Efficient Sparse Polynomial Chaos Expansion Methodology for Computationally-expensive deterministic models

Tamara Al-Bittar [1)], and Abdul-Hamid Soubra[2)]

1) *PhD student, University of Nantes, Bd. de l'université, BP 152, 44603 Saint-Nazaire cedex, France; E-mail: Tamara.Al-Bittar@univ-nantes.fr*
2) *Professor, University of Nantes, Bd. de l'université, BP 152, 44603 Saint-Nazaire cedex, France; E-mail: Abed.Soubra@univ-nantes.fr*

Abstract

The sparse polynomial chaos expansion (SPCE) methodology is an efficient approach that deals with uncertainties propagation in case of high-dimensional problems (i.e. when a large number of random variables is involved). This methodology significantly reduces the computational cost with respect to the classical full polynomial chaos expansion (PCE) methodology. Notice however that when dealing with computationally-expensive deterministic models, the time cost remains important even with the use of the SPCE. In this paper, an efficient combined use of the SPCE methodology and the global sensitivity analysis (GSA) is proposed to solve such a problem. The proposed methodology is validated using a relatively non-expensive deterministic model.

Keywords: Sobol indices; spatial variability; sparse polynomial chaos expansion, global sensitivity analysis.

1. Introduction

An efficient approach to deal with uncertainties propagation in case of high-dimensional problems (i.e. when a large number of random variables is involved) was recently presented by Blatman and Sudret (2010). This approach is based on a Sparse Polynomial Chaos Expansion (SPCE) for the system response and leads to a reduced computational cost as compared to the classical Polynomial Chaos Expansion (PCE) methodology. Notice that both, the PCE and the SPCE methodologies, aim at replacing the original expensive deterministic model which may be an analytical model or a finite element/finite difference model by a meta-model. This allows one to calculate the system response using a simple analytical equation (e.g. Isukapalli et al., 1998; Huang et al., 2009; Mollon et al., 2011; Mao et al., 2012). Notice however that when dealing with computationally-expensive deterministic models with a large number of random variables, the time cost remains important even with the use of the SPCE. Consequently, a method that can reduce once again the cost of the probabilistic analysis is needed. In this paper, an efficient combination between the SPCE methodology and the Global Sensitivity Analysis (GSA) is proposed to solve such a problem. In this method, a small SPCE order is firstly selected to approximate the system response by a meta-model. A GSA based on Sobol indices is then performed on this small SPCE order to determine the weight of each random variable in the variability of the system response. As a result, the variables with very small values of their Sobol

indices (i.e. those that have a small weight in the variability of the system response) can be discarded. Consequently, a response which only depends on a smaller number of random variables is obtained. In other words, one obtains a response with an 'effective dimension'. This dimension is smaller than the initial dimension where the total number of random variables was considered. As it will be shown later, the use of a small SPCE to perform the GSA is not a concern since higher SPCE orders lead to the same influential random variables. Once the 'effective dimension' was determined, a higher SPCE order that makes use of only the most influential random variables can be used. This significantly reduces the computation time. The use of a higher SPCE order is necessary in order to lead to an improved fit of the SPCE.

The proposed methodology is validated using a relatively non-expensive model which was extensively investigated by Al-Bittar and Soubra (2011, 2012). This model involves the computation of the ultimate bearing capacity of a strip footing resting on a weightless spatially varying (c, φ) soil where c is the soil cohesion and φ is the soil angle of internal friction. It should be noticed here that the random fields of c and φ are discretized into a finite number of random variables. This number is small for very large autocorrelation distances and significantly increases for small values of the autocorrelation distances.

The paper is organized as follows: The next two sections aim at briefly presenting both the sparse polynomial chaos expansion (SPCE) and the global sensitivity analysis (GSA). Then, the proposed efficient combination between the SPCE methodology and the GSA is presented. It is followed by the numerical results. The paper ends with a conclusion.

2. Sparse polynomial chaos expansion (SPCE) methodology

In this section, one first presents the polynomial chaos expansion (PCE) and then its extension, the sparse polynomial chaos expansion (SPCE). The Polynomial Chaos Expansion (PCE) methodology allows one to replace an expensive deterministic model which may be an analytical model or a finite element/finite difference numerical model by a meta-model. Thus, the system response may be calculated using a simple analytical equation. This equation is obtained by expanding the system response on a suitable basis which is a series of multivariate polynomials that are orthogonal with respect to the joint probability density function of the random variables.

The PCE theory was originally formulated with standard Gaussian random variables and Hermite polynomials (Ghanem and Spanos, 1989). It was later extended to other types of random variables that use other types of polynomials (Xiu and karniadakis, 2002). In this paper, standard normal random variables in conjunction with Hermite polynomials are used. The coefficients of the PCE may be efficiently computed using a non-intrusive technique where the deterministic calculations are done using for example an analytical model or a finite element/finite difference software treated as a black box. The most used non-intrusive method is the regression approach (e.g. Isukapalli et al., 1998; Huang et al., 2009; Blatman and Sudret, 2010; Mollon et al., 2011; Mao et al., 2012). This method is used in the present work. The PCE methodology can be briefly described as follows:

For a deterministic model Γ with M random variables, the system response can be expressed by a PCE of order p fixed by the user as follows:

$$Y = \Gamma_{PCE}(\xi) = \sum_{\beta=0}^{\infty} a_\beta \Psi_\beta(\xi) \cong \sum_{\beta=0}^{P-1} a_\beta \Psi_\beta(\xi) \tag{1}$$

where P is the number of terms retained in the truncation scheme, $\xi = \{\xi_i\}_{i=1,\ldots,M}$ is a vector of M independent standard random variables that represent the M random variables, a_β are unknown coefficients to be computed and Ψ_β are multivariate Hermite polynomials which are orthogonal with respect to the joint probability density function (PDF) of the standard normal random vector ξ. These multivariate Hermite polynomials can be obtained from the product of one-dimensional Hermite polynomials as follows:

$$\Psi_\beta = \prod_{i=1}^{M} H_{\alpha_i}(\xi_i) \tag{2}$$

where α_i (i=1, …, M) are a sequence of M non-negative integers and $H_{\alpha_i}(.)$ is the α_i^{th} one-dimensional Hermite polynomial. The expressions of the one-dimensional Hermite polynomials are given in Ghanem and spanos (1989) among others.

In practice, the PCE with an infinite number of terms should be truncated by retaining only the multivariate polynomials Ψ_β of degree less than or equal to p. For this purpose, the classical truncation scheme based on the determination of the first order norm $\|\alpha\|_1 = \sum_{i=1}^{M} \alpha_i$ is used. This first order norm should be less than or equal to the order p of the PCE. This leads to a number P of the unknown PCE coefficients equal to $\frac{(M+p)!}{M!\,p!}$. This number is significant in the present case of random fields (especially when considering small values of the autocorrelation distances) and thus, one needs a great number of calls of the deterministic model (see Al-Bittar and Soubra 2011, 2012). The SPCE methodology presented by Blatman and Sudret (2010) is an efficient alternative that can significantly reduce the number of calls of the deterministic model. In this methodology, Blatman and Sudret (2010) have shown that the number of significant terms in a PCE is relatively small since the multivariate polynomials Ψ_β corresponding to high-order interaction (i.e. those resulting from the multiplication of the H_{α_i} with increasing α_i values) are associated with very small values for the coefficients a_β. Thus, a truncation strategy (called the hyperbolic truncation scheme) based on this observation was suggested by these authors. Within this strategy, the multivariate polynomials Ψ_β corresponding to high-order interaction were penalized. This was performed by considering the hyperbolic truncation scheme which suggests that the q-norm should be less than or equal to the order p of the PCE. The q-norm is given by:

$$\|\alpha\|_q = \left(\sum_{i=1}^{M} \alpha_i^q \right)^{1/q} \tag{3}$$

where q is a coefficient (0<q<1). In this formula, q can be chosen arbitrarily. Blatman and Sudret (2010) have shown that sufficient accuracy is obtained for $q \geq 0.5$.

The proposed SPCE methodology leads to a sparse polynomial chaos expansion that contains a small number of unknown coefficients which can be calculated from a reduced number of calls of the deterministic model. This strategy was used in Al-Bittar and Soubra (2011, 2012) and will also be used in this paper to build up a SPCE of the system response. The iterative procedure suggested by Blatman

and Sudret (2010) for building up a SPCE is detailed in Blatman and Sudret (2010) and Al-Bittar and Soubra (2011, 2012) and is not repeated herein. Once the coefficients a_β have been computed, the statistical moments (mean, standard deviation, skewness, and kurtosis) can be calculated with no additional cost. The next subsection is devoted to the method used for the computation of the coefficients a_β of the SPCE using the regression approach.

2.1. COMPUTATION OF THE SPCE COEFFICIENTS BY THE REGRESSION APPROACH

Consider a set of K realizations $\{\xi^{(1)} = (\xi_1,...,\xi_M),...,\xi^{(K)} = (\xi_1,...,\xi_M)\}$ of the standard normal random vector ξ. These realizations are called experimental design (ED) and can be obtained from Monte Carlo (MC) simulations or any other sampling scheme (e.g. Latin Hypercube (LH) sampling or Sobol set). We note $\Gamma = \left\{\Gamma\left(\xi^{(1)}\right),...,\Gamma\left(\xi^{(K)}\right)\right\}$, the corresponding values of the response determined by deterministic calculations.

The computation of the SPCE coefficients using the regression approach is performed using the following equation:

$$\bar{a} = (\eta^T \eta)^{-1} \eta^T \Gamma \tag{4}$$

where the data matrix η is defined by:

$$\eta_{i\beta} = \Psi_\beta(\xi^{(i)}), \qquad i = 1,...,K, \qquad \beta = 0,...,P-1 \tag{5}$$

In order to ensure the numerical stability of the treated problem in Eq.(4), the size K of the ED must be selected in such a way that the matrix $(\eta^T \eta)^{-1}$ is well-conditioned. This implies that the rank of this matrix should be larger than or equal to the number of unknown coefficients. This test was systematically performed while solving the system of equations of the regression approach.

The quality of the output approximation *via* a SPCE closely depends on the SPCE order p. To ensure a good fit between the meta-model and the true deterministic model (i.e. to obtain the optimal SPCE order), the simplest error estimate is the well-known coefficient of determination R^2 given by:

$$R^2 = 1 - \frac{\frac{1}{K}\sum_{i=1}^{K}\left[\Gamma\left(\xi^{(i)}\right) - \Gamma_{SPCE}\left(\xi^{(i)}\right)\right]^2}{\frac{1}{K-1}\sum_{i=1}^{K}\left[\Gamma\left(\xi^{(i)}\right) - \overline{\Gamma}\right]^2} \tag{6}$$

Where

$$\overline{\Gamma} = \frac{1}{K}\sum_{i=1}^{K}\Gamma\left(\xi^{(i)}\right) \tag{7}$$

The value $R^2 = 1$ indicates a perfect fit of the true model response Γ, whereas $R^2 = 0$ indicates a nonlinear relationship between the true model Γ and the SPCE model Γ_{SPCE}. The coefficient R^2 may be a biased estimate since it does not take into account the robustness of the meta-model (i.e. its capability of correctly predicting the model response at any point which does not belong to the experimental design). As a consequence, one makes use of a more reliable and rigorous error estimate, namely the leave-one-out error estimate (Blatman and Sudret, 2010). This error estimate consists in

sequentially removing a point from the experiment design composed of K points. Let $\Gamma_{\xi\backslash i}$ be the meta-model that has been built from the experiment design after removing the i[th] observation and let $\Delta^i = \Gamma(\xi^{(i)}) - \Gamma_{\xi\backslash i}(\xi^{(i)})$ be the predicted residual between the model evaluation at point $\xi^{(i)}$ and its prediction based on $\Gamma_{\xi\backslash i}$. The corresponding coefficient of determination is often denoted by Q^2:

$$Q^2 = 1 - \frac{\frac{1}{K}\sum_{i=1}^{K}\left(\Delta^i\right)^2}{\frac{1}{K-1}\sum_{i=1}^{K}\left[\Gamma\left(\xi^{(i)}\right) - \overline{\Gamma}\right]^2} \tag{8}$$

This coefficient will be used in the present paper to check the accuracy of the fit.

3. Global sensitivity analysis (GSA)

Once the SPCE coefficients are determined, a global sensitivity analysis (GSA) based on Sobol indices can be easily performed. Notice that the first order Sobol index of a given random variable ξ_i (i=1,…, M) gives the contribution of this variable in the variability of the system response. The first order Sobol index is given by Saltelli (2000) and Sobol (2001) as follows:

$$S(\xi_i) = \frac{Var\left[E\left(Y \mid \xi_i\right)\right]}{Var\left(Y\right)} \tag{9}$$

where Y is the system response, $E\left(Y \mid \xi_i\right)$ is the expectation of Y conditional on a fixed value of ξ_i, and Var denotes the variance.

In the present paper, the system response Y is represented by a SPCE. Thus, by replacing Y in Eq.(9) with the SPCE expression, one obtains the Sobol index formula as a function of the different terms of the SPCE (Sudret, 2008). This formula is given by:

$$S(\xi_i) = \frac{\sum_{\beta \in I_i}\left(a_\beta\right)^2 E\left[\left(\Psi_\beta\right)^2\right]}{D_{PC}} \tag{10}$$

where a_β are the obtained SPCE coefficients, Ψ_β are the multivariate Hermite polynomials, $E[.]$ is the expectation operator, and D_{PC} is the variance of the response approximated by the SPCE. The response variance D_{PC} is given by Sudret (2008) as follows:

$$D_{PC} = \sum_{\beta \in [I_1, I_2, \ldots, I_M]}\left(a_\beta\right)^2 E\left[\left(\Psi_\beta\right)^2\right] \tag{11}$$

Notice that the term $E\left[\left(\Psi_\beta\right)^2\right]$ that appears in both Eq. (10) and Eq. (11) is given by Sudret (2008) as follows:

$$E\left(\Psi_\beta^2\right) = \prod_{i=1}^{M}\alpha_i\,! \tag{12}$$

where the α_i are the same sequence of M non-negative integers $\{\alpha_1, \ldots, \alpha_M\}$ used in Eq. (2). Notice finally that I_i in Eq. (10) denotes the set of indices β for which the corresponding Ψ_β is only a function

of the random variable ξ_i (i.e. it only contains the variable ξ_i), and I_i (i=1, …, M) regroup all the indices β for which the corresponding Ψ_β is only a function of the random variable ξ_i (i=1, …, M).

In order to illustrate the PCE theory and the global sensitivity analysis based on Sobol indices in a simple manner, an illustrative example of a PCE of order p=3 using only M=2 random variables (ξ_1 and ξ_2) is presented in Appendix A.

4. Efficient combination between the SPCE methodology and the global sensitivity analysis

As mentioned previously, the time cost of the probabilistic analysis remains important even with the use of the SPCE when dealing with computationally-expensive deterministic models. Consequently, a procedure that can reduce once again this time cost is needed. An efficient combination between the SPCE methodology and the GSA is proposed in this section. The basic idea of this combination is that, for a given discretized random field, the obtained random variables do not have the same weight in the variability of the system response. The variables with a very small contribution in the variability of the system response can be discarded which significantly reduces the dimensionality of the treated problem. This allows one to perform a probabilistic analysis using a reduced Experiment Design (ED) and thus a smaller number of calls of the computationally-expensive deterministic model. The main challenge remains in detecting the most influential random variables in order to reduce the dimensionality of the problem. For this purpose, a procedure that makes use of both the SPCE and the GSA (denoted hereafter by SPCE/GSA) is proposed in this regard. The SPCE/GSA procedure can be summarized by the following steps:

a) Discretize the random field(s): This step was made in this paper using EOLE method and its extensions by Vořechovsky (2008). Let us consider N_{RF} anisotropic non-Gaussian cross-correlated random fields $Z_i^{NG}(x, y)$ ($i = 1,..., N_{RF}$) described by: (i) constant means and standard deviations (μ_i, σ_i; $i = 1,..., N_{RF}$), (ii) non-Gaussian marginal cumulative distribution functions CDFs named G_i ($i = 1,..., N_{RF}$), (iii) a target cross-correlation matrix C^{NG} and (iv) a common square exponential autocorrelation function ρ_Z^{NG}[(x, y), (x', y')] which gives the values of the correlation function between two arbitrary points (x, y) and (x', y'). This autocorrelation function is given as follows:

$$\rho_z^{NG}[(x, y), (x', y')] = \exp\left(-\left(\frac{x - x'}{a_x} \right)^2 - \left(\frac{y - y'}{a_y} \right)^2 \right) \tag{13}$$

where a_x and a_y are the autocorrelation distances along x and y respectively. The Expansion Optimal Linear Estimation method (EOLE) and its extension by Vořechovsky (2008) to cover the case of correlated non-Gaussian random fields are used herein to generate the N_{RF} random fields. Notice that EOLE was first proposed by Li and Der kiureghian (1993) for the case of uncorrelated Gaussian fields, and then extended by Vořechovsky (2008) to cover the case of correlated non-Gaussian fields. In this method, one should first define a stochastic grid composed of q grid points (or nodes) $\{(x_1, y_1), ..., (x_q, y_q)\}$ for which the values of the field are assembled in a vector

$\chi = \left\{ Z(x_1, y_1), ..., Z(x_q, y_q) \right\}$. Then, one should determine the common correlation matrix for which each element $\left(\Sigma_{\chi;\chi} \right)_{i,j}^{NG}$ is calculated as follows:

$$\left(\Sigma_{\chi;\chi} \right)_{i,j}^{NG} = \rho_z^{NG} \left[(x_i, y_i), (x_j, y_j) \right] \tag{14}$$

The common non-Gaussian autocorrelation matrix $\Sigma_{\chi;\chi}^{NG}$ and the target non-Gaussian cross-correlation matrix C^{NG} should be transformed into the Gaussian space using Nataf model (Nataf, 1962) since the discretization of the random fields using EOLE is done in the Gaussian space. As a result, one obtains N_{RF} Gaussian autocorrelation matrices $\Sigma_{\chi;\chi}^{i} (i = 1, ..., N_{RF})$, and a Gaussian cross-correlation matrix C that can be used to discretize the two random fields. The value $\tilde{Z}_i$ of a random field obtained using this method is given by the following equation (cf. Al-Bittar and Soubra, 2011, 2012):

$$\tilde{Z}_i(x, y) \cong \mu_i + \sigma_i \sum_{j=1}^{N} \frac{\kappa_{i,j}^{D}}{\sqrt{\lambda_j^i}} . \left(\phi_j^i \right)^T . \Sigma_{Z(x,y);\chi}^{i} \qquad i = 1, ..., N_{RF} \tag{15}$$

where N_{RF} is the number of random fields, N is the number of terms retained in the series expansion, $\kappa_{i,j}^{D}$ are N_{RF} cross-correlated blocks of independent standard normal random variables obtained using the Gaussian cross-correlation matrix C between the N_{RF} fields, $(\lambda_j^i, \phi_j^i; i = 1, ..., N_{RF})$ are the eigenvalues and eigenvectors of the N_{RF} Gaussian autocorrelation matrices $\Sigma_{\chi;\chi}^{i}$ evaluated at the different points of the stochastic mesh, and $\Sigma_{Z(x,y);\chi}$ is the correlation vector between the value of the field $\tilde{Z}_i$ at an arbitrary point (x, y) and its values at the different points of the stochastic mesh. Notice finally that $\kappa_{i,j}^{D}$, ϕ_j^i, and $\Sigma_{Z(x,y);\chi}$ in Eq.(15) are vectors whose size is equal to N.

Once the two Gaussian random fields are obtained, they should be transformed into the non-Gaussian space (in case of non-Gaussian random fields) by applying the following formula:

$$\tilde{Z}_i^{NG}(x, y) = G_i^{-1} \left\{ \Phi \left[\tilde{Z}_i(x, y) \right] \right\} \qquad i = 1, ..., N_{RF} \tag{16}$$

where $\Phi(.)$ is the standard normal cumulative density function (CDF). For more details about the EOLE method and its extensions to cover the case of cross-correlated non-Gaussian random fields, the reader may refer to Vořechovsky (2008) and Al-Bittar and Soubra (2011, 2012).

After the discretization procedure, a random field is represented by N independent standard normal random variables. For the N_{RF} random fields that have the same autocorrelation function, the total number of random variables is $N_T = N_{RF} \times N$ which can be relatively large especially for small values of the autocorrelation distances.

b) Use a preliminary small order of the sparse polynomial chaos expansion (e.g. p=2) to approximate the system response by a meta-model. The main reason for selecting a small order is the exploration of the most influential random variables (i.e. those that have a significant weight in the variability of the system response) using a small Experiment Design (ED). It should be emphasized here that the

reduced number of the unknown SPCE coefficients related to the small value of the SPCE order leads to a significant decrease in the size of the experiment design, i.e. in the number of calls of the deterministic model.

c) Perform a GSA based on Sobol indices (using the obtained second order SPCE) to determine the weight of each random variable in the variability of the system response. The variables with very small values of their Sobol indices have no significant impact in the variability of the system response and can thus be discarded. Consequently, a response that only depends on a smaller number of random variables is obtained. In other words, one obtains a response with an 'effective dimension' N_e that is smaller than the initial dimension where the total number N_T of random variables was considered. It should be mentioned here that the small SPCE order (i.e. p=2) used firstly to perform the GSA is sufficient to provide the weight of each random variable in the variability of the system response since higher SPCE orders lead to the same influential random variables as will be seen later in the numerical results.
d) Use the same Experiment Design (ED) which was employed in step (b) but this time by only keeping the most influential random variables. By reducing the number of random variables from N_T to N_e ($N_e < N_T$), one has the possibility to use a higher SPCE order (i.e. p>2). The use of a higher SPCE order is necessary to lead to an improved fit of the SPCE since the leave-one-out error estimate Q^2 given in Eq. (8) increases when the SPCE order increases as it will be shown in the numerical results.

As a conclusion, the use of the SPCE/GSA procedure has the advantage of performing a good fit of the deterministic model with a reduced number of model evaluations as compared to the classical SPCE approach.

5. Numerical results

The aim of this section is to validate the present SPCE/GSA procedure. For this purpose, a comparison between the results obtained by using the classical SPCE methodology and those given by the proposed SPCE/GSA procedure is presented. A computationally non expensive deterministic model was used for the validation. The problem used for the validation was presented in Al-Bittar and Soubra (2011, 2012). It aims at computing the ultimate bearing capacity of a strip footing resting on a c-φ spatially varying soil. The input parameters are similar to those considered in Al-Bittar and Soubra (2011, 2012). They are briefly presented in Table 1. For a more detailed description on these data, the reader may refer to Al-Bittar and Soubra (2011, 2012). The deterministic model is based on numerical simulations using FLAC3D and it involves the case of a weightless soil. Thus, one obtains the soil bearing pressure due to only the soil cohesion; the contribution of the soil friction angle being neglected in the present paper. It should be mentioned here that when neglecting the soil weight γ, the computation time decreases from 10 to 5 min per simulation. This significantly reduces the computation time for the validation of the present SPCE/GSA procedure.

As shown in Figure 1, the adopted soil domain considered in the analysis is 15m wide by 6m deep. For the boundary conditions, the horizontal movement on the vertical boundaries of the grid is restrained, while the base of the grid is not allowed to move in both the horizontal and the vertical directions.

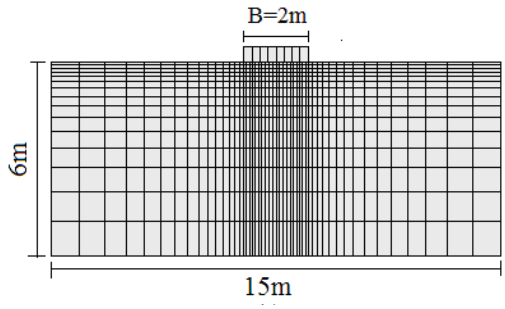

Figure 1. The adopted soil domain

Table 1. Deterministic and probabilistic input parameters

Model	Parameters	Parameters type	Deterministic parameters values	Stochastic parameters		
				Statistical characteristics	Coefficient of correlation *r*	Autocorrelation function $\rho_Z^{NG}[(x,y),(x',y')]$
Soil: Elastic-perfectly plastic (Mohr-Coulomb)	Young modulus E	deterministic	$E = 60MPa$	-	-	-
	Poisson ratio ν	deterministic	$\nu = 0.3$	-	-	-
	Cohesion c	Random field	-	Lognormal $\mu_c = 20kPa$, $Cov_c = 25\%$	Reference case: $C^{NG} = \begin{bmatrix} r(c,c)=0 & r(c,\varphi)=-0.5 \\ r(\varphi,c)=-0.5 & r(\varphi,\varphi)=0 \end{bmatrix}$	Square exponential $\rho_Z^{NG} = \exp\left(-\left(\frac{x-x'}{a_x}\right)^2 - \left(\frac{y-y'}{a_y}\right)^2\right)$
	Friction angle φ	Random field	-	Beta $\mu_\varphi = 30^0$, $Cov_\varphi = 10\%$	Parametric study: $-0.5 \le r(c,\varphi) \le 0$	a_x and a_y are the autocorrelation distances along *x* and y respectively
	Dilation ψ	Random field	-	Beta $\mu_\psi = 20^0$, $Cov_\psi = 10\%$		Reference case: a_x=10m and a_y=1m; Parametric study: $2m \le a_x \le 50m$; $0.5m \le a_y \le 8m$
Foundation: Elastic	Young modulus E	deterministic	$E = 25GPa$	-	-	-
	Poisson ratio ν	deterministic	$\nu = 0.4$	-	-	-
Interface: Elastic-perfectly plastic (Mohr-Coulomb)	Shear stiffness K_s	deterministic	$K_s = 1GPa$	-	-	-
	Normal stiffness K_n	deterministic	$K_n = 1GPa$	-	-	-
	Cohesion c_{int}	deterministic	$c_{int} = 20kPa$	-	-	-
	Friction angle φ_{int}	deterministic	$\varphi_{int} = 30^0$	-	-	-
	Dilation ψ_{int}	deterministic	$\psi_{int} = 20^0$	-	-	-

The validation of the SPCE/GSA procedure is done for the illustrative case [a_x=10m, a_y=1m, r(c, φ) =-0.5] referred to hereafter as the reference case. For this configuration, the discretization of the two random fields c and φ has led to a total number of random variables N_T equal to 24 (12 random variables for each random field as was shown in Al-Bittar and Soubra (2011, 2012)). By using the total number of random variables N_T, Al-Bittar and Soubra (2011, 2012) have shown that a third order SPCE was sufficient to reach a target accuracy of 0.999. An ED involving 800 points was needed to solve the regression problem given in Eq. (4) (i.e. to obtain a well-conditioned regression problem for which the rank of the matrix $(\eta^T\eta)^{-1}$ is larger than or equal to the number of unknown coefficients). By using the present SPCE/GSA procedure, a GSA was performed to detect the most influential random variables. Different SPCE orders (i.e. orders 2, 3, and 4) were considered in order to check if the SPCE order has an impact on the most influential random variables.

Figure 2 depicts the values of Sobol indices for the 24 random variables, as given by SPCEs of orders 2, 3 and 4. The first 12 random variables [i.e. ξ_i for i=1, …, 12] correspond to the cohesion random field and the last 12 random variables [i.e. ξ_i for i=13, …, 24] are those corresponding to the friction angle random field. Figure 2 shows that whatever the SPCE order is, the two first random variables of both fields, (i.e. ξ_1, ξ_2, ξ_{13}, ξ_{14}) are the most influential. For the two random fields, a very fast decay in the weight of the random variables is noticed with quasi negligible values beyond the first two random variables. In fact, the first two random variables of the two random fields, which correspond to the first two eigenmodes of both fields involve 95% of the response variability as may be seen from Table 2. This is logical since the system response (i.e. the ultimate bearing capacity) is an averaged quantity over the soil domain which is therefore quite insensitive to small-scale fluctuations of the spatially varying shear strength parameters c and φ.

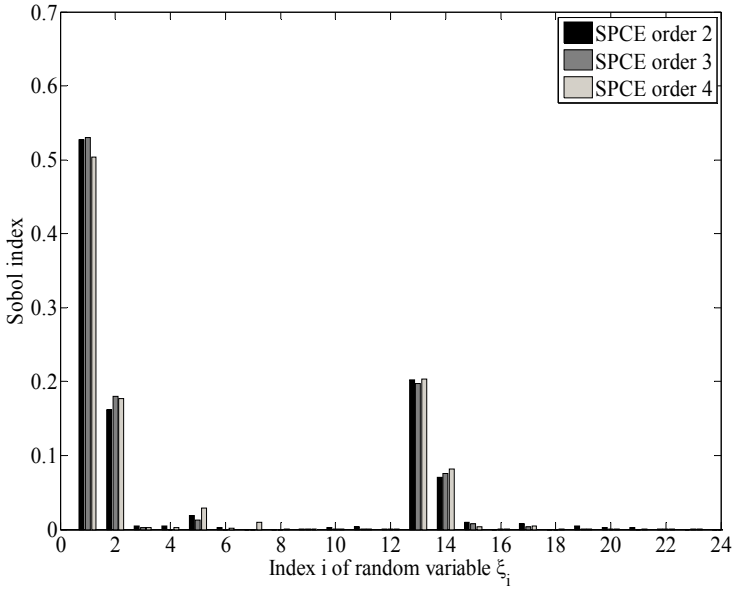

Figure 2. Sobol indices for SPCEs of orders 2, 3 and 4 using the total number of eigenmodes ξ_i (i=1, …, 24)

Table 2. Sobol indices for the reference case where a_x=10m, a_y=1m, and r(c,φ)=-0.5

	ξ_i (i=1, ..., 12) for the cohesion random field											
	ξ_1	ξ_2	ξ_3	ξ_4	ξ_5	ξ_6	ξ_7	ξ_8	ξ_9	ξ_{10}	ξ_{11}	ξ_{12}
Sobol index	0.50	0.17	0.002	0.002	0.03	0.002	0.009	0.0002	0.0002	9×10^{-05}	0.0002	7×10^{-05}
	ξ_i (i=13, ..., 24) for the friction angle random field											
	ξ_{13}	ξ_{14}	ξ_{15}	ξ_{16}	ξ_{17}	ξ_{18}	ξ_{19}	ξ_{20}	ξ_{21}	ξ_{22}	ξ_{23}	ξ_{24}
Sobol index	0.2	0.08	0.001	0.0008	0.002	0.0005	0.0006	0.0003	0.0001	4×10^{-05}	4×10^{-05}	5×10^{-05}

Figure 2 clearly shows that the Sobol indices of the different random variables do not significantly change with the SPCE order. Thus, a second order SPCE is sufficient to identify the influential random variables (i.e. those that have a significant weight in the variability of the ultimate bearing capacity). Increasing the SPCE order has led to the same influential random variables which justify the small SPCE order chosen to perform the preliminary investigations. The main advantage of a small SPCE order is that a small ED is sufficient to solve the regression problem. As shown in Table 3, 150 calls of the deterministic model are needed to solve the regression problem for a second order SPCE. This number attains 800 for a fourth order SPCE. This significant increase is because the number of unknown coefficients significantly increases from 29 to 144 when one chooses a fourth SPCE order instead of a second SPCE order.

Table 3. Number of unknown coefficients and model evaluations for different SPCE order

SPCE order	2	3	4
Number of unknown coefficients P	29	35	144
Number of model evaluations	150	350	800

To choose the number of random variables which will be retained hereafter, the different random variables of the two random fields are firstly sorted in a descending order according to the values of their Sobol indices (cf. first and second columns in Table 4). A threshold of acceptance t_a is then fixed as a percentage of the most influential (weighted) random variable. In the present paper, the most influential random variable is ξ_1 and it has a Sobol index equal to 0.5. Different values of the threshold were tested (cf. first line in Table 4). The random variables having a Sobol index smaller than the prescribed threshold t_a are discarded. In this paper, a threshold of 2% of the Sobol index of the most weighed random variable is considered as sufficient; the corresponding retained random variables provide 98% of the total variance of the system response as may be seen from column 6 of Table 4. For this threshold, an 'effective dimension' N_e=5 is obtained (i.e. 5 random variables are considered to be the most weighed). The 5 retained random variables will now be used with the already existing 150 model evaluations which were firstly employed to approximate the second order SPCE with the total number of random variables N_T=24.

The reduction in the number of random variables from N_T=24 to N_e=5 provides the possibility to use higher SPCE orders (i.e. p>2) with the same ED (i.e. the 150 model evaluations). The use of a higher SPCE order is necessary to lead to an improved fit of the SPCE since the leave-one-out error estimate Q^2 given in Eq. (8) increases when the SPCE order increases as shown in Table 5 for both the classical SPCE approach (using the total number of random variables N_T=24) and the present SPCE/GSA

procedure (where the effective dimension is equal to 5 (i.e. N_e=5)). Using the SPCE/GSA procedure, an SPCE up to p=8 was reached using only 150 model evaluations. From Table 5, one can notice that with the use of the SPCE/GSA procedure, the Q^2 increases with the increase of the SPCE order and stabilizes beyond the order 5. This means that no improvement in the fit is obtained beyond this order. On the other hand, the value of Q^2 given by the present approach is smaller than the classical SPCE approach with a fourth order. This is because 19 random variables were discarded which slightly affect the goodness of the fit.

Figure 3 shows the PDF of the ultimate bearing capacity as obtained by both the classical SPCE approach (with the total number of random variables N_T =24) and the proposed SPCE/GSA procedure (using only five random variables). Table 6 provides the corresponding statistical moments and error estimates. Notice that the results of the present SPCE/GSA approach are given in Table 6 for different values of the model evaluations (from 150 to 800). From this table, one can see that the error estimate of the SPCE/GSA procedure is quasi constant with the increase in the number of model evaluations. This means that 150 model evaluations are sufficient and there is no need for more model evaluations to improve the accuracy of the fit. On the other hand, one can observe (see Figure 3 and Table 6) that the first two statistical moments are well estimated with the present SPCE/GSA approach using the 150 model evaluations. However, the third and fourth statistical moments need more model evaluations (800 model evaluations) in order to converge to their reference values given by the SPCE approach (cf. Table 6). This demonstrates the efficiency of the present SPCE/GSA procedure to compute the first two statistical moments with a much reduced number of the model evaluations with respect to the classical SPCE approach.

As for the Sobol indices of the two random fields c and φ, Table 7 shows that the SPCE/GSA procedure with only 150 model evaluations gives the same results obtained by the classical SPCE approach using 800 model evaluations which demonstrates once again the efficiency of the present SPCE/GSA procedure.

Table 4. Sobol indices of the different random variables and the retained random variables for the different values of the threshold of acceptance

Random variable	Sobol index	t_a=0.5% x ξ_1 =0.0025	t_a=1% x ξ_1 =0.005	t_a=1.5% x ξ_1 =0.0075	t_a=2% x ξ_1 =0.01	t_a=2.5% x ξ_1 =0.0125	t_a=3% x ξ_1 =0.015	t_a=4% x ξ_1 =0.02	t_a=5% x ξ_1 =0.025
ξ_1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
ξ_{13}	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
ξ_2	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17
ξ_{14}	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08
ξ_5	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
ξ_7	0.009	0.009	0.009	0.009					
ξ_6	0.002	0.002							
ξ_{17}	0.002	0.002							
ξ_3	0.002	0.002							
ξ_4	0.002	0.002							
ξ_{15}	0.001								
ξ_{16}	0.0008								
ξ_{19}	0.0006								
ξ_{18}	0.0005								
ξ_{20}	0.0003								
ξ_8	0.0002								
ξ_9	0.0002								
ξ_{11}	0.0002								
ξ_{21}	0.0001								
ξ_{10}	9.0×10^{-05}								
ξ_{12}	7.0×10^{-05}								
ξ_{24}	5.0×10^{-05}								
ξ_{22}	4.0×10^{-05}								
ξ_{23}	4.0×10^{-05}								
Sum of Sobol indices	1.001	0.997	0.989	0.989	0.98	0.98	0.98	0.98	0.98

Table 5. SPCE using the total and the reduced number of random variables

	SPCE order	2	3	4	5	6	7	8
Total number of random variables N_T	Coefficient of determination R^2	0.998	0.999	0.999	-	-	-	-
	Leave-one-out cross-validation Q^2	0.824	0.932	0.9943	-	-	-	-
Reduced number of random variables N_e	Coefficient of determination R^2	0.961	0.963	0.968	0.970	0.972	0.972	0.972
	Leave-one-out cross-validation Q^2	0.791	0.883	0.957	0.961	0.963	0.963	0.963

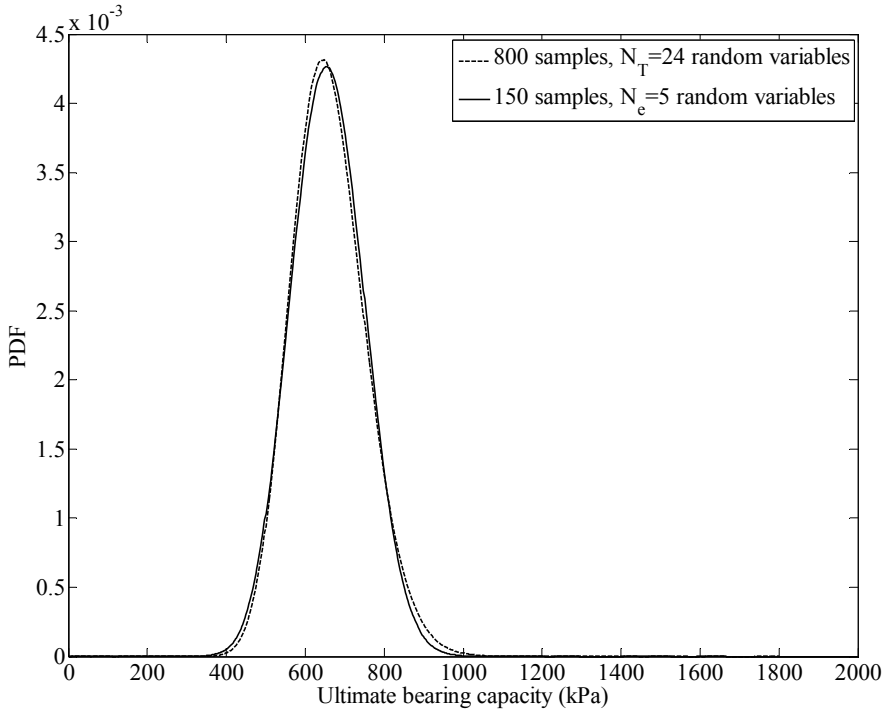

Figure 3. PDF of the ultimate bearing capacity for both the classical SPCE with the total number of random variables N_T =24 and the proposed SPCE/GSA procedure with only five random variables N_e =5.

Table 6. Error estimates of the SPCE and statistical moments of the ultimate bearing capacity as given by the classical SPCE approach and by the present SPCE/GSA procedure

	Number of model evaluations	Mean μ_{qult} (kPa)	Standard deviation σ_{qult} (kPa)	Skewness δ_u (-)	Kurtosis κ_u (-)	R^2	Q^2
With the total number of random variables N_T=24	800	658.2	93.57	0.287	0.163	0.999	0.995
With the reduced number of random variables N_e=5	150	657.84	90.80	0.105	0.0129	0.972	0.957
	200	658.98	91.53	0.168	0.0563	0.972	0.951
	250	659.90	92.10	0.188	0.0630	0.964	0.956
	300	659.73	92.15	0.202	0.0600	0.962	0.963
	400	660.05	90.95	0.291	0.0500	0.969	0.960
	500	659.50	90.81	0.296	0.0430	0.970	0.963
	600	659.75	90.99	0.272	0.116	0.968	0.963
	700	659.50	90.85	0.280	0.1637	0.968	0.963
	800	659.85	91.20	0.30	0.160	0.970	0.967

Table 7. Sobol indices as computed from the classical SPCE approach and the present SPCE/GSA procedure.

	Number of model evaluations	i	S_i (i=1, ..., 12)	i	S_i (i=13, ..., 24)	$S\left(c\right)=\sum_{i=1}^{12} S_i$	$S\left(\varphi\right)=\sum_{i=13}^{24} S_i$
With the total number of random variables N_T=24	800	1	0.5	13	0.2	0.715	0.285
		2	0.17	14	0.08		
		3	0.002	15	0.001		
		4	0.002	16	0.0008		
		5	0.03	17	0.002		
		6	0.002	18	0.0005		
		7	0.009	19	0.0006		
		8	0.0002	20	0.0003		
		9	0.0002	21	0.0001		
		10	9.0×10^{-05}	22	4.0×10^{-05}		
		11	0.0002	23	4.0×10^{-05}		
		12	7.0×10^{-05}	24	5.0×10^{-05}		
	Number of model evaluations	i	S_i (i=1, 2, 3)	i	S_i (i=4, 5)	$S\left(c\right)=\sum_{i=1}^{3} S_i$	$S\left(\varphi\right)=\sum_{i=4}^{5} S_i$
With the reduced number of random variables N_e=5	150	1	0.510	4	0.076	0.721	0.279
		2	0.200	5	0.190		
		3	0.010				

6. Conclusions

An efficient combined use of the SPCE methodology and the global sensitivity analysis (GSA) has been proposed. The aim is to reduce the cost of the probabilistic analysis of computationally-expensive deterministic models. This methodology was validated in this paper using a relatively non-expensive deterministic model. The validation consists in comparing the results of both the classical SPCE methodology with the total number of random variables and the proposed combination between the SPCE and the GSA. Satisfactory results were obtained using a much smaller number of model evaluations with the proposed methodology. The first two statistical moments and the Sobol indices have been well estimated with the very small number of model evaluations. On the other hand, the third and fourth statistical moments need more model evaluations in order to converge to their reference values obtained using the classical SPCE. Since the present SPCE/GSA procedure was shown to be efficient for the probabilistic computation with a reduced calculation cost with respect to the classical SPCE approach, this approach may now be applied with confidence to costly deterministic models.

Appendix A

Illustrative Example

In order to illustrate the PCE theory in a simple manner, a PCE of order p=3 using only M=2 random variables (ξ_1 and ξ_2) will be considered in this illustrative example. Using the classical truncation scheme, Table A.1 presents the retained PCE terms which are those having a first order norm $\|\alpha\|_1$ smaller than or equal to p (i.e. p=3). These terms are presented in Table A.1 in bold characters. As may be easily seen from Table A.1, the PCE basis contains P=10 terms whose expressions are computed using Eq.(2).

Table A.1. Terms retained using the classical truncation scheme for M=2 and p=3

α_1	0	1	0	1	2	0	2	1	2	3	0	3	1	3	3	3
α_2	0	0	1	1	0	2	1	2	2	0	3	1	3	2	3	3
$\|\alpha\|_1$	**0**	**1**	**1**	**2**	**2**	**2**	**3**	**3**	4	**3**	**3**	4	4	5	6	6

Table A.2 presents the expressions of the PCE basis Ψ_β. Using Table A.2, on can write the PCE expression as function of the input random variables (ξ_1 and ξ_2) as follows:

$$Y=\Gamma_{PCE}(\xi)=a_0\Psi_0+a_1\Psi_1+...+a_9\Psi_9= a_0+a_1\xi_1+a_2\xi_2+a_3\xi_1\xi_2+a_4(\xi_1^2-1)+a_5(\xi_2^2-1)+a_6\left(\xi_1^2-1\right)\xi_2+a_7\xi_1\left(\xi_2^2-1\right)+a_8(\xi_1^3-3\xi_1)+a_9(\xi_2^3-3\xi_2) \quad (A.1)$$

In this expression, the unknown coefficients can be computed using Eq.(4) by simulating an ED which contains K initial realizations of the two random variables (ξ_1, ξ_2) and computing the corresponding responses from deterministic calculations. It should be mentioned here that the size K of the ED should ensure the numerical stability of the regression problem and thus it can be enriched each time the matrix $(\eta^T\eta)^{-1}$ is badly-conditioned.

The first order Sobol indices for the two random variable (ξ_1 and ξ_2) can be easily obtained once the coefficients $a_0,\ldots, a_9$ are computed using Eq. (10). The only additional step is to compute $E\left(\Psi_\beta^2\right)$ corresponding to these two random variables. Table A.2 shows the values of $E\left(\Psi_\beta^2\right)$ computed using Eq. (12) for the different Ψ_β terms. The expression of the first order Sobol indices of the two random variables ξ_1 and ξ_2 are written as follows:

$$S(\xi_1)=\frac{a_1^2+2a_4^2+6a_8^2}{a_1^2+2a_4^2+6a_8^2+a_2^2+2a_5^2+6a_9^2}; \qquad S(\xi_2)=\frac{a_2^2+2a_5^2+6a_9^2}{a_1^2+2a_4^2+6a_8^2+a_2^2+2a_5^2+6a_9^2} \quad (A.2)$$

with

$$I_1=(1,4,8); \qquad I_2=(2,5,9) \quad (A.3)$$

Table A.2. Basis of the PCE with the classical truncation scheme for M=2 and p=3

β	PCE order p	$\Psi_\beta = \prod_{i=1}^{M} H_{\alpha_i}(\xi_i)$	$E\left(\Psi_\beta^2\right) = \prod_{i=1}^{M} \alpha_i !$
0	P=0	$H_0(\xi_1)xH_0(\xi_2)=1$	$\alpha_1!$ x $\alpha_2!=0!x0!=1$
1	P=1	$H_1(\xi_1)xH_0(\xi_2)=\xi_1$	$\alpha_1!$ x $\alpha_2!=1!x0!=1$
2		$H_0(\xi_1)\ xH_1(\xi_2)=\xi_2$	$\alpha_1!$ x $\alpha_2!=0!x1!=1$
3	P=2	$H_1(\xi_1)\ xH_1(\xi_2)=\xi_1\ \xi_2$	$\alpha_1!$ x $\alpha_2!=1!x1!=1$
4		$H_2(\xi_1)\ xH_0(\xi_2)=\xi_1^2-1$	$\alpha_1!$ x $\alpha_2!=2!x0!=2$
5		$H0(\xi_1)\ xH_2(\xi_2)=\xi_2^2-1$	$\alpha_1!$ x $\alpha_2!=0!x2!=2$
6	P=3	$H_2(\xi_1)\ xH_1(\xi_2)=\left(\xi_1^2-1\right)\xi_2$	$\alpha_1!$ x $\alpha_2!=2!x1!=2$
7		$H_1(\xi_1)\ xH_2(\xi_2)=\xi_1\left(\xi_2^2-1\right)$	$\alpha_1!$ x $\alpha_2!=1!x2!=2$
8		$H_3(\xi_1)\ xH_0(\xi_2)=\xi_1^3-3\xi_1$	$\alpha_1!$ x $\alpha_2!=3!x0!=6$
9		$H_0(\xi_1)\ xH_3(\xi_2)=\xi_2^3-3\xi_2$	$\alpha_1!$ x $\alpha_2!=0!x3!=6$

References

Al-Bittar, T. and Soubra, A. H. Bearing capacity of strip footing on spatially random soils using sparse polynomial chaos expansion. *International Journal for Numerical and Analytical Methods in Geomechanics*, accepted paper, 2012.

Al-Bittar, T. and Soubra, A. H. Bearing capacity of strip footing on spatially random soils using sparse polynomial chaos expansion. *GeoRisk* (GSP 224), ASCE, Atlanta, USA, 26-28 June 2011.

Blatman, G. and Sudret, B. An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis. *Probabilistic Engineering Mechanics*, 25: 183-197, 2010.

Huang, S. P., Liang, B. and Phoon K. K. Geotechnical probabilistic analysis by collocation-based stochastic response surface method: An Excel Add-in implementation. *Georisk*, 3: 75-86, , 2009.

Isukapalli, S. S., Roy, A. and Georgopoulos, PG. Stochastic response surface methods (SRSMs) for uncertainty propagation: Application to environmental and biological systems. *Risk Analysis* , 18: 357-363, 1998.

Mollon, G., Dias, D. and Soubra, A. H. Probabilistic analysis of pressurized tunnels against face stability using collocation-based stochastic response surface method. *Journal of Geotechnical and Geoenvironmental Engineering ASCE*, 137: 385-397, 2011.

Mao, N., Al-Bittar, T. and Soubra, A. H. Probabilistic analysis and design of strip foundations resting on rocks obeying Hoek-Brown failure criterion. *International Journal of Rock Mechanics and Mining Sciences*, 49: 45-58, 2012.

Nataf, A. Détermination des distributions de probabilités dont les marges sont données. *CR Acad Sci*, 225: 42-3, 1962.

Saltelli, A., Chan, K. and Scott E. M. *Sensitivity Analysis*. John wiley and sons, 2000.

Spanos, P. D. and Ghanem, R. Stochastic finite element expansion for random media. *Journal of Engineering Mechanics*, 115: 1035-1053, 1989.

Sobol, I. M. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mechanics and Computation in Simulation*, 55: 271-280, 2001.

Sudret, B. Global sensitivity analysis using polynomial chaos expansion. *Reliability Engineering and System Safety*, 93: 964-979, 2008.

Vořechovský, M. Simulation of simply cross-correlated random fields by series expansion methods. *Structural Safety*, 30: 337-363, 2008.

Xiu, D. and Karniadakis, G. E. The Wiener-Askey polynomial chaos for stochastic differential equations. *Journal of Scientific Computing*, 24: 619-44, 2002.