

Analysis of covariance (ANCOVA) using polynomial chaos expansions

B. SUDRET

ETH Zürich, Institute of Structural Mechanics, Chair of Risk, Safety & Uncertainty Quantification
Wolfgang-Pauli-Strasse 15, CH-8093 Zürich

Y. CANIOU

Phimeca Engineering, Centre d'Affaires du Zénith, 34 Rue de Sarliève, 63800 Cournon, France

ABSTRACT: Global sensitivity analysis aims at quantifying the uncertainty of the output of a computer model that may be attributed to each input parameter or combination thereof. Variance decomposition techniques that lead to the well-known Sobol' indices are now well established. However this classical framework only holds for *independent* input parameters. Moreover, when the computational model under consideration is costly to evaluate, it is not possible to resort to crude Monte Carlo simulation to evaluate sensitivity indices. In this paper we extend the *polynomial chaos-based* Sobol' indices derived by Sudret (2006, 2008) to the case of dependent input parameters using the covariance decomposition recently proposed by Li *et al.* . The functional decomposition which is natively given by the polynomial chaos expansion is taken advantage of, and the proposed approach is consistent with the classical Sobol' indices when the input variables are independent. The methodology is illustrated on a tolerance analysis problem.

1 INTRODUCTION

Computer simulation has become an inescapable tool of modern engineering. Computational models of ever increasing complexity are built by engineers and scientists in order to describe in the most refined way complex natural or man-made systems. However computer models always represent an idealized vision of the real world. In this context it is important to assess the uncertainty in the model predictions that are caused by the natural variability (aleatory uncertainty) and lack of knowledge (epistemic uncertainty) in the model input parameters.

In the field of uncertainty quantification, sensitivity analysis aims at determining what are the input parameters or combinations thereof that have the largest impact onto the model predictions, *i.e.* those which explain at best the response variability. A review of classical sensitivity analysis can be found in Saltelli *et al.* (2004). In this paper we focus on so-called *variance-based sensitivity indices* originally introduced by Sobol' (1993). This approach is based on the decomposition of the variance of the model response into a sum of contributions that can be related to each single model input parameter, pairs of parameters, triplets, etc. The main drawback of these so-called *Sobol' indices* is that their definition assumes

that the input parameters of the computational model are *statistically independent*.

In engineering problems though, many situations exhibit correlations between model parameters, *e.g.* between loads applied onto a structural system or between material parameters describing a constitutive law. Deriving sensitivity indices that are able to account for the input correlation structure is thus of practical importance. Until recently not much attention has been devoted to this topic. *Distribution-based* sensitivity indices have been introduced by Borgonovo (2007), Borgonovo *et al.* (2011), which are equally defined when the model input parameters are dependent or independent. The interpretation of the resulting δ -indices is not easy though since they do not sum up to 1 as in the case of Sobol' indices. In contrast several approaches to generalize the Sobol' indices to the case of dependent parameters have been recently proposed, see Xu and Gertner (2008), Da Veiga *et al.* (2009), Mara and Tarantola (2012), Kucherenko *et al.* (2012). However no consensus has been attained so far.

In this paper the *covariance decomposition* originally proposed by Li *et al.* (2010) is exploited. In the original paper the authors first build a so-called High-Dimensional Model Representation (HDMR) of the model. In the present paper we propose to make use

of *polynomial chaos expansions* (Ghanem & Spanos 2003) which have already proven efficient for sensitivity analysis *in the case of independent input parameters*, see Sudret (2008), Blatman and Sudret (2010).

As a summary the goal of the paper is to provide a computationally efficient framework for sensitivity analysis in case of dependent parameters which is based on a covariance decomposition of the model output and polynomial chaos expansions. The paper is organized as follows: in Section 2 the basics of polynomial chaos expansions is summarized. In Section 3 the tools of variance-based sensitivity analysis are recalled. In Section 4 the proposed polynomial chaos / covariance-based sensitivity indices are derived. The approach is illustrated by two application examples in Section 5.

2 POLYNOMIAL CHAOS EXPANSIONS

Let us consider a computational model $\mathcal{M} : \mathbf{x} \in \mathcal{D}_{\mathbf{X}} \subset \mathbb{R}^M \mapsto y = \mathcal{M}(\mathbf{x}) \in \mathbb{R}$. The uncertain input parameters are modelled by a random vector \mathbf{X} with $\mathcal{D}_{\mathbf{X}}$ through its probability density function (PDF) denoted by $f_{\mathbf{X}}$. We assume that the model output has a finite variance, *i.e.* $\text{Var}[\mathcal{M}(\mathbf{X})] < +\infty$. In this case $Y = \mathcal{M}(\mathbf{X})$ belongs to the Hilbert space of second order random variables. Thus Y may be represented by its coordinates onto a countably infinite basis (Soize & Ghanem 2004):

$$Y = \mathcal{M}(\mathbf{X}) = \sum_{\alpha \in \mathbb{N}^M} y_{\alpha} \Psi_{\alpha}(\mathbf{X}) \quad (1)$$

where the construction of the basis $\{\Psi_{\alpha}, \alpha \in \mathbb{N}^M\}$ is now described. Assume now that the input random vector \mathbf{X} has independent components, *i.e.* $f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^M f_{X_i}(x_i)$ where f_{X_i} is the marginal distribution of the i -th component X_i . It is possible to define an Hilbertian basis made of multivariate polynomials in the input random vector. More precisely, for any M -tuple $\alpha = \{\alpha_1, \dots, \alpha_M\}$, one defines:

$$\Psi_{\alpha}(\mathbf{X}) = \prod_{i=1}^M P_{\alpha_i}^{(i)}(X_i) \quad (2)$$

where $\{P_n^{(i)}, n \in \mathbb{N}\}$ is the set of orthonormal polynomials with respect to the probability measure $\mathbb{P}(dx_i) = f_{X_i}(x_i) dx_i$. By construction the multivariate basis is orthonormal ($\delta_{\alpha\beta}$ is the Kronecker symbol):

$$\mathbb{E}[\Psi_{\alpha}(\mathbf{X}) \Psi_{\beta}(\mathbf{X})] = \delta_{\alpha\beta} \quad (3)$$

For computational purpose, the series is truncated by selected a truncation set \mathcal{A} , which lead to the *polynomial chaos approximation*:

$$Y_{\mathcal{A}} = \mathcal{M}_{\mathcal{A}}(\mathbf{X}) \stackrel{\text{def}}{=} \sum_{\alpha \in \mathcal{A}} y_{\alpha} \Psi_{\alpha}(\mathbf{X}) \quad (4)$$

For instance, the truncation scheme may correspond to all the multivariate polynomials whose total degree (computed as $\|\alpha\|_1 = \sum_{i=1}^M \alpha_i$) is smaller than or equal to a prescribed p , namely $\mathcal{A}_{M,p} = \{\alpha : \sum_{i=1}^M \alpha_i \leq p\}$. Note that other truncation schemes such as the hyperbolic sets may be more appropriate for large dimensional problems, see Blatman & Sudret (2011).

The set of PC coefficients $\{y_{\alpha}, \alpha \in \mathcal{A}\}$ in Eq.(4) are interpreted as the coordinates of random variable Y in the orthonormal basis made of the Ψ_{α} 's. Many computational methods have been proposed for their evaluation. In practical engineering applications, so-called *non intrusive* methods are most suited, in the sense that they only use a set of runs of the original model \mathcal{M} using selected points in the domain of definition of the input parameters, namely a so-called *experimental design* denoted by $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$. Such non intrusive methods include stochastic collocation approaches based on sparse grids (Xiu 2007), regression (Berveiller, Sudret, & Lemaire 2006) or variable selection techniques such as Least Angle Regression (Blatman & Sudret 2011). In this paper we focus on classical least-square regression for the sake of illustration.

The idea behind least-square regression is to consider the infinite series in Eq.(1) as a sum of the truncated series in Eq.(4) and a residual. The coefficients $\mathbf{y}_{MS} \stackrel{\text{def}}{=} \{y_{\alpha}, \alpha \in \mathcal{A}\}$ are obtained by minimizing the empirical mean-square error between the original model response and its approximation by PC expansion over the selected experimental design:

$$\mathbf{y}_{MS} = \arg \min_{\mathbf{y} \in \mathbb{R}^{\text{card}\mathcal{A}}} \sum_{i=1}^N \left[\mathcal{M}(\mathbf{x}^{(i)}) - \sum_{\alpha \in \mathcal{A}} y_{\alpha} \Psi_{\alpha}(\mathbf{x}^{(i)}) \right]^2 \quad (5)$$

Typical choices of experimental designs are *Latin Hypercube designs* or low-discrepancy sequences, and the size of the experimental design is selected by the thumb rule $N = 2 - 3 \text{card}\mathcal{A}$. Note that robust error estimates based on cross-validation techniques may be derived, see Blatman & Sudret (2010).

As a conclusion, based on an experimental design \mathcal{X} of appropriate size and the model responses $\{\mathcal{M}(\mathbf{x}_i) i = 1, \dots, N\}$, a surrogate model of the original model is obtained, which is nothing but a polynomial response surface with specific orthogonality properties.

3 SENSITIVITY ANALYSIS

3.1 Sobol' indices

Considering a computational model \mathcal{M} with *independent* uncertain input parameters gathered in \mathbf{X} ,

the Sobol' decomposition (also known as functional ANOVA decomposition) reads:

$$\begin{aligned} \mathcal{M}(\mathbf{x}) &= \mathcal{M}_0 + \sum_{i=1}^M \mathcal{M}_i(x_i) \\ &+ \sum_{1 \leq i < j \leq M} \mathcal{M}_{ij}(x_i, x_j) + \dots + \mathcal{M}_{12\dots M}(\mathbf{x}) \end{aligned} \quad (6)$$

In this equation, \mathcal{M}_0 is the mean value of the response, *i.e.* $\mathcal{M}_0 = \mathbb{E}[\mathcal{M}(\mathbf{X})]$ and the other terms are *summands* of increasing order. Introducing the generic index set $u \stackrel{\text{def}}{=} \{i_1, \dots, i_k\} \subset \{1, \dots, M\}$ and denoting by \mathbf{x}_u the subvector of \mathbf{x} obtained by extracting the components labelled by the indices in u , the above equation concisely rewrites:

$$\mathcal{M}(\mathbf{x}) = \mathcal{M}_0 + \sum_{\substack{u \subset \{1, \dots, M\} \\ u \neq \emptyset}} \mathcal{M}_u(\mathbf{x}_u) \quad (7)$$

Imposing that these summands satisfy the following orthogonality property:

$$\mathbb{E}[\mathcal{M}_u(\mathbf{x}_u) \mathcal{M}_v(\mathbf{x}_v)] = 0 \quad \forall u, v \subset \{1, \dots, M\} \quad (8)$$

the above decomposition is *unique*. As a consequence the variance of the model response:

$$\text{Var}[\mathcal{M}(\mathbf{X})] = \mathbb{E}[(\mathcal{M}(\mathbf{X}) - \mathcal{M}_0)^2] \quad (9)$$

can be computed by substituting for Eq.(6) into Eq.(9). It reduces to the sum of the variances of the summands since the cross terms cancel out due to Eq.(8):

$$\text{Var}[\mathcal{M}(\mathbf{X})] = \sum_{u \subset \{1, \dots, M\}} \text{Var}[\mathcal{M}_u(\mathbf{X})] \quad (10)$$

The *Sobol' indices* are then defined by the normalized variances, *i.e.* :

$$S_u = \text{Var}[\mathcal{M}_u(\mathbf{X})] / \text{Var}[\mathcal{M}(\mathbf{X})] \quad (11)$$

By construction they sum up to 1. The first order indices $\{S_i, i = 1, \dots, M\}$ characterize the portion of the output variance that can be attributed to each single parameter X_i . The second order indices $\{S_{ij}, 1 \leq i < j \leq M\}$ correspond to the second-order interaction effects, etc.

3.2 Sobol' indices using polynomial chaos expansions

The traditional approach to evaluating the Sobol' indices defined above is based on Monte Carlo simulation. Following the original estimator developed in Sobol' (1993), several improved estimators have been

proposed, see *e.g.* Janon et al. (2012) for recent developments including asymptotic properties. Whatever the selected estimator, the computational cost in order to get an accurate result is still large due to the sampling procedure.

In contrast *polynomial chaos-based* Sobol' indices provide an exact result for the sensitivity indices once the PC expansion is available. As originally shown in Sudret (2006), Sudret (2008), the truncated expansion in Eq.(4) may be rearranged so as to reflect the decomposition into summands of increasing order. For any non-empty set $u \subset \{1, \dots, M\}$ and any finite truncation set $\mathcal{A} \subset \mathbb{N}^M$ let us define:

$$\mathcal{A}_u = \{\alpha \in \mathcal{A} : k \in u \Leftrightarrow \alpha_k \neq 0, k = 1, \dots, M\} \quad (12)$$

In other words \mathcal{A}_u contains all the multi-indices within the truncation set \mathcal{A} which have non zero components $\alpha_k \neq 0$ if and only if $k \in u$. The sum of the associated PC expansion terms form a function which depends *only* on the input variables \mathbf{x}_u . The reordering of Eq.(4) reads:

$$\mathcal{M}_{\mathcal{A}}(\mathbf{x}) = y_0 + \sum_{\substack{u \subset \{1, \dots, M\} \\ u \neq \emptyset}} \sum_{\alpha \in \mathcal{A}_u} y_{\alpha} \Psi_{\alpha}(\mathbf{x}) \quad (13)$$

As the Sobol' decomposition is unique, the comparison of Eqs.(7) and (13) readily provides the expression of each summand for the PC expansion:

$$\mathcal{M}_u(\mathbf{x}_u) = \sum_{\alpha \in \mathcal{A}_u} y_{\alpha} \Psi_{\alpha}(\mathbf{X}) \quad (14)$$

Due to the orthonormality of the polynomial chaos basis (see Eq.(3)) the variance of the truncated PC expansion (resp. of a given summand) reads:

$$\text{Var}[Y_{\mathcal{A}}] = \sum_{\alpha \in \mathcal{A}, \alpha \neq 0} y_{\alpha}^2 \quad (15)$$

$$\text{Var}[\mathcal{M}_u(\mathbf{X}_u)] = \sum_{\alpha \in \mathcal{A}_u} y_{\alpha}^2$$

and the associated Sobol' indices in Eq.(11) is the mere ratio of these two quantities.

As a conclusion, once a PC expansion is available, the computation of the Sobol' indices reduces to elementary operations on the coefficients of the expansion.

4 COVARIANCE-BASED SENSITIVITY INDICES USING PC EXPANSIONS

4.1 Introduction

As shown in the previous section, a truncated polynomial chaos expansion not only provides a surrogate

model \mathcal{M}_A of the original computational model \mathcal{M} , it also yields a functional decomposition in which the first order effects as well as the second and higher order interaction terms readily appear. When the input variables in \mathbf{X} are independent the polynomial chaos basis is orthonormal by construction so that the various variances that must be computed to evaluate the Sobol' indices are easily obtained.

Consider now the general case where the input random vector has dependent components. Introducing the marginal CDF and PDF of each component X_i , respectively denoted by F_{X_i} and f_{X_i} , the joint PDF of \mathbf{X} may be represented using the copula theory (Nelsen 1999) as follows:

$$f_{\mathbf{X}}(\mathbf{x}) = c(F_{X_1}(x_1), \dots, F_{X_M}(x_M)) \cdot \prod_{i=1}^M f_{X_i}(x_i) \quad (16)$$

where $c(\cdot)$ is the copula density function (which reduces to $c = 1$ in the independent case).

Due to the statistical dependence between the X_i 's it is not possible anymore to derive a unique Sobol'-like decomposition in terms of *orthogonal* summands of increasing order, as Eq.(7) in the independent case, see *e.g.* Chastaings, Gamboa, & Prieur (2012).

However, provided the computational model \mathcal{M} may be given an approximate functional decomposition, the variance of the output may be cast as a *covariance decomposition* as originally shown in Li et al. (2010). The approach proposed in this paper thus consists in:

- building a functional approximation of \mathcal{M} in terms of a polynomial chaos decomposition *as if* the input vector \mathbf{X} had independent components;
- using this functional approximation in order to evaluate the variance of the output under the real input distribution that includes some dependence between the parameters.

4.2 Covariance decomposition

Assume that the PC expansion of the computational model Eq.(4) has been constructed under the hypothesis of independent input random variables. Consider now the true input random vector \mathbf{X} . The variance of the model response reads:

$$\begin{aligned} \text{Var}[\mathcal{M}(\mathbf{X})] &= \text{Cov} \left[\mathcal{M}_0 + \sum_{u \in \{1, \dots, M\}} \mathcal{M}_u(\mathbf{X}_u), \mathcal{M}(\mathbf{X}) \right] \quad (17) \\ &= \sum_{u \in \{1, \dots, M\}} \text{Cov}[\mathcal{M}_u(\mathbf{X}_u), \mathcal{M}(\mathbf{X})] \end{aligned}$$

where each term can in turn be decomposed as follow:

$$\begin{aligned} \text{Cov}[\mathcal{M}_u(\mathbf{X}_u), \mathcal{M}(\mathbf{X})] &= \text{Var}[\mathcal{M}_u(\mathbf{X}_u)] \\ &+ \sum_{v \neq u} \text{Cov}[\mathcal{M}_u(\mathbf{X}_u); \mathcal{M}_v(\mathbf{X}_v)] \end{aligned} \quad (18)$$

By renormalizing the above equation we introduce the *covariance-based total sensitivity index*:

$$S_u^{(cov)} = \frac{\text{Cov}[\mathcal{M}_u(\mathbf{X}_u), \mathcal{M}(\mathbf{X})]}{\text{Var}[\mathcal{M}(\mathbf{X})]} \quad (19)$$

This covariance index is the sum of a *structural* (or *uncorrelated*) sensitivity index $S_u^{(U)}$ and a *correlative* sensitivity index $S_u^{(C)}$ which are defined by:

$$S_u^{(U)} = \frac{\text{Var}[\mathcal{M}_u(\mathbf{X}_u)]}{\text{Var}[\mathcal{M}(\mathbf{X})]} \quad (20)$$

$$S_u^{(C)} = S_u^{(cov)} - S_u^{(S)}$$

In the case of independent input variables, if the functional approximation is a truncated polynomial chaos expansion then the correlative indices vanish due to Eq.(8) and the structural indices reduce to the classical Sobol' indices. The above definition is thus consistent with the classical approach.

4.3 Estimation of the covariance-based indices

Once the functional approximation is available from the PC expansion (see Eqs.(13)-(14)), Monte Carlo estimates of variances and covariances in Eqs.(19)-(20) are easily obtained. Denoting by $\mathcal{X}_{MCS} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ a sample set drawn from the original distribution of the input vector Eq.(16), the empirical mean value and variance of the response reads:

$$\overline{y_A} = \frac{1}{n} \sum_{i=1}^n \mathcal{M}_A(\mathbf{x}_i) \quad (21)$$

$$\widehat{\text{Var}}[Y_A] = \frac{1}{n-1} \sum_{i=1}^n (\mathcal{M}_A(\mathbf{x}_i) - \overline{y_A})^2$$

Note that $\overline{y_A} \neq y_0$ since the expectations of each basis polynomial $\mathbb{E}[\Psi_\alpha]$ under the actual distribution $f_{\mathbf{X}}$ are non zero. For any subset $u \in \{1, \dots, M\}$, the subvector of a sample $\mathbf{x}_i \in \mathcal{X}_{MCS}$ made of the components that are labelled by u is simply denoted by $\mathbf{x}_{u,i}$. As in Eq.(21) the expected value $\mathbb{E}[\mathcal{M}_u(\mathbf{x}_u)]$ is estimated by:

$$\overline{y_u} \stackrel{\text{def}}{=} \widehat{\mathbb{E}}[\mathcal{M}_u(\mathbf{x}_u)] = \frac{1}{n} \sum_{i=1}^n \mathcal{M}_u(\mathbf{x}_{u,i}) \quad (22)$$

The estimate of the covariance-based sensitivity index then reads:

$$\widehat{S}_u^{(cov)} = \frac{\sum_{i=1}^n (\mathcal{M}_A(\mathbf{x}_i) - \overline{y_A}) (\mathcal{M}_u(\mathbf{x}_{u,i}) - \overline{y_u})}{\sum_{i=1}^n (\mathcal{M}_A(\mathbf{x}_i) - \overline{y_A})^2} \quad (23)$$

Similarly the uncorrelated contribution is estimated by:

$$\widehat{S}_u^{(U)} = \frac{\sum_{i=1}^n (\mathcal{M}_u(\mathbf{x}_{u,i}) - \overline{y_u})^2}{\sum_{i=1}^n (\mathcal{M}_A(\mathbf{x}_i) - \overline{y_A})^2} \quad (24)$$

5 APPLICATION EXAMPLE: TOLERANCE ANALYSIS

5.1 Introduction

In industrial mass production, assembled products are made of individual parts that are prone to uncertainty in their dimensions due to the manufacturing processes. Non-assembly problems may occur when the real dimensions of some parts differ (even slightly) from their nominal values, even if they remain in the prescribed tolerance ranges. The customer quality requirements are usually defined in terms of number of out-of-tolerance assemblies (*e.g.* expressed in *ppm*, *i.e.* parts per million). Predicting this probability of non-assembly prior to the mass production is of crucial importance since it may avoid a large wastage during production. Probabilistic methods derived from structural reliability analysis have been recently proposed by Gayton, Beaucaire, Bourinet, Duc, Lemaire, & Gauvrit (2011) in order to address such problems.

Within the tolerance analysis it is also important to detect the *key dimensions*, *i.e.* those dimensions whose uncertainty may affect the most the non-assembly. This question can be solved using global sensitivity analysis. The functional requirement which is typically a function of the various dimensions of the parts in the assembly is considered as the output $Y = \mathcal{M}(\mathbf{X})$ where \mathbf{X} is a random vector whose components model the tolerance range of the parts respective dimensions. The key dimensions correspond to those dimensions which exhibit the largest sensitivity indices.

When complex parts are manufactured it is common that several surfaces are machined during the same operation. The corresponding dimensions (*e.g.* hole or axle diameters) are then strongly correlated: if one dimension is smaller than its nominal value (*e.g.* due to the wear of the tool) the others will most probably also be. Thus the covariance-based sensitivity indices are relevant in this situation. In the sequel we present an application example taken from Caniou (2012).

5.2 Problem statement

The assembly under consideration corresponds to an electrical connector for which a contact clearance is of interest. More specifically the axial deviation of the connector pin is considered which shall not be greater than a given threshold for the sake of correct assembly (see Figure 1).

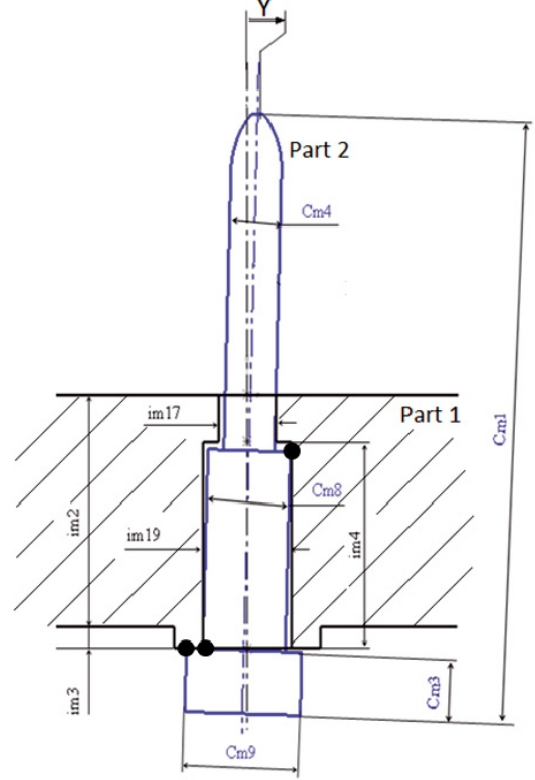


Figure 1: Electric connector – after Gayton *et al.* (2011)

This deviation is computed as a non linear function of the dimensions of the various components and parameters describing their respective positions (see Gayton *et al.* (2011) for details):

$$c = \frac{cm_4 + cm_8 + cm_7}{2} \quad (25)$$

$$i = \frac{ima_{17} + ima_{19} + ima_{20}}{2} \quad (26)$$

$$h = ima_2 + ima_3 \quad (27)$$

$$\alpha = \arccos\left(\frac{c}{\sqrt{i^2 + h^2}}\right) - \arccos\left(\frac{i}{\sqrt{i^2 + h^2}}\right) \quad (28)$$

$$r_2 = \frac{\frac{ima_{19} + ima_{20}}{2} - \frac{cm_8 + cm_7}{2}}{\tan(\alpha)} \quad (29)$$

$$z = \frac{r_2}{\cos(\alpha)} + \left(\frac{cm_9 + cm_{10}}{2} + \frac{cm_7}{4}\right) \tan(\alpha) \quad (30)$$

$$J_1 = (cm_1 - cm_3 - z) \sin(\alpha) \quad (31)$$

$$J_2 = \frac{cm_7}{4} \cos(\alpha) \quad (32)$$

$$J_3 = \frac{cm_5 + cm_6}{2} \cos(\alpha) \quad (33)$$

$$Y = J_1 + J_2 + J_3 \quad (34)$$

The 14 parameters appearing in the above equations are considered as Gaussian random variables, whose parameters are reported in Table 1.

Parameter	Mean μ	σ
cm_1	10.530	0.200/6
cm_3	0.750	0.040/6
cm_4	0.643	0.015/6
cm_5	0.100	0.200/6
cm_6	0.000	0.060/6
cm_7	0.000	0.200/6
cm_8	0.720	0.040/6
cm_9	1.325	0.050/6
cm_{10}	0.000	0.040/6
ima_2	3.020	0.060/6
ima_3	0.400	0.060/6
ima_{17}	0.720	0.040/6
ima_{19}	0.970	0.040/6
ima_{20}	0.000	0.040/6

Table 1: Electric connector – probabilistic modelling of the dimensions and their tolerances (Gaussian distributions, all dimensions in mm).

Because different surfaces are machined during the same manufacturing operation, their dimension are *highly correlated*. The different Spearman correlation coefficients ρ_S are estimated as follows:

$$\rho_S(cm_1, cm_3) = 0.8$$

$$\begin{aligned} \rho_S(cm_4, cm_8) &= \rho_S(cm_4, cm_9) \\ &= \rho_S(cm_8, cm_9) = 0.8 \end{aligned} \quad (35)$$

$$\begin{aligned} \rho_S(ima_{17}, ima_{19}) &= \rho_S(ima_{17}, ima_{20}) \\ &= \rho_S(ima_{19}, ima_{20}) = 0.8 \end{aligned}$$

The joint distribution of this parameters is built as in Eq.(16) from the marginal Gaussian distributions and a *Gaussian copula* (see Nelsen (1999) for details) which is parameterized by a correlation matrix \mathbf{R} derived from Eq.(35):

$$R_{ij} = 2 \sin\left(\frac{\pi}{6} \rho_{S,ij}\right) \quad (36)$$

A covariance-based sensitivity analysis is carried out using polynomial chaos expansion of degree $p =$

2. Due to the dimension of the problem, *i.e.* $M = 14$, the number of coefficients to be computed by mean square regression is high, *i.e.* $P = 120$. The experimental design a Latin Hypercube sample of size $N = 2P = 240$. The various covariance terms are computed using $n = 10^4$ in the Monte Carlo simulations.

Parameter	$S_i^{(cov)}$	S_i^U	S_i^C
cm_1	0.00	0.00	0.00
cm_3	0.00	0.00	0.00
cm_4	0.02	0.01	0.01
cm_5	0.15	0.15	0.00
cm_6	0.02	0.02	0.00
cm_7	0.52	0.51	0.01
cm_8	0.03	0.02	0.01
cm_9	0.00	0.00	0.00
cm_{10}	0.00	0.00	0.00
ima_2	0.00	0.00	0.00
ima_3	0.00	0.00	0.00
ima_{17}	0.11	0.05	0.06
ima_{19}	0.07	0.02	0.05
ima_{20}	0.09	0.03	0.06
Σ	1.00	0.81	0.19

Table 2: Results of the sensitivity analysis for the electrical connectors.

The first order sensitivity indices $S_i, i = 1, \dots, M$ attached to each input variable are reported in Table 2. It appears that the main contributor to the variability of Y are the dimension cm_7 and cm_5 . These parameters are not correlated to others such that $S_{cm_k}^{(cov)} \approx S_{cm_k}^U, k = 5, 7$.

In contrast the three dimensions $ima_k, k = 17, 19, 20$, which contribute to 27% of the output variance are correlated. It is clear from the table that their correlative sensitivity indices S_i^C are greater than the uncorrelated counterpart, which can be explained by the large correlation coefficients between these parameters in Eq.(35).

6 CONCLUSIONS

Taking into account the statistical dependence between model input parameters in global sensitivity analysis is an important problem that has not been given a definitive solution so far. In the context of complex models such as those appearing in mechanical engineering the limited number of runs of the computational model that is affordable precludes the use of simulation-based estimators.

In this paper we use the so-called *covariance decomposition* in order to define sensitivity indices that have both a structural (also called uncorrelated) and a

correlative contribution. These indices are computed by first carrying out a polynomial chaos expansion of the computational model under the assumption of independent input parameters. The expansion is then split into functional summands of increasing order, which in turn are used to compute the required covariances by Monte Carlo simulation.

An application example that is taken from tolerance analysis in mechanical engineering is used in order to illustrate the approach and shows that some important parameters (*i.e.* notably contributing to the output variance) have significant correlative contributions.

REFERENCES

- Berveiller, M., B. Sudret, & M. Lemaire (2006). Stochastic finite elements: a non intrusive approach by regression. *Eur. J. Comput. Mech.* 15(1-3), 81–92.
- Blatman, G. & B. Sudret (2010). Efficient computation of global sensitivity indices using sparse polynomial chaos expansions. *Reliab. Eng. Sys. Safety* 95, 1216–1229.
- Blatman, G. & B. Sudret (2011). Adaptive sparse polynomial chaos expansion based on Least Angle Regression. *J. Comput. Phys* 230, 2345–2367.
- Borgonovo, E. (2007). A new uncertainty importance measure. *Reliab. Eng. Sys. Safety* 92, 771–784.
- Borgonovo, E., W. Castaings, & S. Tarantola (2011). Moment-independent importance measures: new results and analytical test cases. *Risk Analysis* 31(3), 404–428.
- Caniou, Y. (2012). *Global sensitivity analysis for nested and multiscale models*. Ph. D. thesis, Université Blaise Pascal, Clermont-Ferrand.
- Chastaings, G., F. Gamboa, & C. Prieur (2012). Generalized Hoeffding-Sobol decomposition for dependent variables – application to sensitivity analysis. *Electronic Journal of Statistics* 6, 2420–2448.
- Da Veiga, S., F. Wahl, & F. Gamboa (2009). Local polynomial estimation for sensitivity analysis on models with correlated inputs. *Technometrics* 51(4), 452–463.
- Gayton, N., P. Beaucaire, J.-M. Bourinet, E. Duc, M. Lemaire, & L. Gouvrit (2011). APTA: advanced probability-based tolerance analysis of products. *Mécanique & Industries* 12(2), 71–85.
- Ghanem, R. & P. Spanos (2003). *Stochastic Finite Elements : A Spectral Approach*. Courier Dover Publications.
- Janon, A., T. Klein, A. Lagnoux-Renaudie, M. Nodet, & C. Prieur (2012). Asymptotic normality and efficiency of two Sobol’ index estimators. Technical report, INRIA.
- Kucherenko, S., A. Tarantola, & P. Annoni (2012). Estimation of global sensitivity indices for models with dependent variables. *Comput. Phys. Comm.* 183, 937–946.
- Li, G., H. Rabitz, P. Yelvington, O. Oluwole, F. Bacon, C. Kolb, & J. Schoendorf (2010). Global sensitivity analysis for systems with independent and/or correlated inputs. *J. Phys.Chem.* 114, 6022–6032.
- Mara, T. & S. Tarantola (2012). Variance-based sensitivity indices for models with dependent inputs. *Reliab. Eng. Sys. Safety* 107, 125–131.
- Nelsen, R. (1999). *An introduction to copulas*, Volume 139 of *Lecture Notes in Statistics*. Springer.
- Saltelli, A., S. Tarentola, F. Campolongo, & M. Ratto (2004). *Sensitivity analysis in practice – A guide to assessing scientific models*. J. Wiley & Sons.
- Sobol’, I. (1993). Sensitivity estimates for nonlinear mathematical models. *Math. Modeling & Comp. Exp.* 1, 407–414.
- Soize, C. & R. Ghanem (2004). Physical systems with random uncertainties: chaos representations with arbitrary probability measure. *SIAM J. Sci. Comput.* 26(2), 395–410.
- Sudret, B. (2006). Global sensitivity analysis using polynomial chaos expansions. In P. Spanos and G. Deodatis (Eds.), *Proc. 5th Int. Conf. on Comp. Stoch. Mech (CSM5)*, Rhodos, Greece.
- Sudret, B. (2008). Global sensitivity analysis using polynomial chaos expansions. *Reliab. Eng. Sys. Safety* 93, 964–979.
- Xiu, D. (2007). Efficient collocation approach for parametric uncertainty analysis. *Comm. Comput. Phys.* 2(2), 293–309.
- Xu, C. & G. Gertner (2008). Uncertainty and sensitivity analysis for models with correlated parameters. *Reliab. Eng. Sys. Safety* 93, 1563–1573.