Engineering Geology 273 (2020) 105675

Contents lists available at ScienceDirect

Engineering Geology

journal homepage: www.elsevier.com/locate/enggeo

Multivariate probability distribution of Shanghai clay properties

Dongming Zhang[a,*], Yelu Zhou[a], Kok-Kwang Phoon[b], Hongwei Huang[a]

[a] Key Laboratory of Geotechnical and Underground Engineering of Minister of Education and Department of Geotechnical Engineering, Tongji University, Shanghai, China
[b] Dept. of Civil and Environmental Engineering, National Univ. of Singapore, 117576, Singapore

ARTICLE INFO

Keywords:
Shanghai clay
Municipal database
Parameter correlation
Multivariate distribution

ABSTRACT

In this study, a database is compiled for 11 clay parameters covering 50 sites in Shanghai with 4051 data points (labeled as SH-CLAY/11/4051). These sites are distributed in the eight districts of Shanghai covering an area of 145 km^2. This is the first sizeable multivariate soil database compiled at the municipal level. General information, including the geology of Shanghai, the borehole locations and depths, the types of parameters, and the basic statistics associated with these parameters are presented first. The quality of the compiled data is then assessed by comparing them with the data in the published global database CLAY/10/7490 with respect to marginal statistics. Results show that data points in SH-CLAY/11/4051 fall within the range of CLAY/10/7490. It is not surprising that local Shanghai data are more clustered than those in CLAY/10/7490. The range of values spanned by each clay parameter is smaller at a regional/municipal scale than at the global scale. A preliminary study to understand what is "unique" in Shanghai clay data is conducted by comparing global correlations between any two parameters with the corresponding Shanghai versions. Some correlations found in SH-CLAY/11/4051 (e.g., normalized effective vertical stress and normalized undrained shear strength) are observed to be similar to those found in CLAY/10/7490. However, there are other distinct correlations produced by SH-CLAY/11/4051, indicating a regional/municipal effect. This study constructs a multivariate probability distribution based on SH-CLAY/11/4051 to capture these distinct correlations between Shanghai clay parameters. One useful application of this multivariate distribution is act as prior for Bayesian updating.

1. Introduction

One of the critical decisions in geotechnical engineering is to interpret design parameters from site investigation reports (Forster and Culshaw, 1990; Lashkaripour and Ajalloeian, 2003; Cai et al., 2010; Khan et al., 2017). Annex D of ISO2394 (2015) highlighted that "Although the principles of reliability-based design are applicable to geotechnical engineering in general, it is essential to cater for the higher variability depicted by geotechnical design parameters and other site conditions." However, there are often insufficient site-specific data to determine design parameters at the desired project location. Generic (or global) transformation models (Phoon and Kulhawy, 1999a, 1999b) are typically adopted to transform more abundant field measurements to the design parameters needed for geotechnical design. It is widely recognized in practice that local transformation models are more relevant to specific sites. Thus, some local models have been established for specific soils types (Akca, 2003; Long et al., 2010; Lee et al., 2011; Cai et al., 2014). However, the engineer is compelled to use generic models in most cases because these local models are commonly unavailable in routine projects. The main objective of this study is to fill this gap for Shanghai's soft clay. The statistical methodology adopted in this paper follows standard practice (Ching et al., 2016), that require the: (1) compilation of a multivariate database consisting of field and laboratory data and (2) construction of a multivariate probability distribution based on this database. This methodology does not require a genuine multivariate database where all the parameters are measured simultaneously in close proximity, in contrast to conventional regression (Baziar and Jafarian, 2007; Das and Basudhar, 2008; Khanlari et al., 2012; Zhang et al., 2015). Incomplete multivariate databases such as those consisting of multiple sets of bivariate data collected from different locations can be used to construct a multivariate probability distribution.

The multivariate probability distribution presented in this study is valuable, because it is specific to a part of Shanghai. It is reasonable to say that this "local" multivariate probability is more relevant to Shanghai at large than the generic version constructed based on CLAY/10/7490 (Ching and Phoon, 2014a) that encompasses sites from all around the world. It is useful to note that this "site challenge" is much broader than the construction of a local multivariate probability distribution, although this is a critical component. Phoon (2018) described

* Corresponding author.
E-mail addresses: 09zhang@tongji.edu.cn (D. Zhang), yl.zhou@tongji.edu.cn (Y. Zhou), kkphoon@nus.edu.sg (K.-K. Phoon), huanghw@tongji.edu.cn (H. Huang).

https://doi.org/10.1016/j.enggeo.2020.105675
Received 21 August 2019; Received in revised form 8 May 2020; Accepted 9 May 2020
Available online 13 May 2020
0013-7952/

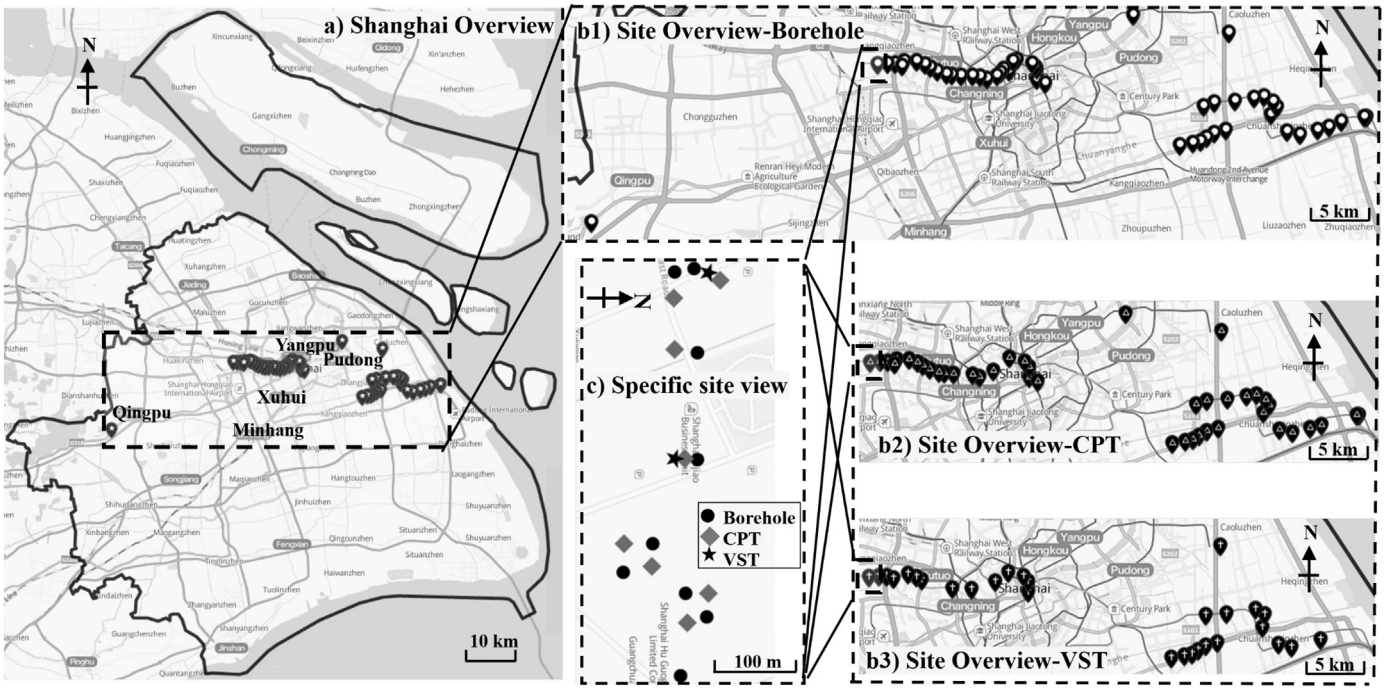


Fig. 1. Locations of the 50 sites in SH-CLAY/11/4051.

the challenge in the context of addressing Multivariate, Uncertain and Unique, Sparse, and InComplete (*MUSIC*) data. Ching and Phoon (2019) demonstrated a rigorous method to construct a multivariate probability distribution from *MUSIC* data. Phoon et al. (2019) suggested that *MUSIC* can be re-interpreted to cover outliers. Ching and Phoon (2020a) subsequently extended *MUSIC* to *MUSIC-X* to account for spatial correlations between two records measured at two close depths. The symbol "*X*" was adopted to emphasize the spatial/temporal dimension in *MUSIC* data. It is noted that this challenge has not been fully addressed, and constitutes a topic of active current research.

It is well known that soil parameters are "related" to each other in our current deterministic practice. The existence of a large number of generic transformation models (Kulhawy and Mayne, 1990) demonstrates the usefulness of bivariate relationships between soil parameters. The concept of a correlation merely expands the deterministic notion of a relationship, such as a mean trend to a more complete probabilistic notion that describes the strength of the relationship on top of the mean trend. This strength is related to the transformation uncertainty about the mean trend. The uncertainty is low when two parameters are strongly correlated. Multivariate correlations extend this probabilistic concept to more than two parameters. This extension is necessary, because it is common to measure more than two soil parameters in close proximity in a site investigation. For example, when undisturbed samples are extracted for an unconfined compression soil test to measure the undrained shear strength, disturbed soil samples nearby are commonly extracted at the same time for the measurement of index properties, such as the liquid limit (LL). Moreover, cone penetration test (CPT) soundings may carry out to provide information on stratification and a more continuous profile of sleeve and tip resistances. The correlations between these different soil parameters produced by laboratory and field tests can thus be determined. These multivariate correlations are pivotal to the construction of a multivariate probability distribution.

Some multivariate probability distributions have been constructed based on different regional or global databases (e.g., Ching and Phoon, 2012, 2014a; D'Ignazio et al., 2016; Liu et al., 2016; Zou et al., 2017). However, the properties of Shanghai's clay are largely absent from these published databases. The purpose of this paper is to compile some properties of Shanghai's clay in a municipal database labeled as SH-CLAY/11/4051. This database contains information from 50 site investigations carried out over a land area of 145 km^2 in Shanghai. A multivariate probability distribution characterizing the statistical relationships between 11 clay parameters was constructed based on this database. To the authors' knowledge, this is the first sizeable multivariate soil database compiled at the municipal level in the literature. This Shanghai database is valuable for both research and practice. One of the major challenge in addressing *MUSIC* data is how to characterize the "uniqueness" of each site quantitatively based on data alone (Phoon, 2020). Building regulations all over the world legally mandate a site investigation at each project site in recognition of this unique or site-specific feature of geotechnical engineering. At present, engineers rely purely on experience to judge which sites are comparable and which sites are not. SH-CLAY/11/4051 presents the most challenging database for research addressing the "*U*" in *MUSIC* data in contrast to global or regional databases, because all sites are broadly similar due to their common geologic origin. In addition, SH-CLAY/11/4051 is a valuable contribution to Shanghai geotechnical engineering practice, because the estimation of design parameters is less uncertain if SH-CLAY/11/4051 and site-specific data are considered together, rather than using highly limited site-specific data alone (Ching et al., 2020).

The rest of this study is organized as follows. Firstly, some background information for SH-CLAY/11/4051 covering the location, test condition, and type of clay parameters are presented. Secondly, the database SH-CLAY/11/4051 is compared with the global database CLAY/10/7490 in terms of marginal statistics and correlations for parameters common to SH-CLAY/11/4051 and CLAY/10/7490. A multivariate probability distribution is then constructed based on SH-CLAY/11/4501 to capture the multivariate correlations between the clay parameters for Shanghai's soft clay. Finally, the practical usefulness of this probability distribution in reducing the uncertainty in the design parameter estimate is demonstrated.

2. SH-CLAY/11/4051 database

Shanghai is located in Eastern China at the mouth of the Yangtze River. It sits on land formed from the deposition of sand, silt, and clay

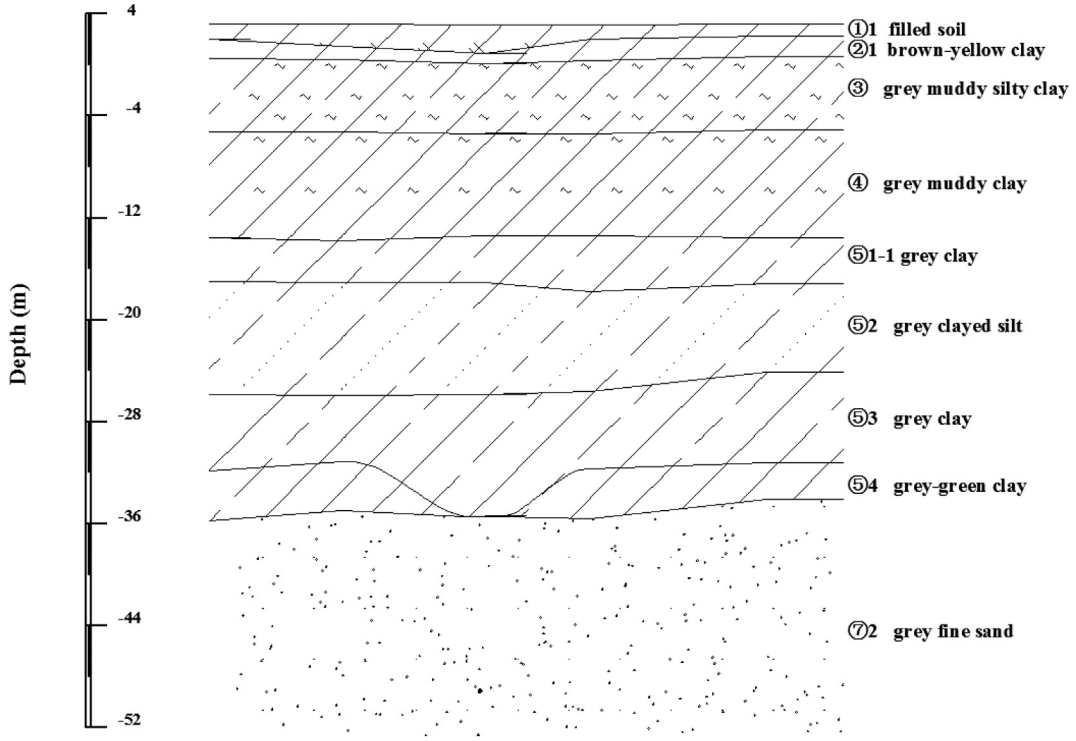


Fig. 2. Typical geological profiles of Shanghai.

that were transported by the Yangtze River.

The data points presented in this study are collected from 50 site investigation reports for Shanghai. These sites cover an approximate area of 145 km^2 in the Pudong, Yangpu, Huangpu, Jing'an, Putuo, Changning, Jiading, and the Qingpu Districts. For comparison, Shanghai occupies around 6341 km^2. The majority of these sites are located to the north-east of the Shanghai Hongqiao International Airport and to the north-west of the Pudong International Airport. In total, there are 553 boreholes, 47 vane shear tests, and 190 cone penetration tests. Fig. 1(a) presents an overview of Shanghai and the locations of the 50 sites. Fig. 1(b1), 1(b2), and 1(b3) shows the distribution of boreholes, vane shear tests, and cone penetration tests on an enlarged scale, respectively. Fig. 1(c) is a typical example showing the distribution of the three types of tests at a specific site in Jiading District.

The Shanghai local code for geotechnical site investigation defines the different geological layers and sublayers constituting Shanghai clay. Fig. 2 shows a typical geological profile of Shanghai that illustrates this stratification. As shown, Shanghai's soft clay mainly consists of grey muddy silty clay, grey muddy clay, grey clay, and grey silty clay. This study compiles a database for Shanghai's soft clay from the ground surface to a maximum depth of 30 m given that many tunnels are constructed within this depth range (Zhang et al., 2015). The histogram of the depth distribution of the boreholes is presented in Fig. 3. Note that the depth of boreholes presented herein pertains to the location of the clay layer whose properties are compiled in this Shanghai database.

In this database, there are eleven parameters: 1) four index parameters, including the liquid limit (LL), plasticity index (PI), liquidity index (LI), and void ratio (e), and 2) seven mechanical parameters, including the at-rest lateral pressure coefficient (K_0), vertical effective stress (σ'_v), undrained shear strength ($S_{u(UCST)}$), the sensitivity ($S_{t(UCST)}$), obtained from the unconfined compression soil test (UCST), undrained shear strength ($S_{u(VST)}$) and sensitivity ($S_{t(VST)}$) obtained from the vane shear test (VST), and specific penetration resistance (p_s) from the CPT.

These 11 parameters were obtained from different types of tests: LL, PI, LI, e, K_0, $S_{u(UCST)}$, and $S_{t(UCST)}$, are produced by laboratory tests on

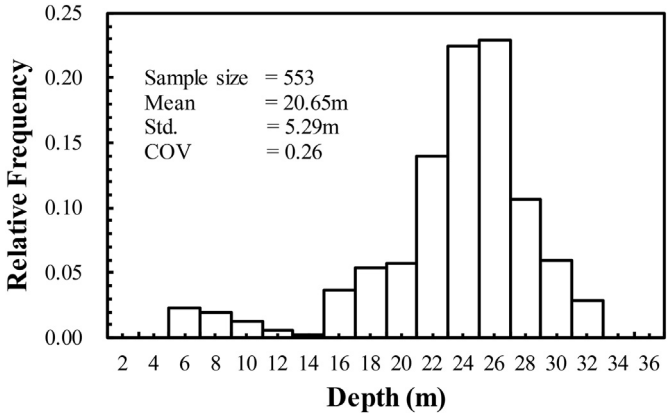


Fig. 3. Depth distribution of the boreholes in SH-CLAY/11/4051.

disturbed and undisturbed clay samples extracted from the boreholes, while $S_{u(VST)}$ and $S_{t(VST)}$ are produced by the VST, and p_s is produced by the CPT. It should be noted that the pore pressure is not measured for CPT tests conducted in China (Wang, 1978) and specific penetration resistance (p_s) is not comparable to the corrected cone tip resistance (q_t) adopted in international practice (Ching et al., 2014). The parameter p_s includes the sleeve friction and tip resistance of the probe, measured with a single bridge strain gauge (Wang, 1978). The effective vertical stress σ'_v is estimated by multiplying the depth of the soil sample by the effective unit weight. Different boreholes within a 35-m radial area are grouped together, and the values of the corresponding parameters at the same depth constitute a multivariate data "point". The reason for using a radial diameter of 35 m for grouping is that soil parameters are strongly correlated within the original correlation distance and the average value for the original correlation distance in Shanghai clay is 35 m (Xiao, 2018; Huang et al., 2017). SH-CLAY/11/4051 contains 4051 data points or records. The entire database thus occupies 4051 rows and 11 columns in EXCEL. It is common to find empty cells in this database, because some tests are not conducted at a particular depth. In

particular, information from boreholes cannot be readily paired with VST because the VST is less frequently conducted in the proximity of the boreholes. The number of empty cells is 26,976. The percentage of incompleteness can be indexed by the ratio, number of empty cells/(4051 × 11) = 60.5%.

The global database CLAY/10/7490 published by Ching and Phoon (2014a) was labeled in accordance with the format of (soil type)/(number of parameters of interest)/(number of data points). Following the same labeling format, the current Shanghai database is thus labeled as SH-CLAY/11/4051. For comparison, the percentage of incompleteness for the global database CLAY/10/7490 is 65.9%. Table A1 presents the general characteristics of the database SH-CLAY/11/4051, including the site number, borehole numbers, sample size, and the upper and lower bounds of the 11 parameters as well as the depth.

Eleven dimensionless parameters are evaluated to facilitate comparison with the global database CLAY/10/7490. They are denoted as LL, PI, LI, e, K_0, σ'_v/P_a (P_a is one atmosphere pressure, i.e., 101.3 kPa), $S_{u(UCST)}/\sigma'_v$, $S_{u(VST)}/\sigma'_v$, $S_{t(UCST)}$, $S_{t(VST)}$, and p_s/σ'_v. Note that the normalized specific penetration resistance ratio p_s/σ'_v is different from the more familiar normalized cone tip resistance $(q_t–\sigma_v)/\sigma'_v$ that is extensively used in the United States of America (USA) and Europe. The corrected cone tip resistance q_t is equal to $q_c + (1–a)\,u_2$, in which q_c is the measured cone tip resistance, a is the area ratio of the cone and u_2 is the pore pressure behind the cone. Table 1 lists the basic statistics of the eleven dimensionless parameters, including the sample size, mean value, coefficient of variation (COV), maximum value (Max), and minimum value (Min). For comparison, the statistics of the corresponding parameters in CLAY/10/7490 are listed in parentheses. Referring to Report EL-6800 (Kulhawy and Mayne, 1990), Shanghai clay is classified as a very soft clay (LI = 0.49–2.19) with slight to medium plasticity (PI = 10.4–26.5), and with medium to high sensitivity (S_t = 2.7–7.8).

3. Comparison between SH-CLAY/11/4051 and CLAY/10/7490

CLAY/10/7490 is the largest published global database for clay parameters compiled thus far (Ching and Phoon, 2014a). It covers sites in 30 countries or regions around the world reported in 251 studies. There are 10 dimensionless clay parameters in the database: LL, PI, LI, normalized vertical effective stress (σ'_v/P_a), normalized preconsolidation stress (σ'_p/P_a), normalized undrained shear strength (S_u/σ'_v), sensitivity (S_t), and three piezocone test (CPTU) parameters, namely, normalized cone tip resistance ($(q_t–\sigma_v)/\sigma'_v$), effective cone tip resistance ($(q_t–u_2)/\sigma'_v$), and pore pressure ratio (B_q). The S_u values in the literature were obtained from various types of tests. These S_u values were all converted to S_u(mob) based on the use of the transformation models proposed by Ching and Phoon (2014a, 2014b). The parameter S_u(mob) is defined as the in situ undrained shear strength mobilized in embankment and slope failure cases. The compiled Shanghai database (SH-CLAY/11/4051) contains four parameters that are also included in the global database (CLAY/10/7490), namely, LL, PI, LI, and σ'_v/P_a. The raw values of $S_u/\sigma'_{v\ (UCST)}$, $S_{t\ (UCST)}$, $S_u/\sigma'_{v\ (VST)}$, and $S_{t\ (VST)}$ can also be obtained from the CLAY/10/7490 database on the TC304 website (ISSMGE TC304 Webpage, 2020). Hence, in this study, the comparison of the SH-CLAY/11/4051 database with the CLAY/10/7490 database is carried out using the abovementioned eight parameters, i.e., LL, PI, LI, σ'_v/P_a, $S_u/\sigma'_{v\ (UCST)}$, $S_{t\ (UCST)}$, $S_u/\sigma'_{v\ (VST)}$, and $S_{t\ (VST)}$. For brevity, the databases SH-CLAY/11/4051 and CLAY/10/7490 in the following text will be referred to as SH-CLAY and ALL-CLAY, respectively.

3.1. Marginal statistics

Fig. 4 shows the boxplots of the eight parameters, namely, LL, PI, LI, σ'_v/P_a, $S_u/\sigma'_{v\ (UCST)}$, $S_{t\ (UCST)}$, $S_u/\sigma'_{v\ (VST)}$, and $S_{t\ (VST)}$. Logarithmic transformations are applied to these parameters to make the differences clearer, considering the different ranges of values covered in these databases. For example, PI lies between 10.4 and 26.5 in SH-CLAY, but covers a wider range between 1.9 and 363 in ALL-CLAY. Note that LI is not transformed owing to the existence of negative values in the ALL-CLAY database. Data for SH-CLAY and ALL-CLAY are shown in the left white boxplot and right grey boxplot in each figure, respectively. Furthermore, the statistics of eight clay parameters common to both databases are listed in Table 1. This clearly shows that all clay parameters in SH-CLAY are more clustered than those in ALL-CLAY, and that no data points are outside the range spanning the lower quartile minus 1.5 times the interquartile range (IQR) (Fig. 4(a)) and the upper quartile plus 1.5 times the IQR in ALL-CLAY. The IQR is calculated as the difference between the upper and lower quartile values. Data in SH-CLAY are not classified as “outliers” with respect to the global database based on this standard exploratory data analysis. This is not surprising given that the global database contains more soil types and covers a wider range of geologic/environmental conditions.

3.2. Correlation between parameter pairs

Given the eight parameters included in both databases, 8 × 7/2 = 28 distinct correlation plots can be generated in total between any two parameters. The scatter plots between two parameters pairs of parameters are shown below to demonstrate the consistency of SH-CLAY data in terms of how: (1) mean trends compare with established correlation equations such as those given in Report EL-6800 (Kulhawy and Mayne, 1990) and (2) bivariate distributions empirically represented by scatter plots compare with those from ALL-CLAY. Note that consistency in the marginal statistics involving one parameter (Fig. 4) does not imply consistency in bivariate dependency between two parameters. Only some parameter pairs, namely, PI with LL, PI with $S_u/\sigma'_{v\ (VST)}$, LI with σ'_v/P_a, LI with $S_{t\ (UCST)}$, LI with $S_{t\ (VST)}$, σ'_v/P_a with $S_u/\sigma'_{v\ (UCST)}$, and σ'_v/P_a with $S_u/\sigma'_{v\ (VST)}$, are presented herein. Details pertaining to other correlation plots can be found in Appendix B (Fig. B1).

Table 1
Basic statistics of the eleven parameters in SH-CLAY/11/4051.

Parameters	n	Mean	COV	Max	Min
LL	2229 (3822)	40.3 (67.7)	0.22 (0.8)	58.7 (515)	26.3 (18.1)
PI	4044 (4265)	18.2 (39.7)	0.18 (1.08)	30.9 (363)	10.3 (1.9)
LI	2067 (3661)	1.15 (1.01)	0.22 (0.78)	2.19 (6.45)	0.49 (−0.75)
e	3875	1.24	0.14	1.86	0.67
K_0	264	0.51	0.09	0.65	0.43
σ'_v/P_a	2658 (3370)	1.30 (1.80)	0.35 (1.47)	2.71 (38.74)	0.28 (4.13E-3)
$S_u/\sigma'_{v(UCST)}$	156 (589)	0.21 (0.6)	0.43 (2.63)	0.69 (27.8)	0.10 (0.04)
$S_{t(UCST)}$	181 (242)	5.2 (48.5)	0.28 (2.35)	7.6 (970)	1.5 (1.25)
$S_u/\sigma'_{v(VST)}$	379 (1607)	0.34(0.57)	0.38 (1.39)	1.14(9.77)	0.22 (0.02)
$S_{t(VST)}$	384 (761)	3.95 (48.8)	0.17(1.90)	7.8(1790)	2.7 (1.25)
p_s/σ'_v	1245	6.3	0.58	46.05	2.55

Note: values shown in the parentheses are the statistics of the corresponding parameters in CLAY/10/7490.

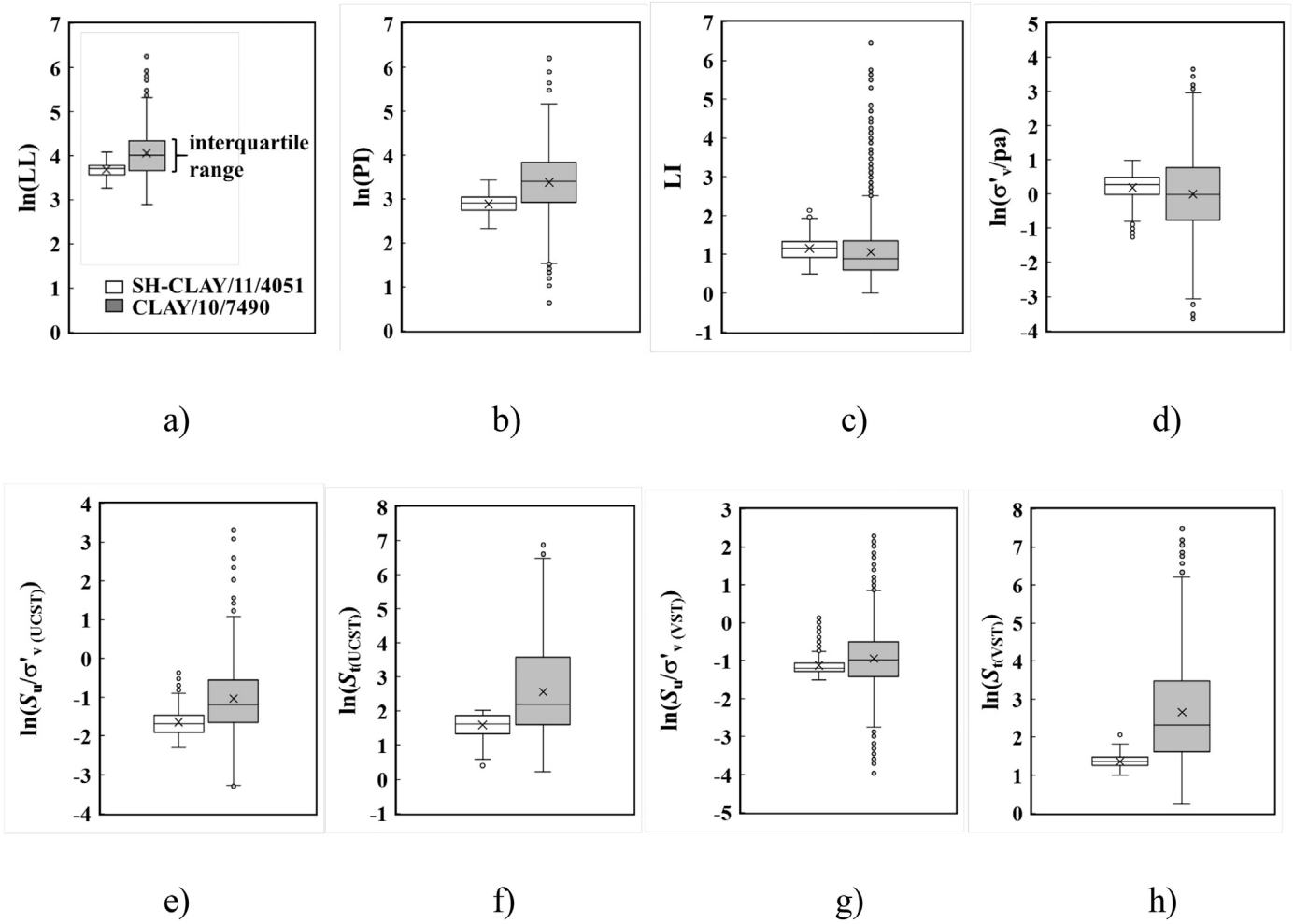


Fig. 4. Boxplots for: a) ln(LL); b) ln(PI); c) ln(LI); d) ln(σ'_v/P_a); e) ln($S_u/\sigma'_{v\ (UCST)}$); f) ln($S_{t\ (UCST)}$); g) ln($S_u/\sigma'_{v\ (VST)}$); and h) ln($S_{t\ (VST)}$).

3.3. PI vs. LL and $S_u/\sigma'_{v\ (VST)}$

Fig. 5(a) plots the plasticity chart (PI v.s. LL) of the two databases. It indicates that most data points in ALL-CLAY are classified as clays, some are classified as clayey silts or silt mixtures, and a few are classified as sand mixtures or sands (Ching and Phoon, 2014a). In contrast, data points in SH-CLAY are all classified as clay, with most of them classified as low-plasticity clay. Fig. 5(b) plots the correlation between PI and $S_u/\sigma'_{v\ (VST)}$. Skempton (1957) proposed the extensively used empirical equation that relates these two parameters for normally consolidated soils. It exhibits a linear and positive trend between PI and $S_u/\sigma'_{v\ (VST)}$. The curve is also plotted in Fig. 5(b). The figures indicate that the correlation between PI and $S_u/\sigma'_{v\ (VST)}$ is positive in ALL-CLAY, which is consistent with the empirical equation proposed by Skempton (1957). However, the correlation between PI and $S_u/\sigma'_{v\ (VST)}$ is distinctly negative in SH-CLAY. It should be noted that Skempton's curve was derived based on data collected mostly from the United Kingdom that may not be representative of global trends, because most data points in ALL-CLAY are above the empirical line. Moreover, the clays in ALL-CLAY are predominantly over-consolidated (OCR = 1–60.23) (Ching and Phoon, 2014a) while the Shanghai clay is normally consolidated or may be slightly overconsolidated (Wei and Hu, 1980). These underlying differences may partly explain the difference in observed trends between the SH-CLAY and ALL-CLAY database. While it is widely accepted as part of our conventional wisdom in practice that similar soils do not behave in the way at different sites because they are naturally occurring (the "site-specific" effect or the "*U*" in *MUSIC* data), it is rare to demonstrate this using two large databases such as SH-CLAY and ALL-CLAY.

3.4. LI vs. σ'_v/P_a, $S_{t\ (UCST)}$ and $S_{t\ (VST)}$

The correlations between LI and σ'_v/P_a, LI and $S_{t\ (UCST)}$ and LI and $S_{t\ (VST)}$ are shown in Fig. 6(a), (b), and (c) respectively. Wood (1983) proposed an empirical equation for LI and σ'_v/P_a, which is also plotted in Fig. 6(a). The LI is negatively related to σ'_v/P_a for both databases, and is consistent with the empirical equation shown in Fig. 6(a). It also indicates that SH-CLAY data fall on the right side of the empirical line. There are many reasons for this bias, but the sensitivity of the clay may be the most important factor. Based on the database presented by Wood (1983), the empirical equation is constructed for insensitive and for clays with low sensitivity, while Shanghai clay is of medium to high sensitivity (sensitivity = 2.7–7.8). Mitchell (1976) has presented relationships between LI and σ'_v/P_a for different levels of clay sensitivity, indicating that the correlation equation will shift gradually to the right side of the LI-σ'_v/P_a plot as the sensitivity increases (as shown by blue dot lines in Fig. 6(a)). This is consistent with the bias exhibited by SH-CLAY data when compared against the empirical equation proposed by Wood (1983).

For the correlations between LI and $S_{t\ (UCST)}$ and LI and $S_{t\ (VST)}$ shown in Fig. 6(b) and (c), respectively, empirical equations also exist in the literature (Bjerrum, 1954; Ching and Phoon, 2012). These empirical equations are plotted in Fig. 6(b) and (c). As shown in Fig. 6(b) and (c), the SH-CLAY data are much more clustered than those of ALL-

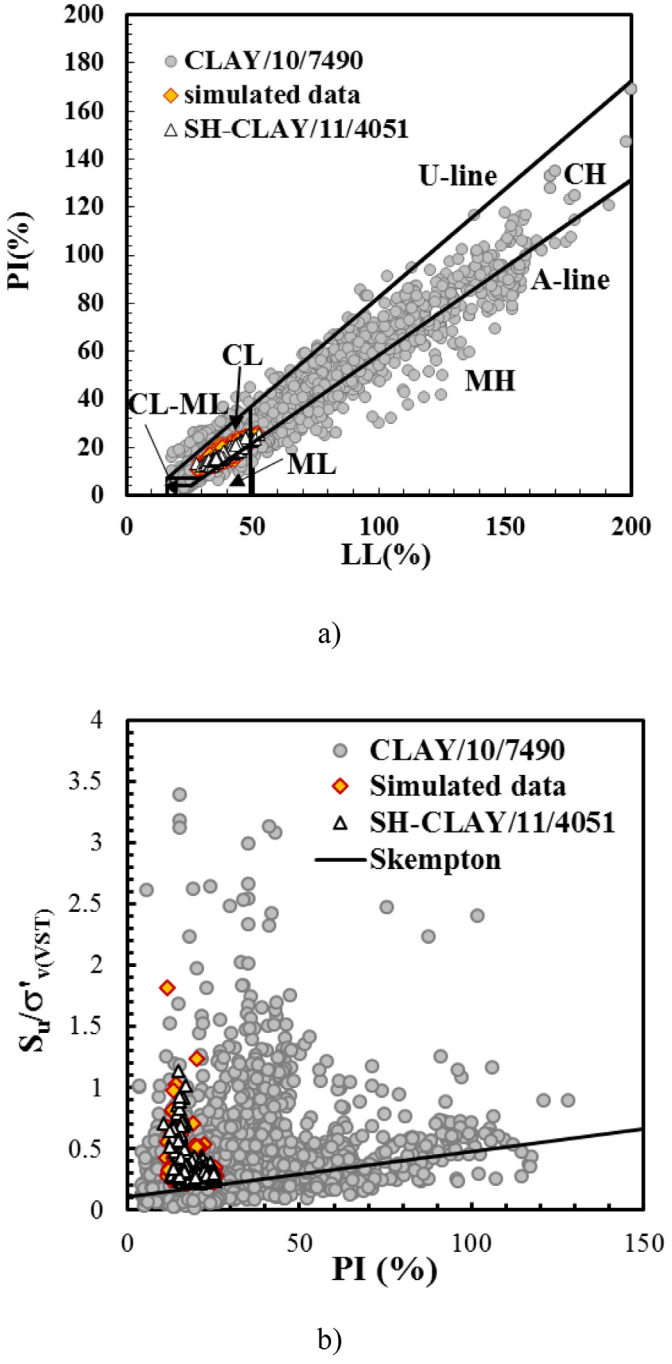


Fig. 5. a) Plasticity chart; CH, high-plasticity clay; CL, low-plasticity clay; MH, high-plasticity silt; ML, low-plasticity silt; b) correlation for PI vs. $S_u/\sigma'_{v\ (VST)}$.

CLAY, and the correlations for SH-CLAY are weaker compared with those for ALL-CLAY. Moreover, it is evident that the SH-CLAY data do not fit the existing equations and that the majority of data fall on the right side of these equations. To the authors' knowledge, there are no past studies discussing this behavior. However, the vast majority of the literature is related to soils in North America, Europe and Japan (Tan et al., 2003, 2006). The two empirical equations are supported by Scandinavian clay data, where the geologic history is distinct from that of Shanghai.

3.5. σ'_v/P_a vs. $S_u/\sigma'_{v\ (UCST)}$ and $S_u/\sigma'_{v\ (VST)}$

Fig. 7 plots the correlations between σ'_v/P_a and $S_u/\sigma'_{v\ (UCST)}$ and $\sigma'_v/$

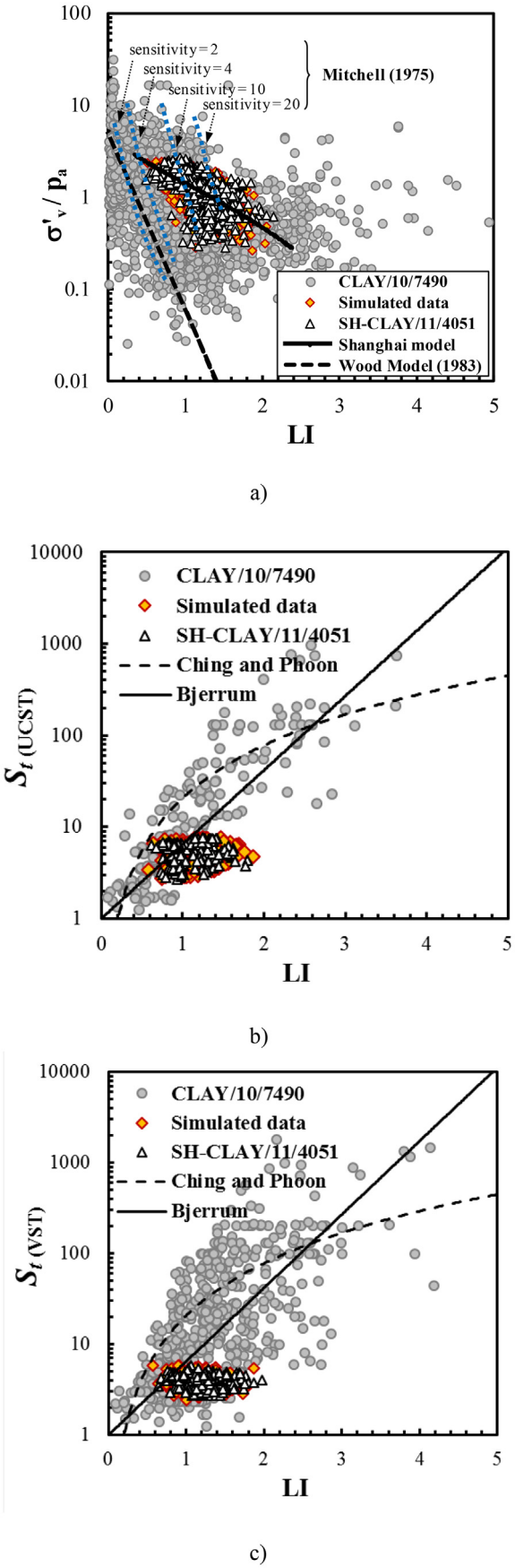
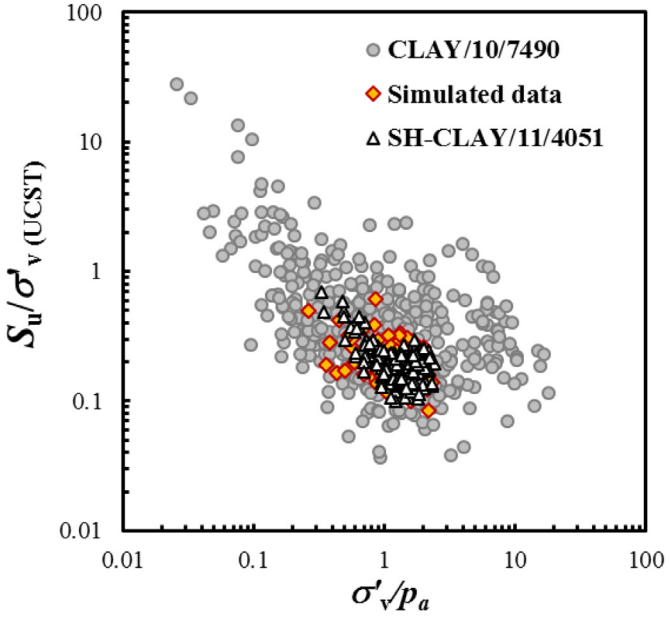
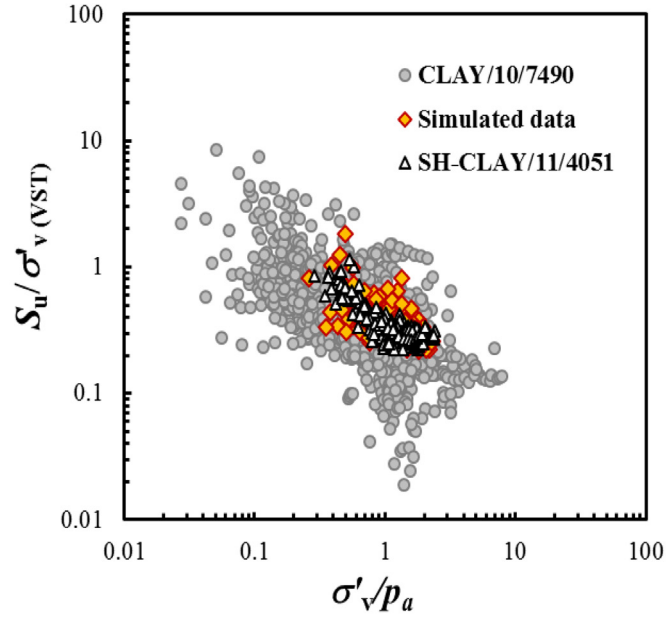


Fig. 6. Correlation for a) LI vs. σ'_v/P_a b) LI vs. $S_{t\ (UCST)}$ and c) LI vs. $S_{t\ (VST)}$.



a)



b)

Fig. 7. Correlation for a) σ'_v/P_a vs. $S_u/\sigma'_{v\ (UCST)}$ b) σ'_v/P_a vs. $S_u/\sigma'_{v\ (VST)}$.

P_a and $S_u/\sigma'_{v\ (VST)}$. Both subplots demonstrate the existence of similar negative correlation trends in the two databases. In general, the SH-CLAY data are more clustered, but they are embedded within the broader data scatter for ALL-CLAY. Visually, SH-CLAY and ALL-CLAY do not look like distinct populations. This is an example showing that certain correlations are not site-specific. This observation produced by two large databases is significant, as explained below.

3.6. Municipal effect

It is clear that the bivariate dependency between any two parameters produced by the two multivariate databases may or may not be comparable, depending on the parameter pair considered (refer to Appendix B). For example, the correlations for LL and PI, LI andσ'_v/P_a, σ'_v/P_a and $S_u/\sigma'_{v\ (UCST)}$, σ'_v/P_a and $S_u/\sigma'_{v\ (VST)}$, and $S_u/\sigma'_{v\ (UCST)}$ and $S_u/\sigma'_{v\ (VST)}$ (refer to Figs. 5–7) are consistent with the correlations shown in ALL-CLAY. It can be inferred that correlations between some specific parameter pairs, such as σ'_v/P_a and S_u/σ'_v in this case, are less affected by the municipal effect (clay type, geologic history, etc.). As a result, it can be concluded that generic transformation models are not always poor when applied to a specific site. The assumption that all transformation models are sensitive to geologic differences is not true and this may partially explain the success of generic models in practice, such as those found in Kulhawy and Mayne (1990). The characterization of the "unique" feature in *MUSIC* data in quantitative terms is clearly challenging when site data may or may not be distinct from global data, and the degree of distinctiveness changes depending on the specific pairs of parameters that are compared. One can hypothesize that this complex relationship between local and global data can extend beyond pairs of parameters (correlation) to higher order dependency structures among various parameters (> 2). It is an ongoing research challenge (e.g. Ching and Phoon, 2020b).

4. Multivariate normal distribution for SH-CLAY/11/4051

4.1. Normalized transformation

The multivariate normal distribution is the most frequently used in reliability analysis and random simulation, because of its tractability and ease of characterization from data (Phoon, 2006). However, many physical parameters do not follow the normal distribution in the marginal sense. To retain the advantages of the multivariate normal distribution, one common strategy is to transform non-normal soil parameters into normal variables. Phoon and Ching (2013) pointed out that it is advantageous to use the Johnson system of distribution to transform the non-normal variable Y_i to its corresponding standard normal random variable X_i, because the transformation is available in closed-form. For an arbitrary non-normal variable, this closed-form relationship between Y_i and X_i is not available. The most familiar member of the Johnson system is the lognormal distribution. Note that Y_i ($i = 1, \ldots, 11$) is defined as the natural logarithm of each of the 11 clay parameters included in SH-CLAY. The closed-form transformation that uses the Johnson distribution is given below,

$$X = \begin{cases} b_X + a_X \sinh^{-1}\left(\dfrac{Y - b_Y}{a_Y}\right) & \text{SU} \\ b_X + a_X \ln\left(\dfrac{Y - b_Y}{a_Y + b_Y - Y}\right) & \text{SB } (b_Y \le Y \le a_Y + b_Y) \\ b_X + a_X \ln\left(\dfrac{Y - b_Y}{a_Y}\right) & \text{SL } (Y \ge b_Y) \end{cases} \quad (1)$$

where (a_X, b_X, a_Y, b_Y) are the four distribution model parameters which can be estimated after identifying the distribution type. There are three types of distributions in the Johnson system (Johnson, 1949): SU (for the "unbounded system"), SB (for the "bounded system") and SL (for the "lognormal system"). Slifker and Shapiro (1980) proposed a robust non-parametric approach based on four percentiles of Y_i for identification of the distribution (SU, SB or SL) and estimation of (a_X, b_X, a_Y, b_Y). Details and examples are given in Ching and Phoon (2015). Following this approach, the distribution type and its corresponding parameters for (Y_1, Y_2 … and Y_{11}) are derived and summarized in Table 2. Note that the transformed normal vector (X_1, X_2 … and X_{11}) does not necessarily follow a multivariate normal distribution, even if

Table 2
Distribution type and distribution parameters for (Y_1, Y_2 … and Y_{11}).

Random variable	Soil parameters	Distribution type	Distribution parameters			
			a_X	b_X	a_Y	b_Y
Y_1	ln(LL)	SB	1.11	−0.31	0.68	3.29
Y_2	ln(PI)	SB	1.06	−0.32	0.96	2.32
Y_3	ln(LI)	SB	2.39	−0.53	2.26	−1.15
Y_4	ln$e$	SB	1.19	−0.41	0.83	−0.29
Y_5	ln(K_0)	SU	3.57	−2.76	0.26	−0.90
Y_6	ln(σ'_v/Pa)	SB	1.63	−1.81	3.48	−2.37
Y_7	ln($S_u/\sigma'_{v(UCST)}$)	SU	1.72	−1.01	0.45	−1.97
Y_8	ln($S_{t(UCST)}$)	SB	0.64	−0.36	1.04	0.99
Y_9	ln($S_u/\sigma'_{v(VST)}$)	SU	0.84	−0.86	0.08	−1.31
Y_{10}	ln($S_{t(VST)}$)	SU	2.49	−0.25	0.39	1.31
Y_{11}	ln(p_s/σ'_v)	SU	0.83	−0.67	0.11	1.56

each X_i is normally distributed. Nevertheless, past experiences revealed that this assumption of a multivariate normal distribution is still physically meaningful for a number of soil databases (Ching and Phoon, 2012, 2014b). It is possible to test for multivariate normality more formally using the square of the Mahalanobis distance (Ching and Phoon, 2015).

4.2. Correlation matrix of multivariate distribution

The multivariate standard normal probability density function (PDF) is solely defined by a correlation matrix **C**

$$f(\mathbf{x}) = |\mathbf{C}|^{-1/2}(2\pi)^{-n/2}\exp\left(-\frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{x}\right) \tag{2}$$

where $\mathbf{x} = (x_1, x_2, \ldots \text{and } x_n)^{\mathrm{T}}$ is a standard normal random vector (each component is a standard normal random variable with zero mean and unit standard deviation), the superscript "T" refers to the vector–matrix transpose, n is the number of variables (= 11 in this paper), and **C** is the correlation matrix composed of δ_{ij} that refers to the Pearson product-moment correlation coefficient between X_i and X_j.

It is clearly more practical to use independent bivariate datasets from various locations to estimate δ_{ij} than to use complete multivariate data. The former only requires the simultaneous measurements of two parameters, which is more commonly encountered than simultaneous measurements of three or more parameters. However, this approach can produce a nonpositive definite correlation matrix which is not valid. Furthermore, each δ_{ij} is associated with some statistical uncertainty, particularly when the sample size is small. Ching and Phoon (2014b) proposed a method to solve these problems by obtaining 1000 samples for each δ_{ij} using the bootstrapping procedure, and by randomly choosing a δ_{ij} sample within its 90% confidence interval to form a new correlation matrix until it becomes positive definite. A positive definite matrix and the corresponding δ_{ij} values will be accepted. Otherwise, it will be rejected. One thousand accepted positive definite matrices and the corresponding δ_{ij} values are generated with this approach, and the final **C** matrix is constructed from the average value of these 1000 valid matrices. Fig. 8 shows the scatter plots between all the possible parameter pairs (11 × 10/2 = 55 pairs). Table 3 presents the final **C** matrix produced by the bootstrapping procedure. Once **C** is determined, Eq. (2) is fully defined. Eq. (2) is useful for Bayesian updating and for developing transformation models specific to Shanghai's clay.

4.3. Simulations

It is straightforward to generate the random vector (x_1, x_2, … and x_{11}), using the multivariate normal distribution with a zero mean vector and the correlation matrix given in Table 3. The transformations from X to Y can be performed as follows,

$$Y = \begin{cases} b_Y + a_Y \sinh\left(\dfrac{X - b_X}{a_X}\right) & \text{SU} \\ b_Y + a_Y\left[1 + \exp\left(-\dfrac{X - b_X}{a_X}\right)\right]^{-1} & \text{SB} \\ b_Y + a_Y \ln\left(\dfrac{X - b_X}{a_X}\right) & \text{SL} \end{cases} \tag{3}$$

After transforming X_i to Y_i using Eq. (3), the simulated soil parameters in the original physical space can be obtained from e^{Yi}.

Note that Eq. (3) is an inverse transformation of Eq. (1). It is also available in closed-form, thus highlighting the attractiveness of the Johnson system of distributions. To demonstrate that the Johnson transformations and the multivariate distribution so constructed are reasonable, the correlation between the simulated variables in the original space is compared with the actual correlation between the measured data in Figs. 5–7 (represented by the solid diamond markers in these figures). For each parameter pair, the number of the simulated data is equal to that of the measured data. Visually, it is reasonable to say that the simulated and measured data belong to the same population in terms of the centroid location and degree of scatter. Despite the assumptions adopted to retain analytical simplicity (Johnson distributions and multivariate normality), the framework is applicable to SHCLAY. As mentioned above, this framework has been shown to be applicable to other soil databases (Ching and Phoon, 2015). The next section shows that the multivariate normal distribution with Johnson transforms produces reasonable estimates of the normalized undrained shear strength.

5. Application of the Shanghai clay multivariate normal distribution

5.1. Bayesian updating

In this section, Bayesian updating is briefly introduced. The constructed multivariate normal distribution serves as a prior distribution, and the posterior distribution can be obtained based on new measured data. The updated distribution is still multivariate normal owing to the conjugacy of the prior distribution. Additionally, the Johnson transform model is still applicable to the updated marginal distributions. The posterior Johnson PDF has the following form,

$$f(Y) = \begin{cases} \dfrac{a'_X}{a_Y\sqrt{2\pi\{1 + [(Y - b_Y)/a_Y]^2\}}} \times \\ \quad \exp\left\{-0.5\left[b'_X + a'_X \sinh^{-1}\left(\dfrac{Y - b_Y}{a_Y}\right)\right]^2\right\} & \text{SU} \\ \dfrac{a'_X a_Y}{\sqrt{2\pi}(Y - b_Y)(a_Y + b_Y - Y)} \times \\ \quad \exp\left\{-0.5\left[b'_X + a'_X \ln\left(\dfrac{Y - b_Y}{a_Y + b_Y - Y}\right)\right]^2\right\} \\ \quad \text{SB}\ (b_Y \le Y \le a_Y + b_Y) \\ \dfrac{a'_X}{\sqrt{2\pi}(Y - b_Y)} \times \exp\left\{-0.5\left[b'_X + a'_X \ln\left(\dfrac{Y - b_Y}{a_Y}\right)\right]^2\right\} \\ \quad \text{SL}\ (Y \ge b_Y) \end{cases} \tag{4}$$

in which (a'_X, b'_X) are the updated parameters after conditioning based on newly measured data, while the parameters (a_Y, b_Y) and the distribution type remain unchanged in the updating process. (a'_X, b'_X) can be calculated as follows:

$$a'_X = a_X/\sigma'_X$$
$$b'_X = (b_X - \mu'_X)/\sigma'_X \tag{5}$$

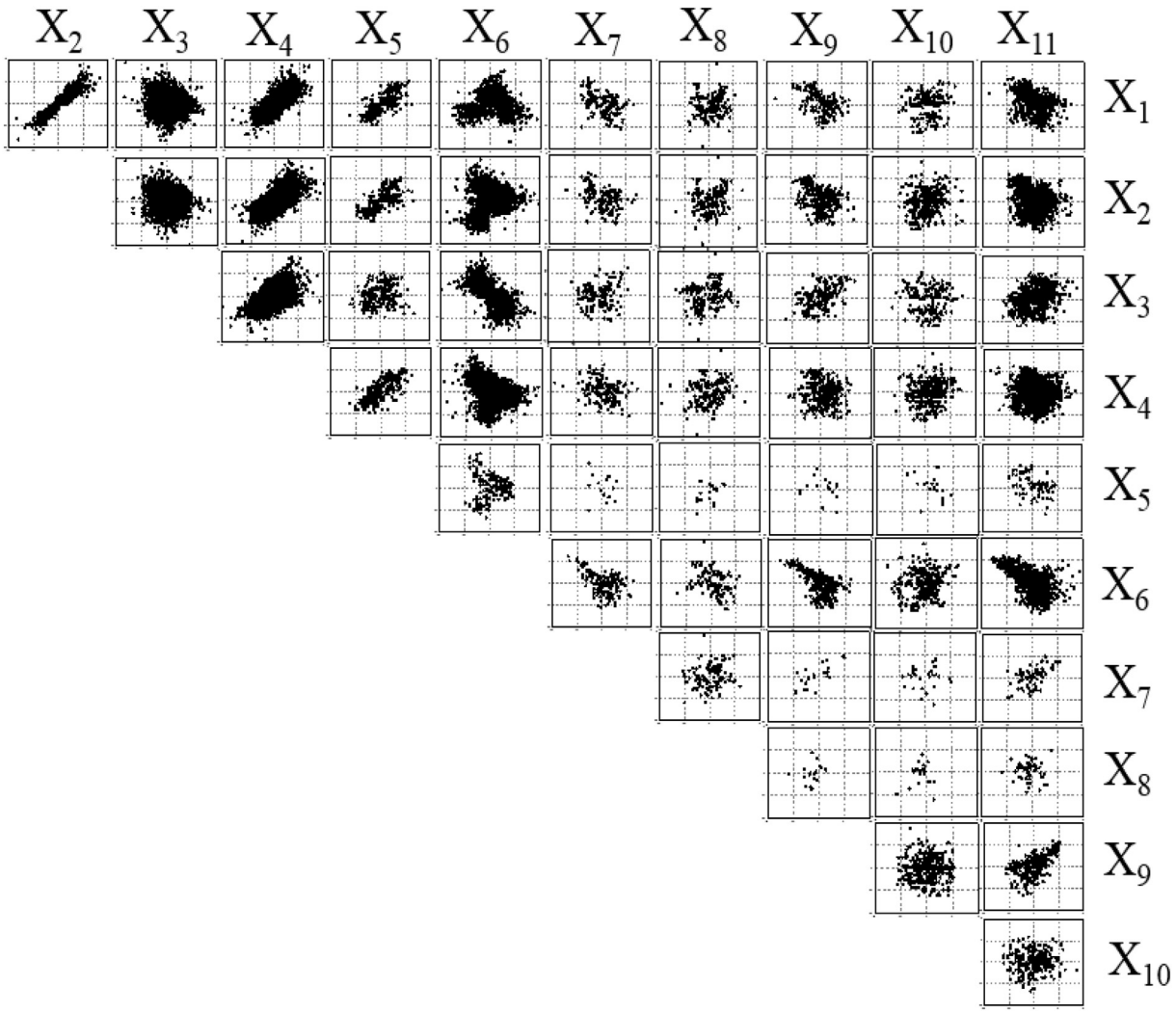


Fig. 8. Scatter plots between X_i and X_j.

in which μ'_x is the updated mean value of X_i and σ'_x is the updated standard deviation. Derivation details are given elsewhere Ching and Phoon (2014b). Eq. (4) is the updated marginal distribution of Y_i, the similar closed-form equations for updated joint distribution of (Y_i, Y_j) can also be developed according to Ching and Phoon (2014b). Given the demonstrated advantages in simulation and Bayesian updating, it is evident that the Johnson system is the preferred choice for transforming non-normal soil parameters to normal variables, unless there is compelling evidence that it is not suitable.

For this application example, the normalized undrained shear strength $S_u/\sigma'_{v\ (UCST)}$ (Y_7) at a specific depth of 13.1 m is updated based on information from other sources at the same depth: LL (Y_1), PI (Y_2), LI (Y_3), e (Y_4), σ'_v/P_a (Y_6), and p_s/σ'_v (Y_{11}). The set of data is measured at a depth of 13.1 m extracted from the site investigation report of this borehole which is located in the Xuhui District: Y_1 = ln(LL) = ln (45.2) = 3.81, Y_2 = ln(PI) = ln(22) = 3.09, Y_3 = ln(LI) = ln (1.35) = 0.30, Y_4 = ln(e) = ln(1.488) = 0.40, Y_6 = ln(σ'_v/P_a) = ln (114.4/101.3) = 0.12 and Y_{11} = ln(p_s/σ'_v) = ln(500/114.4) = 1.47.

The value of Y_7 can be updated based on the other information according to Bayesian theory. The prior and posterior PDFs of Y_7 are

Table 3
C matrix composed of the average of the 1000 accepted δ_{ij}.

X	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
X_1	1.00	0.92	−0.05	0.75	0.72	0.13	−0.43	0.09	−0.44	0.10	−0.30
X_2		1.00	−0.08	0.71	0.67	0.22	−0.37	0.11	−0.37	0.19	−0.24
X_3			1.00	0.55	0.20	−0.67	0.22	0.21	0.51	−0.06	0.31
X_4				1.00	0.69	−0.20	−0.25	0.18	−0.07	0.14	−0.04
X_5					1.00	−0.02	−0.28	0.31	−0.21	−0.17	−0.35
X_6						1.00	−0.49	−0.28	−0.65	0.14	−0.59
X_7		Symmetry					1.00	0.21	0.65	0.02	0.51
X_8								1.00	0.25	−0.02	−0.02
X_9									1.00	−0.02	0.58
X_{10}										1.00	0.08
X_{11}											1.00

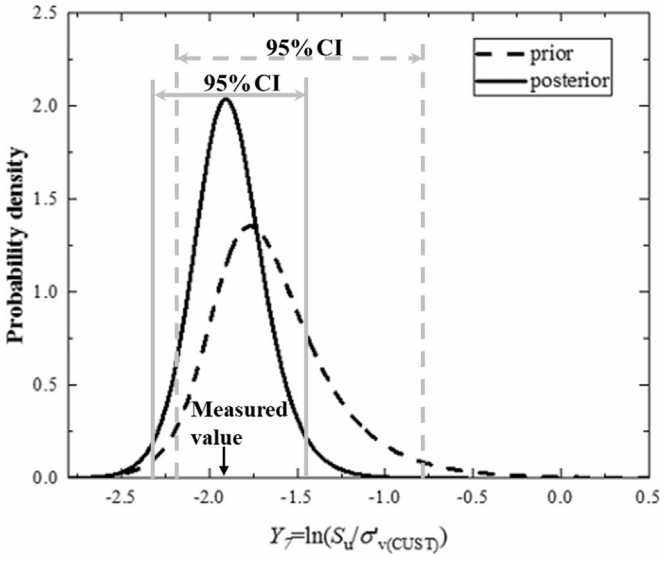


Fig. 9. Bayesian updating of Y_7.

presented in Fig. 9 (the black dash line is the prior distribution, while the black solid line is the posterior distribution). As shown in Fig. 9, the standard deviation decreases. It is worth noting that the true value of Y_7 (i.e., the measured value in situ) is also known as −1.88 according to the site report. Hence, it is obvious that the updated mean of Y_7 is getting closer to its measured value. Moreover, the measured value falls within the 95% confidence interval of the prior and posterior distributions (grey dash lines for prior distribution and grey solid lines for posterior distribution in Fig. 9). Finally, Fig. 9 also shows the 0.05 percentile that is recommended as the characteristic value in Eurocode 7. It is interesting to observe that the 95% confidence interval decreases after updating.

5.2. Developing transformation models for Shanghai clay

As shown in Figs. 5 and 6, existing empirical equations published in the literature are not well suited for Shanghai's clay. Transformation models specific to Shanghai clay can be developed by using the constructed multivariate normal distribution. The median value of the unknown parameter can be estimated by using the following equation:

$$Y' = b_Y + a_Y\left[1 + \exp\left(\frac{b'_X}{a'_X}\right)\right]^{-1} \quad (6)$$

in which a'_X and b'_X are the updated Johnson model parameters obtained from the procedure explained above. The median value Y' given by Eq. (6) is regarded as the best estimate of the unknown parameter Y given the measured values of X_i. Note that Eq. (6) is applicable for one or more X_i, because a'_X andb'_X are the updated parameters based on any number of newly measured data, i.e., not necessarily for only one parameter as the example presented here.

Take the parameters LI (X_3) and σ'$_v$/P_a (X_6) as an example. The conditional mean value of X_6 (σ'$_v$/P_a) equals $\delta_{36} \times X_3 = -0.67 \times {}_3$, and the conditional standard deviation σ'$_x$ equals $(1-(\delta_{36})^2)^{1/2} = 0.74$. The updated parameter $a'_X = 1.63/0.74 = 2.20$, and $b'_X = (-1.81 + 0.67 \times {}_3)/0.74 = -2.45 + 0.91 \times {}_3$. Therefore, the transformation model between LI and σ'$_v$/P_a can be derived as follows,

$$\sigma'_v/pa = \exp\left\{-2.37 + 3.48 \times \left[1 + \exp\left(\frac{-2.93 + 2.71 \times \ln\left(\frac{\ln(LI)+1.15}{1.11-\ln(LI)}\right)}{2.2}\right)\right]^{-1}\right\} \quad (7)$$

Eq. (7) is also plotted in Fig. 6a as a solid line and labeled as "Shanghai model" in the legend. It clearly provides a better transformation model for SH-CLAY/11/4051 than the Wood (1983) equation. Note that the range of the curve is limited ($0.32 \le LI \le 3.03$). This is because the Johnson distribution type of LI is SB, which is a bounded distribution.

6. Conclusion

This study compiles a database for Shanghai soft clay labeled as SH-CLAY/11/4051, whose data points were collected from site investigation reports of 50 sites in Shanghai over an area of approximately 145 km^2 to the north-east of the Shanghai Hongqiao International Airport and to the north-west of the Pudong International Airport. The clay parameters are collected from 553 boreholes, 47 vane shear tests, and 190 cone penetration tests. In this database, there are eleven clay parameters: 1) four index parameters, including the liquid limit (LL), plasticity index (PI), liquidity index (LI), and void ratio (e), and 2) seven mechanical parameters, including the at-rest lateral pressure coefficient (K_0), vertical effective stress (σ'$_v$), undrained shear strength ($S_{u(UCST)}$), the sensitivity ($S_{t(UCST)}$), obtained from the unconfined compression soil test (UCST), undrained shear strength ($S_{u(VST)}$) and sensitivity ($S_{t(VST)}$) obtained from the vane shear test (VST), and specific penetration resistance (p_s) from the CPT.

To the authors' knowledge, SH-CLAY/11/4051 is the first sizeable multivariate soil database compiled at the municipal level. General information is presented, including the geology of Shanghai, the borehole locations, the types of clay parameters, and the basic statistics of these parameters. Furthermore, a comparison is made between the SH-CLAY/11/4051 database and the published global database CLAY/10/7490. As expected, the data points in SH-CLAY/11/4051 are more clustered than those in CLAY/10/7490. Although site-specific effects are widely accepted in practice, they have not been demonstrated using such large databases and they have not been examined from a multivariate perspective (mainly bivariate perspective between two soil parameters). This study only examines site-specificity at the municipal level (called municipal effect), rather than at a project site level.

By comparing the marginal statistics and correlations between parameter pairs, it is concluded that the quality of the data in SH-CLAY/11/4051 is generally satisfactory given that all the data points fall within the range of the global database. It is further observed that the correlations between some parameter pairs (e.g., normalized effective vertical stress (σ'$_v$/P_a) and normalized undrained shear strength (S_u/σ'$_v$), S_u/σ'$_v$ measured by different test procedures) are less affected by the municipal effect. A multivariate probability distribution for the 11 clay parameters has been constructed based on the SH-CLAY/11/4051 database. Bayesian updating and the derivation of transformation models specific to Shanghai clay have been conducted using the constructed multivariate distribution to demonstrate its practical use and performance.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This study is substantially supported by the National Natural Science Foundation of China (No. 51608380, 51538009), by Shanghai Rising-star Program of Shanghai Science and Technology Committee (17QC1400300) and by Key innovation team program of innovation talents promotion plan by MOST of China (No. 2016RA4059). The authors are grateful to these programs. The authors would also like to thank Dr. Rulu Wang and Dr. Hua Shao from Shanghai metro company Ltd., for sharing the valuable database, and the technical committee on Engineering Practice of Risk Assessment & Management of the International Society of Soil Mechanics and Geotechnical Engineering for developing the database 304 dB (http://140.112.12.21/issmge/Database_2010.htm) used in this study and making it available for scientific inquiry. The third author engaged in this research during his sabbatical at the Institute for Risk and Reliability, Leibniz University, which was funded by the Alexander von Humboldt Foundation.

Appendix A

Table A1
General characteristics of the database SH-CLAY/11/4051.

Site ID	BH No.	Sample size	Value	Depth (m)	LL(%)	PI	LI	e	K_0	σ'_v (kPa)	$S_{u(UCST)}$ (kPa)	$S_{t(UCST)}$ (kPa)	$S_{u(VST)}$	$S_{t(VST)}$	p_s (MPa)
0	16 + 1 + 3	137	Min	4	27.2	12.1	0.59	0.77	0.44	48.5	16	3.2	32.6	3.8	0.5
			Max	29	48.9	23.5	1.54	1.55	0.57	262.8	29	4.9	68.7	5.1	4.2
1	9 + 1 + 8	52	Min	11	–	14.2	–	1.01	–	109	–	–	30.5	3.1	0.4
			Max	23		22.1		1.44		206.6			56.1	3.8	1.1
2	11 + 1 + 2	58	Min	12	–	12.3	–	0.97	–	117	–	–	36.8	3.6	0.5
			Max	23		22.8		1.41		204.1			51.1	3.8	0.9
3	11 + 2 + 10	36	Min	11.1	–	13.3	–	0.87	–	111.2	–	–	34.5	3.4	0.5
			Max	21		21.5		1.46		191.2			49	3.9	1.2
4	15 + 0 + 1	38	Min	8.5	–	12.6	–	1	–	80.7	–	–	–	–	0.6
			Max	17.1		23.3		1.51		175					0.7
5	10 + 2 + 8	48	Min	10	–	13.5	–	0.95	–	84.8	–	–	38	3.1	0.5
			Max	22		23.8		1.56		201.3			51.5	3.9	1.5
6	11 + 1 + 2	75	Min	10	–	11.1	–	0.91	–	102.4	–	–	35.9	3.7	0.6
			Max	22.1		24.1		1.46		199.4			55	4	1.2
7	3 + 1 + 3	35	Min	10	–	13	–	0.94	–	98.7	–	–	37.5	3.5	0.5
			Max	22		25		1.58		195			52.9	3.9	0.9
8	6 + 0 + 1	48	Min	8.5	–	13	–	0.88	–	78.7	–	–	–	–	0.7
			Max	21.5		23		1.49		193.1					0.9
9	3 + 0 + 3	4	Min	12.5	–	16.1	–	1.01	–	137	–	–	–	–	0.6
			Max	14		21.3		1.38		137					0.6
10	8 + 0 + 2	21	Min	11	–	14.3	–	1.14	–	105.4	–	–	–	–	0.6
			Max	15		22.8		1.79		146.8					0.8
11	9 + 1 + 7	63	Min	11.5	–	12.3	–	0.93	–	110.5	–	–	32.2	3.3	0.3
			Max	22		24.6		1.48		194.4			48	3.9	1.2
12	10 + 0 + 0	67	Min	10	–	12.5	–	0.92	–	–	–	–	–	–	–
			Max	23.5		20.9		1.51							
13	12 + 2 + 9	121	Min	10.4	31.6	12.1	0.6	0.83	0.44	112	21	3.3	38.6	3.4	0.7
			Max	29.2	48.7	26.4	1.53	1.56	0.63	250.6	55.5	5.6	76.6	5.1	2
14	10 + 0 + 1	89	Min	12	31	10.3	0.77	0.89	0.45	128.8	19.5	1.5	–	–	0.7
			Max	30	47.5	22.7	1.5	1.4	0.52	274.4	40.5	4.9			2
15	9 + 0 + 0	76	Min	12	31.7	11.2	0.71	0.85	0.45	–	20.5	3.1	–	–	–
			Max	29	45.4	22.9	1.4	1.48	0.53		54.5	5.7			
16	7 + 0 + 5	25	Min	13	–	14.4	–	0.93	–	129.6	–	–	–	–	0.5
			Max	21		22.3		1.5		195.9					0.8
17	9 + 1 + 0	59	Min	10	–	12.1	–	0.89	–	–	–	–	42.2	3.7	–
			Max	23.5		23.2		1.51					48.8	3.9	
18	7 + 0 + 0	38	Min	8.7	–	13.4	–	0.89	–	–	–	–	–	–	–
			Max	19		23.3		1.51							
19	10 + 2 + 7	39	Min	10	–	12.5	–	0.85	–	109.2	–	–	33.5	3.3	0.5
			Max	21		23.1		1.5		189.4			57.6	3.7	1.2
21	5 + 2 + 4	14	Min	12	–	16	–	0.95	–	122.7	–	–	31	3.3	0.7
			Max	19.5		20.7		1.42		186.8			48.8	3.8	0.8
22	9 + 0 + 0	74	Min	3	–	12.3	–	0.83	–	51.5	–	–	–	–	–
			Max	26		23.3		1.56		166.1					
23	8 + 1 + 6	49	Min	3	–	12.6	–	0.81	–	39	–	–	29.9	3.2	0.4
			Max	19		24.2		1.61		165.3			41.8	3.7	1.4
24	5 + 0 + 2	35	Min	3	–	12.1	–	0.82	–	40.3	–	–	–	–	0.5
			Max	16		24.4		1.61		144					1.7
25	9 + 2 + 7	40	Min	3	–	13.7	–	1.06	–	36.1	–	–	30.2	3.2	0.4
			Max	16		22.6		1.62		137.7			61.9	4	1.4
26	27 + 2 + 11	205	Min	3.5	31.4	12.2	0.61	0.92	0.43	36.6	19.5	3	28.9	2.7	0.3
			Max	24.2	47.6	22.9	1.84	1.61	0.62	198.9	28.5	5.9	46.4	3.6	1
27	15 + 2 + 2	106	Min	4.1	34.7	13.8	0.64	1	0.46	46.3	14.5	4.2	26.9	2.7	0.4
			Max	23.1	46.7	20.4	1.56	1.5	0.59	198.9	27	6.7	42.3	3.4	1.9
28	9 + 1 + 2	81	Min	6	31.2	13.5	0.54	0.84	0.46	58	18.5	3.2	25.1	2.8	0.4
			Max	23.15	51.1	23.7	1.96	1.66	0.61	193.6	29.5	5.6	44.4	3.3	1.5
29	11 + 2 + 9	95	Min	3.5	32.7	12.3	0.81	0.9	0.44	38	16	3.4	32.5	3.8	0.3
			Max	23.5	51.8	24.7	2.13	1.86	0.6	184.1	27	5.8	52.2	4.8	1.6

(*continued on next page*)

Table A1 (*continued*)

Site ID	BH No.	Sample size	Value	Depth (m)	LL(%)	PI	LI	e	K_0	σ'_v (kPa)	$S_{u(UCST)}$ (kPa)	$S_{t(UCST)}$ (kPa)	$S_{u(VST)}$	$S_{t(VST)}$	p_s (MPa)
30	14 + 2 + 4	82	Min	6	32.3	11.6	0.86	1.02	0.54	61	12.5	3.3	32.1	3.7	0.3
			Max	23	51.2	23.8	1.87	1.71	0.62	179.9	22	6.2	45.9	4.8	0.9
31	20 + 3 + 16	120	Min	4	31.6	13.3	0.71	0.93	0.45	40.2	12	3.5	32.5	3.8	0.4
			Max	24	58.7	30.9	2.04	1.66	0.65	206.4	24	5.6	48.9	4.6	3.7
32	8 + 0 + 1	77	Min	3.5	–	12.6	–	0.75	–	38.6	–	–	–	–	0.3
			Max	26.5		25.8		1.63		206.4					1
33	5 + 1 + 4	38	Min	4.5	–	13.2	–	0.67	–	42.4	–	–	25.3	2.9	0.4
			Max	26		23.6		1.58		217.2			54	4	1.3
34	14 + 1 + 3	135	Min	3	–	12.1	–	0.84	–	43.3	–	–	35.4	3.9	0.4
			Max	28		25.3		1.65		235.8			46.1	4.7	2
35	5 + 0 + 4	47	Min	4	32.6	12.3	0.75	1.02	0.44	38.3	23	–	–	–	0.3
			Max	22	49.1	25	1.73	1.59	0.56	199.5	39.5				1.3
36	11 + 2 + 1	95	Min	3	–	11.4	–	0.86	–	34	–	–	34.9	3	0.9
			Max	24		25.3		1.63		202.1			58.4	4.3	1.3
37	10 + 2 + 9	79	Min	2.5	30.4	12.3	0.76	0.98	0.49	29.2	18.5	–	21	4.7	0.4
			Max	7	50.5	26.5	1.79	1.63	0.55	192.1	42.5		48.6	5.8	1.3
38	13 + 0 + 2	98	Min	2.5	–	10.7	–	0.92	–	28.7	–	–	–	–	0.4
			Max	26		26		1.56		216.4					1
39	9 + 2 + 8	73	Min	3	28.1	10.5	0.77	0.96	0.5	32.7	19.5	–	30.8	3.2	0.3
			Max	27	50.7	25.2	2.19	1.6	0.59	214	33		49.8	5.4	1
40	17 + 2 + 1	158	Min	3	–	10.4	–	0.79	–	37.4	–	–	22.6	3.6	0.3
			Max	26		25.1		1.67		212			68.5	4.8	0.8
41	5 + 2 + 4	56	Min	3.5	–	10.4	–	1.01	–	37.4	–	–	24	3.4	0.4
			Max	25		25.5		1.58		205.4			59	7.8	1
42	12 + 0 + 1	82	Min	3.5	–	10.4	–	0.96	–	43.5	–	–	–	–	0.4
			Max	24		26.2		1.55		199.3					0.8
43	6 + 2 + 6	76	Min	3	30.6	12.2	0.76	0.9	0.48	34.8	15		20.5	3.1	0.3
			Max	23	49.1	23.8	1.75	1.59	0.57	197.3	35.5	—	59.8	4.8	1.7
44	10 + 0 + 0	89	Min	3	–	11.6	–	1.04	–	–	–	–	–	–	–
			Max	24.5		25.6		1.64							
45	5 + 0 + 5	33	Min	8	–	14.2	–	1.16	–	80.9	–	–	–	–	0.5
			Max	25.5		24.3		1.54		223.3					1.4
46	1 + 0 + 1	4	Min	10.2	–	14.2	–	–	–	–	–	–	–	–	0.58
			Max	22.2		22.6									1.04
47	16 + 1 + 3	186	Min	3	26.3	11.2	0.7	0.79	0.46	33.3	19	2.8	35.2	–	0.5
			Max	26.5	51.2	25.5	1.93	1.61	0.64	220	44	4.3	69.8		1.7
48	7 + 0 + 2	113	Min	3	34.1	14.2	0.73	0.92	0.45	36	14.5	6.4	–	–	0.3
			Max	24.6	50.3	25.8	1.39	1.54	0.53	202	52	7.4			2
49	44 + 0 + 0	496	Min	2.5	31.5	11.2	0.49	0.81	0.43	33.6	12	5.7	–	–	–
			Max	30	49.3	23.7	1.65	1.63	0.61	260.9	52	7.6			
52	37 + 0 + 0	86	Min	2.1	31.1	12.2	0.54	0.91	0.46	–	–	–	–	–	–
			Max	17.2	44.9	23.1	1.77	1.42	0.6						

Note: (1) the column heading “Site ID” represents the serial number of each site which is defined by the authors for simplicity; (2) the column heading “BH No.” means the borehole numbers of each site constituted of the number of boreholes+ VSTs+ CPTs; (3) the column heading “sample size” means the total number of data sets and each set of data is a row in the Excel; (4) the rest of the table presents the range of each parameter and “—” means lack of information.

Appendix B

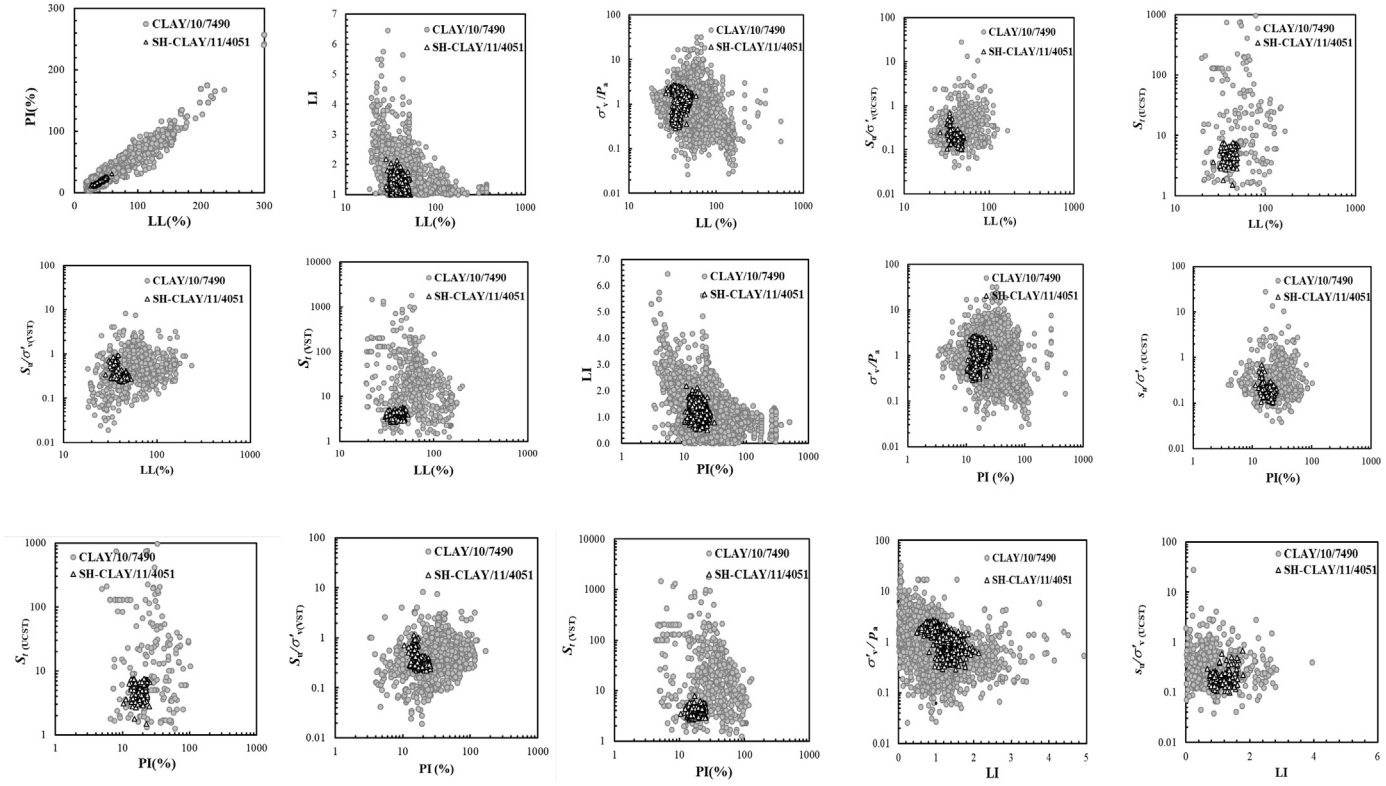
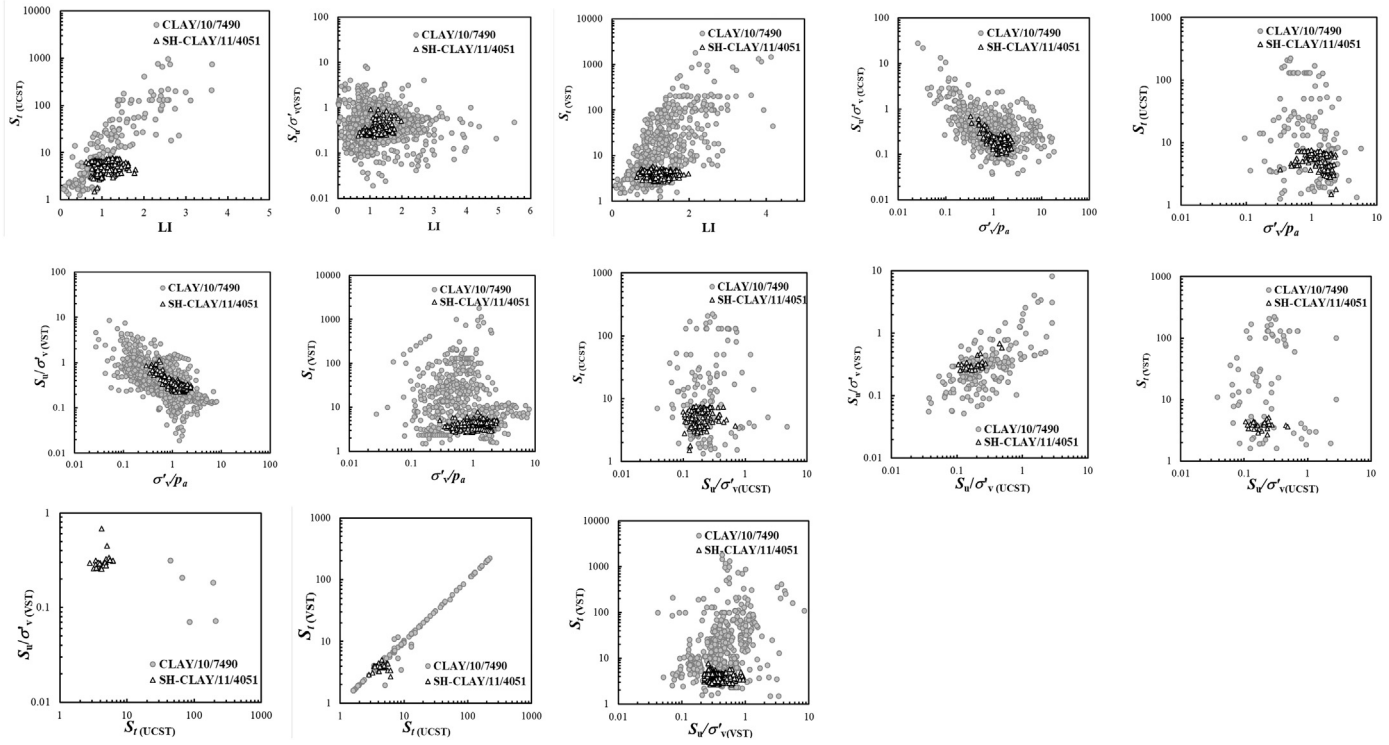


Fig. B1. Correlation plots for 28 parameter pairs.



Referencess

Akca, N., 2003. Correlation of SPT-CPT data from the United Arab Emirates. Eng. Geol. 67, 219–231.

Baziar, M.H., Jafarian, Y., 2007. Assessment of liquefaction triggering using strain energy concept and ANN model: capacity energy. Soil Dyn. Earthq. Eng. 27, 1056–1072.

Bjerrum, Laurits, 1954. Geotechnical properties of norwegian marine clays. Géotechnique 4 (2), 49–69.

Cai, G., Liu, S., Tong, L., 2010. Field evaluation of deformation characteristics of a

lacustrine clay deposit using seismic piezocone tests. Eng. Geol. 116 (3–4), 251–260.
Cai, G.J., Puppala, A.J., Liu, S.Y., 2014. Characterization on the correlation between shear wave velocity and piezocone tip resistance of Jiangsu clays. Eng. Geol. 171, 96–103.
Ching, J., Phoon, K.K., 2012. Modeling parameters of structured clays as a multivariate normal distribution. Can. Geotech. J. 49 (5), 522–545.
Ching, J., Phoon, K.K., 2014a. Transformations and correlations among some clay parameters—the global database. Can. Geotech. J. 51 (6), 663–685.
Ching, J., Phoon, K.K., 2014b. Correlations among some clay parameters–the multivariate distribution. Can. Geotech. J. 51 (6), 686–704.
Ching, J., Phoon, K.K., 2015. Constructing multivariate distribution for soil parameters. In: Chapter 1, Risk and Reliability in Geotechnical Engineering. CRC Press, pp. 3–76.
Ching, J., Phoon, K.K., 2019. Constructing site-specific probabilistic transformation model by Bayesian machine learning. ASCE J. Eng. Mech. 145 (1) 04018126.
Ching, J., Phoon, K.K., 2020a. Constructing a site-specific multivariate probability distribution using sparse, incomplete, and spatially variable (MUSIC-X) data. ASCE J. Eng. Mech. 46 (7) 04020061.
Ching, J. & Phoon, K. K. (2020b). Measuring similarity between site-specific data and records in a geotechnical database. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering, 6(2), 04020011.
Ching, J., Phoon, K.K., Chen, C.H., 2014. Modeling piezocone cone penetration (CPTU) parameters of clays as a multivariate normal distribution. Can. Geotech. J. 51 (1), 77–91.
Ching, J., Li, D.Q., Phoon, K.K., 2016. Statistical characterization of multivariate geotechnical data. In: Chapter 4, Reliability of Geotechnical Structures in ISO2394. CRC Press, Balkema, pp. 89–126.
Ching, j.Y., Phoon, K.K., Khan, Z., Zhang, D.M., Huang, H.W., 2020. Role of municipal database in constructing site-specific multivariate probability distribution. Comput. Geotech. 124, 103623.
Das, S.K., Basudhar, P.K., 2008. Prediction of residual friction angle of clays using artificial neural network. Eng. Geol. 100, 142–145.
D'Ignazio, M., Phoon, K.K., Tan, S.A., Lansivaara, T., 2016. Correlations for undrained shear strength of Finnish soft clays. Can. Geotech. J. 53 (10), 1628–1645.
Forster, A., Culshaw, M.G., 1990. The use of site investigation data for the preparation of engineering geological maps and reports for use by planners and civil engineers. Eng. Geol. 29 (4), 347–354.
Huang, H.W., Xiao, L., Zhang, D.M., Zhang, J., 2017. Influence of spatial variability of soil Young's modulus on tunnel convergence in soft soils. Engineering Geology 228, 357–370. https://doi.org/10.1016/j.enggeo.2017.09.011.
ISO2394, 2015. General Principles on Reliability for Structures. International Organization for Standardization, Geneva, Switzerland.
ISSMGE TC304 Webpage. http://140.112.12.21/issmge/tc304.htm.
Johnson, N.L., 1949. Systems of frequency curves generated by methods of translation. Biometrika 36, 149–176.
Khan, M.S., Hossain, S., Ahmed, A., Faysal, M., 2017. Investigation of a shallow slope failure on expansive clay in Texas. Eng. Geol. 219, 118–129.
Khanlari, G.R., Heidari, M., Momeni, A.A., Abdilor, Y., 2012. Prediction of shear strength parameters of soils using artificial neural networks and multivariate regression methods. Eng. Geol. 131, 11–18.
Kulhawy, F.H., Mayne, P.W., 1990. Manual on estimating soil properties for foundation design. Report EL-6800. In: Electric Power Research Institute. Palo Alto, California.
Lashkaripour, R.G., Ajalloeian, R., 2003. Determination of silica sand stiffness. Eng. Geol. 68 (3), 225–236.
Lee, M.J., Choi, S.K., Kim, M.T., Lee, W., 2011. Effect of stress history on CPT and DMT results in sand. Eng. Geol. 117 (3–4), 259–265.
Liu, S., Zou, H., Cai, G., Bheemasetti, B.V., Puppala, A.J., Lin, J., 2016. Multivariate correlation among resilient modulus and cone penetration test parameters of cohesive subgrade soils. Eng. Geol. 209, 128–142.
Long, M., Gudjonsson, G., Donohue, S., Hagberg, K., 2010. Engineering characterization of Norwegian glaciomarine silt. Eng. Geol. 110, 51–65.
Mitchell, J.K., 1976. Fundamentals of Soil Behavior. J. Wiley & Sons, Toronto.
Phoon, K.K., 2006. Modeling and simulation of stochastic data. In: Proceedings, GeoCongress, ASCE, CDROM.
Phoon, K.K., 2018. Editorial for special collection on probabilistic site characterization. ASCE-ASME J. Risk and Uncertain. Eng. Syst. A Civil Eng. 4 (4), 02018002.
Phoon, K.K., 2020. The story of statistics in geotechnical engineering. Georisk 14 (1), 3–25.
Phoon, K.K., Ching, J., 2013. Multivariate model for soil parameters based on Johnson distributions. In: Foundation Engineering in the Face of Uncertainty: Honoring Fred H. Kulhawy (GSP 229). ASCE, Reston, pp. 337–353.
Phoon, K.K., Kulhawy, F.H., 1999a. Characterization of geotechnical variability. Can. Geotech. J. 36 (4), 612–624.
Phoon, K.K., Kulhawy, F.H., 1999b. Evaluation of geotechnical property variability. Can. Geotech. J. 36 (4), 625–639.
Phoon, K.K., Ching, J., Wang, Y., 2019. Managing risk in geotechnical engineering – from data to digitalization. In: Proceedings, 7th International Symposium on Geotechnical Safety and Risk (ISGSR 2019). Taiwan, Taipei.
Skempton, A.W., 1957. Discussion of "Planning and design of new Hong Kong airport". Proc. Inst. Civ. Eng. 7, 305–307.
Slifker, J.F., Shapiro, S.S., 1980. The Johnson system: selection and parameter estimation. Technometrics 22 (2), 239–246.
Tan, T.S., Phoon, K.K., Hight, D.W., Leroueil, S. (Eds.), 2003. Proceedings, Characterization and Engineering Properties of Natural Soils. Balkema.
Tan, T.S., Phoon, K.K., Hight, D.W., Leroueil, S. (Eds.), 2006. Proceedings, Characterization and Engineering Properties of Natural Soils. Taylor & Francis.
Wang, C.C., 1978. Some experiences with an electrical static penetrometer. Bull. Int. Assoc. Eng. 18 (1), 153–156.
Wei, D.D., Hu, Z.X., 1980. Experimental study of preconsolidation pressure and compressibility of Shanghai subsoil. Chin. J. Geotech. Eng. 2 (4), 13–22 (in Chinese).
Wood, D.M., 1983. Index properties and critical state soil mechanics. In: Proceedings, Symposium on Recent Developments in Laboratory and Field Tests and Analysis of Geotechnical Problems, Bangkok, pp. 301–309.
Xiao, L. (2018). Probabilistic Analysis of Shield Tunnel Convergence by Random Field of Multilayer Soil. M.S. thesis, Tongji University, Shanghai. (in Chinese).
Zhang, W.G., Goh, A.T.C., Zhang, Y.M., Chen, Y.M., Xiao, Y., 2015. Assessment of soil liquefaction based on capacity energy concept and multivariate adaptive regression splines. Eng. Geol. 188, 29–37.
Zhang, D.M., Huang, H.W., Hu, Q.F., Jiang, F., 2015. Influence of multi-layered soil formation on shield tunnel lining behavior. Tunnelling and Underground Space Technology 47, 123–135. https://doi.org/10.1016/j.tust.2014.12.011.
Zou, H., Liu, S., Cai, G., Puppala, A.J., Bheemasetti, T.V., 2017. Multivariate correlation analysis of seismic piezocone penetration (SCPTU) parameters and design properties of Jiangsu quaternary cohesive soils. Eng. Geol. 228, 11–38.