RECOMMENDED PRACTICE

DNV-RP-C207

Statistical Representation of Soil Data

JANUARY 2012

The electronic pdf version of this document found through http://www.dnv.com is the officially binding version

DET NORSKE VERITAS AS

FOREWORD

DET NORSKE VERITAS (DNV) is an autonomous and independent foundation with the objectives of safeguarding life, property and the environment, at sea and onshore. DNV undertakes classification, certification, and other verification and consultancy services relating to quality of ships, offshore units and installations, and onshore industries worldwide, and carries out research in relation to these functions.

DNV service documents consist of among others the following types of documents:

— *Service Specifications.* Procedual requirements.
— *Standards.* Technical requirements.
— *Recommended Practices.* Guidance.

The Standards and Recommended Practices are offered within the following areas:

A) Qualification, Quality and Safety Methodology
B) Materials Technology
C) Structures
D) Systems
E) Special Facilities
F) Pipelines and Risers
G) Asset Operation
H) Marine Operations
J) Cleaner Energy
O) Subsea Systems

Any comments may be sent by e-mail to *rules@dnv.com*

CHANGES

General

This document supersedes DNV-RP-C207, October 2010.

Text affected by the main changes in this edition is highlighted in red colour. However, if the changes involve a whole chapter, section or sub-section, normally only the title will be in red colour.

Main changes

— The expression for $\hat{a}_0$ in 2.4.2.2 has been amended.

Acknowledgment

This Recommended Practice has been developed in close cooperation with the industry. The basis for the Recommended Practice is a Guidance Note for statistical representation of soil data, sponsored by Norsk Hydro and Det Norske Veritas. These contributions are gratefully acknowledged.

CONTENTS

1. Introduction

1.1 General

1.1.1 General

1.1.1.1 This Recommended Practice provides principles, guidance and recommendations for use of statistical methods for analysis and representation of soil data.

1.1.1.2 Soil data to be used as a basis for geotechnical analysis and foundation design are usually encumbered with uncertainty, both in terms of natural variability and in terms of limited amounts of data. Statistical methods form a rational tool for handling of such uncertain data, and this Recommended Practice shows how to apply this tool, in particular with a view to the selection of characteristic values for use in design.

1.1.1.3 The selection of characteristic values of soil properties for use in geotechnical design is often based on subjective judgment and accumulated experience, and the uncertainties which are involved with the soil properties are only to a limited extent brought into the picture when the characteristic values are chosen.

1.1.2 Objective

1.1.2.1 It is a major purpose of this Recommended Practice to demonstrate how characteristic values of soil properties can be extracted from available data in a rational manner by means of statistical methods. Statistical methods allow for quantification of the involved uncertainties and make it possible to account for these uncertainties in a rational manner. The focus is on application of statistical methods to rational selection of characteristic values for design, but other applications of statistical methods are also considered, e.g. development of full probabilistic models of soil properties for use in probabilistic analyses such as structural reliability analyses.

1.1.3 Scope and application

1.1.3.1 The Recommended Practice is applicable to analysis and representation of soil data for all kinds of offshore geotechnical engineering problems.

1.1.3.2 The following major topics are covered by the Recommended Practice:

— types of uncertainties in soil data
— uncertainty modelling, including probability distributions and parameters to describe them
— generic distribution types for soil data
— central estimation of soil parameters
— estimation of soil parameters with confidence
— interpretation of characteristic values from data
— geostatistics, including spatial averaging and spatial estimation
— full probability distribution modelling of soil properties in geotechnical reliability analysis.

1.2 References

1.2.1 Offshore Standards

1.2.1.1 The latest versions of the documents in Table 1-1 apply.

Table 1-1 DNV Offshore Standards

Reference	*Title*
DNV-OS-C101	Design of Offshore Steel Structures, General (LRFD Method)
DNV-OS-F101	Submarine Pipeline Systems
DNV-OS-F201	Dynamic Risers
DNV-OS-J101	Design of Offshore Wind Turbine Structures

1.2.2 Recommended Practices and Classification Notes

1.2.2.1 The Recommended Practices and Classification Notes in Table 1-2 form useful references for application of this document.

Table 1-2 DNV Recommended Practices and Classification Notes

Reference	*Title*
Classification Notes 30.4	Foundations
Classification Notes 30.6	Structural Reliability of Marine Structures

Table 1-2 DNV Recommended Practices and Classification Notes (Continued)	
Reference	*Title*
DNV-RP-E301	Design and Installation of Fluke Anchors in Clay
DNV-RP-E302	Design and Installation of Plate Anchors in Clay
DNV-RP-E303	Geotechnical Design and Installation of Suction Anchors in Clay
DNV-RP-F105	Free Spanning Pipelines

1.3 Definitions

1.3.1 Verbal forms

1.3.1.1 Shall: Indicates a mandatory requirement to be followed for fulfilment or compliance with the present Recommended Practice. Deviations are not permitted unless formally and rigorously justified, and accepted by all relevant contracting parties.

1.3.1.2 Should: Indicates a recommendation that a certain course of action is preferred or is particularly suitable. Alternative courses of action are allowable under the Recommended Practice where agreed between contracting parties, but shall be justified and documented.

1.3.1.3 May: Indicates permission, or an option, which is permitted as part of conformance with the Recommended Practice.

1.3.1.4 Can: Requirements with can are conditional and indicate a possibility to the user of the Recommended Practice.

1.3.1.5 Agreement, or *by agreement*: Unless otherwise indicated, agreed in writing between contractor and purchaser.

1.3.2 Terms

1.3.2.1 Characteristic value: A representative value of a load variable or a resistance variable. For a load variable, it is a high but measurable value with a prescribed probability of not being unfavourably exceeded during some reference period. For a resistance variable it is a low but measurable value with a prescribed probability of being favourably exceeded.

1.3.2.2 Classification Notes: The Classification Notes cover proven technology and solutions which are found to represent good practice by DNV, and which represent one alternative for satisfying the requirements stipulated in the DNV Rules or other codes and standards cited by DNV. The classification notes will in the same manner be applicable for fulfilling the requirements in the DNV offshore standards.

1.4 Abbreviations and symbols

1.4.1 Abbreviations

1.4.1.1 Abbreviations listed in Table 1-3 are used in this Recommended Practice.

Table 1-3 Abbreviations	
Abbreviation	*In full*
CIU	Consolidated isotropic undrained (test)
CPT	Cone penetration test
DSS	Direct simple shear
FORM	First-order reliability method
SORM	Second-order reliability method
UU	Unconsolidated undrained (test)

1.4.2 Symbols

1.4.2.1 Latin characters

a Scale parameter
a Lower bound for bounded variable
a_0 Soil surface intercept
a_1 Depth gradient
b Limit of autocorrelation function for small lag r
b Upper bound for bounded variables
C Covariance, covariance function

C	Covariance matrix
Cov	Covariance
COV	Coefficient of variation
$c_{1-\alpha}$	Quantile in non-central *t* distribution
D	Diameter
E	Mean value operator
f	Probability density function
F	Cumulative distribution function
g	Limit state function
k	Proportionality factor between standard deviation and depth
L	Length
L	Load
m	Mode in probability distribution
n	Sample size, number of data
n	Degrees of freedom
p	Probability, fraction
P	Probability
P_F	Probability of failure
Q	Probability of exceedance, complement of cumulative distribution function
Q	Capacity
r	Parameter in Beta distribution
r	Lag, distance, separation
R	Correlation length
R	Resistance
s	Estimate of standard deviation
s	Position, location coordinates
se	Standard error; standard deviation of estimate
s_u	Undrained shear strength
s_{uC}	Static triaxial compression undrained shear strength
s_{uD}	Static DSS undrained shear strength
s_{uE}	Static triaxial extension undrained shear strength
t	Parameter in Beta distribution
t_n	Quantile in Student's *t* distribution
v	Generic variable in expression for Beta distribution
w	Weighting factor in regression analysis
x	Realisation of stochastic variable
X	Stochastic variable, random variable
X	Vector of stochastic variables
y	Realisation of stochastic variable
Y	Stochastic variable, random variable
z	Depth
Z	Stochastic variable, random variable

1.4.2.2 Greek characters

α	Dimensionless factor for axial pile capacity
α	Scale parameter in Weibull distribution
α	Complement of confidence $1-\alpha$
α_3	Skewness
α_4	Kurtosis
β	Shape parameter in Weibull distribution
β	Reliability index
ε	Residual, difference between observed value and predicted value
Φ	Standard normal cumulative distribution function
γ	Variogram function
γ	Shift parameter in Weibull distribution

λ Variance reduction factor to account for effect of spatial averaging
λ Weighting factor in point estimation by kriging
μ Mean value
ρ Coefficient of correlation, autocorrelation function
$\boldsymbol{\rho}$ Correlation matrix
σ Standard deviation

2. Statistical Representation of Soil Data

2.1 Introduction

2.1.1 General

2.1.1.1 Following an introduction to uncertainty types in soil data and to principles of uncertainty modelling, this section presents statistical methods for analysis and representation of soil data. Formulas for central estimation of soil parameters and formulas for estimation of soil parameters with confidence are given. Examples of application are given with focus on estimation of characteristic values. Guidance for selection of soil parameters for design is given.

2.1.1.2 Issues of special interest in the context of uncertainty modelling for soil data for large volumes, such as spatial averaging and spatial estimation, are dealt with in separate subsections, and examples of application are given.

2.1.1.3 An introduction to full probabilistic modelling of soil properties is given.

2.2 Types of uncertainty

2.2.1 General

2.2.1.1 Uncertainties associated with geotechnical data have many sources and may be divided into the following two main types of uncertainty:

— aleatory uncertainty, i.e. physical uncertainty
— epistemic uncertainty, i.e. uncertainty related to imperfect knowledge.

2.2.2 Aleatory uncertainty

2.2.2.1 Aleatory uncertainty consists of physical uncertainty. Physical uncertainty is also known as inherent uncertainty and intrinsic uncertainty and is a natural randomness of a quantity such as the variability in the soil strength from point to point within a soil volume. Such physical uncertainty or natural variability is a type of uncertainty which cannot be reduced.

2.2.2.2 Physical uncertainty associated with a geotechnical variable is a fundamental property of the variable.

2.2.3 Epistemic uncertainty

2.2.3.1 Epistemic uncertainty consists of statistical uncertainty, model uncertainty and measurement uncertainty, which are all classified as a type of uncertainty associated with limited, insufficient or imprecise knowledge.

2.2.3.2 Epistemic uncertainty can in principle be reduced by collection of more data, by improving engineering models and by employing more accurate methods of measurement.

2.2.3.3 Statistical uncertainty is uncertainty due to limited information such as a limited number of observations of a quantity, e.g. a limited number of soil strength values from a limited number of soil samples tested in the laboratory.

2.2.3.4 Model uncertainty is uncertainty due to imperfections and idealisations made in (1) applied engineering models for representation and prediction of quantities such as pile resistance and foundation capacity and in (2) choices of probability distribution types for representation of uncertain quantities.

2.2.3.5 Model uncertainty involves two elements, viz. (1) a bias if the model systematically leads to overprediction or underprediction of a quantity in question and (2) a randomness associated with the variability in the predictions from one prediction of that quantity to another.

Guidance note:

A geotechnical example of model uncertainty is the uncertainty associated with axial pile capacity predictions from undrained shear strength by the relatively simple α method.

Another example of model uncertainty is the uncertainty associated with the conversion of cone resistance data from CPT tests to undrained shear strength data.

A third example of model uncertainty is the uncertainty associated with the choice of a generic distribution type for the probability distribution of soil properties when data are too scarce to identify the correct distribution type, when no generic distribution type provides a good fit to data, or when a wrong distribution type is used in calculations for simplicity.

---e-n-d---of---G-u-i-d-a-n-c-e---n-o-t-e---

2.2.3.6 Model uncertainty varies from case to case and can be studied by interpretation of pertinent sensitivities.

2.2.3.7 Measurement uncertainty is caused by imperfect instruments and sample disturbance when a quantity is observed. Measurement uncertainty involves two elements, viz. (1) a systematic error associated with the bias in the measurements and (2) a random error associated with the precision in the measurements.

Guidance note:

An example of measurement uncertainty, which may involve both a bias and a random error, is the uncertainty associated with sample disturbance of soil samples collected for subsequent testing in the laboratory, e.g. for determination of the soil strength.

When sample disturbance takes place, the expected tendency will be that a smaller strength than the true strength will be measured in the laboratory, i.e. the sample disturbance implies introduction of a systematic error in terms of a negative bias. If, in addition, the degree of sample disturbance varies from sample to sample, then a random error component in the measured strength is also introduced.

---e-n-d---of---G-u-i-d-a-n-c-e---n-o-t-e---

2.2.3.8 If a considered quantity is not observed directly from measurements, but predicted by imposing some data processing of the observed data, a combined model and measurement uncertainty may result.

Guidance note:

Interpretation of undrained shear strength from cone penetration test results is an example of a procedure that leads to soil strength data which are encumbered with a combined model and measurement uncertainty.

---e-n-d---of---G-u-i-d-a-n-c-e---n-o-t-e---

2.2.3.9 Bias is defined as the difference between the expected value of the prediction of a quantity by measurements, prediction models or other means of estimation, and the true value of the quantity.

2.2.4 Other types of uncertainty

2.2.4.1 Other types of uncertainties exist such as workmanship, human errors and gross errors, but they are not considered here.

Guidance note:

For artificial soil deposits such as a backfill over a pipeline in a trench, the uncertainty associated with the workmanship applied for the installation of the backfill may appear in the same manner as a natural variability. However, for any new backfills to be installed this uncertainty can in principle be reduced by changing and improving the process of work, if this is an option with the available equipment.

---e-n-d---of---G-u-i-d-a-n-c-e---n-o-t-e---

2.2.5 Correlation of uncertainties

2.2.5.1 Some types of uncertainties may be correlated. Other types of uncertainties are uncorrelated or independent. For example, epistemic uncertainties are by nature statistically independent of natural variability.

2.3 Uncertainty modelling

2.3.1 Probability distributions

2.3.1.1 All uncertainty types presented in 2.2.1 - 2.2.3 can in principle be represented by a probability distribution. When a parametric distribution model is used, the distribution can be represented by its generic distribution type and its distribution parameters.

2.3.1.2 In the context of soil properties, the most commonly encountered generic distribution models consist of:

— normal distribution
— lognormal distribution
— weibull distribution
— uniform distribution
— triangular distribution.

Guidance note:

The normal distribution is typically used for variables which come about as sums or averages of a large number of independent quantities. Strength variables often tend to follow the normal distribution. The lognormal distribution is

used for variables which come about as non-negative products of a large number of independent quantities. The uniform distribution often forms a suitable distribution model for physically bounded variables for which no prior information exists.

---e-n-d---of---G-u-i-d-a-n-c-e---n-o-t-e---

2.3.1.3 The most commonly encountered parameters used to characterise a probability distribution consist of the following properties:

— mean value (expected value)
— standard deviation (a measure of the spread or the scatter about the mean)
— coefficient of variation (standard deviation divided by the mean value; a relative measure of the spread or the scatter about the mean)
— lower bound and upper bound (used in the case of bounded variables).

Guidance note:

For some generic distribution types, such as the normal distribution, these parameters are used directly as distribution parameters in the mathematical expressions for the distributions. For other generic distribution types, such as the log-normal distribution and the Weibull distribution, distribution parameters are used which can be derived from the listed characterising properties.

---e-n-d---of---G-u-i-d-a-n-c-e---n-o-t-e---

2.3.1.4 Other terms and properties which are often used in the characterisation of probability distributions include:

— variance, defined as the square of the standard deviation
— skewness, a measure of the asymmetry of the distribution
— kurtosis, a measure of the flatness of the distribution
— standard error, term used for the standard deviation of estimators for constant quantities, such as estimators for distribution parameters like the mean value
— covariance, a measure of the joint variability of two quantities or variables
— correlation, a relative measure of the joint variability of two quantities or variables.

2.3.1.5 Examples of parameters listed in 2.3.1.3 and 2.3.1.4 can be found in 2.3.2.1 through 2.3.2.8.

2.3.2 Definitions, symbols and notions for probability distributions

2.3.2.1 Let X denote a random soil variable such as the strength in a soil deposit, and let x be a realisation of X such as the observed value of X in a particular point within the soil deposit.

2.3.2.2 The cumulative distribution function F_X of X is defined as the probability that X falls short of x,

$$F_X(x) = P[X \le x]$$

where P[] denotes probability.

The probability of exceedance Q_X is defined as the complement of the cumulative distribution function

$$Q_X(x) = 1 - F_X(x) = P[X > x]$$

The probability density function is defined as the derivative of the cumulative distribution function

$$f_X(x) = \frac{dF(x)}{dx}$$

The probability density function is only defined for continuous, differentiable cumulative distribution functions.

The *p* quantile in the distribution of X is the value of X whose cumulative distribution function value is *p*, hence

$$x_p = F_X^{-1}(p)$$

2.3.2.3 The mean value of X is defined as

$$\mu = E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$$

and is also known as the expected value of X.

2.3.2.4 The variance of X is

$$Var[X] = \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x) dx$$

which alternatively can be expressed as

$$Var[X] = E[(X-\mu)^2] = E[X^2] - E[X]^2$$

and the standard deviation of X is

$$\sigma = \sqrt{Var[X]}$$

The coefficient of variation is defined as

$$COV = \frac{\sigma}{\mu}$$

Guidance note:

The coefficient of variation expresses the standard deviation σ as a fraction of the mean value μ. The coefficient of variation is meaningful only for variables which have positive mean values. For geotechnical applications the coefficient of variation is of practical use for variables whose realisations are non-negative, such as soil strengths.

---e-n-d---of---G-u-i-d-a-n-c-e---n-o-t-e---

2.3.2.5 The skewness of X is

$$\alpha_3 = \frac{\int_{-\infty}^{\infty} (x-\mu)^3 f_X(x)dx}{\sigma^3}$$

Guidance note:

Symmetric distributions such as the normal distribution have $\alpha_3 = 0$. Distributions skewed to the right have $\alpha_3 > 0$. Distributions skewed to the left have $\alpha_3 < 0$. Examples are given in Figure 2-1.

---e-n-d---of---G-u-i-d-a-n-c-e---n-o-t-e---

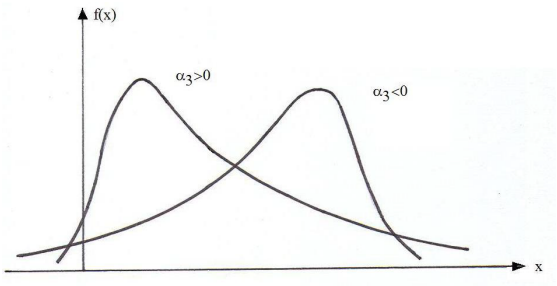

Figure 2-1
Probability density functions to illustrate skewness

2.3.2.6 The kurtosis of X is

$$\alpha_4 = \frac{\int_{-\infty}^{\infty} (x-\mu)^4 f_X(x)dx}{\sigma^4}$$

Guidance note:

The normal distribution has $\alpha_4 = 3$. Long-tailed distributions have $\alpha_4 > 3$. Flat-topped distributions which have a large concentration of outcomes near the mean have $\alpha_4 < 3$. Examples are given in Figure 2-2.

---e-n-d---of---G-u-i-d-a-n-c-e---n-o-t-e---

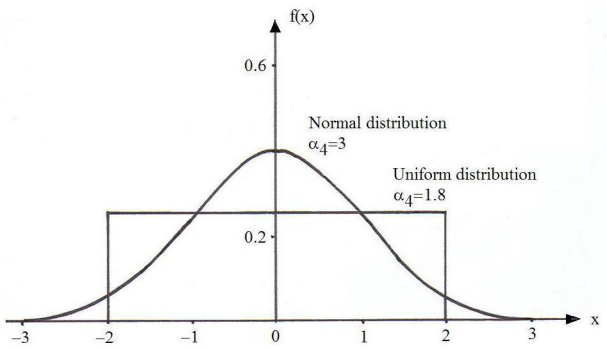

Figure 2-2
Probability density functions to illustrate kurtosis

2.3.2.7 The median of X is the 50% quantile in the distribution of X. For a symmetrical distribution, such as the normal distribution, the median of X coincides with the mean value of X.

2.3.2.8 Let X and Y denote two soil variables. The covariance between X and Y is defined as

$$Cov[X,Y] = E[(X-\mu_X)(Y-\mu_Y)]$$

The covariance matrix for X and Y is defined as

$$\mathbf{C}_{XY} = \begin{bmatrix} Var[X] & Cov[X,Y] \\ Cov[X,Y] & Var[Y] \end{bmatrix}$$

The coefficient of correlation between X and Y is defined as

$$\rho = \frac{Cov[X,Y]}{\sqrt{Var[X]Var[Y]}}$$

and is a relative measure of correlation which assumes values between −1 and 1. A value of 0 indicates no correlation, a value of −1 indicates full negative correlation, and a value of 1 indicates full positive correlation. See Figure 2-3.

The correlation matrix for X and Y is defined as

$$\boldsymbol{\rho}_{XY} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

Guidance note:

The coefficient of correlation is a measure of closeness of linear relationship between two variables X and Y. A positive value indicates a tendency of X and Y to increase together. A negative value indicates that large values of X are associated with small values of Y. Values of either 1 or –1, corresponding to full positive and negative correlation, respectively, imply linear functional relationships between X and Y. The diagram in the upper left corner of Figure 2-3 visualises the situation with full negative correlation between X and Y and the associated linear relationship.

The diagram in the lower right corner of Figure 2-3 visualises that whereas there may be a strong dependency between two variables X and Y, this does not always imply a strong correlation between the variables. For example, $Y = X^2$ is very much dependent on X, but the coefficient of correlation between X and Y is 0.

In general, if X and Y are independent, the coefficient of correlation between them is 0. However, the converse is not necessarily true.

---e-n-d---of---G-u-i-d-a-n-c-e---n-o-t-e---

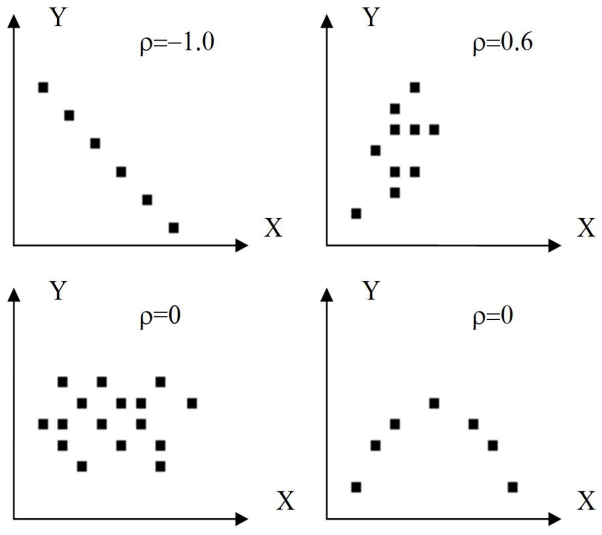

Figure 2-3
Examples of coefficients of correlation

2.3.2.9 A correlation matrix can also be established for the elements of an array of three or more variables. For an array of three variables, $Z = (Z_1, Z_2, Z_3)$, the correlation matrix has the form

$$\boldsymbol{\rho}_{\mathbf{Z}} = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix}$$

Guidance note:

An example of an array of three correlated variables that are commonly encountered in geotechnical problems is the array of undrained shear strengths in compression, direct simple shear and extension, (s_{uC}, s_{uD}, s_{uE}).

---e-n-d---of---G-u-i-d-a-n-c-e---n-o-t-e---

2.3.2.10 Let X denote a soil variable such as the soil strength within a soil volume. X exhibits natural variability and therefore varies from point to point within the soil volume. X is said to form a random field over the soil volume.

Guidance note:

The natural variability from point to point within a soil volume is a result of the natural formation of soil in different depositional environments. This variation can exist even in apparently homogeneous soil units.

---e-n-d---of---G-u-i-d-a-n-c-e---n-o-t-e---

2.3.2.11 The autocorrelation function for X is defined as the coefficient of correlation between the values of X in two points located a distance *r* apart within the soil volume,

$$\rho(r) = \frac{Cov[X(\mathbf{s}), X(\mathbf{s}')]}{Var[X]}$$

where **s** and **s'** are the location coordinates of the two points and $r = |\mathbf{s}-\mathbf{s}'|$ denotes the distance between them.

The autocorrelation function reflects the connectivity in the soil properties and tells something about how the variation of X from point to point is. An autocorrelation function typically decreases for increasing lag *r*.

Guidance note:

A commonly applied model for the autocorrelation function for soil properties is the quadratic exponential decay model,

$$\rho(r) = \exp(-\frac{r^2}{R^2})$$

in which R is a correlation length which expresses the scale of fluctuation in the random field X. Random fields of soil properties are usually anisotropic. The vertical correlation length R_V is usually smaller than the horizontal correlation length R_H. This difference reflects the geological processes that lead to the formation of soil deposits by sedimentary depositing and may amount to as much as an order of magnitude.

Let *r* have components Δx in the horizontal plane and Δz in the vertical direction. For the anisotropic case that $R_V \neq R_H$, the quadratic exponential decay model can then be expressed as

$$\rho(\Delta x, \Delta z) = \exp(-(\frac{\Delta x}{R_H})^2 - (\frac{\Delta z}{R_V})^2)$$

Examples of autocorrelation function models are given in Figure 2-4.

---e-n-d---of---G-u-i-d-a-n-c-e---n-o-t-e---

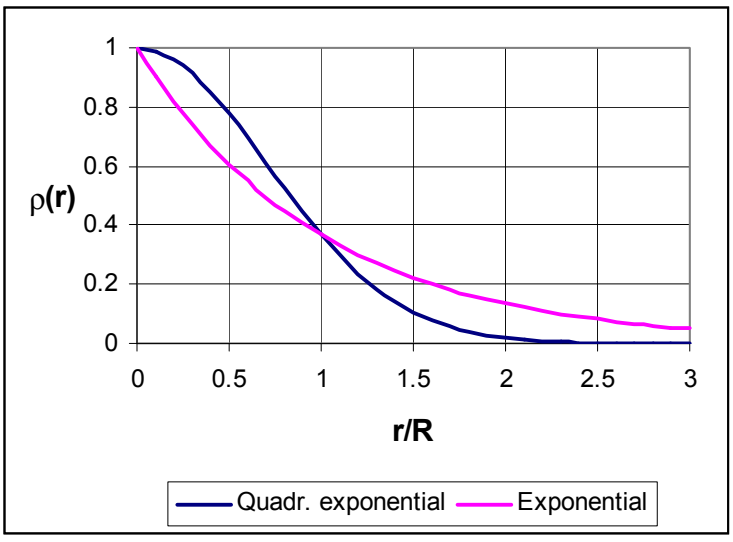

Figure 2-4
Autocorrelation function models

2.3.2.12 The theoretical limit of the autocorrelation function as the lag r approaches zero is unity,

$\rho(r) \to 1$ for $r \to 0$.

However, when $\rho(r)$ is inferred from data, one may observe that

$\rho(r) \to b$ for $r \to 0$ with $0 < b < 1$

The difference $1-b$ is known as "nugget effect". The nugget effect is in most cases due to measurement uncertainty, which by nature does not exhibit any spatial correlation structure.

Guidance note:

The presence of a nugget effect will imply that the quadratic exponential decay model becomes changed to

$$\rho(r) = b \cdot \exp(-\frac{r^2}{R^2}) \text{ for } r > 0$$

---e-n-d---of---G-u-i-d-a-n-c-e---n-o-t-e---

2.3.2.13 The variogram associated with the random field X is an alternative measure of the spatial connectivity of X and is defined as

$$\gamma(r) = C(0) - C(r) = \sigma^2 \cdot (1 - \rho(r))$$

in which

$$C(r) = Cov[X(\mathbf{s}), X(\mathbf{s}')] = \sigma^2 \cdot \rho(r); r = |\mathbf{s} - \mathbf{s}'|$$

is the covariance function with the special case

$$C(0) = Var[X] = \sigma^2$$

where **s** and **s'** denote the location coordinates of two points located a distance $r = |\mathbf{s}-\mathbf{s'}|$ apart.

Figure 2-5 shows a variogram with nugget effect, range and sill. The nugget effect is defined in 2.3.2.12. The range is a distance beyond which the variogram essentially remains constant. The sill is the plateau the variogram reaches at the range.

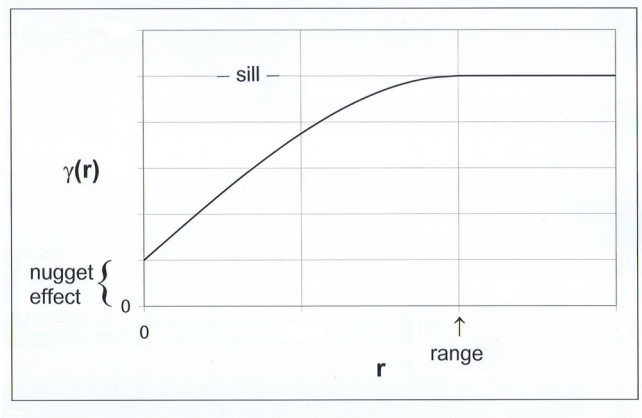

Figure 2-5
Variogram

Guidance note:

The variogram $\gamma(r)$ is slightly more flexible to operate with than the covariance function C(r), because there are more licit variogram functions than there are licit covariance functions. This is in the sense that licit variogram functions, which do not have a sill for large lags r do not have a covariance function counterpart.

The variogram $\gamma(r)$ is sometimes denoted semivariogram, which refers to its alternative expression as $\gamma(r) = ½\text{Var}[X(\mathbf{s})–X(\mathbf{s'})]$.

The most commonly used variogram models with sill are

Spherical model:
$$\gamma(r) \propto \begin{cases} \frac{3}{2}\frac{r}{R} - \frac{1}{2}(\frac{r}{R})^3 & r \le R \\ 1 & r > R \end{cases}$$

Exponential model:
$$\gamma(r) \propto 1 - \exp(-\frac{r}{R})$$

Quadratic exponential decay model: $\gamma(r) \propto 1 - \exp(-(\frac{r}{R})^2)$

Hole effect model: $\gamma(r) \propto 1 - \frac{\sin(ar)}{r}$

The most commonly used variogram models without sill are

Power model: $\gamma(r) \propto |r|^a ; 0 < a < 2$

Hole effect model: $\gamma(r) \propto 1 - \cos(ar)$

The coefficients a and R are scale parameters, and R is sometimes referred to as correlation length.

---e-n-d---of---G-u-i-d-a-n-c-e---n-o-t-e---

2.3.2.14 Autocorrelation functions and variograms can in principle be estimated from data, provided a sufficient amount of data for the relevant parameter is available from the soil volume of interest. Autocorrelation function models and variogram models can then be fitted to the estimated functions. The model that provides the best fit to the estimated function values should be sought after. In practice, the amount of data will often be scarce, in which case one may be left with choosing a model on the basis of experience. Also in practice, one may sometimes find that it is of less importance which model is used as long as the scale parameter is correct. It may facilitate the estimation of autocorrelation functions and variograms from data to sample the data equidistantly, provided such equidistant sampling is feasible.

2.3.2.15 Autocorrelation functions and variograms are used for estimation of soil properties and associated uncertainties in problems that involve large areas and soil volumes. For examples of application, reference is made to 2.7.1 - 2.7.4.

2.3.3 The normal distribution

2.3.3.1 The normal probability distribution is a distribution model which often proves adequate for representation of a number of soil properties, including strength properties such as the undrained shear strength. The normal distribution is fully described by its mean value and its standard deviation, i.e. two parameters are all that is required for a complete description of this distribution model.

2.3.3.2 The cumulative distribution function for a considered soil variable X gives the cumulative probability up to a specified outcome x of X, i.e. the probability that X will fall short of x. For a normally distributed variable X, the cumulative distribution function reads

$$F_X(x) = P[X \le x] = \Phi(\frac{x-\mu}{\sigma}) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-\frac{(x-\mu)^2}{\sigma^2}) dx$$

in which Φ denotes the standard normal cumulative distribution function, μ denotes the mean value of X and σ denotes the standard deviation of X. A table of the standard cumulative normal distribution is given in Appendix A.

2.3.3.3 For the situation that there is only a limited number of laboratory test data on soil samples available for the determination of the distribution of X, the estimates of the two distribution parameters (the mean value and the standard deviation) from the data will be encumbered with statistical uncertainty, i.e. the mean value and the standard deviation will themselves have uncertainties that can be modelled by assigning probability distributions to them. These distributions will have the calculated estimates as their mean values and they will in addition be characterised by their standard errors.

2.3.3.4 Variables that are modelled to follow the normal distribution can take on negative values. This may be unphysical for a number of soil variables such as soil strengths, which never take on negative values, but for which the normal distribution still is an adequate distribution model, e.g. as judged on other physical grounds than the extent of the physically possible space of realisations. Usually, it will not pose any problems that the normal distribution in theory can take on negative values, since for practical purposes the probability of negative values will usually be negligible. Only in cases where the coefficient of variation of a variable in question is particularly large, there may be a risk of getting negative outcomes of the variable when a normal distribution model is used. In such cases it is recommended to apply a truncated normal distribution, where the point of truncation of the lower tail is set at zero.

Guidance note:

Let X denote a normal distribution with mean value μ and standard deviation σ, and let Φ denote the standard normal cumulative distribution function. Let Y be generated as X truncated at x = 0. The cumulative distribution function for Y is then defined as

$$F_Y(y) = \frac{\Phi(\frac{y-\mu}{\sigma}) - \Phi(-\frac{\mu}{\sigma})}{1 - \Phi(-\frac{\mu}{\sigma})};\ y \geq 0$$

---e-n-d---of---G-u-i-d-a-n-c-e---n-o-t-e---

2.3.4 Other distributions

2.3.4.1 The lognormal distribution is defined such that if the variable X is lognormally distributed, then the variable ln$X$ is normally distributed. The lognormal distribution is only defined for positive outcomes of X and is therefore sometimes used to represent distributions of variables that cannot take on negative values, such as strengths. A drawback by using the lognormal distribution for strength variables is that it may sometimes lead to nonconservative predictions of strengths in the lower tail of the distribution.

Guidance note:

The following formulas can be used to transform the mean value and the variance of ln$X$ to mean value and variance of X, when X is lognormally distributed

$$E[X] = \exp(E[\ln X] + \frac{1}{2}Var[\ln X])$$

$$Var[X] = E[X]^2 \cdot \exp(Var[\ln X] - 1)$$

Note that the natural logarithm of the mean value of X is *not* the mean value of ln$X$.

---e-n-d---of---G-u-i-d-a-n-c-e---n-o-t-e---

2.3.4.2 The Weibull distribution has a cumulative distribution function

$$F_X(x) = 1 - \exp(-(\frac{x-\gamma}{\alpha})^{\beta})$$

in which the scale parameter α, the shape parameter β and the shift parameter γ are distribution parameters which all relate to the mean value and standard deviation of the distribution. The Weibull distribution is a flexible distribution which has many applications, e.g. distributions of local maxima of properties that vary in time and space. The exponential distribution and the Rayleigh distribution are special cases of the Weibull distribution for $\beta = 1$ and $\beta = 2$, respectively.

2.3.4.3 The Beta distribution is a flexible distribution model for variables which are upwards and downwards bounded and for which the mean value and the standard deviation are known.

For a variable X which is Beta-distributed over the range [a; b] and has mean value μ and standard deviation σ, the cumulative distribution function reads

$$F_X(x) = \frac{\int_a^x (v-a)^{r-1}(b-v)^{t-r-1}dv}{(b-a)^{t-1}B(r,t-r)};\ a \leq x \leq b \text{ and } 0 < r \leq t$$

where B denotes the Beta function and in which the parameters r and t relate to μ and σ through

$$\mu = a + (b-a)\frac{r}{t} \quad \text{and} \quad \sigma = (b-a)\frac{r}{t}\sqrt{\frac{t-r}{r(t+1)}}$$

2.3.4.4 The uniform distribution is a useful distribution model for variables for which physical upper and lower bounds are known while no other prior information exists.

For a variable X which is uniformly distributed over the range [a; b], the cumulative distribution function reads

$$F_X(x) = \frac{x-a}{b-a}$$

The mean value is

$$E[X] = \frac{a+b}{2}$$

and the standard deviation is

$$\sigma_X = \frac{b-a}{\sqrt{12}}$$

2.3.4.5 The triangular distribution is a useful distribution model for variables for which physical upper and lower bounds are known together with the most likely value.

For a variable X which follows a triangular distribution over the range [a; b] and whose most likely value is m, the cumulative distribution function reads

$$F_X(x) = \begin{cases} \dfrac{(x-a)^2}{(m-a)(b-a)} & a \le x \le m \\ 1 - \dfrac{(b-x)^2}{(b-m)(b-a)} & m \le x \le b \end{cases}$$

The mean value is

$$E[X] = \frac{a+b+m}{3}$$

and the standard deviation is

$$\sigma_X = \sqrt{\frac{a^2+b^2+m^2-ab-am-bm}{18}}$$

2.3.5 Characteristic values

2.3.5.1 For practical purposes in deterministic geotechnical design, one will not apply the entire probability distribution of a quantity in question such as the soil strength, but rather a characteristic value which is then usually defined as a particular quantity associated with the probability distribution. Examples of such quantities which are often used as characteristic values are:

— mean value
— a low quantile in the probability distribution, e.g. the 5% quantile
— mean value minus two standard deviations, which for a normal distribution corresponds to the 2.3% quantile
— most probable value, i.e. the value for which the probability density function is maximum.

Guidance note:

Which definition of the characteristic value applies usually depends on the design code and on the actual application.

For problems which are governed by a local soil strength value, such as the pile tip resistance for an end-bearing pile, a low quantile in the strength distribution is often specified as characteristic value.

For problems, which involve large soil volumes where local strength variations from point to point can be assumed to average out, such as in stability calculations for large gravity-base foundations, the mean value of the soil strength is often specified as characteristic value.

---e-n-d---of---G-u-i-d-a-n-c-e---n-o-t-e---

2.3.5.2 For the case that the distribution of the considered soil strength is estimated based on limited test data such that the distribution parameters become encumbered with statistical uncertainty, also estimates of characteristic values of the distribution become statistically uncertain. To properly account for such statistical uncertainty, it is common to specify that characteristic values shall be estimated with a specified confidence.

2.3.5.3 Estimation with confidence implies that instead of using a central estimate of the quantity in question, a conservative estimate value is used such that the probability of the true value being more favourable than the conservative estimate is at least equal to the specified confidence.

Guidance note:

When the characteristic value in a design code is defined as the 5% quantile in the probability distribution, then an additional requirement usually is that it shall be estimated with a specified confidence. A confidence of 95% is a typical requirement in many codes.

Some design codes, which use the mean value as characteristic value, specify that the characteristic value shall be taken as “a conservatively assessed mean value” or “a cautiously estimated mean value”. The interpretation of this specification is that the characteristic value shall be taken as the mean value estimated with a confidence greater than 50%. Typical confidences would then be 75% or 90%.

---e-n-d---of---G-u-i-d-a-n-c-e---n-o-t-e---

2.3.5.4 Some geotechnical practices define the characteristic value as a lower bound value with reference to some available data set of n observations of the parameter in question.

Guidance note:

Note that a lower bound value is a subjective measure, and if it is literally interpreted as the smallest value among n observed values, a smaller value is likely to be obtained if the number of observations n is increased. This is so, because the larger the number of observations, the higher is the probability that one of them will be made further out

on the lower tail of the distribution. Hence by using the observed lowest value as characteristic value, one may actually loose the benefit of carrying out a larger number of observations.

Note also that a lower bound by its mere definition is not necessarily invariant with the number of observations *n* and does thus not necessarily result in a specific quantile whose exceedance probability is independent of *n*.

A lower bound value specified in a soil report to be used as characteristic value for design is not necessarily a physical lower bound according to the mathematical definition, but rather a recommended low value to be used in design.

It is therefore recommended that the designer always makes sure that such specified lower bound values meet the requirements to the characteristic value as set forth in the design codes and standards which are to be used in design.

When a characteristic value in terms of a so-called lower bound is to be established and is to be presented in a soil report as a lower bound for use in design, it is recommended that it be established on the basis of statistical methods in accordance with the examples of definitions given in 2.3.5.1. A mean value should be used as characteristic value when the average of the soil property in question governs the design, whereas a low quantile should be used when a point value of the property governs the design. An adequate confidence should be used for the estimation of the characteristic value from data.

---e-n-d---of---G-u-i-d-a-n-c-e---n-o-t-e---

2.4 Central estimation of soil parameters

2.4.1 Parameter estimation for independent soil variables

2.4.1.1 Let X be a soil variable, such as a strength, for which *n* observations exist, $x_1,...x_n$. The mean value of X is μ and the standard deviation is σ. Any dependency on depth or on any other conditioning quantity is disregarded here; i.e. X is a so-called independent variable.

2.4.1.2 The mean value μ is estimated by the sample mean

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

2.4.1.3 The standard deviation σ is estimated by the sample standard deviation

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

2.4.1.4 The estimates $\bar{x}$ and s^2 are central estimates of μ and σ^2.

2.4.1.5 The estimates $\bar{x}$ and *s* are encumbered with uncertainty owing to the limited number of data used for the estimation. The uncertainties in the estimates can be represented by the respective standard deviations, denoted standard errors.
If needed, these standard errors can be estimated by

$$se(\bar{x}) = \frac{s}{\sqrt{n}} \text{ and } se(s) = s\sqrt{\frac{\alpha_4 - 1}{4n}}$$

There is also an expression for the standard error of the variance estimate s^2,

$$se(s^2) = s^2\sqrt{\frac{2}{n-1}}$$

2.4.1.6 The median x_m of X is estimated by the sample median. Let the *n* observations of X be arranged in increasing order, $x_1,...x_n$. When *n* is odd, the sample median is the middle value $x_{(n+1)/2}$ in the ordered list of observations. When *n* is even, the sample median is the average of the two middle values $x_{n/2}$ and $x_{(n+2)/2}$ in this list.

Guidance note:

For the special case that the distribution of X is known to be a normal distribution, the median x_m can be estimated by the sample mean

$$\hat{x}_m = \frac{1}{n}\sum_{i=1}^{n} x_i$$

hence utilising that the symmetry of the normal distribution implies that the median is equal to the mean.

For the special case that the distribution of X is known to be a lognormal distribution, the median x_m can be estimated by the sample geometric mean

$$\hat{x}_m = \sqrt[n]{\prod_{i=1}^{n} x_i}$$

Note that the lognormal distribution is not a symmetrical distribution, such that the mean will be different from the median and such that the geometric mean of the observations used to estimate the median *cannot* be used as an estimate of the mean.

Note also that the two median estimates presented for special cases in this guidance note are more accurate estimates than the sample median, because they make use of more information than the sample median does, including the information about distribution type.

---e-n-d---of---G-u-i-d-a-n-c-e---n-o-t-e---

2.4.2 Parameter estimation for dependent soil variables

2.4.2.1 Let Y be a soil variable, such as a strength parameter, whose variation with depth z can reasonably well be modelled as a linear function. Y is then a so-called dependent variable, whereas z is independent. Assume that n observations of pairs (z_i,y_i), $i = 1,...n$, are available from a soil investigation campaign. The variation of Y with depth can now be expressed as

$$Y = a_0 + a_1 z + \varepsilon$$

where the term a_0+a_1z represents the linear mean variation with depth and the term ε represents the natural variability of Y about the mean, and where z denotes the depth below the soil surface. The coefficients a_0 and a_1 represent the surface intercept and the depth gradient, respectively, of the mean of Y. The variability term ε has zero mean and its standard deviation is denoted σ. The standard deviation σ is assumed to be constant, i.e. it is assumed to be independent of depth z. Reference is made to Figure 2-6.

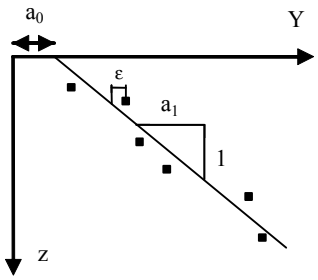

Figure 2-6
Data points (z,Y) and linear model for mean

2.4.2.2 The coefficients a_0 and a_1 in the linear expression can be estimated from the n observed data pairs (z,y) as

$$\hat{a}_1 = \frac{\sum_{i=1}^{n}(z_i - \bar{z})(y_i - \bar{y})}{\sum_{i=1}^{n}(z_i - \bar{z})^2} \text{ and } \hat{a}_0 = \bar{y} - \hat{a}_1\bar{z}$$

in which

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \text{ and } \bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i$$

Guidance note:
The estimates $\hat{a}_0$ and $\hat{a}_1$ are known as best estimates of a_0 and a_1, respectively, and thus provide the best estimate of the straight line that expresses the mean variation of the variable Y with the depth z.

---e-n-d---of---G-u-i-d-a-n-c-e---n-o-t-e---

2.4.2.3 The standard deviation σ of the natural variability term ε is estimated by the sample standard deviation

$$s = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(y_i - (\hat{a}_0 + \hat{a}_1 z_i))^2}$$

2.4.2.4 The estimators $\hat{a}_0$, $\hat{a}_1$ and s are encumbered with uncertainty owing to the limited number of data used for the estimation. The uncertainties in the estimators can be represented by the respective standard deviations, denoted standard errors. If needed, these standard errors can be estimated by

$$se(\hat{a}_0) = s \cdot \sqrt{\frac{1}{n} + \frac{\bar{z}^2}{\sum_{i=1}^{n}(z_i - \bar{z})^2}}$$

$$se(\hat{a}_1) = \frac{s}{\sqrt{\sum_{i=1}^{n}(z_i - \bar{z})^2}}$$

$$se(s^2) = s^2\sqrt{\frac{2}{n-2}}$$

Note that it is the standard error of s^2 rather than of s which is given here.

2.4.2.5 The estimators in 2.4.2.2 may be recognised as the results from a least-squares regression analysis.

2.4.2.6 Situations exist for which the natural variability of the considered soil parameter Y is such that the standard deviation σ of the random variability term ε is not a constant as assumed in 2.4.2.1 through 2.4.2.4, but rather has a variation with depth. A standard deviation which is proportional with depth is commonly encountered. For the case that the standard deviation σ is proportional with the depth z, σ = kz, the coefficients a_0 and a_1 in the linear expression for the mean of Y can be estimated from the n observed data pairs (z,y) as

$$\hat{a}_1 = \frac{\sum_{i=1}^{n} w_i z_i y_i - \frac{\sum_{i=1}^{n} w_i z_i \cdot \sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i}}{\sum_{i=1}^{n} w_i z_i^{\,2} - \frac{(\sum_{i=1}^{n} w_i z_i)^2}{\sum_{i=1}^{n} w_i}}$$

$$\hat{a}_0 = \frac{\sum_{i=1}^{n} w_i y_i - \hat{a}_1 \sum_{i=1}^{n} w_i x_i}{\sum_{i=1|}^{n} w_i}$$

in which $w_i = 1/z_i^2$, where z_i denotes the depth of the ith observation y_i of Y.

2.4.2.7 The proportionality constant k in the expression for the natural variability term is estimated by

$$\hat{k} = \sqrt{\frac{1}{n-2} \sum_{i=1}^{n} w_i (y_i - (\hat{a}_0 + \hat{a}_1 z_i))^2}$$

2.4.2.8 The estimators $\hat{a}_0$, $\hat{a}_1$ and $\hat{k}$ are encumbered with uncertainty owing to the limited number of data used for the estimation. The uncertainties in the estimators can be represented by the respective standard deviations, denoted standard errors. If needed, these standard errors can be estimated by

$$se(\hat{a}_0) = \hat{k} \cdot \sqrt{\frac{\sum_{i=1}^{n} w_i z_i^{\,2}}{\sum_{i=1}^{n} w_i \cdot \sum_{i=1}^{n} w_i z_i^{\,2} - (\sum_{i=1}^{n} w_i z_i)^2}}$$

$$se(\hat{a}_1) = \hat{k} \cdot \sqrt{\frac{\sum_{i=1}^{n} w_i}{\sum_{i=1}^{n} w_i \cdot \sum_{i=1}^{n} w_i z_i^{\,2} - (\sum_{i=1}^{n} w_i z_i)^2}}$$

$$se(\hat{k}^2) = \hat{k}^2 \sqrt{\frac{2}{n-2}}$$

Note that it is the standard error of $\hat{k}^2$ rather than of $\hat{k}$ which is given here.

2.4.3 Calculation example for independent soil variable

2.4.3.1 Soil Data Set #1 consists of 22 observations of the undrained shear strength s_u in a 34 m deep soil deposit, see Figure 2-7. When the ultimate goal is to design a friction pile for installation in this soil deposit to a depth of about 34 m, a visual inspection of the data indicates that it will be a reasonable model to represent the mean strength as a constant value, independent of depth, within this soil deposit.

Guidance note:

The application, i.e. primarily the type of foundation, governs which soil properties are of interest to be estimated. For an axially loaded friction pile as considered in this example, the mean strength over the length of the pile is of interest, forming the basis for prediction of the axial capacity of the pile.

For a shallow foundation to be designed against horizontal sliding or bearing capacity failure, only the strength in a shallow zone below the soil surface would be of interest, and a subset of the 22 strength observations in this example, consisting of the 5 or 6 uppermost observations, would have to be subject to inspection, analysis and interpretation. As the available soil data often reflect the type and size of foundation to be installed, if this is known at the time of the soil investigations, it is likely that the obtained data set would have been different if a shallow foundation were to be installed in the soil deposit rather than a friction pile: The strength observations would have been concentrated in a shallow zone within the soil deposit rather than being approximately uniformly distributed over the 34 m depth of the soil deposit.

It is a prerequisite for application of the data set in this example that the definition of s_u associated with the data set is consistent with the definition of s_u in the design code which is used to predict the axial pile capacity.

---e-n-d---of---G-u-i-d-a-n-c-e---n-o-t-e---

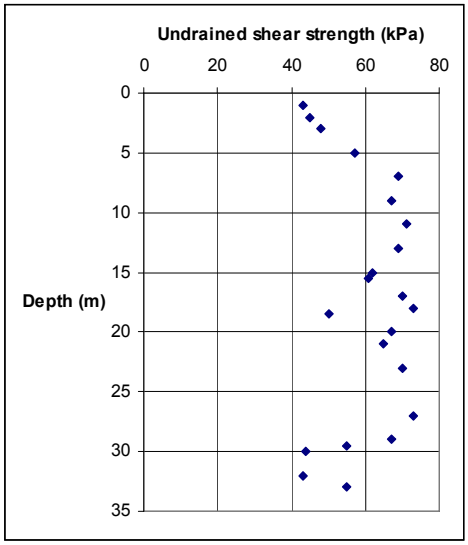

Figure 2-7
Soil Data Set #1: Data for undrained shear strength vs. depth

2.4.3.2 The mean value μ of s_u is estimated by the sample mean of the data, hence

$$\hat{\mu} = \bar{s}_u = \frac{1}{22}\sum_{i=1}^{n} s_{u,i} = 60.2 \text{ kPa}$$

The standard deviation σ of s_u is estimated by the sample standard deviation, hence

$$\hat{\sigma} = s = \sqrt{\frac{1}{22-1}\sum_{i=1}^{22}(s_{u,i} - \bar{s}_u)^2} = 10.6 \text{ kPa}$$

This gives a variance estimate $s^2 = 112.3$ kPa^2. The estimated coefficient of variation is COV = 10.6/60.2 = 0.176.

2.4.3.3 Provided an ideal, accurate method of measuring s_u was applied for observation of the 22 data values, s = 10.6 kPa will be the final estimate of the standard deviation associated with the natural variability of the undrained shear strength s_u. Usually, this is what should be assumed, regardless of the accuracy of the method of measurement, since this accuracy is usually either unknown or is difficult to quantify.

2.4.3.4 Provided the accuracy of the method of measuring s_u is known, e.g. from a calibration of the test equipment, the associated uncertainty can in principle be eliminated from the estimate s, and the final estimate of the standard deviation associated with the natural variability of the undrained shear strength s_u will come out somewhat smaller than the apparent standard deviation of 10.6 kPa. Consider as an example that the standard deviation associated with the accuracy of the measurement method is quantified to $\sigma_{meas} = 6$ kPa. For this case, the net standard deviation associated with the natural variability of s_u becomes reduced to

$$s = \sqrt{10.6^2 - 6^2} = 8.7 \text{ kPa}$$

and the estimated coefficient of variation correspondingly reduces to COV = 8.7/60.2 = 0.145.

2.4.4 Calculation example for dependent soil variable

2.4.4.1 Soil Data Set #2 consists of 51 observations of the undrained shear strength which are available from the depth range 2-39 m in a clay deposit, see Figure 2-8. The data are obtained by two different types of tests, viz. direct simple shear tests and vane tests. The data from the vane tests have been properly transformed to form a consistent population with the data from the direct simple shear tests, cf. 2.4.6.

2.4.4.2 By a visual inspection, the data are seen on average to have a linear variation with depth, $E[s_u] = a_0 + a_1 z$.

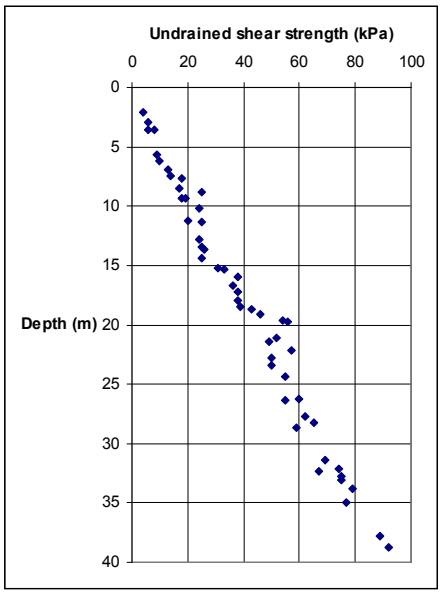

Figure 2-8
Soil Data Set #2: Data for undrained shear strength varying linearly with depth

The variability in the undrained shear strength about this mean is characterised by the standard deviation σ. For an assumption of constant standard deviation with depth, the formulas in 2.4.2.2 gives the following estimates of the coefficients a_0 and a_1

$$\hat{a}_0 = -2.22 \text{ kPa and } \hat{a}_1 = 2.35 \text{ kPa/m}$$

and the standard deviation of the residuals in s_u is estimated by the sample standard deviation in 2.4.2.3

$$\hat{\sigma} = s = 3.76 \text{ kPa}$$

Guidance note:

A negative strength intercept $\hat{a}_0 = -2.22$ kPa at the soil surface has been estimated. This is unphysical and serves as a reminder that the linear model for the strength as a function of depth is valid only within the depth range of the data, i.e. 2-39 m. Within this range, the estimated coefficients will result in only positive mean strength values. The example demonstrates that there may be limitations with respect to extrapolation of statistical estimation results outside the range of the data used for the estimation.

---e-n-d---of---G-u-i-d-a-n-c-e---n-o-t-e---

2.4.5 Outliers

2.4.5.1 Data observations, which are quite far from the bulk of data in a data set, are called outliers. Before carrying out a statistical analysis of a data set, the data set should be inspected for outliers. As a rule of thumb, any data value which is located more than two standard deviations away from the mean value of the data set can be considered an outlier. Graphical presentation is an effective means of detection. Rational, more elaborate methods for identification of outliers can be found in Snedecor and Cochran (1989) and Neter et al. (1990).

2.4.5.2 It is recommended to remove data points that are identified as outliers. However, one should always justify removal of identified outliers by physical evidence, e.g. equipment malfunction, improperly calibrated equipment, and recording and transcription errors. Also evident physical differences between the parent test specimen for a data value identified as an outlier on the one hand and the parent test specimens for the rest of the data set on the other, such as a significantly different sand content, may serve to justify the removal of the identified outlier.

2.4.5.3 Caution should be exercised not to remove data points, which may appear as outliers, but which do not qualify as outliers and which should therefore be kept in the data set. An example of such a data point is a data point far from the bulk of data points in an (X,Y) plane and far from the line that can be estimated from the data for Y's dependency on X. The linear relationship assumed between X and Y might simply not be valid in the region of the (X,Y) plane where the suspicious data point is located.

2.4.6 Data from more than one source

2.4.6.1 Caution should be exercised when data from two or more sources are pooled together to form one data set which is then subject to statistical analysis. For example, soil strength data may arise from different types

of tests on soil samples. Data from different types of tests have to be converted to the same reference scale before any statistical analysis of the pooled data set will be meaningful.

Guidance note:

Data for undrained shear strength of clay are often available from UU tests and CIU tests. To convert UU test results to obtain benchmark CIU test results that can be pooled with results from actual CIU tests, it is common to multiply the strength data from the UU tests by a factor. This factor is often set equal to 1.2.

Conversion of other types of data before pooling may require other transformations than simple multiplication by a conversion factor.

---e-n-d---of---G-u-i-d-a-n-c-e---n-o-t-e---

2.4.7 Bayesian approach to estimation

2.4.7.1 The Bayesian approach to estimation, which capitalises on Bayesian statistics and Bayesian updating, forms an alternative to estimation by the classical statistical approach as presented in 2.4.1 through 2.4.4. The Bayesian approach to estimation allows for inclusion of subjective information and judgment in estimation, e.g. in terms of assumptions of a prior distribution of an unknown parameter to be estimated. According to the Bayesian approach this distribution is then subsequently updated conditional on available sampled data.

2.4.7.2 The Bayesian approach to estimation is useful for estimation in situations where the classical statistical methods may not necessarily suffice, e.g. when data from different sources are to be combined. The Bayesian approach also allows for inclusion of other information than sample data and it allows for updating of prior information when new additional information becomes available.

2.4.7.3 For further details of Bayesian estimation reference is made to Ang and Tang (1975). For further details of Bayesian updating reference is made to Classification Notes No. 30.6.

2.5 Estimation of soil parameters with confidence

2.5.1 Parameter estimation with confidence, independent variables

2.5.1.1 Let X be a soil variable, such as a strength, for which n observations exist, $x_1,...x_n$. When the characteristic value of X is defined as the mean value of X, and when the characteristic value is required to be estimated with a confidence of $1-\alpha$, the following expression for the estimate of the characteristic value applies

$$X_C = \bar{x} - t_{n-1}(\alpha) \cdot \frac{s}{\sqrt{n}}$$

in which $\bar{x}$ is the sample mean of X and s is the sample standard deviation of X based on the n observations, and $t_{n-1}(\alpha)$ is given in Table 2-1. This is based on theory for one-sided confidence intervals. The sample mean and sample standard deviation are to be calculated according to expressions given in 2.4.1.2 and 2.4.1.3.

Table 2-1 Coefficient $t_{n-1}(\alpha)$ for X_C defined as mean value of X

n–1	*1 – α = 0.75*	*1 – α = 0.90*	*1 – α = 0.95*
2	0.82	1.89	2.92
5	0.73	1.48	2.02
10	0.70	1.37	1.81
15	0.69	1.34	1.75
20	0.69	1.33	1.72
30	0.68	1.31	1.70
50	0.68	1.30	1.68

2.5.1.2 When the characteristic value of X is defined as the 5% quantile in the distribution of X, and when the characteristic value is required to be estimated with a confidence of $1-\alpha$, the following expression for the estimate of the characteristic value applies

$$X_C = \bar{x} - c_{1-\alpha}(n) \cdot s$$

in which $\bar{x}$ is the sample mean of X and s is the sample standard deviation of X based on the n observations, and $c_{1-\alpha}(n)$ is given in Table 2-2. This is based on theory for tolerance bounds. The sample mean and sample standard deviation are to be calculated according to expressions given in 2.4.1.2 and 2.4.1.3.

Table 2-2 Coefficient $c_{1-\alpha}(n)$ for X_C defined as the 5% quantile of X

n	*1 – α = 0.75*	*1 – α = 0.90*	*1 – α = 0.95*
3	3.17	5.33	7.66
5	2.47	3.41	4.21
10	2.11	2.57	2.91

Table 2-2 Coefficient $c_{1-\alpha}(n)$ for X_C defined as the 5% quantile of X (Continued)			
n	$1-\alpha = 0.75$	$1-\alpha = 0.90$	$1-\alpha = 0.95$
15	1.99	2.34	2.57
20	1.94	2.21	2.40
30	1.88	2.08	2.22
∞	1.64	1.64	1.64

2.5.1.3 Formulas and tables in 2.5.1.1 and 2.5.1.2 are shown for a single variable such as a strength which can be assumed constant within a soil layer. Other formulas and tables apply for soil properties that vary linearly with depth, see 2.5.2.1 and 2.5.2.2.

2.5.2 Parameter estimation with confidence, dependent variables

2.5.2.1 Let Y be a soil variable, such as a strength, whose variation with depth z on average can reasonably well be modelled as a linear function. Assume that *n* observations of pairs (z_i,y_i), $i = 1,...n$, are available from a soil investigation campaign. The mean variation of Y with depth is expressed as a_0+a_1z, in which the coefficients a_0 and a_1 can be estimated from the observed data pairs as given in 2.4.2.2.

When the characteristic value of Y is defined as the mean value of Y, and when the characteristic value is required to be estimated with a confidence of $1-\alpha$, the following expression may be applied for the characteristic value

$$Y_C = \hat{a}_0 + \hat{a}_1 z - t_{n-2}(\alpha)\cdot s \cdot \sqrt{\frac{1}{n} + \frac{3n}{n^2-1}}$$

Table 2-3 Coefficient $t_{n-2}(\alpha)$ for Y_C defined as mean value of Y			
$n-2$	$1-\alpha = 0.75$	$1-\alpha = 0.90$	$1-\alpha = 0.95$
2	0.82	1.89	2.92
5	0.73	1.48	2.02
10	0.70	1.37	1.81
15	0.69	1.34	1.75
20	0.69	1.33	1.72
30	0.68	1.31	1.70
50	0.68	1.30	1.68

This expression is a linear approximation to the true expression, which results from theory and which is somewhat nonlinear with respect to depth z. The approximate expression is based on the assumption that the *n* data pairs are sampled approximately uniformly over the depth range considered. It gives slightly conservative results in the central part of this depth range, and it is not suitable for extrapolation to characteristic values outside of this depth range. The coefficient estimates $\hat{a}_0$ and $\hat{a}_1$ and the sample standard deviation *s* are to be calculated according to the expressions given in 2.4.2.2 and 2.4.2.3, and $t_{n-2}(\alpha)$ is given in Table 2-3. This is based on theory for one-sided confidence intervals.

2.5.2.2 When the characteristic value of Y is defined as the 5% quantile in the distribution of Y, and when the characteristic value is required to be estimated with a confidence of $1-\alpha$, the following expression for the estimate of the characteristic value applies

$$Y_C = \hat{a}_0 + \hat{a}_1 z - c_{1-\alpha}(n)\cdot s$$

This expression is a linear approximation to the true, somewhat nonlinear expression that results from theory. It is based on the assumption that the *n* data pairs are sampled approximately uniformly over the depth range considered. It gives slightly conservative results in the central part of this depth range, and it is not suitable for extrapolation to characteristic values outside of this depth range. The coefficient $c_{1-\alpha}(n)$ is tabulated in Table 2-4 for various confidence levels $1-\alpha$. For details about the methodology, reference is made to Ronold and Echtermeyer (1996).

Table 2-4 Coefficient $c_{1-\alpha}(n)$ for Y_C defined as the 5% quantile of Y			
n	$1-\alpha = 0.75$	$1-\alpha = 0.90$	$1-\alpha = 0.95$
3	5.2	13.8	27.7
5	2.88	4.41	5.80
10	2.29	2.95	3.44
15	2.14	2.62	2.95

Table 2-4 Coefficient $c_{1-\alpha}(n)$ for Y_C defined as the 5% quantile of Y (Continued)			
n	$1-\alpha = 0.75$	$1-\alpha = 0.90$	$1-\alpha = 0.95$
20	2.06	2.45	2.70
50	1.88	2.10	2.25
100	1.81	1.96	2.06
∞	1.64	1.64	1.64

2.5.2.3 The results presented here for soil properties that vary linearly with depth are based on an assumption of a constant standard deviation of the considered soil property with depth. However, some soil properties exhibit a standard deviation which is proportional with depth. For this case the coefficient $c_{1-\alpha}(n)$ will be a function of depth. It will also depend on how the data are distributed over the depth range considered. Tabulated values for $c_{1-\alpha}(n)$ are not available.

2.5.3 Calculation example for independent soil variable

2.5.3.1 Consider Soil Data Set #1 where $n = 22$ observations of the undrained shear strength s_u in a soil deposit are available, see Figure 2-7. The mean value is estimated to be $\hat{\mu} = 60.2$ kPa and the standard deviation is estimated to be s = 10.2 kPa.

2.5.3.2 A geotechnical problem is encountered for which the characteristic value of the undrained shear strength is to be taken as the mean value estimated with 90% confidence. According to the expression in 2.5.1.1 this characteristic value becomes

$$s_{u,C} = \hat{\mu} - t_{n-1}(\alpha)\cdot\frac{s}{\sqrt{n}} = 60.2 - 1.33\cdot\frac{10.6}{\sqrt{22}} = 57.2\,\text{kPa}$$

2.5.3.3 Another geotechnical problem is encountered for which the characteristic value of the undrained shear strength is to be taken as the 5% quantile with 95% confidence. According to the expression in 2.5.1.2, this characteristic value becomes

$$s_{u,c} = \hat{\mu} - c_{1-\alpha}(n)\cdot s = 60.2 - 2.36\cdot 10.6 = 35.2\,\text{kPa}$$

2.5.4 Calculation example for dependent soil variable

2.5.4.1 Consider Soil Data Set #2 where $n = 51$ observations of the undrained shear strength s_u are available from the depth range 2-39 m in a clay deposit, see Figure 2-8. The mean value is represented as a linear function of depth z, $E[s_u] = a_0 + a_1 z$. The coefficients in the expression for $E[s_u]$ are estimated to be $\hat{a}_0 = -2.22$ kPa and $\hat{a}_1 = 2.35$ kPa/m, respectively, and the standard deviation of the residuals in s_u is estimated by $\hat{\sigma} = s = 3.76$kPa.

2.5.4.2 Consider the situation that the characteristic value of s_u is to be taken as the mean value of s_u estimated with 95% confidence. According to the expression in 2.5.2.1, the characteristic s_u profile then becomes

$$\begin{aligned} s_{u,C} &= \hat{a}_0 + \hat{a}_1 z - t_{n-2}(\alpha)\cdot s\cdot\sqrt{\frac{1}{n}+\frac{3n}{n^2-1}} \\ &= -2.22 + 2.35z - 1.68\cdot 3.76\cdot\sqrt{\frac{1}{51}+\frac{3\cdot 51}{51^2-1}} \\ &= -3.99 + 2.35z \end{aligned}$$

2.5.4.3 Consider the situation that the characteristic value of s_u is to be taken as the 5% quantile of s_u estimated with 95% confidence. According to the expression in 2.5.2.2, the characteristic s_u profile then becomes

$$\begin{aligned} s_{u,C} &= \hat{a}_0 + \hat{a}_1 z - c_{1-\alpha}(n)\cdot s \\ &= -2.22 + 2.35z - 2.25\cdot 3.76 \\ &= -10.68 + 2.35z \end{aligned}$$

Guidance note:

The resulting expression for $s_{u,C}$ gives negative characteristic shear strength in a depth range between the soil surface and about 4.5 m depth. This is unphysical and may be interpreted as an indication that the normal distribution assumption for the undrained shear strength may not necessarily be a good model in the uppermost part of the soil deposit. In the actual case, if it is maintained that the 5% quantile shall be estimated with 95% confidence, the distribution assumption should be revisited and a truncated normal distribution should be considered instead of the normal distribution before the characteristic strength is estimated with the specified confidence. Otherwise, it may be considered to put less emphasis on the statistical interpretation and rather use physical evidence to establish the characteristic value as a physical lower bound. As an example, in the case of the strength of a clay, such physical evidence could consist of knowledge about the thixotropic properties of the clay.

---e-n-d---of---G-u-i-d-a-n-c-e---n-o-t-e---

2.5.5 Minimum sample number

2.5.5.1 From the soil samples obtained during the initial phase of a soil investigation program, it is possible to determine whether or not the number of samples collected was sufficient for the accuracy of prediction required by the design engineer. If not, then an additional phase of soil investigation will be necessary so that additional samples can be obtained.

2.5.5.2 The prediction accuracy required by the design engineer may be expressed as a range of values within which the true but unknown mean value of the population will lie with a probability equal to or greater than a specified value. This specified probability may be recognised as the confidence and the range of values required for accuracy can then be expressed as a two-sided confidence interval,

$$\left[\bar{x} - t_{n-1}(\frac{\alpha}{2}) \cdot \frac{s}{\sqrt{n}}; \bar{x} + t_{n-1}(\frac{\alpha}{2}) \cdot \frac{s}{\sqrt{n}}\right],$$

in which $\bar{x}$ and s are the sample mean and the sample standard deviation, respectively, of the soil parameter in question. The confidence is $1-\alpha$ and $t_{n-1}(\alpha/2)$ is given in Table 2-5.

Table 2-5 Coefficient $t_{n-1}(\alpha/2)$

$n-1$	$1-\alpha = 0.80$	$1-\alpha = 0.90$	$1-\alpha = 0.95$
2	1.89	2.92	4.30
3	1.64	2.35	3.18
4	1.53	2.13	2.78
5	1.48	2.02	2.57
8	1.40	1.86	2.31
10	1.37	1.81	2.23
15	1.34	1.75	2.13
20	1.33	1.72	2.09
30	1.31	1.70	2.04
50	1.30	1.68	1.96

Guidance note:

Assume as an example that four observations of the undrained shear strength are available from tests on four soil samples from a stiff clay deposit: 93, 100, 104 and 107 kPa. The minimum number of soil samples required so that the true, unknown mean value will be within 5% of the mean test result with at least 90% probability is sought.

The mean value estimate is $\bar{x} = 101$ kPa and the standard deviation estimate is $s = 6.06$ kPa. The limits of the 90% confidence interval become $101 \pm 2.35 \cdot 6.06/\sqrt{4} = 93.9$ to 108.1 kPa, i.e. some 7% on either side of the mean.

Four samples are thus not sufficient if the true mean shall be within 5% of the mean test result. This is an interval from 96 to 106 kPa. By combining the expression for the confidence interval with Table 2-5, it can be determined that an interval equal to or narrower than the required interval from 96 to 106 kPa will be achieved when the number of samples is increased by two to a total of six samples, i.e. two more samples must be obtained and tested.

---e-n-d---of---G-u-i-d-a-n-c-e---n-o-t-e---

2.5.6 Bayesian approach to estimation with confidence

2.5.6.1 As an alternative to the classical statistical approach to estimation with confidence presented in 2.5.1 through 2.5.4, a Bayesian approach to estimation with confidence may be applied. This requires that a suitable loss function be chosen. The choice of loss function governs the confidence associated with the estimate that results from the Bayesian estimation. The choice of the right loss function is thus crucial in order to arrive at an estimate that has the desired confidence.

For details of Bayesian estimation with confidence, reference is made to Ang and Tang (1984).

2.6 Combination of variances

2.6.1 Combination rules

2.6.1.1 Variances associated with different uncertainty sources are combined by simple addition when the uncertainty sources are additive and statistically independent.

2.6.1.2 Let X be a soil variable such as a strength. X will exhibit a natural variability from point to point within a soil deposit. The mean and variance of X are estimated on the basis of n observations by measurements on soil samples, so the estimates of the mean and variance of X are encumbered with measurement uncertainty as well as with statistical uncertainty. The measurement uncertainty and the statistical uncertainty are both epistemic uncertainties which are independent of the natural variability of X. In addition the measurement uncertainty and the statistical uncertainty can be assumed to be statistically independent.

2.6.1.3 The observed variance σ_{obs}^2, which is estimated from the n observations of X, forms an apparent variance which is a combination of a variance σ_X^2 associated with natural variability and a variance σ_{meas}^2 associated with measurement uncertainty,

$$\sigma_{obs}^{\ 2} = \sigma_X^{\ 2} + \sigma_{meas}^{\ 2}$$

The total variance associated with the determination of X will include a variance σ_{stat}^2 associated with statistical uncertainty owing to limited data, hence

$$\sigma_{tot}^{\ 2} = \sigma_{obs}^{\ 2} + \sigma_{stat}^{\ 2} = \sigma_X^{\ 2} + \sigma_{meas}^{\ 2} + \sigma_{stat}^{\ 2}$$

If some model is applied to transform n measurements on soil samples to n observations of X before the estimates of the mean and variance of X are established from the data, a model uncertainty will be present. Examples of this are vane test data which are transformed to undrained shear strengths and CPT readings which are converted to undrained shear strengths. For such cases, the variance σ_{model}^2 associated with the applied conversion model are included in the observed variance σ_{obs}^2 and therefore appears in the expressions in the same manner as σ_{meas}^2, hence

$$\sigma_{obs}^{\ 2} = \sigma_X^{\ 2} + \sigma_{meas}^{\ 2} + \sigma_{\text{model}}^{\ 2}$$

and

$$\sigma_{tot}^{\ 2} = \sigma_{obs}^{\ 2} + \sigma_{stat}^{\ 2} = \sigma_X^{\ 2} + \sigma_{meas}^{\ 2} + \sigma_{\text{model}}^{\ 2} + \sigma_{stat}^{\ 2}$$

2.6.1.4 When the variance σ_{meas}^2 associated with the measurement uncertainty can be quantified, it can in principle be eliminated from the observed variance σ_{obs}^2 and the variance σ_X^2 associated with the natural variability of X can be isolated,

$$\sigma_X^{\ 2} = \sigma_{obs}^{\ 2} - \sigma_{meas}^{\ 2}$$

In the case that also a model uncertainty is involved in establishing the n observations of X, the variances associated with both the measurement uncertainty and the model uncertainty need to be quantified in order that the variance associated with the natural variability of X can be isolated,

$$\sigma_X^{\ 2} = \sigma_{obs}^{\ 2} - \sigma_{meas}^{\ 2} - \sigma_{\text{model}}^{\ 2}$$

Guidance note:

It can be dangerous to perform the reduction in the observed variance for possible measurement and model uncertainties.

Caution should be exercised *not* to carry out the favourable reduction of the observed variance σ_{obs}^2 by variances associated with measurement and model uncertainty, unless it is absolutely certain that these uncertainties are present in the data and their variances can be properly quantified and trusted. Otherwise, one may be unfortunate to underestimate σ_X^2 from σ_{obs}^2.

---e-n-d---of---G-u-i-d-a-n-c-e---n-o-t-e---

2.6.1.5 The autocorrelation function $\rho(r)$ for the variable X considered as a random field over the soil deposit can in principle be estimated from spatial observations of X. The estimate of this autocorrelation function can be applied to provide information which can be used to isolate the variance σ_X^2 from the observed variance σ_{obs}^2.

The autocorrelation function shall theoretically attain the value $\rho = 1$ for lag $r = 0$. However, when the autocorrelation function is estimated from measured values of X for various lags $r > 0$, the results can be extrapolated to estimate the value of ρ for $r = 0$. The result of this exercise will usually be that $\rho \to b$ for $r \to 0$, where $b < 1$. The value of the limit b being less than 1 reflects the presence of measurement uncertainty and a possible model uncertainty as described in 2.6.1.2 and 2.6.1.3, and the variance associated with the natural variability of X can be interpreted as

$$\sigma_X^{\ 2} = b \cdot \sigma_{obs}^{\ 2}$$

Caution must be exercised to avoid underestimation of b and thereby of σ_X^2 when b is extrapolated from available estimated $\rho(r)$ values. Estimated $\rho(r)$ values will often be encumbered with large uncertainties, which together with the nonlinearity of $\rho(r)$ with r makes the extrapolation difficult.

2.6.2 Calculation example for independent soil variable

2.6.2.1 Consider Soil Data Set #1 where $n = 22$ observations of the undrained shear strength s_u in a soil layer are available, see Figure 2-7. The mean value of s_u is estimated to be $\hat{\mu} = 60.2$ kPa and the standard deviation of s_u is estimated to be $s = 10.2$ kPa.

2.6.2.2 There is statistical uncertainty associated with the estimates of the mean value and the standard deviation of s_u because only 22 observations of s_u were available for the estimation. The standard error in the estimate of the mean value is

$$se(\bar{s}_u) = \frac{10.6}{\sqrt{22}} = 2.3 \text{ kPa}$$

and the standard error in the estimate of the standard deviation is

$$se(s) = 10.6\sqrt{\frac{2.22-1}{4\cdot 22}} = 1.25 \text{ kPa}$$

which is based on a kurtosis estimate of 2.22.

2.6.2.3 When the soil variable s_u is not known in full by its distribution with known mean value and known standard deviation, but only to the extent that the mean value and standard deviation have been estimated from limited data, it has an associated uncertainty which stems not only from its natural variability but also from the statistical uncertainty associated with the estimates of the mean and the standard deviation. The estimate of the total variance associated with s_u can then be expressed as

$$\begin{aligned}\hat{\sigma}_{tot}^{\ 2} &= s^2 + se(\bar{s}_u)^2 + se(s)^2 \\ &= 10.6^2 + 2.3^2 + 1.25^2 \\ &= 10.9^2 \text{ kPa}^2\end{aligned}$$

2.7 Geostatistics

2.7.1 Spatial averaging

2.7.1.1 When the application of the information about the soil variable X implies that a spatial average of X over large soil volumes or areas shall be used rather than X itself, then the full variance σ_X^2 shall not be used in calculations but rather a reduced variance

$$\sigma_{X_{spatial}}^{\ 2} = \lambda \sigma_X^{\ 2}$$

in which $\lambda < 1$ is a variance reduction factor which reflects the extent of the spatial averaging and which depends on the autocorrelation function for X and the geometry of the involved soil volume or area. In applications, this reduced variance needs to be superimposed by the variance σ_{stat}^2 associated with possible statistical uncertainty owing to limited data and by a variance $\sigma_{applmodel}^2$ associated with a possible model uncertainty involved with the application. When full averaging takes place, i.e. when local fluctuations of X from point to point average out completely over the involved volume or area, then $\lambda = 0$, and the only variance which is left is $\sigma_{stat}^2 + \sigma_{applmodel}^2$. An example of spatial averaging over a continuous soil volume is given in 2.7.2.

Solutions for the variance reduction factor λ for a number of geometries can be found in Ronold (1990).

2.7.1.2 Spatial averaging may also be relevant for problems and mechanisms which involve averages or sums over a number of discrete elements in space rather than averages or integrals over continuous soil volumes.

Guidance note:

An example of a problem which involves spatial averaging of soil properties over a number of discrete elements in space is the lateral capacity of the piled foundation of an offshore jacket platform. This capacity comes about as N times the spatial average of the lateral capacities of the N jacket piles that constitute the foundation. The spatial average of the lateral pile capacities will depend on the autocorrelation function for the soil strength that produces the lateral pile capacities, it will depend on the geometrical configuration of the N pile positions and it will depend on the separations between these positions in the horizontal plane.

---e-n-d---of---G-u-i-d-a-n-c-e---n-o-t-e---

2.7.2 Example of spatial averaging

2.7.2.1 An example of an application, where the spatial average of a soil variable is involved rather than the variable itself, is the axial capacity of a friction pile in clay. Several models exist for prediction of this capacity, including the α method. By this model, the spatial average of the undrained shear strength s_u over the length L of the pile is involved as part of the expression for the axial capacity Q_{axial}, viz.

$$\begin{aligned}Q_{axial} &= \alpha \pi D \cdot \int_{z=0}^{L} s_u(z)dz \\ &= \alpha \pi D \cdot \left(\int_{z=0}^{L} \mu_{s_u}(z)dz + \int_{z=0}^{L} \varepsilon(z)dz \right)\end{aligned}$$

in which α is a dimensionless factor and D is the pile diameter. μ_{Su} denotes the mean value of s_u and ε is a zero-mean term which represents the variability of s_u about μ_{Su}.

2.7.2.2 The reduced variance associated with the spatially averaged undrained shear strength s_u used for prediction of Q_{axial} is $\sigma_{Su,spatial}^2 = \lambda \cdot \sigma_{Su}^2$ and comes about as the variance of the spatial average term

$$\frac{1}{L}\int_{z=0}^{L} \varepsilon(z)dz$$

Guidance note:

When σ_{Su} is constant with depth and when the vertical correlation structure of the strength field can be represented by the quadratic exponential decay model, the variance reduction factor can be expressed as

$$\lambda = \frac{2}{L^2}\left\{LR\sqrt{\pi}\left[\Phi\left(\sqrt{2}\,\frac{L}{R}\right)-\frac{1}{2}\right]+\frac{1}{2}R^2\left[\exp\left(-\frac{L^2}{R^2}\right)-1\right]\right\}$$

in which L is the length of the pile, R is the vertical correlation length of the undrained shear strength s_u, and Φ is the standard normal cumulative distribution function. See Figure 2-9.

---e-n-d---of---G-u-i-d-a-n-c-e---n-o-t-e---

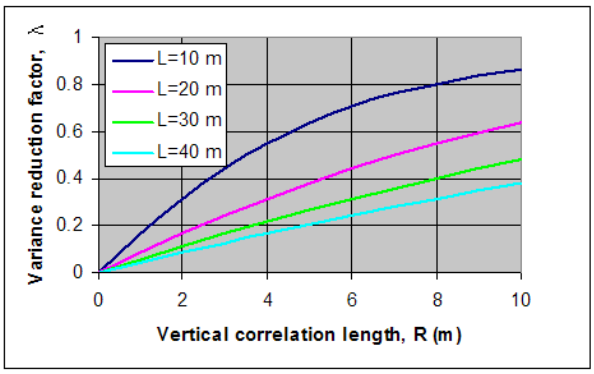

Figure 2-9
Variance reduction factor vs. pile length L and vertical correlation length R

2.7.2.3 Statistical uncertainty associated with the estimation of the mean μ_{Su} and the standard deviation σ_{Su} of s_u comes in addition when the total uncertainty associated with the prediction of Q_{axial} is sought after. Through the expression for Q_{axial}, the mean μ_{Su} of s_u appears to govern the mean value of Q_{axial},

$$E[Q_{axial}] = \alpha\pi D \cdot \int_{z=0}^{L} \mu_{s_u}(z)dz$$

2.7.2.4 The estimation of the mean μ_{Su} and the standard deviation σ_{Su} of s_u and the determination of the involved estimation uncertainties depend on where the friction pile is located relative to available soil borings and CPTs. Three different situations are considered:

1) The position of the pile relative to soil borings and CPTs is unknown, and μ_{Su} and σ_{Su} can be estimated from all soil data available from the soil deposit in which the pile is to be installed.
2) The position of the pile relative to soil borings and CPTs is known and coincides with one of these borings or CPTs. μ_{Su} and σ_{Su} can be estimated from the soil data obtained from the boring or the CPT which is located in the pile position.
3) The position of the pile relative to soil borings and CPTs is known, but does not coincide with any of these borings or CPTs. μ_{Su} and σ_{Su} can be estimated by spatial interpolation between soil data from adjacent borings and CPTs, e.g. by means of a kriging technique. See 2.7.3.

A variance associated with the variability in predictions by the model for Q_{axial} from case to case comes in addition to the reduced variance $\sigma_{Su,spatial}^2$ and the statistical variance σ_{stat}^2.

2.7.3 Spatial estimation – kriging

2.7.3.1 Let X denote a soil variable, such as a strength, which forms a random field over a soil deposit. X has been measured in a number of positions within the soil volume, but is in principle unknown in all other positions. It is assumed that the autocorrelation function for X can be established based on the available observations of X.

2.7.3.2 Kriging is an estimation technique which, based on the measured values of X in conjunction with the autocorrelation function, allows for estimation of the value of X in positions where X has not been measured. For each specified position of estimation, the kriging technique provides both an estimate of X and a standard error in the estimate, accounting for spatial correlations.

2.7.3.3 Assume that X is observed in n positions, $\mathbf{s_i}$, $i = 1,...n$, within the soil volume V and let x_i denote the observed value of X in the position $\mathbf{s_i}$. Assume also that the autocorrelation function $\rho(r)$ for X is available such that the covariance function $C(r) = \sigma^2\rho(r)$ can be established. Here, r denotes the lag between two positions **s** and **s'**, $r = |\mathbf{s}-\mathbf{s'}|$. The point value of X in the unsampled position **s** is to be estimated.

2.7.3.4 The simple kriging estimate of the single point realisation X(**s**) in the position **s** is

$$\hat{x}(\mathbf{s}) = \sum_{i=1}^{n} \lambda_i x_i$$

where λ_i, $i = 1,...n$, are weighting factors, which can be determined as the solution of the n equations

$$\sum_{j=1}^{n} \lambda_j C(|\mathbf{s}_i - \mathbf{s}_j|) = C(|\mathbf{s} - \mathbf{s}_i|) \; ; \; i = 1,...n$$

2.7.3.5 The ordinary kriging estimate of the single point realisation x(s) of X in the position **s** is

$$\hat{x}(\mathbf{s}) = \sum_{i=1}^{n} \lambda_i x_i$$

where λ_i, $i = 1,...n$, are weighting factors, which can be determined as the solution of the n equations

$$\sum_{j=1}^{n} \lambda_j C(|\mathbf{s}_i - \mathbf{s}_j|) + k = C(|\mathbf{s} - \mathbf{s}_i|) \; ; \; i = 1,...n$$

with the unbiasedness constraint

$$\sum_{i=1}^{n} \lambda_i = 1$$

2.7.3.6 The point estimate $\hat{x}(\mathbf{s})$ is encumbered with uncertainty because it is based on a limited number of observations of X in a limited number of positions located some distance away from the point of estimation. Regardless of whether the estimate $\hat{x}(\mathbf{s})$ is established by simple kriging or by ordinary kriging, the standard error of the estimate is

$$se(\hat{x}(\mathbf{s})) = \sqrt{C(0) + \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j C(|\mathbf{s}_i - \mathbf{s}_j|) - 2\sum_{i=1}^{n} \lambda_i C(|\mathbf{s} - \mathbf{s}_i|)}$$

$$= \sigma \sqrt{\rho(0) + \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j \rho(|\mathbf{s}_i - \mathbf{s}_j|) - 2\sum_{i=1}^{n} \lambda_i \rho(|\mathbf{s} - \mathbf{s}_i|)}$$

2.7.3.7 The expressions for the kriging estimates and their standard error are based on assumptions of isotropic autocorrelation and covariance functions. When these functions are anisotropic, e.g. when 3-D problems are encountered and different correlation lengths apply vertically and horizontally, the expressions will become modified as the arguments of these functions, which here appear as a scalar measure of lag, become a vector of lag in space.

2.7.3.8 Kriging can be generalised to apply to prediction of the spatial average of the field X over the soil volume V rather than just estimating the point value in a particular position s. Kriging, which is here presented for the case of a stationary field X with constant mean and variance, can also be generalised to kriging with a trend for the case that the field X has a mean that has a variation with location.

2.7.3.9 Kriging can be applied to planning of soil investigations, e.g. determination of the most optimal position for an additional soil boring when an extra soil boring is to be carried out to supplement an existing array of soil borings in a soil deposit. In this context the most optimal position is the position that will provide as much new knowledge and as little redundant knowledge about the random field X as possible.

Guidance note:

When an additional soil boring is to be carried out, all relevant information, including knowledge of the local geology and results from geophysical surveys, will have to be brought into the picture in addition to the results from the kriging analysis, before the final position of the additional boring can be decided.

---e-n-d---of---G-u-i-d-a-n-c-e---n-o-t-e---

2.7.3.10 Kriging can also be applied to mapping of soil layer boundaries.

2.7.4 Calculation example of kriging

2.7.4.1 The undrained shear strength s_u in a soil layer is measured in 3 positions which form 3 of the corners of a horizontal square whose side length is r = 15 m. The undrained shear strength in the fourth corner of the square is unknown but can be estimated by means of kriging on the basis of the 3 observed values. The autocorrelation function for the strength field is assumed to be $\rho(r) = \exp(-(r/R)^2)$ with R = 30 m. Possible measurement uncertainties and uncertainties associated with sample disturbance are disregarded.

Guidance note:

Kriging can be applied regardless of the spatial geometry of the data and does thus not presume a particular configuration of the data points such as the four corners of a square used in this example. What matters is the distances

between the data points themselves and between the data points and the point of prediction, not the absolute positions of these points.

---e-n-d---of---G-u-i-d-a-n-c-e---n-o-t-e---

2.7.4.2 The measurement positions and the strength measurements are

Position no.	*Position coordinates* **s** *= (x,y)*		s_u *(kPa)*
	x (m)	*y (m)*	
1	0	15	80
2	0	0	85
3	15	0	75

2.7.4.3 Three values of the correlation coefficient ρ are required in the kriging equations, viz. $\rho(0) = 1.0$, $\rho(15) = 0.779$, and $\rho(21.2) = 0.607$.

2.7.4.4 Solution of the set of 3 kriging equations leads to the weights $\lambda_1 = \lambda_3 = 0.779$ and $\lambda_2 = -0.607$, and the simple kriging estimate of s_u in the unsampled position (x,y) = (15,15) becomes

$$\hat{s}_u(15,15) = \sum_{i=1}^{3} \lambda_i s_{u,i} = 69.2 \text{ kPa}$$

The ratio of the estimation variance $se(\hat{s}_u)^2$ to the variance Var[s_u] of the undrained shear strength field s_u is

$$\begin{aligned}\frac{se(\hat{s}_u(15,15))^2}{Var[s_u]} &= 1 + \sum_{i=1}^{3}\sum_{j=1}^{3} \lambda_i \lambda_j \rho(|\mathbf{s}_i, \mathbf{s}_j|) \\ &\quad - 2\sum_{i=1}^{3} \lambda_i \rho(|\mathbf{s}_i, (15,15)|) \\ &= 0.155\end{aligned}$$

2.7.4.5 It appears that the simple kriging applied to point estimation of the s_u field in the position **s** = (15,15) implies an updating of the distribution of s_u in this position from a prior distribution to a posterior distribution by determination of the conditional distribution from observations of s_u in neighbouring locations. In particular it appears that for the present example the updating implies a reduction in the variance of the point estimate of s_u by 84%.

The prior information consist of the mean estimate $\hat{\mu} = 80$ kPa and the standard deviation estimate $\hat{\sigma} = 5$ kPa, both based on the three available observations of s_u given in 2.7.4.2. These are best estimates for the undrained shear strength and the standard deviation of the undrained shear strength, respectively, in any point within the undrained shear strength field regardless of location.

When the additional information about geometry and autocorrelation is included and accounted for through the kriging, the estimate of the undrained shear strength in the considered particular point of prediction becomes updated to 69.2 kPa as outlined in 2.7.4.3 and 2.7.4.4, and the standard deviation of this estimate becomes

$$se(\hat{s}_u(15,15)) = \sqrt{0.155} \cdot 5 = 2.0 \text{ kPa}$$

which is a significant reduction compared to the prior estimate of 5 kPa.

2.7.4.6 It is of interest to compare the simple kriging result presented in 2.7.4.4 with the result of estimation by ordinary kriging. Ordinary kriging implies that an unbiasedness constraint comes into force and solution of the 3 kriging equations of the example leads to altered weights $\lambda_1 = \lambda_3 = 0.875$ and $\lambda_2 = -0.751$. The ordinary kriging estimate of s_u in the unsampled position (x,y) = (15,15) hence becomes

$$\hat{s}_{u,OK}(15,15) = \sum_{i=1}^{3} \lambda_i s_{u,i} = 71.8 \text{ kPa}$$

The difference between the simple kriging estimate 69.2 kPa and the ordinary kriging estimate 71.8 kPa illustrates that simple kriging can be biased. In the present case the bias could well be associated with the fact that the estimation is based on as little as 3 observed field values and that it is an extrapolation rather than an interpolation. The difference between the two estimates demonstrates the added value provided by the unbiasedness constraint of the ordinary kriging.

2.8 Full probability distribution representation of soil properties with examples of application

2.8.1 General

2.8.1.1 A number of applications where uncertainties in soil properties are involved require full probability distribution representation of the soil properties. Such applications include structural reliability analyses, by which the probability of failure for a foundation can be calculated when all governing load and soil variables and their probability distributions are modelled. Full probability distribution representation of soil properties usually consists of a parametric distribution model, i.e. a generic distribution type, such as the normal distribution, and as many distribution parameters as required, e.g. mean value and standard deviation.

2.8.1.2 The most commonly encountered generic distribution types for soil properties are given in 2.3.3 and 2.3.4.

2.8.2 Application to geotechnical reliability analyses

2.8.2.1 In geotechnical design, the reliability of a foundation can be evaluated with respect to relevant failure modes. For each such failure mode, the foundation and its engineering behaviour can be described by a set of stochastic basic variables grouped into one vector **X**, including, e.g., the soil strength, the soil stiffness, the foundation geometry, and the loading. All of these variables are stochastic in the sense that, owing to natural variability and possible other uncertainties, they may take on random realizations according to some probability distribution.

2.8.2.2 For the considered failure mode, the possible realisations of **X** can be separated in two sets; namely the set for which the foundation will be safe, and the set for which it will fail. The surface between the safe set and the failure set in the space of basic variables is denoted the limit state surface, and the reliability problem is conveniently described by a so-called limit state function $g(\mathbf{X})$ which is defined such that

$$g(\mathbf{X})\begin{cases} > 0 \text{ for } \mathbf{X} \text{ in safe set} \\ = 0 \text{ for } \mathbf{X} \text{ on limit state surface} \\ < 0 \text{ for } \mathbf{X} \text{ in failure set} \end{cases}$$

The limit state function is usually based on some mathematical engineering model for the considered limit state, based on the underlying physics, and expressed in terms of the governing load and resistance variables.

2.8.2.3 The failure probability is the probability content in the failure set

$$P_F = P\left[g(\mathbf{X}) \le 0\right] = \int_{g(\mathbf{X})\le 0} f_{\mathbf{X}}(\mathbf{x})d\mathbf{x}$$

where $f_{\mathbf{X}}(\mathbf{x})$ is the joint probability density function for **X** and represents the uncertainty and natural variability in the governing variables **X**.

2.8.2.4 The complementary probability $P_S = 1 - P_F$ is referred to as the reliability and is sometimes also denoted the probability of survival.

2.8.2.5 The reliability may be expressed in terms of the reliability index,

$\beta = -\Phi^{-1}(P_F)$,

where Φ is the standardized cumulative normal distribution function.

2.8.2.6 The failure probability, the reliability, and the reliability index are all suitable measures of the foundation safety with respect to the considered failure mode.

2.8.2.7 For the simple example that **X** consists of two variables, the load L and the resistance R, and the limit state function can be specified as $g(\mathbf{X}) = R - L$, the failure probability becomes a simple convolution integral

$$\begin{aligned} P_F &= P[R - L < 0] \\ &= \int_{R-L<0} f_R(r) f_L(l) dr dl \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{l} f_R(r) f_L(l) dr dl \\ &= \int_{-\infty}^{\infty} F_R(l) f_L(l) dl \end{aligned}$$

where f_R and f_L are the probability density functions of R and L, respectively, and $f_R(r) = dF_R(r)/dr$, where F_R is the cumulative distribution function of R. Reference is made to Figure 2-10.

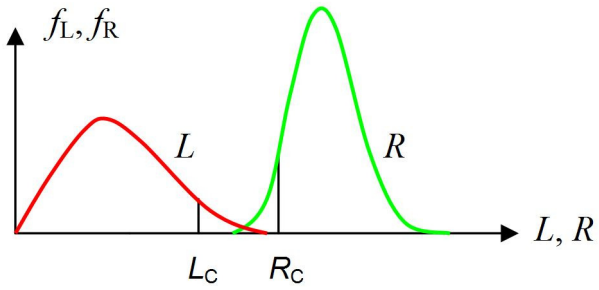

Figure 2-10
Probability density functions and characteristic values for load *L* and resistance *R*

2.8.2.8 The reliability index β can be solved in a structural reliability analysis by means of a reliability method which can be any amongst several available methods, including numerical integration, analytical first- and second-order reliability methods (FORM/SORM), and simulation methods. Simulation methods include the well-known Monte Carlo simulation. Reference is made to Madsen et al. (1986). Some of these methods are approximate methods, which will lead to approximate results for the reliability index. Numerical integration is usually only feasible when **X** consists of very few stochastic variables such as in the example in 2.8.2.7. Analytical first- and second-order solutions to the failure probability are often sufficiently accurate, and they are advantageous to simulation results when the failure probabilities are small.

Guidance note:

When the reliability index as defined in 2.8.2.5 is estimated by a first-order reliability method (FORM), the estimate that results from the estimation is a first-order second-moment reliability index according to Hasofer-Lind's definition, see Madsen et al. (1986).

It is important to distinguish the reliability index as defined in 2.8.2.5 from the so-called simple reliability index. The simple reliability index is defined as the ratio between the mean value and the standard deviation of the limit state function. The simple reliability index is not invariant with the choice of limit state function, it will in general be different from the reliability index as defined in 2.8.2.5, and it is not recommended for use.

---e-n-d---of---G-u-i-d-a-n-c-e---n-o-t-e---

2.8.2.9 A useful byproduct of a structural reliability analysis by these methods is the so-called design point **x***. This is a point on the limit state surface and is the most likely realisation of the stochastic variables **X** at failure. Another useful byproduct is the uncertainty importance factors for **X**. The uncertainty importance factor for the *i*th stochastic variable X_i in **X** roughly gives the fraction of the total uncertainty which is caused by the uncertainty in X_i.

2.8.2.10 Once a target failure probability is specified and once characteristic values of governing load and resistance variables for a foundation are defined, e.g. as quantiles in the respective probability distributions of these variables, structural reliability analyses can be applied to calibrate the necessary requirements to the load and material factors that are part of the design rule for the foundation.

2.9 Design soil parameters

2.9.1 General

2.9.1.1 Soil parameters for use in design are usually given in one or more of the following ways: Characteristic value of the soil property in question, best estimate of the soil property, and upper and lower bounds of the soil property. Soil parameters given in these ways for use in design should be established according to the methodologies presented in 2.4.1 through 2.4.4 and 2.5.1 through 2.5.4.

2.9.1.2 So-called design profiles of characteristic values for soil properties are often established with reference to a specific application, i.e. with reference to a specific type and size of foundation, and will not necessarily be valid for other applications.

Guidance note:

The reason why design profiles established for specific applications will not necessarily be valid for other applications is that the definition of characteristic value may be different for different applications. This depends on the degree to which averaging of soil properties can be assumed to take place. For long friction piles, the degree of averaging may be large, whereas the degree of averaging may be a lot smaller for end-bearing piles and smaller shallow foundations.

The need for close contact between the soil investigation contractor, who shall deliver a soil investigation report and who is most familiar with the assumptions made in establishing the characteristic values and with the uncertainties present, and the geotechnical designer who shall use this report for foundation design, is evident, but is in practice – unfortunately – often neglected.

Caution should be exercised when using prescribed design profiles for soil properties. Some standards refer to both characteristic profiles and design profiles and distinguish between them by means of the material factor. Other

standards refer only to one of the profiles, and for some standards, which refer to the design profile only, the design profile is not necessarily factored, i.e. the design profile may consist of unfactored characteristic values rather than characteristic values divided by the material factor. Hence, prior to using a prescribed design profile, one should always make sure to check whether the soil property values associated with the profile have been divided by a material factor or not.

---e-n-d---of---G-u-i-d-a-n-c-e---n-o-t-e---

2.9.2 Characteristic value

2.9.2.1 Characteristic values are used to represent soil properties such as soil shear strength. The definition of a characteristic value is much governed by the design problem in question and by the geometry. A definition as the mean value of the property typically applies in cases where local fluctuations of the soil property can be assumed to average out over large soil volumes, such as in the case of the axial capacity of long friction piles. A definition as a lower-tail quantile in the distribution of the property typically applies in cases where a local soil strength is governing, such as in the case of the tip resistance of an end-bearing pile.

Guidance note:

In foundation design, the characteristic value of a soil property is always used together with a partial safety factor. The characteristic value and the partial safety factor form a pair. This implies that a partial safety factor, which is meant for use together with a characteristic value defined as the mean value of a soil property, cannot be used together with a characteristic value defined as a low percentile of that soil property.

---e-n-d---of---G-u-i-d-a-n-c-e---n-o-t-e---

2.9.3 Best estimate

2.9.3.1 Best estimates of soil properties are central unbiased estimates with lowest possible standard errors. Best estimates are normally used for assessment of serviceability limit states, i.e. whenever problems are encountered for which predictions of the expected foundation behaviour are of interest. An example of this is the prediction of the expected consolidation settlements for which best estimate values of governing soil deformation parameters are required, e.g. inferred from oedometer tests. This applies to prediction of absolute consolidation settlements as well as to prediction of differential settlements. For prediction of best estimates of differential settlements, the horizontal correlation lengths of the governing soil deformation parameters are of importance.

2.9.4 Lower and upper bounds

2.9.4.1 Soil reports often specify lower and upper bound values for use as characteristic values in design. Lower bounds are usually meant for design against the ultimate limit state where low strengths are unfavourable. Upper bounds are usually meant for considerations where large strengths are unfavourable such as evaluations of skirt penetration resistance for installation of skirted foundations and pile driving resistance for installation ability of piles by means of particular pile driving hammers. Upper bounds are also of relevance for evaluation of pile driving resistance for assessment of fatigue in the pile material. The designer should always make sure that presented lower and upper bounds for soil properties meet the requirements to characteristic values set forth in the codes and standards which are used in design.

For design of structures subjected to cyclic loading or influenced by dynamic behaviour, it may be necessary to perform sensitivity studies for both lower bound values and upper bound values for relevant soil properties for the supporting foundation soils.

2.10 References and literature

2.10.1 List of references

2.10.1.1 Ang. A., and W.H. Tang, *Probability Concepts in Engineering Planning and Design, Vol. I, Basic Principles*, John Wiley and Sons, New York, N.Y., 1975.

2.10.1.2 Ang. A., and W.H. Tang, *Probability Concepts in Engineering Planning and Design, Vol. II, Decision, Risk, and Reliability*, John Wiley and Sons, New York, N.Y., 1984.

2.10.1.3 Madsen, H.O., N.C. Lind, and S. Krenk, *Methods of Structural Safety*, Prentice Hall Inc., Englewood Cliffs, N.J., 1986.

2.10.1.4 Neter, J., W. Wasserman, and M.H. Kutner, *Applied Linear Statistical Models*, Richard D. Irwin, Inc., Homewood, Ill., 1990.

2.10.1.5 Ronold, K.O., "Random Field Modeling of Foundation Failure Modes," *Journal of Geotechnical Engineering*, ASCE, Vol. 116, No. 4, pp. 554-570, 1990.

2.10.1.6 Ronold, K.O., and A. Echtermeyer, "Estimation of fatigue curves for design of composite laminates," *Composites, Part A: Applied Science and Manufacturing*, Elsevier, Vol. 27A, No. 6, pp. 485-491, 1996.

2.10.1.7 Snedecor, G.W., and W.G. Cochran, *Statistical Methods*, Iowa State University Press, Ames, Iowa, 1989.

2.10.2 Other relevant literature

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

2.10.2.2 Cressie, N., *Statistics for Spatial Data*, John Wiley and Sons, New York, N.Y., 1991.

2.10.2.3 Isaaks, E.H., and R.M. Shrivastava, *Applied Geostatistics*, Oxford University Press, New York, N.Y., 1989.

2.10.2.4 Journel, A.G., and C.J. Huijbregts, *Mining Geostatistics*, Academic Press, London, England, 1978.

2.10.2.5 Vanmarcke, E., *Random Fields, Analysis and Synthesis*, The MIT Press, Cambridge, Mass., 1983.

APPENDIX A
TABLES OF PROBABILITY DISTRIBUTIONS

A.1 Table of the Student's *t* distribution

Table A-1 gives the $1-\alpha$ quantiles t_α in the Student's *t* distribution with *n* degrees of freedom, i.e. $P[T < t_\alpha] = 1-\alpha$

Table A-1 Quantiles in the Student's *t* distribution

n	$1-\alpha$				
	0.75	0.90	0.95	0.98	0.99
1	1.000	3.078	6.314	15.895	31.821
2	0.816	1.886	2.920	4.849	6.965
3	0.765	1.638	2.353	3.482	4.541
4	0.741	1.533	2.132	2.999	3.747
5	0.727	1.476	2.015	2.757	3.365
6	0.718	1.440	1.943	2.612	3.143
7	0.711	1.415	1.895	2.517	2.998
8	0.706	1.397	1.860	2.449	2.896
9	0.703	1.383	1.833	2.398	2.821
10	0.700	1.372	1.812	2.359	2.764
11	0.697	1.363	1.796	2.328	2.718
12	0.695	1.356	1.782	2.303	2.681
13	0.694	1.350	1.771	2.282	2.650
14	0.692	1.345	1.761	2.264	2.624
15	0.691	1.341	1.753	2.249	2.602
16	0.690	1.337	1.746	2.235	2.583
17	0.689	1.333	1.740	2.224	2.567
18	0.688	1.330	1.734	2.214	2.552
19	0.688	1.328	1.729	2.205	2.539
20	0.687	1.325	1.725	2.197	2.528
21	0.686	1.323	1.721	2.189	2.518
22	0.686	1.321	1.717	2.183	2.508
23	0.685	1.319	1.714	2.177	2.500
24	0.685	1.318	1.711	2.172	2.492
25	0.684	1.316	1.708	2.167	2.485
26	0.684	1.315	1.706	2.162	2.479
27	0.684	1.314	1.703	2.158	2.473
28	0.683	1.313	1.701	2.154	2.467
29	0.683	1.311	1.699	2.150	2.462
30	0.683	1.310	1.697	2.147	2.457
35	0.682	1.306	1.690	2.133	2.438
40	0.681	1.303	1.684	2.123	2.423
45	0.680	1.301	1.679	2.115	2.412
50	0.679	1.299	1.676	2.109	2.403
60	0.679	1.296	1.671	2.099	2.390
70	0.678	1.294	1.667	2.093	2.381
80	0.678	1.292	1.664	2.088	2.374
100	0.677	1.290	1.660	2.081	2.364
∞	0.674	1.282	1.645	2.054	2.326

A.2 Table of the normal distribution

Table A-2 gives the value of the cumulative normal distribution function $\Phi(x)$ for given quantile x.

Table A-2 The cumulative normal distribution function

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000