Available online at www.CivileJournal.org

Civil Engineering Journal

(E-ISSN: 2476-3055; ISSN: 2676-6957)

Vol. 12, No. 03, March, 2026

Effect of Spatial Variability of Soil Shear Strength on the Displacement of an Axially Loaded Pile

Dung Van Le [1, 2], Kien Trung Nguyen [1, 2*]

[1] *Faculty of Civil Engineering, Ho Chi Minh City University of Technology (HCMUT), Ho Chi Minh City, Vietnam.*

[2] *Vietnam National University Ho Chi Minh City, Linh Xuan Ward, Ho Chi Minh City, Vietnam.*

Received 16 August 2025; Revised 11 February 2026; Accepted 16 February 2026; Published 01 March 2026

Abstract

This manuscript studies the influence of spatial heterogeneity in soil effective cohesion c' on the displacement of an axially loaded pile under static compression. The Monte Carlo method was used to perform statistical evaluation, while the K-Means clustering technique was employed to reduce the necessary number of simulations without compromising the accuracy and reliability of the analysis results. In this study, a two-dimensional correlated random field of c' was generated using the spectral method, assuming a Gaussian distribution with a coefficient of variation of 10%. The spatial correlation lengths of the scaled model were set to 80 mm and 19.4 mm in the horizontal and vertical directions, respectively. The static pile loading process was simulated using the finite element method implemented in the commercial software PLAXIS, with six loading stages ranging from 20 N to 120 N. The influence of soil effective cohesion randomness on pile displacements was evaluated based on statistical data obtained from the numerical simulation of 1000 realizations of random fields. The generated correlated random fields were verified to follow a normal distribution. Meanwhile, the pile displacements at different loading levels tend to follow normal distribution as well, except at the 40 N loading stage, where a deviation from normality was observed. Compared with the homogeneous soil model with a mean value of c', up to 63% of the simulation cases exhibit pile displacement exceeding the allowable limit. Moreover, the effectiveness of the K-means clustering technique in reducing the number of required realizations, and thus the computational workload, was evaluated. The pile displacement results show that when the number of representative clusters reaches 100 or more, the subset data cover more than 98.2% of the full data set, i.e., data from 1000 realizations.

Keywords: Random Field Theory; Effective Cohesion c'; Monte Carlo Simulation; Pile Displacement; K-Means Clustering.

1. Introduction

In the analysis of geotechnical problems, assuming a constant value of soil properties, usually the statistical mean for an entire soil layer, is often inadequate, as material properties typically exhibit spatial variability. Therefore, simulating the spatial variation of soil properties plays a crucial role in predicting and evaluating the reliability of geotechnical structures. One effective approach for considering this variability is the construction of random fields combined with statistical techniques, which have been demonstrated to be suitable in numerous practical studies. This method has been successfully applied in the previous works, such as the study by Tang et al. on the slope stability under spatially variable soil conditions [1], the investigation by Chen et al. on the tunnel face stability and soil arching considering spatial variation of sandy soil [2], and the work by Mert et al. to evaluate the hazard curves of strip foundations resting on sandy soil with spatial randomness of elastic modulus and internal friction angle [3].

* Corresponding author: nguyentrungkien@hcmut.edu.vn

https://doi.org/10.28991/CEJ-2026-012-03-02

The spectral representation method is one of the most effective methods for generating correlated random fields. One noticeable advantage of this method lies in its ability to tightly control the spectral characteristics of the field, allowing accurate simulation and flexibility in adjusting the statistical parameters of the random field, while reducing computational cost by Fourier transforms [4]. This method was first applied in the generation of one-dimensional (1D) random fields, where the field's spectrum is directly controlled through statistical parameters such as the power spectral density function. Mejía and Iturbe used the spectral method to simulate hydrological processes with spatial variability [5], while Shinozuka [6] applied this method to simulate multivariate random processes, enhancing the capability to model complex random fields. The later studies by Shinozuka et al. [7, 8] further extended this method, providing a powerful tool for simulating random processes in various engineering fields.

The extension of the spectral method from 1D to two-dimensional (2D) configuration is an important step in the simulation of random fields with spatial correlation. Gutjahr et al. introduced a joint conditional simulation method combined with Fast Fourier Transform (FFT), which improves the efficiency of 2D random field simulation [9]. Ruan and McLaughlin also used FFT to generate multivariate random fields, optimizing the simulation process for 2D problems [10]. In addition, studies on random fields in geology and seismology, such as the work by Mai and Beroza, have applied spatial random field models to describe the complexity of earthquake slip distributions [11]. In terms of engineering applications, Jin et al. [12] made significant contributions to the simulation of multidimensional random fields, including 2D configurations. Later, Luo and Luo applied 2D rotated anisotropic random fields in the reliability analysis of strip footings, providing an important reference for problems related to structural strength [13].

In the scenarios where three-dimensional (3D) simulation of spatial random fields is required, the spectral method continues to demonstrate its effectiveness. For example, Räss et al. [14] developed an efficient method using parallel computing to generate 3D random fields for large-scale geophysical problems, which helps to reduce computational time and enhance simulation efficiency. The 3D direct Fourier transform was also used in the work by Robin et al. to simulate crossed-correlated random fields for water resources problems [15].

After the generation of a random field, the Monte Carlo simulation method is implemented to carry out a series of computational scenarios corresponding to different random field realizations. Through this process, the output results are statistically summarized and analyzed to clarify the probability distribution and variation trends caused by the heterogeneity of the soil, thereby assessing the influence of soil properties on the structural behavior.

In uncertainty analysis problems using the Monte Carlo method, a large number of simulations are required to ensure the reliability of the results. However, this leads to high computational costs and prolonged processing time. To overcome this limitation, several studies have proposed data clustering techniques to reduce the number of required simulations. According to Jain, K-means technique has been one of the most widely used and influential clustering algorithms in the field of data mining for over 50 years due to its simplicity and efficient handling of large datasets [16]. For example, Xiao et al. [17] combined clustering with surrogate modeling to enhance the efficiency of uncertainty quantification in geotechnical problems. The study shows that the application of K-means not only significantly reduces the number of simulations but also offers a more efficient approach to risk and reliability assessment of complex engineering systems. It is noted that clustering and selecting representative samples for each cluster in K-means approach may result in the loss of some outlier values, particularly in the circumstances of high variability. Given the low probability of encountering such highly distinct values, their omission results in a negligible effect on the calculated reliability probability.

Many studies have applied random fields to assess the reliability of pile foundations in spatially variable soil environments. For instance, Khorramian et al. [18] employed the stochastic finite element method to analyze pile groups in loose sand with random fields simulating soil friction angle and stiffness. However, this study was limited to evaluating the serviceability limit state and did not consider pile head deformation or single pile models. Similarly, the study by Liu et al. [19] developed a non-stationary random field model to simulate the variation of undrained shear strength S_u in embankment foundation soils. However, this model was limited to 1D space and did not fully capture correlated randomness in 2D space when modeling single piles. Other studies, such as those by Xiao-ling et al. [20] and Christodoulou et al. [21] have focused on lateral loading or reliability assessment under field investigation effects, though a 2D correlated random field of effective cohesion of soft soil was not considered. Tan et al. [22] simulated random fields of end bearing resistance and shaft friction to evaluate the reliability of the bearing capacity of single piles. They applied the variance reduction method to reduce the number of computations, while demonstrating the effectiveness of this approach. Gan et al. [23] further extended their study by analyzing the lateral deflection of a single pile under horizontal loading while incorporating the spatial variability of soil stiffness, with the objective of evaluating the probability of exceeding a prescribed deflection limit. Meanwhile, Duan et al. [24] investigated the effect of cone tip resistance random field on the failure probability of the vertically loaded piles. The conditional randomness of soil parameters was also investigated by Xu et al. [25] to evaluate the probabilistic ultimate bearing capacity of vertical piles, considering the deterministic soil properties at observation points.

The spatial variability of undrained shear strength (c') and its effect on single pile deformation and bearing capacity have received limited attention in prior work. This study introduces three primary novelties: (i) an in-depth numerical investigation of this problem using the PLAXIS software, (ii) the automation of these PLAXIS computations, and (iii) the application of the K-means clustering algorithm to efficiently reduce the computational burden.

Finite element method (FEM) was used to investigate the influence of spatial randomness in effective cohesion *c'* on pile head displacement in soft clay. Initially, the numerical FEM model with constant shear strength parameters was validated with a 1:100 scaled physical model [26]. Subsequently, the FEM model was analyzed with randomly distributed values of *c'*, generated using the spectral method with a coefficient of variation of 10% [27], correlation length of 80 mm and 19.4 mm in the horizontal and vertical directions, respectively. These are typical values for random field of *c'* which have been used widely in geotechnical problems [28]. Monte Carlo analysis was performed with 1000 realizations to obtain the probability distribution of pile head displacement. The K-means algorithm was then applied to reduce the number of representative cases, ensuring computational efficiency while preserving the statistical characteristics of the original random field.

The remainder of the paper is organized as follows: section 2 presents the methodology, including the procedure for generating correlated random fields, the experimental model, and the numerical simulation, as well as the automation procedure for analyses in PLAXIS. Meanwhile, section 3 presents the analysis results and discussion. Conclusions are provided in section 4.

2. Research Methodology

2.1. Theoretical Background

Gaussian random fields, also known as normally distributed random fields, are widely-used models in scientific and engineering problems, especially in simulating natural phenomena exhibiting randomness, such as spatial variability in soil, materials, climate processes, and so on. Simulating 2D Gaussian random fields with specified spatial correlation structures serves as a crucial tool in numerical modeling and geotechnical engineering. The discrete Fourier transform was used in this study to generate 2D Gaussian random fields with spatially correlated structures [9-15]. This study presents the theoretical basis for generating a 2D discrete random field with a Gaussian distribution, characterized by a mean $\mu = 0$, a standard deviation $\sigma = 1$, and a Gaussian correlation function. For the sake of completeness, the procedure is briefly described here.

Let's consider a 2D discrete grid of size $N \times M$, where N is the number of rows and M is the number of columns. The random field $Z(x, y)$ is generated in the frequency domain, with each grid value being represented by a complex number:

$$Z(k, l) = a(k, l) + ib(k, l) \tag{1}$$

where, k and l are the frequency indices corresponding to the horizontal x and vertical y directions, with k = 0, 1, 2, …, N-1 and l = 0, 1, 2, …, M-1. The real and imaginary parts of the Fourier spectrum are $a(k, l)$ and $b(k, l)$, respectively. Meanwhile, i is the imaginary unit.

Using Euler's formula, $Z(k, l)$ can be expressed as:

$$Z(k, l) = R(k, l)e^{i\theta(k,l)} \tag{2}$$

where, $R(k, l)$ is the amplitude of the spectrum, representing the contribution of the wave at frequency (k, l) to the random field, and $\theta(k, l)$ is the phase angle, representing the phase shift of the wave. The amplitude and phase angle are calculated as:

$$R(k, l) = \sqrt{\left(a(k, l)\right)^2 + \left(b(k, l)\right)^2} \tag{3}$$

$$\theta(k, l) = arctan\left(\frac{b(k,l)}{a(k,l)}\right) \tag{4}$$

When the random field $Z(x, y)$ in the spatial domain follows a Gaussian distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$, the Fourier spectrum $Z(k, l)$ in the frequency domain is a complex random field. The real $a(k, l)$ and imaginary $b(k, l)$ components at each frequency (except for the case $k = 0, l = 0$) are independent Gaussian random variables with zero mean and unit variance. Consequently, the ratio $b(k, l)/\mathrm{a}(k, l)$ follows a Cauchy distribution, leading to the uniform distribution of phase angle $\theta(k, l)$ over the interval [0, 2π), as presented in [29].

According to the Wiener-Khinchin theorem, the power spectral density *PSD* of the random field is obtained by the Fourier transform of the autocorrelation function $C(x, y)$ in the spatial domain [30-32]. The Gaussian correlation function of a Gaussian random field is expressed as:

$$C(x,y) = e^{-\frac{(x.\Delta x)^2}{\theta_x^2} - \frac{(y.\Delta y)^2}{\theta_y^2}} \quad (5)$$

where, x = 0, 1, 2, …, N-1 and y = 0, 1, 2, …, M-1, θ_x and θ_y are the correlation lengths along the x and y directions. Meanwhile, Δx and Δy are the grid spacings along the x and y directions, respectively. The *PSD* is calculated as:

$$PSD(k,l) = \sum_{x=0}^{N-1}\sum_{y=0}^{M-1} C(x,y)e^{-i2\pi(\frac{k}{N}x+\frac{l}{M}y)} \quad (6)$$

According to the periodogram estimation, the spectral amplitude can be calculated from the power spectrum using the following formula [29, 30]:

$$R(k,l) = \sqrt{PSD(k,l)} \quad (7)$$

Subsequently, the random field in frequency domain is obtained as:

$$Z(k,l) = R(k,l)e^{i2\pi(\frac{k}{N}+\frac{l}{M})} \quad (8)$$

Finally, the random field $Z(x,y)$ is obtained from the complex spectrum by the inverse Fourier transform:

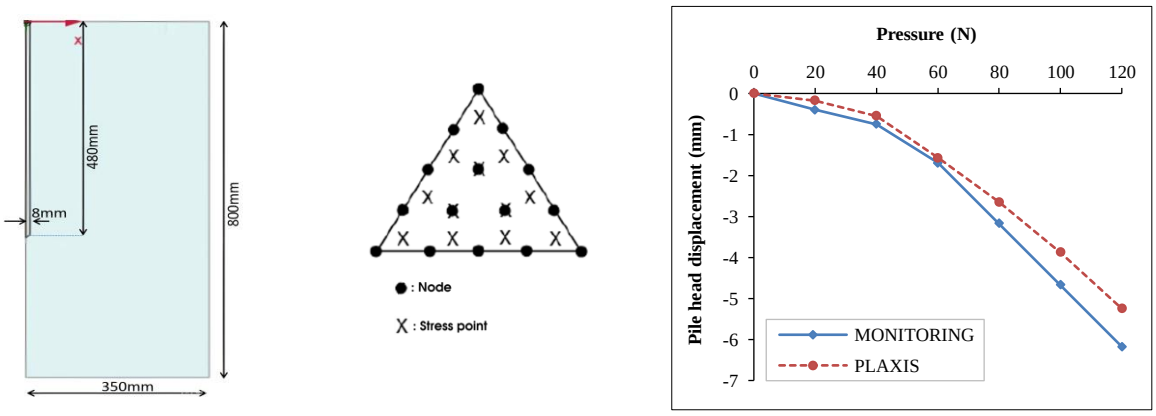
$$Z(x,y) = \frac{1}{NM}\sum_{k=0}^{N-1}\sum_{l=0}^{M-1}\sqrt{PSD(k,l)}e^{i\theta} \quad (9)$$

To ensure the randomness of the field, the phase angle θ is randomly sampled from a uniform distribution over the interval *[0, 2π)*. This phase angle only affects the phase of the frequency components in the complex spectrum $Z(k,l)$, without altering the *PSD*. Consequently, the correlation structure of the random field, determined by *PSD* through the Wiener-Khinchin theorem, remains unaffected by the specific values of the phase angle and thus is fully preserved.

2.2. Experimental and Computational Model

According to the study by Bach et al. [26], a scaled physical model with a ratio of 1:100 of a single pile was established in the laboratory. The pile was made of aluminum, having a circular cross-section with a diameter of 16 mm, and a depth of 480 mm. The container had dimensions of 700 × 700 × 800 mm (length-width-height), filled with soft clay to simulate the natural ground conditions. The clay was prepared using the standard Proctor compaction method with a natural moisture content ranging from 49% to 52% and unit weight ranging from 15.6 to 16.1 kN/m³. A static pile loading test was conducted to evaluate the displacement responses of the single pile under six levels of axial compressive loads: 20 N, 40 N, 60 N, 80 N, 100 N and 120 N.

To simulate the single pile, an axisymmetric model was employed in PLAXIS 2D, as shown in Figure 1(a). This is one of the features supported by the software to model problems with circular symmetry about the vertical axis, such as single piles, caissons, and circular foundations. In the model, the computational domain was discretized using 15-node triangular elements, as shown in Figure 1(b). The soil behaviors were governed by Hardening Soil model, with the properties presented in Table 1. Meanwhile, the pile material was assumed to be linear elastic with properties shown in Table 2. The static loading process was simulated using six different loading increments. As shown in Figure 1(c), the simulated pile head displacements exhibited a trend consistent with the observational measurements, thereby confirming the reliability and suitability of the numerical model.



(a) (b) (c)

Figure 1. (a) Axisymmetric model of a single pile in PLAXIS 2D, (b) 15-node triangular finite element, (c) pile head displacement

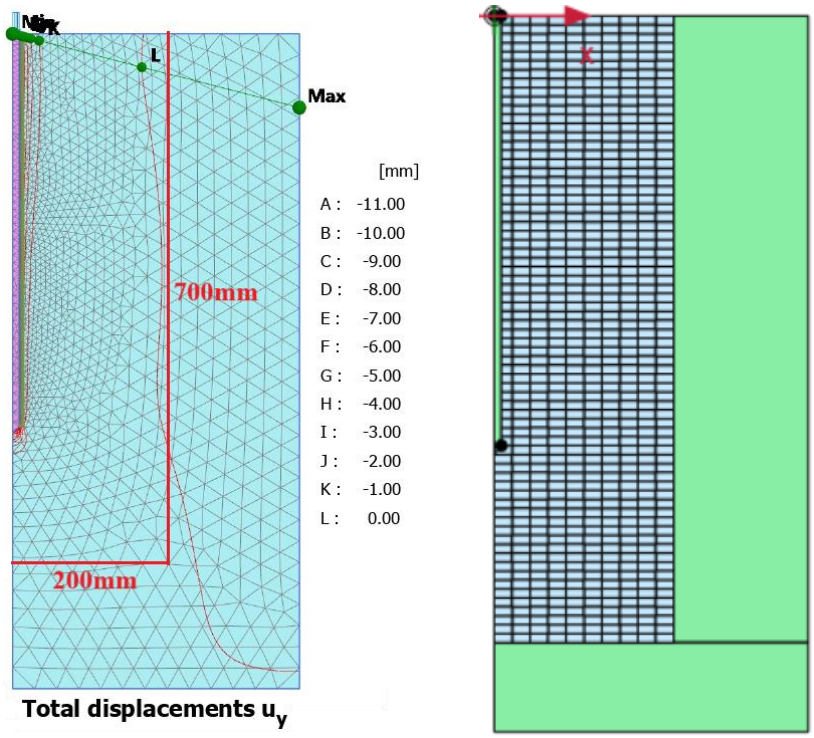
Table 1. Hardening Soil model parameters of the soil

Input parameters	Physical meaning	Unit	Values
γ_{sat}	Saturated unit weight	kN/m^3	15.9
γ_{unsat}	Unsaturated unit weight	kN/m^3	16.8
e_{init}	Initial void ratio	-	2.149
c'_{ref}	Effective reference cohesion	kPa	8.5
φ'	Effective friction angle	degree	19.3
C_c	Compression index	-	0.674
C_s	Swelling index	-	0.253
υ_{ur}	Unloading Poisson's ratio	-	0.15
R_{inter}	Interface reduction factor	-	0.5
OCR	Over consolidation ratio	-	1.5

Table 2. Linear Elastic model parameters of the pile

Input parameters	Physical meaning	Unit	Values
γ	Unit weight	kN/m^3	19.54
E_{ref}	Young's modulus	GPa	95
υ	Poisson's ratio	-	0.15

To reduce the computational time while preserving the effect of spatial randomness, the random field was generated only within the soil region directly influenced by the pile. Outside this zone, the soil properties were assumed to be constant and equal to the mean value of the interested parameter. A preliminary investigation was initially conducted to determine the pile's influence zone based on soil displacement. The results indicate that the influence zone can be defined by a radius of 200 mm around the pile and a depth of 700 mm, as shown in Figure 2(a). Within the influence zone, the random field was generated by discretizing the soil domain into material clusters with a mesh size of 20 × 10 mm, where each element is assigned a distinct value of material property corresponding to random field, as shown in Figure 2(b).



(a) (b)

Figure 2. (a) Influence zone based on pile-induced displacement, (b) Discretization of the influence zone.

Gitirana et al. [27] showed that the coefficient of variation *COV* of effective cohesion *c'* ranges from 10% to 50%. Meanwhile, using the cone penetration test results, Akkaya et al. [28] reported horizontal correlation lengths ranging from 2.0 m to 14.0 m (average of 8.0 m), and vertical correlation lengths ranging from 0.26 m to 3.62 m (average of

1.94 m) for clay soils. Accordingly, in this study, the horizontal and vertical correlation lengths (θ_x and θ_y) were taken as 8.0 m and 1.94 m, respectively, while the *COV* of 10% was selected. With a scale of 1:100, the corresponding correlation lengths in the model are 80 mm and 19.4 mm. Based on the selected parameters, the random field of *c'* was generated.

To assess the distribution of the sample dataset, the study employs a two-step verification: first, constructing probability distribution plots to enable visual observation and identification of the data's distribution trends. Subsequently, for a more in-depth and quantitative evaluation, Q-Q plots are utilized to compare the experimental data points against the theoretical distribution, thereby determining the degree of conformity. Q-Q plot is a graphical tool used to compare the quantiles of experimental data with those of a theoretical distribution (e.g., the normal distribution). If the data conform to the theoretical distribution, the points on the plot will approximately lie along a straight line. Conversely, deviations from the straight line indicate that the data do not fully adhere to the assumed distribution, facilitating the quantification of the degree of discrepancy and the identification of features such as distributional tails or outliers.

As shown in Figure 3, the statistics after the generation process confirm that the field values follow a normal distribution with $R^2 > 0.99$, thereby illustrating the consistency and rationality of the simulation model.

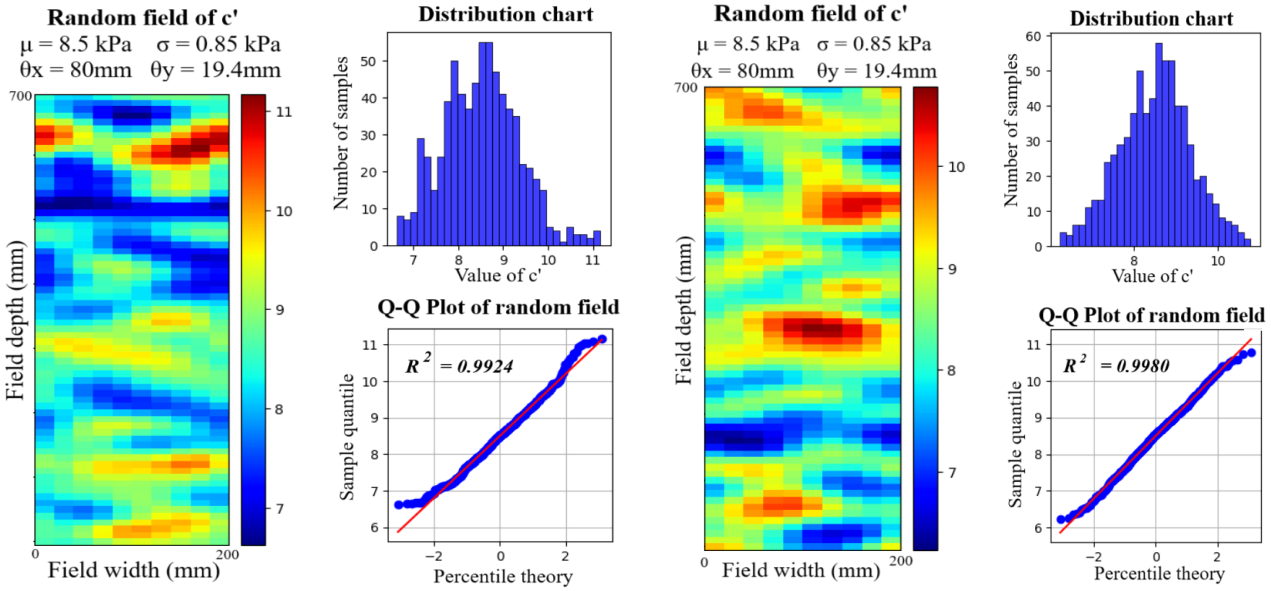


Figure 3. Two realizations of the random field and their corresponding probability distribution histograms

To evaluate the impact of soil heterogeneity on the responses of pile, the Monte Carlo method was applied with 1000 realizations to obtain statistically reliable data.

2.3. An Automated Solution for the Computational Process

In the Monte Carlo process with a large number of simulations, manually updating the values of random variables after each simulation would require an excessive amount of time and effort. Although PLAXIS is not an open-source software, it provides a Python-based interface through PLAXIS API for the users to programmatically interact with it. The use of PLAXIS API enables full automation of the entire process, from model generation to result extraction.

As shown in Figure 4, the automation process consists of the following steps:

(i) Launching PLAXIS Input and PLAXIS Output;

(ii) Initializing Python environment and establishing a connection to PLAXIS Input and Output via PLAXIS API;

(iii) Executing the calculations and extracting results using Python scripting

(iii-a) Defining the computational model;

(iii-b) Generating random field values and assigning them to the model;

(iii-c) Executing the numerical simulation and exporting the corresponding pile head displacement results (U_y) for each realization into text files.

The process from step (iii-b) to (iii-c) is automatically repeated for the desired number of realizations.

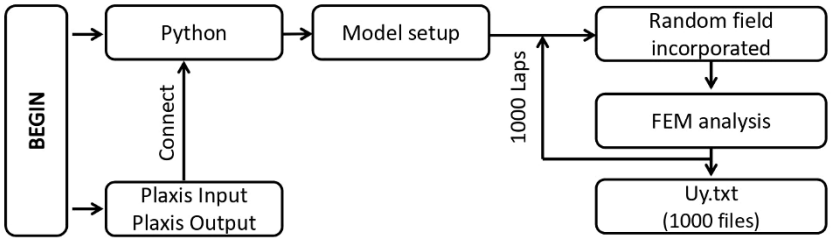
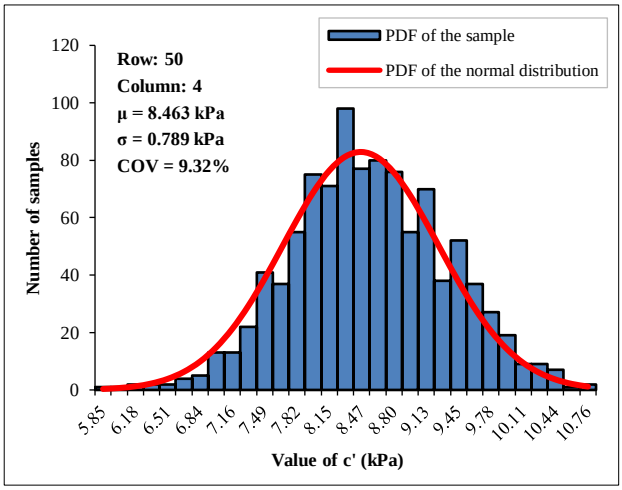


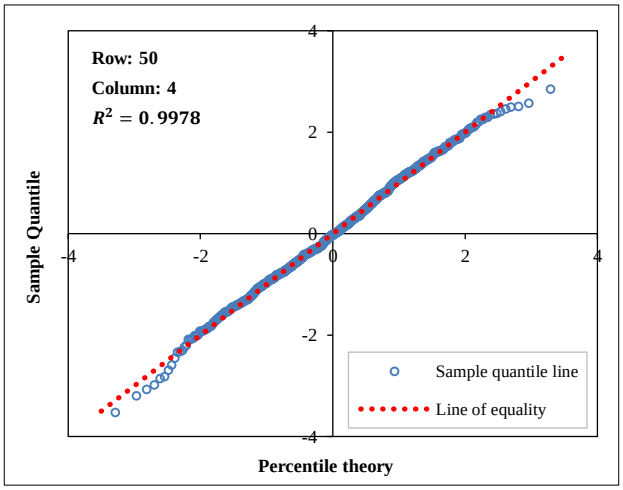
Figure 4. Flowchart of the computational process

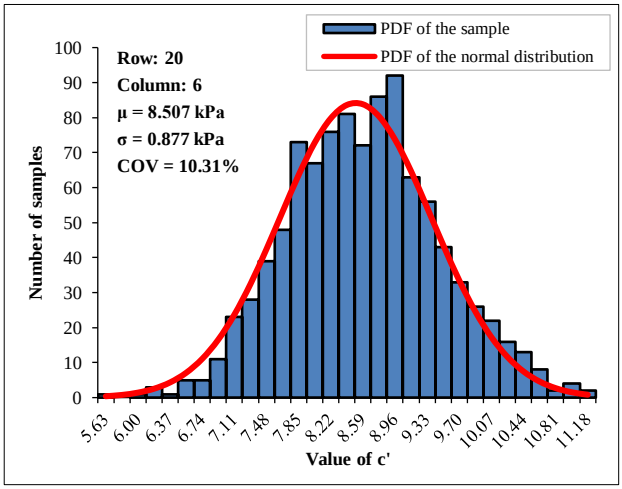
3. Results and Discussion

3.1. Verification of the Distribution of Effective Cohesion

The values of *c'* at two elements in the mesh grid of Figure 2(b) were recorded for 1000 realizations. Figures 5 and 6 indicate that the values of *c'* follow a normal distribution.







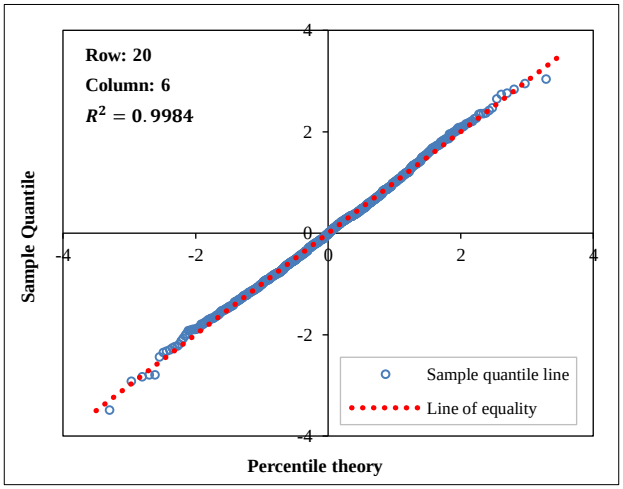


Figure 5. Probability distribution plot of *c'* at two positions in the element mesh grid

Figure 6. Q-Q plot of *c'* at two positions in the element mesh grid

3.2. Probability Distribution of Pile Displacement

As shown in Figure 7, the displacement spectrum obtained from the random field realizations fully covers the range of variation compared with the deterministic model using the mean value of *c'*, demonstrating that the random assignment effectively captures the inherent natural variability of the soil, as previously mentioned in the literature. Similar results were observed in the studies by Dong et al [33] and Wu et al [34]. The proportion of pile head displacement values that are smaller or larger than the displacement corresponding to the mean value of c' exhibits variation across different levels of applied compressive load. This observation indicates a load-dependent sensitivity in the pile-soil system's performance, even when the identical underlying soil random field is used, as shown in Figure 8.

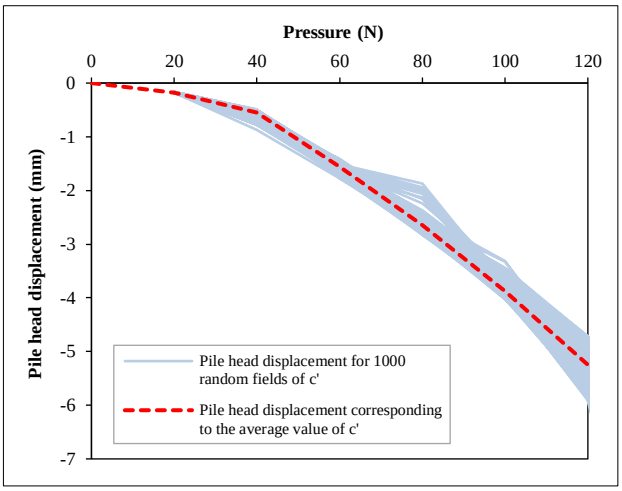


Figure 7. Pile displacement spectrum

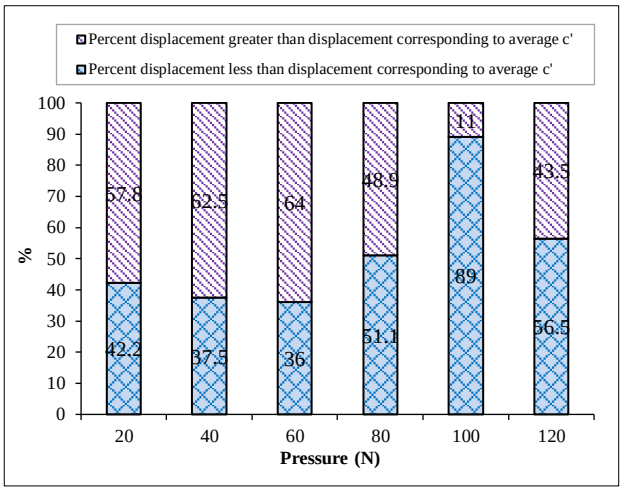
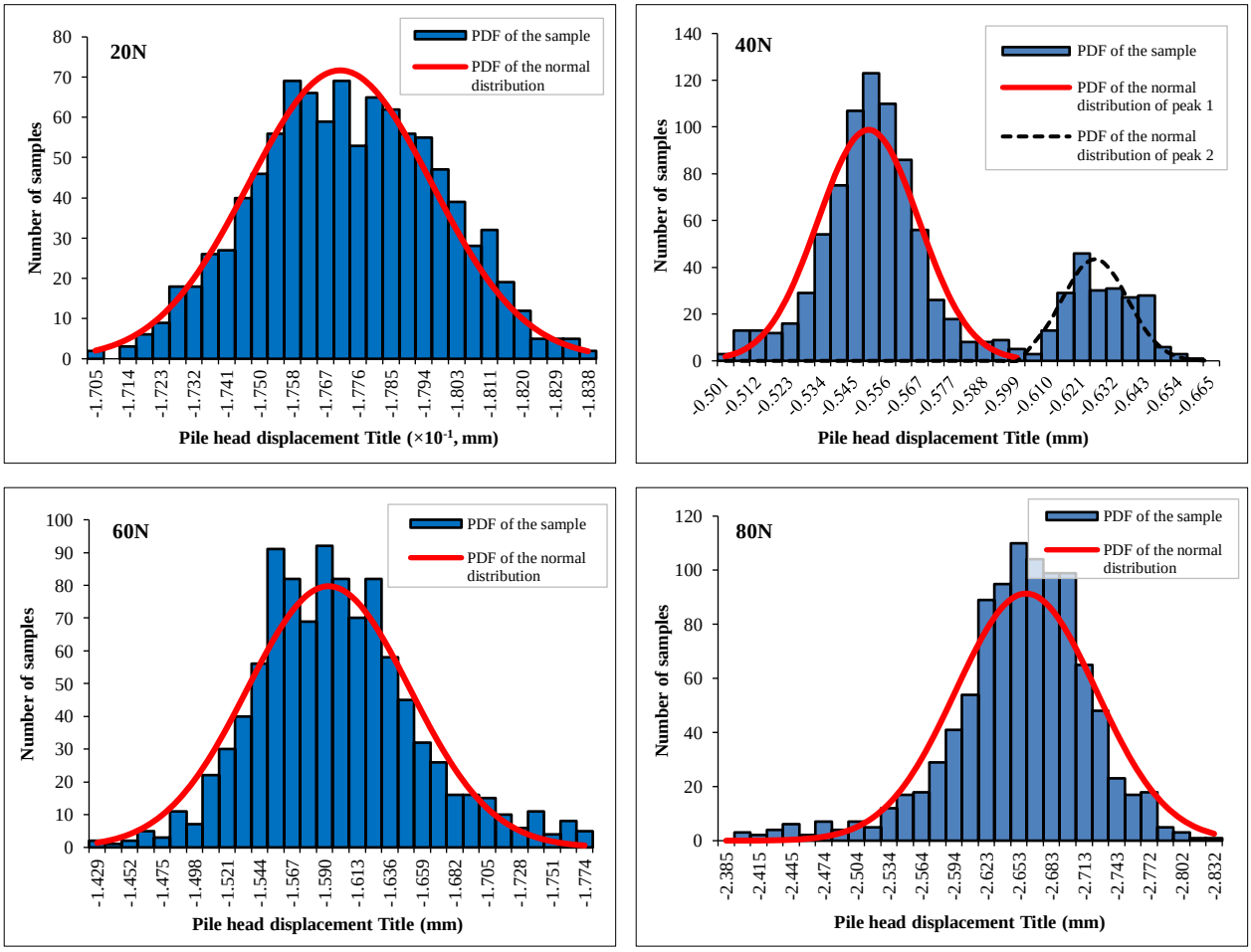


Figure 8. Proportion of displacement greater and less than the displacement corresponding to average c'

Based on the pile head displacement results obtained at compressive load levels of 20N, 40N, 60N, 80N, 100N, and 120N from 1000 realizations, the distribution plots of the pile head displacement values at each compressive load level were constructed, as shown in Figure 9. Meanwhile, Figure 10 shows the corresponding Q-Q plots to evaluate whether these datasets conform to a normal distribution, and if so, the degree of conformity to the normal distribution.

The displacements under different compression loading levels tend to follow a normal distribution with $R^2 > 0.9$, as shown in Figures 9 and 10. However, at the 40 N loading level, the distribution exhibits a bimodal pattern. This is reasonable, as it marks the onset of the transition from elastic to plastic deformation in soil behavior. This transition is shown in the plot of pile head displacement versus static compression loading levels in Figure 1(c), where at a compression load of 40 N, the pile displacement begins to increase at a significantly faster rate than at previous loading levels. Thus, in some realizations the soil remains elastic, while in others it experiences plastic deformation, as shown in Figure 11.



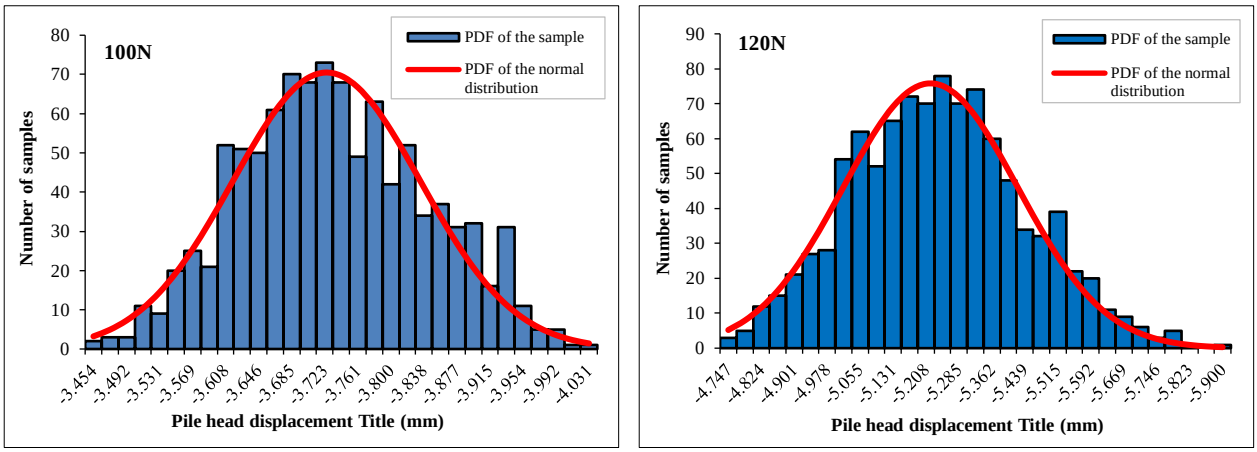


Figure 9. Probability distribution of pile displacement

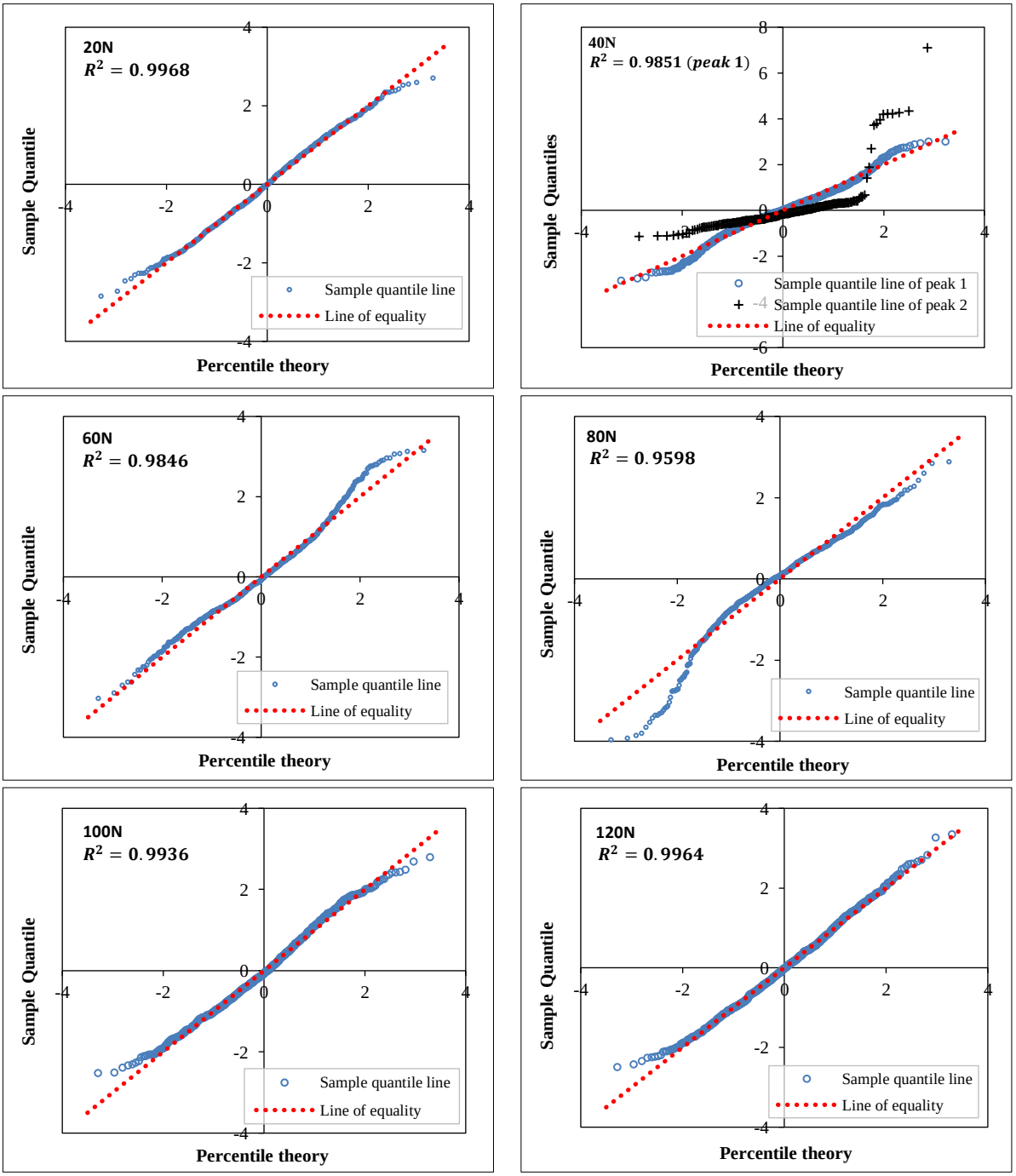


Figure 10. Q-Q plots of pile displacement

The ultimate bearing capacity of the pile, defined at a displacement being equal to 10% of the pile diameter (equivalent to 1.6 mm), was determined by linear interpolation between the two compression loading levels of 60 N and 80 N, yielding a value of 60.57 N. At this loading level, 63% of the random field realizations induces displacements greater than 1.6 mm, while 37% causes displacements less than 1.6 mm, as shown in Figure 12. The statistical results of pile head displacement at the loading level of 60.57 N indicate that this quantity tends to follow a normal distribution, as shown in Figure 13. This observation is further confirmed by the Q-Q plots in Figure 14, with a coefficient of determination R^2=0.9862, reflecting a very high degree of agreement between the experimental data and the assumed normal distribution.

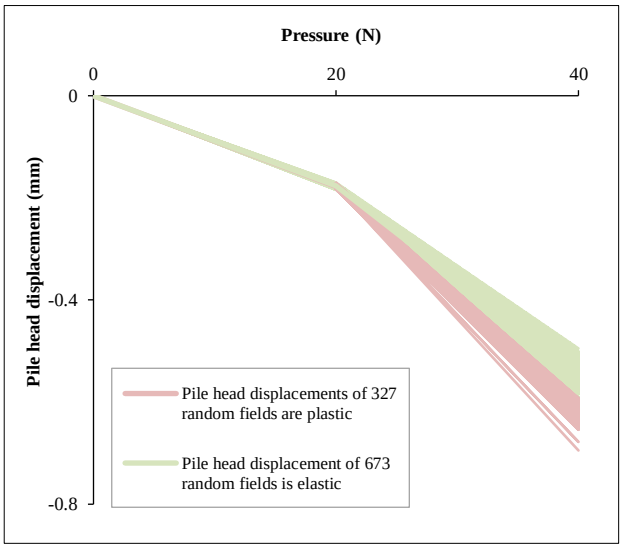


Figure 11. Pile displacement spectrum at 20N and 40N

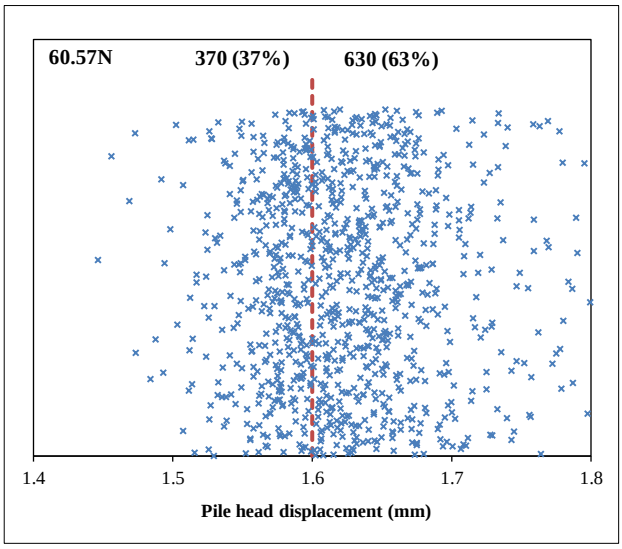


Figure 12. Pile displacement at 60.57N (symbol: x)

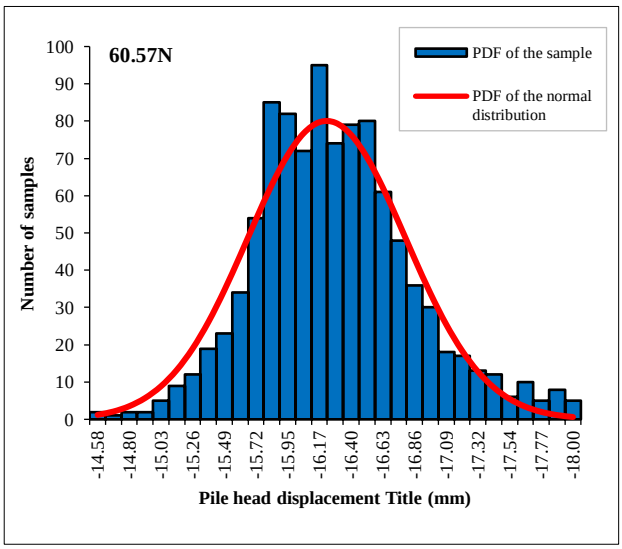


Figure 13. Probability distribution of pile displacement at compressive load level of 60.75 N

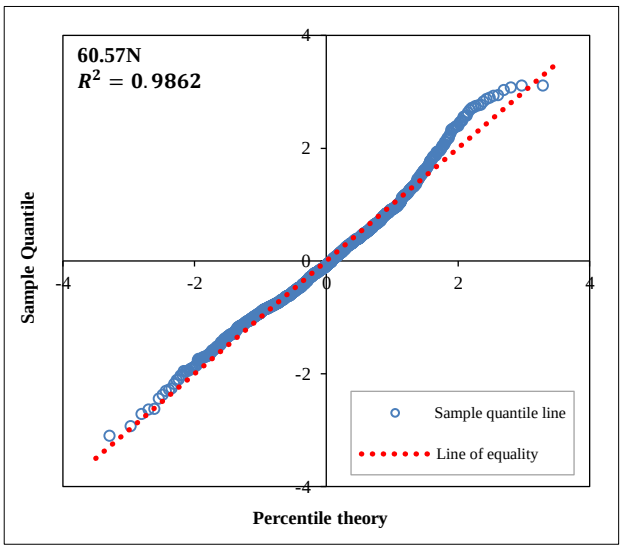


Figure 14. Q-Q plots of pile displacement at compressive load level of 60.75 N

The cumulative distribution function (CDF) for the pile head displacement derived from 1000 Monte Carlo simulation samples was compared against an assumed normal distribution with the mean value and the coefficient of variation being determined directly from the 1000 sample results. As shown in Figure 15, the two CDF curves exhibit a high degree of similarity, confirming that the normal distribution provides a robust and suitable model for characterizing the uncertainty in pile displacement. The observed discrepancy between the empirical CDF and the fitted normal CDF was restricted to a narrow range, specifically 0.95% to 5.68%, as detailed in Figure 15. In addition, the coefficient of variation *COV* of pile head displacement at different loading levels ranges from 1.39% to 7.41%, which is less than the *COV* of 10% of the input parameter c', as illustrated in Figure 16.

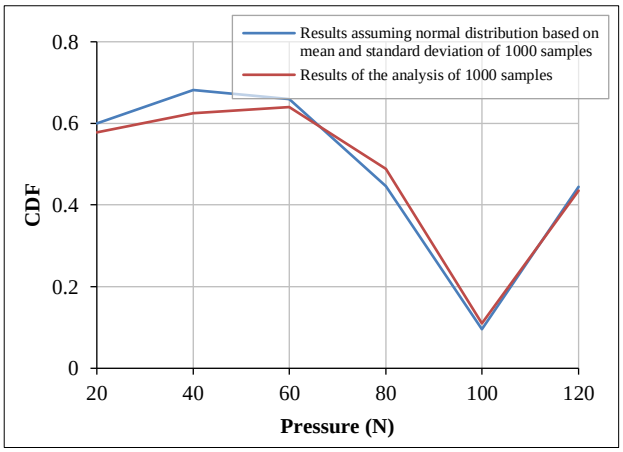


Figure 15. The observed CDF and the CDF of the normal distribution for pile head displacement

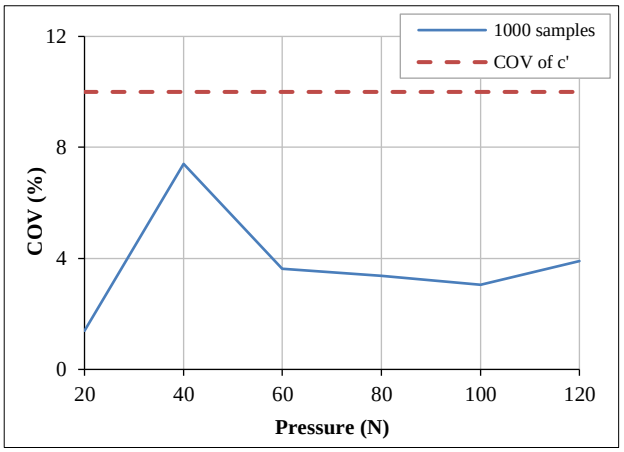
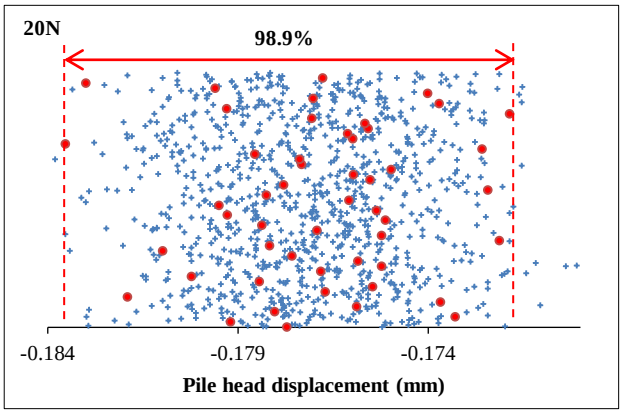


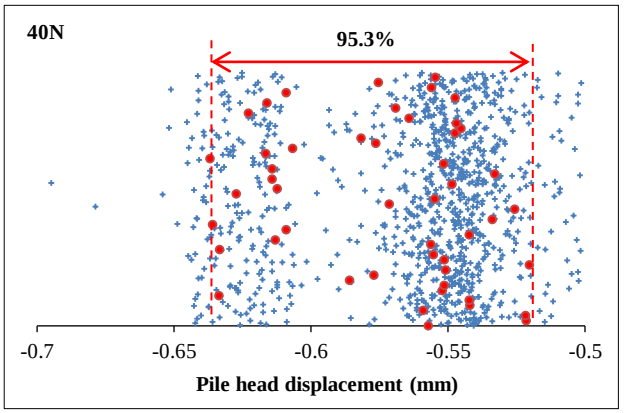
Figure 16. COV of pile head displacement at various compressive load levels

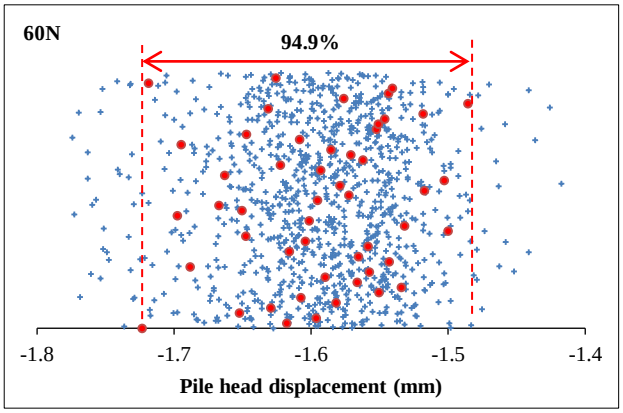
3.3. Reduction of Simulations Using the K-Means Clustering Method

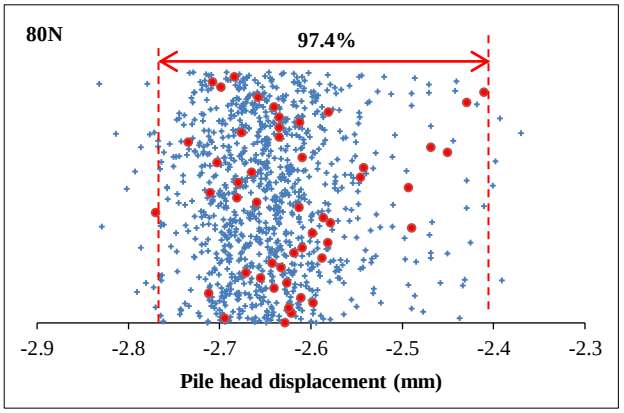
After generating 1000 realizations of the random field, this study applies the K-means clustering algorithm to group and select representative subsets at three scales: 50 samples, 100 samples, and 200 samples (corresponding to 5%, 10%, and 20% of the original dataset, respectively). The objective is to evaluate the representativeness of these clusters in terms of pile head displacement and to compare their responses with those obtained from the full set of 1000 realizations. To visualize the differences, plots are constructed to illustrate the pile head displacement points within each cluster at each loading level, while simultaneously comparing them with the displacement distribution obtained from the full set of realizations. Figures 17, 18, and 19 present the 50 samples cluster, 100 samples cluster, and 200 samples cluster, respectively.

The analysis of pile displacement using representative realizations selected through the K-means clustering method demonstrates a high level of coverage compared with the full dataset. Specifically, with 50 representative samples, the displacement range covers from 94.9% to 98.9% of the full sample set, as shown in Figure 17. When the number of representative cases increases to 100, the coverage improves significantly, ranging between 98.2% and 99.6%, as shown in Figure 18. With 200 representative realizations as in Figure 19, the results nearly converge to the maximum possible coverage, ranging from 98.6% to 99.7% of the entire displacement dataset. It should be noticed that for the bimodal distribution in the 40 N loading case, there are two values of means μ and *COV* corresponding to the elastic and plastic states of the soil displacement.









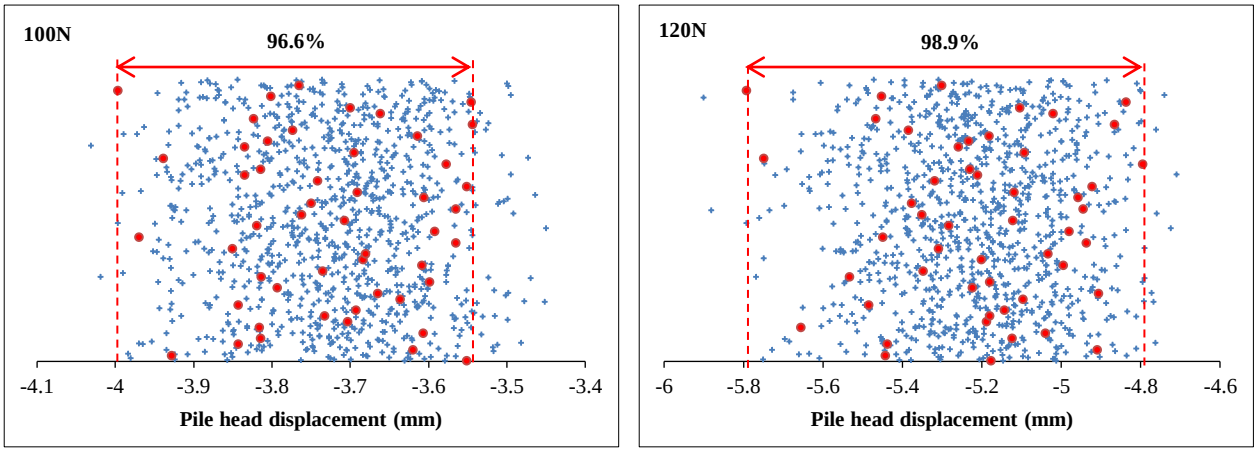


Figure 17. Displacement of pile heads of 50 samples (symbol: o) compared with total 1000 samples (symbol: +)

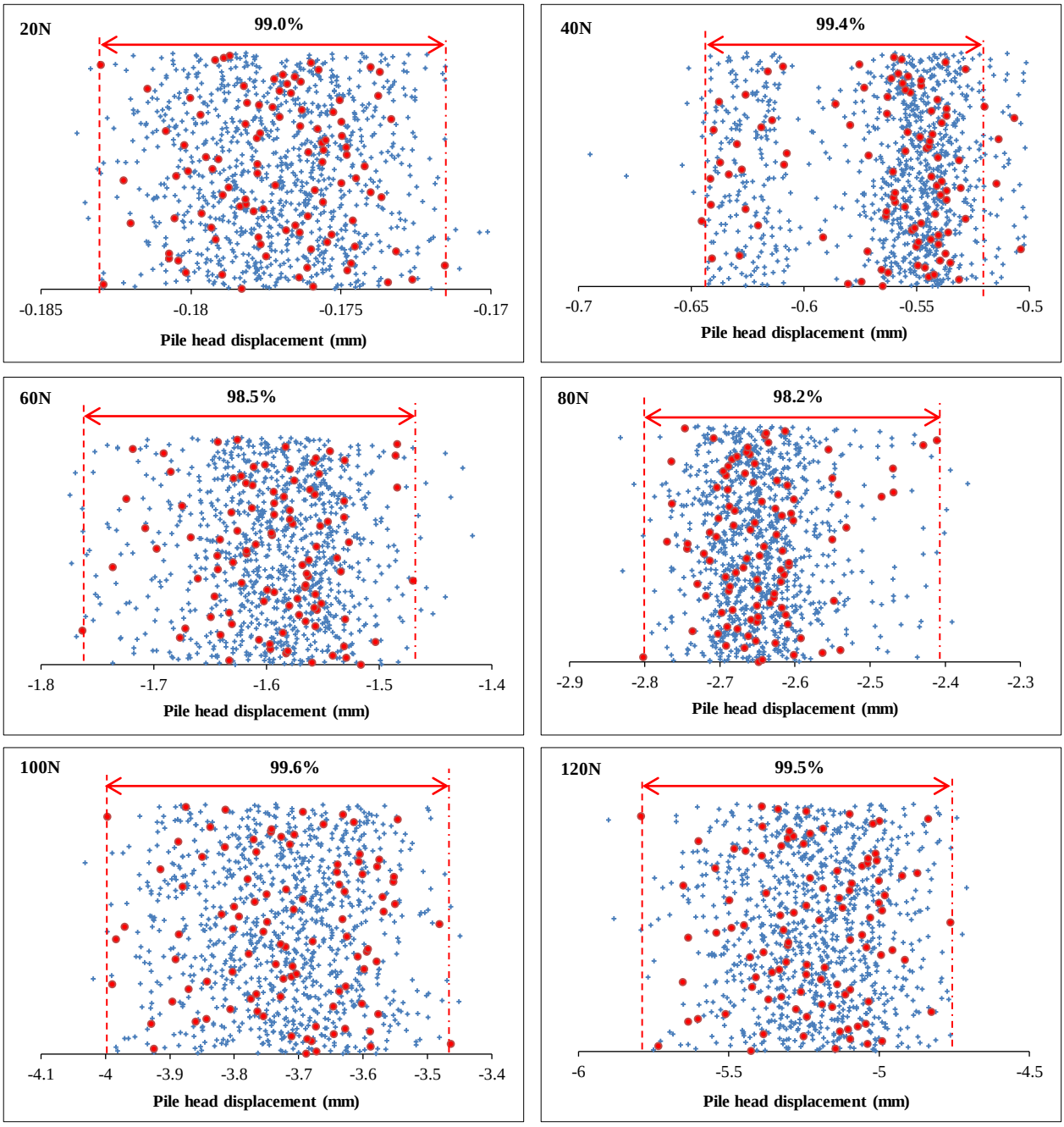


Figure 18. Displacement of pile heads of 100 samples (symbol: o) compared with total 1000 samples (symbol: +)

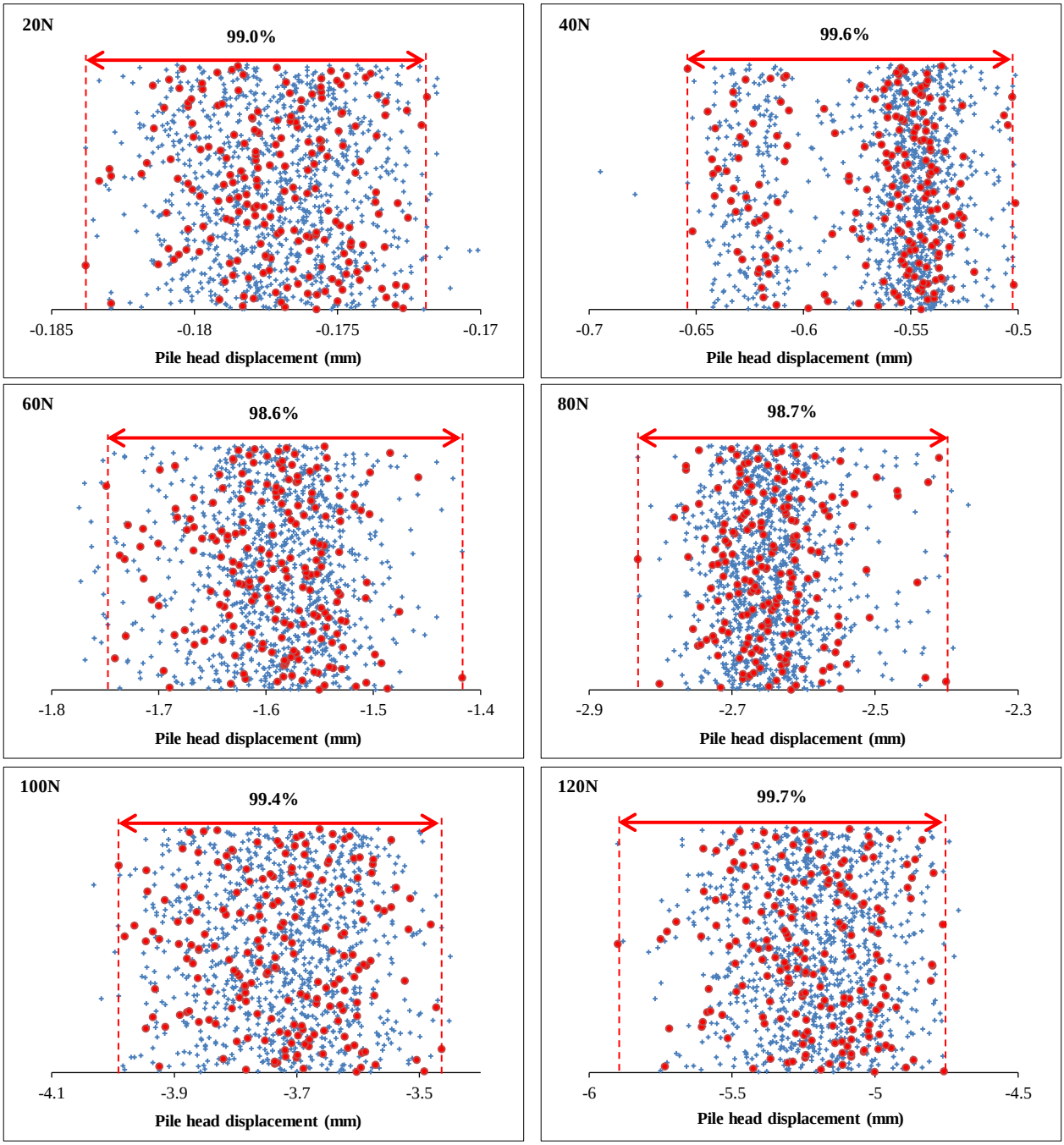


Figure 19. Displacement of pile heads of 200 samples (symbol: o) compared with total 1000 samples (symbol: +)

When computing the weighted mean and coefficient of variation COV for each selected sample group, the results indicate that as the number of representative realizations increases, both the mean and COV of the group approach the corresponding values of the full set of 1000 realizations. At the same time, the variability of these quantities becomes more stable across different loading levels, as shown in Figures 20 and 21.

The relative deviation in the mean pile head displacement compared to a reference value, i.e., the mean of the 1000-sample dataset, exhibits a clear inverse relationship with the sample size N of the K-means subset used for statistical estimation. As shown in Figure 20, for N = 50 samples, the relative deviation of the mean ranges significantly, from 0.1% to 8.8%. When N = 100 samples, the deviation range narrows substantially, from 0.1% to 3.5%. When N = 200 samples, the deviation is further reduced, falling within a tight range of 0.3% to 1.2%. Similar trend is observed for the case of COV, as shown in Figure 21, with the corresponding figures for 50, 100, and 200 samples being 5.6% to 233.5%, 11.6% to 72.9%, and 1.2% to 35.6%, respectively. This consistent trend confirms the principle of statistical convergence, where increasing the sample size provides a more stable and accurate estimate of the true population mean and variance, leading to smaller estimation errors. This suggests that N = 200 samples provide a more reliable representation of the system's expected performance compared to smaller datasets.

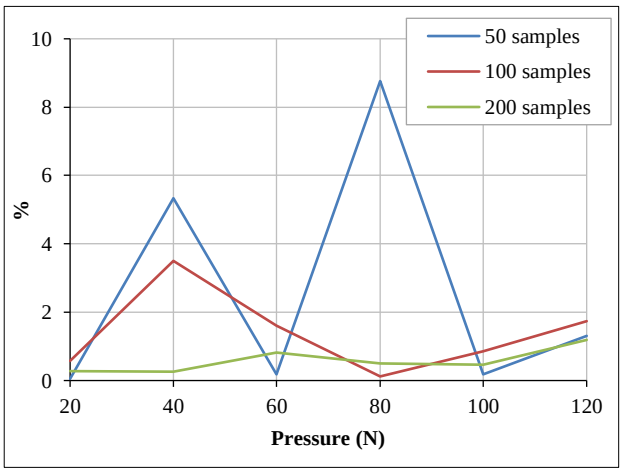


Figure 20. Percentage change of group means relative to the 1000 samples mean

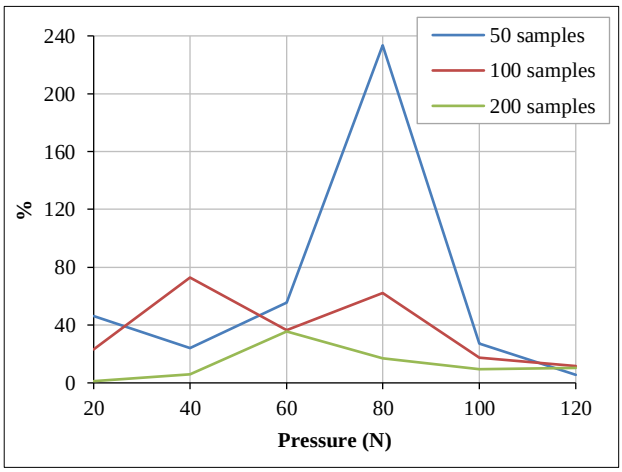


Figure 21. Percentage change of group COV relative to the 1000 samples COV

4. Conclusions

This study investigates the influence of 2D correlated random field of effective cohesion c' on the displacement of an axially loaded pile, using FEM model combined with the Monte Carlo simulations. The spatial heterogeneity of the effective cohesion was initially generated by the spectral method, following a Gaussian distribution with a COV of 10% and Gaussian spatial correlation lengths of 80 mm and 19.4 mm in the horizontal and vertical directions. The values of *c'* at each spatial point in 1000 realizations were verified to follow a normal distribution with high confidence, thereby confirming the reliability and appropriateness of the random field generation process. Subsequently, the random fields were mapped to the finite elements and the static pile loading process was numerically simulated with six loading stages ranging from 20 N to 120 N. The effect of soil spatial variability on the pile head displacement was analyzed, and the application of K-means approach to reduce the necessary number of realizations was also discussed. The following conclusions can be drawn:

- The pile displacements under various loading levels generally show a tendency toward normal distribution with $R^2 > 0.95$, except at the 40 N loading level, where a bimodal distribution is observed.
- The loading level of 40 N reflects a clear transition from the elastic to the plastic state of the surrounding soil. However, the 40 N threshold considered as the boundary between these two behavior regimes is not applicable to all cases, as 23.7% of the realizations reach the plastic deformation stage prior to this loading level. This finding highlights the need to account for the inherent uncertainty and spatial heterogeneity of soil properties when defining the serviceability limit state of the ground.
- The ultimate bearing capacity calculated using the mean value of c′ yielded a result of 60.57 N, calculated at the displacement of 1.6 mm. However, only 37% of the random realizations exhibit displacements smaller than the 1.6 mm threshold, indicating a low safety probability when relying solely on the mean value. This highlights the limitation of deterministic approaches and underscores the necessity of incorporating probabilistic analyses for a more reliable assessment of foundation performance under spatially variable soils.
- The application of the K-means method to select representative fields for the input parameter c′ yields highly encouraging results in terms of the corresponding pile-head displacement output. The findings indicate that representative fields of c′ also produce a similarly comprehensive distribution of output responses, despite the nonlinear and complex nature of the problem. Specifically, 50 representative fields are able to cover more than 94% of the output domain. This demonstrates that the K-means method, when used to select representative input samples, also leads to highly representative output responses.

The study's scope is limited to a single case concerning the correlation length and coefficient of variation, necessitating future research across multiple parameters to assess the consistency of output responses. Furthermore, the pile model simulation was conducted at a small physical scale of 1:100, thus, subsequent investigations on full-scale piles are essential. This helps to validate simulation outcomes against actual monitoring data, thereby enhancing the reliability and applicability of the developed method. Finally, the reliance on the K-means clustering method introduces a known vulnerability to outliers with large deviations, which suggests the need for developing superior K-means sampling strategies or integrating hybrid methodologies to achieve improved sample coverage.

5. Declarations

5.1. Author Contributions

Conceptualization, D.V.L. and K.T.N.; methodology, D.V.L. and K.T.N.; software, D.V.L.; validation, D.V.L.; formal analysis, D.V.L. and K.T.N.; investigation, D.V.L.; writing—original draft preparation, D.V.L.; writing—review and editing, K.T.N.; visualization, D.V.L.; supervision, K.T.N. All authors have read and agreed to the published version of the manuscript.

5.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

5.3. Funding and Acknowledgments

We acknowledge Ho Chi Minh City University of Technology (HCMUT), VNU-HCM for supporting this study.

5.4. Conflicts of Interest

The authors declare no conflict of interest.

6. References

[1] Tang, X. S., Wang, M. X., & Li, D. Q. (2020). Modeling multivariate cross-correlated geotechnical random fields using vine copulas for slope reliability analysis. Computers and Geotechnics, 127, 103784. doi:10.1016/j.compgeo.2020.103784.

[2] Chen, X., Hu, Y., Cheng, P., Fang, P. P., & Yao, K. (2025). Effect of soil arching evolution on tunnel face stability considering spatially variable sandy soils. Transportation Geotechnics, 51, 101496. doi:10.1016/j.trgeo.2025.101496.

[3] Mert, A. C., Yazıcı, G., & Khanbabazadeh, H. (2024). Reliability Assessment of Strip Footings on Spatially Varying Sand. KSCE Journal of Civil Engineering, 28(10), 4258–4271. doi:10.1007/s12205-024-1260-8.

[4] Liu, Y., Li, J., Sun, S., & Yu, B. (2019). Advances in Gaussian random field generation: a review. Computational Geosciences, 23(5), 1011–1047. doi:10.1007/s10596-019-09867-y.

[5] Mejía, J. M., & Rodríguez-Iturbe, I. (1974). On the synthesis of random field sampling from the spectrum: An application to the generation of hydrologic spatial processes. Water Resources Research, 10(4), 705–711. doi:10.1029/WR010i004p00705.

[6] Shinozuka, M. (1971). Simulation of Multivariate and Multidimensional Random Processes. The Journal of the Acoustical Society of America, 49(1B), 357–368. doi:10.1121/1.1912338.

[7] Shinozuka, M., & Deodatis, G. (1996). Simulation of multi-dimensional Gaussian stochastic fields by spectral representation. Applied Mechanics Reviews, 49(1), 29–53. doi:10.1115/1.3101883.

[8] Shinozuka, M., & Jan, C. M. (1972). Digital simulation of random processes and its applications. Journal of Sound and Vibration, 25(1), 111–128. doi:10.1016/0022-460X(72)90600-1.

[9] Gutjahr, A., Bullard, B., & Hatch, S. (1997). General joint conditional simulations using a fast Fourier transform method. Mathematical Geology, 29(3), 361–389. doi:10.1007/BF02769641.

[10] Ruan, F., & McLaughlin, D. (1998). An efficient multivariate random field generator using the fast Fourier transform. Advances in Water Resources, 21(5), 385–399. doi:10.1016/S0309-1708(96)00064-4.

[11] Mai, P. M., & Beroza, G. C. (2002). A spatial random field model to characterize complexity in earthquake slip. Journal of Geophysical Research: Solid Earth, 107(B11). doi:10.1029/2001jb000588.

[12] Jin, S., Lutes, L. D., & Sarkani, S. (1997). Efficient Simulation of Multidimensional Random Fields. Journal of Engineering Mechanics, 123(10), 1082–1089. doi:10.1061/(asce)0733-9399(1997)123:10(1082).

[13] Luo, N., & Luo, Z. (2021). Reliability Analysis of Embedded Strip Footings in Rotated Anisotropic Random Fields. Computers and Geotechnics, 138, 104338. doi:10.1016/j.compgeo.2021.104338.

[14] Räss, L., Kolyukhin, D., & Minakov, A. (2019). Efficient parallel random field generator for large 3-D geophysical problems. Computers & Geosciences, 131, 158–169. doi:10.1016/j.cageo.2019.06.007.

[15] Robin, M. J. L., Gutjahr, A. L., Sudicky, E. A., & Wilson, J. L. (1993). Cross-correlated random field generation with the direct Fourier Transform Method. Water Resources Research, 29(7), 2385–2397. doi:10.1029/93WR00386.

[16] Jain, A. K. (2010). Data clustering: 50 years beyond K-means. Pattern Recognition Letters, 31(8), 651–666. doi:10.1016/j.patrec.2009.09.011.

[17] Xiao, W., Shen, Y., Zhao, J., Lv, L., Chen, J., & Zhao, W. (2025). An Adaptive Multi-Fidelity Surrogate Model for Uncertainty Propagation Analysis. Applied Sciences (Switzerland), 15(6), 3359. doi:10.3390/app15063359.

[18] Khorramian, K., Alhashmi, A.E., & Oudah, F. (2024). Effect of Random Fields in Stochastic FEM on the Structural Reliability Assessment of Pile Groups in Soil. Proceedings of the Canadian Society of Civil Engineering Annual Conference 2022. CSCE 2022. Lecture Notes in Civil Engineering, vol 367. Springer, Cham, Switzerland. doi:10.1007/978-3-031-35471-7_10.

[19] Liu, H., Zhu, D., Bian, X., Wang, Z., & Gu, J. (2025). Analysis of the Vertical Bearing Capacity of Pile Foundations in Backfill Soil Areas Based on Non-Stationary Random Field. Buildings, 15(8), 1314. doi:10.3390/buildings15081314.

[20] Xiao-ling, Z., Bo-han, J., Yan, H., Shong-loong, C., & Xiu-yu, L. (2021). Random field model of soil parameters and the application in reliability analysis of laterally loaded pile. Soil Dynamics and Earthquake Engineering, 147, 106821. doi:10.1016/j.soildyn.2021.106821.

[21] Christodoulou, P., Pantelidis, L., & Gravanis, E. (2020). The Effect of Targeted Field Investigation on the Reliability of Axially Loaded Piles: A Random Field Approach. Geosciences, 10(5), 160. doi:10.3390/geosciences10050160.

[22] Tan, X., Zhang, P., Dong, X., Lin, X., & Lu, Z. (2024). Reliability Analysis of Single Pile in Spatially Variable Soil Based on Variance Reduction Method. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering, 10(2), 4024016. doi:10.1061/ajrua6.rueng-1183.

[23] Gan, Y., Li, Y., Zhu, H., & Zhang, B. (2026). Probabilistic analysis of top deflection of laterally loaded piles considering spatially variable soil stiffness. Computers and Geotechnics, 189, 107592. doi:10.1016/j.compgeo.2025.107592.

[24] Duan, W., Zhao, Z., Cai, G., Zou, H., Chu, Y., Zhang, X., Pu, S., Sun, Y., Chen, R., & Liu, S. (2026). A new CPTU-informed random field approach for probabilistic analysis of vertically loaded piles. Transportation Geotechnics, 56, 101795. doi:10.1016/j.trgeo.2025.101795.

[25] Xu, J., Tan, X., Liu, S., Jiang, Z., & Lu, Z. (2025). Reliability Analysis of Vertically Loaded Piles Using Conditional Random Fields. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering, 11(4). doi:10.1061/ajrua6.rueng-1592.

[26] Bach, L. V. H. & Tran, T. T. (2016). Investigation of pile group effect through static compression tests on scaled physical models. Journal of Construction, no. 6/2016, 191-194. (In Vietnamese).

[27] Gitirana, G. de F. N., & Fredlund, D. G. (2016). Statistical Assessment of Hydraulic Properties of Unsaturated Soils. Soils and Rocks, 39(1), 81–95. doi:10.28927/sr.391081.

[28] Akkaya, A. (Dener), & Vanmarcke, E. H. (2003). Estimation of Spatial Correlation of Soil Parameters Based on Data from the Texas A &M University NGES. Probabilistic Site Characterization at the National Geotechnical Experimentation Sites, 29–40, American Society of Civil Engineers (ASCE), Reston, United States. doi:10.1061/9780784406694.ch03.

[29] Wiener, N. (1930). Generalized harmonic analysis. Acta Mathematica, 55(1), 117–258. doi:10.1007/BF02546511.

[30] Khintchine, A. (1934). Korrelationstheorie der stationären stochastischen Prozesse. Mathematische Annalen, 109(1), 604–615. doi:10.1007/BF01449156.

[31] Oppenheim, A. V, & W.Schafer, R. (1999). Discrete -Time Signal Processing. Prentice Hall, Upper Saddle River, United States.

[32] Bendat, J. S., & Piersol, A. G. (2010). Random Data. Wiley Series in Probability and Statistics, John Wiley & Sons, Hoboken, United States. doi:10.1002/9781118032428.

[33] Dong, X., Tan, X., Lin, X., Guo, W., Zha, F., & Xu, L. (2023). Reliability Analysis and Design of Vertically Loaded Piles in Spatially Variable Soils. International Journal of Geomechanics, 23(10), 4023175. doi:10.1061/ijgnai.gmeng-8426.

[34] Wu, T., Sun, Z., Tan, W., Kasu, C. M., & Gong, J. (2022). Response of vertically-loaded pile in spatially-variable cement-treated soil. PLOS ONE, 17(4), e0266975. doi:10.1371/journal.pone.0266975.