



Extension of subset simulation approach for uncertainty propagation and global sensitivity analysis

Ashraf Abd El-Modby Soliman Ahmed, Abdul-Hamid Soubra

► To cite this version:

Ashraf Abd El-Modby Soliman Ahmed, Abdul-Hamid Soubra. Extension of subset simulation approach for uncertainty propagation and global sensitivity analysis. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, 2012, 6 (3), pp.162-176. 10.1080/17499518.2012.656296 . hal-01006885

HAL Id: hal-01006885

<https://hal.science/hal-01006885>

Submitted on 12 Dec 2018

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.

Extension of subset simulation approach for uncertainty propagation and global sensitivity analysis

Ashraf Ahmed and Abdul-Hamid Soubra*

University of Nantes, Saint-Nazaire, France

The subset simulation (SS) method is a probabilistic approach which is devoted to efficiently calculating a small failure probability. Contrary to Monte Carlo Simulation (MCS) methodology which is very time-expensive when evaluating a small failure probability, the SS method has the advantage of assessing the small failure probability in a much shorter time. However, this approach does not provide any information about the probability density function (PDF) of the system response. In addition, it does not provide any information about the contribution of each input uncertain parameter in the variability of this response. Finally, the SS approach cannot be used to calculate the partial safety factors which are generally obtained from a reliability analysis. To overcome these shortcomings, the SS approach is combined herein with the Collocation-based Stochastic Response Surface Method (CSRSM) to compute these outputs. This combination is carried out by using the different values of the system response obtained by the SS approach for the determination of the unknown coefficients of the polynomial chaos expansion in CSRSM. An example problem that involves the computation of the ultimate bearing capacity of a strip footing is presented to demonstrate the efficiency of the proposed procedure. The validation of the present method is performed by comparison with MCS methodology applied on the original deterministic model. Finally, a probabilistic parametric study is presented and discussed.

Keywords: subset simulation; polynomial chaos expansion; strip footing

1. Introduction

Monte Carlo Simulation (MCS) methodology is well known to be a rigorous and robust tool to calculate the failure probability (P_f) of a mechanical system. However, it becomes very time-consuming when computing a small failure probability. This is due to the large number of calls of the deterministic model required in such a case. As an alternative to MCS methodology, the subset simulation (SS) approach was proposed by Au and Beck (2001) to calculate the small failure probability using a much smaller number of calls of the deterministic model. In this approach, the failure probability is expressed as a product of conditional probabilities of some chosen intermediate failure events. Thus, the problem of evaluating a small failure probability in the original probability space is replaced by a sequence of more frequent events in the conditional probability space.

It should be mentioned here that the SS approach is efficient in computing the failure probability, but it does not provide any information about the probability density function (PDF) of the system response (i.e. it does not allow one to perform an uncertainty propagation from the input variables to the system output). Also, it does not provide any information

about the contribution of each input uncertain parameter to the variability of the system response (i.e. it does not allow one to perform a global sensitivity analysis). In addition, the SS approach does not allow one to calculate the design point (the most probable failure point) which is obtained during a reliability-based analysis. Notice that the computation of the design point has an important practical implication since it provides the partial safety factor corresponding to each input random variable. To overcome the above-mentioned shortcomings, the SS approach is combined herein with the Collocation-based Stochastic Response Surface Method (CSRSM). The procedure of combining SS approach with CSRSM is illustrated in this article through an example problem. In this example, a probabilistic analysis at the ultimate limit state (ULS) of a strip footing resting on a (c, ϕ) soil and subjected to a central vertical load P_s is performed. First, a classical SS computation is performed to calculate the failure probability. Then, the deterministic system responses obtained during this computation are used in a CSRSM analysis with no additional cost to obtain the other outputs cited earlier (i.e. uncertainty propagation, global sensitivity analysis and reliability-based analysis).

*Corresponding author. Email: Abed.Soubra@univ-nantes.fr

In this article, the uncertainties of the soil shear strength parameters (c and ϕ) are taken into account by considering c and ϕ as random variables. The system response adopted in the analysis is the ultimate bearing capacity (q_u). In order to be able to perform a validation of the present approach with MCS methodology applied on the original deterministic model, an analytical deterministic model is used in this study for the computation of the system response. This model is based on the kinematic approach of the limit analysis theory.

The present article is organised as follows: a brief review of both the SS approach and the CSRSM is presented. Then, the combination of the SS approach with CSRSM is described. This is followed by the example problem to illustrate the efficiency of the proposed procedure.

2. Review of the SS approach

Subset simulation (SS) was proposed by Au and Beck (2001) to compute small failure probabilities. This method was used by several authors (Au and Beck 2003, Schuëller *et al.* 2004, Au *et al.* 2007, 2010 among others) and was found to be an efficient tool compared to MCS methodology. The basic idea of SS is that the small failure probability can be expressed as a product of larger conditional failure probabilities. Consider a failure region F defined by the condition $G < 0$ where G is the performance function and let $s_1, \dots, s_k, \dots, s_{N_t}$ be a sample of N_t realisations of a vector ' s ' where ' s ' is composed of M random variables. It is possible to define a sequence of nested failure regions $F_1, \dots, F_j, \dots, F_m$ of decreasing size where $F_1 \supset \dots \supset F_j \supset \dots \supset F_m = F$ (Figure 1). An intermediate failure region F_j can be defined by

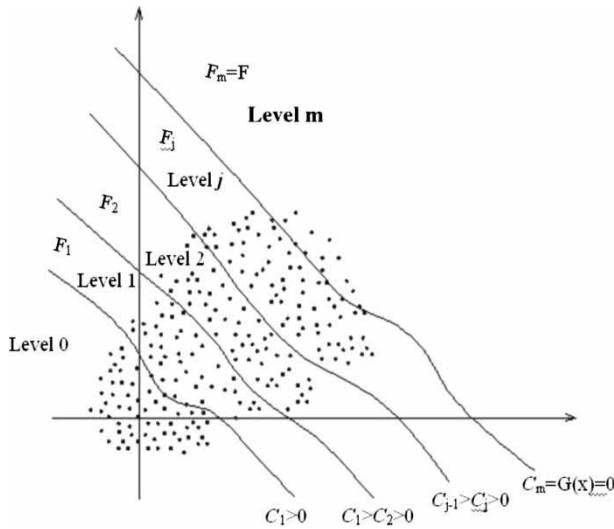


Figure 1. Nested failure domain.

$G < C_j$, where C_j is an intermediate failure threshold whose value is larger than zero. Thus, there is a decreasing sequence of positive failure thresholds $C_1, \dots, C_j, \dots, C_m$ corresponding respectively to $F_1, \dots, F_j, \dots, F_m$, where $C_1 > \dots > C_j > \dots > C_m = 0$. In the SS approach, the space of uncertain parameters is divided into a number m of levels with equal number N_s of realisations ($s_1, \dots, s_k, \dots, s_{N_s}$) in each level, so that the total number N_t of realisations is $N_t = m \times N_s$. An intermediate level j contains a safe region and a failure region defined with respect to a given failure threshold C_j . The failure probability corresponding to this intermediate level j is calculated as follows:

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

where $I_{F_j}(s_k) = 1$ if $s_k \in F_j$ and $I_{F_j}(s_k) = 0$ otherwise. Notice that in the SS approach, the first N_s realisations are generated using MCS methodology according to a target PDF (P_t). The next realisations of the subsequent levels are obtained using the Markov chain method based on a modified Metropolis-Hastings (M-H) algorithm by Santoso *et al.* (2010) using a proposal PDF (P_p). The modified M-H algorithm by Santoso *et al.* (2010) is explained in some detail in Appendix 1. The failure probability $P(F) = P(F_m)$ of the failure region F can be calculated from the sequence of conditional failure probabilities as follows:

$$\begin{aligned} P(F) &= P(F_m) \\ &= P(F_m|F_{m-1}) \times P(F_{m-1}|F_{m-2}) \times P(F_{m-2}|F_{m-3}) \\ &\quad \times \dots \times P(F_2|F_1) \times P(F_1) \end{aligned} \quad (2)$$

It should be noted here that the computation of the failure probability $P(F)$ may be determined using alternatively one of the two following procedures: The first procedure consists firstly in prescribing a sequence of failure thresholds $C_1, \dots, C_j, \dots, C_m$ so that $C_1 > \dots > C_j > \dots > C_m = 0$ and then, calculating the different values of $P(F_j|F_{j-1})$ for the different levels using Equation (1). The second procedure consists first in prescribing a constant conditional failure probability $P(F_j|F_{j-1})$ for the different levels and then, in calculating the different C_j values corresponding to these levels. In the present article, the second procedure is used. Notice that, for simplicity in notations, the constant conditional failure probability will be referred to as p_0 in the following sections.

The algorithm of SS approach can be described by the following steps:

- (1) Generate a realisation of the vector 's' of random variables by MCS according to the target PDF (P_t).
- (2) Using the deterministic model, calculate the system response corresponding to this realisation.
- (3) Repeat steps 1 and 2 until obtaining a prescribed number N_s of realisations of the vector 's' and the corresponding system response values. Then, evaluate the corresponding values of the performance function to obtain the vector $\mathbf{G}_0 = \{G_0^1, \dots, G_0^k, \dots, G_0^{N_s}\}$. Notice that the subscripts '0' refer to the first level (level 0) of SS.
- (4) Prescribe a constant failure probability p_0 for all the failure regions F_j and evaluate the first failure threshold C_1 which corresponds to the failure region F_1 where C_1 is equal to the $[(N_s \times p_0) + 1]$ th value in the increasing list of elements of the vector \mathbf{G}_0 . This ensures that the value of $P(F_1)$ will be equal to the prescribed value of p_0 , i.e. the ratio between the number of realisations corresponding to $G < C_1$ and the number of realisations N_s generated at level 0 is equal to the prescribed value p_0 .
- (5) Among the N_s realisations, 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 realisations are used as 'mother realisations' to generate additional $[(1-p_0)N_s]$ realisations of the vector 's' using Markov chain method based on Metropolis-Hastings algorithm. These new realisations are located in the second level (level 1 in Figure 1).
- (6) The values of the performance function corresponding to the realisations of level 1 are listed in an increasing order and are gathered in the vector of performance functions $\mathbf{G}_1 = \{G_1^1, \dots, G_1^k, \dots, G_1^{N_s}\}$.
- (7) Evaluate the second failure threshold C_2 as the $[(N_s \times p_0) + 1]$ th value in the increasing list of the vector \mathbf{G}_1 .
- (8) Repeat steps 5–7 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:

$$P(F_m|F_{m-1}) = \frac{1}{N_s} \sum_{k=1}^{N_s} I_{F_m}(s_k) \quad (3)$$

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

- (9) Finally, the failure probability $P(F)$ is evaluated according to Equation (2).

It should be mentioned that, in the present study, a Gaussian PDF was used as a target PDF P_t (i.e. it was used to generate the N_s realisations of the first level of SS). Concerning the type of the proposal PDF P_p (which is used to generate the realisations of levels $1, \dots, j, \dots, m-1$), it does not affect the efficiency of the SS approach, and hence, any PDF which can be operated easily can be used as a proposal PDF (see Au and Beck 2001). In this article, similar to Au and Beck (2001, 2003) and Santoso et al. (2010), a uniform PDF was used. The conditional failure probability p_0 was chosen to be equal to 0.1. Notice that the value of p_0 affects the number m of levels required to reach the limit state surface $G=0$. However, it has a very small effect on the total number of realisations N_t (which is a multiple of the number of levels i.e. $N_t = m \times N_s$) required to reach this limit state surface. Indeed, if p_0 is large, the sequence of failure thresholds $C_1, \dots, C_j, \dots, C_m$ will decrease slowly and a large number of levels will be required to reach the limit state surface. In this case, a small number N_s of realisations per level is required to achieve a target high accuracy of $P(F)$. On the contrary, if p_0 is small, the sequence of failure thresholds will reach the limit state surface quickly and a small number of levels will be required. In this case, a large number N_s of realisations per level will be required to attain the same target accuracy of $P(F)$. As a conclusion, an arbitrary value of p_0 can be considered for the probabilistic analysis with a small effect on the computational time.

3. Review of the CSRSRM

CSRSRM involves the propagation of the uncertainties of the input parameters through a computational model to arrive at a random output (Isukapalli *et al.* 1998, Huang *et al.* 2009, Li *et al.* 2011, Mollon *et al.* 2011, Houmadi *et al.* 2012, Mao *et al.* 2012). Thus, this method allows one to determine the PDF of the system response by taking into account the uncertainties of the input parameters *via* their PDFs.

The basic idea of CSRSRM is to approximate the system response by a polynomial chaos expansion (PCE) of a suitable order p (i.e. it replaces the complex finite difference or finite element model by a meta-model). This approximation is carried out in this article in the standard normal space of uncorrelated random variables. Thus, the random input variables (which may be non-normal and/or correlated) should be represented in terms of uncorrelated standard normal random variables. Consequently,

the system response Y may be approximated by Y_{app} as follows:

$$Y \approx Y_{\text{app}} = \sum_{i=0}^{P-1} a_i \Psi_i(\xi) \quad (4)$$

where a_i are the unknown coefficients to be evaluated, P is the size of the expansion (which is equal to the number of the PCE coefficients), ξ is a vector of M standard uncorrelated random variables where M is the number of the uncertain parameters and Ψ_i are multivariate Hermite polynomials. The multivariate Hermite polynomials are given in Mollon *et al.* (2011), among others. The size P of the PCE is given by:

$$P = \frac{(M+p)!}{M!p!} \quad (5)$$

where p is the PCE order. The PCE coefficients can be determined either by a regression or by a projection approach. In this work, a regression approach is used. This approach requires the evaluation of the system response at a given number of collocation points. These collocation points are commonly chosen as the 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 of a system with $M=2$ 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 variable and $(-\sqrt{3}, 0, \sqrt{3})$ for the second variable. In this case, 9 collocation points are available. These available collocation points are $(-\sqrt{3}, -\sqrt{3})$, $(-\sqrt{3}, 0)$, $(-\sqrt{3}, \sqrt{3})$, $(0, -\sqrt{3})$, $(0, 0)$, $(0, \sqrt{3})$, $(\sqrt{3}, -\sqrt{3})$, $(\sqrt{3}, 0)$ and $(\sqrt{3}, \sqrt{3})$. The number N of the available collocation points can be obtained using the following formula:

$$N = (p+1)^M \quad (6)$$

The CSRSM may be summarised by the following steps:

- (1) For a prescribed order p of the PCE, determine the pattern (number and positions) of the available collocation points in the standard space of random variables.
- (2) Transform the collocation points from the standard to the physical space where the input random variables may be non-normal. This can

be carried out by using iso-probabilistic transformation as follows:

$$X = F_X(\Phi(\xi)) \quad (7)$$

where F_X is the Cumulative Density Function (CDF) of the physical random variable X and $\Phi(\xi)$ is the standard Gaussian CDF. Notice that if the physical random variables are correlated, the standard normal random variables should be first correlated using the Cholesky transformation as follows:

$$\begin{bmatrix} \xi_{1c} \\ \xi_{2c} \\ \vdots \\ \xi_{Mc} \end{bmatrix} = \mathbf{H} \cdot \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_M \end{bmatrix} \quad (8)$$

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

- (3) Compute the system response corresponding to each collocation point in the physical space using the deterministic mechanical model.
- (4) Determine the values of the PCE coefficients by regression. Notice that the number N of the available collocation points is generally greater than the number P of the unknown PCE coefficients. This leads to a linear system of equations whose number of equations N is greater than the number P of unknowns. The unknown coefficients of the PCE can be computed by regression using the following equation:

$$\mathbf{a} = (\Psi^T \Psi)^{-1} \cdot \Psi^T \cdot \mathbf{Y} \quad (9)$$

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

$$\Psi = \begin{bmatrix} \Psi_0^1(\xi) & \Psi_1^1(\xi) & \Psi_2^1(\xi) & \cdots & \Psi_{p-1}^1(\xi) \\ \Psi_0^2(\xi) & \Psi_1^2(\xi) & \Psi_2^2(\xi) & \cdots & \Psi_{p-1}^2(\xi) \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ \Psi_0^N(\xi) & \Psi_1^N(\xi) & \Psi_2^N(\xi) & \cdots & \Psi_{p-1}^N(\xi) \end{bmatrix} \quad (10)$$

- (5) The PDF of the approximated system response can be computed by applying MCS method on the PCE determined previously. This is achieved by (i) generating a large number of realisations of the vector $(\xi_1, \xi_2, \dots, \xi_M)$ of standard normal random variables and (ii) calculating the system response corresponding to each realisation by substituting the corresponding vector $(\xi_1, \xi_2, \dots, \xi_M)$ in the PCE. Thus, one obtains a vector of system response values which can be used to provide the PDF of this system response.

Finally, notice that although CSRSM is mainly devoted to the computation of the PDF of the system response, the determination of the PCE in CSRSM has other advantages. One may cite among others:

- (1) The PCE coefficients allow one to easily perform a global sensitivity analysis using the PCE-based Sobol indices (cf. Sudret 2008, Mollon *et al.* 2011). The computation of the Sobol indices is important since it provides the contribution of each random variable to the variability of the system response. This helps the engineers in detecting the uncertain input parameters which have a significant weight in the variability of the system response.
- (2) The obtained PCE allows one to easily perform a reliability-based analysis and a reliability-based design (RBD) since the deterministic model is now described by a simple analytical equation (i.e. the PCE).

4. Extension of the SS approach

This study is devoted to employing the SS approach to compute not only the failure probability but also the PDF of the system response and the corresponding statistical moments without an additional cost. This aim can be achieved by combining the SS approach with CSRSM.

It should be remembered here that in CSRSM, the PCE coefficients are determined by using the values of the system response calculated at several collocation points. Although the roots of the one-dimensional Hermite polynomials are generally used for the determination of the collocation points (Isukapalli *et al.* 1998, Huang *et al.* 2009, Li *et al.* 2011, Mollon *et al.* 2011), this technique is not mandatory. In this article, the combination between SS approach and CSRSM is carried out by using the realisations calculated by SS approach as collocation points to construct the PCE in CSRSM. Notice that the values of the system response corresponding to these realisa-

tions are already calculated and thus, the computation of the PCE coefficients requires no additional calls of the deterministic model.

Once the PCE coefficients are determined, the MCS methodology is applied on the obtained PCE. This allows one to obtain the PDF of the system response. It should be emphasised here that the application of MCS on the PCE requires a very small computation time (since the system response is now represented by an analytical formula); hence, it does not affect the computation efficiency. In addition to computing the PDF of the system response, the proposed procedure has four other advantages:

- (1) The computed PCE coefficients can be used to perform a global sensitivity analysis *via* the PCE-based Sobol indices. It should be mentioned here that the classical computation of Sobol indices is based on MCS methodology (Sobol 2001). This makes it very time-consuming especially when dealing with a large number of random variables. In the present study, the PCE-based Sobol indices are used (e.g. Sudret 2008, Mollon *et al.* 2011). It should be emphasised here that the computation of the PCE-based Sobol indices does not require any additional call of the deterministic model. The PCE-based Sobol indices are computed using an analytical equation which makes use of the PCE coefficients. Although the computation of the PCE-based Sobol indices requires a small computation time for a small number of random variables, it becomes time-consuming in case of a large number of random variables. Even though, this time remains much smaller than that required by the classical approach based on MCS methodology.
- (2) Contrary to the SS approach, the procedure proposed in this study allows the computation of the failure probability for all the values of the failure threshold that are greater than the one considered in the SS analysis without the need to repeat the deterministic calculations (i.e. without an additional cost). This is because the limit state surfaces corresponding to larger values of the failure threshold are closer to the origin of the standard space of random variables and thus, they are included in the sampling zone of the SS methodology as will be seen later.
- (3) The obtained PCE allows one to perform a reliability-based analysis or a RBD. For the reliability-based analysis, the Hasofer-Lind reliability index and the corresponding design point can be easily determined since the PCE is

obtained in the standard uncorrelated space of the random variables. This is done by searching the point which is located on the limit state surface and has the minimum norm in the standard space of random variables. The design point is of great importance since it provides information about the partial safety factors of the different random variables. Concerning the RBD, the obtained PCE makes it easy to compute the dimension of the structure (e.g. the breadth of the footing) corresponding to a target reliability index.

- (4) The obtained PCE allows one to undertake a probabilistic parametric study to show the effect of the different characteristics of the random variables (e.g. coefficient of variation COV, coefficient of correlation ρ and the non-normality) on the PDF of the system response.

5. Example analysis

In this section, the efficiency of the proposed procedure to compute the different outputs cited above is illustrated through an example problem. In this example, a probabilistic analysis of a shallow strip footing of breadth $B = 2$ m resting on a (c, ϕ) soil and subjected to a central vertical load $P_s = 650$ kN/m (i.e. a vertical footing pressure $q_s = 325$ kN/m²) is performed. The analysis is carried out at the ULS. The system response is the ultimate bearing capacity q_u . The uncertain parameters considered in the analysis are the soil shear strength parameters c and ϕ . They are modelled as random variables. Two types of the PDFs are used for these random variables (normal and non-normal, as shown in Table 1). In the case of non-normal random variables, the soil cohesion was assumed to follow a log-normal PDF. However, the soil friction angle was assumed to follow a beta distribution with lower and upper bounds of 0 and 45°, respectively. Also, two cases of uncorrelated (i.e. $\rho_{c,\phi} = 0$) or correlated (i.e. $\rho_{c,\phi} = -0.5$) random variables were also considered in the analysis. The illustrative values used for the statistical parameters of these random variables are those commonly encountered in practice [Wolff (1985) and Phoon and Kulhawy (1999) among others] and they are presented in Table 1. These values will be referred to

hereafter as the reference values. The performance function used to calculate the failure probability is defined as follows:

$$G = (q_u/q_s) - 1 \quad (11)$$

The ultimate bearing capacity q_u is calculated using the deterministic model presented by Soubra (1999). This model is based on the theory of limit analysis. It will be briefly presented in the following subsection.

5.1. Deterministic model

The deterministic model is based on the upper-bound theorem of the limit analysis theory using a kinematically admissible failure mechanism. The approach is simple and self-consistent, and it obtains rigorous upper-bound solutions in the framework of the limit analysis theory. The failure mechanism used for the computation is a translational symmetrical multi-block mechanism (Figure 2). The bearing capacity is obtained by equating the total rate of work of the external forces to the total rate of energy dissipation along the lines of velocity discontinuities. The ultimate bearing capacity (in the absence of a surcharge loading on the ground surface) is given as follows:

$$q_u = \frac{1}{2} \gamma B N_\gamma + c N_c \quad (12)$$

in which B is the footing breadth, γ and c are the soil unit weight and cohesion and N_γ and N_c are the bearing capacity factors due to soil weight and cohesion, respectively. The coefficients N_γ and N_c are functions of the soil friction angle ϕ and the geometrical parameters of the failure mechanism shown in Figure 2. The ultimate bearing capacity of the foundation is obtained by minimisation of Equation (12) with respect to the mechanism's geometrical parameters. For further details on the failure mechanism, the reader can refer to Soubra (1999). It should be mentioned here that although the results given by this approach are upper-bound solutions, they are the smallest ones against the available results given by rigid block mechanisms. Notice also that the computation time of the ultimate bearing capacity is equal to about 0.2 minutes. The small computation time of this model admits the validation of the proposed procedure by comparing with the results given by MCS

Table 1. Statistical characteristics of the random variables.

Random variable	Mean	Coefficient of variation (%)	Type of the probability density function (PDF)	
			Case of normal PDFs	Case of non-normal PDFs
c	20 kPa	20	Normal	Log-normal
ϕ	30°	10	Normal	Beta

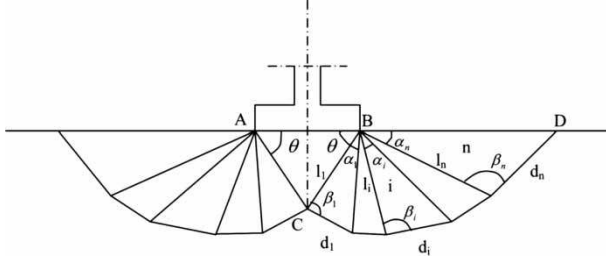


Figure 2. Failure mechanism for the ultimate bearing capacity analysis.

methodology applied on the original deterministic model, as will be shown in the next section. Finally, notice that the deterministic ultimate bearing capacity (i.e. the ultimate bearing capacity obtained using the mean values of c and ϕ given in Table 1) is equal to 1071.72 kN/m². Thus, for the adopted q_s value ($q_s = 325$ kN/m²), the punching safety factor $F_p = q_u/q_s$ is equal to 3.3.

5.2. Validation of the proposed procedure by comparison with MCS methodology

This section is devoted to the validation of the proposed procedure by comparison of the obtained results with those given by MCS methodology applied on the original deterministic model. The comparison involves the values of the failure probability together with the PDF and the statistical moments (mean, coefficient of variation, skewness and kurtosis) of the system response. Notice that normal and uncorrelated random variables are considered herein.

It should be noticed here that in addition to the number and position of the collocation points, the PCE order plays a key role in the accuracy of the approximation by a PCE. The optimal PCE order was chosen in this study. An optimal order was defined as the minimal order that leads to a coefficient of determination R^2 (see Appendix 2) greater than a prescribed value (say 0.9999). In fact, a prescribed small value of p (say $p = 2$) is firstly chosen and the corresponding value of R^2 is calculated. Secondly, p is successively increased until obtaining a sufficiently large value of R^2 (0.9999 in this paper). The numerical results have shown that a third-order PCE is necessary to satisfy the prescribed condition on the coefficient of determination (Table 2). Thus, this PCE order will be used in all subsequent probabilistic calculations performed in this article.

5.2.1. Validation in terms of failure probability (P_f)

To ensure a rigorous computation of the failure probability by SS approach, the number of realisations

Table 2. Coefficient of determination R^2 for PCEs of different orders.

PCE order	R^2
2	0.9971
3	0.9999
4	0.9999

of random variables N_s to be used per level must be sufficient to provide a small value of the coefficient of variation COV_{P_f} of this failure probability. Figure 3 shows the variation of the coefficient of variation of the failure probability computed by SS approach (COV_{P_f}) with the number of realisations N_s to be used per level. Notice that COV_{P_f} is estimated as follows:

$$[COV_{P_f}]^2 = \sum_{j=1}^m [COV_{p_0}^j]^2 \quad (13)$$

where $COV_{p_0}^j$ is the coefficient of variation at an intermediate level j . Details on the computation of $COV_{p_0}^j$ may be found in Au and Beck (2001, 2003). Figure 3 shows that COV_{P_f} decreases (i.e. the accuracy of the calculation increases) with the increase of N_s . It attains a small value (about 10%) when $N_s = 10,000$ realisations per level. Consequently, 10,000 realisations were considered at each level to calculate P_f by SS approach. The corresponding P_f value was found equal to 3.15×10^{-4} . Notice that four levels of SS approach were necessary to calculate this failure probability and thus, the total number of realisations required by SS approach is $N_t = 10,000 + (3 \times 9000) = 37,000$ realisations. It should be emphasised here that the high number of

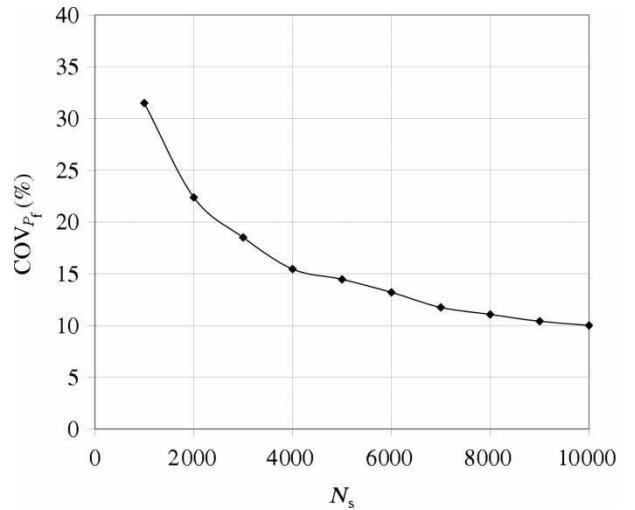


Figure 3. Coefficient of variation of P_f (calculated by SS approach) versus the number of samples N_s used per level.

realisations (i.e. 37,000 realisations) is due to the small value of COV_{P_f} considered in the computation. For practical implementations, a higher value of COV_{P_f} would be acceptable and thus, a smaller number of realisations would be required. For instance, if $N_s = 1000$ realisations per level, COV_{P_f} would be equal to 31.5% and P_f would be equal to 2.56×10^{-4} . This means that for $\text{COV}_{P_f} = 31.5\%$, the number of realisations is reduced by 90% with respect to the one corresponding to $\text{COV}_{P_f} = 10\%$; however, the difference in the P_f value is only 18.7%. The P_f value computed above (i.e. $P_f = 3.15 \times 10^{-4}$) is to be compared with the value of $P_f = 3.22 \times 10^{-4}$ computed by applying MCS methodology on the original deterministic model. One can observe that the two values are very close. Notice that 360,000 realisations were used to calculate the failure probability by applying MCS on the original deterministic model to attain the same COV_{P_f} as that of SS (i.e. about 10% as may be seen from Figure 4). This means that for the same accuracy, the number of calls of the deterministic model required by SS approach is reduced by 89.7% with respect to that required by MCS.

It should be noticed here that contrary to MCS methodology which can be used to compute the failure probabilities corresponding to different values of q_s without repeating the deterministic calculations, the SS approach allows one to calculate the failure probability corresponding to only one q_s value. If the failure probability corresponding to another q_s value is required, one needs to repeat all the deterministic calculations. The combination of SS approach and the CSRSM overcomes this shortcoming. This means that once the PCE (meta-model) is obtained, it

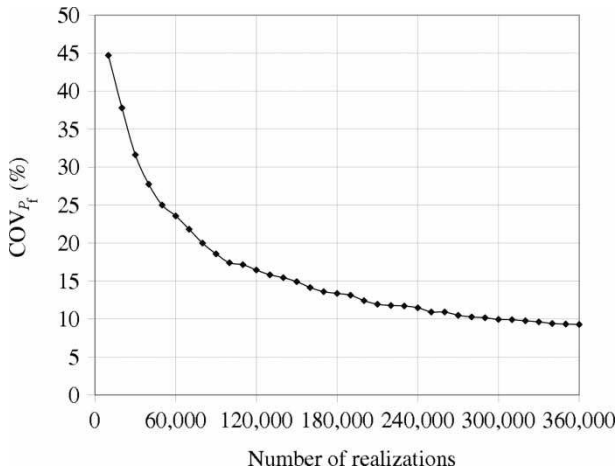


Figure 4. Coefficient of variation of P_f (calculated by applying MCS on the original deterministic model) versus the number of realisations.

can be used to accurately calculate the failure probability corresponding to any q_s value larger than the original one. The accuracy of this computation is ensured by the fact that all the limit state surfaces corresponding to larger q_s values are included in the sampling zone of the SS approach (see Figure 5). Table 3 presents a comparison between the failure probabilities computed by the proposed procedure and those calculated by MCS methodology applied on the original deterministic model for different q_s values. This table shows a good agreement between the two methods. This indicates that contrary to the SS approach, the procedure proposed in this study allows the computation of the failure probability for the values of the footing pressure that are greater than the one considered in the SS analysis.

5.2.2. Validation in terms of PDF

Once the PCE coefficients are determined, the uncertainty propagation can be performed. The PDF, CDF and the statistical moments of the system response can be easily determined by applying MCS methodology on the obtained PCE (meta-model). In order to validate these results, they were compared in Figure 6(a, b) and Table 4 with those obtained by applying MCS on the original deterministic model using (as before) 360,000 realisations of random variables. These results show that there is a good agreement between the proposed procedure and MCS methodology applied on the original deterministic model for both the central part and the distribution tail of the PDF of the system response.

As a conclusion, the proposed procedure allows one to rigorously determine not only the failure

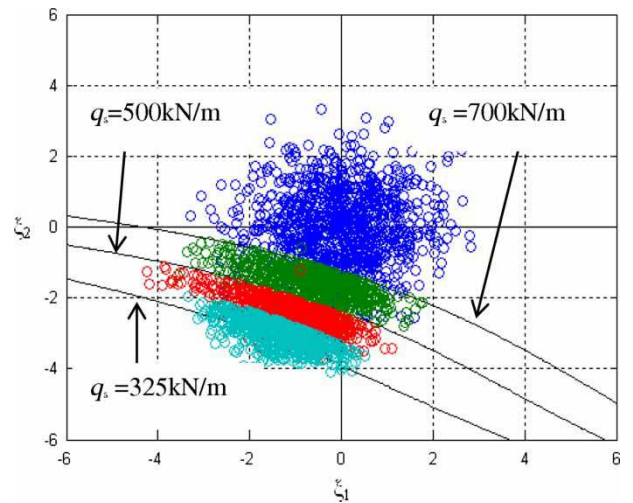


Figure 5. Limit state surfaces corresponding to different values of q_s plotted in the standard space of random variables.

Table 3. Comparison between the failure probability computed by applying MCS on the original deterministic model and that calculated by applying MCS on the meta-model for different values of the footing applied pressure q_s .

q_s (kN/m ²)	Failure probability	
	MCS applied on the original deterministic model	MCS applied on the meta-model
325	3.22×10^{-4}	3.15×10^{-4}
350	6.42×10^{-4}	5.68×10^{-4}
375	1.22×10^{-3}	1.10×10^{-3}
400	2.34×10^{-3}	2.10×10^{-3}
425	3.90×10^{-3}	3.70×10^{-3}
450	6.41×10^{-3}	6.16×10^{-3}
475	9.81×10^{-3}	9.70×10^{-3}
500	1.46×10^{-2}	1.44×10^{-2}
525	2.07×10^{-2}	2.07×10^{-2}
550	2.83×10^{-2}	2.88×10^{-2}
575	3.77×10^{-2}	3.86×10^{-2}
600	4.88×10^{-2}	4.93×10^{-2}
625	6.17×10^{-2}	6.28×10^{-2}
650	7.66×10^{-2}	7.73×10^{-2}
675	9.36×10^{-2}	9.42×10^{-2}
700	1.11×10^{-1}	1.12×10^{-1}

probability but also the statistical moments of the system response with no additional calls of the deterministic model. The next sections are devoted to the presentation of (1) a global sensitivity analysis; (2) a reliability-based analysis and design; and finally (3) a probabilistic parametric study.

5.3. Global sensitivity analysis via PCE-based Sobol indices

As mentioned previously, Sobol indices provide a measure of the contribution of each random variable to the variability of the system response. The Sobol indices of the soil cohesion S_c and the soil friction angle S_ϕ were calculated and were found to be equal to 0.1025 and 0.8975, respectively. This means that, for the statistical moments of the input uncertain parameters considered in this study, the soil friction angle has a significant weight in the variability of the

ultimate bearing capacity. However, the soil cohesion has a relatively small weight in the variability of this response.

5.4. Reliability-based analysis and design

The Hasofer-Lind reliability index β_{HL} is a means by which the safety of a given geotechnical system is measured. It represents the minimum distance between the origin and the limit state surface $G=0$ in the standard space of random variables. The computation of β_{HL} is performed by minimisation under the condition $G=0$, where G is approximated using the PCE. The point (c^*, ϕ^*) resulting from the minimisation is called the design point. It is the most probable failure point corresponding to a given q_s value.

Table 5 presents the reliability index, the design point and the corresponding partial safety factors ($F_c = \mu_c/c^*$, $F_\phi = \tan(\mu_\phi)/\tan\phi^*$) for different q_s values. This table also presents the punching safety factor $F_p = q_u/q_s$, where q_u is the deterministic ultimate bearing capacity (i.e. $q_u = 1071.72$ kN/m²). Remember that for all q_s values larger than the original one (i.e. $q_s = 325$ kN/m²), the reliability index is accurately calculated since the corresponding limit state surfaces are included in the sampling zone. This is ensured by the fact that the distance between the origin and the farthest collocation point in the standard space of random variables is $d_{\max} = 5.04$. This distance is larger than all values of β_{HL} presented in this table.

From Table 5, one can observe that the increase in q_s increases the values of c^* and ϕ^* at the design point. However, the reliability index and the partial safety factors F_c and F_ϕ decrease with the increase in q_s . This is to be expected since the increase in the footing pressure decreases the footing safety and thus provides smaller resistance factors. Notice that for the punching safety factor $F_p = 3$ which is generally used in practice, the corresponding partial safety factors F_c and F_ϕ are, respectively, 1.34 and 1.51. These values are somewhat close to those provided by Eurocode 7 where F_c and F_ϕ are, respectively, equal to 1.4 and 1.25. Notice finally that all the values of the reliability index, design point

Table 4. Comparison between the statistical moments of the ultimate bearing capacity computed by applying MCS on the original deterministic model and those computed by applying MCS on the meta-model.

	Mean value of q_u (kN/m ²)	Standard deviation of q_u (kN/m ²)	Coefficient of variation of q_u (%)	Skewness of q_u	Kurtosis of q_u
MCS applied on the original deterministic model	1150.50	419.50	36.46	1.06	1.82
MCS applied on the meta-model	1150.60	418.30	36.35	1.05	1.78

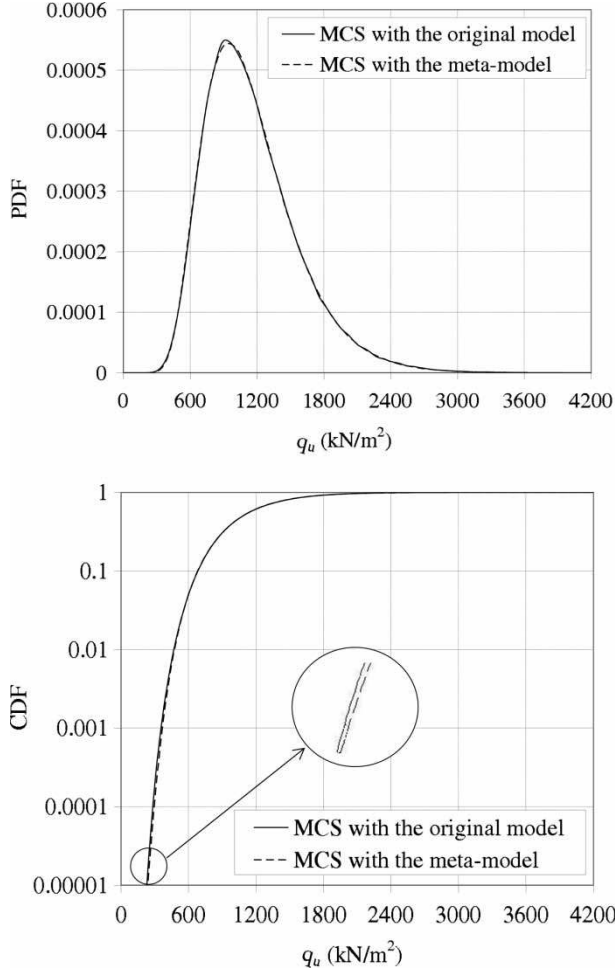


Figure 6. Comparison between the PDF and CDF of the ultimate bearing capacity computed by applying MCS on the original deterministic model and those computed by applying MCS on the meta-model.

and partial safety factors are calculated using only one SS computation (where $P_s = 650$ kN/m). It should be emphasised here that these results could not be obtained using the SS approach since this method does not provide an analytical expression of the system response (or the limit state surface). They are obtained with the use of SS approach combined with CSRSM.

Figure 7 presents two fragility curves in the normal and semi-log scales. These curves provide the variation of the failure probability with the allowable footing pressure q_s [where $q_s = P_u / (B \times F_p)$] when the random variables are non-normal and uncorrelated (the safety factor F_p was taken equal to 3 in this study). These curves can be used to perform either a RBD or a reliability-based analysis.

Concerning the RBD, if for example a strip footing is required to support a service load with a prescribed failure probability of 10^{-3} , from Figure 7(b), the allowable footing pressure is equal to $q_s = 125$ kN/m².

Consequently, the required footing breadth for a service load of 500 kN/m is $B = P_s / q_s = 500 / 125 = 4$ m. For the reliability-based analysis, Figure 7(b) provides the failure probability of a strip footing subjected to a given service load. For instance, if a footing of breadth $B = 2$ m is subjected to a service load of $P_s = 250$ kN/m (i.e. subjected to an allowable pressure of $q_s = 250 / 2 = 125$ kN/m²), the corresponding failure probability is equal to 10^{-3} .

Finally notice that only one SS calculation was performed to compute the fragility curves in Figure 7. This calculation corresponds to the smallest value of q_s . However; for larger q_s values, MCS methodology was applied on the obtained PCE to calculate the failure probability with no additional deterministic calculations. This demonstrates once again the interest of the extension of the SS approach.

5.5. Parametric study

The aim of this section is to investigate the effect of the statistical characteristics (coefficient of variation COV, coefficient of correlation ρ and the type of the PDF) of the random variables on the system response.

5.5.1. Effect of the coefficients of variation (COVs) of the random variables

This section presents the effect of COV_c and COV_ϕ on (1) the PDF of the system response (ultimate bearing capacity) and the corresponding statistical moments and (2) the Sobol indices S_c and S_ϕ . Notice

Table 5. Punching safety factor, reliability index, design point and partial safety factors for different values of the footing applied pressure q_s .

q_s (kN/m ²)	$F_p = q_u / q_s$	β_{HL}	c^* (kN/m ²)	ϕ^* (°)	F_c	F_ϕ
325	3.30	3.48	14.51	20.04	1.38	1.58
350	3.06	3.25	14.88	20.67	1.35	1.53
357.24	3.00	3.19	14.98	20.84	1.34	1.51
375	2.86	3.04	15.23	21.26	1.31	1.48
400	2.68	2.84	15.56	21.81	1.29	1.44
425	2.52	2.65	15.86	22.35	1.26	1.40
450	2.38	2.48	16.14	22.84	1.24	1.37
475	2.26	2.31	16.39	23.31	1.22	1.34
500	2.14	2.16	16.63	23.76	1.20	1.31
525	2.04	2.01	16.85	24.18	1.19	1.29
550	1.95	1.88	17.06	24.58	1.17	1.26
575	1.86	1.75	17.25	24.96	1.16	1.24
600	1.79	1.62	17.43	25.33	1.15	1.22
625	1.71	1.51	17.59	25.67	1.14	1.20
650	1.65	1.39	17.75	26.00	1.13	1.18
675	1.59	1.29	17.90	26.32	1.12	1.17
700	1.53	1.19	18.05	26.63	1.11	1.15

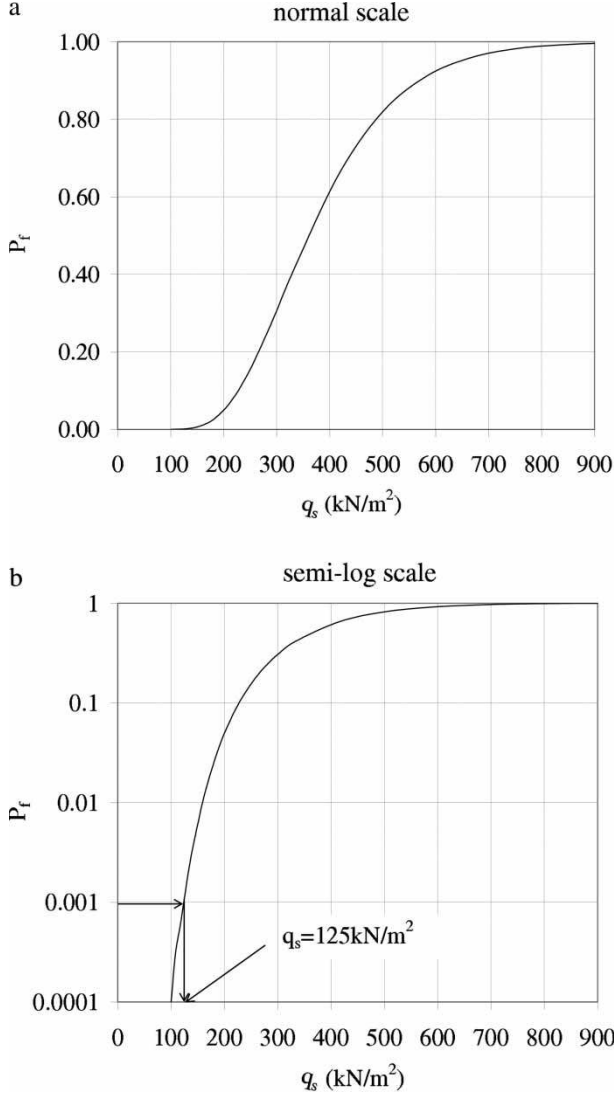


Figure 7. Fragility curves in the case of uncorrelated non-normal random variables using (a) normal scale and (b) semi-log scale.

that, in order to investigate the effect of COV of a certain random variable, its COV is increased or decreased by 50% with respect to its reference value; however, the COV of the other random variable remains constant. Notice also that non-normal and uncorrelated random variables are considered in this section.

Figures 8 and 9 present, respectively, the effect of COV_c and COV_ϕ on the PDF of the system response. The corresponding values of the statistical moments are given in Table 6. This table also provides the effect of COV_c and COV_ϕ on the Sobol indices. From these results, one can observe that COV_c has a negligible effect on the mean value, skewness and kurtosis of the system response; however, it has a small effect on the variability of this response. For

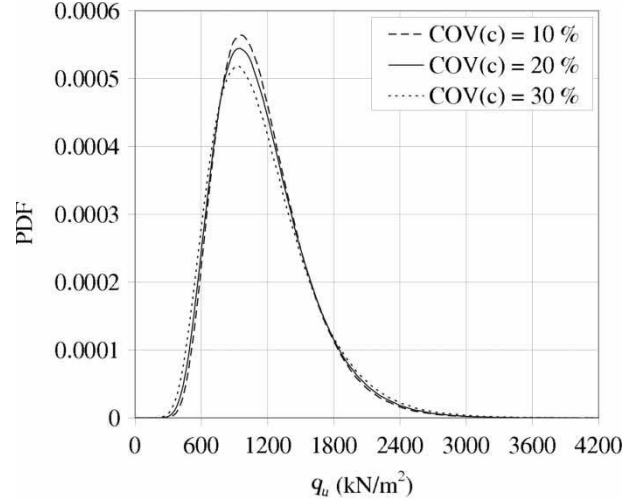


Figure 8. Effect of the coefficient of variation of the soil cohesion (COV_c) on the PDF of the ultimate bearing capacity.

instance, an increase in COV_c by 50% with respect to its reference value increases the COV of the system response by only 6.9%. Concerning COV_ϕ , it was found to have a significant effect on the mean value, skewness and kurtosis of the system response. Also, the variability of the system response was found to be very sensitive to the variability of the soil friction angle (an increase in COV_ϕ by 50% with respect to its reference value increases the COV of the system response by 48.9%). One may observe that the random variable for which the COV is of a significant influence on the variability of the system response (i.e. ϕ) is the one that has the larger Sobol index. Remember that ϕ has a Sobol index of 0.8975 while c has a Sobol index of 0.1025 for the reference case

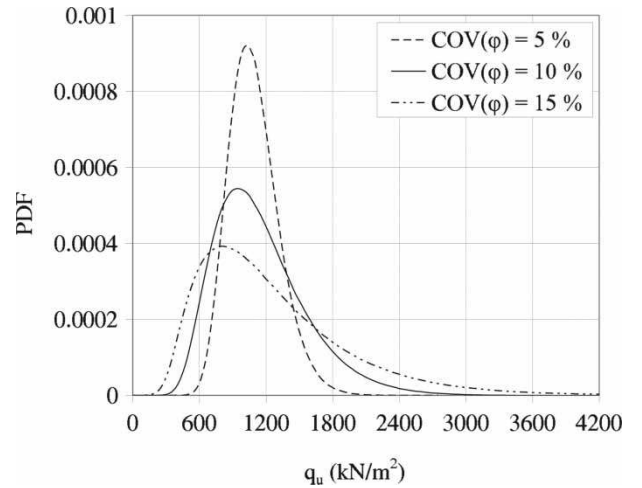


Figure 9. Effect of the coefficient of variation of the soil friction angle (COV_ϕ) on the PDF of the ultimate bearing capacity.

Table 6. Effect of the coefficients of variation of the soil cohesion (COV_c) and the soil friction angle (COV_ϕ) on the statistical moments of the ultimate bearing capacity and on Sobol indices.

COV(%)		Mean value of q_u (kN/m ²)	Standard deviation of q_u (kN/m ²)	Coefficient of variation of q_u (%)	Skewness of q_u	Kurtosis of q_u	Sobol indices	
							S_c	S_ϕ
COV_c (%)	10	1151.05	402.21	34.94	1.04	1.75	0.0278	0.9722
	20	1150.60	418.30	36.35	1.05	1.78	0.1025	0.8975
	30	1149.08	446.52	38.86	1.10	1.98	0.2030	0.7970
COV_ϕ (%)	5	1091.51	225.00	20.61	0.56	0.56	0.3320	0.6680
	10	1150.60	418.30	36.35	1.05	1.78	0.1025	0.8975
	15	1256.11	680.05	54.14	1.51	3.66	0.0431	0.9569

studied before. Finally, Table 6 shows that the increase in COV of a given random variable increases its Sobol index (i.e. its weight in the variability of the system response) and decreases the Sobol index of the other random variable; the COV of the soil friction angle being of greater effect.

5.5.2. Effect of the correlation and the non-normality of the random variables

Figure 10 shows the effect of the correlation and the non-normality of the random variables on the PDF of the ultimate bearing capacity and Table 7 shows the corresponding statistical moments. These results indicate that the mean value of the ultimate bearing capacity is very slightly affected by both the correlation and the non-normality of the random variables. The results also indicate that both assumptions of non-normal variables and negative correlation between these variables have a small effect on the variability of the system response. For instance, the assumption of non-normal random variables decreases the COV of

the system response by 1.9% and 4.7%, respectively, for correlated and uncorrelated random variables. On the other hand, the negative correlation decreases the COV of the system response by 4.5% and 1.8%, respectively, for the cases of normal and non-normal random variables. Concerning the skewness and kurtosis of the system response, they were found to decrease with both the negative correlation and the assumption of non-normal random variables. As a conclusion, these results indicate that the case of normal uncorrelated random variables is conservative since it provides the largest variability of the ultimate bearing capacity.

6. Conclusion

This study presents a combination between the SS approach and the CSRS to calculate not only the failure probability due to an applied footing pressure, but also the PDF of the ultimate bearing capacity. A strip footing of breadth B resting on a (c, ϕ) soil where both c and ϕ are considered as random variables is considered. The combination between the two methods is performed through the use of the realisations generated by the SS approach as collocation points to evaluate the unknown PCE coefficients in CSRS. In addition to the PDF of the ultimate bearing capacity which is computed with no additional calls of the deterministic model, the combination between both methods provides the PCE-based Sobol indices. These indices give the weight of each random variable in the variability of the system response. Finally, the failure probabilities corresponding to q_s values greater than the original one used to perform a SS computation can be calculated based on the obtained PCE of the system response by applying MCS on this PCE. The main findings of this study can be summarised as follows:

- (1) The PDF and the CDF of the ultimate bearing capacity and their corresponding statistical

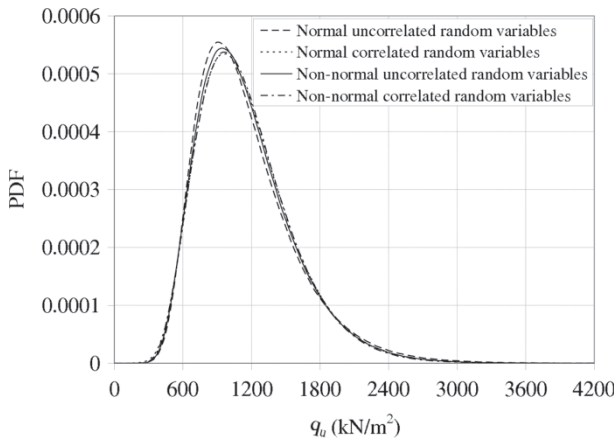


Figure 10. Effect of the correlation between random variables and the type of the probability density function of these variables on the PDF of the ultimate bearing capacity.

Table 7. Effect of the correlation between the random variables and the type of the probability density function (PDF) of these variables on the statistical moments of the ultimate bearing capacity.

Type of the PDF and correlation	Mean value of q_u (kN/m ²)	Standard deviation of q_u (kN/m ²)	Coefficient of variation of q_u (%)	Skewness of q_u	Kurtosis of q_u
Normal uncorrelated variables	1150.51	438.75	38.14	1.28	2.69
Normal correlated variables	1151.09	419.11	36.41	0.96	1.58
Non-normal uncorrelated variables	1150.60	418.30	36.35	1.07	1.78
Non-normal correlated variables	1150.53	410.85	35.71	0.96	1.43

moments, as determined by the proposed procedure, have shown a good agreement with those obtained by applying MCS methodology on the original deterministic model.

- (2) The failure probabilities computed by the proposed procedure and corresponding to q_s values larger than the original one used to perform a SS computation agree well with those computed by applying MCS methodology on the original deterministic model.
- (3) The global sensitivity analysis performed using the PCE-based Sobol indices has shown that the soil friction angle has a significant weight in the variability of the ultimate bearing capacity ($S_\phi = 0.8975$); however, the soil cohesion has a relatively small weight in the variability of this response ($S_c = 0.1025$). This conclusion is valid for the values of the soil uncertainties considered in this article which are the ones frequently encountered in practice for a (c - ϕ) soil.
- (4) The increase in the footing pressure q_s increases the values of c^* and ϕ^* at the design point. However, the reliability index and the partial safety factors F_c and F_ϕ decrease with the increase in the footing pressure. This is to be expected since the increase in the footing pressure decreases the footing safety.
- (5) A fragility curve which can be used to perform either a reliability analysis or a RBD of strip footings was presented. Concerning the reliability analysis, this curve provides the failure probability of a strip footing subjected to a given service load. For the RBD, it allows one to calculate the footing breadth required to support a given service load for a target failure probability.
- (6) A parametric study to investigate the effect of (i) COV_c and COV_ϕ , (ii) the correlation between c and ϕ and (iii) the non-normality of the random variables, on the ultimate bearing capacity has shown that:

- The increase in COV_ϕ considerably increases the variability of the system response; however, the increase in COV_c has a small effect on this variability.
- The random variable for which the COV is of a significant influence on the variability of the system response (i.e. ϕ) is the one that has the greater value of Sobol index.
- The increase in COV of a given random variable increases its Sobol index (i.e. its weight in the variability of the system response) and decreases the Sobol index of the other random variable; the COV of the soil friction angle being of greater effect.
- The variability of the system response was found to decrease with the assumption of non-normal variables with respect to the case of normal variables. This variability also decreases when considering negative correlation between random variables as compared to the case of uncorrelated variables.

References

- Au, S.K. and Beck, J.L., 2001. Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics*, 16, 263–277.
- Au, S.K. and Beck, J.L., 2003. Subset simulation and its application to seismic risk based on dynamic analysis. *Journal of Engineering Mechanics, ASCE*, 129 (8), 901–917.
- Au, S.K., Cao, Z.J., and Wang, Y., 2010. Implementing advanced Monte Carlo simulation under spreadsheet environment. *Structural Safety*, 32 (5), 281–292.
- Au, S.K., Ching, J., and Beck, J.L., 2007. Application of subset simulation methods to reliability benchmark problems. *Structural Safety*, 29, 183–193.
- Houmadi, Y., Ahmed, A., and Soubra, A.-H., 2012. Probabilistic analysis of one-dimensional soil consolidation problem. *Georisk, Assessment and Management of Risk for Engineered Systems and Geohazards*, 6 (1), 36–49.

- Huang, S.P., Liang, B., and Phoon, K.K., 2009. 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.
- Isukapalli, S.S., Roy, A., and Georgopoulos, P.G., 1998. Stochastic response surface methods (SRSM) for uncertainty propagation: application to environmental and biological systems. *Risk Analysis*, 18 (3), 357–363.
- Li, D., et al., 2011. Stochastic response surface method for reliability analysis of rock slopes involving correlated non-normal variables. *Computers and Geotechnics*, 38 (1), 58–68.
- Mao, N., Al-Bittar, T., and Soubra, A.-H., 2012. 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.
- Mollon, G., Dias, D., and Soubra, A.-H., 2011. Probabilistic analysis of pressurized tunnels against face stability using collocation-based stochastic response surface method. *Journal of Geotechnical and Geoenvironmental Engineering, ASCE*, 137 (4), 385–397.
- Phoon, K.K. and Kulhawy, F.H., 1999. Evaluation of geotechnical property variability. *Canadian Geotechnical Journal*, 36, 625–639.
- Santoso, A.M., Phoon, K.K., and Quek, S.T., 2011. Modified Metropolis-Hastings algorithm with reduced chain-correlation for efficient subset simulation. *Probabilistic Engineering Mechanics*, 26 (2), 331–341.
- Schuëller, G.I., Pradlwarter, H.J., and Koutsourelakis, P.S., 2004. A critical appraisal of reliability estimation procedures for high dimensions. *Probabilistic Engineering Mechanics*, 19, 463–474.
- Sobol, I.M., 2001. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55, 271–280.
- Soubra, A.-H., 1999. Upper-bound solution for bearing capacity of foundations. *Journal of Geotechnical and Geoenvironmental Engineering, ASCE*, 125 (1), 59–68.
- Sudret, B., 2008. Global sensitivity analysis using polynomial chaos expansion. *Reliability Engineering and System Safety*, 93, 964–979.
- Wolff, T.H., 1985. Analysis and design of embankment dam slopes: a probabilistic approach. Thesis (PhD). Purdue University, Lafayette, IN.

Appendix 1

Metropolis-Hastings algorithm

The Metropolis–Hastings algorithm is a Markov chain Monte Carlo (MCMC) method. It is used to generate a sequence of new realisations from existing realisations (that follow a target PDF called ‘ P_t ’). Let $s_k \in F_j$ be a current realisation which follows a target PDF ‘ P_t ’. Using a

proposal PDF ‘ P_p ’, a next realisation $s_{k+1} \in F_j$ that follows the target PDF ‘ P_t ’ can be simulated from the current realisation s_k as follows:

- A candidate realisation s is generated using the proposal PDF (P_p). The candidate realisation s is centred at the current realisation s_k .
- Using the deterministic model, evaluate the value of the performance function $G(\hat{s})$ corresponding to the candidate realisation \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 s and set $s_{k+1} = s_k$ (i.e. the current realisation s_k is repeated).
- 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 $r = r_1 r_2$.
- If $r \geq 1$ (i.e. s is distributed according to the P_t), one continues to retain the realisation s_{k+1} obtained in step b; otherwise, reject s and set $s_{k+1} = s_k$ (i.e. the current realisation s_k is repeated).

Notice that in step b, if the candidate realisation \hat{s} does not satisfy the condition $G(\hat{s}) < C_j$, it is rejected and the current realisation s_k is repeated. Also in step d, if the candidate realisation \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 realisation s_k is repeated. The presence of several repeated realisations is not desired as it leads to high probability that the chain of realisations 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. (2010) proposed to modify the classical M-H algorithm as follows:

- A candidate realisation s is generated using the proposal PDF (P_p). The candidate realisation s is centred at the current realisation s_k .
- 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 $r = r_1 r_2$.
- If $r \geq 1$, set $s_{k+1} = \hat{s}$; otherwise, another candidate realisation is generated. Candidate realisations are generated randomly until the condition $r \geq 1$ is satisfied.
- Using the deterministic model, evaluate the value of the performance function $G(s_{k+1})$ of the candidate realisation 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 realisation s_{k+1} obtained in step c; otherwise, reject \hat{s} and set $s_{k+1} = s_k$ (i.e. the current realisation s_k is repeated).

These modifications reduce the repeated realisations and allow one to avoid the computation of the system response of the rejected realisations.

Appendix 2

Computation of the coefficient of determination R^2

Let us consider N realisations $\left\{ \xi^{(I)} = \left(\xi_I^{(I)}, \dots, \xi_M^{(I)} \right), \dots, \xi^{(N)} = \left(\xi_I^{(N)}, \dots, \xi_M^{(N)} \right) \right\}$ of the standard normal random vector ξ , and let $Y = \left\{ Y\left(\xi^{(I)}\right), \dots, Y\left(\xi^{(N)}\right) \right\}$ be the corresponding values of the system response determined by deterministic calculations, then,

$$R^2 = 1 - \Delta_{\text{PCE}}$$

where Δ_{PCE} is given by:

$$\Delta_{\text{PCE}} = \frac{(1/N) \sum_{i=1}^N \left[Y\left(\xi^{(i)}\right) - Y_{\text{app}}\left(\xi^{(i)}\right) \right]^2}{\text{Var}(Y)}$$

and

$$\text{Var}(Y) = \frac{1}{N-1} \sum_{i=1}^N \left[Y\left(\xi^{(i)}\right) - \bar{Y} \right]^2$$

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^N Y\left(\xi^{(i)}\right)$$

Notice that N in these equations is the number of collocation points used to evaluate the unknown coefficients of the PCE. The value $R^2 = 1$ indicates a perfect fit of the true model response Y , whereas $R^2 = 0$ indicates a nonlinear relationship between the true model Y and the PCE model Y_{app} .