Proc. of the 8th International Symposium on Geotechnical Safety and Risk (ISGSR)
Edited by Jinsong Huang, D.V. Griffiths, Shui-Hua Jiang, Anna Giacomini and Richard Kelly

doi: 10.3850/978-981-18-5182-7_13-006-cd

ISGSR 2022

Prior Knowledge on Shear Strength and Compressibility of Glaciolacustrine Sediments in Northern Germany

Julia Sorgatz[1] and Martin Pohl[1]

[1]Geotechnical Engineering North, Bundesanstalt für Wasserbau,
Wedeler Landstraße 157, 22559 Hamburg, Germany.
E-mail: julia.sorgatz@baw.de
E-mail:martin.pohl@baw.de

Abstract:A major challenge of most geotechnical engineering projects is soil data scarcity. This paper aims at extending prior knowledge on shear strength and compressibility of Glaciolacustrine sediments of Northern Germany. Based on triaxial, incremental loading oedometer and complementary laboratory tests on specimens from 13 different locations, the inherent variability of shear strength and compressibility is analyzed; typical ranges and coefficients of variation are established. Prior to variability analysis, k-means clustering, a simple machine learning algorithm, is applied to distinguish soil types by their descriptive properties. This data-driven methodology serves the multivariate character of soil data and allows to provide data on the variability of soil strength and compressibility more accurately. It was found that plasticity index and clay content can be considered to distinguish different soil types. Moreover, it can be shown that mean and variability of shear strength and compressibility are clearly affected by the dominant soil type.

Keywords: soil variability, glaciolacustrine sediments, k-means clustering, multivariate data analysis.

1 Introduction

A major challenge of geotechnical engineering is soil data scarcity or the “the curse of small sample size” (Phoon, 2017) which impedes the selection of characteristic values based on statistical analyses. Commonly, this issue is tackled by experience, engineering judgement and local data repositories. But even with these resources, as outlined by Bond (2011), engineers may not be well trained at predicting the appropriate degree of caution needed to select the characteristic value of a geotechnical parameter. Thus, although still not integral part of everyday engineering practice, the advantages of reliability-based methods such as Bayesian inference in conjunction with prior knowledge are increasingly recognized to account for uncertainty inherent to soil parameters (e. g., Phoon and Kulhawy 1999a, 1999b, Wang et al. 2016, Phoon 2017).

Despite their local uniqueness (Phoon 2019), point statisticsof soils have been investigated by various authors, often for particular applications (e. g. Lumb 1966, 1974, Phoonand Kulhawy 1999a, 1999b, Uzielli et al. 2006). However, as pointed out by Löfman and Korkiala-Tanttu (2019), typical ranges of soil parameters provided in literature can be improved by accounting for local soil characteristics, e. g. materialgenesis. For North German Glaciolacustrine sediments, few information on typical values have been published (Ehlers et al. 2011, Kausch 2020), which, do not cover the materials’ inherent variability.

The presented work focuses on the analysis of shear strength and compressibility of Glaciolacustrine sediments of Northern Germany. Based on triaxial, confined compression and complementary laboratory tests from 13 different locations the inherent variability of shear strength and compressibility is analyzed; typical ranges and coefficients of variation are established. Prior to variability analysis, a simple machine learning algorithm is applied to distinguish different sediments on the basis of their classification properties.

2 Data and methodology

2.1 Characteristics of Glaciolacustrine sediments in Northern Germany

Besides Marine clay and boulder clay, Glaciolacustrine sediments are typical soils in Northern Germany. These sediments were deposited into lakes formed from glacial erosion or deposition. In the area of Northern Germany three glaciation periods are known: Elster glaciation, Saaleglaciation, Weichsel glaciation; the oldest glaciation, the Elsterian, reached furthest south. Only in Western Europe, the ice sheet of the second glaciation, the Saalian, advanced beyond the Elsterian limits. During the last glaciation, the Weichselian, the ice sheet did not cross the Elbe River (Ehlers et al. 2011).

Glaciolacustrine sediments are cohesive soils. By grain size analysis, they are classified as weakly sandy to sandy clays or silts (DIN EN ISO 14688-1:2020-11). Based on their plasticity index I_P and liquid limit w_L, they are classified as CL, OL,ML or CH (DIN 18196:2011-05). With w_L ranging from 20 % to 90 % (on average 46 %), the plasticity of either clay and silt ranges from low to high. Due to local lignite streaks and lenses, Glaciolacustrine sediments are weak to moderate organic; annealing losses between 1 % and 18 % (on average 4 %) are found.

2.2 Shear strength and compressibility properties

This section briefly introduces the most important tests and parameters used for subsequent analyses. For a detailed description of the employed laboratory tests, it is referred to the respective standards. A summary of the employed test data is provided in Table 1. For few locations limited data are available emphasizing the importance of prior knowledge for future engineering projects.

Table 1. Classification properties of the studied sites (assessment of inherent variability).

Site	Depth	Specimens	Clay content	Organic content	Water content w_n	Plasticity index I_P
--	in m	--	in %	in %	in %	--
Brunsbüttel	37 - 40	3	8.0 - 37.0	2.6 -5.6	18.3 - 24.3	0.12 - 0.31
Zerben	6 - 13	2	4.0 - 52.0	5.0	26.2 - 35.9	0.16 - 0.48
Levensau	8 / 37 - 42	4	4.0 - 50.0	1.9 - 4.2	19.2 - 20.3	0.13 - 0.32
Kiel-Holtenau	15 - 19	2	25.0 - 70.0	2.2 - 3.8	18.6 - 21.3	0.31 - 0.32
Kiel-Friedrichsort	17 / 32 - 33	3	14.0 - 26.0	2.4 - 2.5	32.2	0.13
Hunte	5 - 14	2	64.0 - 76.0	7.3 - 7.7	32.2 - 34.3	0.43 - 0.51
Steinhavel	4 - 10 / 21	3	13.0 - 38.0	1.3 - 7.9	24.5 - 29.7	0.16 - 0.53
Ahse	6	2	31.0-36.0	4.0 - 5.1	21.9 - 28.0	0.22 - 0.31
Lauenburg	5 - 45	42	5.0 - 65.0	1.0 - 10.4	27.5 - 34.8	0.06 - 58.2
Niederfinow	2 - 25	17	3.0 - 27.0	2.1 – 5.3	18.5 - 27.5	0.04 - 0.33
Witzeeze	10 - 30	5	11.0 - 36.0	4.2 - 6.2	19.6 - 24.7	0.09 - 0.22
Ems	3 - 15	15	5.0 - 21.0	2.4 - 17.5	21.4 - 32.0	0.06 - 0.31
Niederfinow	10 - 15	2	9.0 - 40.0	--	21.6 - 24.7	0.66

Note: The standards valid at the time of testing apply. The tests were conducted between 1997 and 2013.

The machine learning algorithm is first applied to selected descriptive soil parameters. The liquid limit w_L is determined via the fall cone test; the plastic limit w_P by repeated rolling of an ellipsoidal-sized soil mass (DIN EN ISO 17892-12). The clay content is obtained from sieve and sedimentation tests (DIN EN ISO 17892-4). The organic content results from loss-on-ignition tests (DIN 18128).

By means of incremental loading oedometer tests (DIN EN ISO 17892-5), the compressibility of the soil is investigated. A cylindrical sample is deformed uniaxially. A metal ring prevents the specimen from deviating sideways. A specimen was commonly tested against eight load levels which were doubled after each step and ranged from 17.1 kN/m² to 1021.1 kN/m². Then, the load was relieved to 17.1 kN/m², before the specimen was reloaded to 2040.4 kN/m². In total, results of 90 tests are available for analyses. From the initial loading, the stress-dependent constrainedmodule E_s is obtained as the ratio of change in stress and change in vertical deformation. In the same manner, the stress-dependent constrained reloading module $E_{s,r}$ is obtained from the reloading cycle.

Amongst others, triaxial tests (DIN EN ISO 17892-9) allow to determine the effective shear parameters, effective cohesion c' and effective friction angle ϕ'. The presented data encompasses isotropic consolidated drained triaxial compression (CIDC) and isotropic consolidated undrained triaxial compression (CIUC) tests. In total, 88 tests with three sub-specimens each were conducted; 33 tests of CIUC and 55 tests of CIDC.

2.3 K-means clustering

During field classification, different sedimenttypes may not always be clearly distinguishable or it is ambiguous which type dominates the specimen. However, this distinction may be important to provide accurate information on the soil characteristics. Machine learning tools can assist in classifying sediments based on objective criteria.

K-means clustering is one of the simplest unsupervised machine learning algorithms. It belongs to a family of algorithms which were developed independently by researchers from different disciplines (MacQueen 1967, Steinhaus 1956, Lloyd 1982). Main advantages are its simplicity and its scalability to different sample sizes. K-means clustering partitions n data points in k clusters based on their Euclidean distance to the nearest mean, the cluster centroid.

In simplified terms, the algorithm has three main steps: Firstly, the number of cluster centroids must be provided by the user. The learning process then starts with a group of randomly selected centroids. Subsequently, the algorithm changes the positions of the centroids iteratively until either the difference between old and new centroids reaches a threshold or the defined number of iterations has been reached. The new centroids are computed as the mean value of all of the samples assigned to each previous centroid (Pedregosa et al. 2011).

The presented k-means clustering analyses use Python with the machine learning tools provided by the package scikit-learn(Pedregosa et al. 2011). To account for the different scales of the soil characteristics, the data was normalized before running the analyses. For normalization the L² vector normalization scheme was employed, which is based on the distance of the vector coordinate from the origin of the vector space.

3 Prior knowledge on compressibility and shear strength

3.1 Results of k-means clustering analyses

Figure 1visualizes the results of the k-means clustering analyses. Based on normalizedw_L and normalizedI_P two clusters are identified. When using normalized clay content and normalizedI_P three clusters are found. In the case of the two-cluster solution, one cluster features material of moderate w_L and moderate I_P, whereas the other group features material of high w_L and high I_P. The three-cluster solution is divided in clusters of low, moderate and high clay content and I_P.

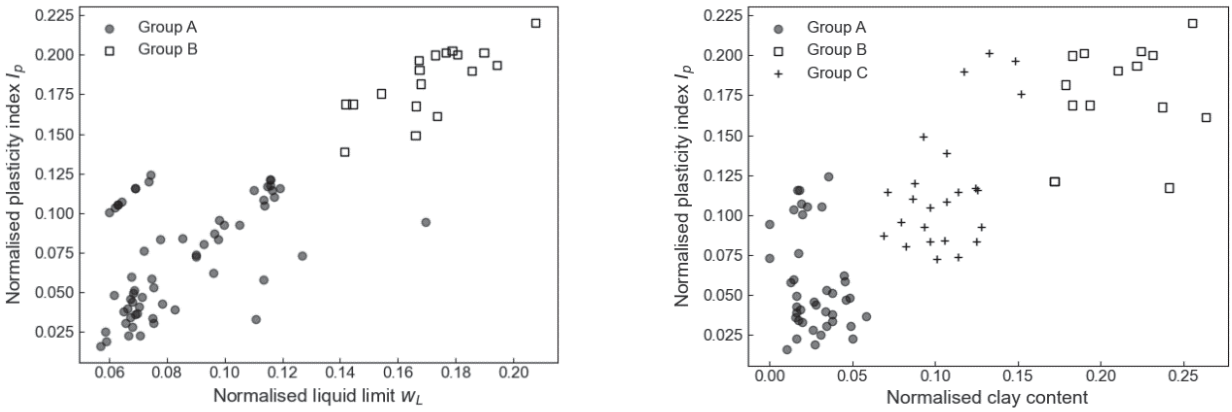

Figure 1. Results of K-means clustering analyses; descriptive properties of the Glaciolacustrine sediments are normalized.

The clusters are validated using further data characteristics. It can be shown that the two-cluster solution corresponds well with the stratigraphic units (Figure 2). In Casagrande's plasticity chart, it can be observed that sediments that were deposited during Weichsel and Saale glaciation are classified as Group A; sediments deposited at the end of the Elster glaciation are classified as Group B. Weichselian and Saalian sediments cannot be clearly distinguished in Casagrande's plasticity chart. Additional parameter investigations did not provide a satisfactory differentiation either. Reasons for this may be, e. g., uncertainties introduced during testing.

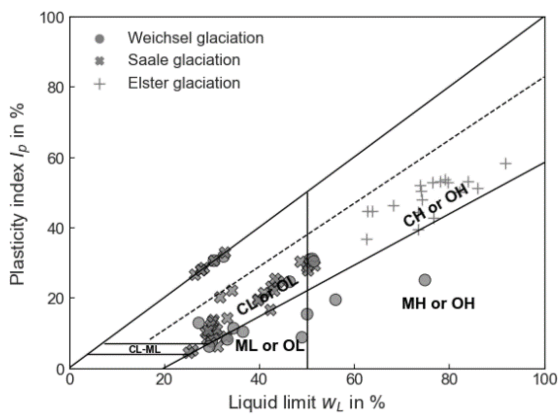

Figure 2. Two-cluster solution plotted against Casagrande's plasticity chart.

The presented k-means clustering analyses show promising results. Yet, it must be noted that the available data of Elsterian sediments primarily contains specimens of the so-called Lauenburg clay, which is a rather distinctclay with varying amounts of silt. In the case of the Saale and Weichsel sediments, the stratigraphic units cannot always be gathered from the geotechnical reports; for single specimens, no information on stratigraphy is available. Since bedload analyses were only carried out in few cases,the existing classification is based on the judgement of senior engineers who are familiar with the regional geology. Still, this may result in erroneous categorizations. Furthermore, it should be noted that k-means clustering is a simplealgorithm which assumes that data belonging to a cluster are located circular around the centroid. However, the clusters may have different shapes.

The differentiation into three clusters cannot be justified with the existing data. This does not necessarily mean that this classification is not accurate, but, at present, it is not clearly supported by the existing data. In the following, the variability of Glaciolacustrine sediments is thus analyzed within the two clusters (see Table 2).

Table 2. Point statistics of classification propertiesfor the two detected clusters.

	Clay content in %	Organic content in %	w_L in %	w_P in %	I_P in %
Group A (Weichselian / Saalian sediments)					
count	61	43	61	53	53
mean	16.44	3.84	37.38	21.18	18.98
standard deviation (std)	14.21	1.82	10.23	7.02	9.01
coefficient of variance (COV)	0.86	47.24	27.36	33.16	47.49
Group B (Elsterian sediments)					
count	18	8	18	18	18
mean	53.47	6.52	75.56	27.02	48.54
standard deviation (std)	14.72	2.65	7.90	5.25	5.61
coefficient of variance (COV)	0.28	40.64	10.45	19.43	11.55

3.2 Typical ranges of mean and variability of compressibility properties

Due to the geological history of origin, it can be assumed that the material was covered by a 500 m to 1000 m ice cover. Thus, the material is geologically preloaded and over-consolidated. On average, a large variability of $E_{s,r}$ is observed within one load level. Among other things, this may be due to inhomogeneous soil specimens and the limited specimen geometry. The water content of the test specimens at the start of the test ranged between 15 % and 40 %; the porosity ranged between 30 % and 75 %.

A linear least-squares regression was conducted to determine the stress-dependent $E_{s,r}$as a function of the mean soil stress σ_m(see Figure 3). R^2 and the mean standard errors (MSE) of the regression parameters are used as a measure of how well the observed outcomes are predicted by the model. A moderate fit is observed as a result of strongly scattering measurements. In the two groups, R^2 is neither significantly reduced nor increased compared to the complete data (Table 3). Thus, there is no loss of fit while the explanatory power of the regressions increases.

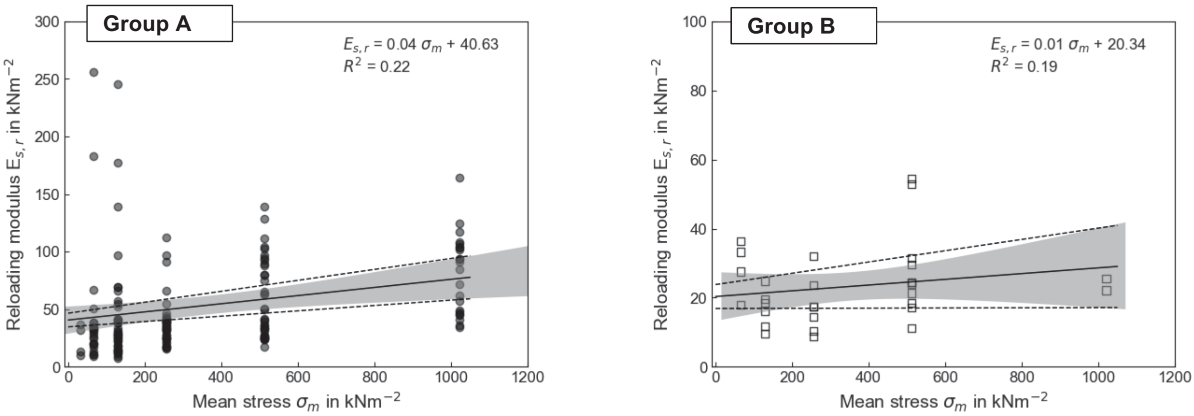

Figure 3. Stress-dependent constrained reloading modulus $E_{s,r}$ over mean soil stress σ_m with 95 % confidence interval (grey shaded areas) of the regression. The dashed lines indicate potential characteristic values.

The variability of the constant *b* expressed via the MSE is higher in Group A than in Group B, whereas the MSE of the slope *a* is the same in both groups (see Figure 3 and Table 3). Another measure for the variability is the sum of squares total (SST) which is the squared differences between the observed dependent variable and its mean. In simple terms, it can be considered as a measure of the total variability of the dataset while considering its linear trend. In the case of the herein presented investigations, the SST is the largest in the total data set (SST = 494526), slightly smaller in Group A (SST = 468032) and significantly reduced in Group B (SST = 3507).

For the definition of characteristic values in engineering practice, the presented equations may be shifted either parallel or slightly inclined along the y-axis until they meet the desired safety level. One potential approach is to add and subtract the MSE to and from the mean values of *a* and *b*, see dashed lines in Figure 3. This allows to account for the uncertainty associated with the regression. Naturally, the uncertainty in the complete data set and in Group A is larger than in Group B. Due to the variability of test results,this approach may yield a too optimistic or too conservative estimate. Depending on the task at hand, in practice, one might therefore choose characteristic values based on engineering judgment.

In summary, Figure 3 and Table 3 illustrate that the differentiation of stratigraphic units by means of cluster analyses allows to provide soil characteristics more accurately. Without differentiation the compressibility of Group A sediments is likely to be underestimated. For Group B, on the other hand, the increase of $E_{s,r}$ and its variability are likely to be overestimated. If the mean minus the standard error of a in Group B is considered, it can even be concluded that the increase, i. e. the dependence of $E_{s,r}$ on σ_m, is not significant.

Table 3. Regression statisticsfor linear regression with $E_{s,r}$ [kN/m²]=$a\cdot\sigma_m + b$.

Parameter	a	b	R^2
Complete data set			
mean	0.03	36.84	0.21
mean standard error (MSE)	0.01	5.27	--
Group A (Weichselian / Saalian sediments)			
mean	0.04	40.63	0.22
mean standard error (MSE)	0.01	5.95	--
Group B (Elsterian sediments)			
mean	0.01	20.34	0.19
mean standard error (MSE)	0.01	3.46	--

3.3 Typical ranges of mean and variability of shear strength properties

Firstly, it is emphasized that the individual test results, e. g., stress-strain relation for each sample, should always be consulted for the determination of shear parameters. In addition, experimental data have shown that the strength envelop for soil is nonlinear. Nevertheless, the linear Mohr–Coulomb strength parameters are widely applied in engineering practice and therefore also used for the following analyses.

Unfortunately, the linear regression with the extended shear diagram (p'-q - diagram) gives negative values of c'. This physically implausible negative c' may result from minor deficiencies during the test procedure and may benefit from a more sophisticated analysis and differentiation. Nevertheless, in the following, information on the variability of ϕ' and c' are derived from the statistical analysis of theindividual test series. For this purpose, each test series consisting of three tests each is evaluated. Then, the resulting ϕ' and c' are statistically described. In total, 61 test series are evaluated. The water content of the specimens at the start of the tests ranged between 18 % and 50 %; the porosity between 33 % and 50 %.

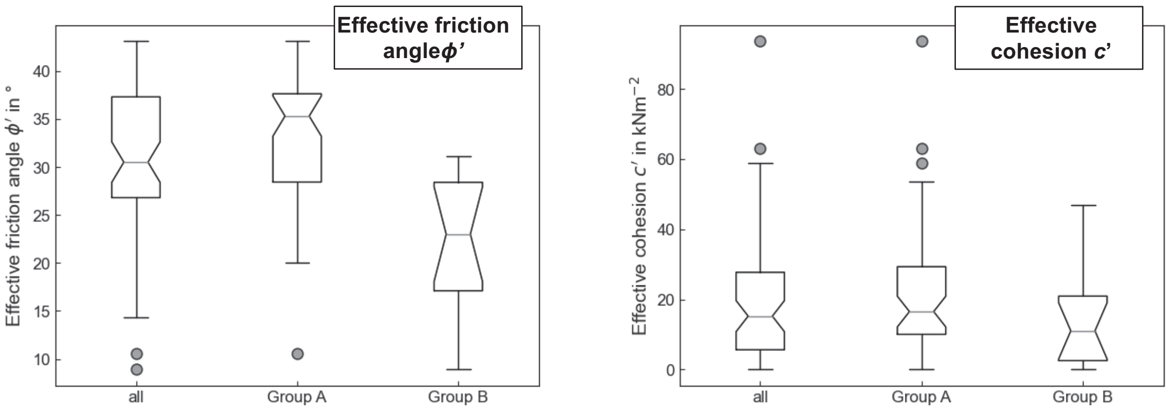

Figure 4. Boxplots of effective friction angle ϕ' (left) and effective cohesion c' (right). In the box 50% of the data are found. The notch shows the median; the whiskers correspond to 1.5 x interquartile range. The circles represent outliers.

Table 4. Summary of point statistics for effective shear strength parameters ϕ' and c'.

	Complete data set		Group A (Weichselian / Saalian sediments)		Group B (Elsterian sediments)	
	ϕ' in °	c' in kN/m²	ϕ' in °	c' in kN/m²	ϕ' in °	c' in kN/m²
count	61	61	48	48	13	13
mean	30.63	20.06	32.78	21.77	22.68	13.71
standard deviation (std)	7.78	18.21	6.64	18.82	6.37	14.01
coefficient of variance (COV)	0.25	0.91	0.20	0.86	0.28	1.02

Again, different material properties are observed within the two groups and the complete data set, but the differentiation into clusters does not reduce the group-inherent variability significantly (see Figure 4 and Table 4). However, it can be assumed that if not differentiated between the stratigraphic layers, the shear strength of Group B (Elsterian sediments) is overestimated; whereas the shear strength of Group A is underestimated (Weichselian /Saaliansediments). To reduce the variability, it may be beneficial to differentiate the clusters further, especially in Group A, e. g., as indicated in Figure 2, which, however, is not fully supported by the present data.

4 Conclusions and outlook

This paper presents a data-driven methodology for a differentiation of soil types that serves the multivariate character of soil data. With the machine learning algorithm k-means clustering two soil clusters with different material properties are identified.Statistics of shear strength and compressibility are determined for each soil type.

The statistical analyses show that a differentiation into soil types reduces the variability within a cluster and, thus, allows for a more precise description of the material. It can be shown that the statistics are affected by the dominant soil type.In the case of oedometer tests, it might be feasibleto use specimens larger in diameter and height to improve the results. In the case of the presented triaxial tests, realistic shear strength properties are only derived if the individual test results are consulted and ambiguous testsare neglected. It is therefore emphasized that clustering as one of many data-driven methods cannot overcome limitations of existing data analysesand uncertainties during testing. It cannot replace engineering judgement.However, it cancertainlyassistthe engineer in describing the soil more precisely, e.g., in order to define characteristic values.

Further investigations should validate the clusters that have been identified. This can be done, e. g., by including additional data or applying more advanced clustering algorithms. In addition, it may be beneficial to differentiate the clusters even further, which, however, is not fully supported by the present data. Based on supplementary studies, the determined mean and variability of the investigated soil types may be reviewed.On a broader basis, it is recommended to support generic databases which store soil data in a structured, machine readable manner. Only in this way, geotechnical engineering will benefit from recent and future developments in the field of data science.

References

Bond, A. (2011). A procedure for determining the characteristic value of a geotechnical parameter. Proc., Geotechnical Safety and Risk. ISGSR 2011. *Bundesanstalt für Wasserbau*, 419–426.

DIN EN ISO 14688-1. GeotechnischeErkundung und Untersuchung — Benennung, Beschreibung und Klassifizierung von Boden - Teil 1: Benennung und Beschreibung. *Beuth, Berlin*.

DIN EN ISO 17892-4. GeotechnischeErkundung und Untersuchung - Laborversuche an Bodenproben - Teil 4: Bestimmung der Korngrößenverteilung. *Beuth, Berlin*.

DIN EN ISO 17892-5. GeotechnischeErkundung und Untersuchung – Laborversuche an Bodenproben – Teil 5: ÖdometerversuchmitstufenweiserBelastung. *Beuth, Berlin*.

DIN EN ISO 17892-9. GeotechnischeErkundung und Untersuchung – Laborversuche an Bodenproben – Teil 9: KonsolidiertetriaxialeKompressionsversuche an wassergesättigtenBöden. *Beuth, Berlin*.

DIN EN ISO 17892-12. GeotechnischeErkundung und Untersuchung –Laborversuche an Bodenproben –Teil 12: Bestimmung der Fließ- und Ausrollgrenzen. *Beuth, Berlin*.

DIN 18128. Baugrund - Untersuchung von Bodenproben - Bestimmung des Glühverlustes.*Beuth, Berlin*.

DIN 18196. Erd- und Grundbau - Bodenklassifikation für bautechnischeZwecke. *Beuth, Berlin*.

Ehlers, J., Grube, A., Stephan, H.-J. andWansa, S. (2011). Pleistocene Glaciations of North Germany—New Results. *Quaternary Glaciations - Extent and Chronology - A Closer Look*, 15, 149–162.

Kausch, F. (2020). GeotechnischeCharakterisierung des Hamburger Baugrundes. 23.10.2020.

Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Trans. Inform. Theory* 28 (2), 129–137.

Löfman, M. and Korkiala-Tanttu, L. (2019). Variability and Typical Value Distributions of Compressibility Properties of Fine-Grained Sediments in Finland. Proc.,*7th International Symposium on Geotechnical Safety and Risk (ISGSR 2019)*, *Research Publishing (S) Pte. Ltd.*, 182–187.

Lumb, P. (1966). The Variability of Natural Soils. *Canadian Geotechnical Journal*, 3 (2), 74–97.

Lumb, P. (1974). Application of Statistics in Soil Mechanics. *Soil Mechanics-New Horizons, London*, 44–111.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proc.,5th Berkeley Symposium on Mathematical Statistics and Probability, Berkley, University of California Press* (Vol 1), 281-297.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O. et al. (2011). Scikit-learn. Machine Learning in Python.

Phoon, K.-K. (2017). Role of reliability calculations in geotechnical design. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, 11(1), 4-21.

Phoon, K.-K. (2019). Managing Risk in Geotechnical Engineering – from Data to Digitalization. Proc.,*7th International Symposium on Geotechnical Safety and Risk (ISGSR 2019), Research Publishing (S) Pte. Ltd.*, 13–34.

Phoon, K.-K. and Kulhawy, F.H. (1999a). Characterization of geotechnical variability. *Canadian Geotechnical Journal*, 36 (4), 612–624.

Phoon, K.-K. and Kulhawy, F. H. (1999b). Evaluation of geotechnical property variability. *Canadian Geotechnical Journal*, 36 (4), 625–639.

Steinhaus, H. (1956). Sur la division des corps matérielsen parties. Bulletin de l'Académie*Polonaise des Sciences. Classe 3* (12), 801–804.

Uzielli, M., Lacasse, S., Phoon, K.-K. and Nadim, F. (2006). Soil variability analysis for geotechnical practice. *Characterisation and Engineering Properties of Natural Soils, Taylor & Francis*.

Wang, Y., Cao, Z. and Li, D. (2016). Bayesian perspective on geotechnical variability and site characterization. *Engineering Geology*, 203, 117-125.