Engineering Geology 209 (2016) 128–142

Contents lists available at ScienceDirect

Engineering Geology

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

Multivariate correlation among resilient modulus and cone penetration test parameters of cohesive subgrade soils

Songyu Liu [a], Haifeng Zou [a], Guojun Cai [a,*], Tejo Vikash Bheemasetti [b], Anand J. Puppala [b], Jun Lin [a]

[a] Institute of Geotechnical Engineering, Southeast University, Nanjing, 210096, Jiangsu, China
[b] Department of Civil Engineering, The University of Texas at Arlington, Arlington, TX 76019, United States

ARTICLE INFO

Article history:
Received 3 December 2015
Received in revised form 16 May 2016
Accepted 28 May 2016
Available online 30 May 2016

Keywords:
Multivariate normal distribution
Resilient modulus
CPTU
Bayesian updating
Box-Cox transformation

ABSTRACT

Extensive research has been conducted to establish empirical equations between the resilient modulus (M_r) and other testing parameters. Despite an increase in the correlation studies, less effort has been made in developing reliable correlations to predict the resilient modulus using cone penetration testing indices. In this research, an attempt has been made to propose new correlations between M_r and the piezocone penetration test (CPTU) indices of clayey soils by using a multivariate normal distribution approach. A database collected from 16 different sites in Jiangsu province, China, was presented to achieve this study. The database contains 124 sets of resilient modulus (M_r) values at the in-situ stress condition, cone tip resistance (q_c), sleeve frictional resistance (f_s), moisture (w) and dry density (γ_d). Four major procedures including the data transformation, correlation analysis, Bayesian updating, and back transformation were applied to derive the correlations among M_r and other indices in the framework of the multivariate normal distribution model. First, each individual parameter was converted to a standard normal variable using the Box-Cox transformation. Later, the correlation matrix of the multivariate normal distribution model was estimated using the product-moment (Pearson) correlation coefficients between all pairwise data. The uncertainties associated with the Box-Cox transformation parameters and the correlation coefficients were evaluated using a bootstrapping technique. Furthermore, the formulas for predicting the posterior mean, coefficient of variation (COV), median, and 95% confidence interval (CI) of M_r conditional on different indices were derived using Bayesian updating combined with back transformation. A new approach based on the Taylor-series expansion was proposed to approximate these statistics in this study. Comparisons between the correlations derived from the multivariate model, the database of Jiangsu clay, and the data collected from the literature demonstrated that the model should be capable of the correlations among the five indices, but it might be slightly biased when applied in the global dataset. This research highlighted a new method to establish the reliable correlations to update M_r using different testing indices.

1. Introduction

Subgrade soils play a prominent role in the design and performance of a pavement structure. Subgrade soil properties such as dry density, moisture content, gradation and shape of soil particle, fines content and stress state have a relatively high degree of impact on the resilient behavior of subgrade soils (Hicks and Monismith, 1971; AASHTO, 1993; Gudishala, 2004; Kim and Labuz, 2007; Kim et al., 2014). The resilient modulus (M_r) property is typically determined using the repeated load triaxial test with a closed loop servo pneumatic loading system (AASHTO, 1993; NCHRP, 2004). However, obtaining the undisturbed soil sample for the laboratory test is time-consuming and not cost effective. Alternatively, researchers have attempted to develop correlations among M_r and in situ testing indices (Heukelom and Klomp, 1962; Duncan and Buchignani, 1976; Mohammad et al., 1999, 2000, 2002, 2007; Dehler and Labuz, 2007).

The piezocone penetration test (CPTU) is an advanced in-situ test and has been widely used in geotechnical engineering due to its high accuracy, good repeatability and low costs (Puppala et al., 1995; Lunne et al., 1997; Mayne, 2007; Liu et al., 2011; Cai et al., 2011, 2012, 2014). The piezocone test provides three separate measurement readings including the cone tip resistance (q_c), sleeve frictional resistance (f_s) and pore water pressure (u), and also gives near continuous information about the field subsurface stratification. Mohammad et al. (2007) investigated the correlations between M_r and ten other testing indices including q_c, f_s, moisture (w), dry density (γ_d), particle gradation, liquid limit, and plasticity index. It was concluded that q_c, f_s, w, and γ_d are the only variables that have a significant pairwise correlation to M_r.

* Corresponding author.
E-mail addresses: liusy@seu.edu.cn (S. Liu), zhf0728@gmail.com (H. Zou), focuscai@163.com (G. Cai), tejovikash.bheemasetti@mavs.uta.edu (T.V. Bheemasetti), anand@uta.edu (A.J. Puppala), aslnjun@163.com (J. Lin).

http://dx.doi.org/10.1016/j.enggeo.2016.05.018
0013-7952/

However, past correlations among M_r and CPTU or laboratory data were empirically established based on either pairwise data or the subjective combination of several selected multivariate testing indices. In this research, the multivariate correlations among M_r and four parameters including q_c, f_s, w, and γ_d are investigated using a new multivariate normal distribution theory. The multivariate normal distribution is an advanced statistical approach to model more than one random variable and it provides a unique advantage of obtaining the multivariate correlations between any two or more testing indices (Phoon, 2006; Phoon and Ching, 2012, 2013; Ching et al., 2011, 2012, 2014; Ching and Phoon, 2012, 2013, 2014).

In this research study, four major steps are involved in developing multivariate correlations: Data transformation, Correlation analysis, Bayesian updating, and Back transformation. The data transformation procedure involves evaluating the statistical distribution in the data and later transforming the non-Gaussian data to Gaussian data using Box-Cox transformations. In previous studies, the data transformation was achieved using the Johnson system of distribution, which is powerful but complicated. The Box-Cox transformation, used in this study, relies on a single power (λ) which is a simple and effective solution for non-Gaussian distributed data. The correlation analysis involves the construction of a multivariate normal distribution model using the transformed variables. Later the Bayesian updating technique combined with the back transformation was performed to determine the posterior statistics such as the mean value of a parameter (e.g., M_r) conditional on different input information (e.g., q_c, f_s, w, or γ_d).

In order to perform the above mentioned study, a database containing one hundred and twenty four (124) sets of $\{M_r, q_c, f_s, w, \gamma_d\}$ was collected and compiled from sixteen sites located at the Jiangsu province, China. Then the data transformation and bivariate correlation analysis was performed to construct the multivariate normal distribution model. The application of the multivariate normal distribution model in updating the statistics of M_r based on the Bayesian theory is then described. To validate the constructed multivariate model and investigate the possible application of the model in a global dataset, a comparison study between the formulas derived from the multivariate normal distribution model and data points collected from global sites was performed. According to the results and analysis, salient findings and conclusions are presented in this paper.

2. Jiangsu clay database

In order to perform multivariate analysis for obtaining multivariate correlations among resilient modulus and piezocone testing indices, a database containing 124 sets of $\{M_r, q_c, f_s, w, \gamma_d\}$ was compiled from Jiangsu province. The geological conditions of the testing sites, laboratory test program, and piezocone penetration tests are presented in the following sections.

2.1. Site description and laboratory tests

The investigation was conducted on the Quaternary clays in Jiangsu province, China. Fig. 1 presents a map of Jiangsu province with nine different cities, where 16 different testing sites were considered for this study. The geological formations of Quaternary clays are mainly Marine, Yangtze River Delta, Floodplain of Long River, Floodplain of Abandoned Yellow River, Lagoon of Lixia River and Lagoon of Taihu Lake. Extensive soil investigations have been performed at these sites over the last decade. The subsoil strata in these test sites mainly consists of soft to stiff clayey soils and silty clay soils with high variability of the strength and stiffness characteristics.

Based on the soil samples from the testing sites, laboratory tests were conducted to determine the moisture content (w), dry density (γ_d) and resilient modulus (M_r) properties. The resilient modulus tests were conducted in accordance with AASHTO standards (AASHTO, 2003). From the 16 different sites, a database is compiled for performing the multivariate analysis. To predict M_r under the in-situ stress level, CPTU soundings have been conducted at adjacent locations near the boreholes where the soil samples were collected. The horizontal distance between the borehole and CPTU sounding is less than 2 m for each set of $\{M_r, q_c, f_s, w, \gamma_d\}$ data. The piezocone testing equipment, representative CPTU profiles and the multivariate database are presented in the subsequent sections.

2.2. Field piezocone penetration testing

Piezocone penetration tests (CPTU) were performed in this research to characterize the subsurface soils in Jiangsu province. All the tests were conducted in accordance with the international standards

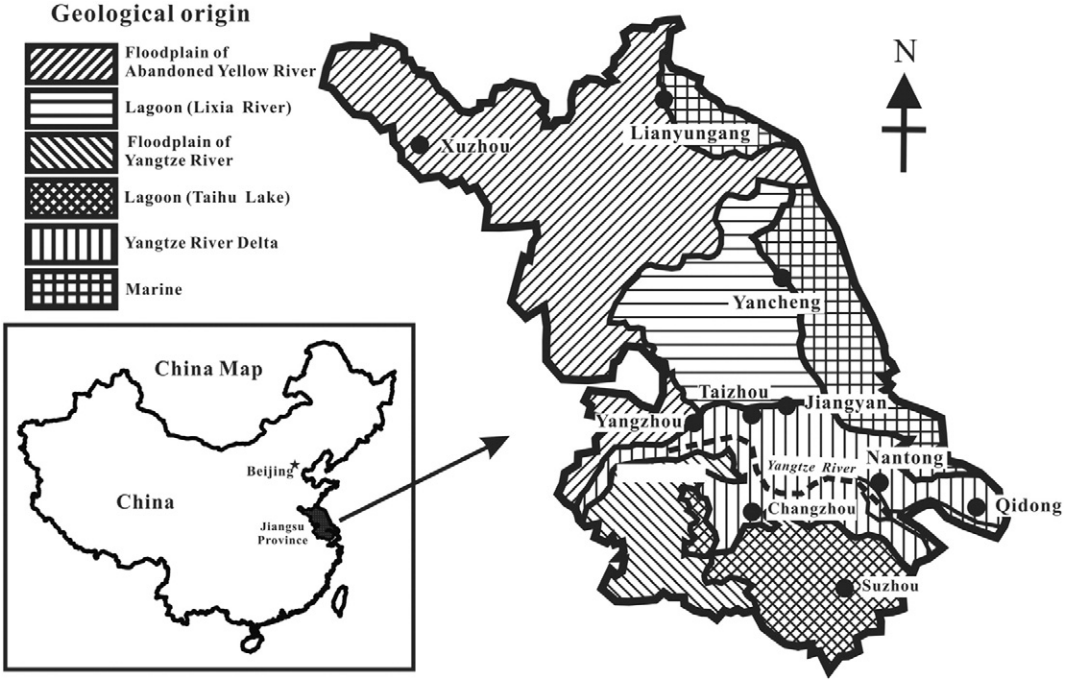

Fig. 1. Distribution and geologic formation of the testing sites.

(ISSMFE, 1989; ASTM D5778, 2012). The section area of the cylindrical cone penetrometer is 10 cm^2 with a tip angle of 60°. The surface area of the friction cylinder is 150 cm^2. The filter made from porous plastic is located at the gap between the cone tip and sleeve, i.e. u_2 position, with a thickness of 5 mm. The ground water table (GWT) at the test locations varied from 0.4 to 4.5 m and was documented immediately after the CPTU tests. A representative profile of the CPTU sounding at the shallow depth is presented in Fig. 2.

2.3. Database of the multiple parameters

In this study, the soil properties considered for the analysis are denoted as $Y_1 = M_r$, $Y_2 = q_c$, $Y_3 = f_s$, $Y_4 = w$, $Y_5 = \gamma_d$. A database containing 124 sets of $\{Y_1, Y_2, \ldots, Y_5\}$ is compiled from 16 sites in Jiangsu Province, as shown in Table 1. The M_r, q_c, and f_s are in MPa, w is in percent (%) and γ_d is in kN/m^3 in this research. Table 2 provides the basic statistics including the mean, COV, maximum value (Max), and minimum value (Min) of the multiple variables in the database. As shown in Table 2, high variability is observed in M_r, q_c, f_s, and w data with the COV ranging from 30% to 48%, whereas the magnitude of variability of γ_d (COV = 13%) is relatively small. Using these data, a multivariate normal distribution model will be constructed to capture the correlations among the five indices. The data of Jiangsu clay shown in Table 1 is referred to as the calibration database in this research.

3. Construction of multivariate normal distribution

This section investigates the application of the multivariate normal distribution in modeling the correlation among $\{Y_1, Y_2, Y_3, Y_4, Y_5\}$ of Jiangsu clay. The multivariate normal distribution model assumes that the marginal probability density distribution of every variable is Gaussian. Therefore, the normality of each variable is tested using the histograms. Based on the observations, the non-normally distributed variables are converted to the standard normal variables using a Box-Cox transformation procedure. Then the correlation matrix of the transformed variables is established using the product-moment (Pearson) correlation coefficients to construct the multivariate normal distribution model.

3.1. Data transformation using Box-Cox method

Fig. 3 presents the histograms of the five indices, $\{Y_1 = M_r, Y_2 = q_c, Y_3 = f_s, Y_4 = w, Y_5 = \gamma_d\}$, in the database. It is shown that the histograms of all five parameters are either left or right skewed with respect to their mean values, depicting a non-normal behavior. The normal distribution in the data is also checked using a formal normality test, the Shapiro-Wilk (SW) test (Shapiro and Wilk, 1965; Razali and Wah, 2011). The criterion of the SW test for rejecting the null hypothesis of the normal distribution is expressed in terms of p-value and significance level (α). The p-value evaluates the probability of the evidence that the null hypothesis is true. The significance level (α) indicates the probability of rejecting the null hypothesis given that the hypothesis is true. In this research the significance level is determined as $\alpha = 0.05$. If the p-value is larger than α, then the normality assumption cannot be rejected at the specific significance level. Using the MATLAB library function, *swtest*, the normality assumption of the five parameters ($Y_1, Y_2, \ldots, Y_5$) should be rejected at the significance level of 0.05 as the p-values are 0.011, 0.009, 0.016, 0.0019, and 0.0197, respectively. Hence to achieve the multivariate normal distribution, all these five indices should be converted to normal variables individually.

Several non-linear transformation methods are available to convert a non-Gaussian variable to a standard normal variable in the statistical literature (e.g., Box and Cox, 1964; Johnston, 1984; Montgomery et al., 2010; Charitidou et al., 2015; Liu et al., 2015). In this study the Box-Cox method is applied because of its practical and effective use in geotechnical research (Box and Cox, 1964; Sakia, 1992; Jaksa, 2007; Bheemasetti, 2015; Puppala et al., 2015). In the Box-Cox method only a single power parameter needs to be determined for converting a non-normal variable to a normal variable. It is more convenient to express the variables in standard formats and thus the covariance matrix of these variables can be simplified to be a correlation matrix, which can facilitate the multivariate normal distribution model. Hence, each non-Gaussian random variable (Y_i) is transformed to approximate a standard normal random variable (X_i) using the Box-Cox method as follows:

$$X_i = \begin{cases} \left[\dfrac{(Y_i)^{\lambda_i} - 1}{\lambda_i} - a_i\right] / b_i \\ [\ln(Y_i) - a_i] / b_i \end{cases} \tag{1}$$

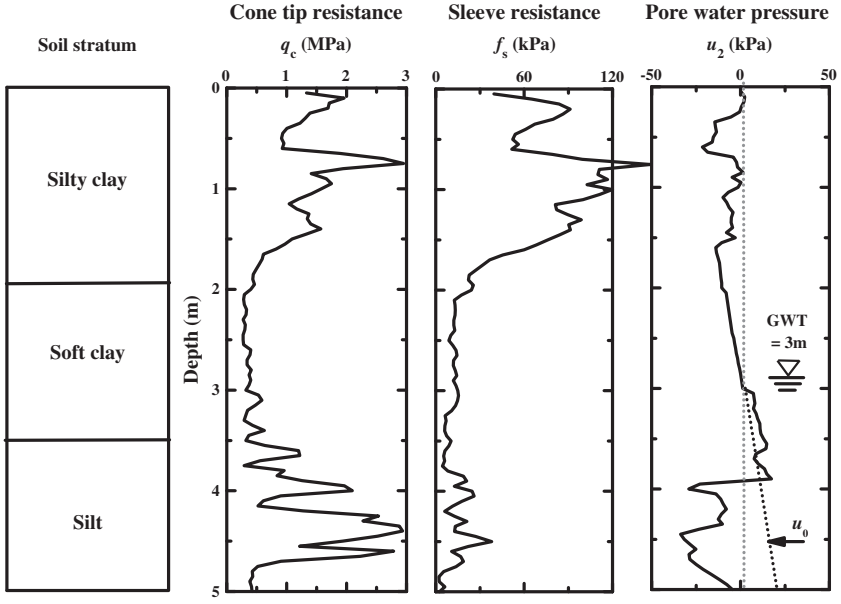

Fig. 2. Representative profiles of piezocone penetration test.

Table 1
Multivariate (calibration) database of the Jiangsu subgrade soils.

No.	Region	Site	M_r (MPa)	q_c (MPa)	f_s (MPa)	w (%)	γ_d (kN/m^{-3})
1			37.50	1.94	0.092	50.2	14.8
2			34.20	2.15	0.085	60.2	14.1
3			32.40	1.70	0.118	66.1	14.5
4			78.70	3.49	0.109	20.4	17.5
5			31.00	1.29	0.066	38.2	13.1
6	Lianyungang & Yancheng	S226 Provincial Expressway	46.50	1.35	0.105	23.3	11.6
7			23.90	1.21	0.071	60.2	14.0
8			28.90	1.52	0.058	57.3	17.2
9			55.00	3.29	0.095	33.0	14.9
10			35.50	1.82	0.126	53.3	14.0
11			65.70	2.97	0.105	25.8	17.6
12			49.20	2.70	0.074	39.4	18.7
13			71.80	2.80	0.130	22.8	17.2
14			25.40	0.62	0.100	39.9	16.2
15			32.10	1.44	0.069	44.8	16.9
16			22.90	0.52	0.066	31.6	13.2
17	Lianyungang & Yancheng	Coastal High-Grade Highway	54.40	1.78	0.131	44.8	19.8
18			27.40	0.78	0.098	43.9	16.6
19			44.50	1.29	0.108	26.4	17.1
20			34.40	1.51	0.108	47.7	15.7
21			56.30	2.41	0.127	44.9	17.3
22			67.10	2.50	0.125	24.6	18.8
23			43.00	2.26	0.099	42.1	16.9
24			53.20	2.87	0.129	34.3	10.6
25			40.90	1.55	0.064	24.4	12.0
26			21.70	1.45	0.047	49.3	10.5
27	Lianyungang & Yancheng	Huaiyan Expressway	26.10	0.55	0.039	23.8	14.2
28			12.50	0.36	0.035	60.3	13.7
29			58.30	3.73	0.031	23.9	15.7
30			13.10	1.01	0.030	78.1	12.6
31			13.80	0.22	0.046	42.2	13.9
32			53.20	2.05	0.092	26.2	15.5
33			54.50	2.28	0.068	22.8	16.8
34	Lianyungang & Yancheng	Linlian Expressway	42.20	1.33	0.108	31.2	16.5
35			49.40	2.31	0.072	28.7	16.8
36			39.00	2.38	0.131	64.6	15.5
37			45.20	1.31	0.091	28.4	17.2
38			34.90	1.22	0.046	24.2	14.5
39			31.80	0.60	0.124	27.2	14.0
40			54.60	1.78	0.130	34.0	16.8
41			45.50	1.41	0.088	31.7	16.4
42			71.90	3.34	0.106	15.3	15.5
43			43.00	1.96	0.103	32.7	13.2
44	Lianyungang & Yancheng	Lian-yan Expressway	44.80	1.95	0.133	51.1	16.4
45			33.10	2.26	0.055	33.5	10.5
46			34.00	0.45	0.110	29.2	18.7
47			31.50	1.29	0.099	51.4	15.5
48			46.30	2.37	0.106	34.9	11.9
49			55.60	3.02	0.122	31.3	14.8
50			72.70	3.34	0.079	15.3	15.4
51			47.60	2.03	0.117	43.4	17.0
52			33.50	1.14	0.094	39.4	17.9
53			24.40	0.57	0.074	34.9	14.3
54	Lianyungang & Yancheng	Lian-yan Railway	23.10	0.75	0.098	50.0	14.0
55			39.70	1.47	0.099	41.5	16.6
56			42.40	2.32	0.101	41.2	15.5
57			50.10	1.76	0.129	36.1	16.2
58			29.40	0.83	0.071	31.6	14.2
59	Changzhou	Xiaocheng Expressway	38.10	1.15	0.060	26.5	16.8
60			47.00	2.37	0.098	27.6	14.3
61			43.30	2.47	0.062	32.9	15.4
62			53.00	2.04	0.108	26.4	15.5
63			44.50	1.51	0.088	28.2	16.5
64	Changzhou	Ningchang Expressway	26.10	0.99	0.136	59.3	14.6
65			36.40	1.77	0.057	21.7	13.8
66			26.60	1.12	0.095	52.6	12.9
67			39.40	1.43	0.058	22.1	12.4
68			42.20	1.72	0.088	27.7	13.0
69	Yangzhou	Ningtong Expressway	36.30	1.02	0.094	26.6	16.3
70			46.10	1.01	0.085	22.4	14.5
71			36.00	1.78	0.062	21.2	13.7
72			57.40	1.18	0.137	15.8	17.4
73	Taizhou	Jiangyan Expressway	57.30	1.03	0.109	17.0	18.6
74			50.50	1.86	0.095	32.7	18.3

(continued on next page)

Table 1 (*continued*)

No.	Region	Site	M_r (MPa)	q_c (MPa)	f_s (MPa)	w (%)	γ_d (kN/m^{-3})
75			36.10	0.79	0.099	43.9	17.0
76			49.40	1.33	0.097	17.1	16.2
77			55.80	2.24	0.100	24.2	16.9
78			55.20	3.29	0.093	32.5	14.8
79			65.70	3.01	0.104	25.1	15.5
80			54.70	1.68	0.098	23.6	19.5
81	Taizhou	Yangtze River Bridge	50.10	1.10	0.093	22.9	19.8
82			42.10	1.01	0.096	19.6	15.2
83			23.60	1.29	0.065	53.4	11.8
84			18.90	0.72	0.047	49.6	15.5
85			57.00	2.42	0.112	24.4	16.2
86			45.30	1.14	0.074	17.9	17.2
87			54.50	1.27	0.071	10.1	15.7
88			70.00	2.30	0.083	13.3	17.5
89	Qidong	Chongqi Bridge	73.70	2.31	0.086	7.9	17.5
90			40.10	1.13	0.115	27.4	13.8
91			88.20	3.41	0.136	15.9	17.3
92			91.50	2.87	0.097	6.9	18.3
93			35.30	1.41	0.063	30.3	15.3
94			58.10	1.95	0.101	22.7	17.3
95			57.40	2.52	0.107	30.9	17.9
96	Qidong	Coastal High-grade Highway	52.30	2.68	0.077	27.5	16.2
97			33.00	0.54	0.041	20.7	18.9
98			45.30	1.30	0.119	29.2	17.2
99			45.40	1.14	0.114	16.9	17.2
100			60.10	2.62	0.091	18.7	14.4
101			46.60	1.34	0.116	21.2	11.6
102			61.70	2.63	0.072	19.2	16.5
103	Xuzhou	Ji-Xu Highway	37.10	0.82	0.034	16.5	16.9
104			33.90	1.58	0.095	40.0	12.7
105			22.80	0.59	0.107	51.4	15.8
106			81.40	2.89	0.140	19.9	18.4
107			85.40	3.10	0.093	11.1	19.2
108			57.90	2.77	0.125	38.5	16.9
109			26.40	0.70	0.058	37.2	17.2
110	Suzhou	Husuzhe Expressway	72.10	2.43	0.120	14.5	17.1
111			62.00	1.19	0.117	12.4	17.5
112			67.30	2.59	0.128	22.0	16.4
113			30.80	1.16	0.099	49.2	17.8
114			29.20	0.52	0.068	27.1	16.4
115			45.10	2.24	0.086	41.9	17.0
116			73.70	2.80	0.098	11.8	15.3
117			73.80	1.88	0.108	10.9	18.0
118	Yancheng	Funing-Jianhu Expressway	85.80	3.07	0.144	22.0	19.9
119			95.80	3.93	0.111	12.1	18.2
120			33.60	0.45	0.089	28.2	18.6
121			39.20	0.66	0.079	19.2	16.6
122			34.50	1.55	0.120	49.6	15.5
123			34.70	1.70	0.122	57.3	16.0
124			34.80	1.18	0.053	32.1	18.6

where Y_i and X_i are respectively the ith soil parameter and corresponding standard normal variable, λ_i is the power exponent for the transformation, and a_i and b_i are respectively the mean and standard deviation of $[(Y_i)^{\lambda i} - 1] / \lambda_i$ (when $\lambda_i \neq 0$) or $\ln(Y_i)$ (when $\lambda_i = 0$).

Estimation of the optimal power exponent (λ_i) has been discussed in a lot of statistical literature (e.g., Box and Cox, 1964; Sakia, 1992) and is also available in common statistical software such as MATLAB (MATLAB, 2013). In this research the *boxcox* function based on the maximum likelihood algorithm is adopted in MATLAB. Table 3 lists the three transformation parameters ($\boldsymbol{\lambda}_i$, a_i, b_i) along with the corresponding SW test result for each X_i variable. The histograms of the five transformed variables are presented in Fig. 4, as well as the curves of the theoretical standard normal distribution which fit well to the histograms for all five transformed indices. Also, the SW p-values presented in Table 3 are larger than the significance level of 0.05. Hence, the null hypothesis of the normal distribution should not be rejected for each Y_i variable.

Table 2
Basic statistics of the multiple variables.

Variables (Unit)	Mean	COV	Min	Max
$Y_1 = M_r$ (MPa)	46.13	0.37	12.54	95.82
$Y_2 = q_c$ (MPa)	1.76	0.48	0.22	3.93
$Y_3 = f_s$ (MPa)	0.09	0.30	0.03	0.14
$Y_5 = w$ (%)	32.00	0.44	6.91	78.11
$Y_6 = \gamma_d$ (kN/m^3)	15.86	0.13	10.47	19.92

The sample volume may be a major concern in the above transformation as the statistics of the geotechnical parameters are generally approximately unbiased, i.e., they are unbiased only if the sample volume is large enough. To address the uncertainties associated with the estimated λ_i, the bootstrapping sampling technique proposed by Ching and Phoon (2014) is applied as follows:

1. n random samples of Y_i are drawn from the calibration database, and n is determined to be 30, 40, …, 120. The minimum number of n is selected as 30 because it is reported to be the minimum sample size required to support a statistically significant conclusion generally (Phoon and Ching, 2012). The maximum number of n is determined as 120 to ensure that the samples could approximate the statistics of λ_i in the large sample volume, and that at the meantime the

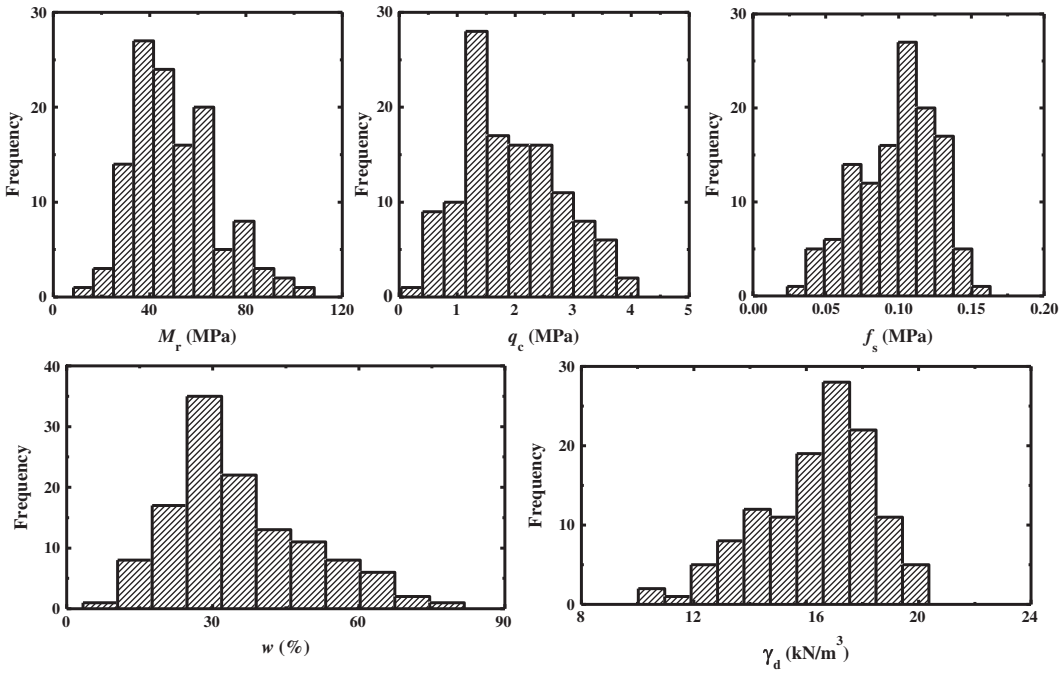

Fig. 3. Histograms of the five selected indices.

bootstrapping sampling would not always generate duplicate samples in different realizations and thus lead to biased analysis;

2. For each sample, the parameter λ_i is estimated using the *boxcox* function in MATLAB;
3. For each sample, identify whether the hypothesis of normality should be rejected for the sample of X_i variable, which is obtained by converting the sample of Y_i using the specific λ_i value listed in Table 3;
4. Repeat the above procedure 4000 times to obtain 4000 bootstrap samples of λ_i and the counts of rejecting the hypothesis of normality;
5. Calculate the mean λ_i and the ratio of rejecting the null hypothesis of normality in the 4000 realizations. The rejecting ratio is determined as the number rejecting the null hypothesis divided by the total number (4000) of the realizations.

If in the 4000 realizations the mean of λ_i approximates the value listed in Table 3 and always remains constant at different sample volumes, then it is supposed that the λ_i value in Table 3 is an unbiased estimate to achieve the Box-Cox transformation. If the rejecting ratio is small enough during the 4000 realizations at each bootstrapping sample, then the specific λ_i value in Table 3 is considered to be always reasonable in converting the Y_i to a normal variable. Fig. 5(a) presents the mean λ_i at each sample volume in the 4000 realizations. Fig. 5(b) illustrates the ratio of rejecting the hypothesis of normality for X_i obtained by converting the sample of Y_i using the specific λ_i value listed in Table 3. Fig. 5(a) shows that the mean values of λ_i at different sample volumes consistently approximate the specific λ_i listed in Table 3. The only exception is that when the sample volume is 30, the mean λ_1 (corresponding to M_r) is about 0.31, which is much smaller than the specific λ_1 (0.41) in Table 3. However, when the sample volume is larger than 60, the mean λ_i becomes almost robust and stable around the corresponding specific values. Moreover, from Fig. 5(b), it can be observed that when the specific λ_i value listed in Table 3 is applied, the ratio of rejecting the hypothesis of normality for the sample of X_i is rather small, consistently less than 8%. These observations indicate that the values listed in Table 3 are acceptable estimates of λ_i to achieve the marginal normal distribution of the soil parameters. The bivariate correlations between any two transformed variables will be discussed subsequently.

Table 3
Estimated Box-Cox transformation parameters of (Y_1, Y_2, …, Y_5).

Soil parameter	λ	*a*	*b*	SW *p*-Value
$Y_1 = M_r$	0.41	9.09	1.82	0.72
$Y_2 = q_c$	0.53	0.58	0.67	0.22
$Y_3 = f_s$	1.40	−0.69	0.01	0.08
$Y_4 = w$	0.34	6.40	1.46	0.78
$Y_5 = \gamma_d$	2.33	275.07	79.27	0.38

3.2. Multivariate correlation between pairwise data

The construction of the multivariate normal distribution model merely relies on the pairwise correlation between X_i and X_j. In the framework of the multivariate normal distribution model, the multivariate probability density function of p variables is determined as follows (Ching et al., 2014):

$$f(\mathbf{X}) = |\mathbf{C}|^{-1/2}(2\pi)^{p/2}e^{-\frac{1}{2}\mathbf{X}^T\mathbf{C}^{-1}\mathbf{X}} \quad (2)$$

where $\mathbf{X}$ is the vector containing $\{X_1, X_2, \ldots, X_p\}$, $\mathbf{C}$ is the covariance matrix of the p variables, $p = 5$ in this study, and the notation "T" refers to the vector-matrix transpose. Since the X_i variable has been standardized, the covariance matrix is simplified to be the correlation matrix.

The bivariate correlations between the pairwise data have to be addressed to construct the model. It should be realized that although each transformed variable follows a standard normal distribution, it is not necessary that the five transformed variables follow the multivariate normal distribution. A fundamental assumption of multivariate normal distribution is that all possible pairs of variables should be correlated linearly. Fig. 6 presents the scatter plots between all possible pairs of X_i and X_j. A visual inspection on the scatter plots shows that non-linear pairwise correlations cannot be observed in X_i variables. Therefore, all the trends within X_i and X_j can be viewed to be linear. There is no strong evidence to reject the underlying null hypothesis of the multivariate normality for the transformed indices. The correlations among multivariate information $\{X_1, X_2, \ldots, X_5\}$ based on the database of Jiangsu clay are further analyzed to construct the multivariate normal distribution model.

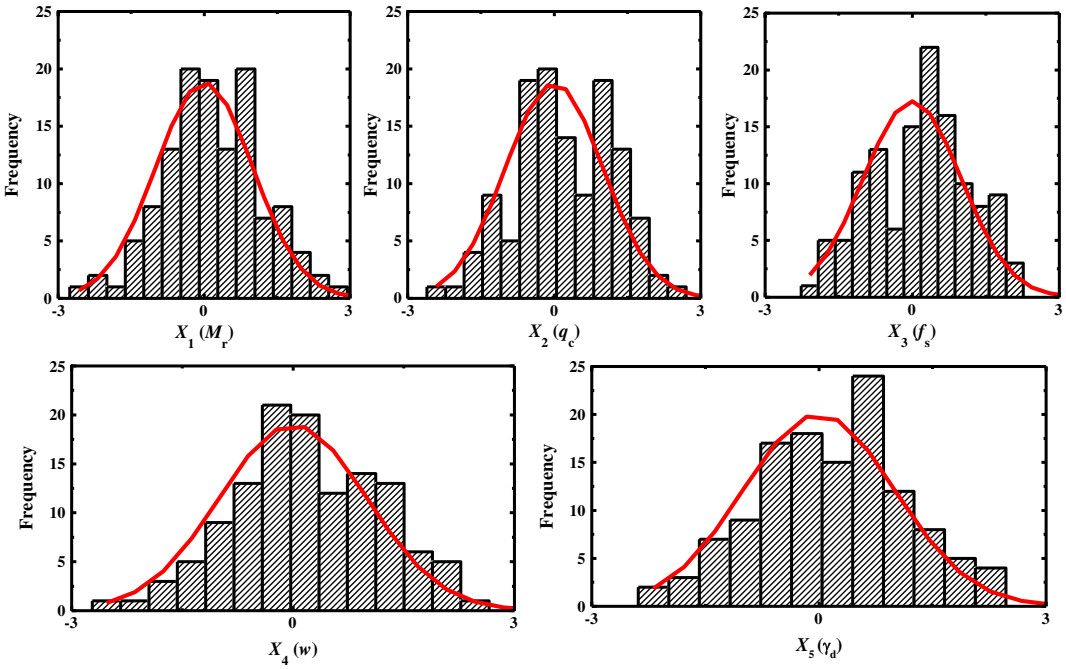

Fig. 4. Histograms of the transformed indices.

The linear correlation between all possible pairs of X_i and X_j data is determined using the product-moment (Pearson) correlations coefficient (δ_{ij}). Estimation of the product-moment correlation coefficients can be simply achieved in either MATLAB or EXCEL software and hence is not presented here. Table 4 lists the estimate of δ_{ij} between X_i and X_j. Since the linear correlation between X_i and X_j is the same as that between X_j and X_i (i.e., $\delta_{ij} = \delta_{ji}$), the number of correlation coefficients is actually $5 \times 4/2 = 10$. Table 4 shows that among the four laboratory and field parameters the q_c (X_2) provides most of the information for the prediction of M_r (X_1) since the correlation coefficient $\delta_{12} = 0.79$ is the highest after transformation. However, the other parameters such as f_s, w, and γ_d also provide valuable information for the prediction of M_r as the correlation coefficients between M_r and these parameters are still large. These findings provide the supportive evidence to incorporate more testing indices in establishing the correlations among M_r and other parameters.

The uncertainties associated with the correlation coefficients listed in Table 4 are not addressed. It is necessary to study the impact of the sample volume on the correlation matrix shown in Table 4. The bootstrapping technique (Ching and Phoon, 2014) is applied to obtain 4000 samples of the δ_{ij} estimates as follows:

1. Each Y_i variable is individually converted to X_i using the Box-Cox transformation with parameters listed in Table 3;
2. Draw n sets of $\{X_1, X_2, \ldots, X_5\}$ samples from the total dataset and estimate δ_{ij} between X_i and X_j. The n is determined to be 30, 40, …, 120;
3. Repeat above procedure for 4000 times and then obtain the mean and 95% confidence interval (CI) of δ_{ij}. The 95% CI describes the interval within which the 95% observations of a variable may fall. The lower and upper bounds of the 95% CI are the 0.025 and 0.975 percentile of the variable, respectively. Estimation of the 95% CI can be achieved using the *percentile* function in EXCEL.

Fig. 7 presents the mean and 95% CI of δ_{ij} in the 4000 realizations at different sample volumes. It is evident that the mean δ_{ij} almost remains constant and approximates the value listed in Table 4. For the same two variables, the maximum difference of mean δ_{ij} values at different sample volumes is consistently less than 0.01, as shown in Fig. 7(a). However, this does not indicate that the mean δ_{ij} can perfectly describe the correlation between X_i and X_j. The 95% CI describes the scatter of the estimates of δ_{ij} and can be used to evaluate the uncertainties associated with the mean value. Fig. 7(b) and (c) show that the impact of the sample volume on the 95% CI of δ_{ij} is quite evident. The 95% CI of δ_{ij} is large when the sample volume is small, but decreases as the sample volume increases. For instance, the 95% CI of δ_{12} is [0.64, 0.88] when the sample volume is 30. But this interval decreases to [0.77, 0.79] when the sample volume is 120. The estimate of mean δ_{12} may be robust at a large sample volume (e.g., 120) as its 95% CI is relatively small. However, for other correlations the estimates of mean δ_{ij} values may be less reliable. For instance, when the sample volume is 120, the 95% CI of δ_{34} is [−0.06, 0.12], which is a relatively wide range compared to the corresponding mean of −0.03.

Based on the sampling results, the mean δ_{13}, δ_{14}, and δ_{15} are estimated respectively to be [0.46, 0.52], [−0.72, −0.69], and [0.45, 0.50], when the sample volume is 120. These 95% CIs can be viewed to be small compared to the corresponding mean values. However, it is found that the correlation coefficients (δ_{ij}, $i > j > 1$) between other indices are more uncertain than δ_{1j} ($j > 1$). A preliminary finding from Fig. 7 is that the 95% CI of δ_{ij} may gradually decrease with the absolute value of the mean δ_{ij}. This may reflect the nature of the scatter within the pairwise data. For instance, if the data points in the scatter plot of X_i and X_j are well clustered about a linear line, it is confident to say that these two variables are strongly correlated and the absolute value of δ_{ij} should be large with small uncertainty. On the contrary, if the scatter of the data points in the plot of X_i and X_j is significant, it is expected that the estimated correlations should be small with large uncertainty. This observation implies that incorporating highly correlated variables is advisable in the construction of the multivariate normal distribution model. If a weak correlation does exist in two variables, it is suggested to increase the sample volume to reduce the statistical uncertainty of the model. However, the correlation matrix shown in Table 4 still provides the best unbiased estimates of δ_{ij} for the current data source. Therefore, this matrix is used in the construction of the multivariate normal distribution model for the Jiangsu clay.

The performance of the constructed multivariate normal distribution model should be evaluated and validated before the practical

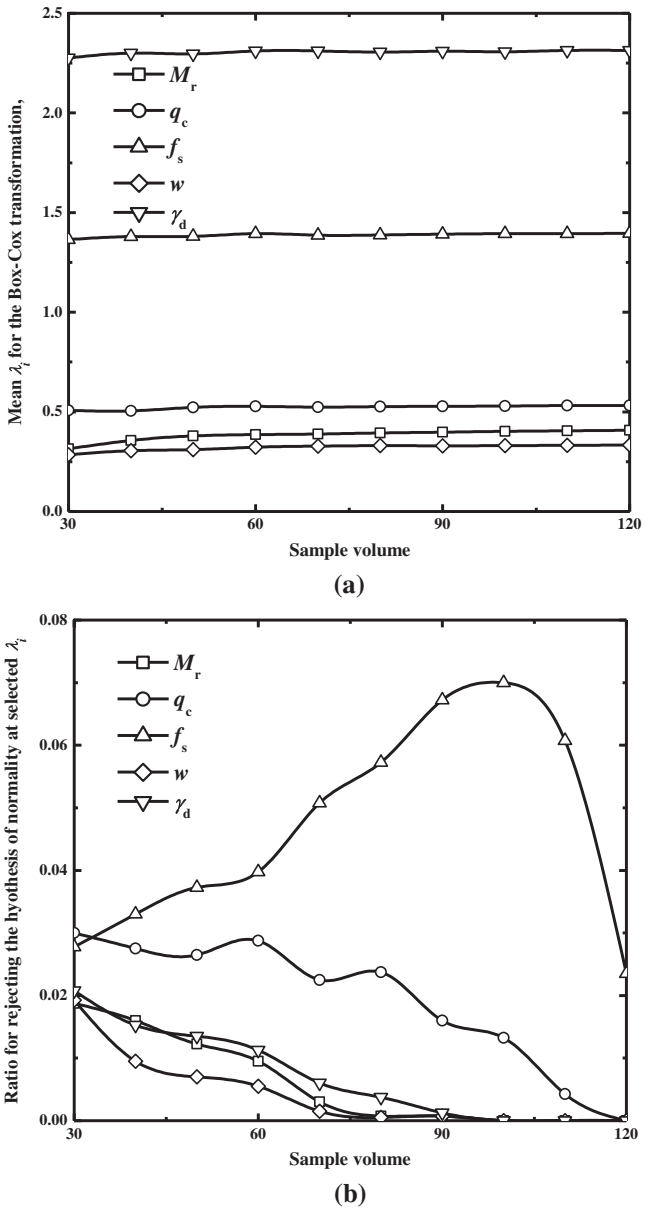

Fig. 5. Variation of the mean of λ and ratio of rejecting the hypothesis of normality with respect to the sample volume.

application. The validation can be performed using statistical tests and geotechnical comparison (Ching et al., 2014). Although several statistical tests to check the multivariate normal distribution were proposed in the literature, previous research revealed that meaningful results could still be obtained for the geotechnical application even if the data do not statistically strictly follow the multivariate normal distribution (Ching et al., 2014). Hence, in this study the validation of statistical features of the constructed multivariate normal distribution model is not conducted. On the contrary, a comparison between the correlations derived from the multivariate normal model and data published in the literature is performed to evaluate whether the model can do justice to the geotechnical practice. Although the model proposed in this study is developed on the basis of a local dataset, it is expected that the correlations derived from the model should at least show similar behavior (e.g. trend and range) to those in the literature. Moreover, it is of interest to investigate whether the constructed model can be extended for possible application in the global sites. Therefore, another multivariate dataset presented by Mohammad et al. (2007) is utilized for the validation.

4. Analysis of multivariate correlations

In this section, the multivariate correlations among the five soil parameters are analyzed using a Bayesian updating technique combined with the back transformation. Then these correlations are validated using the data collected from the literature. To achieve this analysis, the posterior statistics including the mean, COV, median, and 95% CI of Y_i conditional on any other variable are derived analytically.

4.1. Posterior statistics of Y_i

Conditional on any information of $\{X_j, X_k, \ldots, X_s\}$, the updated X_i is still a normal distribution with posterior mean (μ_i) and standard deviation (σ_i) as follows (Ching et al., 2014):

$$\mu_i = E[X_i|X_j, X_k, \ldots, X_s] = \begin{bmatrix}\delta_{ij} & \delta_{ik} & \cdots & \delta_{is}\end{bmatrix}\begin{bmatrix}1 & \delta_{jk} & \cdots & \delta_{js}\\ \delta_{kj} & 1 & \cdots & \delta_{km}\\ \vdots & \vdots & \ddots & \vdots\\ \delta_{mj} & \delta_{mk} & \cdots & 1\end{bmatrix} \tag{3a}$$

$$\sigma_i^2 = \mathrm{VAR}[X_i|X_j, X_k, \ldots, X_s] = 1 - \begin{bmatrix}\delta_{ij} & \delta_{ik} & \cdots & \delta_{is}\end{bmatrix}\begin{bmatrix}1 & \delta_{jk} & \cdots & \delta_{js}\\ \delta_{kj} & 1 & \cdots & \delta_{km}\\ \vdots & \vdots & \ddots & \vdots\\ \delta_{mj} & \delta_{mk} & \cdots & 1\end{bmatrix}^{-1}\begin{bmatrix}\delta_{ij}\\ \delta_{ik}\\ \vdots\\ \delta_{is}\end{bmatrix} \tag{3b}$$

where δ_{ij} is the (i, j) entry of the correlation matrix listed in Table 4, E[] and VAR[] are the expectation and variance operator, respectively.

Since the Box-Cox transformation is monotonic, the order of the observations for each variable is invariant after the transformation and back transformation. Hence the η percentile of Y_i can be estimated using:

$$Y_{i\eta} = \begin{cases}(\lambda_i b_i X_{i\eta} + \lambda_i a_i + 1)^{1/\lambda_i} & \lambda_i \neq 0\\ \exp(b_i X_{i\eta} + a_i) & \lambda_i = 0\end{cases} \tag{4}$$

where $Y_{i\eta}$ and $X_{i\eta}$ are the η percentile values of Y_i and X_i, respectively; $X_{i\eta}$ is obtained according to the posterior mean (μ_i) and standard deviation (σ_i) of updated X_i. Estimation of $X_{i\eta}$ can be achieved using the *norminv* function in either MATLAB or EXCEL. In geotechnical probabilistic analysis, it is of interest to evaluate the mean, COV, median, and 95% CI of a design parameter. The median and 95% CI can be easily derived using the Eq. (4). The η values are selected as 0.025, 0.500, 0.975, which correspond to the lower bound of the 95% CI, median, and upper bound of the 95% CI, respectively.

When $\lambda_i \neq 0$, the formulas for the posterior mean and COV of Y_i can be hardly derived strictly as its posterior marginal probability density distribution function is not accessible using the Box-Cox transformation. Instead, an alternative approach using the Taylor-series expansion is proposed to approximate the posterior mean and COV values. When $\lambda_i \neq 0$, the back transformation of $(Y_i)^l$ can be written as follows:

$$(Y_i)^l = g(X_i) = (\lambda_i b_i X_i + \lambda_i a_i + 1)^{l/\lambda_i} \tag{5}$$

where l is an integer for estimating the posterior lth-order moment of Y_i.

Using the second-order Taylor-series expansion, the expected value of $(Y_i)^l$ after the back transformation can be estimated as follows (Benjamin and Cornell, 1970):

$$E\left[(Y_i)^l\right] = g(\mu_i) + g'(\mu_i)E[X_i - \mu_i] + \frac{1}{2}g''(\mu_i)E\left[(X_i - \mu_i)^2\right] + R_2(X_i) \approx g(\mu_i) + \frac{\sigma_i^2}{2}g''(\mu_i) \tag{6}$$

where $g'(\mu_i)$ and $g''(\mu_i)$ are the first and second derivative functions of $(Y_i)^l$ at the coordinate of $X_i = \mu_i$, respectively, and $R_2(X_i)$ is the second-

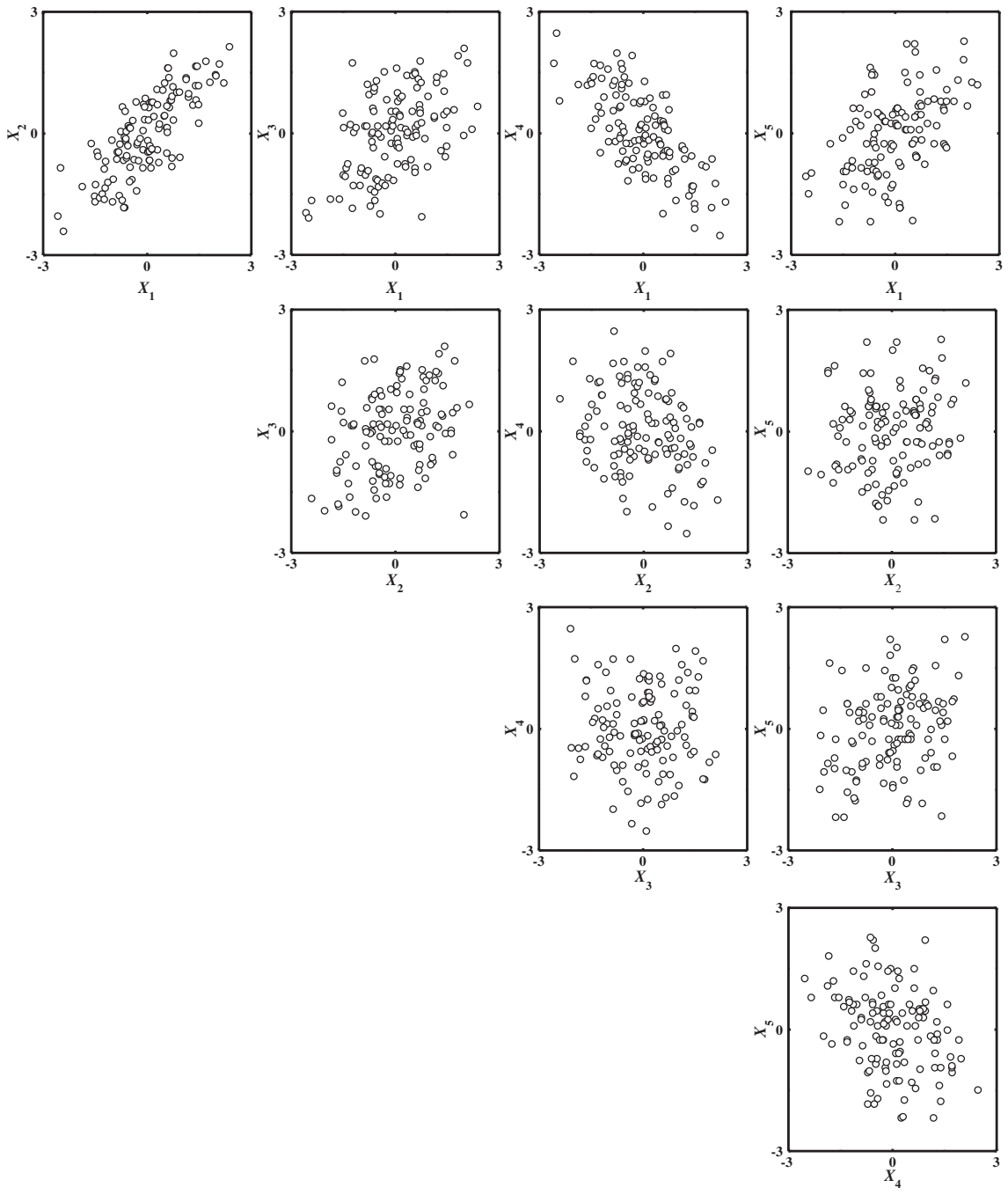

Fig. 6. Examination of linear correlation between X_i and X_j.

order residual of the Taylor-series expansion. According to Eq. (5), the second derivative function of $(Y_i)^l$ is determined as:

$$g''(X_i) = l(l-\lambda_i)b_i^2(\lambda_i b_i X_i + \lambda_i a_i + 1)^{\frac{l-2\lambda_i}{\lambda_i}} \tag{7}$$

Table 4
Product-moment (Pearson) correlation coefficients among $\{X_1, X_2, \ldots, X_5\}$.

	X_1	X_2	X_3	X_4	X_5
X_1	$\delta_{11} = 1.00$	$\delta_{12} = 0.78$	$\delta_{13} = 0.49$	$\delta_{15} = -0.71$	$\delta_{15} = 0.47$
X_2	$\delta_{21} = 0.78$	$\delta_{22} = 1.00$	$\delta_{23} = 0.34$	$\delta_{25} = -0.27$	$\delta_{25} = 0.13$
X_3	$\delta_{31} = 0.49$	$\delta_{32} = 0.34$	$\delta_{33} = 1.00$	$\delta_{35} = -0.03$	$\delta_{35} = 0.27$
X_4	$\delta_{41} = -0.71$	$\delta_{42} = -0.27$	$\delta_{43} = -0.03$	$\delta_{44} = 1.00$	$\delta_{45} = -0.32$
X_5	$\delta_{51} = 0.47$	$\delta_{52} = 0.13$	$\delta_{53} = 0.27$	$\delta_{45} = -0.32$	$\delta_{55} = 1.00$

Submitting Eqs. (5) and (7) into Eq. (6), the expected value of $(Y_i)^l$ can be approximated using the following function:

$$E\left[(Y_i)^l\right] \approx (\lambda_i b_i \mu_i + \lambda_i a_i + 1)^{\frac{l}{\lambda_i}}\left[1 + \frac{l b_i^2 \sigma_i^2 (l-\lambda_i)}{2(\lambda_i b_i \mu_i + \lambda_i a_i + 1)^2}\right] \quad \lambda_i \neq 0 \tag{8}$$

Therefore, the posterior mean of Y_i can be obtained using $l = 1$ as follows:

$$E[Y_i] \approx (\lambda_i b_i \mu_i + \lambda_i a_i + 1)^{\frac{1}{\lambda_i}}\left[1 + \frac{b_i^2 \sigma_i^2 (1-\lambda_i)}{2(\lambda_i b_i \mu_i + \lambda_i a_i + 1)^2}\right] \quad \lambda_i \neq 0 \tag{9}$$

The second moment of Y_i can be determined using $l = 2$ in Eq. (8). Then the posterior variance of Y_i can be determined using the method

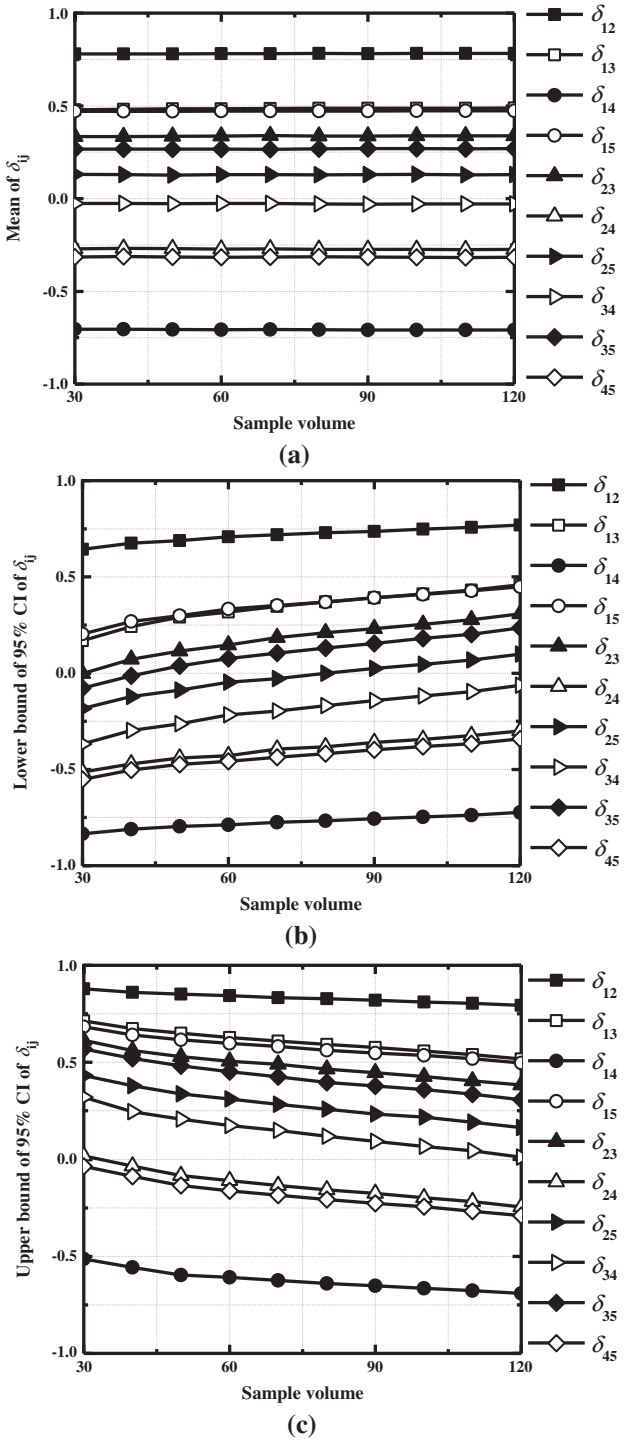

Fig. 7. Mean and 95% confidence interval (CI) of δ_{ij} in the 4000 realizations at different sample volumes for: (a) mean value; (b) lower bound of 95% CI; (c) Upper bound of 95% CI.

of moments as follows:

$$\mathrm{VAR}[Y_i] = E\left[(Y_i)^2\right] - (E[Y_i])^2 \approx b_i^2\sigma_i^2(\lambda_i b_i\mu_i + \lambda_i a_i + 1)^{\frac{2}{\lambda_i}-2}\left[1 - \frac{b_i^2\sigma_i^2(1-\lambda_i)^2}{4(\lambda_i b_i\mu_i + \lambda_i a_i + 1)^4}\right] \quad \lambda_i \neq 0 \tag{10}$$

Hence, combining Eqs. (9) and (10), the posterior COV of Y_i can be estimated using the following formula:

$$\mathrm{COV}[Y_i] = \frac{\sqrt{\mathrm{VAR}[Y_i]}}{E[Y_i]} \approx \frac{b_i\sigma_i\sqrt{4(\lambda_i b_i\mu_i + \lambda_i a_i + 1)^4 - b_i^2\sigma_i^2(1-\lambda_i)^2}}{2(\lambda_i b_i\mu_i + \lambda_i a_i + 1)^3 + b_i^2\sigma_i^2(1-\lambda_i)(\lambda_i b_i\mu_i + \lambda_i a_i + 1)} \quad \lambda_i \neq 0 \tag{11}$$

where COV[] is the COV of the variable in the bracket, λ_i, a_i, and b_i are the transformation parameters listed in Table 3, and μ_i and σ_i are the posterior mean and standard deviation of X_i, respectively.

Eqs. (9) and (11) respectively provide the approximations of the posterior mean and COV of Y_i when $\lambda_i \neq 0$. The accuracy of these two formulas can be simply checked using the parameters in Table 2 and Table 3. For instance, the prior mean and standard deviation of the unconditional X_1 are zero ($\mu_1 = 0$) and unity ($\sigma_1 = 1$), respectively. Submitting the values of $\{\lambda_1, a_1, b_1\}$ shown in Table 3 to the Eqs. (9) and (11), the prior mean and COV of the unconditional M_r are estimated to be 46.12 and 0.369, respectively. These two estimates are quite close to the actual statistics (mean of 46.13 and COV of 0.37) of M_r in the calibration data, as shown in Table 2. Therefore, Eqs. (9) and (11) should be accurate enough to estimate the posterior mean and COV of Y_i. However, caution should be taken when Eq. (6) or (8) is applied for estimating the high-order moment of Y_i as the residual of the Taylor-series expansion may have a remarkable impact on the estimates.

When $\lambda_i = 0$, the posterior marginal probability density function of $\ln(Y_i)$ should be a normal distribution with mean of $(b_i\mu_i + a_i)$ and standard deviation of $b_i\sigma_i$. Then the posterior marginal distribution of Y_i is a lognormal distribution. The posterior mean, variance and COV of Y_i can be obtained using the following formulas when $\lambda_i = 0$ (Low and Tang, 1997; Ching et al., 2012):

$$E(Y_i) = \exp\left(b_i\mu_i + a_i + b_i^2\sigma_i^2/2\right) \quad \lambda_i = 0 \tag{12a}$$

$$\mathrm{VAR}[Y_i] = \exp\left(2b_i\mu_i + 2a_i + b_i^2\sigma_i^2\right)\left[\exp\left(b_i^2\sigma_i^2\right) - 1\right] \quad \lambda_i = 0 \tag{12b}$$

$$\mathrm{cov}[Y_i] = \sqrt{\exp\left(b_i^2\sigma_i^2\right) - 1} \quad \lambda_i = 0 \tag{12c}$$

Using the Eqs. (1), (3a), (3b), (4), (9), (11), (12a), (12b) and (12c) the posterior mean, COV, median and 95% CI of M_r conditional on other testing indices can be estimated. Moreover, datasets containing $\{M_r, q_c\}$ $\{M_r, f_s\}$, $\{M_r, w\}$, and $\{M_r, \gamma_d\}$ are collected from the literature to validate the bivariate and multivariate correlations. The data collected from the literature are independent of the database which is used in the construction of the multivariate normal distribution model and hence can be utilized to check the pertinence of the estimated pairwise correlations. The possibility of extending the constructed model to a global dataset is also discussed using the data from the literature in the subsequent sections. For convenience, the posterior mean, median, and 95% CI of M_r will be discussed together, whereas the posterior COVs of M_r conditional on different parameters will be analyzed in an individual subsection.

4.2. Posterior mean, median and 95% CI of M_r conditional on CPTU data

Fig. 8 presents the estimated correlations among M_r and CPTU data. The calibration database and the data of $\{M_r, q_c, f_s\}$ presented by Mohammad et al. (2007) are also shown in Fig. 8. Given any observation of q_c or f_s, the mean, median and 95% CI of M_r can be obtained by using the Bayesian updating technique combined with the back transformation. Both Eqs. (9) and (12a) show that the posterior mean of M_r not only depends on the posterior mean (μ_1) of X_1 variable, but also is a function of the posterior standard deviation (σ_1) of X_1. Therefore, the explicit expression of the posterior mean of M_r is complicated because the correlation matrix is involved. Instead, the bivariate correlations among M_r, q_c and f_s based on the median ($\eta = 0.5$) are estimated simply as:

$$M_{r,\mathrm{median}} = \left(1.64q_c^{0.53} + 2.58\right)^{2.44} \tag{13a}$$

$$M_{r,\mathrm{median}} = \left(26.11f_s^{1.40} + 3.83\right)^{2.44} \tag{13b}$$

where $M_{r,median}$ is the posterior median of M_r, and $M_{r,median}$, q_c and f_s are all in MPa.

Fig. 8 shows that the medians of the M_r–q_c and M_r–f_s correlations are suitable for the calibration database. The M_r tends to increase with q_c and f_s. This observation is understandable as the q_c and f_s represent the strength and stiffness of the subgrade soil, which should be positively correlated to M_r. The good-of-fitness of the mean and median curve to the calibration database can be quantified using the squared product-moment correlation coefficients (ρ^2) between the predictions and actual observations of M_r. For the mean curves shown in Fig. 8 the ρ^2_{mean} values are estimated to be 0.62 and 0.24 when M_r is conditional on q_c and f_s, respectively. At the meantime, the ρ^2_{median} value of the posterior median curve almost equals to the corresponding ρ^2_{mean} value of the posterior mean curve. Hence, q_c should provide better information for updating M_r in the calibration database.

It can be also observed from Fig. 8 that the posterior mean of M_r (solid curve) is consistently larger than the posterior mean (dash curve) under the same condition. This does not come as a surprise since the histogram of the M_r is right skewed as shown in Fig. 3. The relatively longer tail on the right side in the histogram of M_r contributes to a higher mean value than the median value. However, the difference between the posterior mean and posterior median is not remarkable. On the whole, the ratio of the posterior mean to median ($M_{r,mean}/M_{r,median}$) is approximately 1.01 when the M_r is conditional on q_c. Similar feature can be also found in the M_r–f_s correlation, where the ratio of the posterior mean to posterior median is approximately 1.02. Therefore, the Eqs. (13a) and (13b) which is derived using the median can be applied to conveniently approximate the posterior mean of M_r for the Jiangsu clay database.

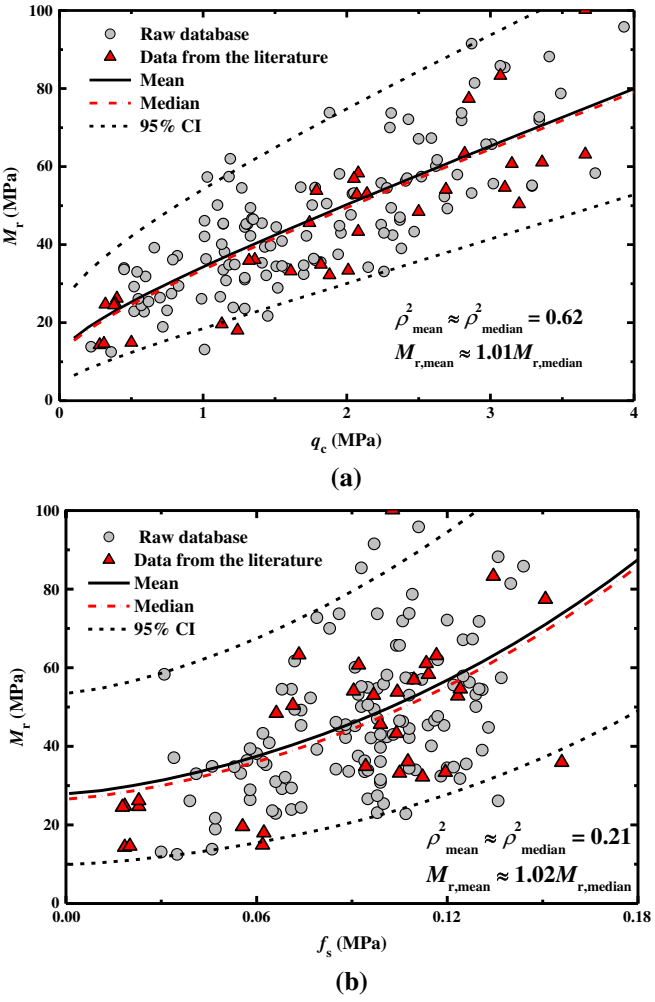

Fig. 8. Comparison between posterior mean, median, 95% CI, actual observations, and data from the literature for: (a) M_r–q_c correlation; (b) M_r–f_s correlation.

Besides, Fig. 8 also shows that the data obtained from literature are almost all within the 95% CIs. The ranges and trends of the correlations derived from the constructed multivariate normal distribution model are comparable to those of the data obtained from the literature. Therefore, the constructed multivariate normal distribution model appears to be capable of the bivariate correlations among M_r, q_c and f_s in the global dataset.

4.3. Posterior mean, median, and 95% CI of M_r conditional on laboratory data

Fig. 9 presents the posterior mean, median, and 95% CI of M_r conditional on the laboratory indices. The calibration database along with the data of {M_r, w, γ_d} presented by Mohammad et al. (2000, 2007) are also presented in Fig. 9. It can be observed that the M_r increases with γ_d, but decreases with w. The bivariate correlations among M_r, w and γ_d based on the median ($\eta = 0.5$) are determined as:

$$M_{r,median} = \left(-1.07w^{0.34} + 8.12\right)^{2.44} \quad (14a)$$

$$M_{r,median} = \left(0.0019\gamma_d^{2.33} + 3.51\right)^{2.44} \quad (14b)$$

where $M_{r,median}$ is in MPa, w is expressed in percent (%), and γ_d is in kN/m^3.

Fig. 9 shows that the posterior mean and posterior median of M_r conditional on either w or γ_d are also close, although the posterior mean values are consistently larger than the posterior median. The ratios of $M_{r,mean}/M_{r,median}$ approximate 1.01 and 1.04 when M_r is conditional on w and γ_d, respectively. These small ratios indicate that the Eqs. (14a) and (14b) can be applied for simple and convenient estimation of the posterior mean of M_r for the Jiangsu clay. For the posterior mean and median curves shown in Fig. 9, the ρ^2_{mean} and ρ^2_{median} values are estimated to be 0.52 for M_r conditional on w and 0.23 for M_r conditional on γ_d. Therefore, it is concluded that the w can provide more information for updating the M_r than the γ_d.

Fig. 9 also shows that the data obtained from the literature are consistent with the 95% CIs developed from the constructed multivariate normal distribution model. The features (i.e., ranges and trends) of the correlations derived from the multivariate normal distribution model are comparable to those of the data obtained from the literature, although the data collected from the literature seem to be more dispersive than the former. Hence the multivariate model appears to be capable of the bivariate correlations among M_r, w and γ_d in the global database.

4.4. Posterior mean, median, and 95% CI of M_r conditional on multiple variables

In engineering practice generally more than one laboratory or in-situ testing index is available. It is important to predict the value of M_r using multiple indices, as the uncertainties associated with the predictions can be further reduced by involving more indices in the framework of the multivariate normal distribution model. The following correlation based on the median ($\eta = 0.5$) can be derived for the case in which only field CPTU data are available:

$$M_{r,median} = \left(1.46q_c^{0.53} + 13.55f_s^{1.4} + 2.36\right)^{2.44} \quad (15a)$$

When only laboratory data are available in practice, the following formula based on the median ($\eta = 0.5$) can be applied to estimate M_r:

$$M_{r,median} = \left(-0.94w^{0.34} + 0.0011\gamma_d^{2.33} + 7.00\right)^{2.44} \quad (15b)$$

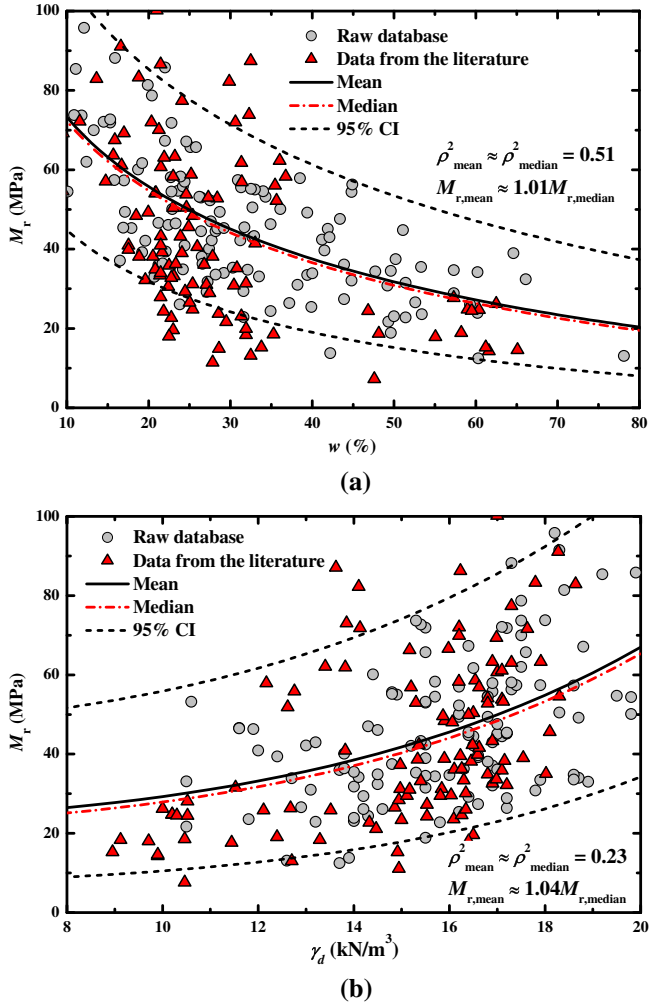

Fig. 9. Comparison between posterior mean, median, 95% CI, actual observations, and data from the literature for: (a) M_r–w correlation; (b) M_r–γ_d correlation.

When all four indices $\{q_c, f_s, w, \gamma_d\}$ are available, the following formula based on the median ($\eta = 0.5$) can be utilized:

$$M_{r,median} = \left(1.13q_c^{0.53} + 13.06f_s^{1.4} - 0.75w^{0.34} + 0.0007\gamma_d^{2.33} + 4.75\right)^{2.44} \quad (15c)$$

where $M_{r,median}$, q_c, and f_s are in MPa, w is in percent (%), and γ_d is in kN/m^3.

The performance of above three functions in updating the posterior median of M_r can be evaluated via comparing the predictions with the actual measurements. Fig. 10 illustrates the comparisons between the predicted posterior mean and median of M_r conditional on $\{q_c, f_s\}$, $\{w, \gamma_d\}$, and $\{q_c, f_s, w, \gamma_d\}$ and the measured M_r in the calibration database. The predicted mean and median values are scattered around the reference lines in Fig. 10. Moreover, the posterior mean and median of M_r are very close, indicating that Eqs. (15a), (15b), and (15c) can be applied for simply and conveniently estimating the posterior mean of M_r for the Jiangsu clay. The ρ^2 value between the posterior mean (and median) and the actual measurements within the calibration database are estimated to be 0.68, 0.58, and 0.98, respectively. It is evident that the accuracy of formulas can be consistently improved by involving more indices, especially when all four input parameters, $\{q_c, f_s, w, \gamma_d\}$, are used for prediction.

It is also found that the accuracies of the formulas for M_r are not significantly improved if the input variables (e.g., q_c and f_s in Eq. 15a) are well correlated. This does not come as a surprise since the predictions of M_r depend on the correlations among all the input variables rather than the pairwise correlations. Considering the extreme case that f_s is perfectly correlated to q_c (i.e., $\delta_{23} = 1$), only one of these two parameter is sufficient for predicting M_r. Hence in establishing multivariate normal models for design parameters in engineering practice, it is suggested to involve more weakly correlated input variables so that the accuracy can be further improved. A combination of the mechanical properties (e.g., q_c) and physical parameters (e.g., w) seems to be a good selection to construct the multivariate normal distribution.

Above comparison illustrates one benefit of the multivariate correlation analysis that the uncertainties associated with M_r can be reduced by incorporating multiple argument variables. However, it does not indicate that the constructed model can be utilized to provide pertinent predictions at a new target site. To ensure meaningful validation, a database completely independent of the calibration database should be used. In this study, the multivariate dataset collected from the literature is applied to achieve the validation (Mohammad et al., 2000). Fig. 11 presents the comparisons between the posterior mean and median of M_r conditional on $\{q_c, f_s\}$, $\{w, \gamma_d\}$, and $\{q_c, f_s, w, \gamma_d\}$ and measured M_r of the dataset from the literature. The trends and ranges of the multivariate correlations developed in this study are consistent with those of the data collected from the literature. However, the posterior mean and median of M_r are still close but may be potentially biased with respect to the measurements. Conditional on $\{q_c, f_s\}$ and $\{q_c, f_s, w, \gamma_d\}$, most data points are evidently above the reference line. A linear line with zero intercept is fitted to the scatter plot of the posterior mean against the measurements for each conditional analysis in Fig. 11. The slope of the straight line evaluates the biases of the predictions with respect to the measured values. It is found that the slopes of the line are 1.03, 0.95, and 1.10 when M_r is conditional on $\{q_c, f_s\}$, $\{w, \gamma_d\}$, and $\{q_c, f_s, w, \gamma_d\}$ in the dataset from the literature, respectively.

Another observation from Fig. 11 is that if more testing indices are involved to predict the posterior mean or median of M_r of the dataset from the literature, the biases seem to be increasing. This potentially indicates that when applied in other datasets, the accuracy of the multivariate correlations derived from the multivariate normal distribution model in this study may decrease if more testing indices are used. A potential explanation is that the constructed model is insufficient to capture the multivariate correlations among the variables, although it is capable of the bivariate correlations between M_r and other testing indices. More data are still necessary to improve the performance of the constructed multivariate model.

4.5. Posterior COV of M_r under different conditions

The uncertainties associated with the posterior mean of M_r are not addressed in above analysis. The posterior COV of Y_i is a major statistics for probabilistic analysis in geotechnical engineering. The posterior COVs of M_r conditional on different testing indices can be estimated using Eqs. (11), (12a), (12b) and (12c). The performance of the testing indices on reducing the uncertainties associated with the posterior mean of M_r is evaluated in this subsection. Fig. 12 presents the posterior COV of M_r conditional on different testing indices against the corresponding mean value. It can be seen that if only bivariate correlations are considered, the posterior COVs of M_r conditional on q_c are consistently smaller than those conditional on other parameters, followed by the moisture (w). Hence, q_c is the most effective parameter to update M_r, whereas the f_s and γ_d provide comparable but the least information for reducing the posterior COV of M_r.

Fig. 12 also shows that if more testing indices are involved in updating the M_r using the multivariate normal distribution model, then the posterior COV of M_r can be further reduced. For instance, when M_r is only conditional on q_c, the posterior COV gradually reduces from 0.32 to 0.19 when the mean of M_r increases from 20 to 80 MPa. When M_r is conditional on $\{q_c, f_s\}$, the posterior COV of M_r decreases

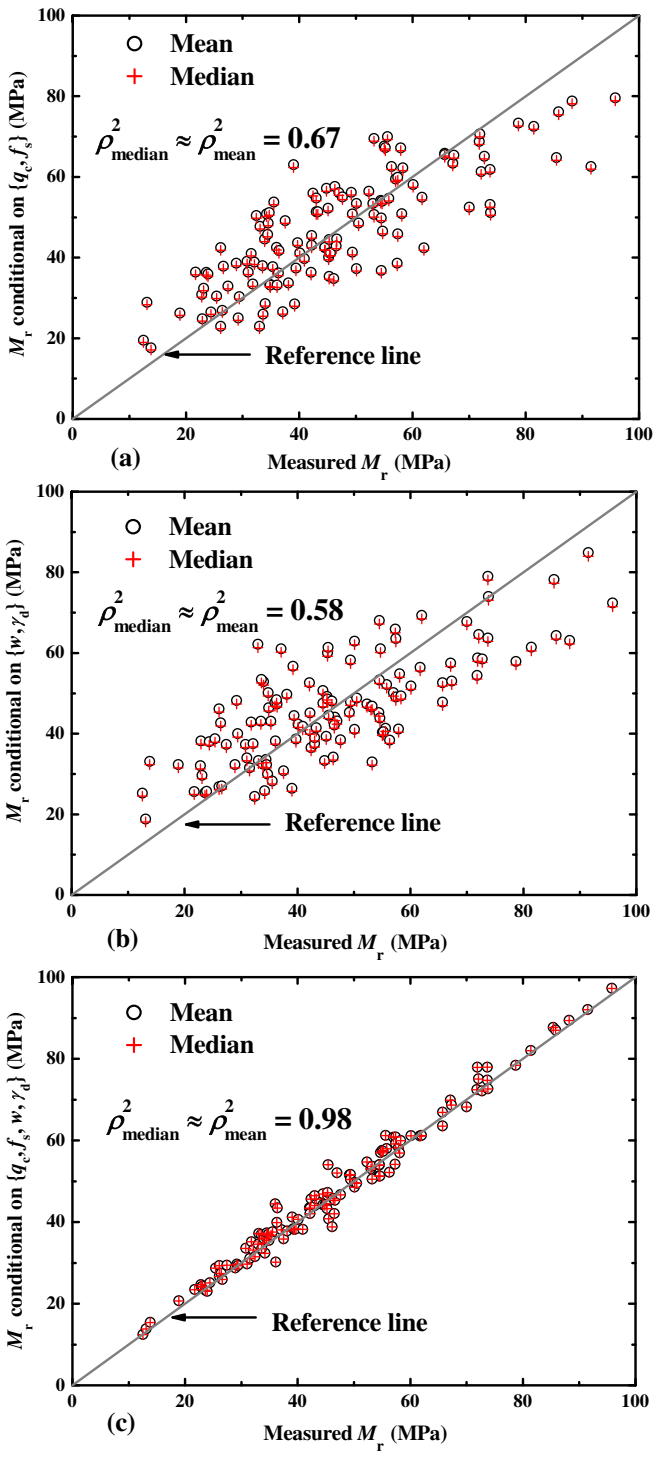

Fig. 10. Actual measurements of the calibration database against the posterior mean and median of M_r conditional on: (a) $\{q_c, f_s\}$; (b) $\{w, \gamma_d\}$; (c) $\{q_c, f_s, w, \gamma_d\}$.

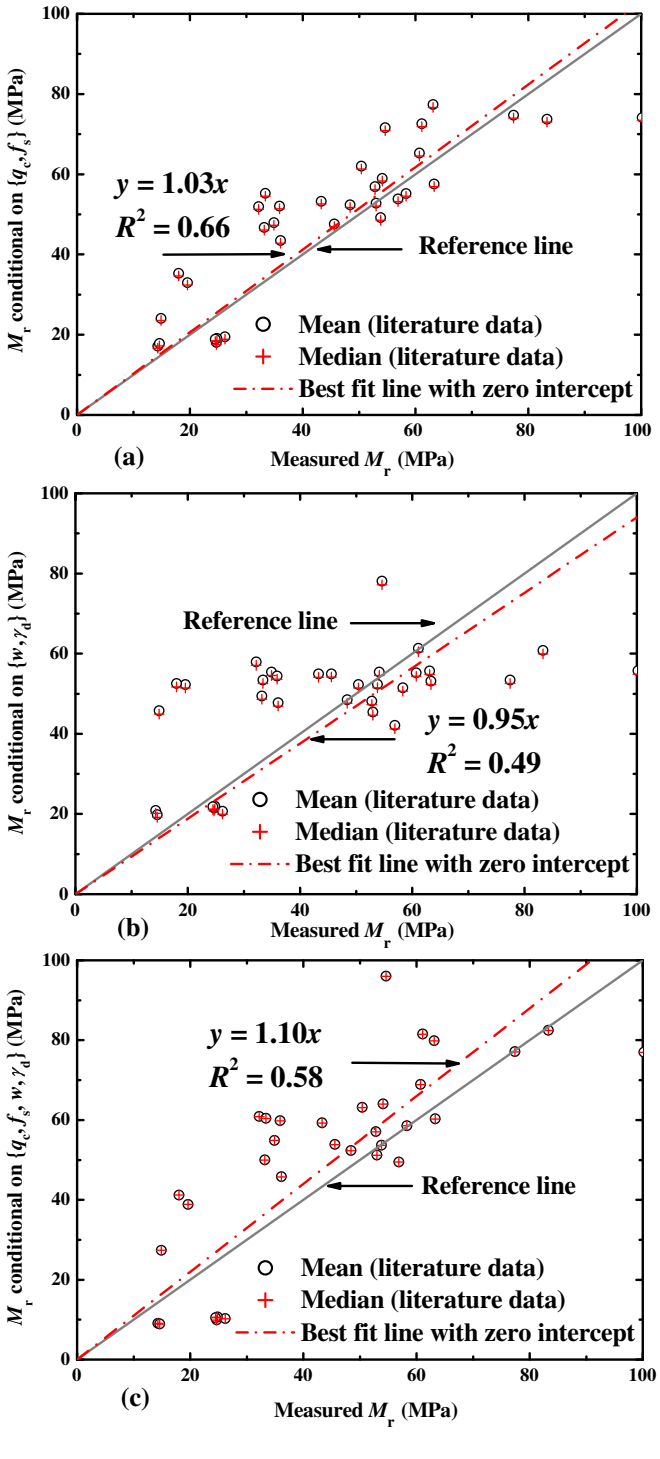

Fig. 11. Actual measurements of the validation database against the posterior mean and median of M_r conditional on: (a) $\{q_c, f_s\}$; (b) $\{w, \gamma_d\}$; (c) $\{q_c, f_s, w, \gamma_d\}$.

from 0.30 to 0.17 correspondingly. When all four testing indices of $\{q_c, f_s, w, \gamma_d\}$ are involved, the posterior COV of M_r decreases from 0.08 to 0.05 correspondingly. This observation demonstrates that the multivariate normal distribution model provides a sound approach to establish the multivariate correlations among design parameters and other testing indices.

5. Summary and conclusions

This paper investigates the feasibility of a multivariate normal distribution in modeling a regional database collected from the Quaternary clay of 16 sites with different geologic formations in Jiangsu province, China. This database contains 124 sets of resilient modulus (M_r) at the in-situ stress condition, piezocone penetration test (CPTU) indices including cone tip resistance (q_c) and sleeve frictional resistance (f_s), and laboratory indices including moisture (w) and dry density (γ_d).

To construct the multivariate normal model, all the geotechnical parameters were individually converted to standard normal variables using the Box-Cox method. A formal normality test combined with the plots of pairwise data shows that the multivariate normal distribution is acceptable for the transformed variables. Then the correlation matrix of the multivariate normal distribution model was obtained by

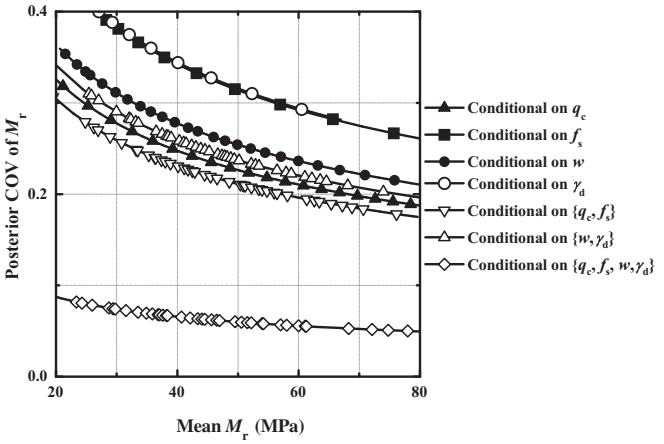

Fig. 12. Posterior COV of M_r conditional on different testing indices.

estimating the product-moment (Pearson) correlation coefficients between all possible pairs of the transformed data. The posterior mean and standard deviation values of the transformed variables were obtained using a Bayesian updating technique. An approximation for estimating the posterior mean, coefficient of variation (COV), median, and 95% confidence interval (CI) of a soil property after back transformation was derived using the second-order Taylor-series expansion. Then the multivariate formulas for predicting M_r using different CPTU testing indices and laboratory properties of the clayey soils were derived. The comparisons between the posterior mean values, medians and 95% CIs of M_r and the measurements of the calibration database demonstrated that the multivariate normal model is capable of the correlations within $\{M_r, q_c, f_s, w, \gamma_d\}$ for the Jiangsu clay. Moreover, if more indices are involved in the multivariate normal model the accuracy of predictions can be consistently improved.

The comparisons between the estimates of M_r and data collected from the literature showed that the constructed model may also be capable of the bivariate correlations between M_r and other variables in the global dataset. But it is potentially unable to capture the multivariate correlations among all five indices in the global dataset because the estimates of M_r seem to be slightly biased, especially when more testing indices are involved. Caution should be taken when the model is used for the application in the global sites. More data should be included to construct a robust multivariate normal distribution model for a global database.

Acknowledgements

Majority of the work presented in this paper was funded by the Foundation for the New Century Excellent Talents of China (Grant No. NCET-13-0118), the Foundation of Jiangsu Province Outstanding Youth (Grant No. BK20140027), the Foundation for the Author of National Excellent Doctoral Dissertation of PR China (Grant No. 201353), The High Level Talent Project of Peak of Six Talents in Jiangsu Province (Grant No. 2015-ZBZZ-001), the Fundamental Research Funds for the Central Universities (Grant No. 2242016K41062) and Jiangsu Key Laboratory of Urban Underground Engineering & Environmental Safety (Southeast University) (Grant No. 2014-10). These supports are gratefully acknowledged.

References

American Association of State Highway and Transportation Officials (AASHTO), 1993f. Guide for Design of Pavement Structures. Washington, D. C.

American Association of State Highway and Transportation Officials (AASHTO), 2003f. Determining the resilient modulus of soils and aggregate materials, in AASHTO T 307-99. Washington, D. C, 2003.

ASTM D5778, 2012. International Standard Test Method for Electronic Friction Cone and Piezocone Penetration Testing of Soils, Annual Book of ASTM Standards. ASTM International, West Conshohocken, PA.

Benjamin, J.R., Cornell, C.A., 1970. Probability, Statistics, and Decision for Civil Engineers. McGraw-Hill, New York.

Bheemasetti, T.V., 2015. Spatial Variability Models and Prediction Analysis of Soil Properties Using Geostatistics. Doctoral Thesis. Univ. of Texas, Arlington, TX, p. 347.

Box, G.E.P., Cox, D.R., 1964. An analysis of transformations. J. R. Stat. Soc. Ser. B 26 (2), 211–252.

Cai, G.J., Liu, S.Y., Puppala, A.J., Tong, L.Y., 2011. Assessment of the coefficient of lateral earth pressure at rest (Ko) from in situ seismic tests. Geotech. Test. J. 34 (4), 310–320.

Cai, G.J., Liu, S.Y., Puppala, A.J., 2012. Reliability assessment of CPTU-based pile capacity predictions in soft clay deposits. Eng. Geol. 141, 84–91.

Cai, G.J., Puppala, A.J., Liu, S.Y., 2014. Characterization on the correlation between shear wave velocity and piezocone tip resistance of Jiangsu soft clays. Eng. Geol. 171, 96–103.

Charitidou, E., Fouskakis, D., Ntzoufras, L., 2015. Bayesian transformation family selection: moving toward a transformed Gaussian universe. Can. J. Stat. 43 (4), 600–623.

Ching, J., Phoon, K.K., Huang, W.-C., 2011. CONSTRUCTING Joint Distributions of Multivariate Geotechnical Data. Georisk 2011 - Geotechnical Risk Assessment and Management, Geotechnical Special Publications 224. ASCE, Reston, pp. 1141–1148.

Ching, J.Y., Phoon, K.K., 2012. Modeling parameters of structured clays as a multivariate normal distribution. Can. Geotech. J. 49 (5), 522–545.

Ching, J.Y., Chen, J.R., Yeh, J.Y., Phoon, K.K., 2012. Updating uncertainties in friction angles of clean sands. J. Geotech. Geoenviron. Eng. ASCE 138 (2), 217–229.

Ching, J., Phoon, K.K., 2013. Multivariate distribution for undrained shear strengths under various test procedures. Can. Geotech. J. 50 (9), 907–923.

Ching, J.Y., Phoon, K.K., 2014. Correlations among some clay parameters – the multivariate distribution. Can. Geotech. J. 51, 686–704.

Ching, J.Y., 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.

Dehler, W., Labuz, J., 2007. Cone penetration testing in pavement design. Report No. MN/RC 2007-36, Minnesota Department of Transportation. University of Minnesota, Minneapolis.

Duncan, J.M., Buchignani, A.L., 1976. An Engineering Manual for Settlement Studies. University of California, Berkeley CA.

Gudishala, R., 2004. Development of Resilient Modulus Prediction Models for Base and Subgrade Pavement Layers from in Situ Devices Test Results. M.S. Thesis. Louisiana State University and Agricultural and Mechanical College Department of Civil Engineering.

Heukelom, W., Klomp, A.J.G., 1962. Dynamic testing as a means of controlling pavement during and after construction. Proceedings of 1st International Conference on the Structural Design of Asphalt Pavement. University of Michigan, Ann Arbor, MI.

Hicks, R.G., Monismith, C.L., 1971. Factors influencing the resilient response of granular materials. Highway Research Record 345. Highway Research Board, Washington, DC, pp. 15–31.

ISSMFE, 1989. International reference test procedure for cone penetration test (CPT). Report of the ISSMFE Technical Committee on Penetration Testing of Soils – TC 16, with Reference to Test Procedures 7. Swedish Geotechnical Institute, Linköping, Information, pp. 6–16.

Jaksa, M.B., 2007. Modeling the natural variability of over-consolidated clay in Adelaide, South Australia. Proceedings of the Second International Workshop on Characterisation and Engineering Properties of Natural Soils. Taylor and Francis, London, pp. 721–2751.

Johnston, J., 1984. Econometric Methods. McGraw-Hill, New York.

Kim, W., Labuz, J.E., 2007. Resilient modulus and strength of base course with recycled bituminous material. Final Report No. MN/RC-2007-05, Minnesota Department of Transportation. University of Minnesota, Minneapolis.

Kim, S.H., Yang, J., Jeong, J.H., 2014. Prediction of subgrade resilient modulus using artificial neural network. KSCE J. Civ. Eng. 18 (5), 1372–1379.

Liu, G., Niu, J., Zhang, C., Guo, G., 2015. Accuracy and uncertainty analysis of soil Bbf spatial distribution estimation at a coking plant-contaminated site based on normalization geostatistical technologies. Environ. Sci. Pollut. Res. 22, 20121–20130.

Liu, S.Y., Cai, G.J., Puppala, A.J., Tu, Q.Z., 2011. Prediction of embankment settlements over marine clay using piezocone penetration tests. Bull. Eng. Geol. Environ. 70 (3), 401–409.

Low, B.K., Tang, W.H., 1997. Efficient reliability evaluation using spreadsheets. J. Eng. Mech. ASCE 123 (7), 749–752.

Lunne, T., Robertson, P.K., Powell, J.J.M., 1997. Cone Penetration Testing in Geotechnical Practice. Blackie Academic and Professional, London.

MATLAB, 2013. [Computer software]. MathWorks, Natick, MA.

Mayne, P.W., 2007. Cone penetration testing: a synthesis of highway practice. NCHRP Synthesis 368. Transportation Research Board, National Academies Press, Washington, D.C.

Mohammad, L.N., Titi, H.H., Herath, A., 1999. Evaluation of Resilient Modulus of Subgrade Soil by Cone Penetration Test. Transportation Research Record No. 1652. pp. 236–245.

Mohammad, L.N., Titi, H.H., Herath, A., 2000. Investigation of the Applicability of Instruction Technology to Estimate the Resilient Modulus of Subgrade Soil. Publication FHWA/LA-00/332. Baton Rouge, Louisiana Transportation Research Center.

Mohammad, L.N., Titi, H.H., Herath, A., 2002. Effect of Moisture Content and Dry Unit Weight on the Resilient Modulus of Subgrade Soils Predicted by Cone Penetration Test. Publication FHWA-LA-00-355. U.S. Department of Transportation, FHWA.

Mohammad, L.N., Herath, A., Abu-Farsakh, M.Y., Gaspard, K., Gudishala, R., 2007. Prediction of resilient modulus of cohesive subgrade soils from dynamic cone penetrometer test parameters. J. Mater. Civ. Eng. 19 (11), 986–992.

Montgomery, D.C., Runger, G.C., Hubele, N.F., 2010. Engineering Statistics. John Wiley & Sons, New York.
National Cooperative Highway Research Program (NCHRP), 2004. Guide for Mechanistic-empirical Design of New and Rehabilitated Pavement Structures. Part 2, Design Inputs. Final Rep. No. NCHRP 1-37A, Washington, D.C..
Phoon, K.K., 2006. Modeling and simulation of stochastic data. GeoCongress 2006, 1–17.
Phoon, K.K., Ching, J., 2012. Beyond Coefficient of Variation for Statistical Characterization of Geotechnical Parameters. Keynote Lecture of Geotechnical and Geophysical Site Characterization 4, ISC 4. pp. 113–130.
Phoon, K.K., Ching, J., 2013. Multivariate Model for Soil Parameters Based on Johnson Distributions. Foundation Engineering in the Face of Uncertainty. pp. 337–353.
Puppala, A.J., Acar, Y.B., Tumay, M.T., 1995. Cone penetration in very weakly cemented sand. J. Geotech. Eng. 121 (8), 589–600.
Puppala, A., Bheemasetti, T., Zou, H.F., Pedarla, A., Yu, X.B., Cai, G.J., 2015. Spatial variability analysis of soil properties using Geostatistics. Chapter 8, Handbook of Research on Advanced Computational Techniques for Simulation-Based Engineering. IGI Global.
Razali, N.M., Wah, Y.B., 2011. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. J. Stat. Model. Anal. 2 (1), 21–33.
Sakia, R.M., 1992. The Box-Cox transformation technique: a review. Statistician 41, 169–1768.
Shapiro, S.S., Wilk, M.B., 1965. An analysis of variance test for normality. Biometrika 52, 591–611.