Multivariate distribution model of compressibility and CPTU indices of Jiangsu inorganic soft clays

Haifeng Zou[1,2], G. Cai[1], S. Liu[1], and A.J. Puppala[3]

[1] Institute of Geotechnical Engineering, Southeast University, Nanjing, 211189, China.
[2] Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong
[3] Department of Civil Engineering, The University of Texas at Arlington, Texas, USA

ABSTRACT

This study developed a multivariate distribution model for the compressibility and piezocone penetration test (CPTU) indices of Jiangsu inorganic soft clays based on a compiled database. Analytical expressions for the marginal probability density distributions of soil variables and correlation coefficients were presented. Based on the constructed model, bivariate and multivariate correlations for predicting the primary and secondary compression indices were proposed. Performance of the developed correlations was evaluated in detailed. The results indicated that the constructed multivariate distribution model provided a favorable way to capture the multivariate dependencies among the compressibility and CPTU indices of Jiangsu clays. Nevertheless, caution shall be exercised when the model and corresponding correlations are used in other soils such as silts and sands.

Keywords: multivariate distribution; clay database; compressibility; piezocone penetration test

1 INTRODUCTION

Soil compressibility has been a major concern for settlement-related issues in geotechnical engineering. In regular site characterization projects, plenty of soil samples have to be collected and thus design costs are significantly increased. To achieve a more efficient and economic evaluation of soil compression behavior, the piezocone penetration testing (CPTU) technique has been gradually used in recent studies (e.g., Bersan et al. 2013; Reid 2015; Tonni et al. 2016). Past research indicated that the three CPTU measurements including cone tip resistance (q_t), sleeve frictional resistance (f_s), and pore water pressure (u_2) are all a function of soil stiffness, fines contents and soil fabrics. These factors also impact the behavior of soil compression, and therefore it is not unreasonable to establish correlations between soil compressibility and CPTU indices.

Despite recent active studies, it seems that most research only focused on bivariate correlations among these soil parameters, such as predicting the slope of the critical state line in e-lgp' space (λ_{10}) using the normalized frictional resistance (F_r) or soil behavior type index (I_c) (e.g., Reid 2015), or predicting the secondary compressibility index (C_α) using the normalized cone tip resistance (Q_t) (e.g., Bersan et al. 2013; Tonni et al. 2016).

This study applied the multivariate distribution model approach to capture the correlations among two compressibility parameters, i.e., primary and secondary compressibility indices (C_c and C_α), and three CPTU indices, including Q_t, F_r and I_c, for Jiangsu inorganic soft clays based on a compiled database. Furthermore, bivariate and multivariate correlations for predicting C_c and C_α using different CPTU indices are derived using the constructed multivariate distribution model, and their performances are examined in detail. Implications and cautions in applying the developed correlations are also discussed. It shall be mentioned that the I_c used in this study follows the definition given by Been and Jefferies (1992) as below, because the involved excess pore water pressure item (B_q) shall be useful to predict C_c and C_α (e.g., Tonni et al. 2016)

$$I_c = \sqrt{\left\{3 - \lg\left[Q_t\left(1 - B_q\right) + 1\right]\right\}^2 + \left(1.5 + 1.3\lg F_r\right)^2} \qquad (1)$$

2 DATABASE

The compiled database only involves data points for Jiangsu inorganic soft clays. Data points corresponding to organic soils were discarded as they show quite different behavior compared to other soils, e.g., the compressibility of peat is generally more significant than clays. All C_c and C_α data were obtained from one dimensional oedometer tests. In some experimental tests, the e-lgp' curves did not approximate a straight line, and thus it is difficult to determine C_c and C_α. A more accurate way is to use the (1+e)-lgp' curves as an alternative (Hong et al. 2012). Then $C_c = C_{cL}(1+e_c)$, where e_c is void ratio at the end of primary consolidation at current effective stress, and C_{cL} is the slope of (1+e)-lgp' curves.

Using the above methods, 72 sets of {Q_t, F_r, I_c, C_c, C_α} were obtained for the Jiangsu inorganic soft clays. The C_c and C_α in the database vary within the ranges of

0.203 – 0.779 and 0.0078 – 0.0287, respectively. These two ranges correspond to medium to very high compressibility according to Mesri (1973). The whole database will be used in the following sections to establish a multivariate distribution model.

3 MARGINAL DISTRIBUTIONS

The first step to develop a multivariate distribution model is to quantify the marginal probability density functions (PDFs) of involved soil parameters. For the convenience of presentation, the five soil parameters are denoted as $Y_1 = Q_t$, $Y_2 = F_r$, $Y_3 = I_c$, $Y_4 = C_c$, and $Y_5 = C_\alpha$. Liu et al. (2016) applied the Box-Cox method to convert a non-Gaussian soil variable (Y) to a standard normal variable (X) as follows:

$$X = \frac{Y^{\lambda} - 1 - \lambda a}{\lambda b} \tag{2}$$

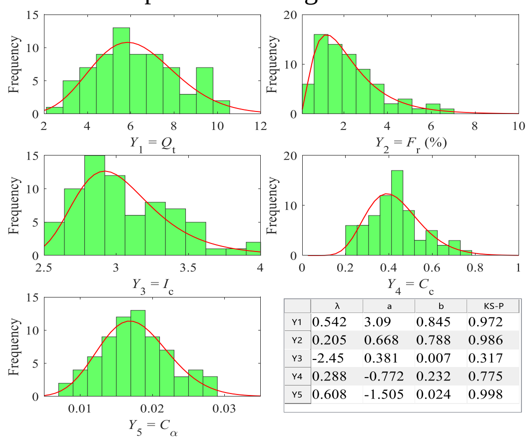
where λ is a transformation power, a and b are scaling and shifting parameters.

The above formula is suitable for $\lambda \neq 0$. For $\lambda = 0$, a natural logarithmic transformation is advised. The estimation of optimal λ value can be readily achieved using Matlab library function *boxcox*. Liu et al. (2016) did not present the PDF of Y. In this study, the following approximation method is proposed to address this issue (Cramer 1999),

$$g(Y) = f(X)\frac{dX}{dY} = \frac{1}{b}Y^{\lambda-1}f(X) \tag{3}$$

where $g(Y)$ and $f(X)$ are PDFs of Y and X, respectively.

The above approximation holds on the basis of the simple idea that the integral of $g(Y)$ over dY always equals to that of $f(X)$ over dX. Using this method, the marginal PDFs of Y variables are estimated and they are compared with sample histograms in Fig. 1. The three transformation parameters (λ, a, b) and P-values of the Kolmogorov-Smirnov test (KS-P) for the transformed X variables are also presented in Fig. 1.



	λ	a	b	KS-P
Y1	0.542	3.09	0.845	0.972
Y2	0.205	0.668	0.788	0.986
Y3	-2.45	0.381	0.007	0.317
Y4	0.288	-0.772	0.232	0.775
Y5	0.608	-1.505	0.024	0.998

Fig. 1. Marginal PDFs of soil variables

Fig. 1 indicates that the estimated PDFs agree well with the sample histograms. The KS-P values are larger than 0.05, indicating that there is no strong evidence to reject the hypothesis that the X variables individually follow a standard normal distribution. Thus the above transformation is effective and the estimated PDFs of Y variables are rational for the investigated database.

4 CORRELATION MATRIX

The next step is to calculate the correlation matrix of X variables, which is formed by the Pearson's correlation coefficients (δ_{ij}) between all pairwise X_i-X_j data. It is considered useful to quantify the uncertainties within δ_{ij} values, especially when sample size is limited. Zou et al. (2017) applied the Fisher's Z transformation to indirectly address this issue as follows,

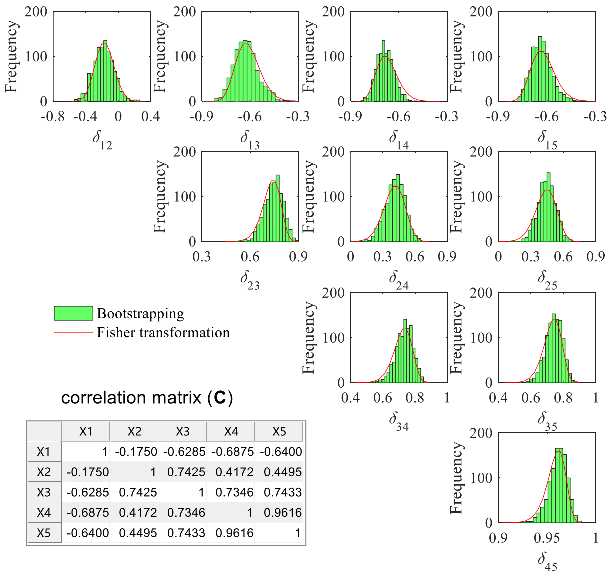
$$z_{ij} = \frac{1}{2}\ln\left(\frac{1+\delta_{ij}}{1-\delta_{ij}}\right), \quad \sigma_{zij} = \frac{1}{\sqrt{N_{ij}-3}} \tag{4}$$

where z_{ij} is the transformed correlation coefficient, σ_{ij} is the standard deviation of z_{ij}, and N_{ij} is the sample size of X_i-X_j data.

It is argued that z_{ij} approximately follows a normal distribution and therefore its PDF is traceable (Cramer 1999). Based on the PDF transformation technique in Equation (3), it is also possible to provide the analytical expression for the PDF of δ_{ij}, as follows,

$$g(\delta_{ij}) = f(z_{ij})\frac{dz_{ij}}{d\delta_{ij}} = \frac{1}{1-\delta_{ij}^{2}}f(z_{ij}) \tag{5}$$

Using Equation (5), the PDFs of δ_{ij} are estimated, and they are compared with the histograms obtained from the bootstrapping techniques suggested by Ching et al. (2014) in Fig. 2. It is indicated that the estimated PDFs match the histograms favorably. Therefore, Equation (5) provides a rational way to report the PDFs of δ_{ij} values.



	X1	X2	X3	X4	X5
X1	1	-0.1750	-0.6285	-0.6875	-0.6400
X2	-0.1750	1	0.7425	0.4172	0.4495
X3	-0.6285	0.7425	1	0.7346	0.7433
X4	-0.6875	0.4172	0.7346	1	0.9616
X5	-0.6400	0.4495	0.7433	0.9616	1

Fig. 2. Estimated PDFs of δ_{ij} for pairwise variables

The estimated median values of δ_{ij} are also tabulated in Fig. 2. A strong correlation between C_c and C_α is observed as δ_{45} reaches 0.96. This is reasonable because the ratio of C_α/C_c generally remains constant for common soils according to Mesri (1973). It is also found that I_c is the most effective CPTU index to predict C_c, followed by Q_t and F_r. This observation is consistent with the previous study conducted by Reid (2015). Therefore, the correlation matrix presented in Fig. 2 shall be reasonable in the context of existing geotechnical knowledge.

5 DERIVED CORRELATIONS

The multivariate distribution model constructed in the previous two sections provides a prior joint PDF for the involved soil parameters (Ching et al. 2014; Liu et al. 2016). It is possible to establish correlations among C_c, C_α and CPTU parameters for the Jiangsu inorganic clays based on the constructed model. This is achieved using a Bayesian updating and a back transformation process. The detailed derivation procedure is available in the literature (e.g., Liu et al. 2016; Zou et al. 2017) and thus it is not presented here. The following four bivariate correlations for Jiangsu inorganic soft clays are derived and compared with data and trends from literature:

- C_α-C_c correlation as shown in Fig. 3(a). The C_α/C_c concept proposed by Mesri (1973) is also illustrated in Fig. 3(a);
- C_α-Q_t correlation as shown in Fig. 3(b). The empirical trend recommended by Bersan et al. (2013) and Tonni et al. (2016) for the Venice Lagoon sands and silts is also given in Fig. 3(b);
- C_c-F_r correlation as shown in Fig. 3(c). The data points of λ_{10} from 31 worldwide sites reported by Reid (2015) are also illustrated in Fig. 3(c). Albeit λ_{10} is different from C_c, i.e., λ_{10} is obtained from isotropic consolidation test whereas C_c is from one-dimensional oedometer test, λ_{10} data are still used here to achieve a qualitative comparison because they share same definition and evaluate similar soil behavior;
- C_c-F_r correlation as shown in Fig. 3(d). Again, the data reported by Reid (2015) are introduced in Fig. 3(d) for a qualitative comparison.

The following conclusions are obtained from Fig. 3:

- The proposed C_α-C_c correlation agrees well with the C_α/C_c concept and the C_α/C_c of Jiangsu inorganic soft clays varying within a narrow range of 0.033 to 0.046.
- The proposed C_α-Q_t correlation agrees with that given by Bersan et al. (2013) only in trend. The empirical correlation suggested by Bersan et al. (2013) systematically underestimates C_α values for the Jiangsu inorganic soft clays. A possible explanation is that their correlation was developed based on sands and silts, which show less notable compressibility than the soft clays in this study.

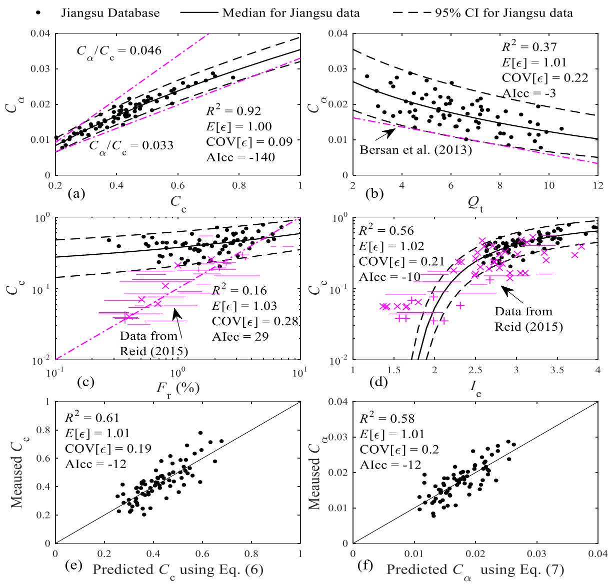


Fig. 3. Developed correlations for predicting C_c and C_α

- The proposed C_c-F_r correlation overestimates the compressibility of soils reported by Reid (2015) significantly. Except for the difference between λ_{10} and C_c, this overestimation is more likely to be attributed to the fact that their data mainly contain tailings, sands, and silty clays, whereas soft clay data are quite limited.
- The proposed C_c-I_c correlation agrees with the data reported by Reid (2015) in trend; nevertheless, the latter are much more scattered than expected.

The above results indicate that existing CPTU-based empirical correlations and data may significantly underestimate the compressibility of Jiangsu inorganic soft clays, perhaps due to the impact of biased sample in soil types.

By comparing the developed bivariate correlations with experimental data, it is found in Fig. 3 that the correlations agree well with the Jiangsu database in both trend and range. Therefore, it seems that constructed multivariate distribution model can do justice to the Jiangsu clay. Based on the constructed model, the following two correlations are derived for predicting C_c and C_α using all three CPTU indices:

$$C_c = \left(-0.05Q_t^{0.542} - 0.026F_r^{0.205} - 2.307I_c^{-2.45} + 1.096\right)^{3.477} \quad (6)$$

$$C_\alpha = \left(-0.008Q_t^{0.542} - 0.006F_r^{0.205} - 0.553I_c^{-2.45} + 0.15\right)^{1.646} \quad (7)$$

Comparisons between measured and predicted C_c and C_α values using the above multivariate correlations are shown in Fig. 3(e) and 3(f), respectively.

To assess the performance of the above proposed correlations, the leave-one-out cross-validation method is used. In this method, each set of sample data is omitted from the database to construct a new multivariate distribution model, and the omitted Y variables are then predicted using the new model. The differences between all omitted and predicted Y values are assumed to approximate the error of the model developed using whole database. This method has demonstrated its effectiveness in Ching and Wu (2017).

The following four performance metrics are selected to achieve a quantitative evaluation and comparison:

(1) R^2, to evaluate the accuracy of prediction in a normalized scale;
(2) $E[\varepsilon]$, mean value of prediction error (ε), to evaluate the systematic bias of prediction error, which is defined as ε = measurement/prediction;
(3) COV[ε], coefficient of variation of ε, which indicates scatter between predicted and measured values. Note that $E[\varepsilon]$ and COV[ε] essentially describe the PDF of ε, which approximately follows lognormal distribution in this study;
(4) AICc, corrected Akaike information criterion, to consider the model complexity due to introducing more model parameters in multivariate analysis (Burnham and Anderson 2002):

$$\mathrm{AICc} = 2n_p - 2\ln(L) + \frac{2n_p^2 + 2n_p}{n_y - n_p - 1} \quad (8)$$

where n_p is number of model parameters in correlations, n_y = 72 is sample size, and L is the maximum value of log-likelihood function for the model. By assuming that ε follows lognormal distribution, ln(L) is calculated by

$$\ln(L) = -\frac{n_y}{2}\ln(2\pi) - \frac{n_y}{2}\ln\left(\frac{1}{n_y}\sum_{i=1}^{n_y}\varepsilon_i^2\right) - \frac{n_y}{2} - \sum_{i=1}^{n_y}\ln\varepsilon_i \quad (9)$$

Eqs. (8) and (9) imply that an increase in prediction error (ε) and number of model parameters (n_p) will lead to the increase of AICc. Therefore, the correlation with lowest AICc value is the optimal one with highest accuracy and least model parameters.

The calculated values of the four performance metrics are also shown in Fig. 3. The following extra conclusions are drawn from Fig. 3:

- All correlations provide unbiased estimates of C_c and C_α because $E[\varepsilon]$ is close to 1;
- The uncertainties within CPTU-based correlations are notable because their R^2 values are low and COV[ε] values are high;
- Multivariate correlations are superior to bivariate correlations because the R^2 values are higher and AICc values are lower for the former. That is to say, the increase of prediction accuracy is more significant than the increase of model complexity.

Based on the above analysis, it is recommended to use Eqs. (6) and (7) to predict the compressibility of Jiangsu inorganic soft clays from CPTU data. However, as discussed previously, caution shall be exercised when these correlations are extended to other soils. Most probably, these correlations may overestimate C_c and C_α values of sands and silts significantly, according to some data reported in the literature.

6 CONCLUSION

This study developed a multivariate distribution model for two compressibility indices (C_c and C_α) and three CPTU indices (Q_t, F_r, I_c) for Jiangsu inorganic soft clays. The following conclusions are obtained:

(1) Among the three CPTU indices, I_c is the most effective parameter for predicting C_c and C_α, followed by Q_t and F_r.
(2) Notable uncertainty in the estimation C_c and C_α of using CPTU indices is observed.
(3) Multivariate correlations are superior to bivariate correlations as the prediction error can be reduced.

ACKNOWLEDGEMENTS

Majority of the work presented in this paper was funded by the National Key R&D Program of China (2016YFC0800200), and the National Natural Science Foundation of China (Grant No. 41672294). These financial supports are gratefully acknowledged.

REFERENCES

Been, K., and Jefferies, M. G. (1992). Towards Systematic CPT Interpretation. Proceedings of Wroth Memorial Symposium, Thomas Telford, London, 1992: 121 - 134.

Bersan, S., Cola, S., and Simonini, P. (2013). Secondary compression of Venice Lagoon sands and silts from CPTU. Geotechnical and Geophysical Site Characterization 4, ISC'4, 1, 383-389.

Burnham, K.P., and Anderson, D.R. 2002. Model Selection and Multimodel Inference: A practical information-theoretic approach (2nd ed.), Springer-Verlag.

Ching, J.Y., Phoon, K.K., and Chen, C.H. (2014). Modeling piezocone cone penetration (CPTU) parameters of clays as a multivariate normal distribution. Canadian Geotechnical Journal 51(1), 77-91.

Ching, J.Y., and Wu, Tsai-Jung. (2017). Probabilistic transformation models for preconsolidation stress based on clay index properties. Engineering Geology, 226, 33-43.

Cramer, H. (1999). Mathematical Methods of Statistics. Princeton University Press, Princeton, NJ.

Hong, Z., Zeng, L., Cui, Y., Cai, Y., and Lin, C. (2012). Compression behaviour of natural and reconstituted clays. Geotechnique, 62(4), 291-301.

Liu, S., Zou, H., Cai, G., Bheemasetti, T., Puppala, A.J., and Lin, J. (2016). Multivariate Correlation among Resilient Modulus and Cone Penetration Test Parameters of Cohesive Subgrade Soils, Engineering Geology, 209, 128-142.

Mesri, G. (1973). Coefficient of Secondary Compression, Journal of the Soil Mechanics and Foundations Division, 99(1), 123-137.

Reid, D. (2015). Estimating slope of critical state line from cone penetration test – an update. Canadian Geotechnical Journal, 52, 46-57.

Tonni, L., Martínez, M.F.G., Simonini, P., and Gottardi, G. (2016). Piezocone-based prediction of secondary compression settlements of coastal defence structures on natural silt mixtures. Ocean Engineering, 116, 101-116.

Zou, H., Liu, S., Cai, G., Puppala, A. J., and Bheemasetti, T. (2017). Multivariate Correlation Analysis of Seismic Piezocone Penetration (SCPTU) Parameters and Design Properties of Jiangsu Quaternary Cohesive Soils. Engineering Geology, 228, 11-38.