



**HAL**  
open science

## Advanced reliability analysis methods

Abdel-Hamid Soubra, Emilio Bastidas-Arteaga

► **To cite this version:**

Abdel-Hamid Soubra, Emilio Bastidas-Arteaga. Advanced reliability analysis methods. ALERT Doctoral School 2014 - Stochastic Analysis and Inverse Modelling, Michael A. Hicks; Cristina Jommi, pp.79-93, 2014, 978-2-9542517-5-2. in2p3-01084374v2

**HAL Id: in2p3-01084374**

**<https://hal.in2p3.fr/in2p3-01084374v2>**

Submitted on 19 Feb 2015

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

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

---

# Advanced reliability analysis methods

**A.-H. Soubra and E. Bastidas-Arteaga**

*University of Nantes – GeM Laboratory - UMR CNRS 6183 – Nantes  
– France*

---

*This chapter presents the subset simulation (SS) approach and the Polynomial Chaos Expansion (PCE) methodology. The SS method is an efficient alternative to the well-known Monte Carlo Simulation (MCS) methodology to calculate small failure probabilities. The basic idea of the SS approach is that the small failure probability can be expressed as a product of larger conditional failure probabilities. On the other hand, the PCE methodology allows one to accurately compute the PDF of a given system response using a reduced number of calls of the deterministic model (as compared to the classical MCS applied on the original complex deterministic model). Indeed, the PCE methodology replaces the computationally-expensive deterministic model by a meta-model. Once the meta-model is determined, MCS can be applied on the obtained PCE to compute the PDF of the system response with a quasi-negligible computation time.*

## **1 Introduction**

The most robust method used for the probabilistic analysis of geotechnical structures is the classical well-known Monte Carlo Simulation (*MCS*) methodology. It should be noted that the probabilistic analysis of an engineering system involves the computation of the PDF of the system response or the calculation of the failure probability for a prescribed threshold of this system response.

*MCS* is not suitable for the computation of the small failure probabilities encountered in the practice of geotechnical engineering (especially when using a computationally-expensive finite element/finite difference deterministic model) due to the large number of simulations required to calculate a small failure probability. As an alternative to *MCS* methodology, [Au01] proposed the subset simulation (*SS*) approach to calculate small failure probabilities. The basic idea of the *SS* approach is that the small failure probability can be expressed as a product of larger conditional failure probabilities.

Similarly, MCS is not suitable for the accurate determination of the PDF of a system response because of the great number of calls of the deterministic model, which are required for such a computation. The PCE methodology allows one to approximate a given system response by a polynomial chaos expansion (PCE) of a suitable order. Thus, the PCE methodology replaces the computationally-expensive deterministic model by a meta-model (i.e. a simple analytical equation). Once the PCE coefficients are determined, MCS can be applied on the obtained PCE to compute the PDF of the system response (and the corresponding statistical moments) with a quasi-negligible computation time.

## 2 Subset simulation (SS) approach

Subset simulation (SS) approach was proposed by [Au01] as alternative to Monte Carlo Simulation (MCS) methodology to compute small failure probabilities. The basic idea of the SS approach is that the small failure probability can be expressed as a product of larger conditional failure probabilities. In this section, one presents a brief description of the steps of SS approach in case of two random variables, the extension to the case of several random variables being straightforward. A detailed description of the SS approach may be found in [Au01], in the chapters 1 and 4 of the book by [Pho08] and in [Ahm12].

The steps of SS approach in case of two random variables ( $V_1, V_2$ ) can be described as follows:

1. Generate a vector of two random variables ( $V_1, V_2$ ) according to a target *PDF* using direct Monte Carlo simulation.
2. Using the deterministic model, calculate the system response corresponding to ( $V_1, V_2$ ).
3. Repeat steps 1 and 2 until obtaining a prescribed number  $N_s$  of vectors of random variables and the corresponding values of the system response.
4. Determine the value of the performance function corresponding to each value of the system response and then, arrange the values of the performance function in an increasing order within a vector  $G_0$  where  $G_0 = \{G_0^1, \dots, G_0^k, \dots, G_0^{N_s}\}$ . Notice that the subscripts ‘ $0$ ’ refer to the first level (level  $0$ ) of the subset simulation approach.
5. Prescribe a constant intermediate conditional failure probability  $p_0$  for the failure regions  $F_j \{j = 1, 2, \dots, m - 1\}$  and evaluate the first failure threshold  $C_1$  which corresponds to the first level of SS approach (see Figure 1). The failure threshold  $C_1$  is equal to the  $[(N_s \times p_0) + 1]^{th}$  value in the increasing list of elements of the vector  $G_0$ . This means that the value of the conditional failure probability of the first level  $P(F_1)$  will be equal to the prescribed  $p_0$  value.
6. Among the  $N_s$  vectors of random variables, there are  $[N_s \times p_0]$  ones whose values of the performance function are less than  $C_1$  (i.e. they are located in the failure region  $F_1$ ). These vectors are used as ‘mother vectors’ to generate  $N_s$  new

vectors of random variables (according to a proposal  $P_p$ ) using Markov chain method based on the modified Metropolis-Hastings algorithm by [San11]. This algorithm is presented in Appendix 1.

7. Using the deterministic model, calculate the values of the system response corresponding to the new vectors of random variables (which are located in level 1). Then, calculate the corresponding values of the performance function. Finally, gather the values of the performance function in an increasing order within a vector  $G_1$  where  $G_1 = \{G_1^1, \dots, G_1^k, \dots, G_1^{N_s}\}$ .
8. Evaluate the second failure threshold  $C_2$  as the  $[(N_s \times p_0) + 1]^{th}$  value in the increasing list of elements of the vector  $G_1$ .
9. Repeat steps 6-8 to evaluate the failure thresholds  $C_3, C_4, \dots, C_m$  corresponding to the failure regions  $F_3, F_4, \dots, F_m$ . Notice that contrary to all other thresholds, the last failure threshold  $C_m$  is negative. Thus,  $C_m$  is set to zero and the conditional failure probability of the last level  $P(F_m | F_{m-1})$  is calculated as:

$$P(F_m | F_{m-1}) = \frac{1}{N_s} \sum_{k=1}^{N_s} I_{F_m}(s_k) \quad (1)$$

where  $I_{F_m} = 1$  if the performance function  $G(s_k)$  is negative and  $I_{F_m} = 0$  otherwise.

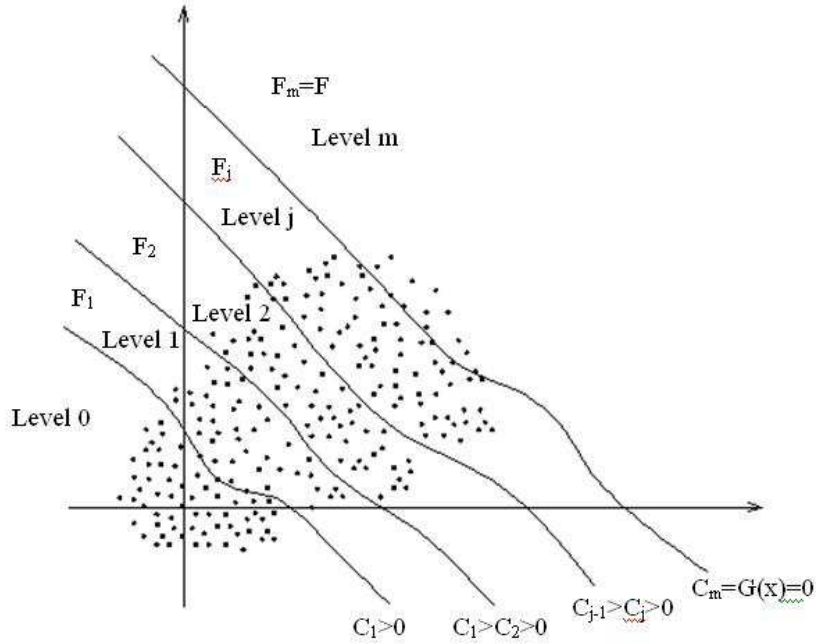


Figure 1: Nested Failure domain.

Finally, the failure probability  $P(F)$  is evaluated as follows:

$$P(F) = P(F_1) \prod_{j=2}^m P(F_j | F_{j-1}) \quad (2)$$

It should be mentioned that a normal *PDF* was used herein as a target probability density function  $P_t$ . However, a uniform *PDF* was used as a proposal probability density function  $P_p$  (for more details, refer to Appendix 1).

## 2.1 Example application

In order to illustrate the algorithm of *SS* methodology in a simple way, a numerical example is provided herein. In this example, *SS* approach was used to calculate the failure probability  $P_f$  against bearing capacity failure of a strip footing of breadth  $B$ . The footing rests on a  $(c, \varphi)$  soil and it is subjected to a service vertical load  $P_s$ . The soil cohesion  $c$  and the soil angle of internal friction  $\varphi$  were considered as random variables. The following formula was used for the computation of the ultimate bearing capacity:

$$q_u = \gamma \frac{B}{2} N_\gamma + c N_c + q N_q \quad (3)$$

in which:

$$N_\gamma = 2(N_q - 1) \tan \varphi \quad (4a)$$

$$N_q = e^{\pi \tan \varphi} \cdot \tan^2 \left( \frac{\pi}{4} + \frac{\varphi}{2} \right) \quad (4b)$$

$$N_c = \frac{N_q - 1}{\tan \varphi} \quad (4c)$$

where  $N_\gamma$ ,  $N_q$  and  $N_c$  are the bearing capacity factors due to soil weight, surcharge loading and cohesion respectively. These coefficients are function of the soil friction angle. On the other hand,  $\gamma$  is the soil unit weight and  $q$  is the surcharge loading. The performance function used in the analysis is:

$$G = \frac{P_u}{P_s} - 1 \quad (5)$$

where  $P_u$  is the ultimate footing load and  $P_s$  is the footing applied load. As mentioned previously, only the soil cohesion and friction angle were considered as random variables. All the other parameters were considered as deterministic. These parameters are given in Table (1).

In this example, the intermediate failure probability  $p_0$  of a given level  $j$  ( $j = 1, 2, \dots, m - 1$ ) was arbitrary chosen equal to 0.2. A small number of samples per level ( $N_s = 10$  samples) was used to facilitate the illustration.

Table 1: Data used for the probabilistic analysis of a strip footing against bearing capacity failure

Parameter	Type of parameter	Mean and coefficient of variation of the parameter
Breadth $B$	Deterministic	2m
Surcharge loading $q$	Deterministic	10kPa
Soil unit weight $\gamma$	Deterministic	20kN/m <sup>3</sup>
Service vertical load $P_s$	Deterministic	1000kN/m
Cohesion $c$	Random normal variable	$\mu_c = 20\text{kPa}$
		$COV_c = 0.3$
Friction angle $\varphi$	Random normal variable	$\mu_\varphi = 30^\circ$
		$COV_\varphi = 0.1$

Table 2 presents (i) the values of  $c$  and  $\varphi$  of each sample for the successive levels (ii) the corresponding values of the performance function and (iii) the values of the failure thresholds  $C_j$  for the different levels. Notice that only the first two levels and the last level for which the failure threshold becomes negative were provided herein for illustration. Table (2) indicates that the failure threshold decreases with the successive levels until reaching a negative value at the last level. This means that the samples generated by the subset simulation successfully progress towards the limit state surface  $G = 0$ . In order to select the failure threshold of a given level, the calculated values of the performance function of this level were arranged in an increasing order as shown in Table (2). Then, the failure threshold was selected as the  $[(N_s \times p_0) + 1]^{th}$  value of the arranged values of the performance function. Since  $N_s = 10$  and  $p_0 = 0.2$ , the failure threshold is equal to the third value of the arranged values of the performance function. The *SS* computation continues until reaching a negative value (or a value of zero) of the failure threshold. In this example, the negative value was reached in the sixth level (where  $C_6 = -0.0936$ ) as shown in Table (2). Theoretically, the last failure threshold should be equal to zero. For this reason,  $C_6$  was set to zero. This means that the last conditional failure probability  $P(F_6|F_5)$  is not equal to  $p_0$ . In this case, the last conditional failure probability  $P(F_6|F_5)$  is calculated as the ratio between the number of samples for which the performance function is negative and the chosen number  $N_s$  of samples (i.e. 10). According to Table (2),  $P(F_6|F_5)$  is equal to  $3/10 = 0.3$ . Thus, the failure probability of the footing under consideration is equal to  $0.2^5 \times 0.3 = 9.6 \times 10^{-5}$ .

Table 2: Results of SS algorithm when  $N_s=10$  and  $p_0=0.2$

Level's number $j$	Cohesion $c$ (kPa)	Angle of internal friction $\varphi$ (deg)	Performance function	Failure threshold $C_j$
1	23.23	26.0	0.9256	1.4875
	31.00	39.1	1.1185	
	6.45	32.2	1.4875	
	25.17	29.9	1.5598	
	21.91	32.1	2.0023	
	12.15	29.4	2.6625	
	17.40	29.6	3.8598	
	22.06	34.5	4.9894	
	41.47	34.3	5.3910	
	36.62	34.3	9.1912	
2	25.83	26.5	0.8587	0.9740
	27.29	26.2	0.9411	
	25.32	26.8	0.9740	
	23.98	25.4	1.0561	
	25.92	26.0	1.0842	
	14.86	28.0	1.1505	
	14.14	28.8	1.1528	
	12.27	28.8	1.1747	
	13.14	29.4	1.1931	
	11.80	30.0	1.2361	
6	15.17	22.9	-0.1604	-0.0936
	14.88	23.0	-0.1003	
	14.88	23.0	-0.0936	
	14.56	22.5	0.0415	
	14.56	22.5	0.0718	
	15.84	22.5	0.0718	
	16.36	21.5	0.1156	
	14.53	20.7	0.1420	
	12.89	20.6	0.1476	
	15.43	20.3	0.1476	
$P_f$	$9.6 \times 10^{-5}$			

It should be emphasized that the failure probability calculated in Table (2) is not accurate due to the small value of  $N_s$ . For an accurate computation of the failure probability,  $N_s$  should be increased. This number should be greater than 100 to provide a small bias in the calculated  $P_f$  value (see chapter 4 by Honjo in [Pho08]).

In order to determine the optimal number of samples  $N_s$  to be used per level, different values of  $N_s$  were considered to calculate  $P_f$  and its coefficient of variation  $COV_{P_f}$  as shown in Table (3). The thresholds corresponding to each  $N_s$  value were calculated and were shown in this table. Table (3) indicates (as was shown before when  $N_s = 10$ ) that for the different values of  $N_s$ , the failure threshold decreases with the successive levels until reaching a negative value at the last level.

Table 3: Evolution of the failure threshold  $C_j$  with the different levels  $j$  and with the number of realizations  $N_s$  when  $p_0 = 0.2$

$C_j$ for level $j$	Number of samples $N_s$ per level						
	10	100	200	1000	2000	2200	2400
$C_1$	1.4875	0.9397	1.0071	1.0638	1.0532	1.0466	1.0803
$C_2$	0.9740	0.4157	0.3969	0.4916	0.4467	0.4466	0.4942
$C_3$	0.7391	0.1011	0.1016	0.1513	0.1434	0.1347	0.1549
$C_4$	0.4007	-0.0491	-0.0437	-0.0307	-0.0616	-0.0536	-0.0564
$C_5$	0.1573	-----	-----	-----	-----	-----	-----
$C_6$	-0.0936	-----	-----	-----	-----	-----	-----
$P_f(\times 10^{-3})$	0.096	2.80	2.72	2.20	2.80	2.60	2.63
$COV_{P_f}(\%)$	221.4	57.9	42.1	18.7	13.3	12.8	12.4

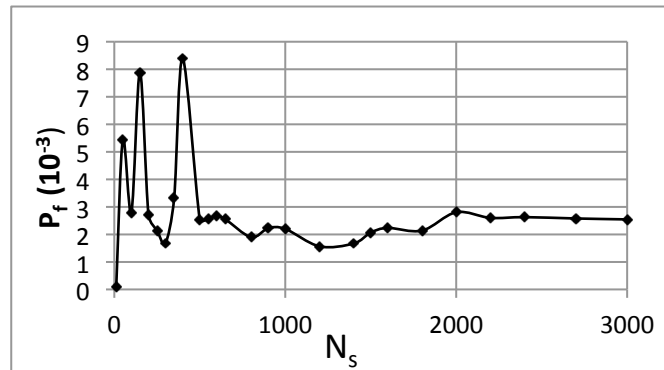
Figure (2a) shows the effect of  $N_s$  on the failure probability. It indicates that for small values of  $N_s$ , the failure probability largely changes with  $N_s$ . However, for high values of  $N_s$ , the failure probability converges to an almost constant value. Figure (2a) also indicates that 2200 samples per level are required to accurately calculate the failure probability. This is because (i) the  $C_j$  values corresponding to  $N_s=2200$  and 2400 samples are quasi similar as it may be seen from Table (3) and (ii) the corresponding final  $P_f$  values are too close (they are respectively equal to  $2.60 \times 10^{-3}$  and  $2.63 \times 10^{-3}$ ).

Figure (2b) shows the effect of  $N_s$  on the coefficient of variation of the failure probability  $COV_{P_f}$ . As expected,  $COV_{P_f}$  decreases with the increase of  $N_s$ . Notice that the values of  $COV_{P_f}$  for  $N_s = 2200$  and 2400 samples are equal to 12.8% and 12.4% which indicates (as expected) that the  $COV_{P_f}$  decreases with the increase in the number of realizations.

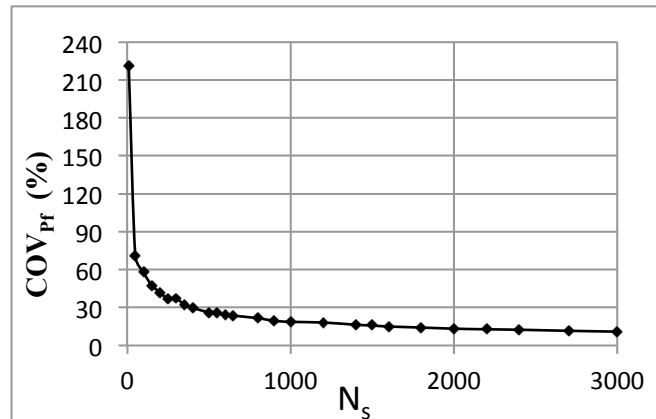
It should be mentioned here that for  $p_0 = 0.2$ , four levels of subset simulation were found necessary to reach the limit state surface  $G = 0$  as may be seen from Table (3). Therefore, when  $N_s=2200$  samples, a total number of  $N_t=2200 \times 4=8800$  samples were required to calculate the final  $P_f$  value. Remember that in this case, the  $COV$  of  $P_f$  was equal to 12.8%. Notice that if the same value of  $COV$  (i.e. 12.8%) is



desired by *MCS* to calculate  $P_f$ , the number of samples would be equal to 20000. This means that, for the same accuracy, the *SS* approach reduces the number of realizations by 56%. On the other hand, if one uses *MCS* with the same number of samples (i.e. 8800 realizations), the value of *COV* of  $P_f$  would be equal to 19.6%. This means that for the same computational effort, the *SS* approach provides a smaller value of  $COV(P_f)$  than *MCS*.



(a)



(b)

Figure 2:  $P_f$  and  $COV_{P_f}$  versus the number of realizations  $N_s$ .

### 3. Polynomial Chaos Expansion (PCE) methodology

The basic idea of this method is to approximate a given system response by a polynomial chaos expansion (PCE) of a suitable order. In other words, the PCE methodology replaces the computationally-expensive deterministic model by a meta-model.

In order to achieve this purpose, all the uncertain parameters (which may have different PDFs) should be represented by a unique chosen PDF. Table (4) presents the usual PDFs and their corresponding families of orthogonal polynomials.

Table 4: Usual probability density functions and their corresponding families of orthogonal polynomials

Probability density functions	Polynomials
Gaussian	Hermite
Gamma	Laguerre
Beta	Jacobi
Uniform	Legendre

Within the framework of the present methodology, the response of a system that involves  $n$  random variables can be expressed by a PCE as follows:

$$\Gamma_{PCE} = \sum_{i=0}^{P-1} a_i \Psi_i(\xi) \quad (6)$$

where  $\Psi_i(\xi)$  are multi-dimensional polynomials defined as the product of one-dimensional polynomials  $\phi_{\alpha_i}$ ,  $(\xi_1, \xi_2, \dots, \xi_n)$  are independent random variables,  $(a_1, a_2, \dots, a_n)$  are unknown coefficients to be evaluated and  $P$  is the size of the PCE.

The size  $P$  of the PCE (which is equal to the number of the unknown PCE coefficients) depends on the number  $n$  of random variables and the order  $p$  of the PCE. It is given as follows:

$$P = \frac{(n+p)!}{n! p!} \quad (7)$$

It should be mentioned here that in this chapter, the random variables are represented in the independent standard normal space. Thus, the suitable corresponding bases are the multidimensional Hermite polynomials as may be seen from Table (4). The expressions of the multi-dimensional Hermite polynomials are given as follows:

$$\Psi_{\alpha} = \prod_{i=1}^n \phi_{\alpha_i}(\xi_i) \quad (8)$$

where  $\alpha = [\alpha_1, \dots, \alpha_n]$  is a sequence of  $n$  non-negative integers and  $\phi_{\alpha_i}(\xi_i)$  are one-dimensional Hermite polynomials. More details on the one-dimensional Hermite polynomials are given in Appendix 2.

For the determination of the PCE unknown coefficients, a non-intrusive technique (in which the deterministic model is treated as a black-box) is used (see [Ahm12] among others). In this chapter, the regression approach is employed. In this approach, it is required to compute the system response at a set of collocation points in order to perform a fit of the PCE using the obtained system response values.

As suggested by several authors (e.g. [Hua09]), the collocation points can be chosen as the result of all possible combinations of the roots of the one-dimensional Hermite polynomial of order  $(p+1)$  for each random variable. For example, if a PCE of order  $p=2$  is used to approximate the response surface of a system with  $n=2$  random variables, the roots of the one-dimensional Hermite Polynomial of order 3 are chosen for each random variable. These roots are  $(-\sqrt{3}, 0, \sqrt{3})$  for the first random variable and  $(-\sqrt{3}, 0, \sqrt{3})$  for the second random variable. In this case, 9 collocation points are available. These collocation points are  $(-\sqrt{3}, -\sqrt{3}), (-\sqrt{3}, 0), (-\sqrt{3}, \sqrt{3}), (0, -\sqrt{3}), (0, 0), (0, \sqrt{3}), (\sqrt{3}, -\sqrt{3}), (\sqrt{3}, 0), (\sqrt{3}, \sqrt{3})$ . In the general case, for a PCE of order  $p$  and for  $n$  random variables, the number  $N$  of the available collocation points can be obtained using the following formula:

$$N=(p+1)^n \quad (9)$$

Referring to Equations (7 and 9), one can observe that the number of the available collocation points is higher than the number of the unknown coefficients. This leads to a linear system of equations whose number  $N$  of equations is greater than the number  $P$  of the unknown coefficients. The regression approach is used to solve this system. This approach is based on a least square minimization between the exact solution  $\Gamma$  and the approximate solution  $\Gamma_{PCE}$  which is based on the PCE. Accordingly, the unknown coefficients of the PCE can be computed using the following equation:

$$\mathbf{a} = (\mathbf{\Psi}^T \mathbf{\Psi})^{-1} \cdot \mathbf{\Psi}^T \cdot \mathbf{\Gamma} \quad (10)$$

in which  $\mathbf{a}$  is a vector containing the PCE coefficients,  $\mathbf{\Gamma}$  is a vector containing the system response values as calculated by the deterministic model at the different collocation points and  $\mathbf{\Psi}$  is a matrix of size  $N \times P$  whose elements are the multivariate Hermite polynomials. It is given as follows:

$$\mathbf{\Psi} = \begin{bmatrix} \psi_0^1(\xi) & \psi_1^1(\xi) & \psi_2^1(\xi) & \dots & \psi_{p-1}^1(\xi) \\ \psi_0^2(\xi) & \psi_1^2(\xi) & \psi_2^2(\xi) & \dots & \psi_{p-1}^2(\xi) \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ \psi_0^N(\xi) & \psi_1^N(\xi) & \psi_2^N(\xi) & \dots & \psi_{p-1}^N(\xi) \end{bmatrix} \quad (11)$$

Notice that in order to calculate the system response corresponding to a given collocation point, the standard normal random variables  $\xi_i$  should be expressed in the original physical space of random variables as follows:

$$x_i = F_{x_i}^{-1} [\Phi(\xi_i)] \quad (12)$$

in which,  $x_i$  is a physical random variable,  $F_{x_i}$  is the CDF of the physical random variable and  $\Phi$  is the CDF of the standard normal random variable. Notice also that if the original physical random variables are correlated, the standard normal random variables should first be correlated using the following equation:

$$\begin{bmatrix} \xi_{1c} \\ \xi_{2c} \\ \cdot \\ \cdot \\ \cdot \\ \xi_{nc} \end{bmatrix} = H \cdot \begin{bmatrix} \xi_1 \\ \xi_2 \\ \cdot \\ \cdot \\ \cdot \\ \xi_n \end{bmatrix} \quad (13)$$

in which  $\{\xi_{1c}, \xi_{2c}, \dots, \xi_{nc}\}$  is the vector of correlated standard normal random variables,  $\{\xi_1, \xi_2, \dots, \xi_n\}$  is the vector of uncorrelated standard normal random variables and  $H$  is the Cholesky transformation of the correlation matrix of the physical random variables.

Once the PCE coefficients are determined, MCS can be applied on the obtained PCE (called meta-model) to compute the PDF of the system response. This is achieved by (i) generating a large number of realizations of the vector  $(\xi_1, \xi_2, \dots, \xi_n)$  of standard normal random variables and (ii) calculating the system response corresponding to each realization by substituting the vector  $(\xi_1, \xi_2, \dots, \xi_n)$  in the meta-model.

### 3.1. Optimal number of collocation points

The number of the available collocation points significantly increases with the increase in the number of random variables (cf. Eq. 9) and it may be very large with respect to the number of the unknown PCE coefficients. This makes it necessary to determine the optimal number of collocation points which is needed by the regression approach to solve the linear system of equations (Eq. 10). Sudret [Sud08] has proposed to successively increase the information matrix  $\mathbf{A}$  [where  $\mathbf{A} = (\Psi^T \Psi)$ ] until it becomes invertible as follows: (a) the  $N$  available collocation points are ordered in a list according to increasing norm, (b) the information matrix  $\mathbf{A}$  is constructed using the first  $P$  collocation points of the ordered list, i.e. the  $P$  collocation points that are the closest ones to the origin of the standard space of random variables and finally

(c) this matrix is successively increased (by adding each time the next collocation point from the ordered list) until it becomes invertible.

### 3.2. Accuracy of the obtained PCE

For a given PCE order, the accuracy of the approximation of the system response by a PCE can be measured by the coefficient of determination. Two types of coefficients of determination exist in literature. These are the coefficient of determination  $R^2$  and the *leave-one-out* coefficient of determination  $Q^2$ .

Let us consider  $J$  realizations of the standard normal random vector  $\xi$  as follows:  $\{\xi^{(1)} = (\xi_1^{(1)}, \dots, \xi_n^{(1)}), \dots, \xi^{(J)} = (\xi_1^{(J)}, \dots, \xi_n^{(J)})\}$ , and let us assume that the vector  $\Gamma = \{\Gamma(\xi^{(1)}), \dots, \Gamma(\xi^{(J)})\}$  includes the corresponding values of the system response determined by deterministic calculations. The coefficient of determination  $R^2$  is calculated as follows:

$$R^2 = 1 - \Delta_{PCE} \quad (14)$$

where  $\Delta_{PCE}$  is given by:

$$\Delta_{PCE} = \frac{(1/J) \sum_{i=1}^J [\Gamma(\xi^{(i)}) - \Gamma_{PCE}(\xi^{(i)})]^2}{Var(\Gamma)} \quad (15)$$

and

$$Var(\Gamma) = \frac{1}{J-1} \sum_{i=1}^J [\Gamma(\xi^{(i)}) - \bar{\Gamma}]^2 \quad (16)$$

Note here that  $J$  is the number of collocation points used to evaluate the unknown coefficients of the PCE. The value  $R^2 = 1$  indicates a perfect approximation of the true system response  $\Gamma$ , whereas  $R^2 = 0$  indicates a nonlinear relationship between the true model  $\Gamma$  and the PCE model  $\Gamma_{PCE}$ .

The coefficient of determination  $R^2$  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 collocation points. As a consequence, a more reliable and rigorous coefficient of determination, called the *leave-one-out* coefficient of determination, was proposed in literature. This coefficient of determination consists in sequentially removing a point from the  $J$  collocation points. Let  $\Gamma_{\xi_i}$  be the meta-model that has been built from  $(J-1)$  collocation points after removing the  $i^{th}$  observation from these collocation points and let  $\Delta^i = \Gamma(\xi^{(i)}) - \Gamma_{\xi_i}(\xi^{(i)})$  be the predicted residual between the model evaluation at

point  $\zeta^{(i)}$  and its prediction at the same point based on  $\Gamma_{\zeta^{(i)}}$ . The empirical error is thus given as follows:

$$\Delta_{PCE}^* = \frac{1}{J} \sum_{i=1}^J (\Delta^i)^2 \quad (17)$$

The corresponding coefficient of determination is often denoted by  $Q^2$  and is called *leave-one-out* coefficient of determination. It is given as follows:

$$Q^2 = 1 - \frac{\Delta_{PCE}^*}{Var(\Gamma)} \quad (18)$$

### 3.3. PCE-based Sobol indices

A Sobol index of a given input random variable is a measure by which the contribution of this input random variable to the variability of the system response can be determined. Sobol indices are generally calculated by MCS methodology. This method is very time-expensive especially when dealing with a large number of random variables. [Sud08] proposed an efficient approach to calculate the Sobol indices based on the coefficients of the PCE. This method is based on ranking the different terms of the PCE and gathering them into groups where each group contains only one random variable or a combination of random variables. For more details on the computation of Sobol indices, the reader may refer to [Ahm12] among others.

## 4. Conclusion

This chapter first presented the subset simulation approach which is an efficient alternative to *MCS* for the computation of a small failure probability. An example application was provided. It aims at showing the practical implementation of the *SS* approach. It was found that for a prescribed accuracy, the *SS* approach significantly reduces the number of realizations as compared to Monte Carlo simulations methodology (the reduction was found equal to 93.3% in the present chapter). In other words, for the same computational effort, the *SS* approach provides a smaller value of the coefficient of variation of  $P_f$  than *MCS*. It should be mentioned that the Matlab code used for the example application is provided in <http://www.univ-nantes.fr/soubra-ah> for practical use.

In a second stage, the Polynomial Chaos Expansion methodology was presented. It was shown that the PCE method replaces the computationally-expensive deterministic model by a meta-model (i.e. a simple analytical equation). Once the PCE coefficients are determined, MCS can be applied on the obtained PCE to easily compute the PDF of the system response with a quasi-negligible computation time.

## APPENDIX 1

### Modified METROPOLIS-HASTINGS algorithm

The Metropolis–Hastings algorithm is a Markov chain Monte Carlo (*MCMC*) method. It is used to generate a sequence of new realizations from existing realizations (that follow a target *PDF* called ' $P_t$ '). Refer to Figure (1) and let  $s_k \in F_j$  be a current realization which follows a target *PDF* ' $P_t$ '. Using a proposal *PDF* ' $P_p$ ', a next realization  $s_{k+1} \in F_j$  that follows the target *PDF* ' $P_t$ ' can be simulated from the current realization  $s_k$  as follows:

- a. A candidate realization  $\hat{s}$  is generated using the proposal *PDF* ( $P_p$ ). The candidate realization  $\hat{s}$  is centered at the current realization  $s_k$ .
- b. Using the deterministic model, evaluate the value of the performance function  $G(\hat{s})$  corresponding to the candidate realization  $\hat{s}$ . If  $G(\hat{s}) < C_j$  (i.e.  $\hat{s}$  is located in the failure region  $F_j$ ), set  $s_{k+1} = \hat{s}$ ; otherwise, reject  $\hat{s}$  and set  $s_{k+1} = s_k$  (i.e. the current realization  $s_k$  is repeated).
- c. If  $G(\hat{s}) < C_j$  in the preceding step, calculate the ratio  $r_1 = P_t(\hat{s})/P_t(s_k)$  and the ratio  $r_2 = P_p(s_k|\hat{s})/P_p(\hat{s}|s_k)$ , then compute the value  $r = r_1 r_2$ .
- d. If  $r \geq 1$  (i.e.  $\hat{s}$  is distributed according to the  $P_t$ ), one continues to retain the realization  $s_{k+1}$  obtained in step b; otherwise, reject  $\hat{s}$  and set  $s_{k+1} = s_k$  (i.e. the current realization  $s_k$  is repeated).

Notice that in step *b*, if the candidate realization  $\hat{s}$  does not satisfy the condition  $G(\hat{s}) < C_j$ , it is rejected and the current realization  $s_k$  is repeated. Also in step *d*, if the candidate realization  $\hat{s}$  does not satisfy the condition  $r \geq 1$  (i.e.  $\hat{s}$  is not distributed according to the  $P_t$ ), it is rejected and the current realization  $s_k$  is repeated. The presence of several repeated realizations is not desired as it leads to high probability that the chain of realizations remains in the current state. This means that there is high probability that the next failure threshold  $C_{j+1}$  is equal to the current failure threshold  $C_j$ . This decreases the efficiency of the subset simulation approach. To overcome this inconvenience, Santoso *et al.* (2011) proposed to modify the classical M-H algorithm as follows:

- a. A candidate realization  $\hat{s}$  is generated using the proposal ( $P_p$ ). The candidate realization  $\hat{s}$  is centered at the current realization  $s_k$ .
- b. Calculate the ratio  $r_1 = P_t(\hat{s})/P_t(s_k)$  and the ratio  $r_2 = P_p(s_k|\hat{s})/P_p(\hat{s}|s_k)$ , then compute the value  $r = r_1 r_2$ .
- c. If  $r \geq 1$ , set  $s_{k+1} = \hat{s}$ ; otherwise, another candidate realization is generated. Candidate realizations are generated randomly until the condition  $r \geq 1$  is satisfied.
- d. Using the deterministic model, evaluate the value of the performance function  $G(s_{k+1})$  of the candidate realization that satisfies the condition  $r \geq 1$ . If  $G(s_{k+1}) < C_j$  (i.e.  $s_{k+1}$  is located in the failure region  $F_j$ ), one continues

to retain the realization  $s_{k+1}$  obtained in step  $c$ ; otherwise, reject  $\hat{s}$  and set  $s_{k+1} = s_k$  (i.e. the current realization  $s_k$  is repeated).

These modifications reduce the repeated realizations and allow one to avoid the computation of the system response of the rejected realizations. This becomes of great importance when the time cost for the computation of the system response is expensive (i.e. for the finite element or finite difference models).

## APPENDIX 2

The one-dimensional Hermite polynomials of orders 0, 1, 2, 3, ...,  $p+1$  are given by:

$$\phi_0(\xi) = 1$$

$$\phi_1(\xi) = \xi$$

$$\phi_2(\xi) = \xi^2 - 1$$

$$\phi_3(\xi) = \xi^3 - 3\xi$$

$$\phi_{p+1}(\xi) = \xi \phi_p(\xi) - p\phi_{p-1}(\xi)$$

where  $\xi$  is a standard normal random variable.

## References

- [Ahm12] A. Ahmed. Simplified and advanced approaches for the probabilistic analysis of shallow foundations. PhD thesis, University of Nantes, 2012.
- [Au01] S.K. Au and J.L. Beck. Estimation of small failure probabilities in high dimensions by subset simulation, *Journal of Probabilistic Engineering Mechanics*, 16: 263-277, 2001.
- [Hua09] H. Huang, B. Liang and KK Phoon. Geotechnical probabilistic analysis by collocation-based stochastic response surface method: An EXCEL add-in implementation. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, 3(2):75-86.
- [Pho08] K.K. Phoon. *Reliability-based design in geotechnical engineering: Computations and applications*, Taylor & Francis, 1-75, 2008.
- [San11] A.M. Santoso, K.K. Phoon and S.T. Quek. Modified Metropolis-Hastings algorithm with reduced chain-correlation for efficient subset simulation. *Probabilistic Engineering Mechanics*, 26(2): 331-341, 2011.
- [Sud08] B. Sudret. Global sensitivity analysis using polynomial chaos expansion. *Reliability Engineering and System Safety*, 93: 964-979.