



Deep learning-based calibration of resistance factors for pile groups with load tests

Yuting Zhang¹ · Jinsong Huang¹ · Jiawei Xie¹ · Shui-Hua Jiang² · Cheng Zeng³

Received: 24 October 2024 / Accepted: 5 April 2025 / Published online: 2 June 2025
© The Author(s) 2025

Abstract

Resistance factors for pile groups are typically derived using empirical methods that do not directly account for system redundancy and overlook the correlation between individual piles, which are inherently influenced by the spatial variability of soils. While rigorous three-dimensional (3D) random finite difference (RFD) or random finite element (RFE) analyses could potentially address these issues, they are constrained by significant computational demands. Therefore, this paper proposes a deep learning-based approach for calibrating resistance factors for pile groups with individual pile load tests. Specifically, a surrogate model based on a convolutional neural network (CNN) is proposed, which is trained and validated using the database generated by RFD analyses. The trained model is further used to derive pile resistances in spatially variable soils. Finally, the resistance factors are calibrated by counting and conditional probability based on the outcomes of load test results. The proposed approach is demonstrated using a pile group example. Results show that the proposed approach effectively captures the impacts of load test results and their corresponding locations, as well as the spatial variability of soil properties, on resistance factors.

Keywords Convolutional neural network (CNN) · Pile groups · Pile load tests · Random finite difference analysis · Resistance factors

1 Introduction

In the load and resistance factor design (LRFD), resistance factors are employed to address uncertainties associated with the pile resistance predictions. These uncertainties arise from variabilities in geotechnical properties, measurement errors in laboratory or field tests, transformation uncertainties between design parameters and test results, and uncertainties inherent in design models [7, 27, 34, 40]. Pile load tests yield a more precise evaluation of pile resistances. Consequently, integrating load tests into the design process can significantly reduce uncertainties of pile resistances, thereby enabling the application of higher resistance factors in designs. For example, the Australian Standard for Piling-Design and Installation [33] specifies that resistance factors range from 0.4 to 0.9, depending on the percentage of piles tested via static load tests, and from 0.4 to 0.8 for piles tested through dynamic load tests. However, the determination of these resistance factors predominantly relies on engineering judgement [11].

✉ Jinsong Huang
jinsong.huang@newcastle.edu.au

Yuting Zhang
yuting.zhang11@uon.edu.au

Jiawei Xie
jjiawei.xie@newcastle.edu.au

Shui-Hua Jiang
sjiangaa@ncu.edu.cn

Cheng Zeng
cheng.zeng@csust.edu.cn

¹ Discipline of Civil, Surveying and Environmental Engineering, Priority Research Centre for Geotechnical Science and Engineering, The University of Newcastle, Callaghan, NSW, Australia

² School of Infrastructure Engineering, Nanchang University, Jiangxi, China

³ School of Traffic and Transportation Engineering, Changsha University of Science and Technology, Changsha, Hunan, China

Several studies have been carried out to calibrate resistance factors for single piles based on load test results [42, 47]. Park, et al. [25] evaluated resistance factors for pipe piles based on proof load tests, suggesting that load tests could lead to higher resistance factors. Zhang, et al. [45] implemented a Bayesian approach to calibrate resistance factors of H-piles using proof load tests, incorporating the consideration of site variabilities and design methods. In contrast, the determination of resistance factors for pile groups mainly depends on empirical approaches. Paikowsky [24] noted that pile groups typically demonstrate higher reliability than individual piles due to system redundancy. Accordingly, previous studies [2, 4, 18] commonly recommended a reliability index of 2.0 to 2.5 for individual piles to attain a target reliability index of 3 for the whole pile group. However, this empirical approach does not directly account for system redundancy. Furthermore, it assumes that resistances of individual piles within the group are perfectly correlated [41]. Yet, such relationships should inherently be influenced by the spatial variability of soils [23].

The calibration of resistance factors for pile groups, based on individual pile load tests, necessitates the consideration of complex pile–soil–pile interaction and the correlations among individual piles. This calibration requires rigorous three-dimensional (3D) random finite difference (RFD) or random finite element (RFE) analyses, when accounting for the spatial variability of soils. However, these analyses are computationally intensive, as the 3D numerical analysis must be repeated numerous times to evaluate the reliability of pile groups. Moreover, the process of calibrating resistance factors necessitates multiple reliability evaluations to achieve a target reliability index, which further escalates computational costs. To address the challenge of computational demand in reliability analysis involving spatially variable soils, various surrogate modelling techniques have been proposed. For example, Jiang, et al. [14] and Jiang and Huang [16] introduced multiple stochastic response surface methods as surrogate models to map the relationship between the factor of safety of slopes and spatial variability of soil properties. Yi, et al. [39] and Khorramian, et al. [17] employed Kriging as surrogate models for evaluating the reliability of pile foundations in spatially variable soils. Nonetheless, these models may experience reduced efficiency and accuracy when dealing with high-dimensional random fields [15], underscoring the need for a robust surrogate model suitable for complex 3D random field problems.

Advanced machine learning algorithms, such as convolutional neural network (CNN), have demonstrated particularly effective in handling high-dimensional data and constructing precise regression models for scenarios involving random fields [35, 38]. For instance, Wang and Goh [37] employed CNN to replace the RFE model for

predicting the probability of slope failures, highlighting the model's computational efficiency and accuracy. Additionally, Wang, et al. [36] utilized three deep learning approaches, including CNN, to evaluate the time-dependent reliability of reservoir slopes, demonstrating that CNN performed the best in the Bazimen landslide case study. Furthermore, Wu, et al. [38] developed a CNN-based surrogate model to predict wall deflection induced by braced excavation in spatially variable soils, while Zhang, et al. [43] applied CNN to analyse tunnel performance in spatially variable soils. Despite these advances, the application of CNN for 3D pile group reliability analyses in spatially variable soils, as well as the integration of CNN models into the LRFD framework for calibrating resistance factors, remains unexplored.

In this paper, a CNN-based approach is proposed to calibrate resistance factors for pile groups with load test results. Initially, a proper number of RFD simulations are conducted, which are used to train a proposed CNN model. Subsequently, the accuracy of this model is validated, and it is employed as a surrogate model to determine resistances for individual piles and pile groups, based on a given random field of soil properties. The trained model facilitates the efficient execution of numerous simulations, allowing for the direct determination of failure probabilities for individual piles and the pile group through counting. Consequently, the calibration of resistance factors is straightforward, enabling the adjustment of these factors to achieve a target reliability index (or probability of failure). Additionally, when load tests are conducted on individual piles, the probability of failure for the pile group is calculated using conditional probability, leading to the recalibration of the resistance factors to achieve the desired reliability index. The structure of the paper is organized as follows: Sect. 2 presents the methodology, providing a detailed description of the proposed approach. Section 3 demonstrates the proposed approach using a 3×3 pile group under vertical loads in spatially variable soils, and the accuracy of the proposed CNN model is validated. Building upon this, Sect. 4 explores the influence of various factors on the calibrated resistance factors, including load test results and their corresponding locations, spatial variability of soil properties, test chains and measurement errors. Section 5 summarizes the major findings of the study.

2 Methodology

2.1 Calibration of resistance factors based on individual proof load tests within LRFD

In LRFD, the resistance and load factors are used to address the uncertainties associated with resistance and

loads, respectively. When considering only dead and live loads, the design equation for pile groups is defined as [3]:

$$\phi R_{gn} = \gamma_D Q_{Dn} + \gamma_L Q_{Ln} \quad (1)$$

where ϕ , γ_D and γ_L represent the resistance factor, dead load factor and live load factor, respectively. Q_{Dn} and Q_{Ln} are the nominal dead load and nominal live load, respectively. R_{gn} is the nominal pile group resistance, which is usually determined by the group efficiency and the summation of individual pile resistances [28]. As noted by Zhang, et al. [46], the group efficiency is a random variable when pile groups are located in spatially variable soils. Therefore, the mean group efficiency, η_n , is utilized to calculate the nominal pile group resistance:

$$R_{gn} = \eta_n \sum_{i=1}^N R_{in} \quad (2)$$

where R_{in} represents the nominal resistance of the i th pile. In this paper, R_{in} is defined as the mean individual pile resistance; thus, $R_n = R_{1n} = \dots = R_{Nn}$.

The limit state function, g , is established as the condition where the pile group resistance equals the sum of loads transferred from the superstructure:

$$g = R_g - Q_D - Q_L = 0 \quad (3)$$

where R_g , Q_D and Q_L denote the pile group resistance, dead load and live load, respectively. However, the actual loads transferred from the superstructures are generally unknown and are typically estimated as [2]:

$$Q_D = \lambda_D Q_{Dn} \quad Q_L = \lambda_L Q_{Ln} \quad (4)$$

where λ_D and λ_L are the dead and live load bias factors, respectively.

By substituting Eqs. (1), (2) and (4) into Eq. (3), the limit state function for pile groups is derived as follows:

$$g = \frac{R_g}{\phi N \eta_n R_n} \times (\gamma_D \kappa + \gamma_L) - (\lambda_D \kappa + \lambda_L) = 0 \quad (5)$$

where N is the number of piles within the group. κ is the ratio of dead to live load, $\kappa = Q_{Dn}/Q_{Ln}$.

In Eq. (5), load factors are specified according to different limit states as outlined by AASHTO [3]. If the Strength I limit state is considered, $\gamma_D = 1.25$ and $\gamma_L = 1.75$. κ typically ranges from 2 to 5 across various superstructures, but it has a minimal impact on resistance factors [1]. Thus, $\kappa = 2$ is considered in this study. For foundation designs, λ_D and λ_L are assumed to follow a lognormal distribution. Their mean (i.e. μ_{λ_D} and μ_{λ_L}) and standard deviations (i.e. σ_{λ_D} and σ_{λ_L}) are derived from Paikowsky [24]. Consequently, a critical task in calibrating resistance factors is determining R_g . Furthermore, when proof load tests are conducted on individual piles within

the group, the corresponding individual pile resistances, $\mathbf{R} = (R_1, R_2, \dots, R_N)$, are crucial in assessing whether the tested piles fail or not.

This paper describes the use of a trained CNN model to generate a large dataset (e.g. one million) for R_g and $\mathbf{R} = (R_1, R_2, \dots, R_N)$. Initially, a limited number of RFD simulations are conducted to create the training database, with the specifics of the RFD analysis detailed in Sect. 2.2. This database is then employed to train CNN models, the architecture of which is outlined in Sect. 2.3. The model inputs are the random fields of soil properties, $\boldsymbol{\theta}$, while the outputs include $\mathbf{R} = (R_1, R_2, \dots, R_N)$ and R_g . Following this, the trained CNN model is applied to calculate the resistances of individual piles and the pile group using the designated random fields of soil properties. After that, the trained CNN model is utilized to derive $\mathbf{R} = (R_1, R_2, \dots, R_N)$ and R_g with the given random fields of soil properties.

Without conducting load tests, the probability of failure for the pile group, $P(F_g)$, is calculated as follows:

$$P(F_g) = \frac{1}{N_T} \sum_{j=1}^{N_T} I[g(\phi_0, \boldsymbol{\theta}_j) < 0] \quad (6)$$

where N_T represents the total number of simulations. $I(\cdot)$ is the indicator function, which is assigned a value of 1 when the specified condition in the bracket is satisfied. $g(\phi_0, \boldsymbol{\theta}_j)$ denotes the limit state function corresponding to the j th random field of soil properties, $\boldsymbol{\theta}_j$, given a trial resistance factor, ϕ_0 .

The reliability index of the pile group is determined as: $\beta = -\Phi^{-1}(P(F_g))$, where Φ^{-1} is the inverse cumulative distribution function of the standard normal distribution. The trial resistance factor, ϕ_0 , is then adjusted to achieve a target reliability index, β_{GT} . In this study, the bisection method is utilized to iteratively adjust ϕ_0 until $|\beta - \beta_{GT}| \leq 0.01$.

With load tests conducted on individual piles, it is assumed that the proof test load is T , and the measurement error is ε . Generally, ε is assumed to follow a normal distribution with a mean of zero, $\mu_\varepsilon = 0$ and a standard deviation of σ_ε . For simplification, it is assumed that σ_ε exhibits a proportional relationship with the corresponding measurements, denoted as $\sigma_\varepsilon = aT$. The value of a depends on the specific load test techniques used and the method employed to interpret the test results. If the load test is conducted on the i th pile, the probability that the pile fails to pass the test is as follows:

$$P(F_i) = \frac{1}{N_T} \sum_{j=1}^{N_T} I[R_i(\boldsymbol{\theta}_j) < T - \varepsilon] \quad (7)$$

where $R_i(\theta_j)$ represents the resistance of the i th pile corresponding to the j th random field of soil properties, θ_j .

Based on Eq. (7), the probability that the pile passes the test is:

$$P(\bar{F}_i) = 1 - P(F_i) \quad (8)$$

Given that the proof load test is conducted on the i th pile, and the pile fails, the probability that the pile group fails is denoted as $P(F_g|F_i)$, which is obtained using conditional probability:

$$P(F_g|F_i) = \frac{P(F_g \cap F_i)}{P(F_i)} \quad (9)$$

where $P(F_g \cap F_i)$ is the probability that both the i th pile and the pile group fail, which can be obtained as follows:

$$P(F_g \cap F_i) = \frac{1}{N_T} \sum_{j=1}^{N_T} I[R_i(\theta_j) < T - \varepsilon, \quad g(\phi_0, \theta_j) < 0] \quad (10)$$

Based on the proof load test result that the i th pile failed, the reliability index of the pile group is determined as: $\beta = -\Phi^{-1}(P(F_g|F_i))$. Finally, the trial resistance factor, ϕ_0 , is adjusted to achieve a target reliability index, β_{GT} .

2.2 Pile resistances in spatially variable soils using RFD analysis

The resistances of individual piles and pile groups in spatially variable soils are assessed through RFD analysis, which combines random field theory with FD analysis. The RFD analysis comprises two main components: the generation of random fields of soil properties and the FD analysis of pile resistances.

Random field theory has been utilized to explicitly model the spatial variability of soils, enhancing the probabilistic analysis and reliability-based design of geotechnical structures [8, 10, 17]. In the context of pile foundations in undrained clay, pile resistances are primarily influenced by the undrained shear strength, c_u , [22, 30]. Consequently, this paper assumes c_u as a random field while maintaining other parameters (e.g. shear modulus) as constants. Using random field theory, a 3D spatially distributed c_u is characterized by specified statistical parameters, including mean, μ_{c_u} , coefficient of variation, COV_{c_u} , and spatial correlation length, Θ . The soil spatial variability displays anisotropic spatial variability, with the horizontal correlation length, Θ_h , is generally much higher than the vertical correlation length, Θ_v . Therefore, an anisotropic random field is adopted and generated using the open-source toolbox 'GSTools' [20].

The FLAC 3D software [13] is employed to determine the resistances for both individual piles and pile groups. To obtain the individual pile resistances, a FD model with the interested pile is first constructed (i.e. Fig. 1a), where the soil is modelled using the Mohr–Coulomb failure criteria and the pile is represented by an inbuilt linear structural element with interface properties. Then, the random field of c_u is mapped into soil elements. Subsequently, a displacement-controlled vertical load is applied to generate a load–displacement curve. The individual pile resistance is determined using the ISSMFE criterion [12], which corresponds to the displacement of piles which is 10% of the pile diameter. Owing to the variation in soils surrounding individual piles, the individual pile analyses need to be repeated for N times to obtain $\mathbf{R} = (R_1, R_2, \dots, R_N)$. For pile group resistance, R_g , a FD model of the pile group is first constructed in the same random field of c_u (i.e. Fig. 1b), followed by the application of a displacement-controlled vertical load. For simplicity, the pile group is considered freestanding, and the pile cap is assumed to be rigid, not undergoing rotation under vertical loads. Thereby, the pile cap does not need to be modelled in the pile group model, and all the individual piles undergo the same displacement under vertical loads. The total load acting on the pile group is calculated as the sum of the loads on individual piles [48]. The same failure criterion [12] utilized for individual piles is applied to determine R_g . A detailed description of the pile resistance analysis using RFD analysis can be found in Zhang, et al. [46].

2.3 Proposed CNN model

Conventional CNN is generally used to process images composed of pixels. Each pixel features three channels corresponding to the primary colours: red, green and blue, with each pixel thus described by three intensity values for these channels. Therefore, the CNN receives this pixel data from the image to be processed. When adapting CNN for processing random fields, the discretized soil elements in FD models serve a role analogous to that of pixels in image processing. In this analogy, the random field of soil properties is equivalent to the channel, and the magnitude of the soil property for each soil element is similar to the channel intensity of a pixel in conventional image processing. Typically, a CNN model comprises six major components: the input layer, convolution layer, pooling layer, activation layer, fully connected layer and output layer. These layers facilitate the construction of diverse architectures tailored to specific problems. The architecture of the CNN model used in this paper is depicted in Fig. 2, while the details of layers and parameters are summarized in Table 1. This configuration was established through a trial-and-error

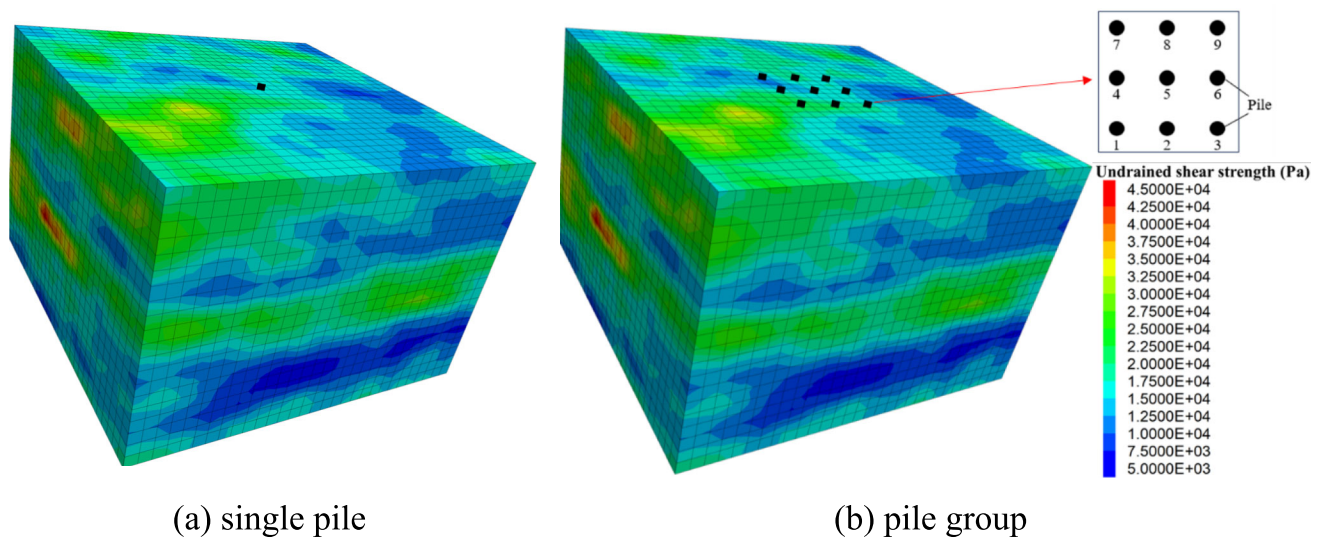


Fig. 1 FD model of single piles and pile groups in spatially variable soils

methodology, informed by previously proposed architecture designs [6, 35]. Key layers in the proposed CNN model are elaborated as follows.

Input layer: the input layer contains the relevant soil property information essential for model construction. In this study, the soil medium is modelled as a cuboid of dimensions $30\text{ m} \times 30\text{ m} \times 20\text{ m}$, which is discretized using cubic elements with a uniform side length of 1 m . Since only the undrained shear strength, c_u , is treated as random fields, the number of channels is one. Therefore, the input layer comprises a size of $30 \times 30 \times 20 \times 1$, where ‘ $30 \times 30 \times 20$ ’ represents the size of the soil domain in FD models, while ‘1’ denotes the random field of c_u .

3D convolution layer: a 3D convolution layer applies sliding cuboidal convolution filters to extract features from the 3D input. Specifically, the 3D filter moves over the input 3D random field of c_u , computing the dot product of

the weights with the input and then adding a bias term. These weights and biases are trained to capture salient features from the input. In the 3D convolution layer, the stride, padding, filter size, and number of filters are common hyperparameters that need to be defined. Stride represents the step size that filters move each time, while padding means adding extra borders of the layer input. A stride size of $1 \times 1 \times 1$ is selected, while a zero padding is chosen to maintain the spatial dimensions of the output. The filter configuration is determined as $2 \times 2 \times 2@32$, which means the filter size is $2 \times 2 \times 2$ and the number of filters is 32. In the proposed CNN model, each 3D convolution layer is followed by a batch normalization layer, which enhances training stability and accelerates network training.

Activation layer: the activation layer is utilized to enhance the nonlinear characterization capability of models. Different activation functions are available, such as the

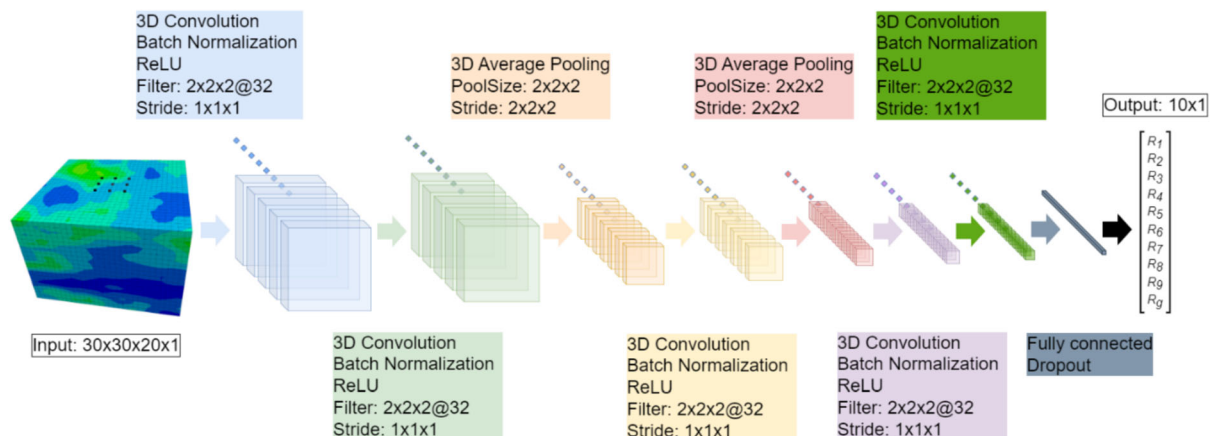


Fig. 2 Architecture of the proposed CNN model

Table 1 Details of layers and parameters of the CNN model

Layer	Filter	Stride	Pooling	Output size
3D Convolution + batch normalization + ReLU	$2 \times 2 \times 2@32$	$1 \times 1 \times 1$	–	$29 \times 29 \times 19 \times 32$
3D Convolution + batch normalization + ReLU	$2 \times 2 \times 2@32$	$1 \times 1 \times 1$	–	$28 \times 28 \times 18 \times 32$
3D Average pooling	–	–	$2 \times 2 \times 2$	$14 \times 14 \times 9 \times 32$
3D Convolution + batch normalization + ReLU	$2 \times 2 \times 2@32$	$1 \times 1 \times 1$	–	$13 \times 13 \times 8 \times 32$
3D Average pooling	–	–	$2 \times 2 \times 2$	$6 \times 6 \times 4 \times 32$
3D Convolution + batch normalization + ReLU	$2 \times 2 \times 2@32$	$1 \times 1 \times 1$	–	$5 \times 5 \times 3 \times 32$
3D Convolution + batch normalization + ReLU	$2 \times 2 \times 2@32$	$1 \times 1 \times 1$	–	$4 \times 4 \times 2 \times 32$
Fully connected	–	–	–	1×10

rectified linear unit (ReLU), $f(x) = \max(0, x)$, tanh function, $f(x) = \tanh(x)$, and sigmoid function, $f(x) = 1/(1 + e^{-x})$. A distinctive feature of the ReLU is its unbounded output for positive inputs, unlike the bounded outputs of sigmoid and tanh functions. Additionally, the gradients of the ReLU are either zero or one, which significantly accelerates computational processes compared to sigmoidal functions [6]. Furthermore, the ReLU demonstrates superior convergence properties in stochastic gradient descent (SGD) algorithms, surpassing both sigmoid and tanh functions in this regard [26, 29]. Therefore, the ReLU is adopted in this study.

Pooling layer: the fundamental concept behind pooling layers is downsampling, which effectively reduces the spatial dimensions of data. There are two principal types of pooling: max pooling and average pooling. Max pooling selects the maximum value from each subarray, whereas average pooling computes the mean value. Average pooling layers are adopted instead of max pooling layers, as the pile resistances are primarily governed by the overall soil properties along the pile depth rather than extreme value for specific locations. In this model, the average pooling operation is selected with a pooling size of $2 \times 2 \times 2$ and stride of $2 \times 2 \times 2$.

Fully connected layer: a fully connected layer is adopted at the end of the proposed CNN model, where all neurons in one layer are connected with every neuron in the subsequent layer. The fully connected layer processes the flattened tensor received from the preceding hidden layer, transforming these inputs into the desired outputs. For this specific application, the layer outputs vectors corresponding to individual pile resistances and the pile group resistance.

2.4 Procedure of the proposed approach

The procedure of the proposed approach is summarized as follows:

Step 1: Generate random fields of c_u based on specified statistical parameters, via random field theory.

Step 2: Map the values of c_u to the individual pile and pile group models. Compute individual pile resistances and pile group resistance, according to Sect. 2.2.

Step 3: Repeat steps 1 and 2 for N_1 times to generate the training database.

Step 4: Divide the database generated in Step 3 into training and validation datasets. Utilize these datasets to train the proposed CNN model, as described in Sect. 2.3. The CNN model is designed to learn the mapping relationship between the input soil properties and the output pile resistances.

Step 5: Generate additional N_T random fields of c_u . These random fields are then used as inputs for the trained CNN model, to derive N_T sets of values for $\mathbf{R} = (R_1, R_2, \dots, R_N)$ and R_g .

Step 6: In cases where load tests have not been conducted, the resistance factor is calibrated using Eq. (6) and the methodology outlined in Sect. 2.1.

Step 7: In cases where load tests have been conducted, the resistance factor is calibrated following Eqs. (7) to (10) and the methodology described in Sect. 2.1.

3 Example

To demonstrate the proposed approach, a 3×3 pile group subjected to vertical loads in undrained clay is utilized. The pile length (L) is 10.5 m, with 10 m embedded in clay. The pile diameter (D) is 1.0 m while the pile spacing (d) is 3 m.

The elastic modulus (E_p) is determined to be 2.2×10^7 kPa, and the Poisson ratio is set at 0.3. For the soil medium, only the undrained shear strength, c_u , is treated as a random variable, with a mean value $\mu_{c_u} = 20$ kPa. The general range of COV_{c_u} was 10%–55% [27], and $\text{COV}_{c_u} = 50\%$ is adopted for probabilistic analyses. Moreover, an anisotropic spatial correlation length is adopted [27] with the horizontal correlation length, $\Theta_h = 40$ m, significantly exceeding the horizontal correlation length, $\Theta_v = 5$ m. The values of shear modulus, G , and bulk modulus, K , are determined as 1.3×10^3 kPa and 6.0×10^3 kPa, respectively, following Bowles [5].

For training the proposed CNN model, an initial set of 1000 RFD simulations is conducted following Sect. 2.2, to generate the individual pile resistance and the pile group resistance. The initial dataset of 1000 samples is then divided into training and validation datasets with a ratio of 80:20 [44] to train and validate the proposed CNN model. Once the CNN model is trained, an additional 5000 random fields of soil properties are generated, and corresponding random finite difference (RFD) analyses are performed to obtain the true resistance factors. These true resistance factors are then compared with the resistance factors calibrated using the pile resistances predicted by the trained CNN model. This comparison is conducted to further evaluate the accuracy and reliability of the proposed deep learning-based approach for calibration purposes. The configurable hyperparameters of the CNN model utilized in this study are detailed in Table 2. The stochastic gradient descent (SGD) algorithm [21] is employed to optimize the filter weights and biases during model training. The accuracy of the CNN model is quantitatively assessed using the root mean square error (RMSE) and mean absolute percentage error (MAPE), which are defined in Eqs. (11) and (12), respectively.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y} - y_i)^2} \quad (11)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y} - y_i}{y_i} \right| \times 100\% \quad (12)$$

where $\hat{y}$ and y_i are the values predicted by the CNN model and calculated by RFD analyses, respectively.

3.1 Validation of FDM

To validate the FD models, a deterministic analysis is performed with the mean undrained shear strength and $\text{COV}_{c_u} = 0$. The results from this deterministic analysis are then compared with those derived from empirical static formulas as proposed by Poulos and Davis [28].

According to Poulos and Davis [28], for piles in clays, the individual pile resistance, R , is calculated using the following formula:

$$R = \int_0^L U c_u \alpha dz + A_b c_u N_c \quad (13)$$

where U is the pile perimeter, α is the undrained pile–soil adhesion factor, N_c is the bearing capacity factor.

To calculate the pile group resistance, R_g , the following empirical relationship is suggested by Poulos and Davis [28]:

$$\frac{1}{R_g^2} = \frac{1}{n^2 R^2} + \frac{1}{R_B^2} \quad (14)$$

$$R_B = B_r L_r c_u N_c + 2(B_r + L_r) L \bar{c}_u \quad (15)$$

where R_B is the bearing capacity for block failure of the group; $\bar{c}_u$ is the average undrained shear strength. For a 3×3 pile group considered herein, $L_r = B_r = 2d + D$.

The undrained pile–soil adhesion factor, α , varies considerably with many factors, such as the types of piles, the soil conditions and the pile installation methods [28]. The typical relationship between α and c_u for driven piles has been reported by McClelland [19]. It is generally accepted that $\alpha = 1$ for soft clays (e.g. $c_u \leq 24$ kPa). The value of the bearing capacity factor, N_c , has been proposed by Skempton [32]. When calculating R , the value of N_c is limited to a maximum value of 9 for the pile length larger than four times of the pile diameter [9]. When calculating R_B , the value of N_c is a function of L_r/B_r and L/B_r , and the curve is provided in Poulos and Davis [28].

Based on Eqs. (13) to (15), the individual pile resistance and pile group resistance are 770 kN and 6168 kN, respectively. These values are comparable to the results obtained from FLAC3D, which are 776 kN for the individual pile resistance and 6623 kN for pile group resistance, respectively, validating the adopted FD models.

Table 2 Configurable hyperparameters of the CNN

Learning rate	Maximum epoch	Minimum batch size	Data shuffle	L2-regularization	Validation frequency
0.01	500	64	Every epoch	0.0001	50

3.2 Accuracy of the CNN model

This subsection evaluates the accuracy of the CNN model trained by different sample sizes (e.g. 100, 200, ..., and 1000 samples). For example, Fig. 3a indicates 100 samples are utilized for training the CNN model, which means that 80 samples are utilized for training and 20 samples are employed for validation during the training process. The trained CNN model is subsequently adopted to predict the individual pile resistances and pile group resistance using the 5000 random fields in the testing dataset. Figure 3 shows the CNN-predicted centre pile resistance (i.e. pile 5) and the corresponding RFD-calculated resistances, which demonstrates that data points become more closely aligned with the 1:1 line as the number of training samples increases, indicating enhanced predictive performance of the CNN model. The RMSE, MAPE, and coefficient of determination (R^2) are calculated for each sample size and depicted in Fig. 3. A decreasing trend in RMSE and MAPE and an increasing trend in R^2 values are observed as the training sample size increases. Specifically, RMSE and MAPE decrease from 1.7×10^5 N and 13.1% with 100 samples to 3.6×10^4 N and 3.6% with 1000 samples, respectively. Conversely, the R^2 rises from 0.718 to 0.985 as the sample size increases from 100 to 1000. These trends suggest that larger training datasets enable the CNN model to learn more comprehensive features associated with the spatial variation of the random properties, resulting in improved prediction accuracy.

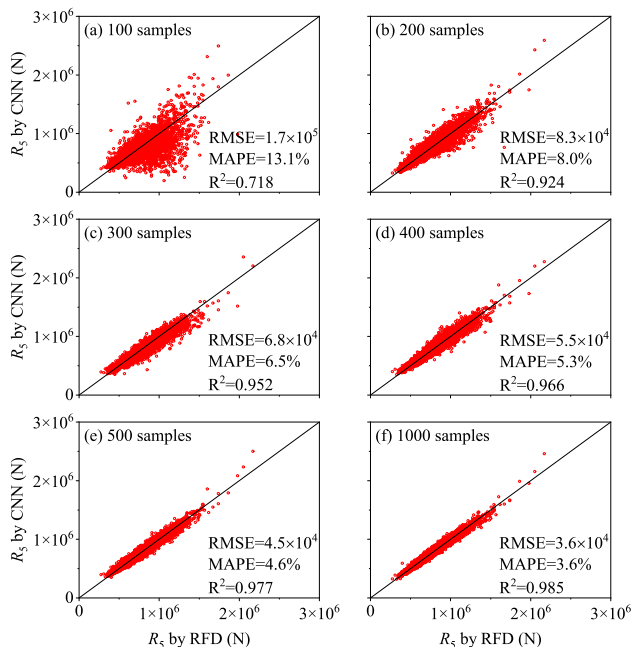


Fig. 3 RFD versus CNN for centre pile (pile 5) resistance with different sample sizes

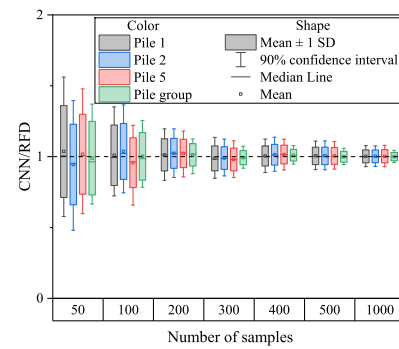


Fig. 4 Boxplot comparison of pile resistances across different sample sizes for individual piles and the pile group

Figure 4 displays a boxplot comparison of pile resistances across different sample sizes for individual piles and the pile group. Results illustrate that as the number of training samples increases from 50 to 1000, both the mean and median values of the CNN-predicted/RFD-calculated ratios gradually converge towards 1. Simultaneously, the standard deviation (SD) of the ratio, which serves as an indicator of estimation uncertainty, also decreases with an increase in sample size.

The accuracy of the proposed CNN model is further evaluated by comparing the real resistance factors with those calibrated using the trained CNN model. Specifically, for the pile group to achieve a target reliability index of 2.33, without conducting load tests on individual piles, the real resistance factor and the calibrated resistance factor based on the trained CNN model are 0.73 and 0.72, respectively. When one proof load test is conducted on pile 1 that passes, the real and the one calibrated using the CNN model increase to 1.03 and 1.05, respectively. Conversely, if one proof load test is conducted on pile 1 that fails, the real and the one calibrated using the CNN model decrease to 0.67 and 0.66, respectively. The good agreement between these resistance factors demonstrates the accuracy and reliability of the proposed CNN model.

3.3 Important soil zones detected by the CNN model

The local soils surrounding piles play a more critical role in determining pile resistance than the broader soil domain. While this study uses the entire soil domain as input features, the CNN model effectively captures the significant influence of the local soils surrounding the pile. This is demonstrated through the application of the Saliency Maps technique [31], which calculates the gradients of the output pile resistance with respect to the input soil properties and visualizes the relative importance of different regions based on these gradients. Areas with higher gradient magnitudes are interpreted as having a greater influence on

the model's predictions, thereby identifying the critical regions of soil properties that most significantly affect the prediction of pile capacities. Figure 5 illustrates gradient maps across various depths, revealing that the soils within or surrounding the pile group area contribute more substantially to the predicted pile resistances compared to soils located farther from the pile group.

4 Results

4.1 Effect of the test results and corresponding test locations

This subsection assesses the impact of load test results and their corresponding test locations on the calibrated resistance factors. The measurement error is assumed to be $\sigma_\varepsilon = 0.1 T$. For demonstration purposes, it is assumed that three pile load tests are performed on piles 1, 5, and 9, with a variety of resulting outcomes being observed and analysed. It is noted that when one test fails among three tests,

and the failed pile would be either pile 1, pile 5 or pile 9, the symmetry of the pile group implies that the results for piles 1 and 9 are identical. Consequently, two distinct resistance factors are derived when one test fails among three tests.

Figure 6 presents the impact of load test locations on resistance factors. Specifically, if only one pile failed among three load tests, the failure of pile 5 results in a lower resistance factor compared to failures in pile 1 or pile 9. This is attributed to the central position of pile 5, which holds more critical information regarding the reliability of the pile group. As such, a failure at the central location suggests a higher likelihood that the entire pile group might fail, thereby necessitating a lower resistance factor to meet the target reliability index [46].

It is observed from Fig. 6 that resistance factors generally decrease as the number of failed tests increases, as also noted by Zhang, et al. [45]. For comparative purposes, the resistance factor without load tests is also included in Fig. 6, indicating that only when all three piles fail in the tests does the resistance factor fall below that obtained

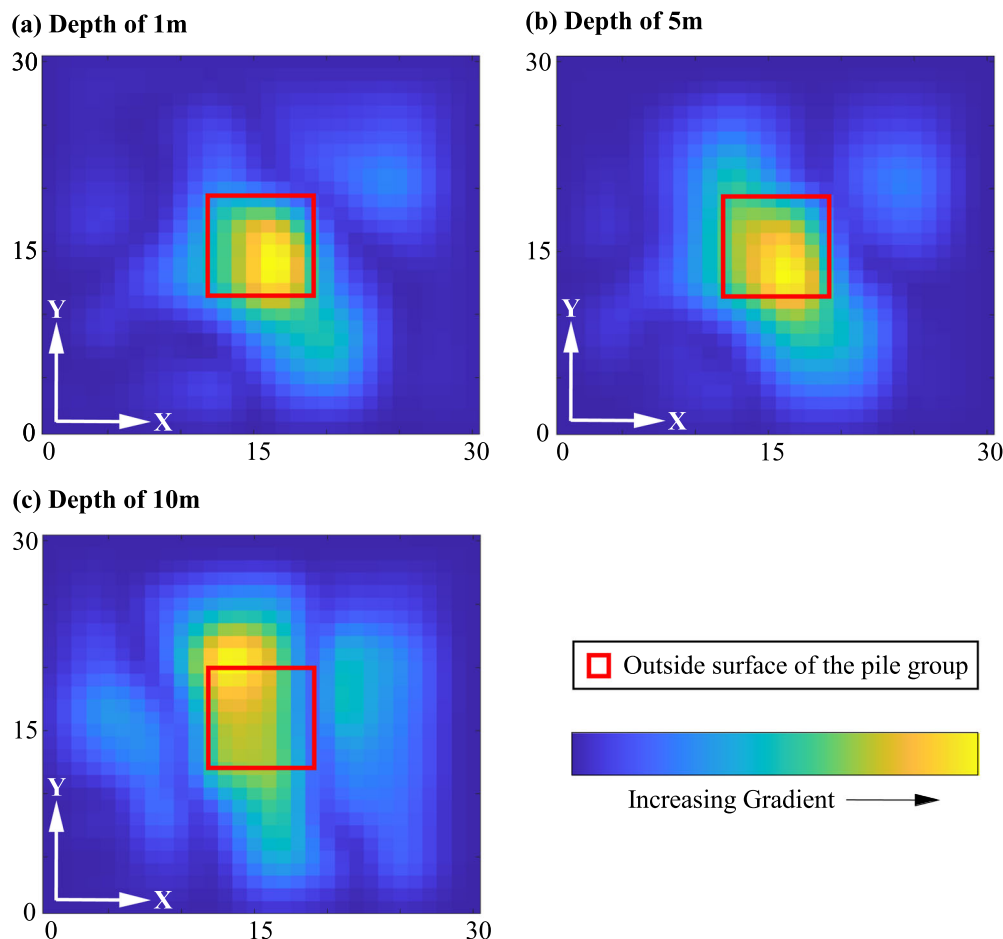


Fig. 5 Gradient maps at various depths

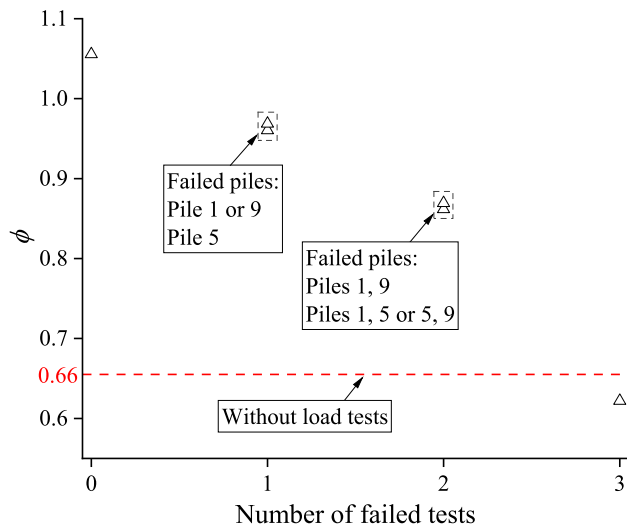


Fig. 6 Resistance factors with different test results and test locations

without conducting load tests. This observation underscores the significant potential of load tests to yield higher resistance factors, potentially reducing pile construction costs.

4.2 Effect of the horizontal correlation length

This subsection evaluates the influence of the horizontal correlation length, Θ_h , on the calibrated resistance factors. Again, it is assumed that the measurement error is $\sigma_\varepsilon = 0.1 T$. Four different values of Θ_h are considered, specifically 5 m, 10 m, 20 m and 40 m, while Θ_v is fixed at 5 m. Figure 7 illustrates the calibrated resistance factors for various values of Θ_h under different load test scenarios. In particular, Fig. 7a presents cases where all tested piles yield identical test results (i.e. all fail or all pass) and cases where no tests are conducted. In contrast, Fig. 7b depicts cases where tested piles yield different test results (i.e. some piles pass while others fail). In these figures, the

notation ‘F’ means the tested pile fails the test while the notation ‘P’ represents the tested pile passes the test. The prefix number identifies the specific pile tested. For example, ‘1P5F’ indicates that load tests were conducted on pile 1 and pile 5, with pile 1 passing and pile 5 failing.

In Fig. 7a, when no tests are conducted, resistance factors decrease as Θ_h increases. This decrease occurs because a larger Θ_h leads to greater similarity in soil properties around individual piles, causing the resistances of individual piles more likely to be uniformly high or low. As a result, the pile group system exhibits lower overall reliability, necessitating a lower resistance factor to achieve the target reliability index. In contrast, when Θ_h is small, the resistances of individual piles are less dependent on each other. In this scenario, weaker piles can be compensated by stronger ones, enhancing the overall reliability of the pile group, and thereby allowing for a higher resistance factor. Similarly, when all tested piles either pass or fail, the untested piles are more likely to exhibit similar outcomes as Θ_h increases, enabling the adoption of higher or lower resistance factors, respectively.

In Fig. 7b, when the tested piles yield different test results (i.e. some piles pass while others fail), the resistance factors exhibit a non-monotonic trend. Specifically, the resistance factors initially decrease and subsequently increase, which aligns with observations reported in [46]. For instance, when two load tests are conducted on Pile 1 and Pile 5, where Pile 1 passes the proof test and Pile 5 fails, the resistance factor decreases from 0.88 to 0.85 as Θ_h increases from 5 to 20 m. However, as Θ_h further increases to 40 m, the resistance factor increases to 0.89.

4.3 Effect of the coefficient of variation

This subsection investigates the influence of the coefficient of variation of undrained shear strength, COV_{c_u} , on the calibrated resistance factors. It is assumed that the

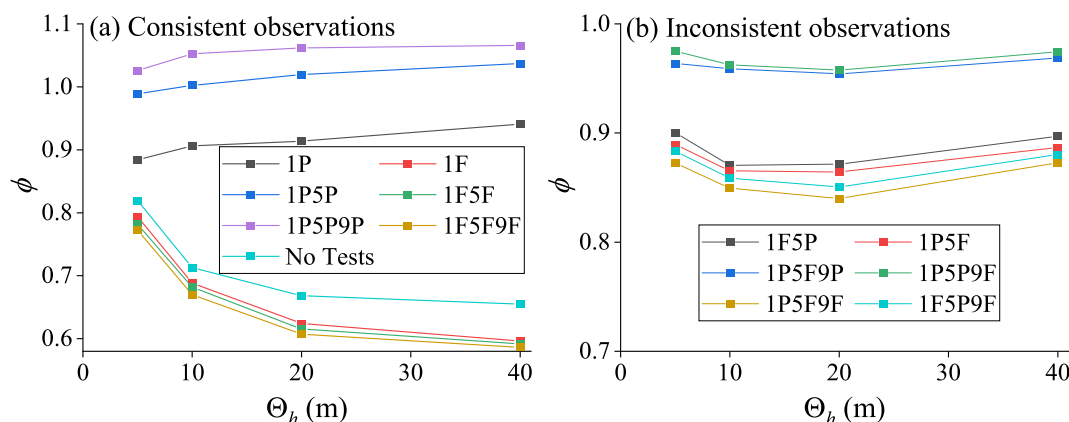


Fig. 7 Resistance factors as a function of Θ_h and test results: **a** consistent observations and **b** inconsistent observations

measurement error is $\sigma_\varepsilon = 0.1 T$, while $\Theta_h = 40$ m and $\Theta_v = 5$ m. Figure 8 illustrates the variation in resistance factors for different values of COV_{cu} under various test scenarios. Results indicate that resistance factors decrease as COV_{cu} increases, regardless of the number of tests or their corresponding outcomes. For example, when one load test is conducted on Pile 1 and passes, the resistance factor decreases from 0.96 to 0.90 as COV_{cu} increases from 20 to 100%. Similarly, when three load tests are conducted on Pile 1, Pile 5 and Pile 9, with Pile 1 passing while Pile 5 and Pile 9 fail, the resistance factor decreases from 0.91 to 0.79. This trend is primarily attributed to the increase in COV_{cu} , which leads to greater variability in individual pile resistances. Consequently, the reliability of the pile group system decreases, necessitating a lower resistance factor to achieve the target reliability index.

4.4 Effect of the test chains

This subsection examines the effect of test chains on the calibrated resistance factors. It is assumed that $\Theta_h = 40$ m and $\Theta_v = 5$ m. Additionally, two distinct measurement errors (i.e. $\sigma_\varepsilon = 0.1 T$ and $\sigma_\varepsilon = 0.3 T$) are considered to represent varying levels of accuracy in testing approaches. Figure 9 illustrates resistance factors as a function of test chain configurations and measurement errors. As shown in Fig. 9, resistance factors vary significantly across different test chains. Specifically, for $\sigma_\varepsilon = 0.1 T$, conducting one load test on pile 1 that passes results in increased resistance factors compared to scenarios without load tests (i.e. $\phi = 0.66$). Performing a subsequent test on pile 5 that fails slightly reduces the resistance factors. However, if the second test on pile 5 also passes, the resistance factor increases further. In cases where the two tests have been

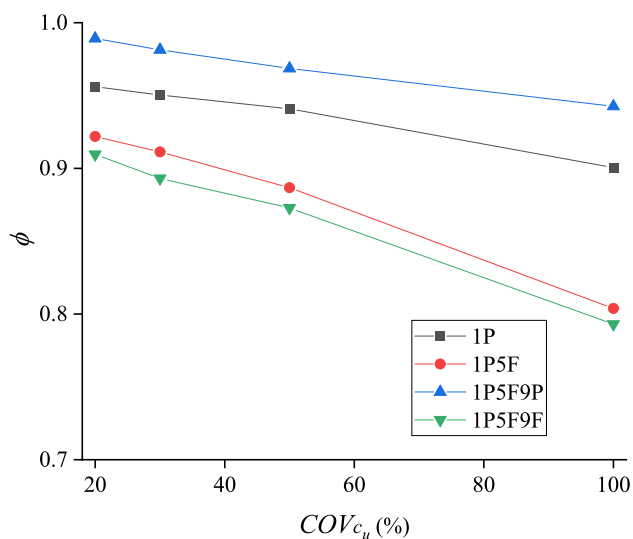


Fig. 8 Resistance factors with various COV_{cu}

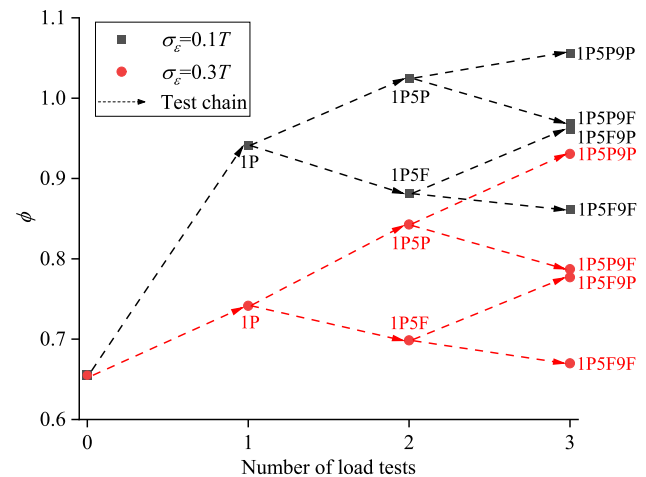


Fig. 9 Resistance factors with various test chains and measurement errors

conducted with both passing, a third test on pile 9 that fails leads to a slight reduction in the resistance factor. Nevertheless, the resistance factor remains higher than the value obtained when only one load test is conducted on pile 1 that passes. Similarly, in cases where two tests have been conducted with pile 1 passing and pile 5 failing, a third test on pile 9 that passes results in the increase in resistance factors, surpassing the value obtained from only one load test conducted on pile 1 that passes. These observations highlight the critical role of load tests in reducing the uncertainty associated with pile resistance.

Additionally, Fig. 9 indicates that lower measurement errors yield higher resistance factors for a given chain. This trend is attributed to the higher accuracy of the testing method, which enhances confidence in the results, thereby justifying the use of higher resistance factors.

5 Conclusions

This paper proposes a CNN-based approach to calibrate resistance factors for pile groups with individual pile load tests. A novel CNN model is developed and demonstrated to accurately substitute the computationally demanding RFD analyses of pile groups. By employing the trained CNN model, a comprehensive dataset of individual pile resistances and pile group resistances in spatially variable soils is generated. Subsequently, resistance factors are calibrated to achieve a specified reliability index through direct counting and the application of conditional probability based on the outcomes of individual pile load tests. To validate the proposed approach, a pile group is analysed, and the key findings are summarized as follows:

- (1) The proposed CNN-based approach effectively captures the impact of load test locations on resistance factors. Specifically, the failure of the central pile results in a lower resistance factor compared to failure at the corner piles, whereas the success of the central pile yields a higher resistance factor than a similar outcome at the corner piles.
- (2) The effect of horizontal spatial correlation length on calibrated resistance factors is dependent on load test outcomes. When all tested piles pass, resistance factors increase as the horizontal spatial correlation length increases. Conversely, when all tested piles fail, resistance factors decrease as the correlation length increases. Additionally, when multiple load tests yield mixed results (i.e. some piles pass while others fail), there exists a critical spatial correlation length, typically between 10 and 20 m, corresponding to the lowest resistance factors.
- (3) The coefficient of variation of soil properties and the measurement error in load tests exhibit a similar effect on the calibrated resistance factors. Specifically, resistance factors decrease as either the coefficient of variation of soil properties or the measurement error increases.

Acknowledgements This work was supported by the National Natural Science Foundation of China (Project Nos. 42272326, 52179103, and 52222905), the Jiangxi Provincial Natural Science Foundation (Grant Nos. 20242BAB24001 and 20232ACB204031) and the Hunan Provincial Department of Education (Project No. 23C0141).

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Data availability Some or all data, models, or code that supports the findings of this study is available from the corresponding author upon reasonable request.

References

1. AbdelSalam SS et al (2011) LRFD resistance factors for design of driven H-piles in layered soils. *J Bridge Eng* 16(6):739–748
2. Allen TM et al (2005) Calibration to determine load and resistance factors for geotechnical and structural design. In: Transportation research circular
3. AASHTO (2020) LRFD bridge design specifications, 9th edn. American Association of State Highway and Transportation Officials
4. Becker DE (1996) Eighteenth Canadian geotechnical colloquium: limit states design for foundations. Part II. Development for the national building code of Canada. *Can Geotech J* 33(6):984–1007. <https://doi.org/10.1139/t96-125>
5. Bowles JE (1995) Foundation analysis and design
6. Cha Y-J et al (2017) Deep learning-based crack damage detection using convolutional neural networks. *Comput Aided Civ Infrastruct Eng* 32(5):361–378
7. Ching J et al (2016) Spatial correlation for transformation uncertainty and its applications. *Georisk Assess Manag Risk Eng Syst Geohazards* 10(4):294–311
8. Griffiths DV et al (2002) Bearing capacity of rough rigid strip footing on cohesive soil: probabilistic study. *J Geotech Geoenviron Eng* 128(9):743–755
9. Houlsby GT, Martin CM (2003) Undrained bearing capacity factors for conical footings on clay. *Géotechnique* 53(5):513–520
10. Huang JS et al (2010) System reliability of slopes by RFEM. *Soils Found* 50(3):343–353
11. Huang JS et al (2016) Updating reliability of single piles and pile groups by load tests. *Comput Geotech* 73:221–230
12. ISSMFE (1985) Axial pile loading test—part 1: static loading. *Geotech Test J* 8(2):79–90
13. Itasca Consulting Group I (2017) FLAC3D—fast lagrangian analysis of continua in three-dimensions, Ver. 6.0. Minneapolis: Itasca
14. Jiang S-H et al (2015) Efficient system reliability analysis of slope stability in spatially variable soils using monte carlo simulation. *J Geotech Geoenviron Eng* vol 141(2)
15. Jiang S-H et al (2023) Data augmentation for CNN-based probabilistic slope stability analysis in spatially variable soils. *Comput Geotech* 160:105501
16. Jiang S-H, Huang J-S (2016) Efficient slope reliability analysis at low-probability levels in spatially variable soils. *Comput Geotech* 75:18–27
17. Khorramian K et al (2023) Optimized active learning Kriging reliability based assessment of laterally loaded pile groups modeled using random finite element analysis. *Comput Geotech* 154:105135
18. Kwak K et al (2010) Reliability-based calibration of resistance factors for static bearing capacity of driven steel pipe piles. *Can Geotech J* 47(5):528–538
19. McClelland B (1974) Design of deep penetration piles for ocean structures. *J Geotech Eng Div* 100(7):709–747
20. Müller S et al (2021) GSTools v1. 3: a toolbox for geostatistical modelling in Python. *Geosci Model Dev Discuss*, pp 1–33
21. Murphy KP (2012) Machine learning: a probabilistic perspective. MIT press
22. Naghibi M, Fenton GA (2011) Geotechnical resistance factors for ultimate limit state design of deep foundations in cohesive soils. *Can Geotech J* 48(11):1729–1741
23. Naghibi F, Fenton GA (2017) Target geotechnical reliability for redundant foundation systems. *Can Geotech J* 54(7):945–952
24. Paikowsky SG (2004) NCHRP report 507: load and resistance factor design (LRFD) for deep foundations. Transportation Research Board
25. Park JH et al (2012) Implementation of Bayesian theory on LRFD of axially loaded driven piles. *Comput Geotech* 42:73–80
26. Peng Y et al (2024) Asphalt pavement raveling recognition based on deep learning algorithms. *Intell Transp Infrastruct*. <https://doi.org/10.1093/iti/liae013>

27. Phoon KK, Kulhawy FH (1999) Characterization of geotechnical variability. *Can Geotech J* 36(4):612–624
28. Poulos HG, Davis EH (1980) Pile foundation analysis and design
29. Rao C, Liu Y (2020) Three-dimensional convolutional neural network (3D-CNN) for heterogeneous material homogenization. *Comput Mater Sci* 184:109850
30. Ronold KO (2016) Characteristic soil strength for axial pile capacity and its estimation with confidence for offshore applications. *Struct Saf* 63:81–89
31. Simonyan K et al (2013) Deep inside convolutional networks: visualising image classification models and saliency maps. *arXiv preprint [arXiv:1312.6034](https://arxiv.org/abs/1312.6034)*
32. Skempton AW (1951) The bearing capacity of clays. In: *Selected papers on soil mechanics*, pp 50–59
33. AS (2009) Piling-design and installation. Standards Association of Australia
34. Tang C et al (2019) Statistical analyses of model factors in reliability-based limit-state design of drilled shafts under axial loading. *J Geotech Geoenviron Eng* 145(9):04019042
35. Wang ZZ (2022) Deep learning for geotechnical reliability analysis with multiple uncertainties. *J Geotech Geoenviron Eng* 148(4):06022001
36. Wang L et al (2023) Deep learning methods for time-dependent reliability analysis of reservoir slopes in spatially variable soils. *Comput Geotech* 159:105413
37. Wang Z-Z, Goh SH (2021) Novel approach to efficient slope reliability analysis in spatially variable soils. *Eng Geol* 281:105989
38. Wu C et al (2022) Prediction of wall deflection induced by braced excavation in spatially variable soils via convolutional neural network. *Gondwana Res* 123:184–197
39. Yi P et al (2023) Probabilistic failure envelopes of monopiles in scoured seabed based on a new nonstationary random field model. *J Geotech Geoenviron Eng* vol 149(6)
40. Zhang Y et al (2025) Robust calibration of shaft and base resistance factors for piles based on multiobjective optimization. *J Geotech Geoenviron Eng* 151(3)
41. Zhang LM et al (2001) Reliability of axially loaded driven pile groups. *J Geotech Geoenviron Eng* 127(12):1051–1060
42. Zhang J et al (2014) Calibrating cross-site variability for reliability-based design of pile foundations. *Comput Geotech* 62:154–163
43. Zhang J et al (2021) Deep learning-based evaluation of factor of safety with confidence interval for tunnel deformation in spatially variable soil. *J Rock Mech Geotech Eng* 13(6):1358–1367
44. Zhang JZ et al (2022) Hybrid machine learning model with random field and limited CPT data to quantify horizontal scale of fluctuation of soil spatial variability. *Acta Geotech* 17(4):1129–1145
45. Zhang Y et al (2023) Bayesian updating on resistance factors of H-Piles with axial load tests. *Comput Geotech* 159:105421
46. Zhang Y et al (2024) Updating reliability of pile groups with load tests considering spatially variable soils. *Georisk Assess Manag Risk Eng Syst Geohazards* 18:1–15
47. Zhang Y, Huang J (2022) Calibration of resistance factor based on pile load test conducted to failure. In: *8th International symposium on geotechnical safety and risk (ISGSR)*, Newcastle, Australia
48. Zhang Q-Q, Zhang Z-M (2012) Simplified calculation approach for settlement of single pile and pile groups. *J Comput Civ Eng* 26(6):750–758

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.