An Efficient Method to Compute the Failure Probability

Wenping Gong
Research Assistant Professor, Glenn Dept. of Civil Engineering, Clemson University, Clemson, USA

C. Hsein Juang
Professor, Glenn Dept. of Civil Engineering, Clemson University, Clemson, USA

James R. Martin, II
Professor, Glenn Dept. of Civil Engineering, Clemson University, Clemson, USA

Jie Zhang
Associate Professor, Department of Geotechnical Engineering, Tongji University, Shanghai, China

ABSTRACT: This paper presents an efficient method to compute the failure probability of a geotechnical system, which is based upon numerical integration of the cumulative distribution function (CDF) of the performance function. This new method is inspired by the concept of the vertex method often used in conjunction with fuzzy sets theory; however, new approach is taken to account for the probabilistic feature of the uncertain input parameters. In the new method, only the deterministic analysis of the system performance and the evaluation of the joint probability of the uncertain input parameters are required. The proposed new method is a deterministic approach, easy to follow and apply; no Monte Carlo simulation is required. Through an example study of a shallow strip foundation, the effectiveness and the efficiency of the proposed new method, in terms of the accuracy and the computational effort, respectively, are demonstrated.

1. INTRODUCTION

The failure probability of a geotechnical system (P_f) may be computed as a multi-fold probability integral, expressed as follows:

$$P_f = \Pr\left[g(\boldsymbol{X}) \le 0\right] = \int_{g(\boldsymbol{X})\le 0} f(\boldsymbol{X})d\boldsymbol{X} \tag{1}$$

where $\boldsymbol{X} = [X_1, X_2, \ldots, X_n]^T$, is a vector of the uncertain input parameters X_1, X_2, …, and X_n, in which the subscript n is the number of the uncertain input parameters; $f(\boldsymbol{X})$ is the joint probability density function (PDF) of the uncertain parameters X; $g(\boldsymbol{X})$ is the performance function, which is formulated such that $g(\boldsymbol{X}) \le 0$ denotes the failure of the geotechnical system; and, $\Pr\left[g(\boldsymbol{X}) \le 0\right]$ is the conditional probability of $g(\boldsymbol{X}) \le 0$.

Difficulty in evaluating the multi-fold probability integral in Eq. (1) has led to many approximation methods, such as mean value first order second moment method (Ang and Tang 2007), advanced first order second moment method or first order reliability method (Hasofer and Lind 1974; Melchers 1987; Lee and Kwak 1987), and point estimate-based moment method (Zhao and Ono 2000). Although these methods have been widely applied in engineering practices, there is room for improvement in a few aspects. First, the accuracy of these approximate methods may be a problem if the performance function is highly nonlinear and/or high-dimensional. Second, the computation of the partial derivative of the performance function may be a challenge, especially in the situations where the system performance can only be evaluated using the numerical methods such as finite element method (FEM). Third, since the distribution of the performance function is approximated with its moments of finite order, the evaluation of the moments may introduce errors. Fourth, in search for the minimum reliability index (β), a local minimum, rather than the global minimum, may be possible. The failure probability is related to the reliability index as follows:

$$P_f = \Phi(-\beta) \tag{2}$$

where $\Phi(\cdot)$ is the cumulative distribution function (CDF) of the standard normal variable.

To avoid the shortcomings of these methods, the sampling method such as Monte Carlo simulation (MCS) may be used alternatively. Although MCS could yield a more accurate determination of the failure probability of the geotechnical system, the required number of simulations of the system performance may be too large, especially for a system of low failure probability. In many cases, the application of MCS may be limited due to its prohibitive computational demand. The low computational efficiency of MCS becomes more profound when it is applied to a system, the performance of which can only be analyzed using numerical methods. In order to improve the computational efficiency of MCS, various sampling techniques have been studied, such as Latin hypercube sampling (Florian 1992), importance sampling (Grooteman 2011), and subset simulation (Au and Beck 2001; Ching et al. 2005). It is noted that while the computational efficiency of MCS can be greatly improved through the use of these sampling techniques, the knowledge of advanced probability theory and programing skills can be a barrier to the practicing engineer.

In this paper, a new method, based upon the numerical integration of the CDF of the performance function, is created for computing the failure probability of a geotechnical system. Here, the performance function and the joint probabilities of the integration grids, through the vertex combinations of the uncertain input parameters, are computed to construct the CDF of the performance function. The new method is formulated in a deterministic manner, which does not require the computation of the partial derivative and the moments of the performance function, and nor does it require an iterative process to minimize the reliability index. Thus, the proposed new method is easy to apply. As will be shown later, the new method can produce results that agree very well with those obtained from MCS, and yet, it is much more efficient.

2. NEW METHOD TO COMPUTE THE FAILURE PROBABILITY

Based upon the CDF of the performance function, the failure probability of a system is expressed as follows:

$$P_f = F\left[g(\boldsymbol{X}) = 0\right] \tag{3}$$

where $F[g(\boldsymbol{X})]$ is the CDF of the performance function $g(\boldsymbol{X})$, the detailed construction of which is formulated later.

2.1. Numerical integration of the CDF of the performance function

The analytical (or direct) integration of the CDF of the performance function, $F[g(\boldsymbol{X})]$, can be a challenge if the performance function is highly nonlinear and/or high dimensional. Thus, the numerical integration method is employed herein to construct the CDF of the performance function. Here, each and every uncertain input parameter is first discretized into a set of discrete vertices using the α-cut concept of fuzzy sets theory (e.g., Dong and Wong 1987; Juang et al. 1998; Gong et al. 2014&2015). The obtained vertices are then combined to represent the domain of the uncertain input parameters. Finally, the CDF of the performance function is constructed with the computed performance function and joint probabilities of the vertex combinations of the uncertain input parameters. In short, the CDF of the performance function may be established with following steps:

1. Discretize the truncated standard normal variable of [-5, 5] into a set of discrete vertices using the α-cut concept of fuzzy sets theory. For example, a set of discrete vertices, denoted as $\{x_0, x_1^-, x_1^+, x_2^-, x_2^+, \ldots, x_{(m-1)}^-, x_{(m-1)}^+\}$, is obtained if the number of α-cut levels is m. Plotted in Figure 1 are the resulting vertices of the standard normal variable x_1. As such, the distribution of the uncertain

parameter x_1 is represented with a set of discrete vertices: $[x_{10}, f(x_{10})]$, $[x_{11}^-, f(x_{11}^-)]$, and $[x_{11}^+, f(x_{11}^+)]$ for $m = 2$; $[x_{10}, f(x_{10})]$, $[x_{11}^-, f(x_{11}^-)]$, $[x_{11}^+, f(x_{11}^+)]$, $[x_{12}^-, f(x_{12}^-)]$, and $[x_{12}^+, f(x_{12}^+)]$ for $m = 3$; or, $[x_{10}, f(x_{10})]$, $[x_{11}^-, f(x_{11}^-)]$, $[x_{11}^+, f(x_{11}^+)]$, $[x_{13}^-, f(x_{13}^-)]$, $[x_{13}^+, f(x_{13}^+)]$, $[x_{14}^-, f(x_{14}^-)]$, and $[x_{14}^+, f(x_{14}^+)]$ for $m = 4$, as shown in Figure 1. Here, the truncated range is set at [-5, 5] so that the probability of the uncertain input parameters falling outside of this range is less than 5.733×10^{-7}, a relatively low probability that can be ignored.

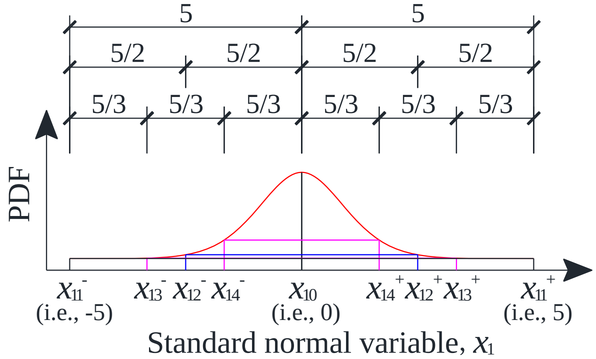


Figure 1: α-cut concept of the uncertain input parameter.

2. At each α-cut level, compute the vertices of the uncertain input parameters in the original distribution spaces of the uncertain input parameters, denoted as $\{X_{i0}, X_{i1}^-, X_{i1}^+, X_{i2}^-, X_{i2}^+, \ldots, X_{i(m-1)}^-, X_{i(m-1)}^+\}$, using the transformation that was suggested in Low and Tang (2007).
3. Combine the vertices of the uncertain input parameters that obtained in Step 2. Note that the number of the vertex combinations is $(2m-1)^n$, where m is the number of α-cut levels and n is the number of uncertain parameters.
4. Calculate the performance function, $g(\boldsymbol{X}_i)$, and the joint probability, $f(\boldsymbol{X}_i)$, of each vertex combination of the uncertain parameters, $\boldsymbol{X}_i$. This process is repeated for all $(2m-1)^n$ vertex combinations of the uncertain parameters.
5. Construct the CDF of the performance function, $F[g(\boldsymbol{X})]$, which is expressed as:

$$F[g(\boldsymbol{X})] = \int_{g(\boldsymbol{Y}) \le g(\boldsymbol{X})} f(\boldsymbol{Y}) d\boldsymbol{Y} \tag{4}$$

where $\boldsymbol{Y} = [X_1, X_2, \ldots, X_n]^T$. Given the $(2m-1)^n$ pairs of $g(\boldsymbol{X}_i)$ and $f(\boldsymbol{X}_i)$ values obtained in Step 4, the integral in Eq. (4) can be approximated with:

$$F[g(\boldsymbol{X})] = \frac{\sum_{g(\boldsymbol{X}_i) \le g(\boldsymbol{X})} f(\boldsymbol{X}_i)}{\sum f(\boldsymbol{X}_i)} \tag{5}$$

Then, the failure probability of the system, P_f, is obtained using linear interpolation.

Note that in the evaluation of the joint probability, $f(\boldsymbol{X}_i)$, of the vertex combination of the uncertain input parameters, $\boldsymbol{X}_i$, the standard normal space and the standard normal variables, in terms of $\{x_0, x_1^-, x_1^+, x_2^-, x_2^+, \ldots, x_{(m-1)}^-, x_{(m-1)}^+\}$, should be used. Further, the correlation between (or among) the uncertain input parameters should be considered, in which the original correlation coefficient should be modified in line with the equivalent normal transformation, as suggested in Der Kiureghian and Liu (1986).

2.2. Optimization of the number of cut levels of the uncertain input parameters

It is noted with the increase of the number of cut levels (m), the CDF of the performance function can be more accurately constructed and thus the failure probability. The CDF of the performance function and the failure probability of the system would converge to the true value (or analytical solution) if the selected number of cut levels (m) is sufficiently large. However, an increase in m beyond some certain level can lead to a drastic increase in the computational effort. Here, the optimal number of the cut levels (m) can be determined with following procedures:

1. Set the initial m as 2, denoted $m_0 = 2$, and compute the initial failure probability of the system, denoted as P_{f0}, using the procedures in Section 2.1.
2. Set a new m as $(m_0 + 1)$, denoted as $m_1 = m_0+1$, and compute the corresponding failure probability of the system, denoted as P_{f1}.
3. Set a new m of $(m_1 + 1)$, denoted as $m_2 = m_1+1$, and compute the corresponding failure

probability of the system, denoted as P_{f2}. Then, the relative variation (or error) of the computed failure probability is computed as follows:

$$\varepsilon_1 = \frac{\left|P_{f0} - P_{f1}\right|}{P_{f1}} \times 100\% \tag{6}$$

$$\varepsilon_2 = \frac{\left|P_{f1} - P_{f2}\right|}{P_{f2}} \times 100\% \tag{7}$$

where ε_1 and ε_2 are the relative variation of the computed failure probability with respect to the number of cut levels m_1 and m_2, respectively.

4. Determine whether the number of cut levels m_2 is acceptable with the following acceptance rule: if $\varepsilon_1 < 1.0\%$ and $\varepsilon_2 < 1.0\%$, then m_2 is acceptable and the failure probability of the system is P_{f2}, denoted as $m = m_2$ and $P_f = P_{f2}$; otherwise, set $m_1 = m_2$, $P_{f0} = P_{f1}$, and $P_{f1} = P_{f2}$, and then go back to Step 3. The convergence criterion, in terms of $\varepsilon_1 < 1.0\%$ and $\varepsilon_2 < 1.0\%$, is used herein only for illustration purposes. The selection of convergence criterion is problem-specific; for example, 5.0% may be acceptable in other problems.

Note that in the construction of the CDF of the performance function, $F[g(\boldsymbol{X})]$, at a given number of cut levels m ($m > 2$), the results obtained at the number of cut levels of (m -1) should be utilized. In other words, only the new vertex combinations of the uncertain input parameters that are formed at the current number of cut levels need to be analyzed, while the vertex combinations that has been analyzed previously are not required to be studied. In reference to Figure 2, for a system with 2 uncertain parameters, 9 vertex combinations of these parameters need to be studied at the level of m = 2; 16 new vertex combinations need to be analyzed at the level of m = 3; and, 40 new vertex combinations need to be studied at the level of m = 4. However, 25 pairs of the performance function and joint probability values are utilized to establish the CDF of the performance function for m = 3; and, 65 pairs of the performance function and joint probability values are utilized to establish the CDF of the performance function for m = 4.

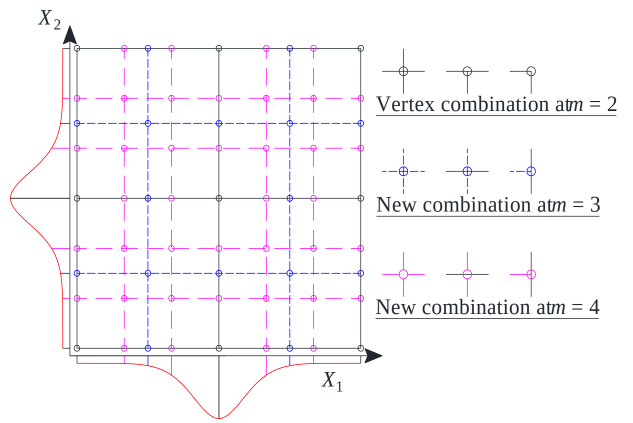


Figure 2: Vertex combinations of the uncertain input parameters.

3. ILLUSTRATIVE EXAMPLE

To illustrate the effectiveness and the efficiency of the proposed new method in computing the failure probability of a geotechnical system, a shallow strip foundation is examined.

3.1. Bearing capacity of the shallow strip foundation

The ultimate bearing capacity of the shallow strip foundation is evaluated with the following expression:

$$q_{\text{ult}} = 0.5\gamma_1 B N_\gamma + c N_c + \gamma_2 D N_q \tag{8}$$

where B is the width of the foundation; D is the depth of the foundation relative to the ground level; γ_1 is the unit weight of the soil under the foundation base; γ_2 is unit weight of the soil above the foundation base; c is the effective cohesion; and, N_c, N_q, and N_γ are the bearing capacity factors, which are estimated as follows:

$$N_\gamma = 1.8\left(N_q - 1\right)\tan\phi \tag{9}$$

$$N_q = \tan^2\left(\frac{\pi}{4} + \frac{\phi}{2}\right) e^{\pi \tan\phi} \tag{10}$$

$$N_c = \left(N_q - 1\right)\cot\phi \tag{11}$$

where ϕ is the effective friction angle. No ground water table is considered herein for simplicity.

The performance function of this geotechnical problem is formulated as follows:

$$g = q_{\text{load}} - q_{\text{ult}} \tag{12}$$

where q_{load} is the load effect.

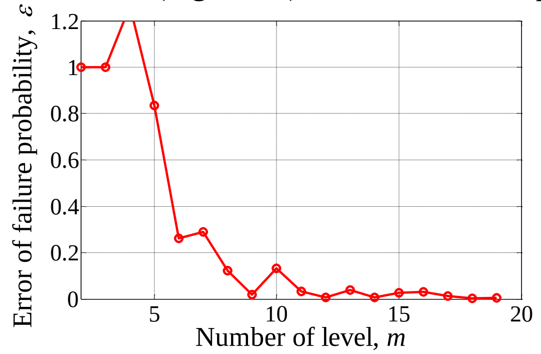
In this example, the geometry of the foundation, the load on the foundation, and the unit weight of the soil are treated as deterministic input parameters: q_{load} = 300 kPa, B = 1.5 m, D = 1.2 m, and $\gamma_1 = \gamma_2$ = 17.3 kN/m^3. The effective cohesion and friction angle are taken as uncertain input parameters. The mean and the standard deviation of the effective cohesion are 14.4 kPa and 1.7 kPa, respectively (i.e., μ_c = 14.4 kPa and σ_c = 1.7 kPa). The mean and the standard deviation of the effective friction angle are 20° and 1.2°, respectively (i.e., μ_ϕ = 20° and σ_ϕ = 1.2°). The effective cohesion and the effective friction are negatively correlated, $\rho_{c,\phi}$ = -0.5. Note that the parameters setting of this example is based upon Cherubini (2000).

3.2. Reliability analysis of the bearing capacity using the proposed new method

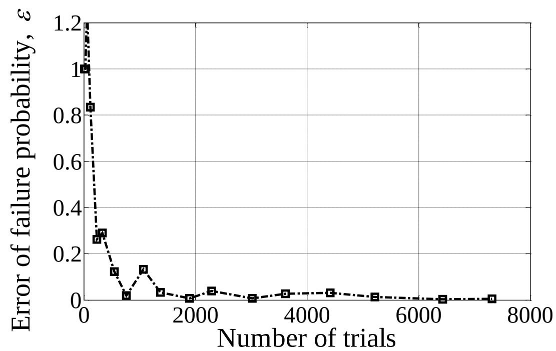
Following the procedures of the proposed new method that outlined in Section 2.2, the failure probability and the reliability index of the shallow foundation are readily computed. The obtained failure probability is $P_f = 1.661\times10^{-3}$ and the corresponding reliability index is β = 2.936, whereas, the "exact" failure probability obtained with 1,000,000MCS runs is $P_{f\text{MCS}} = 1.666\times10^{-3}$ and the corresponding reliability index is β_{MCS} = 2.935. Here, the relative error of the reliability (ε), defined below, is only 0.03%.

$$\varepsilon = \left| \frac{\beta - \beta_{\text{MCS}}}{\beta_{\text{MCS}}} \right| \times 100\% \tag{13}$$

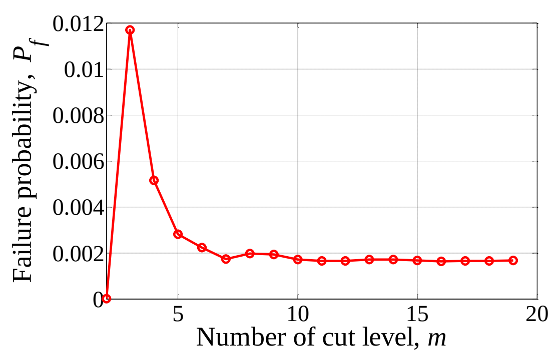
Figure 3 shows the convergence of the analysis results, in terms of failure probability (P_f), and the relative variation of the failure probability using the proposed new method with the adopted number of cut levels (m) and the number of trials. As can be seen, the relative variation of the computed failure probability decreases with the number of cut levels of the uncertain input parameters (Figure 3a) and the number of trials (or simulations) of the system performance (Figure 3b). As such, the computed



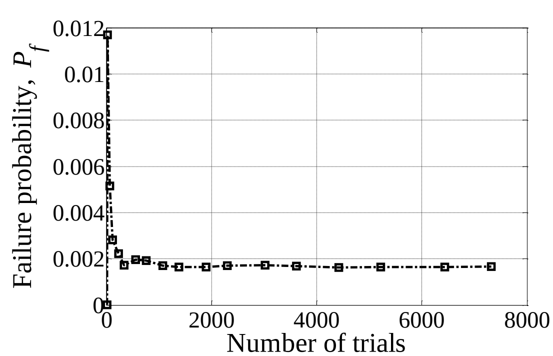
(a) Relative variation versus number of cut levels



(b) Relative variation versus number of trials



(c) Failure probability versus number of cut levels

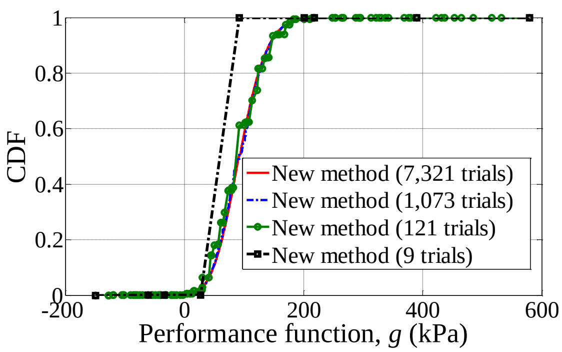


(d) Failure probability versus number of trials

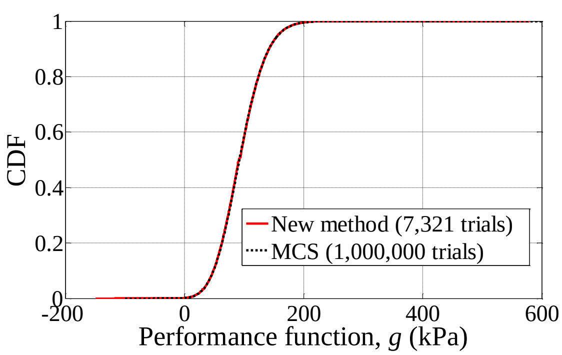
Figure 3: Convergence of the analysis results using the proposed new method.

failure probability converges with the number of cut levels (Figure 3c) and the number of trials for the system performance (Figure 3d).

For this example, the optimal number of cut levels for the set of uncertain input parameters is 19 and the number of trials for the system performance is 7,321. Actually, the variation of the computed failure probability is relatively low when the number of cut levels is greater than 10 (i.e., $m > 10$) and the number of trials (simulations) for the system performance is greater than 1,073, as demonstrated in Figure 3.



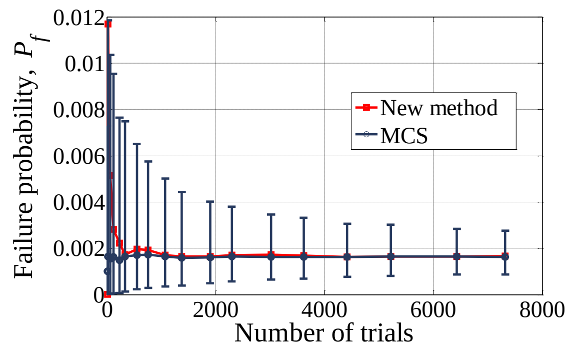
(a) CDF of the performance function obtained with different number of trials



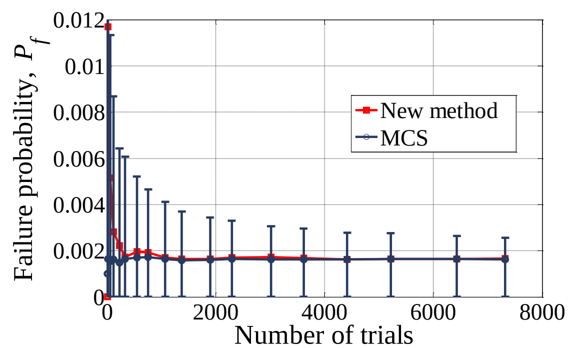
(b) CDF of the performance function obtained with 1,000,000 MCSs

Figure 4: CDF of the performance function using the proposed new method.

Further plotted in Figure 4(a) are the CDFs of the performance function that are obtained with different numbers of cut levels of the uncertain input parameters (indicated herein by the number of trials for the system performance). It shows that with the increase of the number of cut levels, the obtained CDF of the performance function converges. As noted, there is little difference between the CDF of the performance function obtained with 10 cut levels (represented with the number of trials of 1,073) and that obtained with 19 cut levels (represented with the number of trials of 7,321). Further, the CDF of the performance function obtained with 19 cut levels agrees well with that obtained using 1,000,000 samples of MCS (Figure 4b). Therefore, the effectiveness and efficiency of the proposed new method are clearly demonstrated.



(a) P_f is lognormally distributed



(b) P_f is truncated normally distributed

Figure 5. Comparison of the failure probability between the proposed new method and MCS.

Note that in this paper, the sample number of 1,000,000 is selected such that the coefficient variation of the failure probability obtained with MCS, defined below (Eq. 14), is deemed low and negligible.

$$\mathrm{COV}_{P_f} = \sqrt{\frac{1-P_f}{n_{\mathrm{MCS}} P_f}} \tag{14}$$

where COV_{P_f} is the coefficient of variation of the failure probability obtained with MCS, and n_{MCS} is the sample number required in the MCS analysis.

Table 1: Reliability analysis results with different parameters setting

$\rho_{c,\phi}$	New method: Number of trials	New method: β	1,000,000 MCSs: β_{MCS}	Relative error, ε_β
Scenario 1: σ_c = 1.7 kPa and σ_ϕ = 1.2°				
0.0	7,321	2.135	2.145	0.47%
-0.25	7,321	2.442	2.450	0.32%
-0.5	7,321	2.936	2.935	0.03%
-0.75	5,217	3.935	3.895	1.01%
Scenario 2: σ_c = 1.7 kPa and σ_ϕ = 1.2°				
0.0	2,297	1.181	1.201	1.61%
-0.25	7,321	1.349	1.350	0.11%
-0.5	7,321	1.571	1.574	0.16%
-0.75	7,321	1.955	1.964	0.47%
Scenario 3: σ_c = 2.2 kPa and σ_ϕ = 1.5°				
0.0	4,417	1.683	1.696	0.72%
-0.25	4,417	1.932	1.934	0.13%
-0.5	5,217	2.333	2.340	0.30%
-0.75	10,017	3.154	3.148	0.17%
Scenario 4: σ_c = 3.4 kPa and σ_ϕ = 3.1°				
0.0	1,377	0.863	0.889	2.87%
-0.25	3,017	0.978	1.000	2.18%
-0.5	2,297	1.141	1.162	1.85%
-0.75	2,297	1.423	1.454	2.10%

Additional series of analyses is carried out to compare the proposed method with MCS, and the results are presented in the following. It is well recognized that some variation of the failure probability obtained with MCS is expected. Thus 1,000 repeats of MCS are carried out herein to derive a 95% confidence level of the failure probability. Figure 5 shows the comparison between the failure probability obtained with the proposed new method and the 95% confidence level of the failure probability obtained with 1,000 repeats of MCS. The failure probability is assumed to be lognormally distributed in Figure 5(a), while the truncated normal distribution is assumed in Figure 5(b).

Figure 5 shows that the variation of the failure probability obtained with MCS (indicated by the bars of the 95% confidence intervals) decreases with the number of MCS samples; while the mean of the failure probability obtained with MCS agrees well with that obtained with the new method. Therefore, compared with MCS, the proposed method is more effective (i.e., little variation/uncertainty) in evaluating the failure probability. The new method also converges quicker.

3.3. *Further assessment of the proposed new method*

In order to investigate the influence of the variation of the uncertain input parameters and the correlation between (or among) the uncertain input parameters, different parameters settings are adopted for this geotechnical problem and the analysis results are listed in Table 1. This parameters setting is based on Cherubini (2000).

The results show that while the two methods yield approximately the same reliability index values for the shallow strip foundation examined in different parameters settings, the proposed method requires much less number of trials than does the MCS. Thus, the effectiveness (i.e., accuracy) and the efficiency (i.e., computational saving) of the new method in evaluating the failure probability are not influenced by different parameter settings. Therefore, the conclusions reached previously about the proposed method are further confirmed.

In fact, the effectiveness and the efficiency of the proposed new method are also not affected by the distribution of the uncertain input parameters, the nonlinearity of the performance function, and the form of the performance function. These features are addressed in an ongoing study at Clemson University.

4. CONCLUSIONS

Based upon the results presented, the effectiveness (i.e., accuracy) and the efficiency (i.e., computational effort) of the proposed new method are established. This finding is not influenced by the variation of the uncertain input parameters and the correlation between (or among) the uncertain input parameters. Apart from the failure probability of the system, the CDF of the performance function can also be effectively and efficiently constructed using the proposed new method.

Obviously, the proposed new method is not perfect. A potential limitation of the proposed method is that the computational efficiency may be reduced when the problem involves a large number of uncertain input parameters, which is the nature of the vertex methods. Indeed, the new method needs further investigations but it shows a potential.

5. REFERENCES

Ang, A.H.S., and Tang, W.H. (2007). *Probability Concepts in Engineering: Emphasis on Applications to Civil and Environmental Engineering*, 2nd Ed., Wiley, New York.

Au, S. K., and Beck, J. L. (2001). Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics*, 16(4), 263-277.

Cherubini, C. (2000). "Reliability evaluation of shallow foundation bearing capacity on c' φ'soils." *Canadian Geotechnical Journal*, 37(1), 264-269.

Ching, J., Beck, J.L., and Au, S.K. (2005). "Hybrid subset simulation method for reliability estimation of dynamical systems subject to stochastic excitation." *Probabilistic Engineering Mechanics*, 20(3), 199-214.

Der Kiureghian, A., and Liu, P.L. (1986). "Structural reliability under incomplete probability information." *Journal of Engineering Mechanics*, 112(1), 85-104.

Dong, W.M., and Wong, F.S. (1987). "Fuzzy weighted averages and implementation of the extension principle." *Fuzzy Sets and Systems*, 21(2), 183-199.

Florian, A. (1992). "An efficient sampling scheme: updated latin hypercube sampling." *Probabilistic Engineering Mechanics*, 7(2), 123-130.

Gong, W., Wang, L., Juang, C.H., Zhang, J., and Huang, H. (2014). "Robust geotechnical design of shield-driven tunnels." *Computers and Geotechnics*, 56, 191-201.

Gong, W., Wang, L., Khoshnevisan, S., Juang, C. H., Huang, H., and Zhang, J. (2015). "Robust geotechnical design of earth slopes using fuzzy sets." *Journal of Geotechnical and Geoenvironmental Engineering*, 141(1), 04014084.

Grooteman, F. (2011). "An adaptive directional importance sampling method for structural reliability." *Probabilistic Engineering Mechanics*, 26(2), 134-141.

Hasofer, A.M., and Lind, N.C. (1974). "Exact and invariant second-moment code format." *Journal of the Engineering Mechanics Division*, 100(1), 111-121.

Juang, C.H., Jhi, Y.Y., and Lee, D.H. (1998). "Stability analysis of existing slopes considering uncertainty." *Engineering Geology*, 49(2), 111-122.

Lee, T.W., and Kwak, B.M. (1987). "A reliability-based optimal design using advanced first order second moment method." *Journal of Structural Mechanics*, 15(4), 523-542.

Low, B.K., and Tang, W.H. (2007). "Efficient spreadsheet algorithm for first-order reliability method." *Journal of Engineering Mechanics*, 133(12), 1378-1387.

Melchers, R. E. (1987). *Structural Reliability*: Analysis and Prediction. Ellis Horwood Ltd., Chichester, West Sussex, U.K.

Zhao, Y. G., and Ono, T. (2000). "New point estimates for probability moments." *Journal of Engineering Mechanics*, 126(4), 433-436.