Multivariate probabilistic assessment of a regional database in Copenhagen

Efthymios Panagiotis[1#], *Irene* Rocchi[1], and *Varvara* Zania[1]

[1]*Technical University of Denmark, Department of Environmental and Resource Engineering, Kongens Lyngby, Denmark*
[#]*Corresponding author: eftpan@dtu.dk*

ABSTRACT

Extensive research has been carried out to establish empirical equations between two geotechnical properties and in spite of the increase in correlation studies, a construction of a multivariate distribution function of more than two parameters is still rare. The objective of this study is to investigate the possibility of constructing two multivariate distribution functions, each consisting of 5 geotechnical parameters based on a) triaxial and soil classification tests and b) oedometer and soil classification tests in stiff over-consolidated clay till from Copenhagen. For this purpose, laboratory measurements from twenty-six sites in Copenhagen are utilized and two multivariate databases are constructed. The correlations among the ratio of deviatoric and mean effective stresses (q/p′), over-consolidation ratio (OCR), secant modulus at reference stress 100 kPa (E_{50}^{ref}), initial void ratio (e_o) and liquid limit (LL) are investigated in multivariate model A, while multivariate model B demonstrates the correlations among oedometer modulus at 100 kPa reference stress ($E_{oed(ur)}^{ref}$), over-consolidation ratio (OCR), permeability change index (C_k), initial void ratio (e_o) and liquid limit (LL). The Nataf transformation model is used for the construction of the multivariate distributions, which are then used to simulate the correlations between geotechnical properties by generating artificial samples. Finally, the artificial samples are compared with the original database for an initial validation of the model. The constructed multivariate models obtained as a result of this study can act as prior for Bayesian updating in multivariate distribution functions when additional geotechnical tests are carried out.

Keywords: multivariate distribution; nataf transformation; regional database; Copenhagen clay till.

1. Introduction

In recent years, great attention has been paid to the application of probabilistic methods in soil characterization, as the interpretation leading to design parameters from site investigation reports is a critical decision in geotechnical engineering and probabilistic methods allow geotechnical engineers to quantify uncertainties, which unavoidably arise in a geotechnical project. Soil characterization uncertainties are associated with inherent variability of soils and rocks, measurement and/or testing errors, transformation models and limited available data from geotechnical tests owing to budget constraints (Baecher and Christian 2005, Ching and Phoon 2012). Deterministic approaches neglect soil characterization uncertainties, which leads to overestimation of failure probabilities, increased design costs and deteriorate sustainable footprint Luo et al. (2018). Luo et al. (2018) highlighted the influence of accounting for the variation of the soil properties by investigating the effect of soil spatial variability on structural reliability assessment in deep excavation in soft silty clay deposits. Different levels of coefficient of variation were considered for undrained shear strength and secant modulus and their effect on bending moment and strut forces was evaluated through a FEM model. Simplistic probabilistic models, such as marginal probability distributions, are not sufficient to measure uncertainties introduced by the soil inherent variability, due to the complexity of soil behavior and/or incomplete soil databases (Phoon and Kulhawy 1999).

On the other hand, advanced probabilistic models, such as the bivariate and the multivariate approaches, can provide valuable insights on the statistical dependence of soil properties, which can be a powerful tool for quantifying adequately geotechnical risks (Tang et al. 2013, Wu 2013). Bivariate correlations express the relation between two soil properties, while multivariate correlations expand this probabilistic concept to more than two parameters. However, it is common to measure more than two soil parameters in close proximity during site investigation phase e.g., undrained shear strength from undrained triaxial tests and index properties, such as liquid (LL) or plastic (PL) limit. Multivariate probabilistic models are considered more representative, when data is available in a sufficient quantity, as the intrinsic and mechanical soil properties are correlated to a greater extent (Phoon and Ching 2015). Various multivariate probability distributions have been constructed based on regional and global databases (Ching and Phoon 2012b, Liu et al. 2016, Zang et al. 2020). Ching and Phoon (2012b) constructed a multivariate distribution of five geotechnical properties related to soil strength (intact and remolded undrained shear strength), stress level (pre-consolidation and in-situ stress) and soil classification (Liquidity index) in clays based on a global database. The resulting distribution was compared with another independent database of structured clays and empirical equations, and good predictability of the parameters in the independent database was achieved with the constructed model. Liu et al. (2016) constructed a multivariate normal model based on a regional database from the Quaternary clays in Jiangsu province, China with their geological formations to be Marine, Yangtze River Delta, Floodplain of Long River, Floodplain of Abandoned Yellow River, Lagoon of Lixia River and Lagoon of Taihu Lake. The model contains five parameters (resilient modulus, tip resistance, sleeve friction, water content and dry density) and it aimed to accurately predict the resilient modulus based on different testing indices. Zhang et al. (2020) established a multivariate ground model in Shanghai soft clay conditions, consisting of 11 soil parameters (four index and seven mechanical parameters) obtained by vane shear tests and penetration resistance from the CPT.

This paper aims to formulate multivariate models using a regional geotechnical database for Copenhagen's Quaternary glacial deposits (Upper clay till), which is a soil type less studied from this point of view and for which fewer empirical correlations exist. The database contains index and mechanical properties of soil (compressibility, shear strength and stiffness) from laboratory tests performed as part of the ground investigation campaign of the Copenhagen metro (M3 and M4 lines) and the Svanemøllen Skybrudstunnel. As the database was sparse and incomplete, it was considered more reliable to assess two different multivariate models based on a) triaxial and soil classification tests and b) oedometer and soil classification tests. The constructed multivariate ground models are intended to be used for reliability design of deep excavations in the future and therefore, the selection of the parameters was in line with that goal. The parameters included in these models are required for the validation of Hardening-small strain stiffness (HS-small), which is the constitutive model selected to perform reliability analysis in the context mentioned (Schanz 1999). Specifically, two multivariate probability distributions showing the statistical dependence among five (deviatoric over mean effective stresses ratio at 10% axial strain $\left[\frac{q}{p}\right]_{\varepsilon_\alpha=10\%}$, over-consolidation ratio OCR, secant modulus at a reference stress of 100 kPa E_{50}^{ref}, initial void ratio e_o and liquid limit LL) and five (unloading oedometer modulus at a reference stress of 100 kPa $E_{oed(ur)}^{ref}$, over-consolidation ratio OCR, permeability index C_k, initial void ratio e_o and liquid limit LL) geotechnical parameters respectively were constructed based on this database. Finally, samples were generated from the constructed multivariate models and compared with the original database for validation.

2. Description of regional database

Extensive ground investigations were carried out as part of the construction of the Cityringen metro line (M3) and the extension to Sydhavn (M4) in Copenhagen. In general, data from 26 different testing sites were considered for the construction of the multivariate model A (triaxial and soil classification tests) and 15 testing sites for the construction of the multivariate model B (oedometer and soil classification). The ground conditions at these sites mainly consist of recent deposits (inhomogeneous fill layer), post and late glacial deposits locally, glacial deposits (sand till, clay till and meltwater sand and gravel) and Danian Limestone. Clay and sand tills are highly over-consolidated, which is reflected in the high variability in their strength and stiffness characteristics, while meltwater sand and gravel are characterized as dense to very dense deposits. Laboratory tests were carried out on samples taken from the testing sites so as to determine soil classification parameters (clay content CC, mean diameter d_{50} and uniformity coefficient C_u), index properties (liquid limit LL and plastic limit PL), strength (φ', c', ψ, c_u) and stiffness properties (E_{50}^{ref}) from triaxial tests and compressibility and stress level properties from oedometer tests ($E_{oed(ur)}^{ref}$, permeability C_k, pre-consolidation stress σ'_c). The samples that underwent the aforementioned tests were used as well for the estimation of the void ratio (e_o), bulk and dry density (ρ_b and ρ_d) and water content (w). The database used for establishing the multivariate models contains 33 sets of parameters for the multivariate distribution A (20 undrained and 13 drained triaxial tests) and 22 for the distribution B. Triaxial tests were carried out according to ISO/TS 17892-9 (2004) and oedometer tests according to BS 1377-5 (1990). All the specimens tested in triaxial apparatus were firstly saturated until a sufficient B-value was achieved, followed by a phase of anisotropic consolidation at the pre-consolidation stress, which was derived either by oedometer or vane test, an unloading phase to in-situ stress and finally a shearing phase.

3. Construction of the multivariate models

3.1. Methodology

The construction of a multivariate distribution becomes less complicated when the model parameters follow the same marginal distributions, especially for the case of the normal distribution, where probability density function (PDF) can be easily estimated by a mathematical equation. However, most of the geotechnical properties do not follow normal distribution and not necessarily the same type of distribution. When the dependent random variables do not follow the same marginal distribution, a translation method should be used Li et al. (2008). The most common method for the construction of a multivariate distribution is based on Gaussian copula concept (Nelsen 2006, Straub 2014). When the dependent variables can be described by Gaussian copula, then the Nataf transformation model can be used. It should be mentioned that other transformation models do exist such as the Rosenblatt and the Hermite polynomials transformation. However, the Nataf model was used for the construction of the multivariate distribution in this study because the Rosenblatt transformation requires that the joint PDF of a random vector is known beforehand, which is impossible in most engineering problems and it is highly affected by the transformational order of random vector (Shuang et al. 2008) and the application of the Hermite polynomial transformation technique requires large amount of data when it comes to dependent non-normally distributed variables.

3.1.1. Nataf transformation model

Consider a vector of random variables X= (X_1, X_2, ..., X_n) ^{T} with known marginal distributions and cumulative density functions (CDFs) $FX_i(x_i)$, i=1, …., n. The Nataf transformation model can be used to transform the correlated variables X_i from their physical space to the standard normal space and let Y_i denote the equivalent standard normal variable. The transformation can be achieved by Eq. (1).

$$X_i \rightarrow Y_i = \Phi^{-1}[FX_i(x_i)] \quad (1)$$

Where Φ^{-1} corresponds to the inverse standard normal CDF. The transformed variables are assumed to be jointly normally distributed, which allows to define a multivariate distribution based on the marginal distributions of the given parameters and a symmetric correlation matrix (C_{XX}) that can be estimated by using Pearson correlation coefficients from the database. The Pearson's correlation coefficient (ρ_{ij}) between a pair of variables can be defined as shown in Eq. (2), where ∈ denotes the expectation, μ the mean of each variable, σ the standard deviation and Cov (X_i, X_j) the covariance of the variables X_i and X_j.

$$\rho_{ij} = \frac{Cov(Xi,Xj)}{\sigma_{Xi}\cdot\sigma_{Xj}} = \in \left[\left(\frac{X_i-\mu\cdot x_i}{\sigma_{X_i}}\right)\cdot\left(\frac{X_j-\mu\cdot x_j}{\sigma_{X_j}}\right)\right] \quad (2)$$

The Pearson coefficient expresses the linear dependency between two random variables X_i and X_{i+1} and can vary within the range of [-1,1]. It is clear that if ρ_{ij}=0 then there is no correlation between the two variables while ρ_{ij}=±1 indicates that a perfect linear correlation exists between the two variables. Positive correlation coefficients indicate that an increase in one parameter leads to the increase of the other, while the opposite is observed when the correlation coefficient is negative.

The correlation matrix of the transformed variables is not the same as the correlation matrix of the uncertain variables X_i. For the estimation of the correlation matrix of the transformed variables Cholesky decomposition is applied. The correlation matrix should be positive-definite, which means that every eigenvalue should be positive, for establishing a valid multivariate distribution. However, reliability design methods in geotechnical engineering often require a large number of random variables. For this purpose, inverse Nataf transformation model is applied and dependent samples from the previously developed joint distribution are generated.

3.2. Marginal statistics

3.2.1. Reference stiffness modulus E_{50}^{ref}

Initially, the secant stiffness modulus corresponding to 50% of the ultimate deviatoric stress q_f was estimated based on drained triaxial tests. However, the secant modulus at a reference stress of 100 kPa (E_{50}^{ref}) is of interest since it is required for the soil calibration of HS small strain constitutive model and provides a stress independent stiffness parameter. The equation that relates the E_{50} estimated by the triaxial tests and the one at a reference stress of 100 kPa is shown in Eq. (3).

$$E_{50} = E_{50}^{ref}\left(\frac{\sigma'_3+c'\cdot\cot\varphi'_p}{\sigma'_{ref}+c'\cdot\cot\varphi'_p}\right)^m \quad (3)$$

Where σ'_3 is the minor principal stress that the triaxial test was carried out, c′ is the cohesion, φ'_p is the peak angle of shearing resistance and m is the stress exponent. A power fitting was carried out between E_{50}, and the fraction shown in the parenthesis of Eq. (3) so as to estimate the stress exponent for power law m. It can be seen in Eq. (3) that the effective strength parameters are required for the calibration of the model. Clay tills in Copenhagen are identified as transitional soils with properties between a typical cohesive and a typical cohesionless soil. In particular, both clay tills of low cohesion-high angle of shearing resistance (c′<10 kPa) and high cohesion-low angle of shearing resistance (c≈10-60 kPa) have been found at the testing sites. Even though the majority of the samples appear to belong to the first group an investigation was carried out by considering various observed pairs of c′-φ′. Finally, the median values of the effective strength parameters were selected and introduced in the estimation of m exponent. Fig. 1 illustrates the secant modulus E_{50} with respect to the normalized stress, where both axes are in logarithmic scale. The stress exponent m was considered to be constant in this study, and particularly a value of 0.9 was adopted. For the justification of the decision of a constant m value, different m values were considered and insignificant effect on the type of the fitted distribution of E_{50}^{ref} was observed. After the stress exponent m was derived, the reference secant modulus was back calculated for each test by using Eq. (3). Hereafter, a statistical analysis was carried out as for the rest of the parameters.

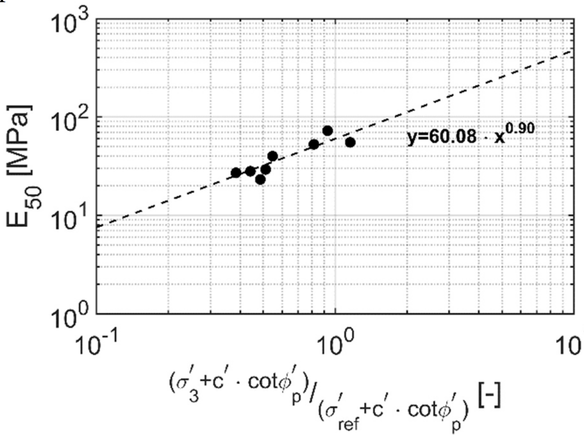

Figure 1. Secant modulus corresponding to 50% of q_f versus normalized stress level.

3.2.2. Reference oedometer modulus $E_{oed(ur)}^{ref}$

A similar process to the one described previously, has been followed for establishing the marginal information of the reference oedometer modulus. The oedometer modulus was estimated based on the first unloading-reloading stage, since the application of interest is excavation problems (unloading). As unloading was carried out in stages, the oedometer modulus was calculated for each unloading stage individually (tangent), which in turn increases the sample population at the stress space and eliminates the bias that might be introduced by the limited available oedometer tests. By using the equation of HS small strain stiffness constitutive model for oedometer modulus, shown in Eq. (4), the power exponent m is firstly derived.

$$E_{oed\,(ur)} = E_{oed(ur)}^{ref}\left(\frac{\sigma'_1 + c'\cdot\cot\varphi'_p}{\sigma'_{ref} + c'\cdot\cot\varphi'_p}\right)^m \quad (4)$$

Where, σ'_1 is the vertical stress at each unloading increment. The tangent unloading oedometer modulus with respect to the normalized axial stress is shown in Fig. 2. The axial stress used in the equation is the final axial stress at each stage. It is important to highlight that m was found to be the same as the one estimated from the reference secant modulus (m=0.9). This reinforces the decision of considering m constant in the constructed multivariate distributions.

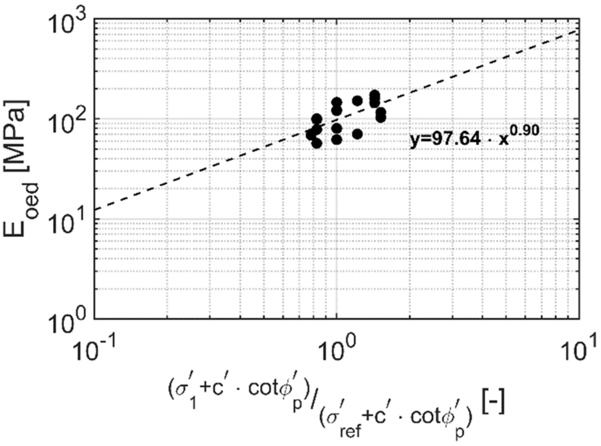

Figure 2. Tangent oedometer modulus E_{oed} estimated by the first unloading-reloading stage with respect to the normalized axial stress over the reference stress of 100 kPa.

3.2.3. Permeability change index C_k

The void ratio dependency of the permeability was taken into consideration in the multivariate model B. This can be achieved by specifying an appropriate value for the C_k parameter and the initial void ratio e_o. The equation that expresses the relationship among these three parameters is shown in Eq. (5). The C_k index corresponds to the slope of the linear fit between Δe and $\log(k/k_o)$ during primary loading. Six to eight points were used to estimate the C_k index for each test with the coefficient of determination R^2 to range between 0.60 and 0.95. After C_k was estimated, a statistical analysis was performed.

$$\log\left(\frac{k}{k_o}\right) = \frac{\Delta e}{C_k} \quad (5)$$

4. Results

4.1. Marginal statistics

The basic statistics of the multiple variables of the multivariate model A, including mean μ, coefficient of variation COV, maximum and minimum values are presented in Table 1. It can be seen that q/p′, e_o and LL are the least variable parameters (COV is smaller than 0.25) while OCR and E_{50}^{ref} are the most variable ones (COV is bigger than 0.35). Additionally, the fitted marginal distributions of the five parameters are also shown in Table 1. The fitted distributions with the histograms constructed by the regional database are presented later on.

Table 1. Basic statistics and marginal distibutions of the multiple variables for the multivariate model A

Parameters	Mean μ	COV	Max	Min	Marginal distribution
q/p′ [-]	1.63	0.11	2.19	1.37	Lognormal
OCR [-]	9.68	0.57	31.09	0.99	Rayleigh
E_{50}^{ref} [MPa]	49.21	0.37	77.09	16.02	Weibull
e_o [-]	0.39	0.23	0.630	0.27	Lognormal
LL [%]	24.65	0.16	35.7	19.10	Lognormal

Similarly, the basic statistics and the best fitted distributions of the five parameters of the multivariate ground model B are shown in Table 2. The least variable parameters in this case are $E_{oed(ur)}^{ref}$, e_o and LL while OCR and C_k are the most variable ones.

It is worthy to highlight that initial void ratio and liquid limit measurements are highly consistent between the two populations. The fitted distributions of e_o found to be different in the two constructed multivariate models which is mainly attributed to the smaller population of the multivariate model B.

Table 2. Basic statistics and marginal distributions of the five parameters for the multivariate model B

Parameters	Mean μ	COV	Max	Min	Marginal distribution
$E_{oed(ur)}^{ref}$ [MPa]	105.78	0.26	146.05	59.92	Weibull
OCR [-]	13.67	0.86	41.17	4.5	Lognormal
C_k [-]	0.015	0.51	0.03	0.007	Weibull
e_o [-]	0.43	0.17	0.63	0.31	Weibull
LL [%]	24.93	0.11	30.9	20.7	Lognormal

Note that because two separate multivariate models are developed due to lack of all tests considered at a same point, it is important to evaluate the overlap of the databases use in each subset to assess representativeness. The liquid limit measurements presented in Table 1 and Table 2 seem to be consistent. However, soil classification parameters, such as LL and CC, from the samples that were tested on Triaxial and Oedometer apparatus are additionally compared with the soil classification database that comprises of 142 grain size distribution curves. Fig. 3 and Fig. 4 illustrate the representativeness of the triaxial and oedometer database correspondingly. It can be seen that the samples used for establishing the two multivariate distributions can be considered representative. It should be also mentioned that both databases include tests of lower clay content (CC< 17%), where LL measurements were not available, and they potentially capture the southwest part of the data cloud. Even though the population used for the establishment of the multivariate functions were 33 and 22 respectively, the amount of datapoints shown in Fig. 3 and Fig. 4 is lower because there are tests carried out in samples taken from the same borehole and similar depth and as a consequence there have been assigned the same values of LL and CC or no LL measurement is reported.

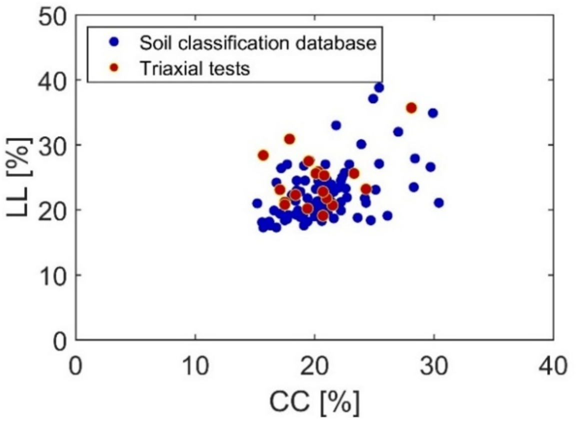

Figure 3: Representativeness of triaxial database.

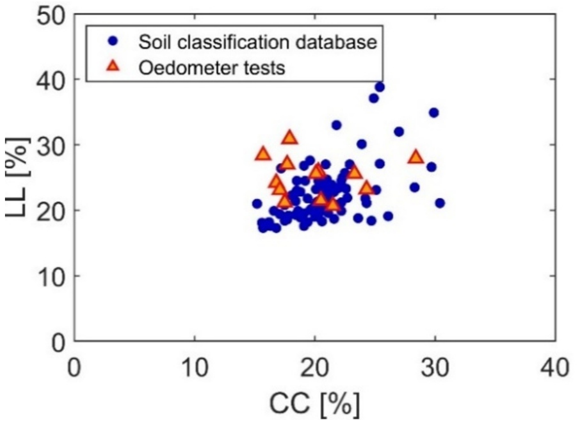

Figure 4: Representativeness of oedometer database.

4.2. Cross-correlation

The next step is to assess the correlation between multiple soil parameters, since it is not realistic to assume that geotechnical properties are not correlated. The Pearson's coefficients were estimated as it was described in Subsection 2.1.1 and the results are presented in Table 3. According to Evans (1996) recommendation for classifying the correlations among the variables based on the magnitude of Pearson's coefficients, E_{50}^{ref} is very strong correlated with e_o and strongly correlated with LL. Moderate correlation was found between LL and e_o. The remaining correlations are characterized as weak ($0.2 \leq |\rho| < 0.4$) or very weak ($|\rho| < 0.2$). An illustration of the correlations among the properties is presented in Fig. 5, where the diagonal plots display the histogram of the five variables with their fitted distributions.

The very strong correlation observed between E_{50}^{ref} and e_o is in agreement with previous stiffness correlations Kulhawy and Mayne (1990). The strong correlation between E_{50}^{ref} and LL can be explained by the void ratio dependency on LL and the correlation between E_{50}^{ref} and e_o. However, Kulhawy and Mayne (1990) reported an OCR dependency on the deformation properties of soils which has not been observed in the present case study as their correlation was found to be weak. The positive sign of the Pearson's coefficient though indicates that stiffness increases with OCR which is in line with Kulhawy and Mayne (1990).

Table 3: Cross correlation coefficients among the selected parameters for multivariate model A. (i) Undrained and Drained tests were used for the estimation of ρ_{ij} and (ii) Only Drained tests were used for the estimation of ρ_{ij}

C_{xx}	q/p´ (i)	OCR (i)	E_{50}^{ref} (ii)	e_o (i)	LL (i)
q/p´	1	-0.163	0.368	-0.112	-0.041
OCR	-0.163	1	0.325	-0.290	-0.251
E_{50}^{ref}	0.368	0.325	1	-0.931	-0.692
e_o	-0.112	-0.290	-0.931	1	0.534
LL	-0.041	-0.251	-0.692	0.534	1

The q/p´ ratio and OCR were found to not correlate well (moderately or better) with any of the other parameters, which was not anticipated. However, the correlations were improved when the peak q/p´ ratio was considered. In particular, the pairs of q/p´-e_o and q/p´-OCR were found to be moderately correlated (ρ_{ij}=-0.392 and ρ_{ij}=+0.383 respectively). A strong correlation was expected between q/p´ and LL, because of the existing empirical equations, that relate the φ´ or the C_u with the LL, the I_p and/or the CC (Ladd et al. 1977, Terzaghi 1996, Stark et al. 2005, Sørensen and Okkels 2013). However, this was not captured in the present database and the correlation was not improved when the peak q/p´ ratio was considered instead of q/p´ at ε_α=10%.

Similarly, the Pearson's coefficients among the five parameters for the multivariate model B are shown in Table 4. A visual representation of the correlation matrix is shown in Fig. 6, including the histograms with the marginal distributions of each parameter in the diagonal plots. The permeability index C_k is strongly correlated with the index properties (e_o and LL) and e_o is moderately correlated with LL which agrees with the observations from Tavenas et al. (1983), where the vertical permeability in clays found to correlate well with e_o, LL and CC. However, poor correlations were found among $E_{oed(ur)}^{ref}$, OCR and the rest of the parameters. A higher correlation coefficient was expected between $E_{oed(ur)}^{ref}$ and e_o due to the void ratio dependency on deformation properties discussed previously, however, the negative sign of the coefficient indicates that the expected relationship $E_{oed(ur)}^{ref}$ and e_o is captured properly (e.g., as e_o increases, $E_{oed(ur)}^{ref}$ decreases).

Table 4. Cross correlation coefficients among the selected parameters for the multivariate model B

C_{xx}	$E_{oed(ur)}^{ref}$	OCR	C_k	e_o	LL
$E_{oed(ur)}^{ref}$	1	0.011	-0.211	-0.147	-0.227
OCR	0.011	1	0.366	0.049	0.280
C_k	-0.211	0.366	1	0.601	0.679
e_o	-0.147	0.049	0.601	1	0.511
LL	-0.227	0.280	0.679	0.511	1

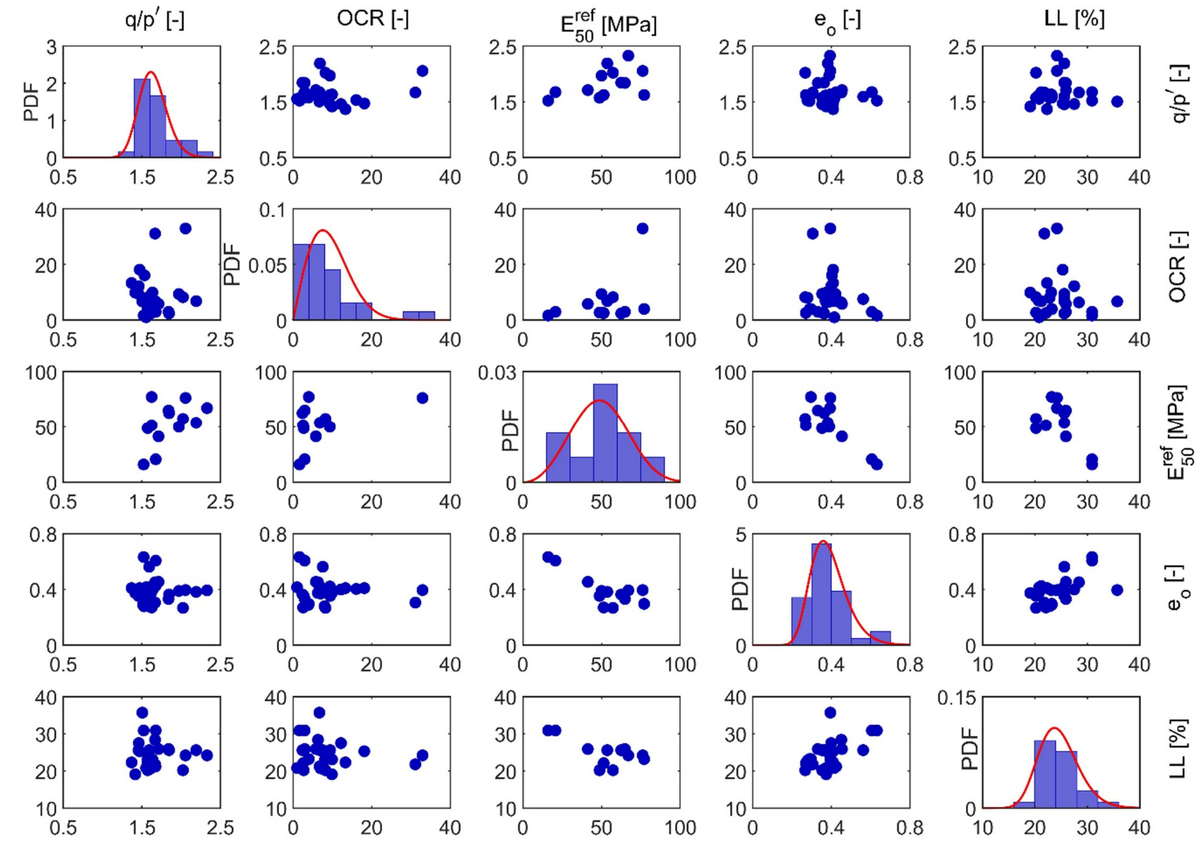

Figure 5. Scatter plots between the physical random variables X_i and X_j for multivariate model A and normalized histograms with probability density function for the five selected parameters. Marginal distributions are also displayed.

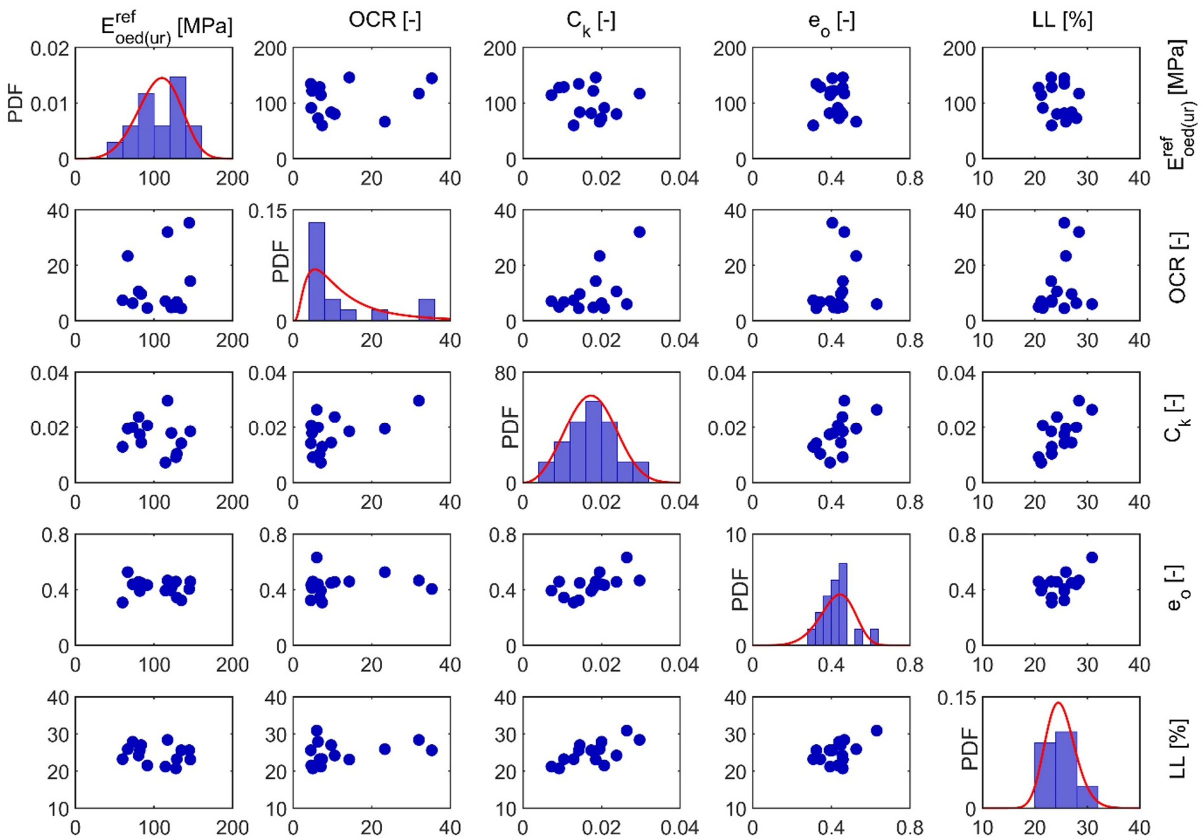

Figure 6. Scatter plots between the physical random variables X_i and X_j for multivariate model B and normalized histograms with probability density function for the five selected parameters. Marginal distributions are also displayed.

4.3. Validation of the multivariate distributions

The constructed multivariate distributions by applying the Nataf transformation model should be validated. The validation of the model in this study was carried out by generating artificial data points from the multivariate distributions and comparing them with the initial database. The comparison of 1000 simulated data with the original database is illustrated in Fig.7 for the multivariate model A and in Fig. 8 for the multivariate model B.

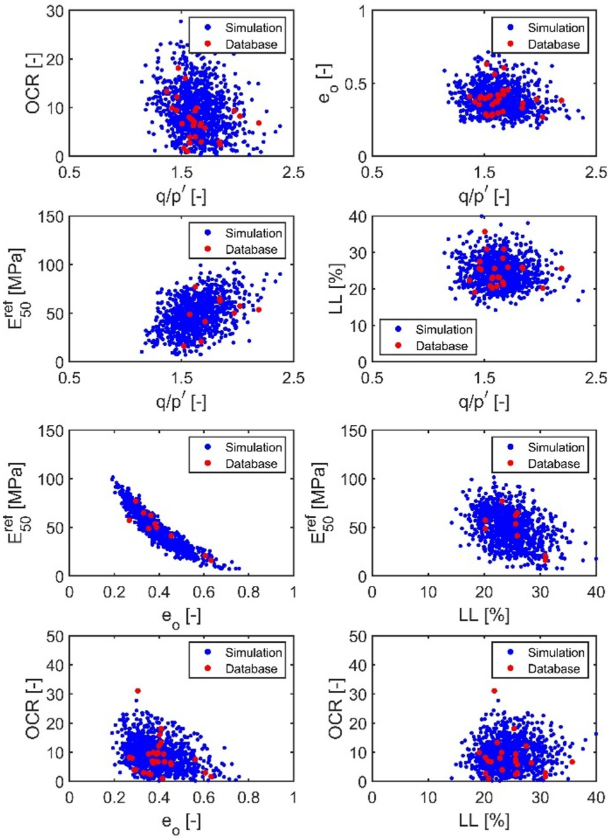

Figure 7. Comparison between the original database and 1000 simulated data based on multivariate model A.

It is clear that bigger scatter is observed in the generated samples of multivariate distribution B compared to distribution A. This is mainly associated with the bigger coefficient of variation of the parameters, which is in turn related to the smaller population used for the analysis. The COVs of the geotechnical parameters have a clear impact on the resulting scatter, e.g. minor scatter is observed between E_{50}^{ref} and e_o (see Fig. 7) where COVs are relatively small and Pearson's coefficient is the highest observed ($\rho_{ij} \approx -0.93$). A much wider scatter can be observed in the case of $E_{oed(ur)}^{ref}$ and OCR, where OCR has a big COV.

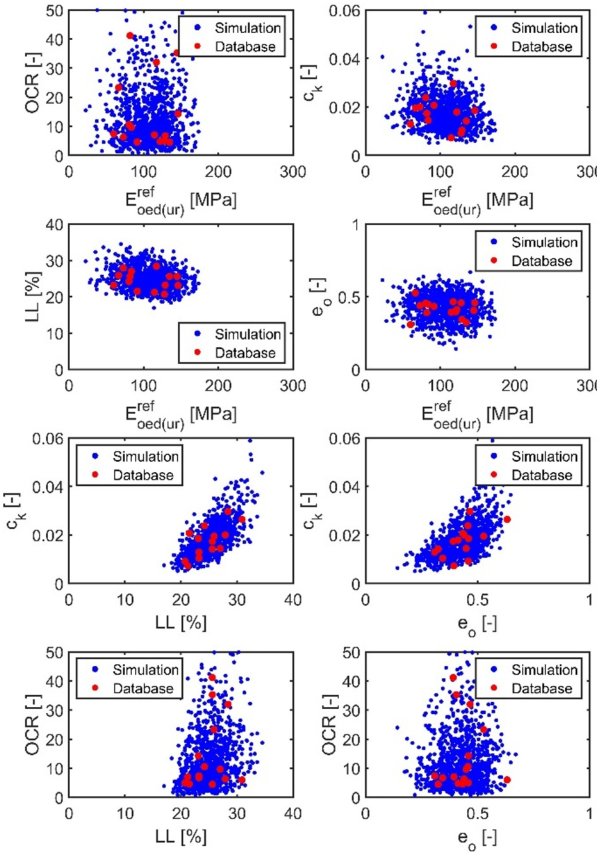

Figure 8. Comparison between the original database and 1000 simulated database based on multivariate model B.

5. Conclusion

This study assembles two databases for Copenhagen upper clay till, whose data points were collected from site investigation reports of 25 sites in Copenhagen that were carried out as part of the construction of the last two metro lines (M3 & M4) and 1 construction site of the Svanemøllen Skybrudstunnel. The databases comprise of geotechnical properties that were derived from anisotropically consolidated drained and undrained triaxial tests, oedometer and soil classification tests.

The concept of the multivariate distribution was applied for five parameters of each database so as to investigate and model the correlations among them. The multivariate model A includes over-consolidation ratio (OCR), secant modulus at reference stress of 100 kPa (E_{50}^{ref}), the ratio of deviatoric and mean effective stresses $(q/p')_{\varepsilon a=10\%}$, initial void ratio ($e_o$) and liquid limit (LL) and multivariate distribution function B comprises of $E_{oed(ur)}^{ref}$, e_o, LL, OCR and C_k. To do so, the basic and marginal statistics of each parameter was firstly evaluated. The most variable parameters found to be C_k and OCR. The basic statistics of e_0 and LL were found to be similar among the databases, which indicates representativeness of the obtained results. Furthermore, the results have been compared with a bigger database consisting of 142 soil classification tests and since the basic statistics derived for both distributions are close to these statistical properties, this indicates that the distributions obtained can be considered representative of the clay till behavior.

By comparing the correlations between pairs in the multivariate distribution functions, very strong correlation found between E_{50}^{ref} and e_o, while strong correlations observed in the pairs of E_{50}^{ref} – LL, C_k-e_o and C_k-LL. The rest of the correlations were characterized as weak to very weak, except the correlation between e_o and LL, which was found to be moderate. The obtained correlations were discussed with reference to existing relationships reported in the literature. It should be stressed out that the correlations observed between deformation and compressibility properties of clay till with index and classification properties are in line with reported correlations while the strength properties of clay till found to not follow the existing relationships.

Finally, the multivariate distribution functions were constructed by applying the Nataf transformation model since the selected parameters are described by different marginal distributions. Models were validated by generating artificial samples, then they were compared with the original database and satisfactory results were obtained. The constructed multivariate distributions can act in the future as prior for Bayesian update. However, further validation of the multivariate distributions is intended to be carried out prior to any applications.

Acknowledgements

The authors are grateful to Metroselskabet and ARUP for providing the laboratory geotechnical tests and allowing publication.

References

Baecher, G. B., Christian, J. T. "Introduction-Uncertainty and Risk in Geotechnical Engineering." In: Reliability and Statistics in Geotechnical Engineering, John Wiley & Sons Ltd, West Sussex, England, 2005, pp. 3-16.

British Standards Institution (BSI) " BS 1377-5: Methods of test for Soils for civil engineering purposes- Part 5: Compressibility, permeability and durability tests." BSI, London, United Kingdom, 1990.

Ching, J, Phoon, K.-K. "Value of Geotechnical Site Investigation in Reliability-Based Design.", Adv. Struct. Eng., 15 (11), pp. 1935–1945, 2012 https://doi.org/10.1260/1369-4332.15.11.1935

Ching, J, Phoon, K.-K. "Modelling parameters of structured clays as a multivariate normal distribution.", Can. Geotech. J. 49, pp. 522-545, 2012. https://doi.org/10.1139/t2012-015

Evans, J. "Straightforward statistics for the behavioural sciences.", J. Am. Stat. Assoc., 91 (436), pp. 1750-1751, 1996. https://doi-org.proxy.findit.cvt.dk/10.2307/2291607

International Organization for Standardization (ISO) " ISO/TS 17892-9: Geotechnical investigation and testing- Laboratory testing of soil — Part 9: Consolidated triaxial compression tests on water saturated soils." ISO, Geneva, Switzerland, 2004.

Kulhawy, F. H., Mayne, P. W. "Manual on Estimating Soil properties for Foundation Design.", Electric Power Research Institute EL-6800, Project 1493-6, Electric Power Research Institute, Palo Alto, Calif., 1990.

Ladd, C. C, Foott, R., Ishihara, K., Schlosser, F. and Poulos, H. G. "Stress-deformation and strength characteristics, state of the art report. " In: Proceedings of the 9th International Conference on Soil Mechanics and Foundation Engineering, Tokyo, Japan, 1977, 2, pp. 421–494.

Liu, S., Zou, H., Cai, G., Bheemasetti, T. V., Puppala, A. J., Lin, J. "Multivariate correlation among resilient modulus and cone penetration tests parameters of cohesive subgrade soils.", Eng. Geol., 209, pp. 128-142, 2016. https://doi.org/10.1016/j.enggeo.2016.05.018

Luo, Z., Di, H., Kamalzare, M. and Li, Y. "Effects of soil spatial variability on structural reliability assessment in excavations.", Undergr. Space., 5 (1), pp. 71-83, 2018. https://doi.org/10.1016/j.undsp.2018.10.004

Nelsen, R. B. " Methods of Constructing Copulas.", In An Introduction to Copulas, edited by P. Buhlmann, P. Diggle, U. Gather and S. Zeger, Springer Science +, New York, USA, 2006, pp. 51-105.

Phoon, K. K., Ching, J. "Constructing multivariate distributions for soil parameters.", In Risk and Reliability in Geotechnical Engineering, edited by K. K. Phoon and J. Ching, Taylor & Francis Group, Sound Parkway NW, 2015, pp. 3-72. https://doi-org.proxy.findit.cvt.dk/10.1201/b17970

Phoon, K. K., Kulhawy, F. H. " Characterization of geotechnical variability. ", Can. Geotech. J., 36, pp. 612-624, 1999. https://doi.org/10.1139/t99-038

Schanz, T., Vermeer, P. A., Bonnier, P. G. The Hardening Soil Model: Formulation and Verification. Presented at Beyond 2000 in Computational Geotechnics-10 years of PLAXIS, Balkema, Rotterdam, 1999. https://doi.org/10.1201/9781315138206

Sørensen, K. K., Okkels, N. " Correlation between drained shear strength and plasticity index of undisturbed overconsolidated clays.", In: Proceedings of the 18th International Conference on Soil Mechanics and Geotechnical Engineering, Paris, France, 2013, pp. 1–6.

Stark, T. D., Choi, H., McCone, S. " Drained shear strength parameters for analysis of landslides." J. Geotech. Geoenviron. Eng., 131 (5), pp. 575–588, 2005. https://doi.org/10.1061/(ASCE)10900241(2005)131:5(575)

Straub, D. "Value of information analysis with structural reliability methods.", Struct. Saf., 49, pp. 75-85, 2014. https://doi.org/10.1016/j.strusafe.2013.08.006

Li, H. S., Lu, Z. Z., Yuan, X. K. "Nataf transformation-based point estimate method.", Chin. Sci. Bull., 53 (17), pp. 2586-2592, 2008. https://doi.org/10.1007/s11434-008-0351-0

Tang, X.-S., Li, D.-Q., Rong, G., Phoon, K.-K., Zhou, C.-B. "Impact of copula selection on geotechnical reliability under incomplete probability information.", Comput. Geotech.,49,pp. 264-278, 2013. https://doi.org/10.1016/j.compgeo.2012.12.002

Tavenas, F., Jean, P., LeBlond, P. and Leroeil, S. "The permeability of natural soft clays. Part II, Permeability Characteristics.", Can. Geotech. J., 20 (4), pp. 645-660, 1983. https://doi.org/10.1139/t83-073

Terzaghi, K., Peck, R. B, Mesri, G. "Soil Mechanics in Engineering Practice.", 3rd edn. Wiley-Interscience, New York, NY, USA, 1996.

Wu, X. Z. "Probabilistic slope stability analysis by a copula-based sampling method." Comput. Geosci.,17 (5), pp. 739-755, 2013. 10.1007/s10596-013-9353-3

Zhang, D., Zhou, Y., Phoon, K. K., Huang, H. "Multivariate probability distribution of Shanghai clay properties.", Eng. Geol.,273, 2020. https://doi.org/10.1016/j.enggeo.2020.105675