The uncorrelated total sensitivity indices indicate that all the inputs are important (cf. Table 3). The first noticeable result is that the total effects of x_1 and x_2 are identical and remain unchanged whatever the input order. This is explained by the fact that they are not correlated with the other ones and that they just both interact. For $\{x_3, x_4\}$ and $\{x_5, x_6\}$ we find the same behaviour in the results: the total sensitivity indices of x_5 is equal to 0.45 as long as it precedes x_6 in the set, otherwise its effect is 0.10 (the gap equals the uncorrelated marginal effect 0.35). This is easily explained since, as long as x_5 precedes x_6 in the set, only x_{6-5} is provided by the decorrelation procedure. So, the sensitivity indices estimated are the full sensitivity indices of x_5 and the uncorrelated sensitivity indices of x_6. Finally, let us note that $\{0.27, 0.28, 0.45\}$ which are the total sensitivity indices of $\{x_1, x_3, x_5\}$ (equivalently of $\{x_2, x_4, x_6\}$) sum up to 1 and are the estimated values of $\{S^c_{12}, S^c_{34}, S^c_{56}\}$. Actually the analytical values are $\{0.2681, 0.2922, 0.4397\}$ respectively.

It can be concluded that the model cannot be further reduced. All the inputs are relevant for the output variance ($ST^u_i \geqslant 0.1, \forall i = 1, 2, 3$). The significant gap between the marginal effects and the total effects means that interactions are preponderant. Since $S^c_{12} + S^c_{34} + S^c_{56} = 1$, fixing both $\{x_5, x_6\}$ to any values within their range of uncertainty will reduce the response variance by an amount of $S^c_{56} = 44\%$ given that the model is additive in the pairs $\{x_1, x_2\}$, $\{x_3, x_4\}$, $\{x_5, x_6\}$.

[**Insert Table 2 about here**]

[**Insert Table 3 about here**]

5. Conclusion

In this work we have described a computational method to perform sensitivity analysis of pairwise dependent inputs that relies on the Gram-Schmidt decorrelation procedure and the subsequent use of polynomial chaos expansion (PCE). This procedure generalizes the Xu & Gertner 2008b approach, which was proposed for linearly dependent inputs, to the case of conditionally dependent inputs. The PCE is an ANOVA decomposition of the original model and provides accurate estimates of variance-based sensitivity indices. Nevertheless, such an ANOVA representation is not unique in the case of dependent variables and in our approach it is directly linked to the decorrelation procedure. Indeed, several independent variable sets can be generated depending on the input ordering in the set and several variance-based sensitivity indices can be computed. In practice, according to the SA setting addressed, only a few indices are of interest. In particular, in the model simplification setting, small values for the total sensitivity indices ST_i^u identify non important inputs, and allow the analyst to build a metamodel with less inputs than the original model. We have tested the method on two numerical case study. The estimation of all the sensitivity indices has only required one sample of input-output points of size N= 4,096.

Acknowledgment: The authors are indebted to the anonymous referees for their insightful comments.

6. Appendix

Let us consider the linear model $y = x_1 + x_2 + x_3$, where the x_i's are dependent standard normal random variables characterized by the following product moment correlation matrix,

$$\mathbf{C} = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix}.$$

In the following, we derive the possible ANOVA-representations of this model from which the new variance-based sensitivity indices can be computed.

Let us note $\mathbf{W}$ the inverse correlation matrix. In the sequel $i \neq j \neq k$ and $W_{ij} = W_{ji}$ ($\mathbf{W}$ is symmetric). Then, we find,

$$W_{ii} = \frac{\rho_{jk}^2 - 1}{-2\rho_{ij}\rho_{ik}\rho_{jk} + \rho_{ij}^2 + \rho_{ik}^2 + \rho_{jk}^2 - 1}, \tag{14}$$

$$W_{ij} = \frac{\rho_{ij} - \rho_{ik}\rho_{jk}}{-2\rho_{ij}\rho_{ik}\rho_{kj} + \rho_{ij}^2 + \rho_{jk}^2 + \rho_{kj}^2 - 1}. \tag{15}$$

Besides, it is straightforward to prove that,

$$y \sim \mathcal{N}(0, V_y), \tag{16}$$

$$V_y = 3 + 2\left(\rho_{12} + \rho_{13} + \rho_{23}\right), \tag{17}$$

while the conditional expectations are,

$$E\left[y|x_i\right] = (1 + \rho_{ij} + \rho_{ik})x_i, \tag{18}$$

$$E\left[x_i|x_j\right] = \rho_{ij}x_j. \tag{19}$$

So, the first decorrelated random variable writes,

$$x_{i-j} = x_i - \rho_{ij}x_j, \tag{20}$$

$$x_{i-j} \sim \mathcal{N}(0, 1 - \rho_{ij}^2). \tag{21}$$

It is less easy to prove that,

$$x_k|x_i, x_j \sim \mathcal{N}(E_{k|i,j}, V_{k|i,j}), \tag{22}$$

$$E_{k|i,j} = -\frac{W_{ki}x_i + W_{kj}x_j}{W_{kk}}, \tag{23}$$

$$V_{k|i,j} = \frac{1}{W_{kk}}. \tag{24}$$

But, from this latter result, we infer that,

$$E\left[y|x_i, x_j\right] = x_i + x_j + E\left[x_k|x_i, x_j\right], \tag{25}$$

$$\Leftrightarrow E\left[y|x_i, x_j\right] = x_i + x_j - \frac{W_{ki}x_i + W_{kj}x_j}{W_{kk}}, \tag{26}$$

$$\Leftrightarrow E\left[y|x_i, x_j\right] = \left(1 - \frac{W_{ki}}{W_{kk}}\right)x_i + \left(1 - \frac{W_{kj}}{W_{kk}}\right)x_j. \tag{27}$$

Since the model is linear as well as the dependence amongst the original random variables, it is expected that,

$$E\left[y|x_i, x_j\right] = E\left[y|x_i\right] + E\left[y|x_{j-i}\right]. \tag{28}$$

Indeed, if we calculate $E\left[y|x_i, x_j\right] - E\left[y|x_i\right]$ by accounting for equations (14-15,23,20), after some developments, we find that,

$$E\left[y|x_i, x_j\right] - E\left[y|x_i\right] = \left(1 + \frac{\rho_{jk}}{1-\rho_{ij}^2} - \frac{\rho_{ij}\rho_{ik}}{1-\rho_{ij}^2}\right)x_{j-i}, \tag{29}$$

$$\Leftrightarrow E\left[y|x_i, x_j\right] = (1 + \rho_{ij} + \rho_{ik})x_i + \left(1 + \frac{\rho_{jk}}{1-\rho_{ij}^2} - \frac{\rho_{ij}\rho_{ik}}{1-\rho_{ij}^2}\right)x_{j-i}. \tag{30}$$

The last result is obtained thanks to equation (18).

Finally, since $E\left[y|x_1, x_2, x_3\right] - E\left[y|x_i, x_j\right] = x_k - E\left[y|x_i, x_j\right] = x_{k-ij}$ by definition (see equation (10)), we prove that, $y = E\left[y|x_1, x_2, x_3\right] = x_1 + x_2 + x_3 = x_{k-ij} + E\left[y|x_i, x_j\right] = (1 + \rho_{ij} + \rho_{ik})x_i + \left(1 + \frac{\rho_{jk}}{1-\rho_{ij}^2} - \frac{\rho_{ij}\rho_{ik}}{1-\rho_{ij}^2}\right)x_{j-i} + x_{k-ij}$. Only the last equality is an ANOVA-HDMR as defined by Sobol (2001).

7. References

[1] Bedford, T. (1998) Sensitivity indices for (Tree)-dependent variables. In: Proceedings of the second international symposium on sensitivity analysis of model output, Venice Italy, 17–20.

[2] Borgonovo, E. and Tarantola, S. (2008) Moment independent and variance-based sensitivity analysis with correlations: An application to the stability of a chemical reactor. *International Journal of Chemical Kinetics*, **40**:687–698.

[3] Iman, R. I. and Conover, W. J. (1982) A distribution-free approach to inducing rank correlation among input variables. *Communications in Statistics: Simulation and Computation*, **11**:311–334.

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

[5] Hastie, T. J. and Tibshirani, R. J. (1990) *Generalized additive models.* Chapman & Hall, London.

[6] Lebrun, R. and Dutfoy, A. (2009) Do Rosenblatt and Nataf isoprobabilistic transformations really differ? *Probabilistic Engineering Mechanics*, **24**:577–584.

[7] Lewandowski, D., Cooke, R. M. and Tebbens, R. J. D. (2007) Sample-based estimation of correlation ratio with polynomial approximation. *ACM Transactions on Modeling and Computer Simulation*, **5**:1–16.

[8] Li, G., Rabitz, H., Yelvington, P. E., Oluwole, O. O., Bacon, F., Kolb C. E. and Schoendorf, J. (2010) Global Sensitivity Analysis for Sys-

tems with Independent and/or Correlated Inputs. *J. Chemical Physics*, **114**:6022–6032.

[9] Li, G., Rabitz H., Hu, J., Chen, Z. and Ju, Y. (2008) Regularized random-sampling high dimensional model representation (RS-HDMR). *J. Mathematical Chemistry*, **43**:1207–1232.

[10] 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.

[11] Mara,T. A. and Rakoto Joseph, O. (2008) Comparison of some efficient methods to evaluate the main effect of computer model factors. *Journal of Statistical Computation and Simulation*, **78**:167–178.

[12] McKay, M. D. (1996) Variance-based methods for assessing uncertainty importance. Technical Report NUREG-1150,UR-1996-2695, Los Alamos National Laboratory.

[13] Morris, M. D. (1991) Factorial sampling plans for preliminary computational experiments. *Technometrics*, **33**:161–174.

[14] Nataf, A.(1962) Détermination des distributions de probabilité dont les marges sont données. *Comptes Rendus de l'Académie des Sciences*, **225**:42–43.

[15] Nelsen, R. B. (2006) *An introduction to copulas.* Springer series in statistics, second edition, N.Y.

[16] Oakley, J. E. and O'Hagan, A.(2004) Probabilistic sensitivity analysis of complex models: a Bayesian approach. *J. Royal Stat. Soc. B*, **66**:751–769.

[17] Rabitz, H., Alis, O., Shorter, J., and Shim, K. (1999) Efficient input-output model representations. *Computer Physics Communication*, **117**:11–20.

[18] Rosenblatt, M. (1952) Remarks on the multivariate transformation. *Annals of Mathematics and Statistic*, **43**:470–472.

[19] Saltelli, A., Tarantola, S. and Chan, K. (1999) A quantitative model independent method for global sensitivity analysis of model output. *Technometrics*, **41**:39–56.

[20] Saltelli, A. (2002) Making best use of model evaluations to compute sensitivity indices. *Computational Physics Communications*, **145**:280–297.

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

[22] Saltelli, A., Ratto, M., Tarantola, S., Campolongo, F. (2006) Sensitivity analysis practices: Strategies for model-based inference. *Reliability Engineering and System Safety*, **91**:1109–1125.

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

[24] Sobol, I. M. (2001) Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, **55**:271–280.

[25] Storlie, C. B. and Helton, J. C. (2008) Multiple predictor smoothing

methods for sensitivity analysis: Description of techniques *Reliability Engineering and System Safety*, **93**:28–54.

[26] Sudret, B. (2008) Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering and System Safety*, **93**:964–979.

[27] Tarantola, S., Gatelli, D. and Mara, T. A. (2006) Random balance designs for the estimation of first-order sensitivity indices. *Reliability Engineering and System Safety*, **91**:717–727.

[28] Xu, C. and Gertner, G. Z. (2008a) A general first-order global sensitivity analysis method. *Reliability Engineering and System Safety*, **93**:1060–1071.

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

8. Tables

Table 1: Test case 1 analytical first-order sensitivity indices for different correlation structures.

$(\rho_{12}, \rho_{13}, \rho_{23})$	Input	S_i	S_i^u
$(0.5, 0.8, 0)$	x_1	0.94	0.02
	x_2	0.40	0.05
	x_3	0.58	0.03
$(-0.5, 0.2, -0.7)$	x_1	0.49	0.72
	x_2	0.04	0.37
	x_3	0.25	0.48
$(-0.49, -0.49, -0.49)$	x_1	0.00	0.98
	x_2	0.00	0.98
	x_3	0.00	0.98

Table 2: Marginal sensitivity indices of the non-linear polynomial model with a sparse correlation matrix. Although the original polynomial is of the form $f_{12}(x_1, x_2) + f_{34}(x_3, x_4) + f_{56}(x_5, x_6)$ the marginal contribution of x_5 and x_6 are about 35% due to their high correlation coefficient.

x_1	x_2	x_3	x_4	x_5	x_6
0.00	0.00	0.02	0.00	0.35	**0.00**
$(\hat{S}_1)$	$(\hat{S}_{2-1})$	$(\hat{S}_{3-12})$	$(\hat{S}_{4-123})$	$(\hat{S}_{5-1234})$	$(\hat{S}^u_6)$
0.00	*0.00*	0.02	0.00	0.35	0.00
$(\hat{S}^u_1)$	$(\hat{S}_2)$	$(\hat{S}_{3-2})$	$(\hat{S}_{4-23})$	$(\hat{S}_{5-234})$	$(\hat{S}_{6-2345})$
0.00	**0.00**	*0.02*	0.00	0.35	0.00
$(\hat{S}_{1-3456})$	$(\hat{S}^u_2)$	$(\hat{S}_3)$	$(\hat{S}_{4-3})$	$(\hat{S}_{5-34})$	$(\hat{S}_{6-345})$
0.00	0.00	**0.00**	*0.02*	0.35	0.00
$(\hat{S}_{1-456})$	$(\hat{S}_{2-4561})$	$(\hat{S}^u_3)$	$(\hat{S}_4)$	$(\hat{S}_{5-4})$	$(\hat{S}_{6-45})$
0.00	0.00	0.02	**0.00**	*0.35*	0.00
$(\hat{S}_{1-56})$	$(\hat{S}_{2-561})$	$(\hat{S}_{3-5612})$	$(\hat{S}^u_4)$	$(\hat{S}_5)$	$(\hat{S}_{6-5})$
0.00	0.00	0.02	0.00	**0.00**	*0.35*
$(\hat{S}_{1-6})$	$(\hat{S}_{2-61})$	$(\hat{S}_{3-612})$	$(\hat{S}_{4-6123})$	$(\hat{S}^u_5)$	$(\hat{S}_6)$

Table 3: Total sensitivity indices of the non-linear polynomial model with a sparse correlation matrix. All the inputs are important as all the sensitivity indices are greater than 10%. The fact that the sensitivity indices of x_1 (x_2 resp.) remain unchanged indicates that they are not correlated with the other inputs.

x_1	x_2	x_3	x_4	x_5	x_6
0.27	0.27	0.28	0.26	0.45	**0.10**
$(\widehat{ST}_1)$	$(\widehat{ST}_{2-1})$	$(\widehat{ST}_{3-12})$	$(\widehat{ST}_{4-123})$	$(\widehat{ST}_{5-1234})$	$(\widehat{ST}_6^u)$
0.27	*0.27*	0.28	0.26	0.45	0.10
$(\widehat{ST}_1^u)$	$(\widehat{ST}_2)$	$(\widehat{ST}_{3-2})$	$(\widehat{ST}_{4-23})$	$(\widehat{ST}_{5-234})$	$(\widehat{ST}_{6-2345})$
0.27	**0.27**	*0.28*	0.26	0.45	0.10
$(\widehat{ST}_{1-3456})$	$(\widehat{ST}_2^u)$	$(\widehat{ST}_3)$	$(\widehat{ST}_{4-3})$	$(\widehat{ST}_{5-34})$	$(\widehat{ST}_{6-345})$
0.27	0.27	**0.26**	*0.28*	0.45	0.10
$(\widehat{ST}_{1-456})$	$(\widehat{ST}_{2-4561})$	$(\widehat{ST}_3^u)$	$(\widehat{ST}_4)$	$(\widehat{ST}_{5-4})$	$(\widehat{ST}_{6-45})$
0.27	0.27	0.28	**0.26**	*0.45*	0.10
$(\widehat{ST}_{1-56})$	$(\widehat{ST}_{2-561})$	$(\widehat{ST}_{3-5612})$	$(\widehat{ST}_4^u)$	$(\widehat{ST}_5)$	$(\widehat{ST}_{6-5})$
0.27	0.27	0.28	0.26	**0.10**	*0.45*
$(\widehat{ST}_{1-6})$	$(\widehat{ST}_{2-61})$	$(\widehat{ST}_{3-612})$	$(\widehat{ST}_{4-6123})$	$(\widehat{ST}_5^u)$	$(\widehat{ST}_6)$

9. Figures

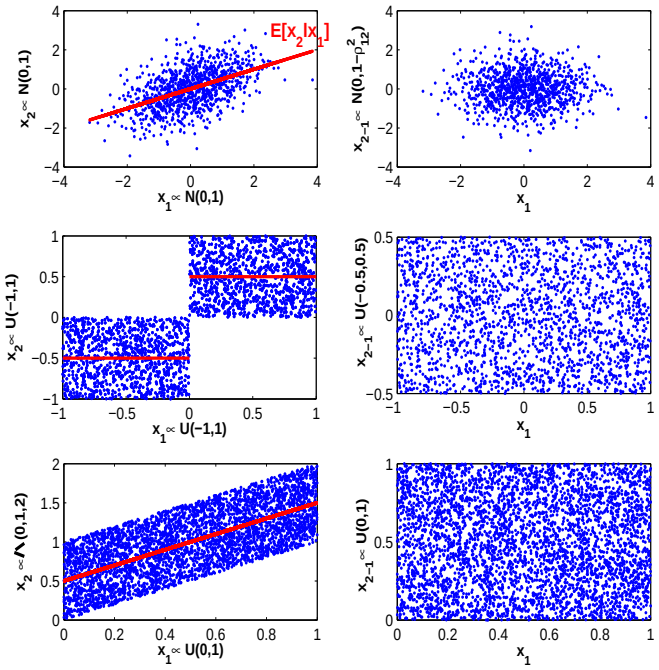


Figure 1: Scatterplots of x_2 versus x_1 for several joint distributions (on the left). Scatterplots of x_{2-1}, the sample of x_2 decorrelated from x_1, versus x_1 (on the right).