

## Research Article

# Correlation Analysis and Prediction of the Physical and Mechanical Properties of Coastal Soft Soil in the Jiangdong New District, Haikou, China

Yongchang Yang,<sup>1,2</sup> Xinying Song,<sup>3</sup> Shuai Zhang ,<sup>4</sup> Jun Hu ,<sup>3</sup> Ming Ruan,<sup>5</sup> Dongling Zeng,<sup>1,2</sup> Han Luo,<sup>1,2</sup> Jiangsi Wang,<sup>1,2</sup> and Zhixin Wang<sup>1,2</sup>

<sup>1</sup>Hainan Key Laboratory of Marine Geological Resources and Environment, Haikou 570206, China

<sup>2</sup>Hainan Institute of Hydrological and Geological Engineering Exploration, Haikou 571100, China

<sup>3</sup>School of Civil Engineering and Architecture, Hainan University, Haikou 570228, China

<sup>4</sup>School of Information and Communication Engineering, Hainan University, Haikou 570228, China

<sup>5</sup>Hainan Institute of Geological Survey, Haikou 571100, China

Correspondence should be addressed to Shuai Zhang; zhangshuai@hainanu.edu.cn

Received 9 October 2023; Revised 6 December 2023; Accepted 11 December 2023; Published 3 January 2024

Academic Editor: Yi Zhang

Copyright © 2024 Yongchang Yang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The compressibility and shear strength of soil play a crucial role in engineering design and construction. For this study, samples were collected from the indoor geotechnical tests conducted on the fourth layer of the third series of the Haikou Formation. By conducting a correlation analysis of various physical properties of soil and utilizing the random forest algorithm, we developed a predictive model for the compressibility and shear strength of coastal soft soil. Initially, we proposed an empirical formula that utilizes mathematical statistical analysis methods to characterize the correlation between the indicators of this soil. Subsequently, we employed the feature selection guided by the aforementioned data analysis results to establish a random forest model. This model predicts the compressive modulus, compressibility coefficient, cohesion, and internal friction angle of the soil. The results indicate that the established model exhibits strong predictive capabilities, with the mean squared error values of compression modulus (0.012), compression coefficient ( $1.21 \times 10^{-6}$ ), cohesion (0.081), and internal friction angle (0.003). The data analysis methods, fitting parameters, empirical formulas, and random forest model employed in this study hold substantial value in guiding the preliminary evaluation stage of engineering projects with limited data. This study helps to save time and cost of geotechnical investigation for soft soils in the area.

## 1. Introduction

Estimating the mechanical parameters for civil construction projects based on measured soil's physical parameters is crucial for proposing appropriate design parameters, establishing scientific and reasonable calculation models, selecting foundation pits with favorable safety and economic adaptability, and determining the related support modes [1]. Moreover, it can help to avoid tedious, time-consuming, and expensive laboratory measurements while also reducing construction time and cost. Therefore, it is of significant theoretical and practical importance to collect mathematical statistics on the physical and mechanical properties of regional and representative strata and establish their predictive model [2–4].

Many scholars have conducted a great deal of research on the correlation between rock's physical and mechanical indexes and achieved substantial results. As early as the 1950s, Chinese researchers have systematically concluded the correlations of soft soil's shear strength indexes with void ratio and plasticity index, which have played an important role in engineering construction in Shanghai [5]. According to related engineering geological data of Shanghai, several sets of practical correlations between physical and mechanical properties are derived [6]. Bai et al. [7] investigated the effect of plastic index on the compression deformation parameters of saturated soft clay and concluded the linear fitting relations of compression index, swelling index, and secondary consolidation coefficient with the plastic

index. Tian et al. [8] focused on clay soil in Beijing and performed statistics of physical and mechanical indexes. Li et al. [9] analyzed the relations of the internal friction angle of soft soil in the southern part of Kunming with various physical indexes. Xiaoliang et al. [10] investigated the relations of soil's compression indexes with the number of loading or unloading cycles and saturability. Jing et al. [11] concluded the correlations of the variations of internal friction angle and cohesive force with some physical parameters including plasticity and liquidity index. Xuchang et al. [12] preliminarily investigated the correlation between physical and mechanical performance indexes of soil in Yangzhou. Linping et al. [13] investigated the correlation between soil's physical and mechanical indexes of clay soil in Binhai new district of Tianjin. Xianwei et al. [14] presents the relevance and correlation analysis on the physical and mechanical indexes of Zhanjiang clay.

The above study extensively analyzes the correlation between the physical and mechanical properties of soil, which greatly simplifies geotechnical engineering analysis. However, most of these studies mainly focus on analyzing the correlation of individual factors. Considering multiple factors simultaneously and obtaining accurate correlations are challenging due to the complexity and uncertainty of the soil [15].

Machine learning methods like artificial neural networks (ANN), support vector machines, and random forests have gained significant attention in geotechnical engineering due to their ability to efficiently and accurately map highly nonlinear problems [16, 17]. Zhang et al. [18] developed a non-parametric ensemble artificial intelligence approach to calculate the  $E_s$  of soft clay. The mean squared error and correlation coefficient of the model applied to the testing set were 0.13 and 0.91. Pham et al. [19] investigated and compared the performance of four machine learning methods, particle swarm optimization—adaptive network based fuzzy inference system (PANFIS), genetic algorithm—adaptive network based fuzzy inference system (GANFIS), support vector regression, and ANN, for predicting the strength of soft soils. And concluded that out of four models, the PANFIS indicates as a promising technique for prediction of the strength of soft soils [19]. Taffese and Abegaz [20] used machine learning techniques to predict the compaction and strength properties of amended soil. Li et al. [21] compared the performance of random forest regression and artificial neural network, two commonly used machine learning methods, for predicting soil properties. They found that the random forest regression method generally yielded smaller prediction errors [21].

Previous studies on machine learning mainly focus on predicting individual indices such as compression parameters and strength parameters. The analysis lacks machine learning models that can comprehensively predict compression and shear strength indices. Moreover, these studies have not yet analyzed the prediction of compression and shear strength indices specifically for soft soils in Haikou City, Hainan Province. Jiangdong new district in Haikou, as a pilot demonstration region in Hainan free trade port, has begun a great number of engineering plans and constructions at present. In the future, the construction scale of various projects will be continuously expanded. In order to better exploit and

develop underground space, reasonably save the engineering construction cost, shorten engineering construction period, and accumulate rich regional empirical parameters, this study focused on the fourth member sedimentary soil in the Tertiary Haikou Formation (with the lithology of cohesive soil and extensive distribution in Jiangdong district, Haikou) and conducted geotechnical tests for statistical analysis. The main contributions of this study can be summarized as follows.

(1) A dataset was prepared based on investigation reports in Haikou. (2) Studied the correlation among the physical indexes, compressibility indexes, and shear strength indexes of this stratum by means of mathematical statistical analysis. (3) A random forest regression algorithm in machine learning was used to develop a model that can predict soil compression and shear strength indicators. (4) The predictive performance of ML methods and engineering measured data was compared for evaluating model accuracy.

## 2. Sampling and Test Statistics

*2.1. Engineering Situations Data Source.* The present geotechnical data were sourced from the project Investigation and Evaluation of Underground Space Development and Utilization Potential for Jiangdong New District, Haikou. The project is a subproject of the Comprehensive Survey of Urban Geology in Jiangdong New District, Haikou, organized and implemented by Hainan Provincial Bureau. According to the regional geological data of Jiangdong New Area, the rock and soil mass that overlays the sedimentary soil layer in the fourth section of the Tertiary Haikou Formation is predominantly composed of quaternary sea-land alternating sedimentary soil. This layer is known to be problematic for engineering purposes, as it consists mostly of severely liquefied sand and seismic soft soil. The sedimentary soil in the fourth section of the Tertiary Haikou Formation serves as the primary pile end bearing layer for regional engineering construction, as well as the main layer for the development and utilization of underground space. The sampling was conducted at a depth of 100 m, with a total of 182 boreholes drilled.

*2.2. Test Method.* The limit moisture ratio was measured by cone penetrometer via rolling. The moisture content when the cone with a weight of 76 g sank by 10 mm was set as the liquid limit, while the moisture content when the fractures appeared as the soil stripe was rubbed to 3 mm and fracture was set as the plastic limit. The difference between liquid limit and plastic limit was defined as the plasticity index. The test data of compressibility indexes were measured with standard consolidation test (at a pressure of 100–200 kPa). According to Standard for Geotechnical Testing Method (GB/T 50123-1999), consolidated quick direct shear test was performed.

*2.3. Sampling Method and Statistical Analysis.* During the present field drilling process, rotary drilling with mud protection wall was adopted. Rock samples were collected with the core barrel ( $\phi 91$  mm) while soil samples were collected with single-action triple tube. In this study, 279 sedimentary soil samples in the fourth member of the Tertiary Haikou

Formation were collected for statistics. The soil samples were mainly cohesive soil (silty clay or clay). Considering the actual engineering applications at present, the distribution range and possible depth of the soil layer, the sampling depth was controlled within 100 m (ranging from 7.50 to 97.00 m). According to the parameter statistical method as described in Code for Investigation of Geotechnical Engineering (GB50021-2001, the 2019 Edition), the statistics of basic physical and mechanical parameters of this layer of soil were obtained and listed in Table 1.

Apparently, the variation coefficients of some physical indexes including soil density, moisture content, and void ratio were all smaller than 0.300, suggesting the reasonability of soil layer division. The variation coefficients of compressibility and shear strength indexes were mostly larger than 0.300. This is mainly due to that partial perturbation or soil stress release may exist in postpreparation of samples, thereby leading to great variations in sample parameters.

Overall, the test data of samples were quite reliable and practical. It was feasible to perform correlation analysis on these data for regional design suggestions and empirical calculation.

### 3. Overall Analysis of Correlation of Soil's Parameters

The engineering characteristics of soil in soil mechanics are mainly directly reflected by its physical and mechanical indexes. Therefore, statistical analysis and summary of the measured indexes of the same stratum in the region is of great practical significance for the accumulation of regional geological experience and engineering practice experience.

Based on previous statistical analysis experiences of geotechnical data, the correlation among soil's physical and mechanical indexes generally can be described by linear models [22]. This study adopted least square linear fitting and unary linear regression for analysis. First, the correlations among various indexes of the collected soil samples were judged. Overall correlation analysis between soil sampling position and various test indexes was performed, and the results are shown in Table 2. Various indexes ranked from weak to strong correlation. The detailed parameters were then analyzed.

Based on the above statistics of correlation coefficients and significance test results, the following conclusions can be drawn.

- (1) Except the moderate correlation with wet density, moisture content, and void ratio, soil's sampling depth was weakly correlated with consistency, plasticity index, compressibility, and shear strength indexes.
- (2) Wet density and moisture content were highly correlated with void ratio, which can satisfy basic conversion indexes of three-phase indexes. Wet density, moisture content, and void ratio were moderately–strongly correlated with compressibility and shear strength indexes.

- (3) The compressibility indexes including the modulus of compression and compression coefficient showed exactly different correlations with all parameters. This is consistent with the definitions of soil's compressibility indexes. A greater modulus of compression suggests stronger deformation resistance ability, which corresponds to a smaller compression coefficient.
- (4) The shear intensity indexes showed consistent positive/negative correlations with the other parameters. The shear strength indexes exhibited negatively moderate–strong correlations with moisture content and void ratio, as well as negative weak–moderate correlations with liquidity and plasticity indexes.
- (5) The compressibility indexes were overall moderately–strongly correlated with the shear strength indexes. More favorable compressibility indexes suggest stronger shear strength indexes and better geological properties of soil engineering.

Based on the above preliminary analysis results, the correlation of soil's three-phase measured indexes with sampling depth, shear strength, and compressibility indexes were analyzed in depth. In addition, high-pressure consolidation test was performed on 120 soil samples for exploring the correlation between preconsolidation pressure and the foundation's bearing capacity.

### 4. Correlation between Soil's Three-Phase Indexes and Sampling Depth

The correlations among the measured wet density, moisture content and void ratio, and the sampling depths of all soil samples were investigated, as the statistical scatter diagrams are shown in Figures 1–4.

Overall, wet density was in direct proportion to the sampling depth, while moisture content and void ratio were inversely proportional to the sampling depth. Through preliminary analysis, the soil samples were collected from old clay and can be regarded as normal consolidated–overconsolidated soil. The mechanical indexes such as plasticity, compressibility, and shear strength indexes were relatively stable. As the sampling depth and overlying soil pressure increased, natural density increased gradually while moisture content and void ratio dropped gradually. In terms of negative/positive correlation, as the sampling depth increased, both compressibility and shear strength indexes were improved. This also conforms to soil's sedimentary rules of underconsolidated–consolidated–overconsolidated transition.

Meanwhile, it can be observed from the scatter diagrams that void ratio and moisture content overall showed identical variation rules. Figures 3 and 4 show the scatter diagram of the correlation between void ratio and moisture content. Linear correlation can be observed, suggesting that pores in the soil were almost filled by water, with almost no void content. The results fit well with the measured saturation value from 79 to 100, with a mean value of 93.93. The soil can be judged as saturated soil.

TABLE 1: Statistics of physical and mechanical indexes of the fourth member sedimentary soil of the tertiary Haikou Formation.

Statistical method	Wet density		Dry density	Moisture content		Porosity ratio		Saturation		Specific gravity of soil particles		Liquid limit
	$\rho$ (g/cm <sup>3</sup> )	$\rho_w$ (g/cm <sup>3</sup> )	$\rho_d$ (g/cm <sup>3</sup> )	$\omega_0$ (%)	$\omega_1$ (%)	$e_0$	Sr (%)	$G_s$	$G_s$	$\omega_L$ (%)		
Number of samples	235	235	235	279	279	176	176	176	279	279	279	
Maximum	2.09	1.11	1.11	20.6	20.6	0.545	83	83	2.65	23.3	23.3	
Minimum	1.66	1.72	1.72	51.7	51.7	1.433	100	100	2.73	68.1	68.1	
Mean	1.88	1.44	1.44	30.8	30.8	0.885	94	94	2.7	36.4	36.4	
Standard deviation	0.084	0.14	0.14	7.737	7.737	0.204	3.789	3.789	0.019	10.321	10.321	
Coefficient of variation	0.044	0.097	0.097	0.251	0.251	0.231	0.04	0.04	0.007	0.284	0.284	
Standard value	—	—	—	—	—	—	—	—	—	—	—	
Statistical method	Plastic limit		Plasticity index	Liquidity index		Modulus of compression		Compression coefficient		Cohesive force		Internal friction angle
	$\omega_p$ (%)	$I_p$	$I_p$	$I_L$ (%)	$Es_{1-2}$ (MPa)	$Es_{1-2}$ (MPa)	$\sigma_{1-2}$ (MPa) <sup>-1</sup>	C	C	$\phi$ (Degree)		
Number of samples	279	279	279	279	177	177	177	154	154	154		
Maximum	13.2	8	8	0.13	3.86	3.86	0.025	20.2	20.2	6		
Minimum	37.3	33.4	33.4	2.29	68.24	68.24	0.602	168	168	38		
Mean	20	16.4	16.4	0.71	21.64	21.64	0.127	78.4	78.4	23.2		
Standard deviation	4.961	5.828	5.828	0.316	12.059	12.059	0.097	35.388	35.388	6.727		
Coefficient of variation	0.248	0.356	0.356	0.447	0.557	0.557	0.766	0.451	0.451	0.289		
Standard value	—	—	—	—	—	—	—	73.5	73.5	22.3		

TABLE 2: Statistics of the Pearson's correlation coefficient among sample's measured parameters.

	Sampling depth	Wet density	Moisture content	Porosity ratio	Plasticity index	Liquidity index	Modulus of compression	Compression coefficient	Cohesive force	Internal friction angle
Sampling depth	1	—	—	—	—	—	—	—	—	—
Wet density	0.529	1	—	—	—	—	—	—	—	—
Moisture content	-0.455	-0.912	1	—	—	—	—	—	—	—
Porosity ratio	-0.476	-0.957	0.998	1	—	—	—	—	—	—
Plasticity index	-0.158	-0.649	0.754	0.752	1	—	—	—	—	—
Liquidity index	-0.344	-0.239	0.182	0.170	-0.426	1	—	—	—	—
Modulus of compression	0.252	-0.426	-0.389	-0.409	-0.297	-0.125	1	—	—	—
Compression coefficient	-0.216	-0.491	0.524	-0.547	-0.298	-0.236	-0.532	1	—	—
Cohesive force	0.233	0.459	-0.453	-0.454	-0.269	-0.243	0.788	-0.440	1	—
Internal friction angle	0.213	0.593	-0.686	-0.681	-0.675	-0.088	0.532	-0.496	0.613	1

Note: For the calculated correlation coefficient, the values of over 0.8, 0.6~0.8, 0.4~0.6, 0.2~0.4, and below 0.2 are indicative of high correlation, strong correlation, moderate correlation, weak correlation, and extremely weak correlation or no correlation, respectively.

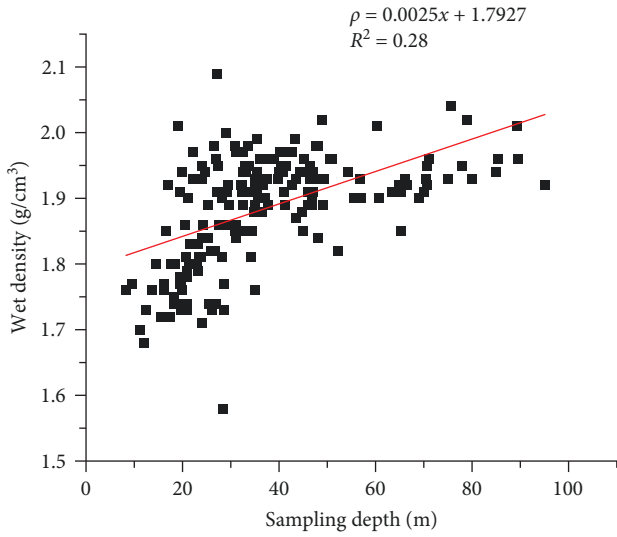


FIGURE 1: Scatter plot of wet density and sampling depth.

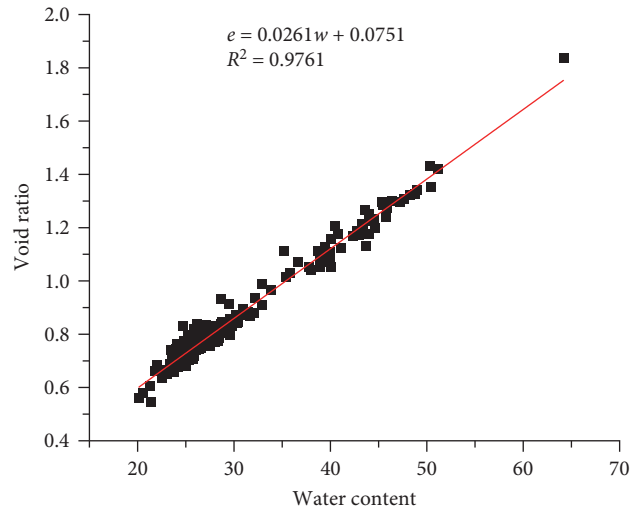


FIGURE 4: Scatter plot of void ratio and water content.

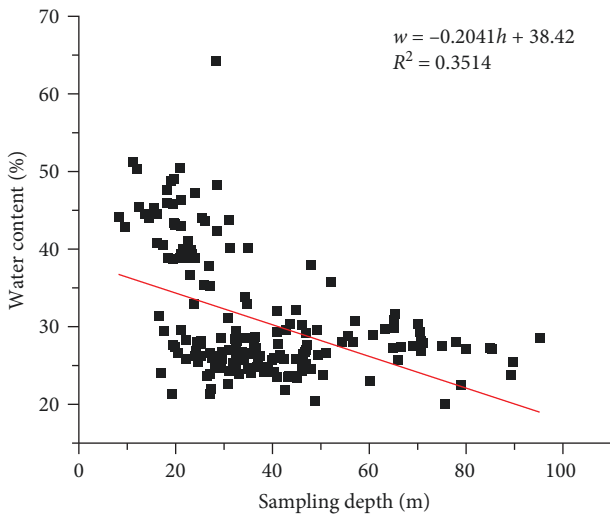


FIGURE 2: Scatter plot of water content and sampling depth.

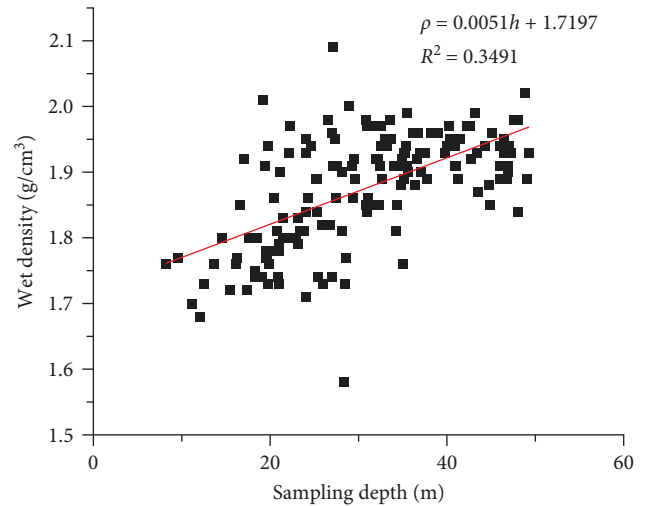


FIGURE 5: Scatter plot of wet density and sampling depth.

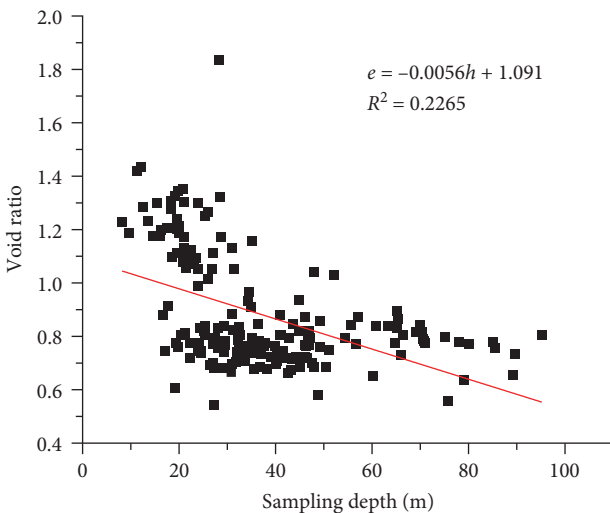


FIGURE 3: Scatter plot of void ratio and sampling depth.

In terms of consolidation procedure of consolidated soil, when soil was consolidated to a certain degree, various pores and voids were almost compressed or filled by bound water. Accordingly, both moisture content and void ratio were gradually fixed. As shown in Figures 1–3, for the soil samples with a sampling depth of above 50 m, the correlations of wet density, moisture content, and void ratio with the sampling depth were higher than the correlations for the samples with a sampling depth of below 50 m. After eliminating the samples with a depth of below 50 m, linear fitting was performed on the correlations of wet density, moisture content, and void ratio with the sampling depth, as the fitting formulas and correlation coefficients listed in Figures 5–7.

Based on the above analysis results, for the soil samples from the fourth member of the Tertiary Haikou Formation from Jiangdong new district, Haikou, the related empirical formulas that describe the correlations of dry density, moisture content, and void ratio with the sampling depth can be written as follows:



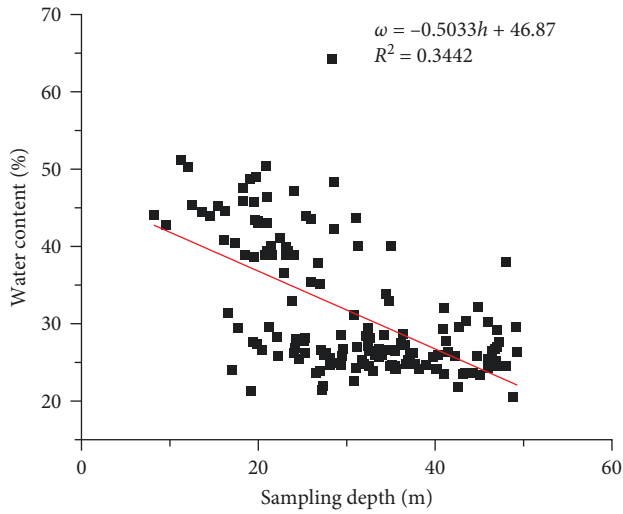


FIGURE 6: Scatter plot of water content and sampling depth.

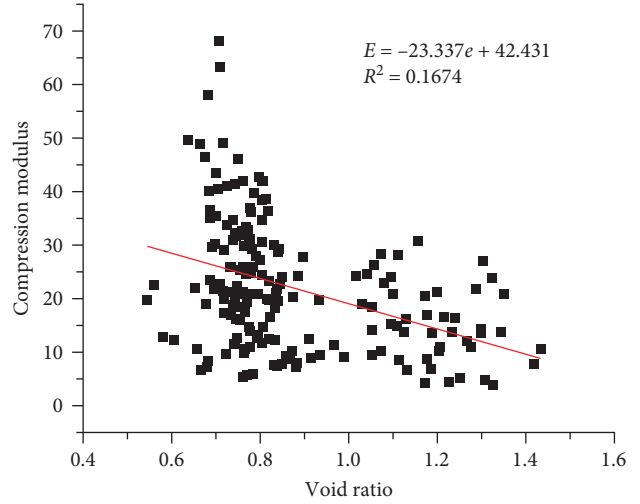


FIGURE 8: Scatter plot of void ratio and compression modulus.

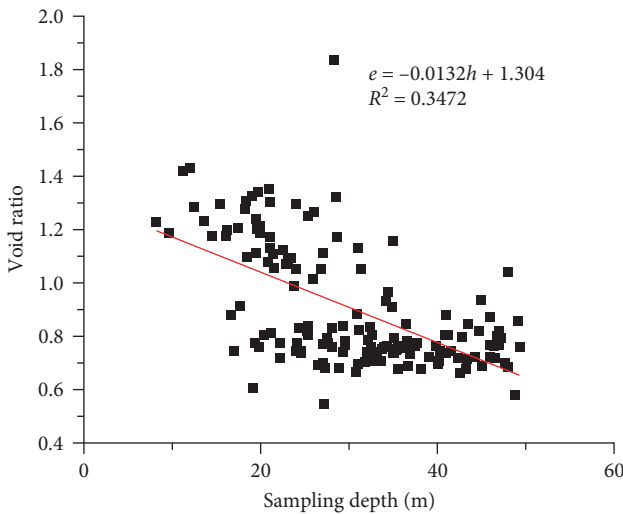


FIGURE 7: Scatter plot of void ratio and sampling depth.

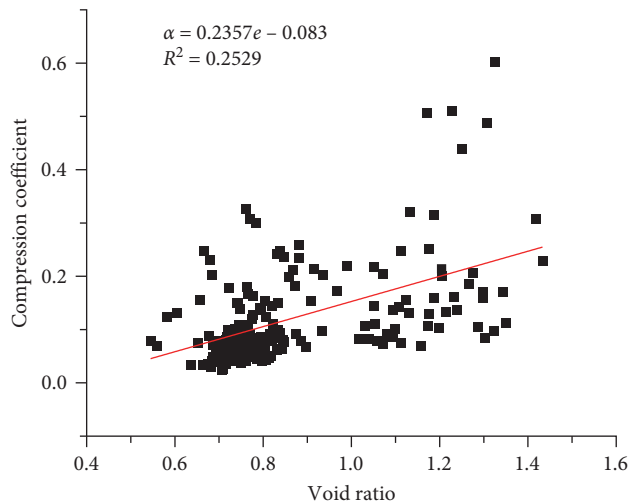


FIGURE 9: Scatter plot of void ratio and compression coefficient.

(1) When  $h \leq 50$  m:

$$\rho = 0.0051h + 1.7194, \omega = 0.5033h + 46.87, e = 0.0132h + 1.304. \quad (1)$$

When  $h > 50$  m,  $\rho$ ,  $\omega$ , and  $e$  can be directly set as 1.969, 28.5, and 0.787. The values are basically coincident with the statistical averages of 1.930, 27.6, and 0.784 in the 48 group samples below 50 m.

(2) The empirical formula between void and moisture content can be written as follows:

$$e = 0.0261\omega + 0.0751. \quad (2)$$

### 5. Correlations of Void Ratio with Soil's Mechanics and Displacement Index

According to previous research results, for the soil samples collected from this layer, the void ratio is moderately–strongly correlation with compressibility and shear strength indexes, as the detailed statistics are shown in Figures 8–11.

Generally, under the additional stress, free-state underground can be discharged from the pores of soil on account of the fluidity, thereby leading to volume reduction and inducing compression [23, 24]. This can account for soil compressibility. The deformation is then referred to as consolidation. For ordinary foundations, the sedimentation and deformation are always designed by considering both compression modulus and coefficient [25].

Soil's shear strength refers to soil's capability to resist shear failure and equals to the shear stress on the sliding surface when shear failure occurs in soil. Certainly, whether

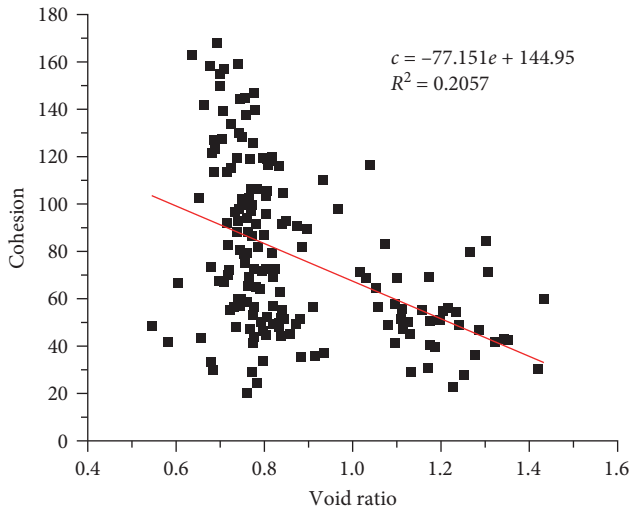


FIGURE 10: Scatter plot of void ratio and cohesion.

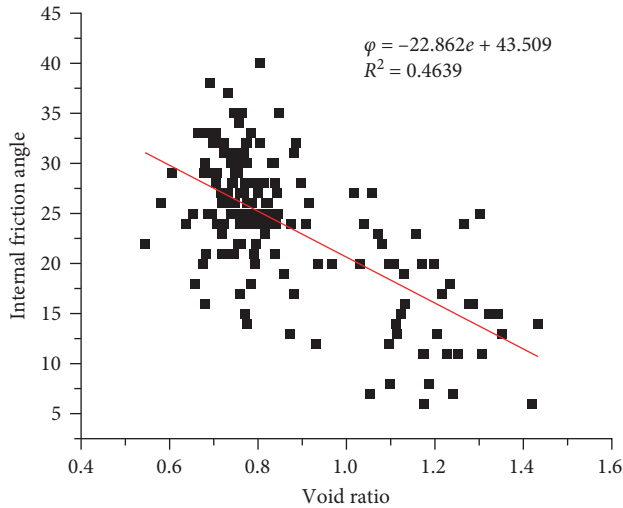


FIGURE 11: Scatter diagram of void ratio and internal friction angle.

soil reaches the shear failure state not only depends on soil properties but also is closely correlation with the applied stress combination [26]. Therefore, the indexes should be selected in combination with actual engineering condition (mainly, the drainage condition). According to the three-phase measured results, the collected soil samples were saturated soil and water in soil was mostly bound water, with poor drained. The quick direct shear indexes in this study are of great significance to practical applications [27].

However, in actual production, compressibility and shear tests are always time-consuming, with great difficulty in sample collection. At the preliminary engineering design phase, the compressibility and shear strength indexes can be reasonably derived in combination with the detailed burial depth and void ratio for further estimation of foundation sedimentation and stability. This is quite significant for the design of engineering exploration schemes and foundations.

It can be observed in Figures 8–11, for the fourth member sedimentary soil of Haikou Formation from Haikou new

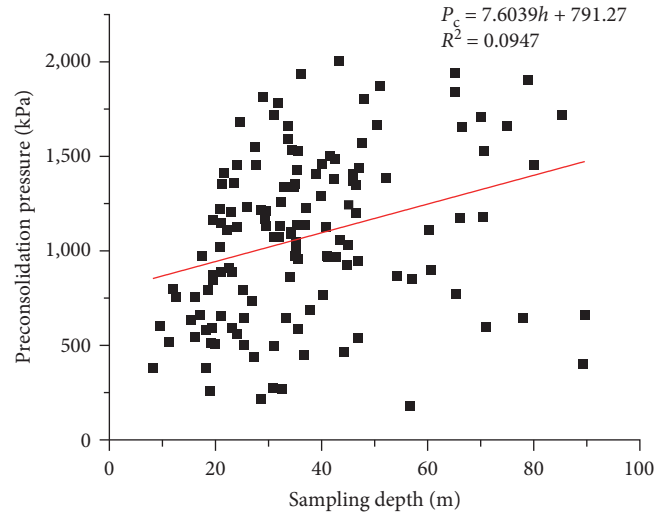


FIGURE 12: Scatter diagram of preconsolidation pressure and sampling depth.

district, Haikou, the empirical formulas of compressibility and shear strength indexes can be written as follows.

Compressive indexes:

$$eEs = -23.337e + 42.431, \alpha_{1-2} = 0.2357e - 0.083. \quad (3)$$

Shear strength indexes (direct shear test):

$$c_k = -77.151e + 144.95, \varphi_k = -22.862e + 43.509. \quad (4)$$

Soil layer in natural world has undergone ever-changing consolidation history in long geological history; however, the soil has endured the maximum pressure and reached certain consolidation degree. The maximum pressure is exactly the abovementioned preconsolidation pressure. Considering that overconsolidated soil samples were collected, 120 samples with measured preconsolidation pressures and consolidation indexes were selected for statistics for gain better understanding of soil's sedimentary history, estimate foundation sedimentation, evaluating the characteristic value of foundation bearing capacity and propose reasonable and economic foundation scheme. The correlation between preconsolidation pressure and sampling depth and the correlations of void ratio with preconsolidation pressure and consolidation index were analyzed, as the results shown in Figures 12–14.

As shown in Figure 12, the preconsolidation pressure was quite weakly correlated with the sampling depth, indicating that soil was subjected to low external force in sedimentary history, under relatively stable state. The statistical results of  $P_c$  can be described below, a range from 178.5 to 2003.7 kPa, a mean value of 1074.4kPa, and a variation coefficient of 0.408.

Generally, high-pressure consolidation test is performed to measure the pressure and the compressibility indexes. The test process and parameter calculation are quite time-consuming. It can be observed from Figures 13 and 14 that soil's consolidation index was quite strongly correlated with void



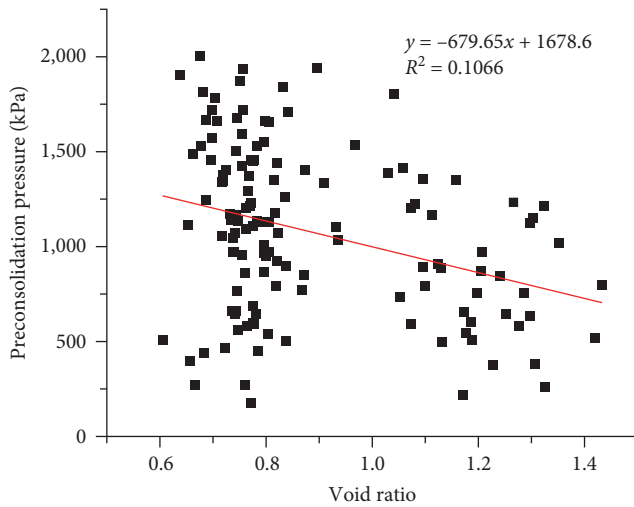


FIGURE 13: Scatter diagram of void ratio and preconsolidation pressure.

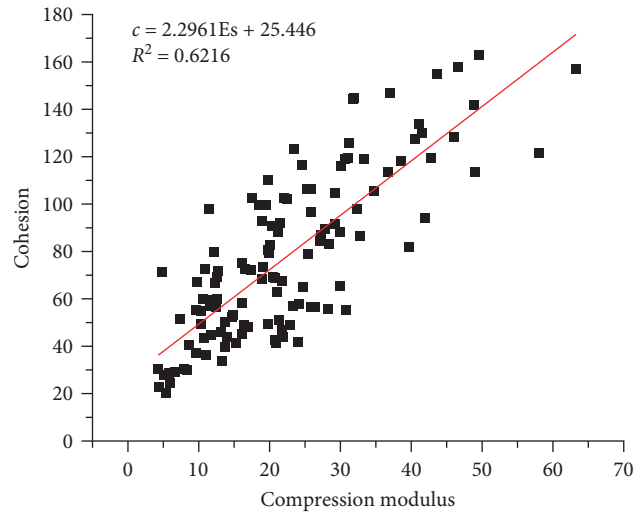


FIGURE 15: Scatter plot of cohesion and compression modulus.

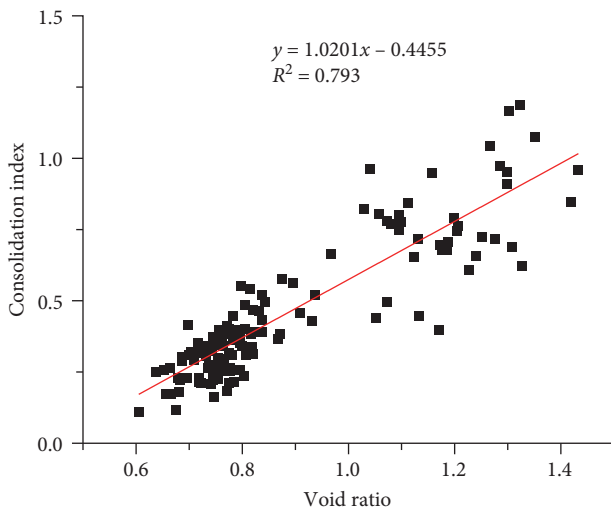


FIGURE 14: Scatter plot of void ratio and consolidation index.

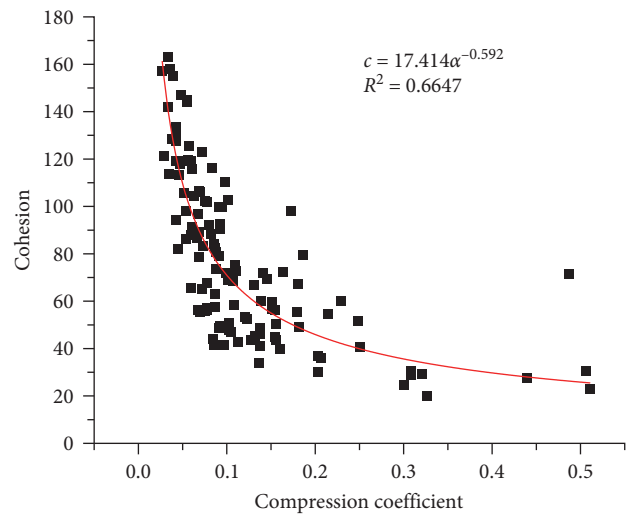


FIGURE 16: Scatter plot of cohesion and compression coefficient.

ratio, and moderately correlated with the preconsolidation pressure. The following empirical formula can be used at the beginning or in preliminary stage of engineering construction:

$$P_c = -67.65e + 1678, C_c = 1.0201e - 0.4455. \quad (5)$$

### 6. Correlation between Shear Strength and Compressibility Indexes

The compressibility index reflects soil’s consolidation-induced deformation while the shear strength index reflects soil’s shear-induced deformation [28, 29]. Under the deformation induced by both consolidation and shear, pores in the soil can be compressed, water flows out and soil particles move [30, 31]. Although the different deformation mechanisms, these two parameters are correlated to certain degree,

as listed in Table 2. The internal correlation should be analyzed, which can also be used for cross-verification in parameter design at the early stage of engineering construction. Figures 15–18 display the statistical results.

It can be observed that the linear correlation coefficients of the modulus of compression with internal friction angle and cohesive force were 0.5317 and 0.7888, respectively, suggesting moderate–strong correlation; the correlation of compression coefficient with internal friction angle and cohesive force can be reasonably fitted by power functions, with a correlation coefficient of 0.6489 and 0.8152, respectively, suggesting strong correlation. Overall, strong correlation between compressive and shear strength indexes can be observed.

The main reasons can be described below.

- (1) Soil samples can be regarded as saturated clay soil. During the consolidation-induced deformation process, pores were almost filled by water, with quite small voids.

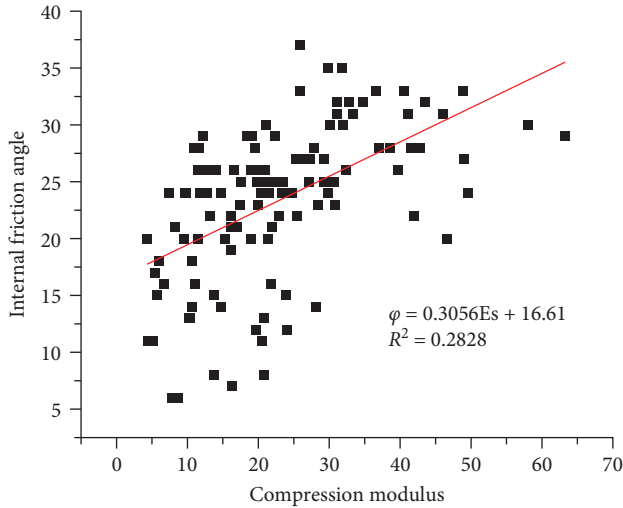


FIGURE 17: Scatter diagram of internal friction angle and compression modulus.

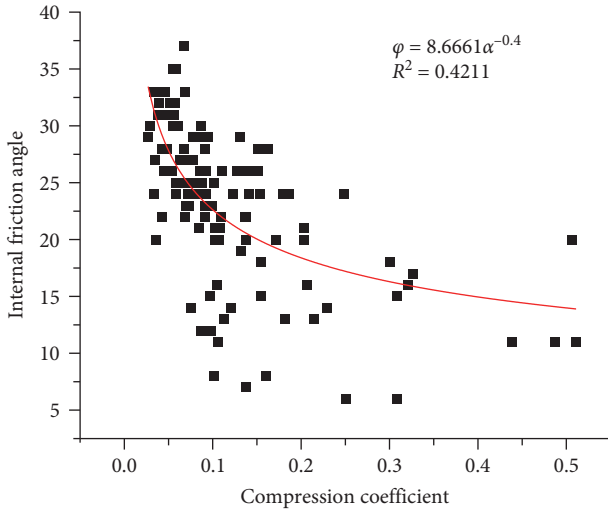


FIGURE 18: Scatter diagram of internal friction angle and compression coefficient.

Water discharge under squeezing played a dominant role in consolidation deformation under compression.

- (2) Soil particles were extruded during soil's vertical consolidation deformation process. There existed mutual shearing and friction among particles during the deformation process.
- (3) Water among soil particles was mostly bound water, with strong adsorption capacity with soil particles. Great shear friction can be found in water drainage process.
- (4) Soil compression on macroscopic level corresponds to shear deformation among soil particles on microscopic level. Accordingly, compressibility and shear strength indexes show strong correlation.

In actual engineering applications, when the measured indexes show great difference under great disturbance on

sample collection, the indexes can be validated according to the following empirical formulas:

$$\begin{aligned} C_k &= 2.2961Es_{1-2} + 25.446, C_k = 17.414\alpha_{1-2}^{-0.592}, \\ \varphi_k &= 0.3056Es_{1-2} + 16.61, \varphi_k = 8.6661\alpha_{1-2}^{-0.4}. \end{aligned} \quad (6)$$

## 7. Predict Soft Soil Parameters Using Random Forest Regression

Based on the analysis presented above, it is evident that soil's three-phase index and other factors are correlated with its compression and shear strength indicators. Previous sections have quantitatively examined these relationships. However, Figures 1–18 shows a significant amount of dispersion in the measured data. Consequently, the quantitative relationship curves for various parameters are deficient in accurately representing the data. Estimating the compression and strength properties of soils based only on a single metric and a simple linear fit is problematic. The deformation and strength indicators of soil exhibit complex nonlinear relationships with soil properties. Thus, there is a need to comprehensively characterize the relationships between soil properties and its deformation and strength indicators. Given the exceptional performance of machine learning in fitting nonlinear complex relationships, the random forest algorithm in machine learning is chosen as a regression method specifically designed for high-dimensional soil parameters data. The training process of random forest involves “randomness” and “ensemble” effects, enabling it to accurately capture the randomness and diversity of soil parameters. The specific modeling process (Figure 19) is described below:

- (1) Preparing the dataset, based on the results of Sections 3–6 on the correlation of soil parameters, the following parameters are selected as inputs for the model: depth of sampling, wet density, moisture content, void ratio, plasticity index, and liquidity index. The output variables for prediction are compression modulus, compression coefficient, cohesion, and internal friction angle. Overall, a total of 185 data points were assembled for analysis. The dataset was then divided into a training set, comprising 70% of the data, to facilitate model training, and a test set, accounting for 30% of the data, to assess the model's generalization ability.
- (2) The bootstrap sampling method is used to randomly select samples with replacement from the original data set collected from the site of the engineering project. This method aims to reduce the sample dependency of the data set, thereby improving the robustness of the model. In each sampling, a subset of features is randomly selected. By controlling the number and types of features, it is possible to effectively reduce the complexity of the model and avoid overfitting.

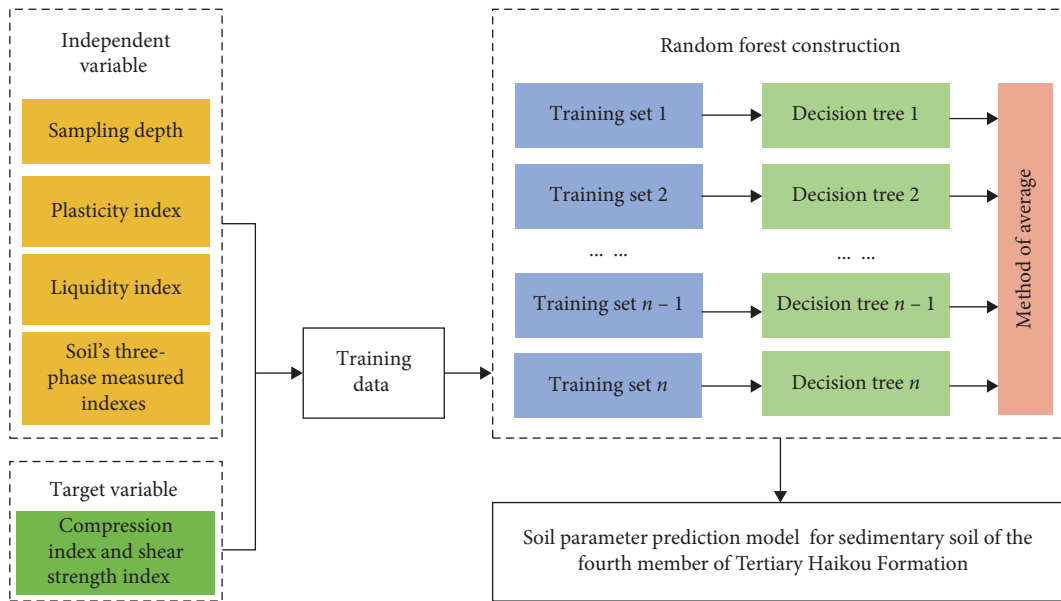


FIGURE 19: Random forest regression model workflow.

TABLE 3: MSE error of measured values and predicted value of random forest regression models.

Parameters	Compression modulus	Compression coefficient	Cohesion	Internal friction angle
MSE	0.012	$1.21 \times 10^{-6}$	0.081	0.003

- (3) Develop a decision tree model based on the chosen samples and features using CART algorithm. CART builds a binary tree structure recursively by partitioning the input space into subsets based on the values of input features. The tree is constructed in a way that each internal node represents a decision based on a feature, and each leaf node represents the output (class label for classification or numerical value for regression) for the corresponding subset of data. Since this study focuses on regression analysis, the decision tree construction process employs the Gini coefficient as the split criterion to identify the most suitable feature for node splitting. The Gini impurity measures the likelihood of misclassification. For a given node, it calculates the probability of misclassifying a randomly chosen element if it were randomly labeled according to the distribution of classes in the node. Consequently, each subtree is able to effectively capture and comprehend the regression characteristics inherent in the data.
- (4) Bootstrap sampling is used to construct subsequent decision trees until the predefined number of trees is reached (set at 500 for this study). During the construction of each tree, the previously created trees are combined through ensembling, and their average probabilities are calculated to obtain the final prediction result.

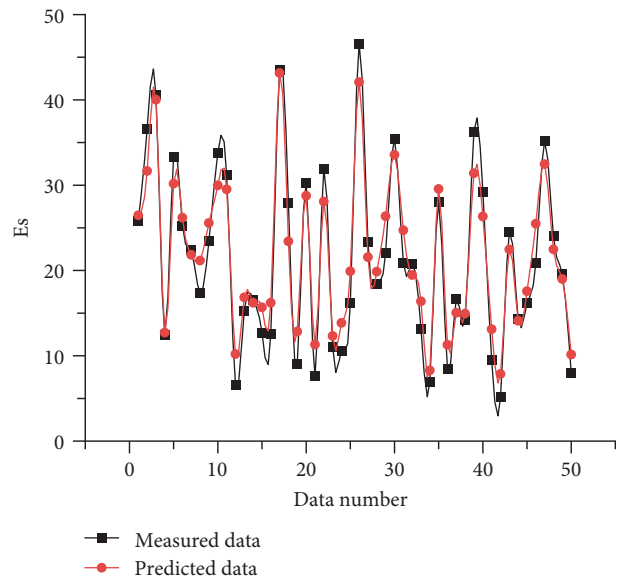


FIGURE 20: Comparison of predicted and measured values of compression modulus.

- (5) Train the model by inputting the training set and starting the training program. Upon completion of training, the model calculates the predicted results by reading the test set data. The mean square error (MSE) is used to evaluate the discrepancy between

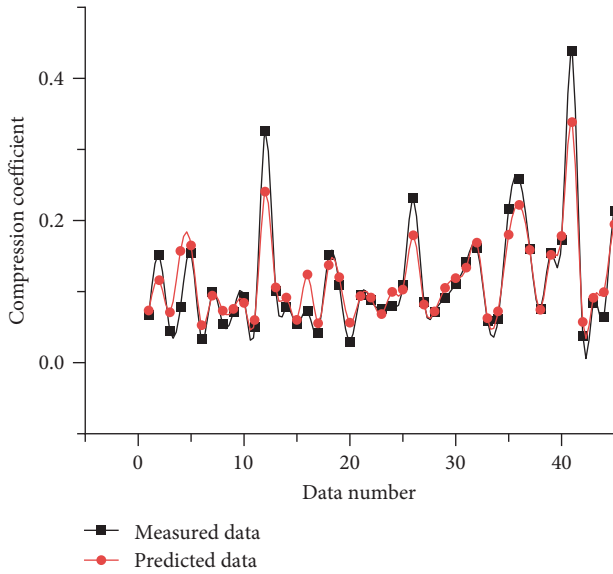


FIGURE 21: Comparison of predicted and measured values of compression coefficient.

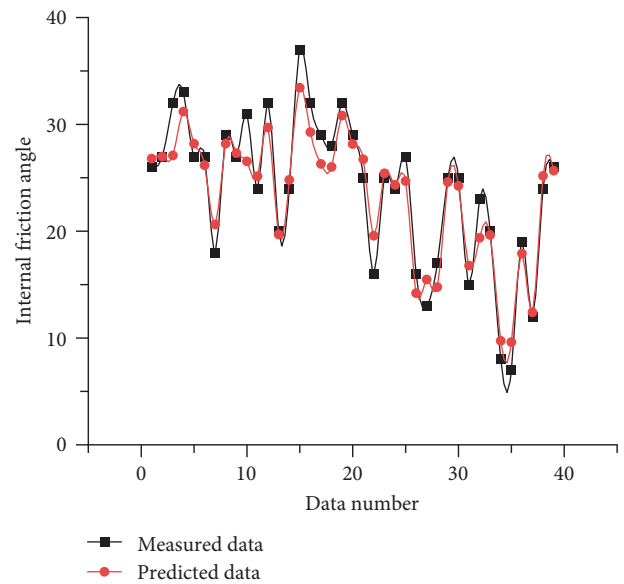


FIGURE 23: Comparison of predicted and measured values of internal friction angle.

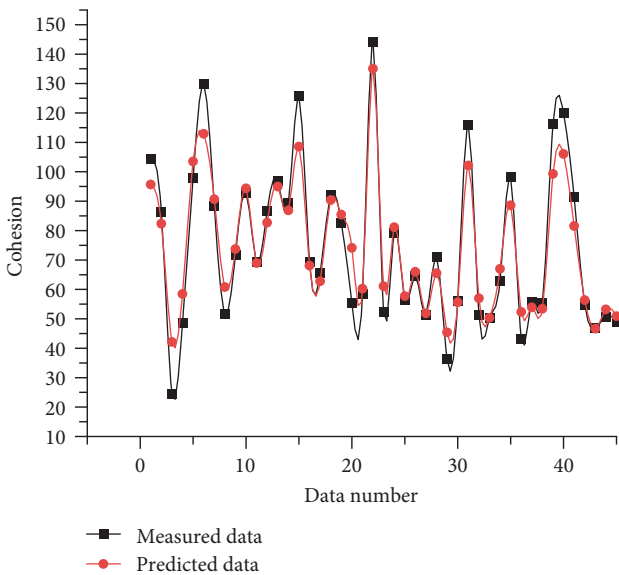


FIGURE 22: Comparison of predicted and measured values of cohesion.

the actual data and the model's predictions. The formula for MSE is given by the following equation:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (7)$$

where  $n$  is the total number of samples,  $y_i$  is the measured results, and  $\hat{y}_i$  is the predicted results of the model. After training, the model's prediction accuracy for compressibility modulus, compression coefficient, cohesive strength, and internal friction angle is shown in Table 3.

Table 3 illustrates that the mean squared error (MSE) values of the CART-based random forest regression model for predicting each parameter are all below 0.1. This implies that the model performs significantly better in comprehensively fitting each parameter compared to the correlation fitting curves discussed earlier in the section.

The test set data are inputted into the trained model to calculate the predicted output. Then, these predicted values are compared with the measured values. The results of this comparison can be found in Figures 20–23. The figure reveals a strong consistency between the predicted and measured values of compression modulus, compression coefficient, cohesion, and angle of internal friction. This consistency indicates that the model possesses a certain degree of generalization capability, allowing it to be applied in predicting the parameters of soft soil in the fourth layer of the third series of the Haikou Formation.

## 8. Conclusions

This study focused on the fourth member sedimentary soil of Haikou Formation collected from Jiangdong new district, Haikou, and conducted soil test on 279 samples. Through analysis, various fitting formulas that describe the correlations of three-phase indexes and sampling depth, void ratio and compressibility indexes, void ratio and shear strength indexes, void ratio and preconsolidation pressure, and compressibility indexes and shear strength indexes were derived. A random forest model was established to synthesize the above parameters, realizing the prediction of compression, and shear strength indexes. The present research results can provide reference for in-depth understanding of basic physical and mechanical parameters, design of exploration scheme and foundation, and the calculation of sedimentation deformation. The main conclusions are described below.

- (1) At a sampling depth of below 50 m, the three-phase measured indexes (wet density, moisture content, and void ratio) of soil samples were well correlated with the sampling depth; as the sampling depth exceeded 50 m, the three-phase measured indexes were almost fixed.
- (2) Overall, soil's void ratio was in good correlation with compressibility and shear strength indexes. As the void ratio decreased, the compression modulus increased, the compression coefficient dropped, while both cohesive force and internal friction angle increased. A smaller void ratio was indicative of greater preconsolidation pressure and smaller consolidation index.
- (3) Among the correlations between compressibility and shear strength indexes, the compression modulus was linearly correlated with cohesive force and internal friction angle, while the correlations of compression coefficient with cohesive force and internal friction angle can be described by power functions.
- (4) The data were first subjected to correlation analysis for random forest model parameter selection. Subsequently, the random forest model was developed to predict the compressibility index and shear strength index of soft soil. The model demonstrated a high level of accuracy in predicting the indices and exhibited excellent generalization ability. The research outcomes are particularly helpful in the planning and initial design stages for soft soil projects in saving time and cost.

In summary, the machine learning algorithm based on random forest regression can well predict the bearing capacity parameters and deformation parameters of coastal soft soil. However, due to the different causes and environments of the soil, this model can only be applied to Jiangdong New District of Haikou. In the future, a large amount of data support is needed to obtain a machine learning model with a wider range of applications.

### Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

### Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

### Acknowledgments

This research was funded by an Independent project of the Hainan Key Laboratory of Marine Geological Resources and Environment: Study on the Constitutive Model and Engineering Usability of Typical Sea Sands in Hainan (22-HNHYDZZYHJZZ005), the National Natural Science Foundation of China (51968019), the Hainan Provincial Natural Science Foundation Innovation Research Team Project (522CXTD511), and the High Technology Direction Project

of the Key Research & Development Science and Technology of Hainan Province, China (ZDYF2021GXJS020).

### References

- [1] Y. Wang, J. Li, Q. Jiang, Y. Huang, and X. Li, "Experimental study on variation law and mechanism of soil shear strength parameters along the slope," *Advances in Civil Engineering*, vol. 2019, Article ID 3586054, 11 pages, 2019.
- [2] P. Cheng, Y. Liu, Y. P. Li, and J. T. Yi, "A large deformation finite element analysis of uplift behaviour for helical anchor in spatially variable clay," *Computers and Geotechnics*, vol. 141, Article ID 104542, 2022.
- [3] P. Cheng, Y. Liu, Y. Zhang, X. Liu, S. Yin, and K. Yao, "Numerical modelling of the keying process for a suction embedded plate anchor in spatially varying clays," in *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, pp. 1–19, 2023.
- [4] P. Cheng, J. Guo, K. Yao, and X. Chen, "Numerical investigation on pullout capacity of helical piles under combined loading in spatially random clay," *Marine Georesources & Geotechnology*, vol. 41, no. 10, pp. 1118–1131, 2023.
- [5] S. M. Huang and D. Z. Gao, *Foundation and Underground Engineering in Soft Ground*, China Architecture and Building Press, Beijing, 2005.
- [6] L. M. Zhu and X. D. Weng, "Correlation between physical and mechanical properties of clay in Shanghai area," in *China Civil Engineering Society, et al, Proceedings of the First National Conference on Soil Mechanics and Foundation Engineering*, China Industry Press, Beijing, 1964.
- [7] B. Bai, J. Zhou, and G. Zhang, "Effect of plasticity index of saturated soft clay on its deformation parameters," *Journal of Hydraulic Engineering*, vol. 11, pp. 51–55, 2001.
- [8] X. F. Tian, J. Z. Sun, and B. Zhang, "Statistics analysis on physical and mechanical indexes of clay in Beijing," *Resources & Industries*, vol. 04, pp. 80–84, 2007.
- [9] C. C. Li, S. G. Xu, and J. Wu, "The analysis of the correlativity between the internal frictional angle and physical parameter of the soft soil in the south city, Kunming, Yunnan Province," *Science Technology and Engineering*, vol. 14, no. 3, 2014.
- [10] H. Xiaoliang, T. Xiaohui, L. Zeyong, and T. Longyu, "Compressive indices of cohesive soil and relation to degree of saturation," *Journal of Engineering Geology*, vol. 25, no. 5, pp. 1336–1343, 2017.
- [11] Z. Jing, J. Liangwei, L. Qiang, Z. Wensheng, and W. Xingzheng, "Correlative analysis between variation level of soil strength parameters and influencing factors," *Journal of Engineering Geology*, vol. 26, no. 3, pp. 592–601, 2018.
- [12] L. Xuchang, S. De'an, and X. Yongfu, "Study on correlations of physical and mechanical property indexes of Yangzhou clays," *Journal of Engineering Geology*, vol. 27, no. 2, pp. 333–340, 2019.
- [13] G. Linping, M. Kun, Y. Aiwu, and L. Xiaoqiang, "Correlation analysis of the physical-mechanical properties of clayey soil in Binhai new area in Tianjin," *Chinese Journal of Underground Space and Engineering*, vol. 15, no. 2, pp. 452–457, 2019.
- [14] Z. Xianwei, K. Lingwei, G. Aiguo, and T. Yongfei, "Physical and mechanical indexes and correlation analysis of clay with noticeable structures in Zhanjiang area," *Journal of Engineering Geology*, vol. 19, no. 4, pp. 447–454, 2011.
- [15] J. Ching and K.-K. Phoon, "Correlations among some clay parameters—the multivariate distribution," *Canadian Geotechnical Journal*, vol. 51, no. 6, 2014.



- [16] D. Arditi and T. Pulket, "Predicting the outcome of construction litigation using an integrated artificial intelligence model," *Journal of Computing in Civil Engineering*, vol. 24, no. 1, 2010.
- [17] J. Chen, M. Vissinga, Y. Shen, S. Hu, E. Beal, and J. Newlin, "Machine learning-based digital integration of geotechnical and ultrahigh-frequency geophysical data for offshore site characterizations," *Journal of Geotechnical and Geoenvironmental Engineering*, vol. 147, no. 12, 2021.
- [18] D.-M. Zhang, J.-Z. Zhang, H.-W. Huang, C.-C. Qi, and C.-Y. Chang, "Machine learning-based prediction of soil compression modulus with application of 1D settlement," *Journal of Zhejiang University-SCIENCE A*, vol. 21, pp. 430–444, 2020.
- [19] B. T. Pham, L. H. Son, T.-A. Hoang, D.-M. Nguyen, and D. Tien Bui, "Prediction of shear strength of soft soil using machine learning methods," *CATENA*, vol. 166, pp. 181–191, 2018.
- [20] W. Z. Taffese and K. A. Abegaz, "Prediction of compaction and strength properties of amended soil using machine learning," *Buildings*, vol. 12, no. 5, p. 613, 2022.
- [21] Y. Li, H. Rahardjo, A. Satyanaga, S. Rangarajan, and D. T.-T. Lee, "Soil database development with the application of machine learning methods in soil properties prediction," *Engineering Geology*, vol. 306, Article ID 106769, 2022.
- [22] P. He, W. D. Wang, and Z. H. Xu, "Empirical relationship between compression index and rebound index of Shanghai clay," *Geotechnical Mechanics*, vol. 39, no. 10, pp. 3773–3782, 2018.
- [23] D. J. Kong and W. H. Sun, "Statistics analysis on friction angle of shear strength for clay," *Port & Waterway Engineering*, vol. 2, pp. 10–13, 2000.
- [24] B. M. Das, *Advanced Soil Mechanics*, CRC press, London, 2019.
- [25] L. Yi, "Investigation on the swelling properties and microstructure mechanism of compacted Gaomiaozhi bentonite," *Journal of Engineering Geology*, vol. 24, no. 3, pp. 451–458, 2016.
- [26] M. Jing, G. Yanbin, and Y. Junyun, "Study on the effective angle of internal friction of remolded clay in Shanghai," *Journal of Engineering Geology*, vol. 16, no. S1, pp. 677–680, 2008.
- [27] C.-J. Wu, G.-L. Ye, and J. H. Wang, "Relationship between compression index and natural water content of Shanghai clay," *Rock and Soil Mechanics*, vol. 35, no. 11, pp. 3184–3190, 2014.
- [28] W. T. Liu, Y. H. Shi, and H. F. Zeng, "Assessment of geotechnical characteristics of clay in the Houshui Bay HaiNan," *Natural Science Journal of HaiNan University*, vol. 31, no. 4, pp. 317–322, 2013.
- [29] L. N. Wu, Q. Q. Xie, and Y. Y. Fang, "Correlation analysis of the physical mechanical indexes of cohesive soil in Hefei area," *Hydrogeology & Engineering Geology*, vol. 29, pp. 43–45, 2002.
- [30] T. Y. Pan, "Physical mechanics and engineering characteristic index analysis of soil," *Journal of Water Conservancy and Construction Engineering*, vol. 9, no. 5, pp. 87–91, 2011.
- [31] X. Du, Y. Sun, Y. Song, and C. Zhu, "In-situ observation of wave-induced pore water pressure in seabed silt in the yellow river estuary of China," *Journal of Marine Environmental Engineering*, vol. 10, no. 4, pp. 305–317, 2021.