NRC Research Press

ARTICLE

Modeling piezocone cone penetration (CPTU) parameters of clays as a multivariate normal distribution

Jianye Ching, Kok-Kwang Phoon, and Chih-Hao Chen

Abstract: This study examines the possibility of modeling piezocone cone penetration (CPTU) cone tip resistance, excessive pore pressure behind the cone, undrained shear strength, and overconsolidation ratio of lightly overconsolidated clays as a multivariate normal distribution. This is part of a continuing study to develop a multivariate distribution that could be used to simulate common soil parameters at a clay site. This study compiles a large database consisting of 535 data points in which the CPTU parameters, undrained shear strength, and overconsolidation ratio are simultaneously measured in close proximity. A multivariate normal distribution is then used to capture the correlations between soil parameters of interest and to derive useful equations for Bayesian inference. This constructed multivariate normal distribution and equations are further validated by another independent database consisting of 594 data points as well as by empirical equations proposed in literature. The most useful outcome of this study is to provide a systematic and *analytical* method for updating the distributions of the normalized undrained shear strength and the overconsolidation ratio in the presence of CPTU parameters.

Key words: correlation, multivariate normal distribution, site characterization, piezocone cone penetration (CPTU) tests, reliability-based design.

Résumé : Cette étude examine la possibilité de modéliser la résistance de la pointe du cône d'un essai de pénétration au cône (CPTU), la pression interstitielle excessive derrière le cône, la résistance au cisaillement non drainé, et le coefficient de surconsolidation d'argiles légèrement surconsolidées en tant de distribution normale multi-variable. Ceci fait partie d'une étude en cours visant à développer une distribution multi-variable qui pourrait être utilisée pour simuler des paramètres communs des sols sur un site d'argile. Cette étude compile une vaste base de données comprenant 535 points où les paramètres CPTU, la résistance au cisaillement non drainé, et le coefficient de surconsolidation sont mesurés simultanément à des endroits rapprochés. Une distribution normale multi-variable est ensuite utilisée pour illustrer les corrélations entre les paramètres du sol d'intérêt et pour dériver des équations utiles pour les inférences bayésiennes. La distribution normale multi-variable ainsi construite et les équations sont validées avec une autre base de données indépendante contenant 594 points, de même qu'avec des équations empiriques proposées dans la littérature. Le résultat le plus pertinent de cette étude est de fournir une méthode systématique et analytique pour la mise à jour des distributions de la résistance au cisaillement non drainé normalisée et du coefficient de surconsolidation en présence des paramètres CPTU. [Traduit par la Rédaction]

Mots-clés : corrélation, distribution normale multi-variable, caractérisation de site, essais de pénétration au cône (CPTU), conception basée sur la fiabilité.

Introduction

Piezocone cone penetration (CPTU) tests have been popular for decades because of their ability to construct nearly continuous vertical soil profiles. Multivariate information is available in a typical CPTU sounding, including the pore pressure behind the cone, u_2, cone tip resistance, q_c (or the corrected cone tip resistance $q_t = q_c + (1 - a)u_2$, in which a is the area ratio of the cone), and sleeve friction, f_s. These test indices could be correlated to the undrained shear strength (s_u) and overconsolidation ratio (OCR) by pairwise correlations, e.g., s_u–q_t, s_u–u_2, and OCR–q_t correlations (Lunne et al. 1997). Although it is relatively easy to update the first two moments (mean and coefficient of variation) of s_u and OCR based on a single test index, e.g., q_t, it is less obvious how to do so in the presence of multiple test indices, e.g., q_t and u_2. An option is to discard all test indices, but keep the most relevant and (or) most accurate test index to update the first two moments of s_u or OCR. However, it is uneconomical to eliminate costly information in the updating process because the abandoned test indices may further reduce the coefficient of variation of s_u. In addition, it is good geotechnical practice to cross-validate interpretation of soil properties from different sources of information, given the significant assumptions and empiricism underlying most pairwise correlations. In fact, the geotechnical engineering literature is replete with such pairwise correlations.

To update the means and coefficients of variation of s_u and OCR conditioning on multiple CPTU test indices, the rigorous approach is to construct a multivariate probability distribution function from the multivariate information. This paper examines the feasibility of adopting the multivariate normal distribution to model the correlation structure among normalized undrained shear strength (s_u/σ'_v, where σ'_v is the vertical effective stress), overconsolidation ratio (OCR = σ'_p/σ'_v, where σ'_p is the preconsolidation stress), normalized cone tip resistance $(q_t - \sigma_v)/\sigma'_v$ (where σ_v is the vertical total stress), normalized effective cone tip resistance $(q_t - u_2)/\sigma'_v$, normalized excess pore pressure $(u_2 - u_0)/\sigma'_v$

Received 11 July 2012. Accepted 10 October 2013.

J. Ching and C.-H. Chen. Department of Civil Engineering, National Taiwan University, Taipei, Taiwan.
K.-K. Phoon. Department of Civil and Environmental Engineering, National University of Singapore, Singapore; National Taiwan University of Science and Technology, Taiwan.
Corresponding author: Jianye Ching (e-mail: jyching@gmail.com).

(where u_0 is the static pore pressure), and pore pressure ratio $B_q = (u_2 - u_0)/(q_t - \sigma_v)$.

A large database consisting of the following six dimensionless parameters simultaneously measured in close proximity in lightly overconsolidated clays is first compiled:

1. $Y_1 = s_u/\sigma'_v$
2. $Y_2 = \text{OCR}$
3. $Y_3 = (q_t - \sigma_v)/\sigma'_v$
4. $Y_4 = (q_t - u_2)/\sigma'_v$
5. $Y_5 = (u_2 - u_0)/\sigma'_v$
6. $Y_6 = B_q$

The database contains 535 sets of $\{Y_1, Y_2, \ldots, Y_6\}$, which can be viewed as realizations of a multivariate normal distribution. This database is labeled as Clay/6/535, based on the notation (soil type)/(number of parameters of interest)/(number of data points). This database is larger than the previous database (Clay/5/345) compiled by the authors in Ching and Phoon (2012*a*), where five parameters were of concern. The multivariate normal distribution can be used to capture the correlations among $\{Y_1, \ldots, Y_6\}$ and to derive useful equations for further Bayesian inference. To ensure the analysis results are realistic, the conclusions are further validated by another independent large database consisting of 594 data points where $\{Y_1, \ldots, Y_6\}$ are only partially known (i.e., one or more components are not reported in the original source) and also by empirical equations proposed in literature.

When multivariate geotechnical data exist in sufficient amounts, it is of significant practical usefulness to construct a multivariate probability distribution function. The useful characteristics are: (*i*) it is possible to derive the mean and coefficient of variation (COV) of any parameter given the information contained in a subset with possibly *more than one* parameter and (*ii*) it may be possible to perform a study theoretically using the multivariate distribution if new strong pairwise correlations can be found either among the original components or some derived components. For the former, it is likely for the COV of a design parameter, say s_u, to *reduce* when other parameters, say $(q_t - \sigma_v)/\sigma'_v$ and OCR, have been measured. This aspect is significant for reliability-based design. In fact, COV reduction can be viewed as a value of information and may eventually provide a sensible method for deciding if it is worthwhile to measure an additional parameter. For the latter, the ability to predict the existence of new correlations not included as part of model calibration provides a stronger scientific underpinning to correlation studies in geotechnical engineering. The reason is that these predictions can be falsified by taking new observations, which is the cornerstone of the scientific method. In other words, it is more difficult to develop multivariate models, but if they do stand the test of time, they are usually more robust.

Existing pairwise correlations among $\{Y_1, Y_2, \ldots, Y_6\}$

Pairwise correlations among $\{Y_1, Y_2, \ldots, Y_6\} = \{s_u/\sigma'_v, \text{OCR}, (q_t - \sigma_v)/\sigma'_v, (q_t - u_2)/\sigma'_v, (u_2 - u_0)/\sigma'_v, B_q\}$ have been studied for decades. These correlations are classified into two categories: (*i*) those based on empirical regression and (*ii*) those based on analytical methods.

Correlations based on empirical regression

Robertson et al. (1986) and Senneset et al. (1989) adopted the following three empirical models between s_u/σ'_v and CPTU test indices:

$$N_{kT} = \frac{(q_t - \sigma_v)/\sigma'_v}{s_u/\sigma'_v}$$
$$N_{kE} = \frac{(q_t - u_2)/\sigma'_v}{s_u/\sigma'_v} \quad (1)$$
$$N_{\Delta u} = \frac{(u_2 - u_0)/\sigma'_v}{s_u/\sigma'_v}$$

where "N" variables represent proper empirical constants. Similar empirical models for OCR were proposed by Chen and Mayne (1996):

$$\text{OCR} = 0.259\left(\frac{q_t - \sigma_v}{\sigma'_v}\right)^{1.107}$$
$$\text{OCR} = 0.545\left(\frac{q_t - u_2}{\sigma'_v}\right)^{0.969} \quad (2)$$
$$\text{OCR} = 1.026\left[\frac{(u_2 - u_0)/\sigma'_v}{(q_t - \sigma_v)/\sigma'_v}\right]^{-1.077}$$

Correlations based on analytical methods

Konrad and Law (1987*a*) derived an expression to evaluate the vertical yield stress of a soil during cone penetration:

$$\sigma'_{vc} = \frac{q_t - \alpha u_2}{1 + \delta \tan\phi' \cot\theta} \quad (3)$$

where $\alpha = u_1/u_2$ is a factor to convert the measured pore pressure (u_2) to the pore pressure induced in the failure zone (u_1), δ is the friction coefficient between the soil and cone surface, ϕ' is the effective friction angle of soil, and θ is the apex angle of the cone. Adoption of $\alpha = 1.0$, $\delta = 1.0$, $\theta = 30^o$, and $\phi' = 30^o$ into the above equation and normalization by σ'_v leads to (Robertson et al. 1988)

$$\text{OCR} = 0.5\left(\frac{q_t - u_2}{\sigma'_v}\right) \quad (4)$$

Based on the spherical cavity expansion theory (Vesic 1972),

$$q_t - \sigma_v = s_u[(4/3)(\ln I_r + 1) + \pi/2 + 1]$$
$$\Delta u_{oct} = (4/3)\ln I_r s_u \quad (5)$$

where rigidity index $I_r = G/s_u$, G is the shear modulus of the clay, and Δu_{oct} is the excessive pore pressure due to octahedral stress. The above equation can be combined with the following well-known relation derived from the modified Cam Clay model

$$s_u = (M/2)(\text{OCR}/2)^{\Lambda}\sigma'_v$$
$$\Delta u_{shear} = \sigma'_v[1 - (\text{OCR}/2)^{\Lambda}] \quad (6)$$

where $M = 6\sin\phi'/(3 - \sin\phi')$, Λ is the plastic volume strain ratio or the SHANSEP exponent, and Δu_{shear} is the excessive pore pressure due to shear. Combining eqs. (5) and (6) yields (Mayne 1986; Wroth 1988)

$$\text{OCR} = 2\left\{\frac{(2/M)(q_t - \sigma_v)/\sigma'_v}{(4/3)[\ln(I_r) + 1] + \pi/2 + 1}\right\}^{1/\Lambda} \quad (7)$$

Alternatively, the following relation can be reached by considering the fact that $u_2 = \Delta u_{oct} + \Delta u_{shear} + u_0$ and by taking the typical value $\Lambda = 0.75$ (Chen and Mayne 1994)

$$\text{OCR} = 2\left[(1.95M + 1)\frac{q_t - u_2}{\sigma'_v}\right]^{1.33} \quad (8)$$

Mayne and Holtz (1988) combined eq. (6) with the SHANSEP concept to derive the following expression for OCR:

$$\text{OCR} = \left[\frac{0.75\Delta u_2/\sigma'_v}{(s_u/\sigma'_v)_{NC}\ln(I_r)}\right]^{1/\Lambda} \quad (9)$$

where $(s_u/\sigma'_v)_{NC}$ is the normalized undrained shear strength of a normally consolidated clay. They further adopted typical values of $I_r = 400$, $\Lambda = 0.7$, and $(s_u/\sigma'_v)_{NC} = 0.35$ to reach the following approximation:

$$\text{OCR} = 0.48(\Delta u_2/\sigma'_v)^{1.43} \quad (10)$$

Alternatively, the modified Cam Clay model can be used to derive the following expression (Mayne and Bachus 1988):

$$\text{OCR} = 2\left[\frac{\Delta u_2/\sigma'_v - 1}{(M/2)\ln(I_r) - 1}\right]^{1/\Lambda} = \left[\frac{\Delta u_2/\sigma'_v - 1}{(s_u/\sigma'_v)\ln(I_r) - (1/2)^{\Lambda}}\right]^{1/\Lambda} \quad (11)$$

There are at least two limitations in many of the above pairwise correlations. The first limitation has been discussed earlier: pairwise correlations are unable to address the scenarios where there is multivariate information. Another limitation is that many of the above correlations only provide point estimates of soil parameters. From a design point of view, deterministic point estimates may be insufficient because the resulting design safety level should depend on the degree of parametric uncertainties among others. This is especially true for reliability-based design or comparable simplified methodologies, such as load-resistance factor design (LRFD), limit-state design, and partial-factor design. These design methodologies are increasingly being adopted in new design codes.

Calibration and validation databases

This study compiles two databases from the literature: the Clay/6/535 database consisting of 535 lightly overconsolidated clay data points with complete measurement of $\{Y_1, Y_2, \ldots, Y_6\}$ at close proximity and another database consisting of 594 data points with partially measured information. The former is for the purpose of constructing a multivariate probability distribution for the random vector $\{Y_1, Y_2, \ldots, Y_6\}$, while the latter is for subsequent validation. To ensure meaningful validation, the sources for these two databases are completely independent. In the following, the Clay/6/535 database will be referred to as the calibration database, while the latter is referred to as the validation database.

Calibration database Clay/6/535

Table 1 shows the basic information for the calibration database. There are 535 data points with complete $\{Y_1, Y_2, \ldots, Y_6\}$ information from 40 sites. The geographical regions cover Brazil, Canada, Hong Kong, Italy, Malaysia, Norway, Singapore, Sweden, UK, USA, and Venezuela. The clay properties cover a wide range of OCR (mostly 1~6 except for five sites) and wide range of plasticity index, PI (10~168). Highly overconsolidated (OC) (fissured) and organic clays are nearly absent in this database. The first two moments (mean, COV) and the range of each component Y_i is shown in Table 2. It is apparent that the range covered by each component is fairly large.

The calibration data points are pre-processed based on available information to ensure that subsequent analyses will be as consistent as practically possible:

1. All q_c are converted into q_t to correct for the effect of pore pressure generated behind the cone. This requires the knowledge of u_2 and area ratio, a. There are 15 sites where a is not documented. For these cases, an average a value of 0.7 is assumed, because a typically ranges from 0.55 to 0.9 (Lunne et al. 1997) and because 0.7 is also the average a value for our database. There are also two sites where the pore pressure is not measured. For these cases, the correlation equations suggested by Mayne et al. (1990) are adopted to estimate u_2 based on q_c.
2. All measured pore pressures are converted into u_2. There are nine sites where u_1 (pore pressure at cone tip) rather than u_2 is measured. In this case, the correlation equation suggested by Mayne et al. (1990) is used to convert u_1 into u_2.
3. For all cases, the reported s_u values were tested based on various types of tests, including unconsolidated undrained compression (UU), unconfined compression (UC), triaxial isotropically consolidated undrained compression (CIUC), triaxial K_0-consolidated undrained compression (CK_0UC), and field vane (FV). These values cannot be compared directly because s_u typically depends on stress state, strain rate, sampling disturbance, etc. All measured s_u values are converted into the equivalent CIUC values: CK_0UC to CIUC based on Kulhawy and Mayne (1990), UU and UC to CIUC based on Chen and Kulhawy (1993), and FV to field value based on Bjerrum (1972) (see Table 3). Note that the FV value is similar to a DSS (direct simple shear) value. These DSS values are further converted to CIUC values using the transformation equations suggested by Kulhawy and Mayne (1990).

Ideally, only data sources with complete documentation should be used. Nonetheless, perfect data sources are rare and all current statistical characterization studies in the literature involve making appropriate assumptions to ensure that the sample size is large enough to produce reasonably robust statistics. Ultimately, the appropriateness of these assumptions would be tested using an independent validation database.

Validation database

In literature, most data points do not contain complete $\{Y_1, Y_2, \ldots, Y_6\}$ information: they contain only partial information, i.e., a subset of $\{Y_1, Y_2, \ldots, Y_6\}$ is known. The most common data points would be those containing pairwise information, $\{Y_i, Y_j\}$. Among these pairwise data points, some would be more prevalent than others because they are deemed to be more strongly correlated. It is unfortunate that weaker correlations are automatically dropped out of the geotechnical engineering literature as they still contain valuable information if viewed from a probabilistic sense of uncertainty reduction, albeit to a lesser degree than stronger correlations. In fact, one would hope that the eventual popularization of multivariate analysis such as that attempted in this study would allow all information to be retained in the literature rather than be presented as selective retention of some information deemed useful in the deterministic sense.

The "incomplete" data points in the validation database are left out of the calibration database because the development of a multivariate probability distribution for $\{Y_1, Y_2, \ldots, Y_6\}$ requires complete information. Nonetheless, these data points can be applied in a useful way to validate the conclusions derived from the calibration database. The basic idea here is that once a complete multivariate probability distribution has been constructed, it is possible to draw conclusions for any information subset, including the ubiquitous pairwise information. Clearly, these conclusions can be tested against the appropriate "incomplete" data points in the validation database.

For the validation database, there are 594 data points from 32 sites. Among these data points, 412 contain simultaneous information of $\{Y_1, Y_3, Y_4, Y_5, Y_6\}$, i.e., OCR is missing, and 456 contain simultaneous information of $\{Y_2, Y_3, Y_4, Y_5, Y_6\}$, i.e., s_u is missing. The geographical coverage is quite wide: two sites in Germany, five in Norway, two in Poland, three in Singapore, six in Sweden, four in the USA, one in Japan, others in Hong Kong, Canada, Italy, Korea, Australia, etc. The clay properties cover a wide range of OCR (mostly 1~6 except for two sites) and a wide range of PI (11~86). The clay types are also quite broad. Organic and highly OC (fissured) clays are absent in this database. The same data

Table 1. Basic information for the cases in the Clay/6/535 calibration database.

	Region	Site	OCR	No. of data points	PI	Area ratio	Filter location	UC	UU	CIUC	CAUC	FV	References
1	Ottawa (Canada)	Arnprior	2.3–9.7	21	24–34	0.66	u_2					*	Konrad and Law (1987*a*, 1987*b*)
2		Gloucester	2.8–6.1	18	19–31	0.7	u_1			*		*	Konrad and Law (1987*a*, 1987*b*)
3		STP	3.8–4.9	8	15–37	0.66	u_2		*			*	Konrad and Law (1987*a*, 1987*b*)
4	Quebec (Canada)	Berthierville	1.1–1.5	10	15–27	0.82	u_2					*	Rochelle et al. (1988)
5		Louisville	1.2–2.9	13	36–44	0.82	—					*	Rochelle et al. (1988)
6		NRCC	2–3	5	28–46	0.66	u_2					*	Konrad and Law (1987*a*, 1987*b*); Eden and Law (1980)
7	North Sea	Brage	1.7–3.6	3	21–25	—	u_1				*		Rad and Lunne (1988)
8		Troll	1.6	5	19–41	—	u_1				*		Rad and Lunne (1988)
9	Scotland (UK)	Bothkennar	1–3.5	23	24–75	—	u_2		*			*	Hight et al. (1992); Jacobs and Coutus (1992)
10	Yorkshire (England)	Cowden	2.6–4.5	4	16–22	—	u_2				*		Rad and Lunne (1988)
11	Norway	Drammen	1.1–1.6	22	19–71	0.78	u_2				*	*	Lacasse and Lunne (1982); Rad and Lunne (1988)
12		Emmerstad	3.9	1	10	0.38	u_2				*		Rad and Lunne (1988)
13		Haga	2–4.5	4	12–35	—	u_2				*		Rad and Lunne (1988)
14		Haltenbanken	4.6–8.6	3	18	—	u_2				*		Rad and Lunne (1988)
15		Onsoy	1–1.9	5	35–47	0.8	u_2				*	*	Rad and Lunne (1988)
16		Troll East	1.6–1.8	21	20–45	—	u_2				*	*	Aas et al. (1986)
17	Illinois (USA)	Evanston	1–4.42	18	15–21	0.7	u_1		*			*	Finno (1989)
18	New Brunswick (Canada)	Fredericton	2.8–6.1	6	10.1–13	0.66	u_2					*	Konrad and Law (1987*a*); Valsangar et al. (1985)
19	UK	Grangemouth	1.4–3	11	23–57	0.8	u_2	*			*		Powell and Quarterman (1988); Valsangar et al. (1985)
20	Gothenburg (Sweden)	Gota Alv	1.1–2.2	11	34–46	—	u_2					*	Leroueil et al. (1983)
21		Stora An	1.2–4.2	7	58–86	—	u_2					*	Leroueil et al. (1983)
22	Hong Kong	—	1.9–4	25	12–65	0.7	u_1					*	Koutsoftas et al. (1987)
23	Houston (USA)	—	3.7–5	3	11–29	—	—		*				Mahar and O'Neill (1983)
24	London (UK)	Kingston Bridge	7.2–9.5	4	44–47	0.75	u_2					*	Long and O'Riordan (1988)
25	Malaysian Peninsula	Muar	1.4–3.5	11	33–79	—	u_2					*	Chang (1992)
26	Singapore	Norfolk Road	0.8–1.8	21	17–87	—	u_1		*			*	Chang (1992)
27		KJ_BH	1–8.8	112	13–60	0.8	u_2		*			*	Phoon (2012)
28	Norrkaping (Sweden)	Norrkaping	1.2–1.6	30	11–61	—	u_2					*	Leroueil et al. (1983)
29	Recife city (Brazil)	RR1	1.2–2.2	10	27–79	—	u_1		*	*		*	Coutinho (2007)
30		RR2	1–3.3	14	48–168	—	u_2		*	*		*	Coutinho (2007)
31	Brazil	Rio de Janeiro	1.6–2.12	3	60–87	0.75	u_2				*	*	Rad and Lunne (1988); Rocha-Filho and Alencar (1985); Sills et al. (1988)
32	Orinoco Delta (Venezuela)	Orinoco E1	1.1	5	44–56	0.7-	u_2					*	Azzouz et al. (1983)
33		Orinoco F1	1.2	5	42–57	0.7-	u_2					*	Azzouz et al. (1983)
34	Montreal (Canada)	St. Hilaire	1.1–6.1	10	33–47	0.7	u_2					*	Lafleur et al. (1988)
35	Massachusetts (USA)	Saugus	1.2–5.4	15	13–29	0.7	u_2			*		*	Baligh et al. (1980)
36	Atchafalaya Basin, Louisiana (USA)	Station 1381	1.1–2.1	20	37–117	0.7-	u_2			*		*	Baligh et al. (1980)
37	Stockholm (Sweden)	Svartiolandet	1.1–2.6	16	29–60	—	u_2					*	Leroueil et al. (1983)
38		Upplands-Vasby	1.1–1.4	15	48–83	—	u_2					*	Battaglio et al. (1986); Bayne and Tjelta (1987)
39	Italy	Valdi Chiana	1–8.3	4	31–39	—	u_1		*			*	Cancelli and Cividini (1984)
40	Vancouver (Canada)	—	1.5–1.6	1	16	—	u_1				*		Rad and Lunne (1988)

Table 2. Statistics of (Y_1, Y_2, …, Y_6) based on the 535 data points.

	Mean	COV	Max	Min
$Y_1 = s_u/\sigma'_v$	0.641	0.596	3.041	0.105
Y_2 = OCR	2.353	0.657	9.693	1.000
$Y_3 = (q_t - \sigma_v)/\sigma'_v$	9.350	0.678	58.878	2.550
$Y_4 = (q_t - u_2)/\sigma'_v$	5.280	0.885	43.694	0.605
$Y_5 = (u_2 - u_0)/\sigma'_v$	4.709	0.574	21.720	0.236
$Y_6 = B_q$	0.556	0.338	1.072	−0.093

pre-processing mentioned above for the calibration database is applied to the validation database to ensure consistent interpretation of data.

Construction of multivariate normal distribution

The main purpose of this section is to demonstrate how the multivariate data points in the calibration database, after an appropriate nonlinear transform, can be modeled as a multivariate normal distribution. The multivariate normal probability density function is available analytically and can be defined uniquely by a mean vector and a covariance matrix:

$$f(\underline{\mathbf{X}}) = |\mathbf{C}|^{-1/2}(2\pi)^{-n/2}e^{-(1/2)(\underline{\mathbf{X}}-\underline{\boldsymbol{\mu}})'\mathbf{C}^{-1}(\underline{\mathbf{X}}-\underline{\boldsymbol{\mu}})} \tag{12}$$

in which $\underline{\mathbf{X}} = (X_1, X_2, \ldots X_n)'$ is a normal random vector with n components, $\underline{\boldsymbol{\mu}}$ is the mean vector, and $\mathbf{C}$ is the covariance matrix. For $n = 3$, the mean vector and covariance matrix are given by

$$\underline{\boldsymbol{\mu}} = \begin{Bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{Bmatrix} \quad \mathbf{C} = \begin{bmatrix} \sigma_1^2 & \delta_{12}\sigma_1\sigma_2 & \delta_{13}\sigma_1\sigma_3 \\ \delta_{12}\sigma_1\sigma_2 & \sigma_2^2 & \delta_{23}\sigma_2\sigma_3 \\ \delta_{13}\sigma_1\sigma_3 & \delta_{23}\sigma_2\sigma_3 & \sigma_3^2 \end{bmatrix} \tag{13}$$

where μ_i and σ_i are the mean and standard deviation of X_i, respectively, and δ_{ij} is the product-moment (Pearson) correlation between X_i and X_j. If X_i is a standard normal random variable with zero mean and unit standard deviation (i.e., $\mu_i = 0$ and $\sigma_i = 1$), eq. (13) reduces to

$$\underline{\boldsymbol{\mu}} = \begin{Bmatrix} 0 \\ 0 \\ 0 \end{Bmatrix} \quad \mathbf{C} = \begin{bmatrix} 1 & \delta_{12} & \delta_{13} \\ \delta_{12} & 1 & \delta_{23} \\ \delta_{13} & \delta_{23} & 1 \end{bmatrix} \tag{14}$$

It is clear that the full multivariate dependency structure of a normal random vector only depends on a covariance matrix ($\mathbf{C}$) containing bivariate information (correlations) between all possible pairs of components, namely X_1 and X_2, X_1 and X_3, and X_2 and X_3. The practical advantage of capturing multivariate dependencies in any dimension (i.e., any number of random variables) using only bivariate dependency information is obvious. The most evident disadvantage is that most soil parameters are not normally distributed.

Multivariate non-normal distribution

Let Y_1, Y_2, and Y_3 denote three non-normally distributed soil parameters. One well-known cumulative distribution function (CDF) transform approach for constructing a valid multivariate distribution for these soil parameters is

1. Define $X_i = \Phi^{-1}[F_i(Y_i)]$, in which $\Phi^{-1}(\cdot)$ is the inverse standard normal cumulative distribution function and $F_i(\cdot)$ is the cumulative distribution function of Y_i. By definition, X_1, X_2, and X_3 are *individually* standard normal random variables. That is, the histogram of any component, X_i, will look normal (bell-shaped). For brevity, we denote the CDF transform $X_i = \Phi^{-1}[F_i(Y_i)]$ by $X_i = T_i(Y_i)$, i.e., $T_i(.) = \Phi^{-1}[F_i(.)]$.
2. Assume (X_1, X_2, X_3) follows a multivariate normal distribution as defined by eq. (12) with $\underline{\boldsymbol{\mu}}$ and $\mathbf{C}$ given by eq. (13). It is crucial to note here that *collectively* (X_1, X_2, X_3) does not necessarily follow a multivariate normal distribution even if each component is normally distributed. For example, if the scatter plot of X_i versus X_j shows a distinct nonlinear trend, then the multivariate normal distribution assumption is incorrect.

Simulation

It is simple to obtain realizations of *independent* standard normal random variables (U_1, U_2, U_3) using library functions in many software. Realizations of *correlated* standard normal random variables (X_1, X_2, X_3) can be obtained using $\underline{\mathbf{X}} = \mathbf{L}\underline{\mathbf{U}}$, in which L is the lower triangular Cholesky factor satisfying $\mathbf{C} = \mathbf{L}\mathbf{L}'$. Finally, each soil parameter is obtained using $Y_i = T_i^{-1}(X_i)$.

Application to calibration database

The aforementioned CDF transform approach is applied to the calibration database, where $Y_1 = s_u/\sigma'_v$, Y_2 = OCR, $Y_3 = (q_t - \sigma_v)/\sigma'_v$, $Y_4 = (q_t - u_2)/\sigma'_v$, $Y_5 = (u_2 - u_0)/\sigma'_v$, and $Y_6 = B_q$. Each set of {Y_1, Y_2, …, Y_6} constitutes one data point. The marginal distributions F_1, F_2, …, and F_6 are chosen among the Johnson system of distributions (Slifker and Shapiro 1980; Phoon and Ching 2013). The Johnson system of distributions contains three families of distributions that can be generated as a transformation from a standard normal random variable X:

$$X_n = (X - b_x)/a_x = \kappa(Y; a_y, b_y) \tag{15}$$

where X_n is the shifted and scaled version of the standard normal variable X; a_i and b_i are, respectively the scale and shift parameters in the X or Y space; the transformation function κ has the following forms for the SU, SB, and SL families:

$$\kappa(Y; a_y, b_y) = \begin{cases} \sinh^{-1}(Y_n) & \text{for SU} \\ \ln[Y_n/(1 - Y_n)] & \text{for SB} \\ \ln(Y_n) & \text{for SL} \end{cases} \tag{16}$$

where $Y_n = (Y - b_y)/a_y$ is the shifted and scaled version of Y. The distribution SU is an unbounded distribution that is defined on [−∞, ∞], SB is a bounded distribution defined on [b_y, $a_y + b_y$], and SL is a lower bounded distribution defined on [b_y, ∞]. Their probability density functions are as follows:

$$f(Y) = \begin{cases} (a_x/a_y)\exp\{-0.5[b_x + a_x \sinh^{-1}(Y_n)]^2\}/\sqrt{2\pi(1 + Y_n^2)} & \text{for SU} \\ (a_x/a_y)\exp\langle -0.5\{b_x + a_x \ln[Y_n/(1 - Y_n)]\}^2\rangle/[\sqrt{2\pi}Y_n(1 - Y_n)] & \text{for SB} \\ (a_x/a_y)\exp\{-0.5[b_x + a_x \ln(a_y) + a_x \ln(Y_n)]^2\}/(\sqrt{2\pi}Y_n) = a_x \exp\{-0.5[b_x^* + a_x \ln(Y - b_y)]^2\}/[\sqrt{2\pi}(Y - b_y)] & \text{for SL} \end{cases} \tag{17}$$

Table 3. Equations used for the conversion of undrained shear strength from different test types to equivalent CIUC values.

Test type	Equation	Reference
UU	$s_u(\text{CIUC}) \approx 0.243\sigma'_v + 0.821 s_u(\text{UU})$	Chen and Kulhawy (1993)
UC	$s_u(\text{CIUC}) \approx 0.237\sigma'_v + 0.853 s_u(\text{UC})$	Chen and Kulhawy (1993)
FV	$s_u(\text{DSS}) \approx s_u(\text{field}) \approx \lambda s_u(\text{FV})$; λ depends on PI	Bjerrum (1972)
	$s_u(\text{CIUC}) \approx [s_u(\text{DSS})/(0.77 - 0.0064\phi')/\text{OCR}^{0.776}]\text{OCR}^{0.709}$	Kulhawy and Mayne (1990)
CK_0UC	$s_u(\text{CIUC}) \approx [s_u(CK_0UC)/(1.13 - 0.0094\phi')/\text{OCR}^{0.738}]\text{OCR}^{0.709}$	Kulhawy and Mayne (1990)

Note: Average is taken over available soundings at a site. λ, field vane correction factor proposed by Bjerrum (1972).

where $b_x^* = b_x - a_x \ln(a_y)$.

Figure 1 shows some distributions in the Johnson system — this system can generate distributions with a wide range of mean value, COV, skewness, and kurtosis. The transformation equations $X = T(Y)$ and the inverse transforms $Y = T^{-1}(X)$ for the Johnson distributions are listed in Table 4. Slifker and Shapiro (1980) showed that it is possible to identify the family type (SU, SB, SL) and estimate the four model parameters $\{a_x, b_x, a_y, b_y\}$ based on four sample quantiles of Y. Let us denote the η sample quantile of Y by Y_η. These four sample quantiles are $Y_{\Phi(-3z)}$, $Y_{\Phi(-z)}$, $Y_{\Phi(z)}$, and $Y_{\Phi(3z)}$, where Φ is the cumulative density function of the standard normal distribution. For $z = 0.7$, these are respectively 0.018, 0.242, 0.758, and 0.982 sample quantiles. With a properly chosen z value (0.7 is chosen in this study), Slifker and Shapiro (1980) proposed the following steps: (*i*) first calculate $m = Y_{\Phi(3z)} - Y_{\Phi(z)}$, $n = Y_{\Phi(z)} - Y_{\Phi(-3z)}$, $p = Y_{\Phi(z)} - Y_{\Phi(-z)}$; (*ii*) calculate $D = m(n/p^2)$; (*iii*) the family type can be identified based on D: SU if $D > 1$, SB if $D < 1$, and SL if $D = 1$; (*iv*) the four parameters $\{a_x, b_x, a_y, b_y\}$ can be estimated based on the sample quantiles as well:

$$
\begin{aligned}
&\text{For SU } (D > 1)\\
&\quad a_x = 2z/\cosh^{-1}\{0.5[(m/p) + (n/p)]\}\\
&\quad b_x = a_x \sinh^{-1}\{[(n/p) - (m/p)]/[2(D-1)^{0.5}]\}\\
&\quad a_y = 2p(D-1)^{0.5}/\{[(m/p) + (n/p) - 2][(m/p) + (n/p) + 2]^{0.5}\}\\
&\quad b_y = [(y_z + y_{-z})/2] + p[(n/p) - (m/p)]/\{2[(m/p) + (n/p) - 2]\}\\
&\text{For SB } (D < 1)\\
&\quad a_x = z/\cosh^{-1}\langle 0.5\{[1 + (p/m)][1 + (p/n)]\}^{0.5}\rangle\\
&\quad b_x = a_x \sinh^{-1}\langle [(p/n) - (p/m)]\{[1 + (p/m)][1 + (p/n)] - 4\}^{0.5}/[2(D^{-1} - 1)]\rangle\\
&\quad a_y = p\langle \{[1 + (p/m)][1 + (p/n)] - 2\}^2 - 4\rangle^{0.5}/(D^{-1} - 1)\\
&\quad b_y = [(y_z + y_{-z})/2] - (a_y/2) + p[(p/n) - (p/m)]/[2(D^{-1} - 1)]\\
&\text{For SL } (D = 1)\\
&\quad a_x = 2z/\ln(m/p)\\
&\quad b_x^* = a_x \ln\{[(m/p) - 1]/[p(m/p)^{0.5}]\}\\
&\quad a_y = [(y_z + y_{-z})/2] - 0.5p[(m/p) + 1]/[(m/p) - 1]
\end{aligned}
\tag{18}
$$

Note that the lognormal family (SL) is effectively parameterized by three parameters only (a_x, b_x^*, b_y). Table 5 shows the identified Johnson family types and estimated parameters for Y_1, Y_2, …, and Y_6. Among $\{Y_1, Y_2, \ldots, Y_6\}$, most variables are distributed as SU. The only exception is Y_2 = OCR, which is distributed as SB. The CDF transform for SU will be denoted by T_{SU}, and that for SB will be denoted by T_{SB}. The more well-known lognormal distribution (type SL) does not provide a good fit to most variables. Figure 2 shows histograms of Y_1, Y_2, …, and Y_6 as well as the fitted Johnson distributions.

Y_1, Y_2, …, and Y_6 are transformed into X_1, X_2, …, and X_6 by using the CDF transforms $X = T(Y)$ listed in Table 4. X_1, X_3, X_4, X_5, and X_6 are obtained using T_{SU}, and X_2 is obtained using T_{SB} with the model parameters listed in Table 2. The resulting X_1, X_2, …, and X_6 are roughly standard normal, with the mean of X_i (denoted by μ_i) being zero and the standard deviation of X_i (denoted by σ_i) being unity. The scatter plots between X_i and X_j are shown in Fig. 3 for the 535 calibration data points. It can be seen that the trends are fairly linear. Hence, for this particular soil database, there is no strong evidence to reject the underlying multivariate normality. The product-moment correlation matrix for $(X_1, X_2, \ldots, X_6)$ can be readily estimated, as shown in Table 6.

Spearman (1904) derived the relationship between the product-moment correlation and the rank correlation between bivariate normal random variables

$$\delta_{ij} = 2\sin[(\pi/6)\rho_{ij}] \tag{19}$$

where ρ_{ij} is the rank correlation between X_i and X_j. Note that the rank correlation between X_i and X_j is the same as the rank correlation between Y_i and Y_j because the transformation $X = T(Y)$ does not alter the ranks. The rank correlation matrix of $(Y_1, \ldots, Y_6)$ (or equivalently the rank correlation matrix of $(X_1, \ldots, X_6)$) is shown in Table 7. Table 8 shows the "predicted" product-moment correlation matrix based on the rank correlations in Table 7 and eq. (19). If the underlying multivariate normality does hold exactly, Table 8 must be close to the actual product-moment correlation matrix for $(X_1, \ldots, X_6)$. It is evident that the two matrices in Tables 6 and 8 are similar, suggesting that the underlying multivariate normality may be plausible.

Simulation results

The CDF transform approach is employed to simulate samples of $(Y_1, \ldots, Y_6) = \{s_u/\sigma'_v, \text{OCR}, (q_t - \sigma_v)/\sigma'_v, (q_t - u_2)/\sigma'_v, (u_2 - u_0)/\sigma'_v, B_q\}$ satisfying the multivariate probability distribution in the second-moment sense. A set of 1000 simulated samples, together with the calibration database, are shown in Fig. 4 for illustration. The performance of the CDF transform approach in constructing the multivariate probability distribution is adequate insofar as pairwise correlations are concerned. The simulated samples closely mimic the correlation behaviors of the calibration database.

Fig. 1. Some distributions in the Johnson system: (*a*) SU family, (*b*) SB family, and (*c*) SL family.

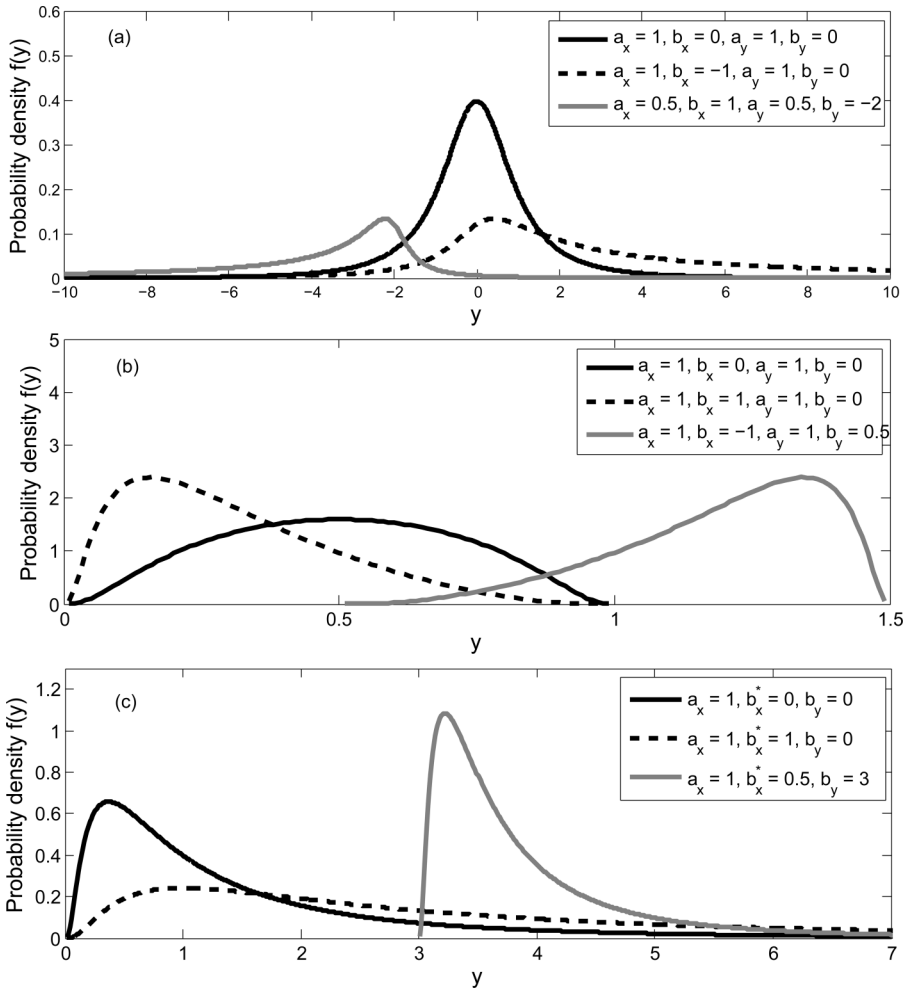


Table 4. Transformation equations $X = T(Y)$ and $Y = T^{-1}(X)$ for the Johnson distributions.

Family	CDF transform $X = T(Y)$	Inverse CDF transform $Y = T^{-1}(X)$
SU	$T_{SU}(Y) = b_x + a_x \sinh^{-1}(Y_n)$	$T^{-1}{}_{SU}(X) = b_y + a_y \sinh(X_n)$
SB	$T_{SB}(Y) = b_x + a_x \ln[Y_n/(1 - Y_n)]$	$T^{-1}{}_{SB}(X) = [b_y + (a_y + b_y) \times \exp(X_n)]/[1 + \exp(X_n)]$
SL	$T_{SL}(Y) = b_x^* + a_x \ln(Y - b_y)$	$T^{-1}{}_{SL}(X) = b_y + \exp(X_n)$

Note: $X_n = (X - b_x)/a_x$ and $Y_n = (Y - b_y)/a_y$ are the shifted and scaled X and Y; $b_x^* = b_x - a_x \ln(a_y)$.

Table 5. Identified Johnson family types and parameters for $(Y_1, Y_2, \ldots, Y_6)$ based on the 535 data points.

	Family	Johnson parameters a_x	b_x	a_y	b_y
$Y_1 = s_u/\sigma'_v$	SU	1.222	−1.742	0.141	0.250
Y_2 = OCR	SB	0.709	1.887	12.724	0.954
$Y_3 = (q_t - \sigma_v)/\sigma'_v$	SU	1.033	−1.438	1.723	4.157
$Y_4 = (q_t - u_2)/\sigma'_v$	SU	0.989	−1.593	0.868	1.638
$Y_5 = (u_2 - u_0)/\sigma'_v$	SU	0.971	−0.762	1.116	3.123
$Y_6 = B_q$	SU	2.961	0.049	0.544	0.570

Higher moments

Figure 4 visually illustrates the similarity between the simulated $(Y_1, \ldots, Y_6)$ and the actual $(Y_1, \ldots, Y_6)$. It is desirable to examine the similarity to a greater depth. The constructed multivariate probability distribution is used to simulate 10 000 batches of $(Y_1, \ldots, Y_6)$ samples, each batch consisting of 535 simulated data points. For each batch, the first four sample moments, including marginal moments such as $E[Y_1^2]$ (where E denotes mean) and cross moments such as $E[Y_1^2 \times Y_5 \times Y_6]$, can be simulated — these moments are called the simulated moments. At the end, there will be 10 000 sets of simulated moments. The moments for the actual calibration data points can be also estimated — these moments are called the actual moments. The first four moments are herein considered because they are related to mean (central position), variance (spread), skewness (degree of asymmetry), and kurtosis (tail thickness) of a probability distribution. The relative locations of the actual moments with respect to the 10 000 sets of simulated moments can be quantified by the quantile position, denoted by P. For instance, if 2500 simulated moments of $E[Y_1]$ out of 10 000 are less than the actual moment, the P position is 2500/10 000 = 0.25. If the P value for a certain moment falls within the $(1 - \alpha) \times 100\%$ confidence interval $[\alpha/2, 1 - \alpha/2]$, the hypothesis that "the constructed multivariate probability distribution correctly reflects the actual moment" cannot be rejected at $100 \times \alpha\%$ level of significance.

Figure 5 shows the P locations for all moments relevant to Y_1. There are in total 39 moments: four marginal moments $E[Y_1]$, $E[Y_1^2]$, $E[Y_1^3]$, and $E[Y_1^4]$, five second-order cross moments $E[Y_1 Y_j]$ ($j = 2$ to 6), 15 third-order cross moments $E[Y_1^2 Y_j]$ and $E[Y_1 Y_j Y_k]$, and 15 fourth-order cross moments $E[Y_1^3 Y_j]$ and $E[Y_1^2 Y_j Y_k]$. It is clear that

Fig. 2. Histograms of Y_1, Y_2, …, and Y_6 and the fitted distributions.

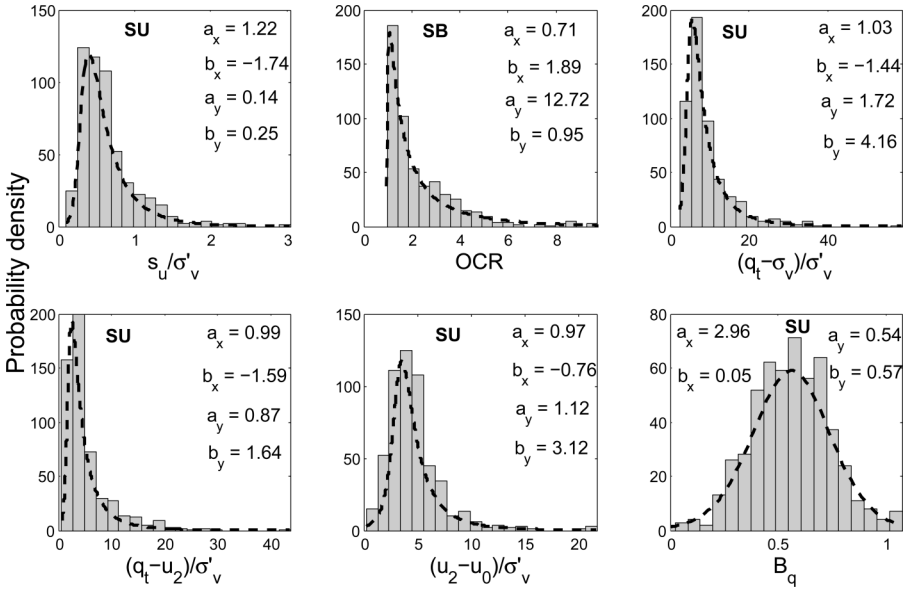
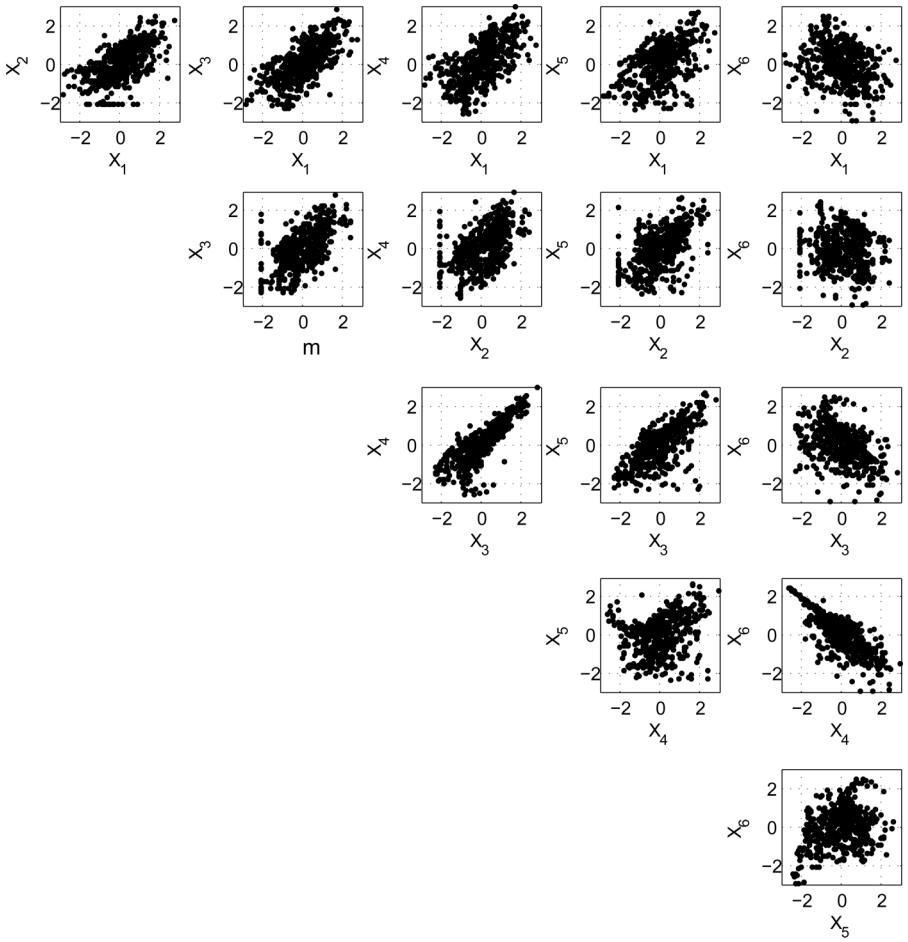


Fig. 3. Examination of multivariate normality between $X_i = T_i(Y_i)$ and $X_j = T_j(Y_j)$ for the calibration database.



all P values are within the 90% confidence interval [0.05, 0.95] (the two horizontal dashed lines in Fig. 5). The same verification is taken for Y_2, Y_3, …, Y_6. The conclusion is similar: nearly all P values are within the 90% confidence interval, except five (out of 39) moments for Y_2, one moment for Y_4, and two moments for Y_6. It is worthy to note that statistical uncertainty increases with the order of the moment being estimated, but this sample size aspect has already been included in the above simulated 90% confidence interval, which is essentially obtained using a parametric bootstrapping approach.

Table 6. Product-moment (Pearson) correlations (**C**) among (X_1, X_2, …, X_6).

	X_1	X_2	X_3	X_4	X_5	X_6
X_1	1.00	$\delta_{12} = 0.62$	$\delta_{13} = 0.67$	$\delta_{14} = 0.61$	$\delta_{15} = 0.49$	$\delta_{16} = -0.28$
X_2	$\delta_{21} = 0.62$	1.00	$\delta_{23} = 0.61$	$\delta_{24} = 0.51$	$\delta_{25} = 0.54$	$\delta_{26} = -0.15$
X_3	$\delta_{31} = 0.67$	$\delta_{32} = 0.61$	1.00	$\delta_{34} = 0.83$	$\delta_{35} = 0.70$	$\delta_{36} = -0.45$
X_4	$\delta_{41} = 0.61$	$\delta_{42} = 0.51$	$\delta_{43} = 0.83$	1.00	$\delta_{45} = 0.31$	$\delta_{46} = -0.77$
X_5	$\delta_{51} = 0.49$	$\delta_{52} = 0.54$	$\delta_{53} = 0.70$	$\delta_{54} = 0.31$	1.00	$\delta_{56} = 0.28$
X_6	$\delta_{61} = -0.28$	$\delta_{62} = -0.15$	$\delta_{63} = -0.45$	$\delta_{64} = -0.77$	$\delta_{65} = 0.28$	1.00

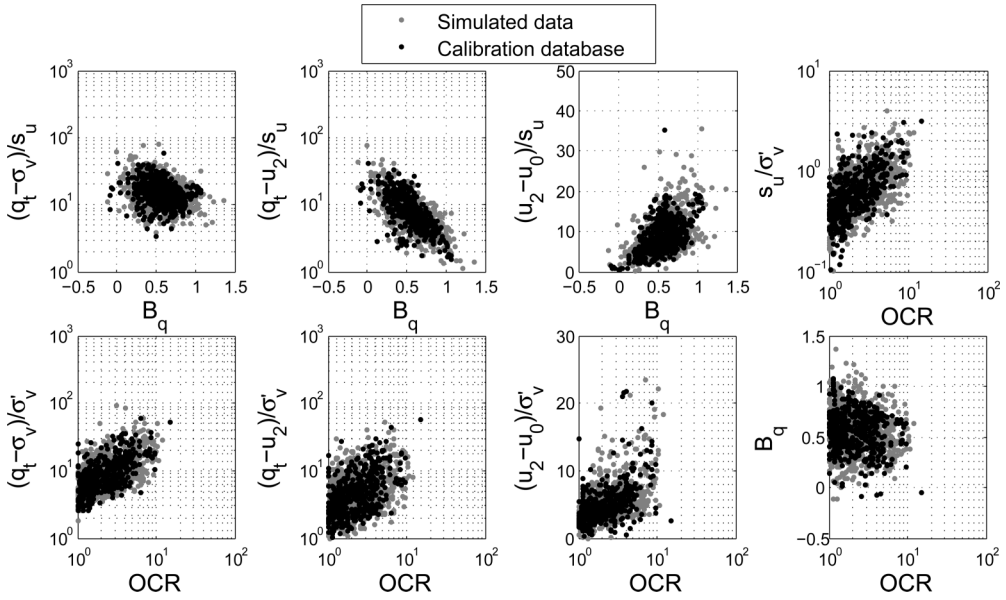
Table 7. Rank (Spearman) correlations (**R**) among (Y_1, Y_2, …, Y_6) (or among (X_1, X_2, …, X_6)).

	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6
Y_1	1.00	$\rho_{12} = 0.65$	$\rho_{13} = 0.68$	$\rho_{14} = 0.64$	$\rho_{15} = 0.47$	$\rho_{16} = -0.30$
Y_2	$\rho_{21} = 0.65$	1.00	$\rho_{23} = 0.58$	$\rho_{24} = 0.51$	$\rho_{25} = 0.54$	$\rho_{26} = -0.11$
Y_3	$\rho_{31} = 0.68$	$\rho_{32} = 0.58$	1.00	$\rho_{34} = 0.85$	$\rho_{35} = 0.68$	$\rho_{36} = -0.48$
Y_4	$\rho_{41} = 0.64$	$\rho_{42} = 0.51$	$\rho_{43} = 0.85$	1.00	$\rho_{45} = 0.35$	$\rho_{46} = -0.74$
Y_5	$\rho_{51} = 0.47$	$\rho_{52} = 0.54$	$\rho_{53} = 0.68$	$\rho_{54} = 0.35$	1.00	$\rho_{56} = 0.22$
Y_6	$\rho_{61} = -0.30$	$\rho_{62} = -0.11$	$\rho_{63} = -0.48$	$\rho_{64} = -0.74$	$\rho_{65} = 0.22$	1.00

Table 8. Predicted product-moment (Pearson) correlations (**C**′) among (X_1, X_2, …, X_6).

	X_1	X_2	X_3	X_4	X_5	X_6
X_1	1.00	$\delta_{12} = 0.64$	$\delta_{13} = 0.69$	$\delta_{14} = 0.63$	$\delta_{15} = 0.51$	$\delta_{16} = -0.29$
X_2	$\delta_{21} = 0.64$	1.00	$\delta_{23} = 0.63$	$\delta_{24} = 0.53$	$\delta_{25} = 0.56$	$\delta_{26} = -0.15$
X_3	$\delta_{31} = 0.69$	$\delta_{32} = 0.63$	1.00	$\delta_{34} = 0.84$	$\delta_{35} = 0.71$	$\delta_{36} = -0.47$
X_4	$\delta_{41} = 0.63$	$\delta_{42} = 0.53$	$\delta_{43} = 0.84$	1.00	$\delta_{45} = 0.33$	$\delta_{46} = -0.78$
X_5	$\delta_{51} = 0.51$	$\delta_{52} = 0.56$	$\delta_{53} = 0.71$	$\delta_{54} = 0.33$	1.00	$\delta_{56} = 0.29$
X_6	$\delta_{61} = -0.29$	$\delta_{62} = -0.15$	$\delta_{63} = -0.47$	$\delta_{64} = -0.78$	$\delta_{65} = 0.29$	1.00

Fig. 4. Comparisons between the calibration database and the 1000 simulated data points.



Validation of constructed multivariate distribution

The multivariate distribution constructed by the CDF transform approach must be validated. In this study, two types of qualitative validations are demonstrated:

1. Comparison between the simulated data points from the constructed multivariate distribution with the validation database. As mentioned earlier in the "Calibration and validation databases" section, an independent validation database has been compiled in addition to the calibration database. The data points in the validation database do not overlap with those in the calibration database. It is insightful to examine visually if the validation database follows the trends produced by the simulated data points.
2. Comparison between the simulated data points from the multivariate distribution with existing empirical correlation equations in the literature. Note that these equations were not used in the calibration of the multivariate distribution. They can be considered independent information summarized in the form of equations and a representation

Fig. 5. P locations for all moments relevant to Y_1.

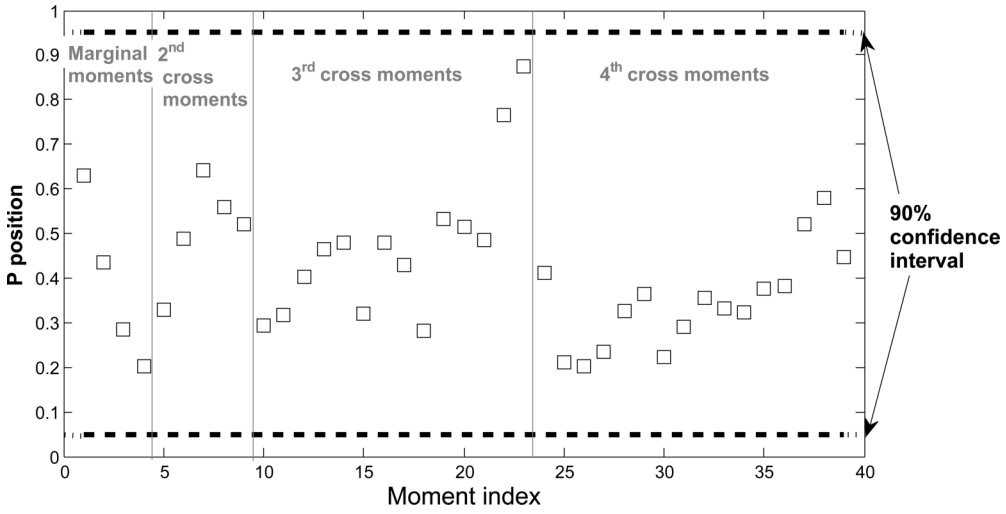
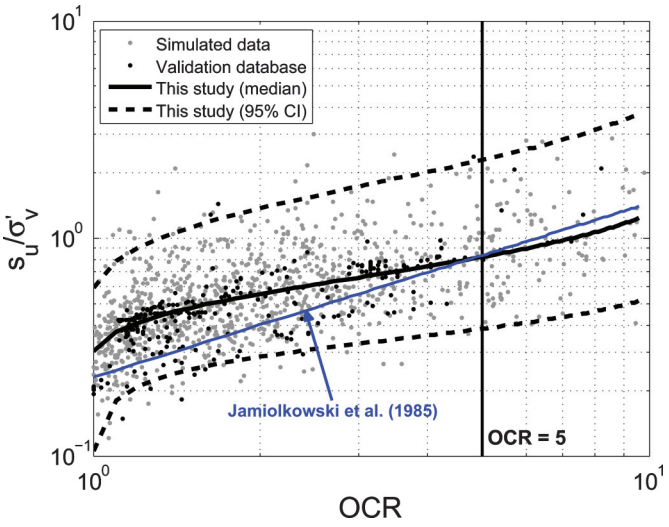


Fig. 6. Verification for the (s_u/σ'_v)–OCR correlation.



of our current state of knowledge. It would be desirable for the constructed multivariate distribution to be largely consistent with our existing body of knowledge.

OCR–s_u/σ'_v correlations

It is well known that s_u/σ'_v is strongly correlated with OCR based on the SHANSEP concept (Ladd and Foott 1974). Figure 6 shows the simulated data and the validation database for the OCR–(s_u/σ'_v) correlations. The empirical equation provided by Jamiolkowski et al. (1985) is also plotted. Jamiolkowski et al. (1985) suggested that $s_u(\text{field})/\sigma'_v \approx 0.23 \times \text{OCR}^{0.8}$. Note that s_u in this study is s_u(CIUC), which is larger than s_u(field). The behaviors of the validation database (dark dots) are similar to those of the simulated data (grey dots). The posterior quantiles (the theoretical quantiles for the simulated data) can be determined by Bayesian analysis (to be presented in a later section) and are plotted in the figure — the solid line is the posterior median and dashed lines are the posterior 95% confidence interval (CI). The empirical equation by Jamiolkowski et al. (1985) seems largely consistent with the simulated data and the validation database. It seems to produce s_u/σ'_v estimates that are more conservative (smaller) than the posterior median of the simulated data points (thin solid line).

$[(q_t - \sigma_v)/\sigma'_v]$–OCR, $[(q_t - u_2)/\sigma'_v]$–OCR, and $[(u_2 - u_0)/\sigma'_v]$–OCR correlations

Figure 7 shows the comparisons between the simulated data and the validation database for the $[(q_t - \sigma_v)/\sigma'_v]$–OCR, $[(q_t - u_2)/\sigma'_v]$–OCR, and $[(u_2 - u_0)/\sigma'_v]$–OCR correlations. The posterior quantiles for the simulated data are also plotted in the figure — the solid line is the posterior median and dashed lines are the 95% CI. The behaviors of the validation database are similar to those of the simulated data. The equations given by literature, including Mayne (1986), Konrad and Law (1987*a*), Mayne and Bachus (1988), Mayne and Holtz (1988), Wroth (1988), and Chen and Mayne (1994) are also plotted — all the equations generally produce OCR estimates that are larger than the posterior median value of the simulated data points (solid line).

N_{kT}–B_q, N_{kE}–B_q, and $N_{\Delta u}$–B_q correlations

Figure 8 compares the simulated data with the validation database for the N_{kT}–B_q, N_{kE}–B_q, and $N_{\Delta u}$–B_q correlations, indicating that the validation data behave similarly to the simulated data. The posterior quantiles for the simulated data are also plotted in the figure — the solid line is the posterior median and dashed lines are the 95% CI. The simulated data points as well as the posterior median (solid line) are consistent to those developed by Ching and Phoon (2012*b*), but the data scatter is somewhat larger than the 95% CIs developed by Ching and Phoon (2012*b*). This may be because the CIs developed by Ching and Phoon (2012*b*) are for measurement-error-free N_{kT}, N_{kE}, and $N_{\Delta u}$.

Bayesian updating for soil parameters

Given the constructed multivariate normal distribution, it is possible to infer any parameter Y_i of interest given the information contained in a subset of other parameters (Y_j, Y_k, …) *analytically*.

Updating in the X space

Before the updating, X_i is standard normal with (prior) mean $\mu_i = 0$ and (prior) standard deviation $\sigma_i = 1$. After the updating (with the information of (X_j, X_k, …)), it can be shown that X_i is

Fig. 7. Verification for the OCR–$[(q_t - \sigma_v)/\sigma'_v]$, OCR–$[(q_t - u_2)/\sigma'_v]$, and OCR–$[(u_2 - u_o)/\sigma'_v]$ correlations.

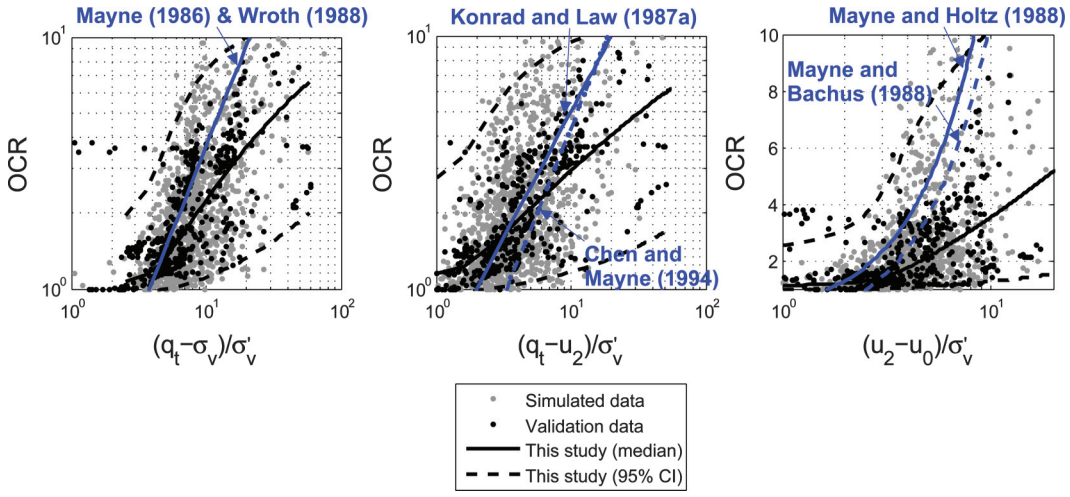
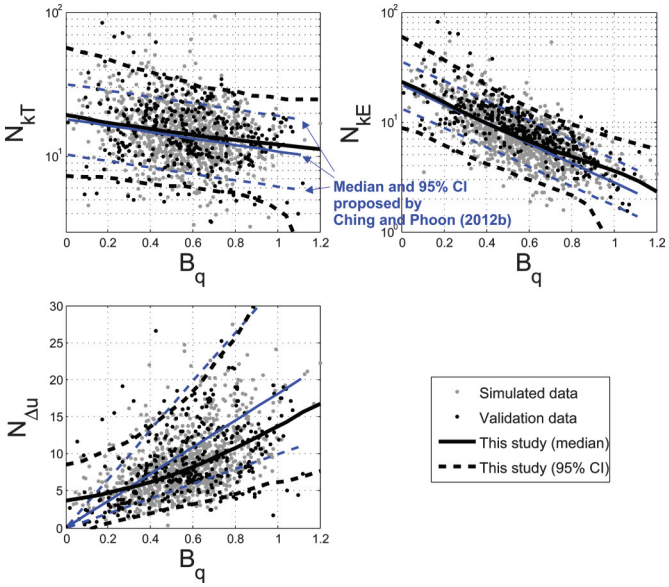


Fig. 8. Validation for the N_{kT}–B_q, N_{kE}–B_q, and $N_{\Delta u}$–B_q correlations.



still normally distributed, but with updated mean = μ'_i and updated standard deviation = σ'_i

$$\mu'_i = \mathrm{E}(X_i|X_j, X_k, \ldots) = \begin{bmatrix}\delta_{ij} & \delta_{ik} & \cdots\end{bmatrix}\begin{bmatrix}1 & \delta_{jk} & \cdots \\ \delta_{jk} & 1 & \\ \vdots & & \ddots\end{bmatrix}^{-1}\begin{bmatrix}X_j \\ X_k \\ \vdots\end{bmatrix}$$

$$\sigma'^2_i = \mathrm{VAR}(X_i|X_j, X_k, \ldots) = 1 - \begin{bmatrix}\delta_{ij} & \delta_{ik} & \cdots\end{bmatrix}\begin{bmatrix}1 & \delta_{jk} & \cdots \\ \delta_{jk} & 1 & \\ \vdots & & \ddots\end{bmatrix}^{-1}\begin{bmatrix}\delta_{ij} \\ \delta_{ik} \\ \vdots\end{bmatrix} \quad (20)$$

where E and VAR denote mean and variance, respectively, and δ_{ij} is the product-moment correlation between X_i and X_j (see Table 6). The multivariate normal is popular in part because of its analytical simplicity as exhibited by eq. (20). In practice, s_u/σ'_v and OCR are of interest, namely X_1 and X_2. Table 9 shows the updated mean μ'_1 and updated standard deviation σ'_1 for X_1 given the information of subsets of (X_2, X_3, X_4, X_5, X_6). Table 10 shows the updated mean μ'_2 and updated standard deviation σ'_2 for X_2 given the information of subsets of (X_3, X_4, X_5, X_6). These

Table 9. Updated means and standard deviations of X_1.

Information	μ'_1	σ'_1
X_2	$0.619X_2$	0.786
X_3	$0.671X_3$	0.742
X_4	$0.612X_4$	0.791
X_5	$0.493X_5$	0.870
X_3, X_6	$0.683X_3 + 0.0266X_6$	0.741
X_4, X_6	$0.972X_4 + 0.468X_6$	0.732
X_5, X_6	$0.618X_5 - 0.451X_6$	0.755
X_3, X_4, X_5, X_6	$0.443X_3 + 0.609X_4 - 0.124X_5 + 0.422X_6$	0.714

Table 10. Updated means and standard deviations of X_2.

Information	μ'_2	σ'_2
X_3	$0.610X_3$	0.793
X_4	$0.514X_4$	0.858
X_5	$0.542X_5$	0.840
X_6	$-0.148X_6$	0.989
X_3, X_6	$0.681X_3 + 0.158X_6$	0.780
X_4, X_6	$0.982X_4 + 0.608X_6$	0.765
X_3, X_4, X_5, X_6	$0.257X_3 + 0.602X_4 + 0.0589X_5 + 0.415X_6$	0.751

updated means and standard deviations are derived analytically based on eq. (20).

Updating in the *Y* space

Recall that Y_i is related to X_i through eq. (15)

$$(X_i - b_{x,i})/a_{x,i} = \kappa(Y_i; a_{y,i}, b_{y,i}) \quad (21)$$

where $a_{x,i}$, $b_{x,i}$, $a_{y,i}$, and $b_{y,i}$ are the parameters for the Johnson distribution of Y_i. After the updating, X_i is still normal with mean = μ'_i and standard deviation = σ'_i. Equation (21) becomes

$$(X_i - b_{x,i})/a_{x,i} = (\mu'_i + \sigma'_i Z - b_{x,i})/a_{x,i} = [Z - (b_{x,i} - \mu'_i)/\sigma'_i]/(a_{x,i}/\sigma'_i) = \kappa(Y_i; a_{y,i}, b_{y,i}) \quad (22)$$

where Z is standard normal. As a result, the posterior distribution of Y_i is still a Johnson distribution of the same family type (SU, SB, or SL) with the following posterior parameters:

$$a'_{x,i} = a_{x,i}/\sigma'_i \quad b'_{x,i} = (b_{x,i} - \mu'_i)/\sigma'_i \quad a'_{y,i} = a_{y,i} \quad b'_{y,i} = b_{y,i} \quad (23)$$

Note that $a'_y = a_y$ and $b'_y = b_y$ parameters remain unchanged — only a'_x and b'_x are updated. For the SL distribution, the second parameter in eq. (23) should be written as $b^{*\prime}_{x,i} = (b^*_{x,i} - \mu'_i)/\sigma'_i$. It is remarkable that the Johnson distribution is a "conjugate prior distribution" with respect to the multivariate normal framework proposed in this study — the posterior distribution is still Johnson.

With the above results, one can easily calculate the parameters of the posterior Johnson distribution for Y_1 or Y_2. For instance, suppose $Y_3 = (q_t - \sigma_v)/\sigma'_v = 3.6$ and $Y_6 = B_q = 0.5$ are known from a CPTU test, and the parameters of the posterior Johnson distribution for $Y_1 = s_u/\sigma'_v$ can be obtained by the following steps:

Step 1: Convert (Y_3, Y_6) to $(Y_{n,3}, Y_{n,6})$:

$$\begin{aligned} Y_{n,3} &= (Y_3 - b_{y,3})/a_{y,3} = (3.6 - 4.157)/1.723 = -0.323 \\ Y_{n,6} &= (Y_6 - b_{y,6})/a_{y,6} = (0.5 - 0.570)/0.544 = -0.129 \end{aligned} \tag{24}$$

Step 2: Convert $(Y_{n,3}, Y_{n,6})$ to (X_3, X_6) (see the T_{SU} transform in Table 4):

$$\begin{aligned} X_3 &= b_{x,3} + a_{x,3}\sinh^{-1}(Y_{n,3}) \\ &= -1.438 + 1.033\sinh^{-1}(-0.323) = -1.766 \\ X_6 &= b_{x,6} + a_{x,6}\sinh^{-1}(Y_{n,6}) \\ &= 0.049 + 2.961\sinh^{-1}(-0.129) = -0.331 \end{aligned} \tag{25}$$

where the parameters $(a_{x,3}, b_{x,3}, a_{y,3}, b_{y,3})$ and $(a_{x,6}, b_{x,6}, a_{y,6}, b_{y,6})$ are listed in Table 2.

Step 3: Compute μ'_1 and σ'_1 based on (X_3, X_6) (see the fifth row in Table 9):

$$\mu'_1 = 0.683X_3 + 0.0266X_6 = -1.215 \quad \sigma'_1 = 0.741 \tag{26}$$

Step 4: Determine the parameters of the posterior Johnson distribution for Y_1 according to eq. (23):

$$\begin{aligned} a'_{x,1} &= a_{x,1}/\sigma'_1 = 1.222/0.741 = 1.649 \\ b'_{x,1} &= (b_{x,1} - \mu'_1)/\sigma'_1 = (-1.742 + 1.215)/0.741 = -0.711 \\ a'_{y,1} &= a_{y,1} = 0.141 \\ b'_{y,1} &= b_{y,1} = 0.250 \end{aligned} \tag{27}$$

Note that the posterior distribution for Y_1 is still Johnson SU.

Figure 9 shows examples of the posterior probability density functions of $Y_1 = s_u/\sigma'_v$. The formulas for the Johnson probability density functions are given in eq. (17). The upper left plot shows the prior density function of Y_1 (no information), while the other plots show the posterior density functions of Y_1 given the various information. The posterior mean values and posterior COVs of the density functions are also tabulated in the plots. It is clear that incorporating more information may result in smaller posterior COVs.

It is also possible to compute the η-quantile for the posterior Johnson distribution analytically

$$Y_\eta = \begin{cases} b'_y + a'_y \sinh\{[\Phi^{-1}(\eta) - b'_x]/a'_x\} & \text{for SU} \\ \langle b'_y + (a'_y + b'_y)\exp\{[\Phi^{-1}(\eta) - b'_x]/a'_x\}\rangle/\langle 1 + \exp\{[\Phi^{-1}(\eta) - b'_x]/a'_x\}\rangle & \text{for SB} \\ b'_y + a'_y \exp\{[\Phi^{-1}(\eta) - b'_x]/a'_x\} = b'_y + \exp\{[\Phi^{-1}(\eta) - b^{*\prime}_x]/a'_x\} & \text{for SL} \end{cases} \tag{28}$$

where Y_η is the posterior η-quantile of Y. To showcase the posterior quantile Y_η, Fig. 6 shows the case where $Y_1 = s_u/\sigma'_v$ is to be updated with information Y_2 = OCR (the first scenario in Table 9). Consider the information Y_2 = OCR = 5 (the vertical solid line in Fig. 6). The following five steps can be used to determine the posterior η-quantile of Y_1. The first four steps are very similar to the four steps presented in the previous paragraph.

Step 1: Convert Y_2 to $Y_{n,2}$:

$$Y_{n,2} = (Y_2 - b_{y,2})/a_{y,2} = (5 - 0.954)/12.724 = 0.318 \tag{29}$$

Step 2: Convert $Y_{n,2}$ to X_2 (recall Y_2 is distributed as Johnson SB; see T_{SB} in Table 4):

$$\begin{aligned} X_2 &= b_{x,2} + a_{x,2}\ln[Y_{n,2}/(1 - Y_{n,2})] = 1.887 \\ &\quad + 0.709\ln[0.318/(1 - 0.318)] = 1.346 \end{aligned} \tag{30}$$

where the parameters $(a_{x,2}, b_{x,2}, a_{y,2}, b_{y,2})$ are listed in Table 2.

Step 3: Compute μ'_1 and σ'_1 based on X_2 (see the first row in Table 9):,

$$\mu'_1 = 0.619X_2 = 0.833 \quad \sigma'_1 = 0.786 \tag{31}$$

Step 4: Determine the parameters of the posterior Johnson distribution for Y_1 according to eq. (23):

$$\begin{aligned} a'_{x,1} &= a_{x,1}/\sigma'_1 = 1.222/0.786 = 1.555 \\ b'_{x,1} &= (b_{x,1} - \mu'_1)/\sigma'_1 = (-1.742 - 0.833)/0.786 = -3.276 \\ a'_{y,1} &= a_{y,1} = 0.141 \\ b'_{y,1} &= b_{y,1} = 0.250 \end{aligned} \tag{32}$$

Note that the posterior distribution for Y_1 is still Johnson SU.

Step 5: Determine the posterior η-quantile of Y_1 according to eq. (23) (recall Y_1 is Johnson SU):

$$\begin{aligned} Y_\eta &= b'_{y,1} + a'_{y,1}\sinh\{[\Phi^{-1}(\eta) - b'_{x,1}]/a'_{x,1}\} \\ &= 0.250 + 0.141\sinh\{[\Phi^{-1}(\eta) + 3.276]/1.555\} \end{aligned} \tag{33}$$

As a result, the updated median value $Y_{0.5}$ is about 0.82, whereas the updated $Y_{0.025}$ and $Y_{0.975}$ are, respectively, 0.38 and 2.29, hence the updated 95% confidence interval (CI) is [0.38, 2.29].

The above calculation is repeated for various values of OCR. The resulting relationship between OCR and the updated median value of s_u/σ'_v (Y_1) is shown as the solid line in Fig. 6, whereas the relationship between OCR and the updated 95% CI is shown as the dashed lines. The comparison between the 95% CI and the validation database shows acceptable consistency with 3% outliers. The downward trends for the posterior quantiles near OCR = 1 may not reflect real behaviors of the calibration data, but are due to the fact that the SB distribution is used for OCR. The SB distribution is bounded from below at OCR = 0.954 ($b_{y,2} = 0.954$ in Table 2). OCR = 1 is close to the lower bound and is a very low quantile position. When OCR = 1 is mapped to the standard normal space, the

Fig. 9. Posterior probability density functions of $Y_1 = s_u/\sigma'_v$ given the information of subsets of (Y_3, Y_4, Y_5, Y_6).

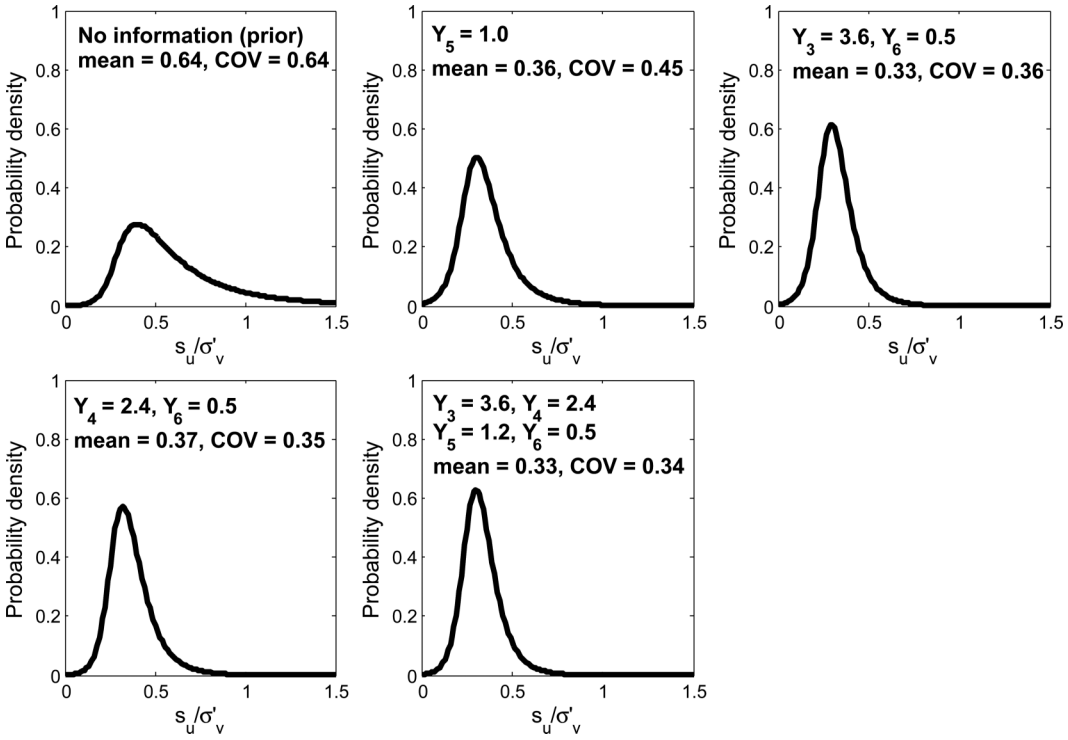
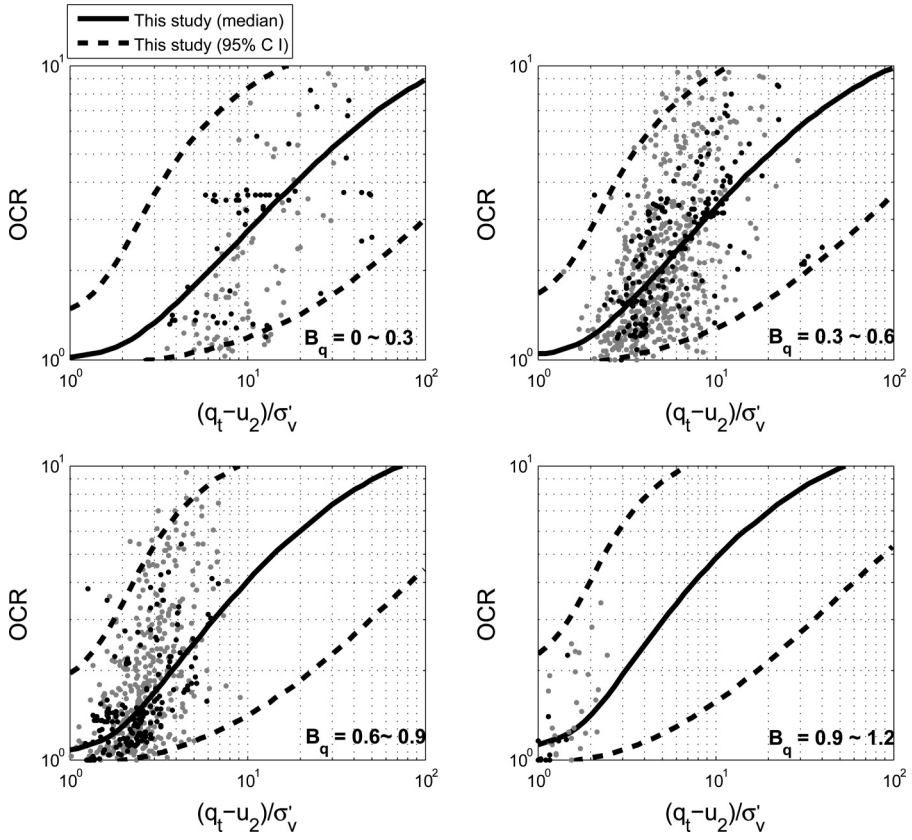


Fig. 10. Comparisons between the 95% CIs restricted to various B_q ranges with the validation data within each range. The 95% CI in each plot is based on B_q at the middle value of the range.



corresponding X_2 will be very negative. Because X_1 and X_2 are positively correlated ($\delta_{12} = 0.62$), the posterior quantile for X_2 will be fairly negative, too. This explains the downward trends in the posterior quantiles near OCR = 1 in Fig. 6.

To further showcase the Bayesian updating, the fifth, sixth, and seventh rows in Table 9 are used to derive the posterior quantiles for the N_{kT}–B_q, N_{kE}–B_q, and $N_{\Delta u}$–B_q relations, respectively, shown in Fig. 8. The updated median and 95% CI are plotted in this figure, and the % outliers produced by the validation database are found to be 7%, 5%, and 7%, respectively. Other scenarios in Table 9 also perform satisfactorily, with % outliers for the validation database ranging from 3% to 7%. These outlier percentages are deemed reasonable because these percentages are close to the 5% theoretical value (100%–95% = 5% outliers).

The first, second, and third rows in Table 10 can be used to derive the posterior quantiles for the relations shown in Fig. 7. The updated median and 95% CI are plotted in this figure with the validation database — the consistency is high. The % outliers for the validation database are found to be 2%, 2%, and 6%, respectively. Other equations in Table 10 also perform satisfactorily, with % outliers for the validation database ranging from 1% to 6%.

Beyond pairwise correlations

The constructed multivariate distribution is essentially founded on all possible pairwise correlations between any two components in the multivariate data. For example, in Table 6, it is clear that 15 separate correlations are needed to populate the correlation matrix for a six-dimensional random vector. It is interesting to explore if this relatively simple simulation method can capture higher-order dependency information in the database. It is difficult to visualize dependency between three or more components and there is no common statistical measure to quantify such higher-order dependencies. Nevertheless, it is of practical interest to study one aspect of higher-order dependency that can be found occasionally in practice; namely, a pairwise correlation equation may occasionally depend on the value of a third variable. In other words, the coefficients of the correlation equation are not constants, but functions of a third variable.

This section studies one such example: $[(q_t - u_2)/\sigma'_v]$–OCR correlation. The goal is to study how these correlation equations are affected by a third variable, B_q. Figure 10 compares the medians and 95% CIs restricted to various B_q ranges with the validation data (dark dots) within each range. For instance, the upper left plot in the figure shows the validation data points with $B_q = 0 \sim 0.3$, and the posterior median and 95% CI shown in the plot are derived based on $B_q = (0 + 0.3)/2 = 0.15$. It is evident that the 95% CI moves slightly upwards for increasing B_q. This behavior seems to be consistent with the validation database. This limited study appears to imply that third-order information is reasonably captured by the constructed multivariate distribution.

Conclusion

In this study, 535 data points in which CPTU parameters, undrained shear strength, and overconsolidation ratio are simultaneously measured in close proximity in lightly overconsolidated clays are compiled for the purpose of constructing a multivariate probability distribution. Each data point in this calibration database consists of six components: (*i*) normalized undrained shear strength (s_u/σ'_v), (*ii*) overconsolidation ratio (OCR), (*iii*) normalized cone tip resistance $(q_t - \sigma_v)/\sigma'_v$, (*iv*) normalized effective cone tip resistance $(q_t - u_2)/\sigma'_v$, (*v*) normalized excess pore pressure $(u_2 - u_0)/\sigma'_v$, and (*vi*) pore pressure ratio B_q. Based on a proposed CDF transform approach, the above six parameters and the relationships between all possible subset of parameters can be succinctly characterized as a multivariate distribution.

A second database consisting of 594 data points from 32 sites with lightly overconsolidated clays is compiled to validate the consistency of the constructed multivariate distribution and the resulting predictive equations for s_u/σ'_v and OCR. Among the 594 data points in this validation database, 412 contain simultaneous measurement of all parameters except OCR and 456 contain simultaneous measurement of all parameters except that s_u is missing. This validation database is entirely independent of the calibration database. The validation shows evidence of consistency.

The behavior of the multivariate distribution is also compared to existing empirical equations proposed in literature. The behavior is largely consistent with these equations, which can be viewed as representing our existing body of knowledge. Once the multivariate distribution is constructed and validated, a set of Bayesian equations that are useful for updating the distributions of s_u/σ'_v and OCR based on the availability of CPTU parameters are derived.

It is important to note that the constructed six-dimensional multivariate distribution is theoretically capable of predicting a wealth of relationships between parameters not considered in the literature. These relationships exist to maintain probabilistic consistency within the multivariate distribution. These relationships may be falsified in the presence of new data. The multivariate distribution has a significant theoretical advantage over existing bivariate regression equations precisely because it can be falsified and hence its generality can be improved. Existing bivariate regression equations are always correct as long as they are used within the confines of the calibration database. There is no systematic method to improve their generality beyond the calibration database. They also focus on strong correlations. Weak correlations are typically discarded. Their value in reducing the coefficient of variation of design parameters is not well appreciated. Site investigation is a costly exercise and ideally, one should exploit all measured geotechnical data for design. The multivariate distribution is an elegant model to summarize all available information, including uncertainties and their correlations. Bayesian inference is a systematic method for updating any parameter of interest from information gathered in other parameters, taking into account the "value" of these additional inputs correctly.

References

Aas, G., Lacasse, S., Lunne, T., and Hoeg, K. 1986. Use of in-situ tests for foundation design on clay. *In* Use of in-situ tests in geotechnical engineering. GSP 6. ASCE, Blacksburg, Va. pp. 1–30.

Azzouz, A., Baligh, M.M., and Ladd, C.C. 1983. Cone penetration and engineering properties of soft Orinoco clay. *In* Proceedings of the 3rd International Conference on Behavior of Offshore Structure, Cambridge, Mass. Vol. 1, pp. 161–180.

Baligh, M.M., Vivatrat, V., and Ladd, C.C. 1980. Cone penetration in soil profiling. Journal of the Soil Mechanics and Foundations Division, ASCE, **106**(4): 447–461.

Battaglio, M., Bruzzi, D., Jamiolkowski, M., and Lancellotta, R. 1986. Interpretation of CPT's and CPTU's. *In* Proceedings of the 4th International Geotechnical Seminar: Field Instrumentation and In-Situ Measurements, Singapore. pp. 129–143.

Bayne, J.M., and Tjelta, T.I. 1987. Advanced cone penetrometer development for in-situ testing at Gullfaks C. *In* Proceedings of the 19th Offshore Technology Conference, Houston. Vol. 4, pp. 531–540. doi:10.4043/5420-MS.

Bjerrum, L. 1972. Embankment on soft ground. *In* Proceedings of the ASCE Specialty Conference on Performance of Earth and Earth-Supported Structures, Lafayette.

Cancelli, A., and Cividini, A. 1984. An embankment on soft clays with sand drains: numerical characterization of the parameters from in-situ measurements. *In* Proceedings of the International Conference on Case Histories in Geotechnical Engineering. pp. 637–643.

Chang, M.F. 1992. Interpretation of overconsolidation ratio from *in situ* tests in recent clay deposits in Singapore and Malaysia: Reply. Canadian Geotechnical Journal, **29**(1): 168. doi:10.1139/t92-020.

Chen, Y.-J., and Kulhawy, F.H. 1993. Undrained strength interrelationships among CIUC, UU, and UC tests. Journal of Geotechnical Engineering, **119**(11): 1732–1750. doi:10.1061/(ASCE)0733-9410(1993)119:11(1732).

Chen, B.S.-Y. and Mayne, P.W. 1994. Profiling the overconsolidation ratio of clays by piezocone tests. Report No. GIT-CEE/GEO-94-1 submitted to the National Science Foundation by Georgia Institute of Technology, Atlanta, August 1994.

Chen, B.S.Y., and Mayne, P.W. 1996. Statistical relationships between piezocone measurements and stress history of clays. Canadian Geotechnical Journal, **33**(3): 488–498. doi:10.1139/t96-070.

Ching, J., and Phoon, K.-K. 2012*a*. Modeling parameters of structured clays as a multivariate normal distribution. Canadian Geotechnical Journal, **49**(5): 522–545. doi:10.1139/t2012-015.

Ching, J., and Phoon, K.-K. 2012*b*. Establishment of generic transformations for geotechnical design parameters. Structural Safety, **35**: 52–62. doi:10.1016/j.strusafe.2011.12.003.

Coutinho, R.Q. 2007. Characterization and engineering properties of Recife soft clays – Brazil. *In* Proceedings of the International Workshop on Characterisation and Engineering Properties of Natural Soil, Singapore. pp. 2049–2099. doi:10.1201/NOE0415426916.ch12.

Eden, W.J., and Law, K.T. 1980. Comparison of undrained shear strength results obtained by different test methods in soft clays. Canadian Geotechnical Journal, **17**(3): 369–381. doi:10.1139/t80-044.

Finno, R.J. 1989. Subsurface conditions and pile installation data: 1989 foundation engineering congress test section. *In* Predicted and observed axial behavior of piles: results of a pile prediction symposium. Geotechnical Special Publication 23. ASCE, New York. pp. 1–74.

Hight, D.W., Bond, A.J., and Legge, J.D. 1992. Characterization of the Bothkennar clay: an overview. Géotechnique, **42**(2): 303–347. doi:10.1680/geot.1992.42.2.303.

Jacobs, P.A., and Coutus, J.S. 1992. A comparison of electric piezocone tips at the Bothkennar test site. Géotechnique, **42**(2): 369–375. doi:10.1680/geot.1992.42.2.369.

Jamiolkowski, M., Ladd, C.C., Germain, J.T., and Lancellotta, R. 1985. New developments in field and laboratory testing of soils. *In* Proceedings of the 11th International Conference on Soil Mechanics and Foundation Engineering, San Francisco. Vol. 1, pp. 57–153.

Konrad, J.-M., and Law, K.T. 1987*a*. Preconsolidation pressure from piezocone tests in marine clays. Géotechnique, **37**(2): 177–190. doi:10.1680/geot.1987.37.2.177.

Konrad, J.-M., and Law, K.T. 1987*b*. Undrained shear strength from piezocone tests. Canadian Geotechnical Journal, **24**(3): 392–405. doi:10.1139/t87-050.

Koutsoftas, D.C., Foott, R., and Handfelt, L.D. 1987. Geotechnical investigations offshore Hong Kong. Journal of Geotechnical Engineering, **113**(2): 87–105. doi:10.1061/(ASCE)0733-9410(1987)113:2(87).

Kulhawy, F.H., and Mayne, P.W. 1990. Manual on estimating soil properties for foundation design. EPRI. Report EL6800, August.

Lacasse, S., and Lunne, T. 1982. Penetration tests in two Norwegian clays. *In* Proceedings of the 2nd European Symposium on Penetration Testing, Amsterdam. pp. 661–669.

Ladd, C.C., and Foott, R. 1974. New design procedure for stability of soft clays. Journal of the Soil Mechanics and Foundations Division, ASCE, **100**(7): 763–786.

Lafleur, J., Silvestri, V., Asselin, R., and Soulié, M. 1988. Behaviour of a test excavation in soft Champlain Sea clay. Canadian Geotechnical Journal, **25**(4): 705–715. doi:10.1139/t88-081.

Leroueil, S., Tavenas, F., Samson, L., and Morin, P. 1983. Preconsolidation pressure of Champlain clays. Part II. Laboratory determination. Canadian Geotechnical Journal, **20**(4): 803–816. doi:10.1139/t83-084.

Long, M.M., and O'Riordan, N.J. 1988. The use of piezocone in the design of a deep basement in London clay. *In* Penetration testing in the U.K. Thomas Telford, London. pp. 173–176.

Lunne, T., Robertson, P.K., and Powell, J.J.M. 1997. Cone penetrating testing: in geotechnical practice. Taylor & Francis.

Mahar, L.J., and O'Neill, M.W. 1983. Geotechnical characterization of desiccated clay. Journal of Geotechnical Engineering, **109**(1):56–71. doi:10.1061/(ASCE)0733-9410(1983)109:1(56).

Mayne, P.W. 1986. CPT indexing of in situ OCR in clays. *In* Use of In Situ Tests in Geotechnical Engineering. GSP 6. pp. 780–793.

Mayne, P.W., and Bachus, R.C. 1988. Profiling OCR in clays by piezocone. *In* Proceedings of the 1st International Symposium on Penetration Testing. Balkema, Rotterdam. Vol. 2, pp. 857–864.

Mayne, P.W., and Holtz, R.D. 1988. Profiling stress history from piezocone soundings. Soils and Foundations, **28**(1): 16–28. doi:10.3208/sandf1972.28.16.

Mayne, P.W., Kulhawy, F.H., and Kay, J.N. 1990. Observations on the development of pore-water stresses during piezocone penetration in clays. Canadian Geotechnical Journal, **27**(4): 418–428. doi:10.1139/t90-058.

Phoon, K.K. 2012. Personal communication.

Phoon, K.K., and Ching, J. 2013. Multivariate model for soil parameters based on Johnson distributions. *In* Foundation engineering in the face of uncertainty. Geotechnical Special Publication honoring Professor F.H. Kulhawy. pp. 337–353. doi:10.1061/9780784412763.027.

Powell, J.J.M., and Quarterman, R.S.T. 1988. The Interpretation of cone penetration tests in clays, with particular reference to rate effects. *In* Proceedings of the 1st International Symposium on Penetration Testing, Orland., Vol. 2, pp. 903–909.

Rad, N.S., and Lunne, T. 1988. Direct correlations between piezocone test results and undrained shear strength of clay. *In* Proceedings of the 1st International Symposium on Penetration Testing, Orlando, Fla. Vol. 2, pp. 911–917.

Robertson, P.K., Campanella, R.G., Gillespie, D., and Greig, J. 1986. Use of piezometer cone data. *In* Use of in situ tests in geotechnical engineering. Geotechnical Special Publication 6. pp. 1263–1280.

Rocha-Filho, P., and Alencar, J.A. 1985. Piezocone tests in the Rio De Janeiro soft clay deposit. *In* Proceedings of the 11th International Conference on Soil Mechanics and Foundation Engineering, San Francisco. Vol. 2, pp. 859–862.

Rochelle, P.L., Zebdi, M., Leroueil, S., Tavenas, F., and Virely, D. 1988. Piezocone tests in sensitive clays of eastern Canada. *In* Proceedings of the 1st International Symposium on Penetration Testing, Orlando, Fla. Vol. 2, pp. 831–841.

Robertson, P.K., Howie, J.A., Sully, J.P., Gillespie, D.G., and Campanella, R.G. 1988. Discussion on Preconsolidation pressure from piezocone tests in marine clay by J.M. Konrad and K. Law. Geotechnique, **38**(3): 455–465. doi:10.1680/geot.1988.38.3.455.

Senneset, K., Sandven, R., and Janbu, N. 1989. Evaluation of soil parameters from piezocone tests. *In* In situ testing of soil properties for transportation. Transportation Research Record No. 1235. pp. 24–37.

Sills, G.C., Almeida, M.S.S., and Danziger, F.A.B. 1988. Coefficient of consolidation from piezocone dissipation tests in a very soft clay. *In* Proceedings of the 1st International Symposium on Penetration Testing, Orlando, Fla. Vol. 2, pp. 967–974.

Slifker, J.F., and Shapiro, S.S. 1980. The Johnson system: selection and parameter estimation. Technometrics, **22**(2): 239–246. doi:10.1080/00401706.1980.10486139.

Spearman, C. 1904. "General intelligence", objectively determined and measured. American Journal of Psychology, **15**: 201–293. doi:10.2307/1412107.

Valsangar, A.J., Landva, A.O., and Alkins, J.C. 1985. Performance of a raft foundation supporting a multi-storey building. *In* Proceedings of the 38th Canadian Geotechnical Conference, Edmonton, Alta. pp. 315–323.

Vesic, A.S. 1972. Expansion of cavities in infinite soil mass. Journal of the Soil Mechanics and Foundations Division, ASCE, **98**(3): 265–290.

Wroth, C.P. 1988. Penetration testing: a rigorous approach to interpretation. *In* Proceedings of the 1st International Symposium on Penetration Testing, Orlando, Fla. Vol. 1, pp. 303–311.