HENRY

Hydraulic Engineering Repository

Ein Service der Bundesanstalt für Wasserbau

Conference Paper, Published Version

Pohl, Carsten

Determination of characteristic soil values by statistical methods

Verfügbar unter/Available at: https://hdl.handle.net/20.500.11970/99593

Vorgeschlagene Zitierweise/Suggested citation:
Pohl, Carsten (2011): Determination of characteristic soil values by statistical methods. In: Vogt, Norbert; Schuppener, Bernd; Straub, Daniel; Bräu, Gerhardt (Hg.): Geotechnical Safety and Risk. ISGSR 2011. Karlsruhe: Bundesanstalt für Wasserbau. S. 427-434.

Determination of characteristic soil values by statistical methods

C. Pohl
ELE – Consulting Engineers Ltd., Essen, Germany

ABSTRACT: According to new European standard, EC 7, characteristic values are the fundamental input data for geotechnical verifications using the concept of partial safety factors. Characteristic values have to be determined for impacts as well as for soil properties. Although, European standard, EC 7, and German supplementary rules, specified in DIN 1054, give a consistent definition of characteristic values of a soil property, first experiences with the concept of partial safety factors have shown some problems with the interpretation of these definitions.

Besides, the geotechnical expert has significant discretion in the determination of characteristic soil values and this influences the results of geotechnical verifications. If results of field- and laboratory tests are available in an adequate sample size, statistical methods are an effective tool to determine characteristic soil values in a verifiable way and to get best possible information from realized site investigations. The following paper points out simple statistical methods and gives recommendations for their practical application. All procedures are demonstrated with the help of examples.

Keywords: characteristic value, laboratory test, site investigation, standard, statistical analysis

1 INTRODUCTION

In general, geotechnical stability could be verified, as shown in figure 1, by deterministic or probabilistic procedures. Deterministic procedures involve comparisons of single values for design impacts (F) and design resistances (R); however, probabilistic procedures take the distributions of all parameters (soil values, forces, etc.) and results in the probability p_F of a certain occurrence. In geotechnical practice probabilistic procedure is still only applied in exceptional cases. Regarding to deterministic procedures, the determination of single appropriate values for parameters that are characterized by statistical spread influences the results of geotechnical analysis essentially.
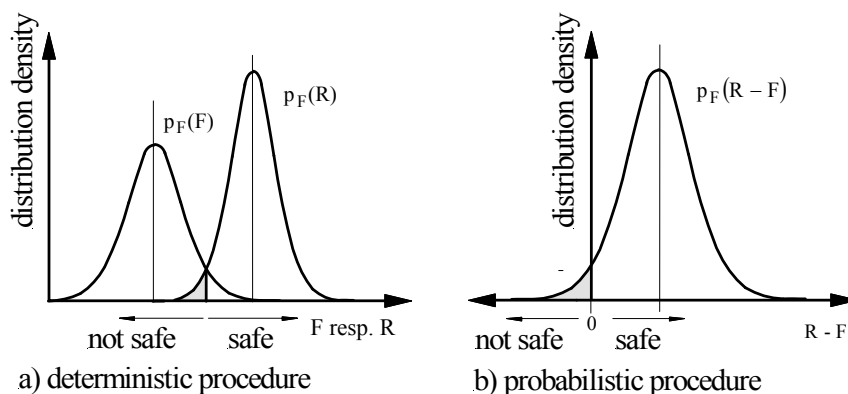

Figure 1. Deterministic vs. probabilistic procedure

Thus, with harmonization of European standards towards the concept of partial safety factors, as stipulated in EC 7 (DIN EN 1997), the phrase of characteristic soil values has been anchored in geotechnical verifications. Characteristic soil values should be representative values for soil properties of a homogeneous zone. Indeed, this approach is well known by geotechnical experts, but new European standards give much more precise definitions how to determine characteristic soil values than former national standards.

Finally, the determination of characteristic soil values is complex. As a consequence of its spatial variability and its local dispersion, there is much more uncertainty in soil properties than for other materials in civil engineering, like steel or concrete. Especially for spatial extensive projects or large depots, the soil is, as result of its history, often inhomogeneous and the properties – even of a homogeneous zone – spread widely. Furthermore, economic arguments may sometimes prevent an acceptable size of samplings.

As consequence to the restriction of DIN EN 1997 and DIN 1054 on qualitative definitions of characteristic soil values, geotechnical expert has a considerable discretion. Hence, it is hardly surprising, that characteristic soil values are still frequently determined by individual experience. Often, only the median values of test results are used. Nonetheless statistical methods can be an effective tool to determine characteristic soil values in a verifiable way, an adequate sample number presumed. Reasonably applied, statistical methods increase information content of site investigation and contribute to minimizing soil risk.

2 CHARACTERISTIC VALUES OF SOIL PROPERTIES

According to DIN 1055-100, the characteristic value is the fundamental representative value for actions. The analog intent must be taken for characteristic values of soil properties.

Nevertheless, no all-embracing definition of characteristic soil values is available up to now. Rules and specifications for determination of characteristic soil values have to be composed from several consistent standards. DIN EN 1997-1 gives within the general rules in section 2.4.3(5), the following basic definition:

"The characteristic value of a soil or rock parameter shall be selected as a cautious estimate of the value affecting the occurrence of the limit state."

This formulation has been chosen consciously. *Cautious estimate* should underline principle of caution. *Selection* is a hint to necessary geotechnical expertise. Reference to *limit state* emphasizes an obligatory regard to the respective construction and limit state. With this background, characteristic values of soil properties have to be determined in case of complex constructions in cooperation with the structural engineer. The basis of characteristic soil values must always be formed by field and laboratory tests, completed by local information and by experience if available.
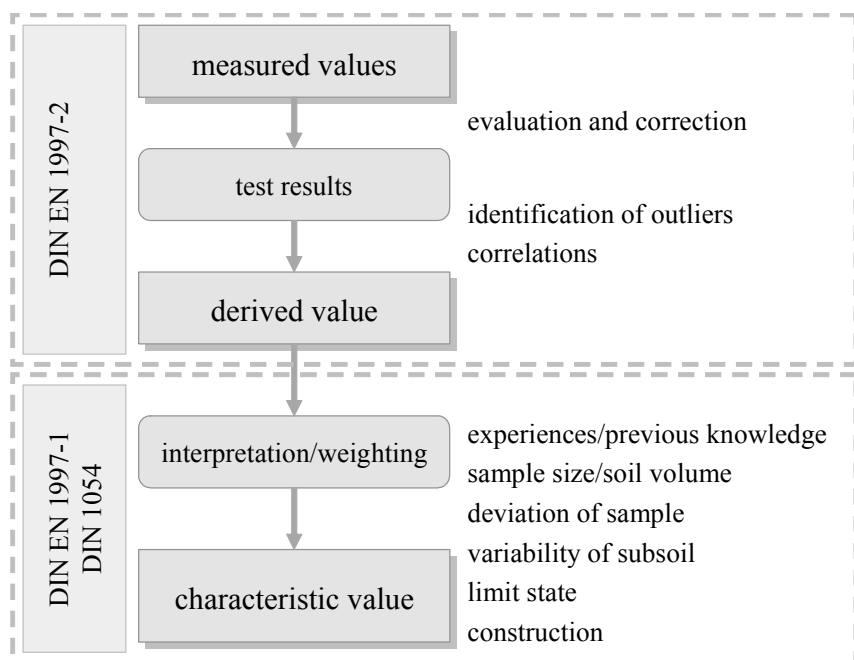

Figure 2. Characteristic value of a soil property

More practical advice for the deduction of characteristic soil values of a homogeneous soil layer from field and laboratory tests is given in DIN EN 1997-2. Principally, it is recommendable to proceed as shown in figure 2 in a simplified way in accordance to Bauduin (2001). First, tests results have to be ascertained by evaluation and correction of measurements, e.g. shear strength from a direct shear test. Afterwards, outliers have to be identified and correlated values from indirect tests, e.g. undrained shear parameters from penetration tests, have to be added to get so-called derived values.

Characteristic values of a soil property follow then by interpretation and weighting of these derived values. Useful experiences as well as previous knowledge of regional subsoil could be regarded, above all, sample size of tests, statistical spread of results and variability of subsoil have to be taken into consideration. Limit state and ability of the construction to rearrange impacts are furthermore important factors for interpretation.

This general procedure, which is shown in figure 2 in accordance with European standards and German supplementary rules, is analogously valid for site investigations everywhere in the world, because local spreading and regional variability of subsoil is naturally a common problem.

3 STATISTICAL METHODS

Site investigations are always just samples. The sample size controls information content and therefore representativeness of investigation. Statistical methods allow surveys of test results (descriptive statistic) as well as founded estimation from sample to population (deductive statistic).

With regard to DIN EN 1997-1 (sec. 2.4.5.2(11)), the arithmetic probability for a worse value should not exceed 5 % in the regarded limit state if statistical methods are applied. In this context a cautious estimation of the mean of a limited set of geotechnical parameters corresponds to a mean with a confidence level of 95 %. In cases when a local failure cannot be excluded, a cautious estimation of the low value corresponds to the 5 % fractile.

3.1 *Descriptive statistic*

The survey quality of samples with n single values x_i depends strongly on the distribution function that is chosen. In the case of soil properties, a normal distribution often already shows an adequate compliance. This distribution has the favorable attribute that every linear combination of normal distributed values is again normally distributed. Thus, if two or more normal distributed parameters are summarized within a linear relationship to a resultant resistance, the resistance still keeps the normal distribution.

Normal distribution is described by the arithmetical mean:

$$\overline{x} = \frac{\sum x_i}{n} \tag{1}$$

and by standard deviation:

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2} \ . \tag{2}$$

For soil parameters, which show typically a large scattering, as for example the water permeability, the lognormal distribution is preferable, because it does not take any negative values.

Depending on the available data set, much more complex distributions, as e.g. the Weibull distribution or the beta distribution, could give a better compliance with derived values, but in fact, normal and lognormal distribution are well-known and their compliance is often already satisfying.

Hence, in geotechnical practice, statistical analysis is actually restricted most times on these two simple distributions. Compliance of the selected distribution should be verified by visual verification or by a test of goodness of fit.

3.2 Deductive statistic

Deductive statistic permits an estimation of the mean or the lower value of the population on the basis of available samples from a site investigation by definition of confidence intervals, as shown in figure 3. Width of the confidence interval is designated by the requested probability α.

For small sample sizes, as it is often unavoidable in the case of site investigations, the student-distribution (t-distribution) allows an estimation of the variance (deviation of the population). Characteristic values x_k as cautious estimation of the mean are then given by:

$$x_k = \overline{x} \pm t_{n-1}^{\alpha} \frac{s}{\sqrt{n}}.$$ (3)

Student-distribution is tabulated as a function of probability α and of degree of freedom, which is equal to the sample size n reduced by one.

Figure 3. Gaussian normal distribution and confidence intervals for 5 % fractile and for 50 % fractile

For larger sample sizes the t-distribution decreases as a consequence of better information content of site investigation; the level of confidence becomes smaller. For endless sample sizes, the variance can be assumed to be known and the t-distribution tends to the 5 % fractile of the standard normal distribution ($t_{\infty}^{\alpha} \rightarrow 1{,}645$).

The necessary sample sizes for this assumption are regularly only available, if geotechnical expert have access to a regional database or to correlations with the results of indirect investigations, which could often be performed in much higher quantity than direct field or laboratory tests.

Characteristic values of soil properties as 50 % fractiles are only under the condition tolerable, that the construction shows an appropriate ductility. If for example a superstructure, which is founded on individual footings, could not compensate differences in soil strains, a low value (5 % fractile) of the stiffness has to be taken into account. Estimation of variance follows in this case by a Taylor series expansion.

Pre-factors K, which already summarize estimation function t and sample size n as well as probability α, have been established to disburden application of deductive statistic. Through this facilitation of eq. 3, characteristic values x_k could be determined by:

$$x_k = \overline{x} \pm K \cdot s.$$ (4)

Values for K are documented in table 1.

Table 1. Values for K of a normal distributed attribute

n	$K_{50\,\%}$	$K_{5\,\%}$
10	0.580	2.911
20	0.387	2.369
40	0.266	2.126
100	0.166	1.927
∞	0	1.645

3.3 Test of goodness of fit

Authenticity of the selected (theoretical) distribution with the actual (empirical) distribution should be verified. First, a visual comparison of empirical and theoretical distribution is recommendable. If both distributions show a satisfying fitting visually, a hypothesis test should also check the theoretical distribution. In this context the Kolmogorow-Smirnow-Test according to Lilliefors has become popular, because of its very simple procedure (Hartung et al., 1989).

The hypothesis "The existing differences between empirical and theoretical distribution are not stochastic justified" has to be dismissed, if maximum discrepancy L_n^{norm} (Figure 4) of both distributions does not exceed a critical value $l_{n,1-\alpha}^{norm}$:

$$\sqrt{n} \cdot L_n^{norm} \leq l_{n;1-\alpha}^{norm}$$ (5)

with

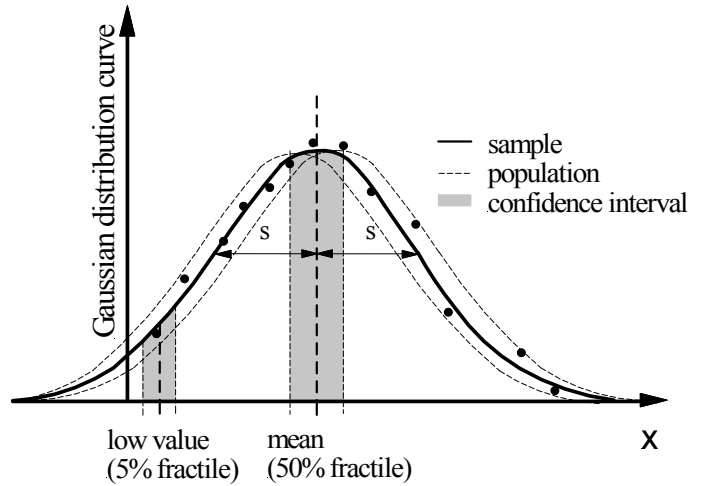$$L_n^{norm} = \sup_x \left| \phi\left(\frac{x - \overline{x}}{s} \right) - F_n^0(x) \right|$$ (6)

and

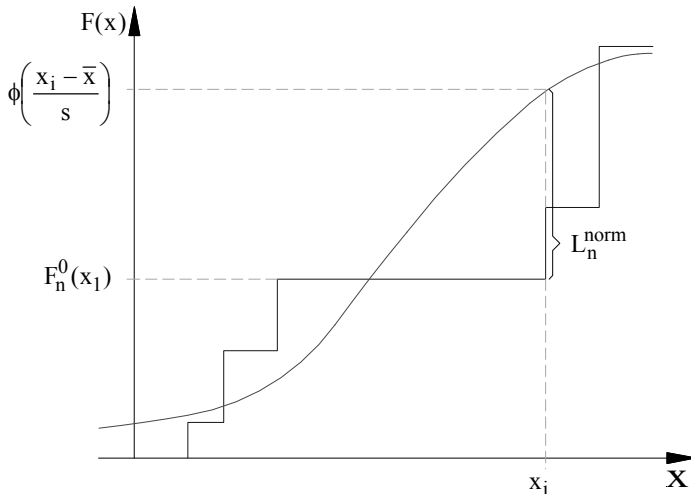|		
$F_n^0(x)$	=	value of empirical distribution at point x
$\phi\left(\dfrac{x - \overline{x}}{s} \right)$	=	value of theoretical (normal) distribution at point x
$l_{n;1-\alpha}^{norm}$	=	critical value for the test of unspecified normal distribution according to tab. 2
$1 - \alpha$	=	level of significance (usually 0.95).

Figure 4. Kolmogorow-Smirnow-Test according to Lilliefors

Table 2. Values for $l_{n,1-\alpha}^{norm}$ of a normal distributed attribute (Hartung et al., 1989)

n	5	8	10	20	30	> 30
$l_{n;0.90}^{norm}$	0.72	0.74	0.76	0.79	0.80	0.81
$l_{n;0.95}^{norm}$	0.76	0.81	0.82	0.85	0.88	0.89
$l_{n;0.99}^{norm}$	0.91	0.94	0.94	1.03	1.03	1.04

3.4 *Characteristic values with regard to trend analysis*

Some soil mechanical properties follow a trend, e.g. drained shear resistance increases approximately linear with stress level resp. to depth.

Assuming, that no useful experience is available, the confidence level of a linear regression curve for the mean could be concluded by estimation of standard deviation s_1 (Bauduin, 2001):

$$s_1^2 = \frac{1}{n-2} \cdot \left[\frac{1}{n} + \frac{(z - \overline{z})^2}{\sum\limits_{i=1}^{n}(z_i - \overline{z})^2} \right] \cdot \sum_{i=1}^{n} [(x_i - \overline{x}) - b(z_i - \overline{z})]^2$$ (7)

resp. for the 5 % fractile:

$$s_2^2 = \frac{1}{n-2} \cdot \left[1 + \frac{1}{n} + \frac{(z - \overline{z})^2}{\sum\limits_{i=1}^{n}(z_i - \overline{z})^2} \right] \cdot \sum_{i=1}^{n} [(x_i - \overline{x}) - b(z_i - \overline{z})]^2 .$$ (8)

Hence, characteristic value x_k at depth z could be written by:

$$x_k = \overline{x} + b(z - \overline{z}) \pm t_{n-2}^{\alpha} s_1$$ (9)

resp.:

$$x_k = \overline{x} + b(z - \overline{z}) \pm t_{n-2}^{\alpha} s_2 ,$$ (10)

where

$$b = \frac{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2 (z_i - \bar{z})^2}{\sum\limits_{i=1}^{n}(z_i - \bar{z})^2}.$$ (11)

As a consequence of the additional variable z, the value of the t-distribution has to be metered with sample size n reduced by two (n − 2).

Figure 5 emphasizes estimation of the characteristic soil value with regard to a trend analysis with the help of a shear stress − normal stress diagram; the underlying data has been taken from Kruse (2003). Altogether 25 single direct shear tests of a marl have been performed at 5 stress levels, varying in a range from 100 kN/m² to 500 kN/m². The linear regression results in a medium friction angle of $\varphi'_{medium} = 33.8°$ and a cohesion of $c'_{medium} = 23.2$ kN/m².

linear trend	
φ'_{medium}	33.8°
s_φ	2.41
$\varphi'_{k,normal}$	33.9°
$\varphi'_{k,log}$	34.2°
c'_{medium}	23.2 kN/m²
s_c	13.01
$c'_{k,normal}$	15.6 kN/m²
$c'_{k,log}$	5.9 kN/m²

♦ ♦ ♦ measured values
—— regression curve
▬ confidence interval
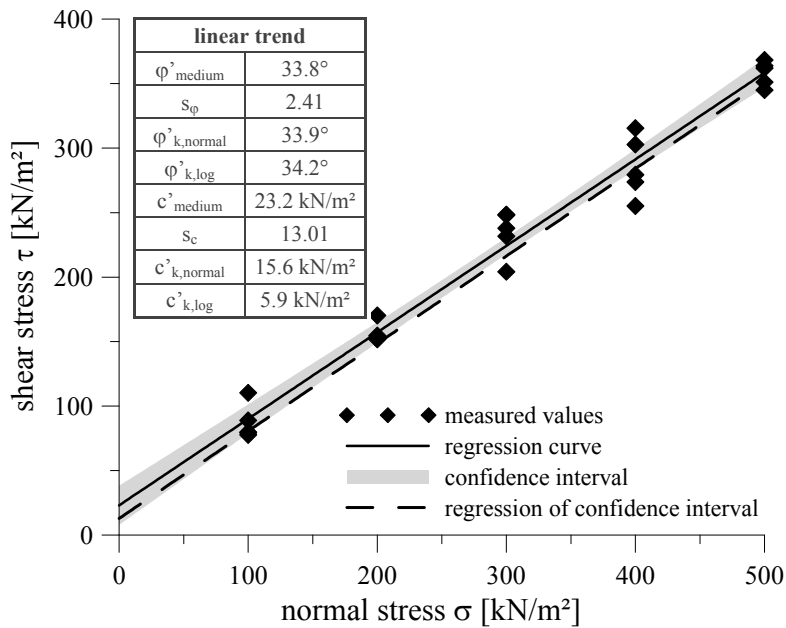- - - regression of confidence interval

Figure 5. Characteristic values, taking a linear trend into consideration (values from Kruse (2003))

The confidence intervals of the linear regression curve are slightly hyperbolic functions due to standard deviation s_1 resp. s_2 (Fellin et al., 2008; Fellin 2005). The hyperbolic relationship could be linearized again, if absolutely necessary. Characteristic shear parameters are then conform to inclination and zero crossing of the regression of the confidence interval; due to the hyperbolic relationship is $\varphi'_{k, normal} = 33.9°$ close to the medium friction angle, but c_k' has decreased to $c'_{k, normal} = 15.6$ kN/m².

Assuming a lognormal distribution of derived values, c_k' decreases to $c'_{k, log} = 5.9$ kN/m², whereas φ_k' increases a little to $\varphi'_{k, log} = 34.2°$.

4 RECOMMENDATIONS

Statistical methods are always a great tool, if they are applied professionally and if circumstances are compatible. The procedures, which have been arranged in section 3, should give a little reminder, that simple statistical methods can increase the information content of site investigations significantly. Nevertheless, statistical methods cannot replace necessary expertise.

The indispensable condition for the implementation of statistical methods for the determination of characteristic values of soil properties is a sufficient sample size. Mathematical considerations provide a much larger size than established in site investigation practice.

For a given tolerable discrepancy of the upper value from the mean $\Delta\hat{x}_p = \hat{x}_p - \hat{x}_p^u$, the statistical sample size could be estimated by:

$$n \geq \left(\frac{2\, t_{n-1,\, 1-\alpha/2} \cdot s}{\Delta\hat{x}_p} \right)^2 . \tag{12}$$

Eq. 12 demands knowledge of the standard deviation. Thus, the investigation size would have to be extended iteratively until eq. 12 is fulfilled, something that is in practice not compatible with activities in a building place and that is often contrary to tolerable costs and time effort.

In general, the estimation of 5 % fractiles requires larger samples than the estimation of 50 % fractiles. However, soil values, which are usually used as 5 % fractiles, frequently tend to result in extensive test procedures. Fischer (2001) stipulates sample sizes above 10. If on the one hand technical and economical aspects are passable and on the other hand the specific soil composition does not require larger sample sizes, 30 repetitions per type of soil seem to be adequate.

Although the t-distribution is an effective instrument for the estimation of variance, this estimation is still risky. Useful experience as well as results of indirect site investigations can help to reduce this uncertainty. Variance could be implemented from results of indirect tests, which are frequently performed in much higher quantity than direct tests. To consider useful experiences, a decision has to be made if this information is representative enough to assume a fully known variance. Otherwise useful experience could be implemented by a Bayesian analysis. Additionally, correlations could exceed the information content.

It has to be noticed, that using best possible information from available data is also in the interest of the client. Hence, he should be insistent that mean value, standard deviation and characteristic values as 50 % fractile and as 5 % fractile are declared.

5 EXAMPLE

Figure 6 illustrates the consequence of interpretation strategy (cf. Kisse et al. 2008). Drained shear parameters of Frankfurt clay have been taken from Moormann (2002). The database comprehends $n = 56$ resp. $n = 57$ values for the friction angle φ' and for the cohesion c'.

Applying the statistical methods according to sec. 3.1 to 3.3, the characteristic mean values (50 % fractile) of the normal distributed attributes exceed always the analogous values of the lognormal distributed attributes.

Characteristic lower values (5 % fractiles) behave contrarily. 5 % fractiles of the lognormal distributed attributes exceed most times the 5 % fractile of the normal distributed attributes. This tendency becomes very obvious in the case of the cohesion, because 5 % fractile of the normal distributed cohesion take a nonsensical negative value. In contrast to this, the 5 % fractile of the lognormal distributed cohesion is still conform to a positive value $c_k' = 9{,}3$ kN/m².

In figure 6, the available data of Frankfurt clay has been handled as results of a local site investigation, but in fact these data have only the character of a regional experience, which has to be supplemented in the practical case by local subsoil data. If previous knowledge does not suffice for conclusion on variance, characteristic values have to be determined by Bayesian estimation. The Bayesian theorem implicates probability of independent data of previous knowledge and local site investigation.

If for example 5 pairs of shear parameters of local subsoil have been ascertained, with $\overline{\varphi}' = 22{,}5°$, $s_\varphi = 4{,}0$ and $\overline{c}' = 35$ kN/m² and $s_c = 7{,}5$ kN/m², characteristic values are calculated to $\varphi_k' = 21{,}5°$ and $c_k' = 33{,}4$ kN/m².
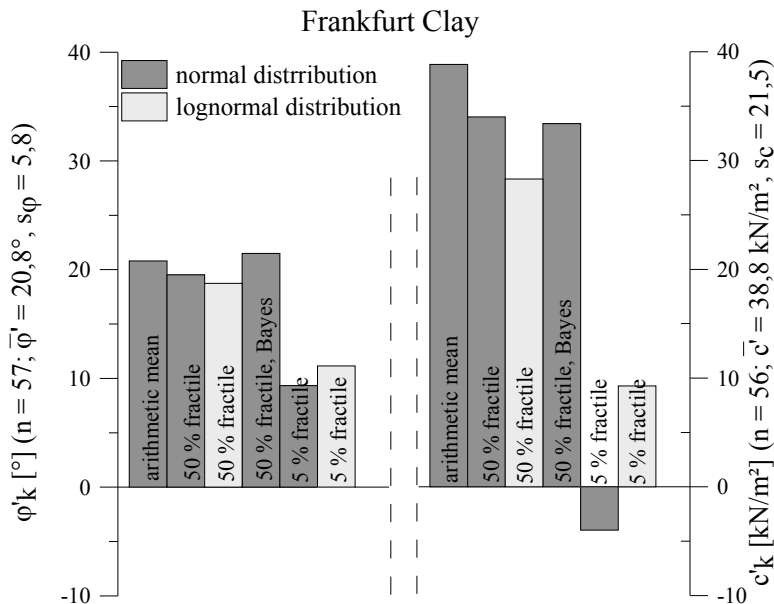
Figure 6. Example Frankfurt clay (values from Moormann (2002))

6 CONCLUSIONS

Current geotechnical design practice is mainly based on a deterministic procedure. Single (characteristic) values for impacts and resistances are the fundamental input data for geotechnical verifications (Schneider, 1993). Nevertheless, statistical methods are a reasonable instrument for determination of these characteristic values from the results of field and laboratory tests. Perhaps, application of statistical methods is the first step for a change to aspired probabilistic procedures in the future.

In the previous sections, well known and in other parts of engineering well-established statistical methods have been arranged. Examples have shown simplicity of their application. Unfortunately, due to the necessary sample sizes, statistical methods are rarely practiced in geotechnics up to now, even though DIN EN 1997 underlines their optional application explicitly.

Information content of site investigations is optimized by statistical methods and process of determination of characteristic soil values becomes verifiable.

REFERENCES

Bauduin, C. 2001. Determination of characteristic values. Geotechnical Handbook – Volume 1: Fundamentals. Smoltczyk (Ed.). Berlin: Ernst & Sohn.

Fellin, W., Berghamer, S., Renk, D. 2009. Konfidenzgrenzen der Scherfestigkeit als Grundlage zur Festlegung charakteristischer Scherparameter. Geotechnik 32/1. pp. 30-36. (in German)

Fellin, W. 2005. Assessment of characteristic shear strength parameters of soil and its implication in geotechnical design. In: Analyzing Uncertainty in Civil Engineering. Fellin, W., Lessmann, Oberguggenberger, M., H., Vieider, R. (Ed.). pp. 33-49.

Fischer, L. 2001. Das neue Sicherheitskonzept im Bauwesen. Bautechnik Spezial. Berlin: Ernst & Sohn. (in German)

Hartung, J., Elpelt, B., Klösener, K. H. 1989. Statistik – Lehr- und Handbuch der angewandten Statistik. 7. Auflage. R. Oldenbourg Verlag GmbH (in German)

Kisse, A., Pohl, C., Richwien, W. 2008. Konsequenzen der Festlegung von charakteristischen Bodenkennwerten für geotechnische Nachweise. Baugrundtagung 2008. DGGT (Ed.). (in German)

Kruse, B. 2003. Status quo bei der Festlegung charakteristischer Werte von Bodenkenngrößen. In: Stochastische Prozesse in der Geotechnik. Ziegler, M. (Ed.). Schriftenreihe Geotechnik im Bauwesen. TU Aachen. pp. 43-53 (in German)

Moormann, Ch. 2002. Trag- und Verformungsverhalten tiefer Baugruben in bindigen Böden unter besonderer Berücksichtigung der Baugrund-Tragwerk- und der Baugrund-Grundwasser-Interaktion. Katzenbach (Ed.). Mitteilungen des Institutes und der Versuchsanstalt für Geotechnik der Technischen Universität Darmstadt. (in German)

Schneider, H. R. 1993. Defintion and determination of characteristic soil properties. Proceedings of the 14th International Conference on Soil Mechanics and Foundation Engineering. Hamburg, Balkema. pp. 2271-2274.