PACIFIC EARTHQUAKE ENGINEERING
RESEARCH CENTER

Estimation of Uncertainty in Geotechnical Properties for Performance-Based Earthquake Engineering

Allen L. Jones

Steven L. Kramer

and

Pedro Arduino
University of Washington

PEER 2002/16
DEC. 2002

Estimation of Uncertainty in Geotechnical Properties for Performanced-Based Earthquake Engineering

Allen L. Jones
Department of Civil and Environmental Engineering
University of Washington

Steven L. Kramer
Department of Civil and Environmental Engineering
University of Washington

Pedro Arduino
Department of Civil and Environmental Engineering
University of Washington

This work was supported in part by the Pacific Earthquake Engineering Research Center through the Earthquake Engineering Research Centers Program of the National Science Foundation under Award number EEC-9701568

PEER Report 2002/16
Pacific Earthquake Engineering Research Center
College of Engineering
University of California, Berkeley
December 2002

ABSTRACT

Prediction of the performance of structures during earthquakes requires accurate modeling of the geotechnical components of soil-structure systems. Geotechnical performance is strongly dependent on the properties of the soil beneath and adjacent to the structure of interest. These soil properties can be described by using deterministic and probabilistic models. Deterministic models typically use a simple discrete descriptor for the parameter of interest. Probabilistic models account for uncertainty in soil properties by describing those properties by using discrete statistical descriptors or probability distribution functions.

This report describes sources and types of uncertainty in geotechnical engineering practice, and introduces the basic concepts and terminology of the theory of probability. Statistical parameters and the probability distributions most commonly used to describe geotechnical parameters are reviewed. The report then presents tabulated statistical parameters for soil properties that have been reported in the literature; both laboratory and field-measured parameters describing moisture-density, plasticity, strength, and compressibility characteristics are presented. The theory of regionalized variables, including concepts of autocorrelation, variograms, and stationarity are presented, along with tabulated values of parameters describing spatial variability that have been reported in the literature. Finally, procedures and tools for estimation and simulation of spatially variable soil properties are presented.

ACKNOWLEDGMENTS

This work was supported in part by the Pacific Earthquake Engineering Research Center through the Earthquake Engineering Research Centers Program of the National Science Foundation under Award number EEC-9701568. Their financial support is gratefully acknowledged.

The information provided by Prof. Kenneth H. Stokoe and Dr. M.B. Darendian, of the University of Texas is also gratefully acknowledged.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURE

Abbreviations Used Frequently:

CDF	Cumulative distribution function
COV	Coefficient of variation
cov(X,Y)	Covariance between *X* and *Y*
MGF	Moment-generating function
MCM	Monte Carlo Methods
pdf	Probability density function

Non-Greek Notation Used Frequently:

r_h	Autocorrelation function
C(h)	Autocovariance function
$F_X(x)$	Cumulative density function
X_i, x_i	Discrete data of a random or regionalized variable
E(x)	Expected Value or mean
Z(X)	Random or regionalized variable function
h	Lag or separation distance
m(s)	Moment-generating function
a	Nugget parameter
P(x)	Probability
$f_X(x)$	Probability density function
X, x	Random or regionalized variable
h_r	Range of influence
V(x)	Variance

Greek Notation Used Frequently:

Ω	Outcome set
ϑ	Events
$\mathbb{R}$	The set of real numbers
μ	Expected value or mean

σ	Variance
μ_k	k^{th} moment of X
c_k	k^{th} central moment of X
ρ	correlation coefficient
$\gamma(h)$	Variogram function

1 Introduction

Prediction of the performance of structures during earthquakes requires accurate modeling of the geotechnical components of soil-structure systems. Geotechnical performance is strongly dependent on the properties of the soil beneath and adjacent to the structure of interest. These soil properties can be described using deterministic and/or probabilistic models. Deterministic models typically use a single discrete descriptor for the parameter of interest. Probabilistic models describe parameters by using discrete statistical descriptors or probability distribution (density) functions. The spatial variation of the properties can be described by using stochastic interpolation methods.

This overview provides a basis for describing the spatial uncertainty of geotechnical soil properties. Sources and types of uncertainty are first presented, followed by a discussion of the probabilistic treatment of geotechnical data using geostatistics. Finally, generation of realizations of random functions using simulation is discussed. The latter two topics are the primary focus of this overview.

The spatial distribution of geotechnical properties in natural soil deposits is difficult to predict deterministically. Limited sampling, especially in subsurface drilling, further complicates prediction of soil properties. Prediction of the spatial occurrence of soil properties in either an optimal best estimate or within a probabilistic framework is necessary for effective numerical modeling of soils with heterogeneous properties.

Applied geostatistical estimation and simulation techniques can be used to model spatial variability from limited sample sets (or from known distributions of data). While traditional statistics generally assumes independence between samples, geostatistics takes advantage of the fact that samples located in proximity to one another are often more similar than those obtained at large separation distances. Geostatistics provides a means of quantifying this spatial correlation in soil properties and then of using that information for both estimation and stochastic simulation.

Geostatistical estimation refers to techniques that provide the best linear unbiased estimators of unknown properties. When the structural information of the data is known, then the estimator can be defined. This description requires more than the simple first-order moments of the random variable of interest. Geostatistical simulation is a spatial Monte Carlo process where a random "draw" from a local cumulative distribution function simulates a value of a property at a given location. The simulation process is run multiple times to produce a series of realizations all of which correspond to the observed data at the sample locations, the univariate distribution, and the spatial correlation of the observed data. In essence, each realization is a probable representation of the underlying reality given the available data. These multiple realizations can be used as input to a transfer function (e.g., multiple realizations of shear strength as input to slope stability analyses), or processed to provide a map of the probability of a given situation being true (e.g., a plot of the probability that a given factor of safety will be exceeded). The results of the transfer function can often be evaluated in terms of economic loss and/or risk.

This report is organized into several distinct sections. The next section discusses the sources and types of uncertainty. The following section discusses the theory of probability and is followed by a qualitative summary of soil property variability. Soil property variability is followed by a discussion of the theory of regionalized variables and a quantitative presentation of values for soil property spatial dependence. The section following discusses estimation and simulation, and is followed by a summary of available tools and examples.

2 Sources and Types of Uncertainty

"Certainty" refers to situations in which the outcome of an event or the value of a parameter is known with unit probability. Conversely, uncertainty occurs when a collection of values associated with respective uncertain "states of nature" occur with strictly non-negative probabilities for at least two different possible values; the simplest examples are tossing a coin or rolling a die. Uncertainty analysis is an emerging approach that uses estimation and simulation techniques to consider the variability of available data and to estimate the frequency with which values of interest are likely to be exceeded. While it has not yet been widely applied in geotechnical engineering practice, this approach offers insight into existing data for heterogeneous geotechnical systems.

Uncertainty pervades many aspects of geotechnical earthquake engineering, particularly in the characterization of soil properties. In general, some of this uncertainty may be due to the difficulty in making accurate measurements and some may be due to uncertainty in the models, equations, and understanding of the systems involved. Additional uncertainty can result from the spatial variability of the system.

Uncertainty in geotechnical soil properties can be formally grouped into aleatory and epistemic uncertainty (Lacasse et al., 1996). Aleatory uncertainty represents the natural randomness of a property and, as such, is a function of the spatial variability of the property. Recognizing spatial variability is important because it can help distinguish the distances over which it occurs compared to the scale of the data of interest (Whitman, 1996). Epistemic uncertainty results from a lack of information and shortcomings in measurement and/or calculation. Epistemic uncertainty includes the systematic error resulting from factors such as the method of property measurement, the quantity of available data, and modeling errors. Figure 2.1 illustrates the types of uncertainty in geotechnical soil properties. Human error would be considered a third source of uncertainty, however it is not considered in this overview because it is difficult to isolate and its effects on probability are usually included in compilations of statistics on aleatory uncertainty.

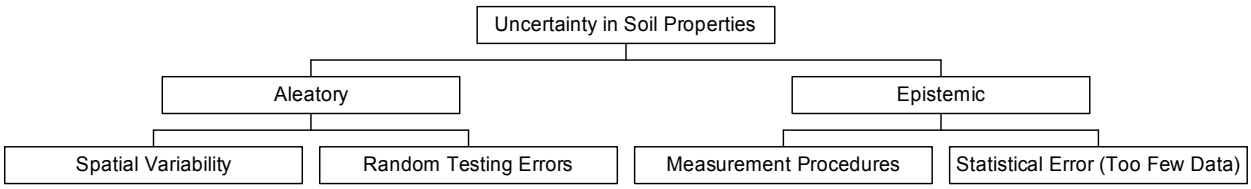


Fig. 2.1 Sources of uncertainty in geotechnical soil properties (adapted from Whitman, 1996)

As an example, consider SPT sampling in a deposit of loose sand. Sources of aleatory uncertainty in the measured SPT resistance would include the natural variability of the soil deposit and random testing errors, such as that caused by a single soil defect (e.g., an erratic boulder). Sources of epistemic uncertainty could include non-standard equipment (such as the sampler size, deformed samplers or rods, rod length, hammer drop system, hammer weight, etc., not conforming to the SPT standard), and insufficient data to form reasonable statistics, such as one boring over a large site. It is important to note that epistemic uncertainty can usually be reduced by acquisition of additional data or improvements in measurement procedures. Aleatory uncertainty, on the other hand, is inherent to the variable and cannot be reduced by additional information.

3 The Theory of Probability

A basic understanding of some fundamental aspects of probability is required to quantify the uncertainty in geotechnical soil properties. These concepts include probability space, distributions, and moments. A brief overview of the basic concepts and terminology of probability are presented in the following sections; more detailed descriptions of this information can be found in numerous introductory textbooks on probability such as Benjamin and Cornell (1970), Ott (1984), Kelly (1994), and Mendenhall and Beaver (1991).

3.1 PROBABILITY SPACE

The notion of *probability space* is fundamental to probability theory. Probability space consists of:

(1) An *outcome set*, Ω, that contains all elements of all sets of data under consideration;

(2) A collection of *events*, ϑ, which are subsets of Ω; and

(3) A procedure for computing the probabilities of events called the *probability mass function*, *p(x),* or the *probability density function,* $f_X(x)$. Mass functions and density functions are discussed in Section 3.3.

$\mathtt{R}$ is defined as the set of all real numbers. In $\mathtt{R}$, $\{x: 3 \leq x < 6\}$ is the set of all real x greater or equal to three and less than six. The terms *sample space* and the *outcome set* are synonymous. Note that for sets (events) to be useful, they must have an assignable probability. Two general types of probability space exist:

(1) Finite space (also known as discrete space) is comprised of a known, finite number of possible samples. An example might be the percentage of drill holes at a site that penetrate a peat layer. If N borings are drilled, there are exactly $N+1$ possible values of the percentage of borings that penetrate peat. In a finite space, the probability of an event, ϑ, occurring is equal to or greater than zero, and the sum of all probabilities of each event occurring is 1:

$$p(\vartheta) \geq 0 \text{ for all } \vartheta \in \Omega \quad \text{and} \quad \sum_{\vartheta \in \Omega} p(\vartheta) = 1. \tag{3.1}$$

(2) Infinite space (also known as continuous space) consists of an infinite, unknown number of possible samples taken from a continuous interval. An example in terms of geotechnical soil properties would be the distribution of shear strength in a soil layer across a site. In terms of probability, an absolutely continuous probability space in R consists of the outcome set, Ω, which is an interval, and a function $f_X(x)$ such that:

$$f_X(x) \geq 0 \text{ for all } x \in \Omega \quad \text{and} \quad \int_{\Omega} f_X(x)dx = 1. \tag{3.2}$$

In other words, continuous probability space contains an unknown number of possible samples and the samples are not countable, but are taken from a continuous interval. Most geotechnical phenomena occur in continuous space.

3.2 RANDOM VARIABLES

A *random variable* is a variable that can take on multiple values. The domain of a random variable is the outcome set and its range is the set of possible values. Mathematically, a random variable can be expressed as a real function, $Z(x)$, that associates a real number, x_i, with each element in the outcome set, $x_i \in \Omega$. The real number, x_i, will correspond to every outcome of an experiment, the function $\{Z(x) \leq x_i\}$ is an event for any real number x_i, and the probabilities $P\{Z(x) = +\infty\}$ *and* $P\{Z(x) = -\infty\}$ will be zero for the random variable. The random variable can be visualized as a function that maps a sample space onto a real line as shown on Figure 3.1.

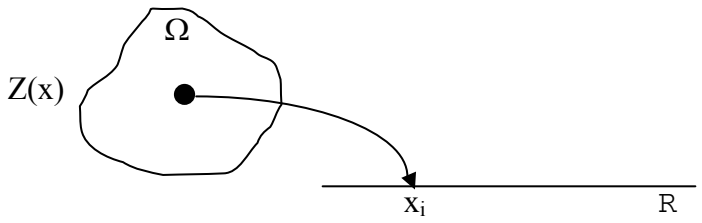


Fig. 3.1 Schematic illustration of random variable — each element of the sample space, Ω, corresponds to a real value, x_i, of the random variable

Two types of random variables exist: discrete and continuous. Discrete random variables take on values from discrete probability space, i.e., they represent quantities, such as SPT resis-

tance or USCS soil type, that can assume only a finite number of different values. Continuous random variables take on values from continuous probability space; common examples include undrained strength, friction angle, permeability, exit gradients at the toe of a levee, and time to occurrence of any event.

3.3 PROBABILITY DISTRIBUTIONS

Probability distributions are recipes for computing the probabilities of all possible events for a particular random variable. These recipes will specify the probabilities over the entire sample space.

3.3.1 Probability Functions

The *probability mass function* (pmf) and the *probability density function* (pdf) are functions that assign a probability to every interval of the outcome set for discrete and continuous random variables, respectively. The pmf and the pdf are denoted $p_X(x)$ and $f_X(x)$, respectively, where X is the random variable itself and x is the value that the random variable can take on. Probability functions have the following properties:

(1) $p_X(x) \geq 0$ (discrete)

$f_X(x) \geq 0$ (continuous)

(2) $\sum_{X \in \Omega} p(x) = 1$ (discrete)

$\int_{-\infty}^{\infty} f_X(x)dx = 1$ (continuous)

(3) $P[I] = \sum_{x \in I} p(x)$ (discrete)

$P[a \leq x \leq b] = \int_a^b f_X(x)dx$ (continuous)

It is important to note that any single value in a continuous probability space has zero probability and that the probabilities of intervals are the same whether they contain their endpoints or not.

3.3.2 The Cumulative Distribution Function

The *cumulative distribution function* (CDF) of a probability distribution is a function that defines the probability of the random variable taking on values less than x, or:

$$F(x) = P(X \le x) \tag{3.3}$$

is commonly referred to as the distribution function, and is denoted $F_X(x)$. It is appropriate to include the descriptor cumulative to remind us that $F_X(x)$ is the *accumulated* probability of all numbers less than or equal to x. As x increases, the probability continues to correspondingly increase. The CDF applies to both discrete and continuous random variables; $F_X(x)$ is a non-negative, monotonically increasing function such that $0 \le F_X(x) \le 1$, $F_X(-\infty) = 0$ and $F_X(+\infty) = 1$. For real a and b, with $a < b$, $P(a \le X \le b) = F_X(b) - F_X(a)$.

Many distributions can be systematically grouped into customary distributions. These distributions are covered later in this chapter.

3.4 MATHEMATICAL EXPECTATION

The predicted value of random variables is generally described using mathematical expectation. The expected value of a random variable, X, is a number computed from the pdf of x that represents the expected long-term average observed value of X. The expected value of a continuous random variable is given by:

$$E(X) = \int_{-\infty}^{\infty} x\, f_X(x)dx\,. \tag{3.4}$$

The expected value is commonly referred to as the *mean* or *average*, and is given the symbol, μ.

3.5 VARIANCE AND MOMENTS

The "dispersion" of the random variable about the mean is described by the *variance*, which is defined as:

$$V(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)dx\,. \tag{3.5}$$

The variance can also be described as the expected square, minus the squared expected value (or as some remember it, mean square minus squared mean), i.e.,

$$V(X) = E(X^2) - [E(X)]^2 . \tag{3.6}$$

It is more common to express dispersion using the *standard deviation*, $\sigma = \sqrt{V(X)}$, so that the units of the descriptor and the random variable are the same. Random variables are often acted upon by functions, and it is useful to be able to relate the expectation of the function to that of the random variable. For example, a linear function of the random variable X will describe a transformed function. The mean and variance of the function exhibit the following important properties:

$$E(aX + b) = aE(X) + b \text{, and} \tag{3.7}$$

$$V(aX + b) = a^2 V(X) . \tag{3.8}$$

Many of the essential characteristics of a pdf can be described by a set of relatively simple scalar quantities. The expected value of the various powers of a random variable, X, $E(X)$, $E(X^2)$, $E(X^3)$, etc.) are called the moments of X. Moments are important in describing regionalized variables discussed later in this text. For any non-negative integer k, the k^{th} moment of X is defined by:

$$\mu_k = E(X^k) . \tag{3.9}$$

The moments of the random variable $X - E(X)$ are also important. They are called the central moments of X, because $E(X)$ is regarded as the "center" for the distribution X. The central moments are also referred to as the moments of X about $E(X)$, and are denoted by c_0, c_1, c_2, and so forth, where:

$$c_k = E(X - \mu_k)^k, \text{ where } \mu_k = E(X). \tag{3.10}$$

Note that the zero-th and first central moments are 1 and 0, respectively. The second central moment is the variance, the third central moment measures skewness, and the fourth central moment measures the steepness of the peak of the pdf near its center (kurtosis). In terms of moments, the mean and variance can be expressed as $E(X) = \mu_1$ and $V(X) = E(X - \mu_1)^2$.

The *coefficient of variation* represents a relative (and dimensionless) measure of dispersion and is expressed as:

$$COV = \frac{\sigma_X}{\mu_X} \times 100\% \,. \tag{3.11}$$

The *COV* has been commonly used to describe the variation of many geotechnical soil properties and insitu test parameters. Note that the mean, standard deviation, and *COV* are interdependent — knowing any two will give the third.

In practice, it becomes convenient to estimate moments of geotechnical soil parameters where little data are available (sparse data) by assuming that the *COV* is similar to previously measured values from other data sets of the same parameter. A summary of *COV* values reported in the literature is presented in the next chapter.

The dispersion of sparse data can also be estimated by using other methods such as the "three-sigma rule." This method for approximating the variance recognizes that 99.73% of all values of a normally distributed parameter are within three standard deviations of the average. According to the three-sigma rule, the standard deviations can be approximated by dividing the range (highest value minus the lowest value) by 6. Duncan (2000) points out that engineers have a tendency to underestimate the range between the lowest and highest conceivable values, therefore, a conscious effort should be made to make the range between the two as high as seemingly possible. Christian et al. (2001) also caution that engineers with less experience in statistics will estimate the lowest and highest conceivable values with significant unconservative bias. The field of order statistics offers an alternative to the three-sigma rule that accounts for the quantity of available data. According to Burlington and May (1978), the standard deviation of a normally distributed random variable can be estimated by dividing the range of measured values by the value *N* shown in Table 3.1.

.

Table 3.1 Number of standard deviations, *N*, in expected sample range as function of number of measurements, *n* (after Burlington and May, 1978)

n	*N*
2	1.128
3	1.693
4	2.059
5	2.326
6	2.534
7	2.704
8	2.847
9	2.970
10	3.079
15	3.472
20	3.735
30	4.090

In the above discussion, it is implicitly assumed that random variables are normally distributed as discussed in Section 3.6.2. If the variables are not normally distributed, Equation 3.11 cannot be used. To solve this problem, the technique of normal tail approximation can be used to transform a non-normal distribution into an equivalent normal distribution. This results in the concept of the equivalent *COV*. The reader is referred to Paleheimo and Hannus (1974) for this procedure.

3.5.1 Moments of Two Random Variables

Joint moments can be used to describe the relationship between two random variables. The first joint moment about the mean is a measure of the interdependence between two random variables, X and Y, and is called the *covariance* of X and Y. The covariance between two random variables is defined as:

$$\operatorname{cov}(X,Y) = E\left[(X-\mu_X)(Y-\mu_Y)\right] = E(XY) - \mu_X\mu_Y . \tag{3.12}$$

If X and Y are independent, $\operatorname{cov}(X,Y) = 0$; however, the converse is not true. A positive covariance means that one random variable increases as the other increases. A negative covariance means that one random variable increases when the other decreases. Therefore, the covariance is said to be a relative measure of the degree of positive or negative correlation between X and Y.

For n random variables, the set of covariances can be expressed in the form of a covariance matrix:

$$\boldsymbol{C} = \begin{bmatrix} c_{11} & c_{12} & c_{13} & \cdots & c_{1n} \\ c_{21} & c_{22} & c_{23} & \cdots & c_{2n} \\ c_{31} & c_{32} & c_{33} & & \\ \vdots & \vdots & & \ddots & \\ c_{n1} & c_{n2} & & & \sigma_{nn} \end{bmatrix} \tag{3.13}$$

where: $c_{ij} = \mathrm{cov}(x_i, x_j)$, and $c_{11} = \sigma_1^2, c_{22} = \sigma_2^2, \ldots, c_{nn} = \sigma_n^2$. To obtain a relative measure of correlation between *X* and *Y*, the covariance is divided by the square root of the product of the variances to yield the *correlation coefficient*:

$$\rho = \frac{\mathrm{cov}(X,Y)}{\sqrt{V(X)V(Y)}}. \tag{3.14}$$

The correlation coefficient will take on values between –1 and 1. In general, the correlation coefficient expresses the relative strength of the association between two parameters. The distribution of one of a group of correlated parameters, given specific values of the other parameters, will be influenced by the degree of correlation. For example, Lumb et al. (Lumb, 1970; Grivas, 1981 and Wolff, 1985) have shown that the shear strength parameters *c* and ϕ are often negatively correlated with correlation coefficient ranging from -0.72 to 0.35. The distribution of *c* for a particular ϕ value and will be different than the distribution of all *c* values.

Consider a boring in which samples have been obtained at eight different depths within the same soil unit (Harr, 1987). Water contents are measured for each of the samples, with the results plotted as shown in tabular form in Figure 3.2.

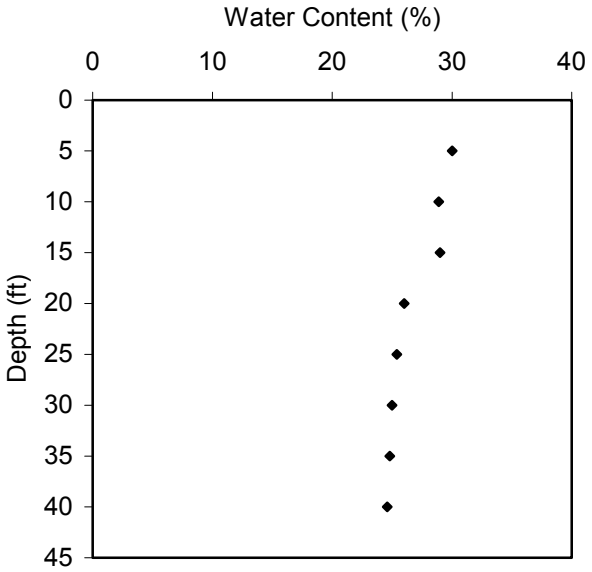


Fig. 3.2 Measured water content data

Water content, w (%)	Depth, d (ft)	w^d	w^2	d^2
30.0	5	150	900.00	25
28.9	10	289	835.21	100
29.0	15	435	841.00	225
26.0	20	520	676.00	400
25.4	25	635	645.16	625
25.0	30	750	625.00	900
24.8	35	868	615.04	1225
24.6	40	984	605.16	1600
$\Sigma w = 213.7$	$\Sigma d = 180$	$\Sigma w^d = 4631$	$\Sigma w^2 = 5742.57$	$\Sigma d^2 = 5100$

$$\rho = \frac{\mathrm{cov}(X,Y)}{\sqrt{V(X)V(Y)}} = \frac{N\sum wd - \sum w \sum d}{\sqrt{\left[N\sum w^2 - \left(\sum w\right)^2\right]\left[N\sum d^2 - \left(\sum d\right)^2\right]}} =$$

$$\frac{8(4631) - 213.7(180)}{\sqrt{\left[8(5742.57) - (213.7)^2\right]\left[8(5100) - (180)^2\right]}} = -0.937 .$$

The correlation coefficient shows strong negative correlation: the water content decreases as the depth increases.

3.5.2 Moment-Generating Functions

A moment-generating function (MGF) is used to find the moments of a particular distribution. Although MGFs apply to both discrete and continuous probability spaces, this summary focuses on continuous distributions because they are more applicable to geotechnical soil properties. The MGF is defined by:

$$m(s) = E(e^{sX}) \tag{3.15}$$

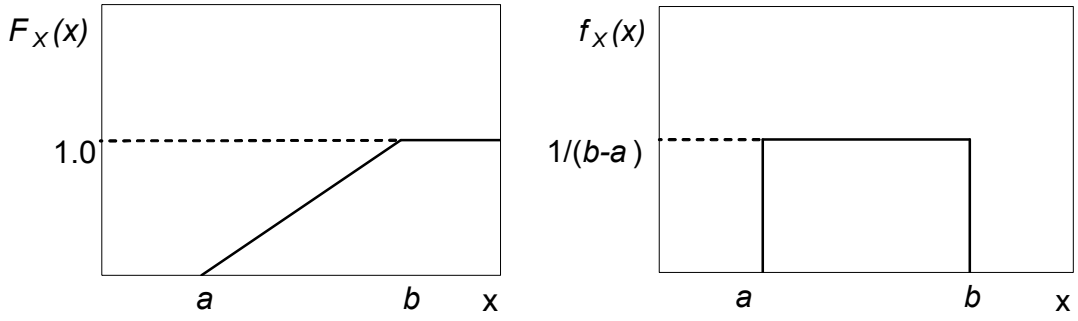
for only those numbers s for which this expected value exists. Moments can be generated by using the n^{th} derivative to find the n^{th} moment, or expected values to the power n. In general, $\frac{d^n m}{ds^n} = E(X^n e^{sX})$, and $\frac{d^n m}{ds^n} = E(X^n)$ evaluated at $s = 0$. Therefore, $\frac{dm}{ds} = E(X)$ and $\frac{d^2 m}{ds^2} = E(X^2)$ evaluated at $s = 0$.

3.6 SPECIAL PROBABILITY DISTRIBUTIONS

Many special probability distributions exist, most notably the family of normal distributions. Other important distributions include the uniform, exponential, and gamma distributions. Rather than focus on the derivations, this section presents useful properties of these distributions. Because few, if any, geotechnical properties will behave as a discrete probability space, special discrete distributions are not presented herein. The reader is referred to Ott (1984) and Kelly (1994) for a discussion of discrete systems.

3.6.1 Uniform Distribution

Some random variables are equally likely to take on any value within an interval. The uniform distribution models a number chosen at random from the interval (a,b) such that no part of the interval is favored over any other part of the same size. Such a random variable, X, is said to be uniformly distributed with parameters a and b, and is denoted $X \sim U(a,b)$. The probability density function (pdf), cumulative distribution function (CDF), moment-generating function (MGF), $E(X)$, and $V(X)$ are shown in Table 3.3. The pdf and CDF for a uniform distribution are illustrated in Figure 3.3.

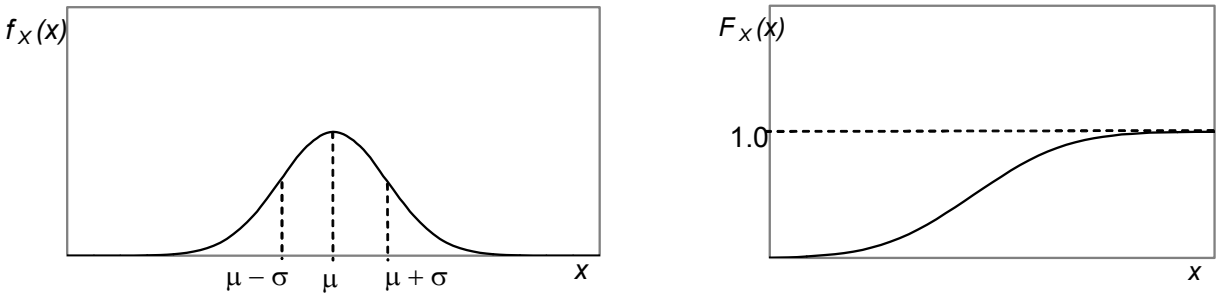


(a) (b)

Fig. 3.3 Uniform distribution: (a) probability density function, (b) cumulative distribution function

3.6.2 Normal Distribution

A family of random variables called "normal random variables" model randomly chosen members of some large population as outlined in the *Central Limit Theorem* (refer to Kelly, 1984 or Ott, 1984). The random variable X is said to be normally distributed with parameters μ (mean) and σ^2 (variance), and is denoted $X \sim N(\mu,\sigma^2)$. The pdf, CDF, MGF, *E(X)*, and *V(X)* are shown in Table 3.3. The pdf and CDF for a normal distribution are illustrated in Figure 3.4. Normally distributed random variables vary from $-\infty$ to $+\infty$.



(a) (b)

Fig. 3.4 Normal distribution: (a) probability density function, (b) cumulative distribution function

The moment-generating function can be used to find the mean and variance. For a normal distribution, the MGF is $m(s) = e^{\mu s + \sigma^2 s^2 / 2}$. To find *E(X)*, the first derivative is evaluated at *s* = 0. Therefore, $m'(s) = (\mu + 2\sigma^2 s / 2) e^{\mu s + \sigma^2 s^2 / 2}$ evaluated at $s = 0$ is $E(X) = \mu$. By using this methodology, the variance, σ^2, can be obtained. The $E(X^2)$ is first obtained by taking the second derivative of *m(s)* and evaluating at $s = 0$. This results in $\sigma^2 + \mu^2$. By definition, the variance is $E(X^2) - [E(X)^2]$; therefore the variance is σ^2.

Any normally distributed random variable, *X*, can be transformed to a standard normal variable, *Z*, by $Z = (X - \mu)/\sigma$. The standard normal variable has the properties $\mu = 0$ and $\sigma^2 = 1$; its distribution is called the "standard normal distribution." The CDF of the standard normal distribution is given in Table 3.2.

Table 3.2 Values of the CDF of the standard normal distribution, $F_Z(Z) = 1 - F_Z(-Z)$

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
-3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
-3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0005	.0005	.0005	.0005
-3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
-3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
-2.9	.0019	.0018	.0017	.0017	.0016	.0016	.0015	.0015	.0014	.0014
-2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
-2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
-2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
-2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
-2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
-2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
-2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
-2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
-2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
-1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
-1.8	.0359	.0352	.0344	.0336	.0329	.0322	.0314	.0304	.0301	.0294
-1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
-1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
-1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
-1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0722	.0708	.0694	.0681
-1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0859	.0853	.0838	.0823
-1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
-1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
-1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
-0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
-0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
-0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
-0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
-0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
-0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
-0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
-0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
-0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4365	.4325	.4286	.4247
-0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

3.6.3 Lognormal Distribution

A parameter is said to be lognormally distributed if its logarithm is normally distributed. The *log normal* distribution is shown in Figure 3.5. Therefore, if X is lognormally distributed, then $Y = \ln X$ is normally distributed. In this case, the statistical parameter Z, is:

$$Z = \frac{\ln X - \mu_{\ln x}}{\sigma_{\ln x}}. \tag{3.16}$$

Therefore, the lower bound of the lognormal distribution is zero and the upper bound is infinite. The lognormal distribution provides a convenient model for random variables with relatively large coefficients of variation (> 30%) for which an assumption of normality would imply a significant probability of negative values (USACE, 1999). Random variables often assumed to be lognormally distributed include the coefficient of permeability, the undrained shear strength of clay, and factors of safety (USACE, 1999).

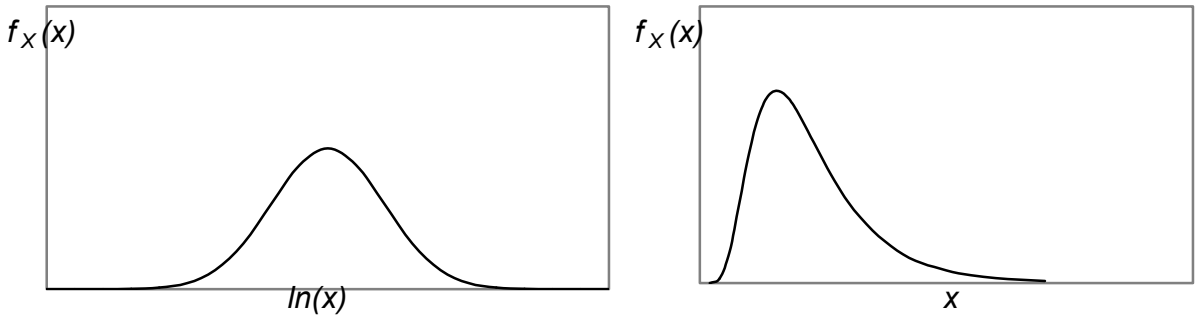
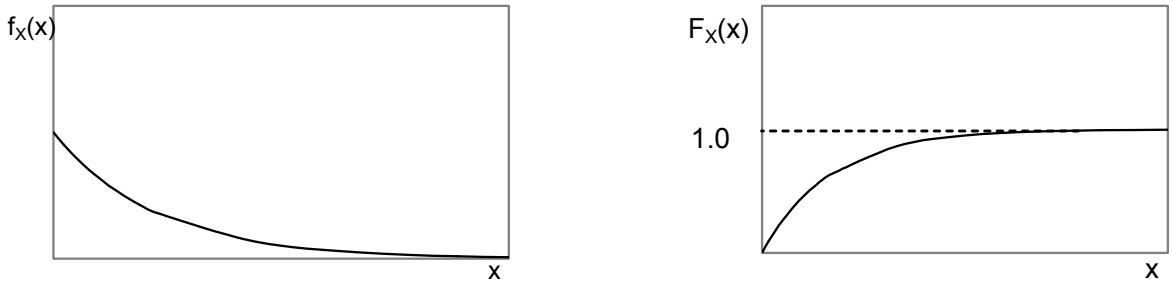


Fig. 3.5 Lognormal distribution: (a) probability density function of ln X, (b) probability density function of X showing no negative values and asymmetry

3.6.4 Exponential Distribution

The exponential distribution can be used to model a number of physical phenomena, such as the time, t, for a component to fail, or the distance, d, that an object travels before a collision. Exponential random variables are most commonly used to model a time-dependent process (e.g., Poisson process) with an arrival rate of λ arrivals per unit time. The random variable X is said to be exponentially distributed with parameter λ, and is denoted $X \sim \exp(\lambda)$. The pdf, CDF, MGF,

E(X), and *V(X)* are shown in Table 3.3. The pdf and CDF for an exponential distribution are illustrated in Figure 3.6.

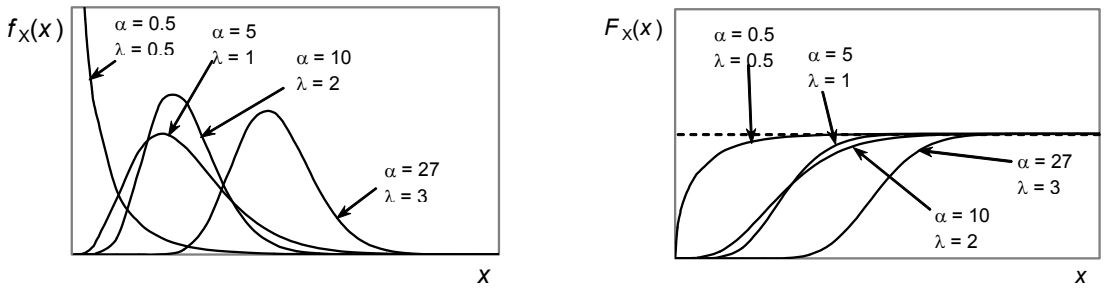


(a) (b)

Fig. 3.6 Exponential distribution: (a) probability density function, (b) cumulative distribution function

3.6.5 Gamma Distribution

The gamma distribution is the distribution of the sum of the squares of *n* independent, normally distributed random variables. Gamma distributions with various combinations of parameters are useful to model properties with positive values that tend to cluster near some value, but tend to have some very large values that produce a long right tail. The random variable *X* is said to be gamma distributed with parameters α and λ, and is denoted $X \sim \text{gam}(\alpha, \lambda)$. The pdf, CDF, MGF, *E(X)*, and *V(X)* are shown in Table 3.3. The pdf and CDF will vary for varying α and λ values as shown in Figure 3.7. Note that when $\alpha = 1$, the gamma density becomes the exp(λ) density.



(a) (b)

Fig. 3.7 Gamma distribution: (a) probability density function, (b) cumulative distribution function

Table 3.3 Properties of special distribution functions

Distribution Name	pdf, $f_X(x)$	CDF, $F_X(x)$	MGF, $m(s)$	$E(X)$	$V(X)$
Uniform	$\begin{cases} \frac{1}{b-a}, if\ a < x < b \\ 0, \quad otherwise \end{cases}$	$\begin{cases} 0, \quad if\ x \le 0 \\ \frac{x-a}{b-a}, if\ 0 < x < 1 \\ 1, \quad if\ x \ge 1 \end{cases}$	$\frac{e^{bs} - e^{as}}{s(b-a)}\ for\ s \neq 0$, and $m(0)=1$	$\frac{b+a}{2}$	$\frac{(b-a)^2}{12}$
Normal	$\frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$ for all real x	$\int_{-\infty}^{z} \varphi(z)dz$, where $\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$, $Z = (X-\mu)/\sigma$	$m(s) = e^{\mu s} + e^{\sigma^2 s^2/2}$ for all real s	μ	σ^2
Exponential	$\begin{cases} \lambda e^{-\lambda x}, if\ x > 0 \\ 0, \quad otherwise \end{cases}$	$\begin{cases} 0, \quad if\ x \le 0 \\ 1 - e^{-\lambda x}, if\ x < 0 \end{cases}$	$\frac{\lambda}{\lambda - s}\ for\ s < \lambda$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gamma	$\begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, if\ x, \lambda, \beta > 0 \\ 0, \quad otherwise \end{cases}$ $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx = \frac{n!}{\lambda^{n+1}}$ for any $\lambda > 0$ and n=0,1,2,…	There is no convenient formula unless α is positive.	$\left(\frac{\lambda}{\lambda - s}\right)^\alpha\ for\ s < \lambda$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$

3.6.6 Estimating Probability Distributions

It is sometimes adequate to know only estimates of the mean and standard deviation of the random variable, and knowledge of the form of the probability density function may not be necessary. However, in order to ensure that these estimates are reasonable and check assumptions regarding the shape of the distribution, it is recommended that the shape of the distribution be plotted as a check. A suggested method to assign or check assumed moments of random variables (USACE, 1995) is to:

- Assume trial values for the expected value and standard deviation and take the random variable to be normal or lognormal.
- Plot the resulting density function, and tabulate and plot the resulting CDF.

- Assess the reasonableness of the shape of the pdf and the values of the CDF.
- Repeat the above steps with successively improved estimates of the expected value and standard deviation until and appropriate pdf and CDF are obtained.

3.7 SUMMARY

A basic understanding of the theory of probability can be used to compute basic statistical parameters of a soil property. Probability distributions, expectation, and moments are the basic statistical descriptors of a random variable. These descriptors can be used to estimate the variability of geotechnical soil properties.

4 Uncertainty in Soil Properties

Most soils are naturally formed in many different depositional environments; therefore their physical properties will vary from point to point. This variation can exist even in an apparently homogeneous soil unit. Variability of soil properties is a major contributor to the uncertainty in geotechnical engineering analyses. Laboratory test results on natural soils indicate that most soil properties can be considered as random variables conforming to the normal distribution function (Lumb, 1966; Tan et al., 1993). This section summarizes modeling of uncertainty in soil properties and presents values of *COV* from various sources for index properties, laboratory-measured properties, and field-measured properties.

4.1 QUANTIFYING UNCERTAINTY IN SOIL PROPERTIES

Insitu soil properties may vary vertically and horizontally for a variety of reasons, including:

- Depositional environment — in general, fine-grained soils are deposited in low-energy environments and are therefore more uniform than course-grained soils, which are usually deposited in high-energy environments.
- Degree of weathering — soil properties can be influenced by weathering, a factor that affects soil at the ground surface most strongly and that decreases with depth below the ground surface. However, factors such as erosion and locally variable rates of deposition can produce soil profiles with variable weathering effects.
- Physical environment — most soils exhibit properties, such as inherent and induced anisotropy, that are influenced by their physical environment. Because stress changes can occur locally during the lifetime of a soil deposit, their effects can introduce uncertainty into measured soil properties.

The spatial variation of soil properties consists of several components and can be represented by a simple model (Phoon and Kulhawy, 1999):

$$\xi(z) = t(z) + w(z) + e(z) \tag{4.1}$$

where ξ = insitu soil property, t = deterministic trend component, w = random component, e = measurement error, and z = depth. The trend and random components are illustrated graphically in Figure 4.1.

Basic principles of soil mechanics indicate that many soil properties of interest are strongly influenced by effective confining pressure. Because effective confining pressures generally increase with depth, these properties should be expected to exhibit some regular, predictable trend with depth. The trend can be determined by fitting (in a least squares sense) a smooth deterministic function (e.g., a straight line, parabola, or exponential) to the data or by a moving average procedure.

The random component of soil variability is also referred to as "inherent soil variability," which is expressed relative to the deterministic trend of the property as illustrated in Figure 4.1. Measurement error, whether it be from laboratory or field measurements, can introduce additional variability into soil properties. Measurement error can arise from equipment, operator, and random testing effects (Phoon and Kulhawy, 1999).

The scale of fluctuation is a term that describes the spatial fluctuation of the property of interest about the trend (Fig. 4.1). A parameter with a short scale of fluctuation changes rapidly with position, one with a long scale of fluctuation changes over greater distances. A procedure for calculation of the scale of fluctuation is described in Chapter 5, and typical values of the scale of fluctuation are presented in Chapter 6.

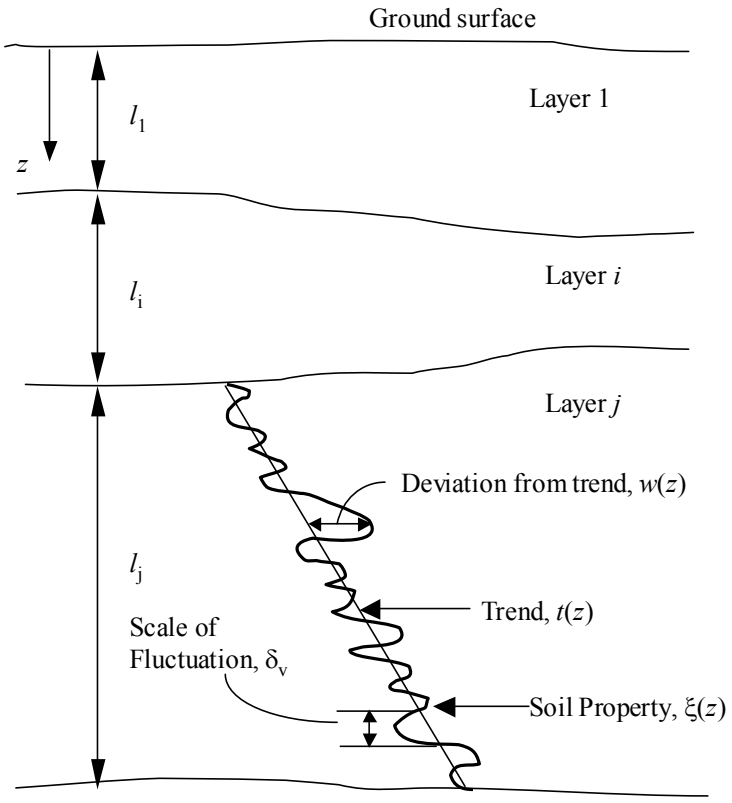


Fig. 4.1 Inherent soil variability (after Phoon and Kulhawy, 1999)

The following sections present a summary of text and tables to assist PEER researchers in estimating variability in geotechnical soil properties. Soil properties are summarized in three main categories: laboratory-measured properties, field-measured properties, and properties obtained from other types of testing. Each category presents various forms of variability consisting of inherent variability, measurement variability, and an estimate of spatial correlation (as expressed in terms of scale of fluctuation). The primary purpose of these sections is to present *COV* data; parameter values are presented to simply allow PEER researchers to judge the conditions for which the *COV* values are applicable. Total variability can be computed using the following equation:

$$COV(total) = \sqrt{COV(measurement)^2 + COV(random)^2} \ . \tag{4.2}$$

4.2 LABORATORY-MEASURED PROPERTIES

Laboratory measured soil properties are important for determining soil indices, strength parameters, and consolidation characteristics.

4.2.1 Inherent Variability

This section presents tabulated *COV* values of inherent variability of various laboratory-measured properties.

4.2.1.1 Moisture-Density Characteristics

The unit weight (or density) of a soil is important for determining states of stress beneath and adjacent to structures. Unit weights are typically determined by measuring the weight and volume of soil samples in the laboratory.

Lacasse and Nadim (1996) suggest the pdf for unit weight is normally distributed for all soil types. Table 4.1 presents tabulated *COV* data for natural water content, w_n, total unit weight, γ, dry unit weight, γ_d, buoyant (submerged) unit weight, γ_b, relative density, D_r, specific gravity, G_s, and degree of saturation, S.

Table 4.1 COV of inherent soil variability for moisture content, unit weight, and relative density

Property (units)	Soil Type	No. of Data Groups	No. of Tests Per Group		Property Value		Property COV (%)		Note
			Range	Mean	Range	Mean	Range	Mean	
w_n (%)	Fine-grained	40	17 – 439	252	13 – 105	29	7 – 46	18	1
	Silty clay	*	*	*	*	*	20	*	3
	clay	*	*	*	*	*	13	*	
	*	18	*	*	*	*	*	17.7	4
γ (kN/m^3)	Fine-grained	6	5 – 3200	564	14 – 20	17.5	3 – 20	9	1
γ	*	*	*	*	*	*	3	*	3
γ	*	12	*	*	*	*	*	7.1	4
γ_d (kN/m^3)	Fine-grained	8	4 – 315	122	13 – 18	15.7	2 – 13	7	1
γ_b (kN/m^3)	All soils	*	*	*	5 – 11	*	0 – 10	*	2
$D_r^{(a)}$ (%)	Sand	5	*	*	30 – 70	50	11 – 36	19	1
$D_r^{(b)}$ (%)	Sand	5	*	*	30 – 70	50	49 – 74	61	
G_s	*	*	*	*	*	*	2	*	3
S	*	*	*	*	*	*	10	*	3

*Not reported.
(a) Total variability for direct method of determination.
(b) Total variability for indirect determination using SPT values.
Notes:
(1) Phoon and Kulhawy (1999).
(2) Lacasse and Nadim (1996). No comments made on whether measurement variability was included.
(3) Harr (1987). No comments made on whether measurement variability was included.
(4) Kulhawy (1992). No comments made on whether measurement variability was included.

4.2.1.2 Plasticity Characteristics

Plasticity characteristics are important for classification of soil types and for determining engineering behavior. Plasticity indices are usually measured in terms of Atterberg limits in the laboratory.

Lacasse and Nadim (1996) suggest that the plastic limit, *PL*, and the liquid limit, *LL*, are normally distributed in clays. Table 4.2 presents tabulated *COV* data for liquid limit, plastic limit, plasticity index, *PI*, and liquidity index, *LI*.

Table 4.2 COV of inherent soil variability for plasticity indices

Property (%)	Soil Type	No. of Data Groups	No. of Tests Per Group		Property Value		Property COV (%)		Note
			Range	Mean	Range	Mean	Range	Mean	
LL	Fine-grained	38	15 – 299	129	27 – 89	51	7 – 39	18	1
	Clay	*	*	*	30 – 80	*	3 – 20	*	2
	*	28	*	*	*	*	*	11.3	4
PL	Fine-grained	23	32 – 299	201	14 – 27	22	6 – 34	16	1
	Clay	*	*	*	13 – 23	*	3 – 20	*	2
	*	27	*	*	*	*	*	11.3	4
PI	Fine-grained	33	15 – 299	120	12 – 44	25	9 – 57	29	1
LI	Clay, silt	2	32 – 118	75	*	0.094	60 – 88	74	1

* Not reported.
Notes:
(1) Phoon and Kulhawy (1999).
(2) Lacasse and Nadim (1996). No comments made on whether measurement variability was included.
(3) Harr (1987). No comments made on whether measurement variability was included.
(4) Kulhawy (1992). No comments made on whether measurement variability was included.

4.2.1.3 Strength Characteristics

Strength parameters of soils are probably the most important parameters used in geotechnical engineering. Most design methodologies rely on the strength of soils as input. Therefore, recognition of uncertainty in soil strength properties is very important. Various laboratory methods exist for measuring strength parameters, and the inherent variability in the strength parameters they produce is summarized in this section.

Lacasse and Nadim (1996) and Wolff et al. (1996) suggest that the pdf for friction angle, ϕ, is normally distributed in sands. They also suggest that a lognormal pdf be used for undrained shear strength, s_u, in clays, and that a normal pdf be used for s_u in clayey silts. Lacasse and Nadim (1996) suggest the pdf for undrained strength ratio, s_u/σ'_{v0}, is either normally or lognormally distributed for clay soils. Table 4.3 presents tabulated *COV* data for drained friction angle, $\overline{\phi}$, tangent drained friction angle, $\tan\overline{\phi}$, undrained shear strength, cohesion, *c*, and undrained strength ratio.

Table 4.3 COV of inherent soil variability for strength parameters

Property (units)	Soil Type	No. of Data Groups	No. of Tests Per Group		Property Value		Property COV (%)		Note
			Range	Mean	Range	Mean	Range	Mean	
$\overline{\phi}$ (°)	Sand	7	29 – 136	62	35 – 41	37.6	5 – 11	9	1
	Clay, silt	12	5 – 51	16	9 – 33	15.3	10 – 56	21	
	Clay, silt	9	*	*	17 – 41	33.3	4 – 12	9	
	*	20	*	*	*	*	*	12.6	4
$\tan\overline{\phi}$	Clay, silt	4	*	*	0.24 – 0.69	0.509	6 – 46	20	1
$\tan\overline{\phi}$	Clay, silt	3	*	*	*	0.615	6 – 46	23	
$\tan\overline{\phi}$	Sand	13	6 – 111	45	0.65 – 0.92	0.744	5 – 14	9	1
	*	7	*	*	*	*	*	11.3	4
$\overline{\phi}$ (°)	Sand	*	*	*	*	*	2 – 5	*	2
ϕ (°)	Gravel	*	*	*	*	*	7	*	3
	Sand	*	*	*	*	*	12	*	
s_u[(a)] (kPa)	Fine-grained	38	2 – 538	101	6 – 412	100	6 – 56	33	1
s_u[(b)] (kPa)	Clay, Silt	13	14 – 82	33	15 – 363	276	11 – 49	22	
s_u[(c)] (kPa)	Clay	10	12 – 86	47	130 – 713	405	18 – 42	32	
s_u[(d)] (kPa)	Clay	42	24 – 124	48	8 – 638	112	6 – 80	32	
	*	38	*	*	*	*	*	33.8	3
s_u[(e)] (kPa)	Clay	*	*	*	*	*	5 – 20	*	2
s_u[(f)] (kPa)	Clay	*	*	*	*	*	10 – 35	*	
s_u[(d)] (kPa)	Clayey silt	*	*	*	*	*	10 – 30	*	
c[(g)]	*	*	*	*	*	*	40	*	3
s_u/σ'_{v0}	Clay	*	*	*	*	*	5 – 15	*	2

* Not reported.
(a) Unconfined compression test.
(b) Unconsolidated-undrained triaxial compression test.
(c) Consolidated isotropic undrained triaxial compression test.
(d) Laboratory test not reported.
(e) Triaxial test.
(f) Index s_u.
(g) No specification on how the parameter was defined.
Notes:
(1) Phoon and Kulhawy (1999).
(2) Lacasse and Nadim (1996). No comments made on whether measurement variability was included.
(3) Harr (1987). No comments made on whether measurement variability was included.
(4) Kulhawy (1992). No comments made on whether measurement variability was included.

4.2.1.4 Consolidation and Permeability Characteristics

Consolidation and permeability characteristics are important to quantify stress/strain relations and the time-dependent behavior of soils. This section presents typical uncertainties in the various laboratory parameters that are used to characterize consolidation and permeability behavior.

Lacasse and Nadim (1996) suggest that the pdf for the overconsolidation ratio is either normally or lognormally distributed for clay soils. They also suggest (1996) that the pdf for void

ratio, porosity, and initial void ratio is normally distributed for all soil types. Table 4.4 presents tabulated *COV* data for compression index, C_c, preconsolidation pressure, p_c', overconsolidation ratio, *OCR*, coefficient of permeability, k, coefficient of consolidation, c_v, void ratio, e, and porosity, n.

Table 4.4 COV of inherent soil variability for consolidation and permeability parameters

Property (units)	Soil Type	No. of Data Groups	No. of Tests Per Group		Property Value		Property COV (%)		Note
			Range	Mean	Range	Mean	Range	Mean	
C_c	Sandy clay	*	*	*	*	*	26	*	1
	Clay	*	*	*	*	*	30	*	
	*	*	*	*	*	*	37	*	2
p_c'	*	*	*	*	*	*	19	*	1
OCR	*	*	*	*	*	*	10 – 35	*	3
k	*	*	*	*	*	*	240[(a)]	*	1
	*	*	*	*	*	*	90[(b)]	*	
c_v	*	*	*	*	*	*	33 – 68	*	4
e, n, e_o	All soil types	*	*	*	*	*	7 – 30	*	5
n	*	*	*	*	*	*	10	*	1

* Not reported.
(a) 80% saturation.
(b) 100% saturation.
Notes:
(1) Harr (1987).
(2) Kulhawy (1992). No comments made on whether measurement variability was included.
(3) Lacasse and Nadim (1996). No comments made on whether measurement variability was included.
(4) Duncan (2000).
(5) Lacasse and Nadim (1996).

4.2.1.5 Stiffness and Damping Characteristics

The seismic response of soil deposits is strongly influenced by the stiffness and damping characteristics of the soil. These characteristics are typically described in an equivalent linear framework, i.e., by maximum shear modulus, modulus reduction curves, and damping curves. While data and procedures for deterministic prediction of these parameters have been reported, very little explicit information on their uncertainty is available.

Seed and Idriss (1970) presented experimental data on the variation of shear modulus and damping ratio with cyclic strain amplitude, and used the data to develop their widely used modulus reduction and damping curves for sands and clays. The data for sands came from many sources that used different types of testing equipment and different (and, in many cases, large) ranges of effective confining pressure. For sands, shear modulus results were expressed in terms

of the parameter $K_2 = G/1000(\sigma'_m)^{1/2}$ where G and σ'_m are in psf. Shear moduli for clays were normalized by undrained strength. Damping ratios for sands and clays were also presented graphically. The plots of experimental modulus reduction and damping data in Seed and Idriss (1970) suggest substantial uncertainty, but the ranges of effective confining pressures (for tests on sands) and plasticity characteristics (for clay specimens) are now known to have a systematic effect on modulus reduction and damping behavior. As a result, statistical parameters computed directly from this data would likely overpredict the inherent variability of these properties.

About ten years ago, the Electric Power Research Institute (EPRI) undertook an investigation of appropriate methods for estimating earthquake ground motion in eastern North America. This work, which involved numerous investigators, included extensive, high-quality field and laboratory testing of soils at more than 200 different sites. The measured shear wave velocities were shown to be lognormally distributed with $\sigma_{\ln V} = 0.39$ (velocity measured in m/sec). Laboratory modulus reduction and damping data were used to estimate and model variability in modulus reduction and damping curves. The modulus reduction ratio, G/G_{max}, at a cyclic shear strain of 0.03% was determined to be lognormally distributed with $\sigma_{\ln G/Gmax} = 0.35$ (truncated at 2σ). Monte Carlo analyses were performed with the median modulus reduction and damping curves scaled by a constant value (thereby retaining the shapes of the median curves) that produced the target variability at $\gamma = 0.03\%$. It should be noted that the characterization of constant standard deviation implies that the COV of G/G_{max} increases relatively rapidly with increasing shear strain amplitude.

Very high-quality testing (resonant column and torsional shear) has been performed under the direction of Prof. Kenneth H. Stokoe at the University of Texas for many research projects including the previously described EPRI project, a Savannah River Site investigation, and the ROSRINE program (*http://rccg03.usc.edu/Rosrine/*). Darendian (2001) used data from more than 20 sites to develop a model for modulus reduction and damping behavior that included estimates of uncertainty in G/G_{max} and damping ratio. The model, which expresses G/G_{max} as a function of shear strain, plasticity index, OCR, effective confining pressure, and soil type, was calibrated using a first-order, second-moment Bayesian technique. The variance in G/G_{max} and damping ratio was observed to vary with strain level and with soil type; examples are presented in Figure 4.2. Details of the Darendian-Stokoe model will be presented in technical papers that are nearing completion at the time of this report.

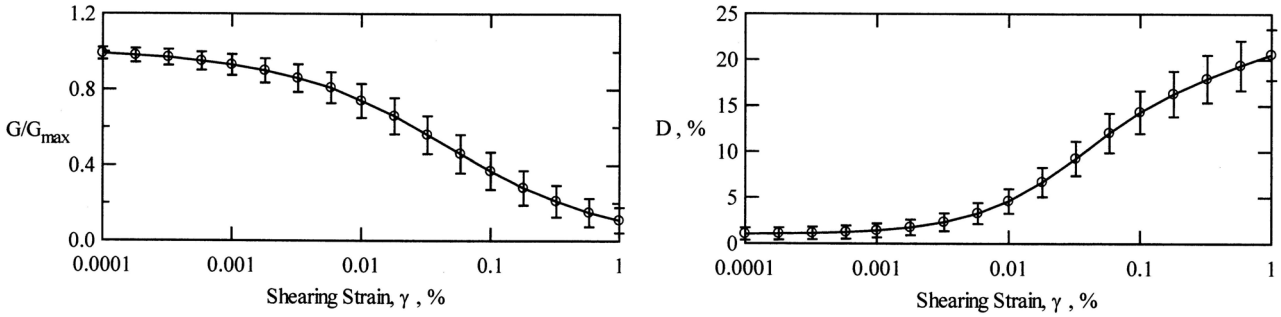


Fig. 4.2 Variation with shear strain amplitude of mean and standard deviation of typical modulus reduction and damping curves (Darendian, 2001)

4.2.2 Measurement Variability

The process of quantifying uncertainty includes a component due to measurement variability, which can arise from errors in the laboratory equipment, errors by the person conducting the laboratory test, and random testing effects that cannot be separately measured. Table 4.5 presents tabulated *COV* data for measurement error for some laboratory-measured properties.

Table 4.5 Summary of total measurement error for laboratory-measured properties (after Phoon and Kulhawy (1999)

Property (units)	Soil Type	No. of Data Groups	No. of Tests Per Group		Property Value		Property COV (%)	
			Range	Mean	Range	Mean	Range	Mean
$s_u^{(a)}$ (kPa)	Clay, silt	11	*	13	7 – 407	125	8 – 38	19
$s_u^{(b)}$ (kPa)	Clay, silt	2	13 – 17	15	108 – 130	119	19 – 20	20
$s_u^{(c)}$ (kPa)	Clay	15	*	*	4 – 123	29	5 – 37	13
$\overline{\phi}^{(a)}$ (°)	Clay, silt	4	9 – 13	10	2 – 27	19.1	7 – 56	24
$\overline{\phi}^{(b)}$ (°)	Clay, silt	5	9 – 13	11	24 – 40	33.3	3 – 29	13
$\overline{\phi}^{(b)}$ (°)	Sand	2	26	26	30 – 35	32.7	13 – 14	14
$\tan\overline{\phi}^{(a)}$ (°)	Clay, silt	6	*	*	*	*	2 – 22	8
$\tan\overline{\phi}^{(b)}$ (°)	Clay	2	*	*	*	*	6 – 22	14
w_n (%)	Fine-grained	3	82 – 88	85	16 – 21	18	6 – 12	8
LL, (%)	Fine-grained	26	41 – 89	64	17 – 113	36	3 – 11	7
PL, (%)	Fine-grained	26	41 – 89	62	12 – 35	21	7 – 18	10
PI, (%)	Fine-grained	10	41 – 89	61	4 – 44	23	5 – 51	24
γ, (kN/m^2)	Fine-grained	3	82 – 88	85	16 – 17	17.0	1 – 2	1

* Not reported.
(a) Triaxial compression test.
(b) Direct shear test.
(c) Laboratory vane shear test.

4.3 FIELD-MEASURED PROPERTIES

In many cases, soil properties are determined or inferred from the results of field tests. Insitu field tests can be quite useful, particularly in soils that are difficult to sample without disturbance. This section presents a discussion of uncertainty for insitu testing and estimated values for quantifying uncertainty in the form of summarized *COV* values for various properties.

4.3.1 Inherent Variability

This section presents tabulated *COV* values of inherent variability of various insitu measured properties.

4.3.1.1 Standard Penetration Test (SPT) Resistance

SPT resistance is one of the most common forms of insitu testing conducted for geotechnical engineering. Therefore, quantifying uncertainty for SPT resistance is particularly useful. SPT resistance is typically collected using standardized sampling equipment consisting of a hammer, rods and split-spoon sampler. Table 4.6 presents tabulated data for SPT measured resistance, *N*.

Table 4.6 COV of inherent soil variability for SPT resistance

Property	Soil Type	No. of Data Groups	No. of Tests Per Group		Property Value		Property COV (%)		Note
			Range	Mean	Range	Mean	Range	Mean	
N	Clay & Sand	*	*	*	10 – 70	*	25 – 50	*	1
N	Sand	22	2 – 300	123	7 – 74	35	19 – 62	54	2
N	Clay Loam	2	2 – 61	32	7 – 63	32	37 – 57	44	
N	*	*	*	*	*	*	26	*	3

Notes:
(1) Phoon and Kulhawy (1996).
(2) Phoon and Kulhawy (1999).
(3) Harr (1987). No comment made on whether measurement variability was included.

4.3.1.2 Cone Penetration Test (CPT) and Electric Cone Penetration Test (ECPT) Resistance

CPT resistance is probably the second most frequently used test for collecting insitu soil measurements. CPT resistance is usually measured using mechanical means (MCPT) or electrically

(ECPT). CPT testing is usually conducted by pushing a standardized cone into the soil subsurface to measure resistance of the tip and friction of the sleeve.

Lacasse and Nadim (1996) suggest that a lognormal pdf be used for MCPT testing in sands and a normal or lognormal pdf be used for MCPT testing in clays. Table 4.7 presents tabulated *COV* data for CPT-corrected tip resistance, q_T, and tip resistance, q_c.

Table 4.7 COV of inherent soil variability for CPT measurements

Property (units)	Soil Type	No. of Data Groups	No. of Tests Per Group		Property Value		Property COV (%)		Note
			Range	Mean	Range	Mean	Range	Mean	
q_T (MPa)	Clay	*	*	*	0.5 – 2.5	*	< 20	*	1
	Clay	9	*	*	0.4 – 2.6	1.32	2 – 17	8	2
q_c (MPa)	Clay	*	*	*	0.5 – 2.0	*	20 – 40	*	1
	Sand	*	*	*	0.5 – 30.0	*	20 – 60	*	1
	Sand	57	10 – 2039	115	0.4 – 29.2	4.10	10 – 81	38	2
	Silty Clay	12	30 – 53	43	0.5 – 2.1	1.59	5 – 40	27	2
*	*	*	*	*	*	*	37	*	3

* Not reported
Notes:
(1) Phoon and Kulhawy (1996).
(2) Phoon and Kulhawy (1999).
(3) Harr (1987). No comment made on whether measurement variability was included.

4.3.1.3 Vane Shear Testing (VST) Undrained Shear Strength

Table 4.8 presented tabulated *COV* data of inherent soil variability for undrained shear strength as measured by VST measurement.

Table 4.8 COV of inherent soil variability of undrained shear strength using VST measurement

Property (units)	Soil Type	No. of Data Groups	No. of Tests Per Group		Property Value		Property COV (%)		Note
			Range	Mean	Range	Mean	Range	Mean	
s_u (kPa)	Clay	*	*	*	5 – 400	*	10 – 40	*	1
s_u (kPa)	Clay	31	4 – 31	16	6 – 375	105	4 – 44	24	2

* Not reported
Notes:
(1) Phoon and Kulhawy (1996).
(2) Phoon and Kulhawy (1999).

4.3.1.4 Dilatometer Test (DMT) Parameter

Table 4.9 presents tabulated *COV* data for DMT measurement parameters, *A* and *B*, material index, I_D, horizontal stress index, K_D, and modulus, E_D.

Table 4.9 COV of inherent soil variability of DMT measurement parameters

Property (units)	Soil Type	No. of Data Groups	No. of Tests Per Group		Property Value		Property COV (%)		Note
			Range	Mean	Range	Mean	Range	Mean	
A (kPa)	Clay	*	*	*	100 – 450	*	10 – 35	*	1
	Sand	*	*	*	60 – 1300	*	20 – 50	*	
	Sand to clayey sand	15	12 – 25	17	64 – 1335	512	20 – 53	33	2
	Clay	13	10 – 20	17	119 – 455	358	12 – 32	20	
B (kPa)	Clay	*	*	*	500 – 880	*	10 – 35	*	1
	Sand	*	*	*	350 – 2400	*	20 – 50	*	
	Sand to clayey sand	15	12 – 25	17	346 – 2435	1337	13 – 59	37	2
	Clay	13	10 – 20	17	502 – 876	690	12 – 38	20	
I_D	Sand	*	*	*	1 – 8	*	20 – 60	*	1
	Sand to clayey sand	15	10 – 25	15	0.8 – 8.4	2.85	16 – 130	53	2
	Sand, silt	16	*	*	2.1 – 5.4	3.89	8 – 48	30	
K_D	Sand	*	*	*	2 – 30	*	20 – 60	*	1
	Sand to clayey sand	15	10 – 25	15	1.9 – 28.3	15.1	20 – 99	44	2
	Sand, silt	16	*	*	1.3 – 9.3	4.1	17 – 67	38	
E_D (MPa)	Sand	*	*	*	10 – 50	*	15 – 65	*	1
	Sand to clayey sand	15	10 – 25	15	9.4 – 46.1	25.4	9 – 92	50	2
	Sand, silt	16	*	*	10.4 – 53.4	21.6	7 – 67	36	

* Not reported
Notes:
(1) Phoon and Kulhawy (1996).
(2) Phoon and Kulhawy (1999).

4.3.1.5 Pressuremeter Test (PMT) Parameters

Table 4.10 presents tabulated *COV* data for PMT limit stress, p_L, and modulus, E_{PMT}.

Table 4.10 COV of inherent soil variability of PMT parameters

Property (units)	Soil Type	No. of Data Groups	No. of Tests Per Group		Property Value		Property COV (%)		Note
			Range	Mean	Range	Mean	Range	Mean	
p_L (kPa)	Clay	*	*	*	400 – 2800	*	10 – 35	*	1
	Sand	*	*	*	1600 – 3500	*	20 – 50	*	
	Sand	4	*	*	1617 – 3566	2284	23 – 50	40	2
	Cohesive	5	10 – 25	*	428 – 2779	1084	10 – 32	15	
E_{PMT} (MPa)	Sand	*	*	*	5 – 15	*	15 – 65	*	1
	Sand	4	*	*	5.2 – 15.6	8.97	28 – 68	42	2

* Not reported.
Notes:
(1) Phoon and Kulhawy (1996).
(2) Phoon and Kulhawy (1999).

4.3.2 Measurement Variability

Measurement variability arises from errors in insitu testing equipment, by the person operating the equipment, and by random testing effects that cannot be separately measured.

4.3.2.1 Standard Penetration Test (SPT) Resistance

A number of factors can influence the measurement variability of SPT results; the most important of these are presented in Table 4.11. Uncertainty estimates of the *COV* (in percent) for the various factors include the following (Kulhawy and Trautmann, 1996):

(1) Equipment: 5 – 75 (best to worst case);

(2) Procedure: 5 – 75 (best to worst case);

(3) Random: 12 – 15;

(4) Total: 14 – 100, where the total is computed by $COV(total) = \left[COV(equipment)^2 + COV(procedure)^2 + COV(random)^2\right]^{1/2}$; and

(5) Range: 15 – 45 where the range represents probable magnitudes of field test measurement error considering limited data and judgment involved in estimating COV.

Table 4.11 Sources of variability in SPT test results (Kulhawy and Trautmann, 1996)

SPT Variable		Relative Effect on Test Results
Type	**Item**	
Equipment	Non-standard sampler	Moderate
	Deformed or damaged sampler	Moderate
	Rod diameter/weight	Minor
	Rod length	Minor
	Deformed drill rods	Minor
	Hammer type	Moderate to significant
	Hammer drop system	Significant
	Hammer weight	Minor
	Anvil size	Moderate to significant
	Drill rig type	Minor
Procedural/ Operator	Borehole size	Moderate
	Method of maintaining hole	Minor to significant
	Borehole cleaning	Moderate to significant
	Insufficient hydrostatic head	Moderate to significant
	Seating of sampler	Moderate to significant
	Hammer drop method	Moderate to significant
	Error in counting blows	Minor

4.3.2.2 Cone Penetration Test (CPT) Resistance

The major factors influencing the variability of CPT results are presented in Table 4.12.

Table 4.12 Sources of variability in CPT test results (Kulhawy and Trautmann, 1996)

CPT Variable		Relative Effect on Test Results
Type	**Item**	
Equipment	Cone type (MCPT or ECPT)	Moderate to significant
	Cone size	Minor
	Cone angle	Moderate to significant
	Rod compression (MCPT)	Significant
	Manufacturing defects	Minor to moderate
	Leaky seals (ECPT)	Minor
	Excessive cone wear	Minor to moderate
Procedural/ Operator	Telescoping vs. continuous penetration	Moderate to significant
	Calibration error	Minor to moderate
	Penetration rate	Minor
	Inclined penetration	Moderate to significant

MCPT

Uncertainty estimates of the *COV* (in percent) for the various factors include the following (Kulhawy and Trautmann, 1996):

(1) Equipment: 5;

(2) Procedure: 10 – 15 (tip and side resistances, respectively);

(3) Random: 10 – 15 (tip and side resistances, respectively);

(4) Total: 15 – 22 (tip and side resistances, respectively), where the total is computed by $COV(total) = \left[COV(equipment)^2 + COV(procedure)^2 + COV(random)^2\right]^{1/2}$; and

(5) Range: 15 – 25 where the range represents probable magnitudes of field test measurement error considering limited data and judgment involved in estimating COV.

ECPT

Uncertainty estimates of the *COV* (in percent) for the various factors include the following (Kulhawy and Trautmann, 1996):

(1) Equipment: 3;

(2) Procedure: 5;

(3) Random: 5 – 10 (tip and side resistances, respectively);

(4) Total: 8 – 12 (tip and side resistances, respectively), where the total is computed by $COV(total) = \left[COV(equipment)^2 + COV(procedure)^2 + COV(random)^2\right]^{1/2}$; and

(5) Range: 5 – 15 where the range represents probable magnitudes of field test measurement error considering limited data and judgment involved in estimating COV.

4.3.2.3 Vane Shear Testing (VST) Undrained Shear Strength

The major factors influencing the variability of VST undrained shear strength are presented in Table 4.13. Uncertainty estimates of the COV (in percent) for the various factors consist of the following (Kulhawy and Trautmann, 1996):

(1) Equipment: 5;

(2) Procedure: 8;

(3) Random: 10;

(4) Total: 14, where the total is computed as above; and

(5) Range: 10 – 20 where the range represents probable magnitudes of field test measurement error considering limited data and judgment involved in estimating COV.

Table 4.13 Sources of variability in VST test results (Kulhawy and Trautmann, 1996)

VST Variable		Relative Effect on Test Results
Type	**Item**	
Equipment	Vane length	Minor
	Height/diameter ratio	Moderate
	Blade thickness	Moderate
	Torque measuring device	Moderate to significant
	Damaged vane	Moderate to significant
Procedural/ Operator	Vane insertion method	Moderate to significant
	Rod friction calibration	Moderate to significant
	Time delay between insertion and testing	Minor to moderate, to significant in soft clays
	Vane rotation rate	Moderate

4.3.2.4 Dilatometer Test (DMT) Parameter

The major factors influencing the variability of DMT results are presented in Table 4.14. Uncertainty estimates of the COV (in percent) for the various factors consist of the following (Kulhawy and Trautmann, 1996):

(1) Equipment: 5;

(2) Procedure: 5;

(3) Random: 8;

(4) Total: 11, where the total is computed as above; and

(5) Range: 5 – 15 where the range represents probable magnitudes of field test measurement error considering limited data and judgment involved in estimating COV.

Table 4.14 Sources of variability in DMT test results (Kulhawy and Trautmann, 1996)

DMT Variable		Relative Effect on Test Results
Type	**Item**	
Equipment	Leaking seals	Minor
	Deformed membrane	Moderate
	Bent or deformed push rods	Minor to moderate
	Damaged blade	Minor
Procedural/ Operator	Push rod inclination	Minor to moderate
	Testing rate	Moderate to significant
	Driving method	Minor to moderate
	Rod friction	Minor
	Calibration error	Minor to moderate

4.3.2.5 Pressuremeter Test (PMT) Parameter

The major factors influencing the variability of PMT results are presented in Table 4.15.

Table 4.15 Sources of variability in PMT test results (Kulhawy and Trautmann, 1996)

PMT Variable		Relative Effect on Test Results
Type	**Item**	
Equipment	Gage error	Minor
	Expansion of tubing	Minor to moderate
	Friction loss in tubing	Minor
	Probe dimensions	Minor to moderate
	Probe design (PMT)	Minor
	Membrane aging	Minor
	Cutting shoe size (SBPMT)	Minor to moderate
	Cutting position (SBPMT)	Minor
	Probe shape (SBPMT)	Minor to moderate
	Drilling equipment (SBPMT)	Minor to moderate
	Electrical sensor compliance (SBPMT)	Minor
Procedural/ Operator	Drilling method and borehole preparation (PMT)	Significant
	Probe inflation rate	Minor to moderate
	Relaxation time (SBPMT)	Moderate to significant
	Probe advance rate (SBPMT)	Moderate to significant

PMT

Uncertainty estimates of the *COV* (in percent) for the various factors include the following (Kulhawy and Trautmann, 1996):

(1) Equipment: 5;

(2) Procedure: 12;

(3) Random: 10;

(4) Total:16, where the total is computed by $COV(total) = \left[COV(equipment)^2 + COV(procedure)^2 + COV(random)^2\right]^{1/2}$; and

(5) Range: 10 – 20 where the range represents probable magnitudes of field test measurement error considering limited data and judgment involved in estimating COV. Results may differ for p_0, p_f, and p_l, but data are insufficient to clarify.

SBPMT

Uncertainty estimates of the *COV* (in percent) for the various factors include the following (Kulhawy and Trautmann, 1996):

(1) Equipment: 8;

(2) Procedure: 15;

(3) Random: 8;

(4) Total: 19, where the total is computed by $COV(total) = \left[COV(equipment)^2 + COV(procedure)^2 + COV(random)^2\right]^{1/2}$; and

(5) Range: 15 – 25 where the range represents probable magnitudes of field test measurement error considering limited data and judgment involved in estimating COV. Results may differ for p_0, p_f, and p_l, but data are insufficient to clarify.

4.4 OTHER TYPES OF TESTING

Values for fractal dimensional data are given by Vallejo (1996). Fractal dimension is a measure of the degree of the irregularity of plots representing the variability of engineering properties verses the depth of a soil profile. Values for variability of ground-penetrating radar in sands are given by Young and Doucette (1996).

5 Spatial Variability

The terms used to characterize uncertainty thus far have disregarded correlation in space. Soil properties, however, can be expected to be spatially correlated. Characterization of the spatial distribution of soil properties requires the use of *regionalized variables*, which have a particular structure with properties between a truly random variable and one that is completely deterministic. This means that the properties of a regionalized variable at locations X and $X + \Delta h$, where Δh is a separation distance, are correlated. The correlation of a variable with itself is referred to as *autocorrelation*. The autocorrelation depends on the relative positions of X_i and $X_i + \Delta h$, both in distance and direction, and on the particular property being considered. The size, shape, orientation, and spatial arrangement of the samples constitute *supports* of the regionalized variable, therefore; the regionalized variable will change if any of the supports change (Davis, 1986). The independence of the property at X_i and $X_i + \Delta h$ beyond a certain distance is simply a particular case of autocorrelation, in which the properties are purely random.

This chapter introduces some basic terms and procedures that can be used to describe the spatial variability of soil parameters. These include trends, scales of fluctuation, and correlation functions/variograms.

5.1 TREND OR DRIFT

The *trend* or *drift* is the expected value of the random variable at a point, $\overline{X}_i$, or computationally, the weighted average of all the points within the neighborhood around point *i* (Davis, 1986). If the drift is subtracted from the regionalized variable, X_i, the residuals, $X_i - \overline{X}_i$, will also be a regionalized variable with a local mean equal to zero. If the residuals are stationary, a semivariogram can be computed. Trends can be modeled as linear, quadratic, or higher-order

functions in one or more dimensions using least-squares techniques. Davis (1986) outlines trend surface models and statistical tests of those types of trends.

5.2 SCALE OF FLUCTUATION

The scale of fluctuation (Fig. 4.1) provides a measure of the estimated distance over which a soil property shows strong correlation. The scale of fluctuation uses the behavior of the normalized variance under successive local averaging or smoothing to describe the distance over which the soil properties show strong correlation. A high value of the scale of fluctuation, δ, indicates the slowly varying nature of the property about the trend (low spatial variability). The steps involved in calculating the scale of fluctuation are as follows:

(1) Calculate the variance for the series of data; this is the reference variance, σ_r;

(2) Smooth the series of data by applying a moving average window of length w and substitute the original data value with the new smoothed value, x_n* (e.g., for a window size, $w = 3$, $x_n^* = (x_{n-1} + x_n + x_{n+1})/w$);

(3) Calculate the variance for the smoothed data: this is the windowed variance, σ_w and will be lower than σ_r due to canceling out of fluctuations due to spatial averaging (Wickremesinghe, 1993);

(4) Normalize the windowed variance by the reference variance and multiply by the window length to obtain: $SOF = (\sigma_w / \sigma_r) * w$;

(5) Repeat Steps 2 – 4 incrementing the width of the window until the smoothing window is greater than about half the length of the data series;

(6) Plot out the SOF as a function of window length;

(7) Observe the behavior of the curve and take the first peak value as an estimate of the correlation length or scale of fluctuation, δ.

As an example, Figure 5.1 presents tip resistance of an electric cone penetration test (CPT) for a site on an alluvial plain where soils consist of interbedded loose sands and soft silts. The data set consists of 498 measurements of tip resistance, qc, digitized at a 0.02 m interval. Note that the data in Figure 5.1 exhibit significant variability, particularly in the top meter of the soil profile. However, there are ranges of depth in which the tip resistance seems to be relatively uniform. To determine the average length of those depth ranges when observations are likely to be correlated, the procedure outlined above is used to calculate the scale of fluctuation. Figure 5.2 shows how the scale of fluctuation varies with the length of the moving average window. The plot peaks at 0.26 m; therefore the scale of fluctuation is taken to be 0.26 m. This estimate of correla-

tion length in the tip resistance corresponds to the average thickness of interbedded layers of the sand and silt.

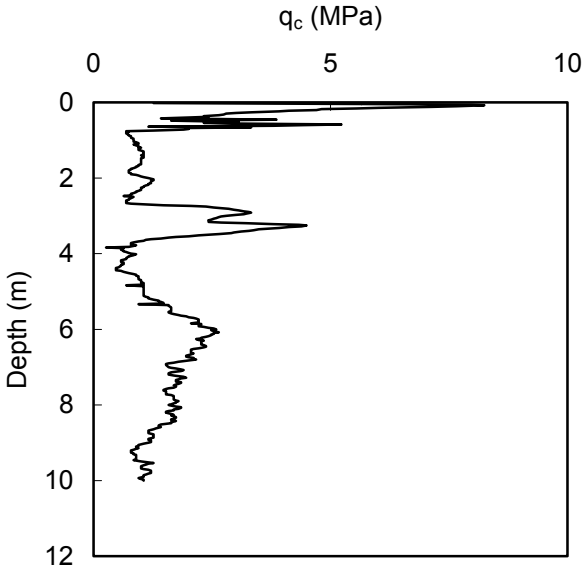


Fig. 5.1 CPT tip resistance log

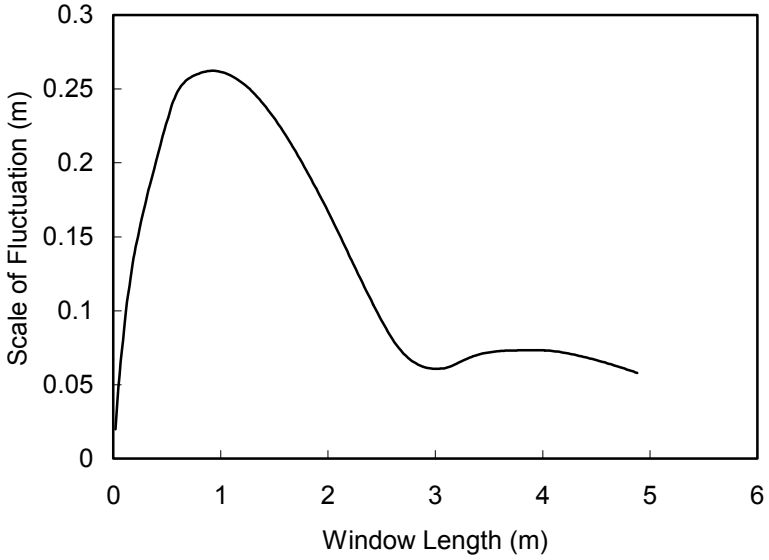


Fig. 5.2 Plot of scale of fluctuation

Tabulated values of scale of fluctuation reported in the literature are presented in Chapter 6.

5.3 CORRELATION FUNCTIONS AND VARIOGRAMS

In spatial analysis, it becomes convenient to use a notation that represents data separated by a predefined distance, called the "separation distance" (also referred to as the *lag* distance), denoted h. Furthermore, it is also convenient to use a notation where data points X_i and X_{i+h} are separated by h intervals. In this notation, X_i is a measurement of the random variable at location i, and X_{i+h} is another measurement taken at h intervals. When h appears in an equation other than a subscript, it represents the numbers of data points separated by the interval. This shorthand notation becomes convenient when presenting functions in future sections of this report.

Correlation of spatial data uses a subset of the *probability space* called the *sample population*. Sampling distributions of the random variable X is also a random variable with a *sample* mean and variance calculated from a *finite* set of data. The sample mean, $E(X)$, and variance, $V(X)$ are given by:

$$E(X) = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ , and} \tag{5.1}$$

$$V(X) = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\sum_{i=1}^{n}(X_i)^2\right) \tag{5.2}$$

where n is the number of samples in the finite probability space. The covariance between samples of population random variables, X and Y, would correspondingly be:

$$\text{cov}(X,Y) = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i Y_i - \frac{1}{n}\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i\right). \tag{5.3}$$

Therefore, the autocovariance of X as a function of the separation distance, Δh, becomes:

$$\text{cov}(X_i, X_{i+h}) = \frac{\left(\sum X_i X_{i+h}\right) - \frac{1}{n-h}\sum X_i \frac{1}{n}\sum X_{i+h}}{n-h-1} \tag{5.4}$$

and is illustrated in Figure 5.3. The autocovariance is calculated between a support or series and itself displaced by a lag of distance. The autocovariance at lag 0 is simply the variance of the random variable. The autocovariance is typically calculated for lags from about 0 to $n/4$.

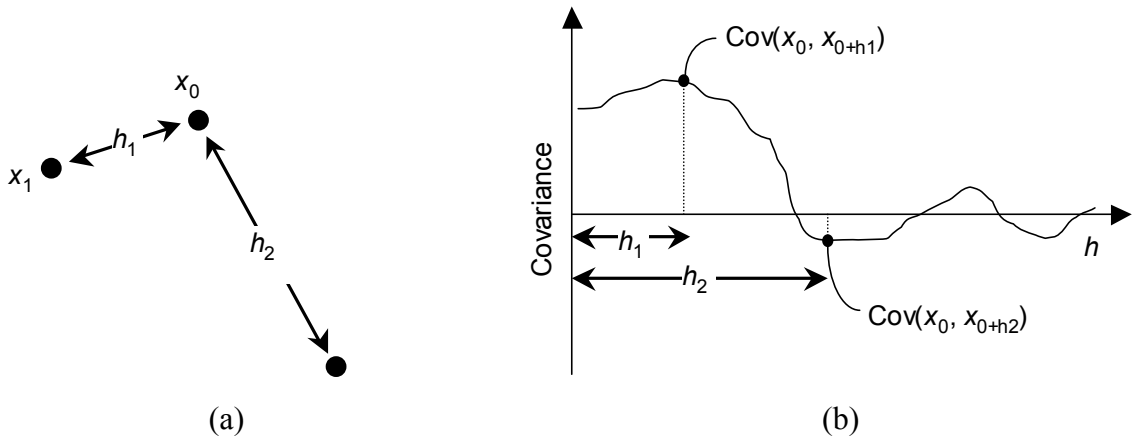


Fig. 5.3 (a) Spatial covariance represented by data points, *x,* and separation distances of *h*; (b) hypothetical data showing that autocorrelation should be higher for low *h* than for high *h*

The autocovariance can be symbolically represented by cov_h, $C(h)$, or σ^2_h. The autocorrelation function is obtained by normalizing the autocovariance by the variance of the variable itself:

$$r_h = \frac{\text{cov}(X_i, X_{i+h})}{V(X)} . \tag{5.5}$$

It should be clear that for $h = 0$, the value of $\text{cov}(X_i, X_{i+h})$ is equal to $V(X)$ and the autocorrelation function has a value of 1.0. As the lag distance increases, the autocorrelation function should be expected to decrease (unless the variable happens to have a constant value), at least for small lag distances. Variables with periodic characteristics will have autocorrelations functions that decrease and increase periodically with lag distance.

5.3.1 Autocorrelation Functions

Autocorrelation functions can be plotted as a function of the lag distance to produce an autocorrelogram. Figures 5.4 through 5.9 show several idealized time series and their corresponding autocorrelograms (adapted from Davis, 1986). Figures 5.4 and 5.5 show a constant and linearly increasing value property, respectively. For a constant function, the autocorrelation function is always equal to 1 because a shift in the function of any lag distance does not change the value of the function. The linearly increasing time series shown in Figure 5.5 will have negative, or linear decreasing, correlation with increasing separation distance. Zero correlation will occur at the

point where the property is zero. Beyond a value of 50, the autocorrelation function would then become negative.

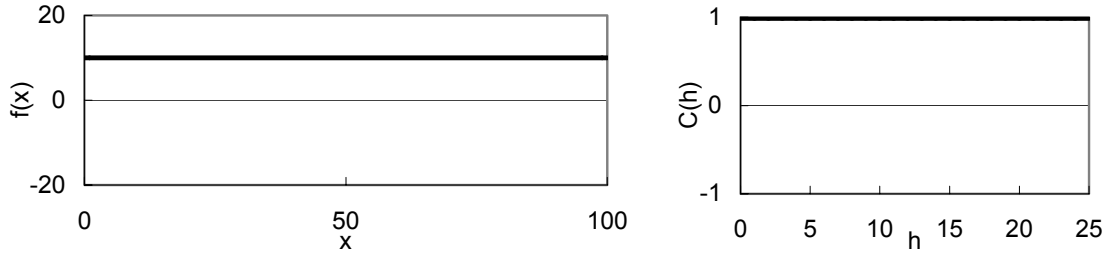


Fig. 5.4 Idealized autocorrelogram showing a constant property variable

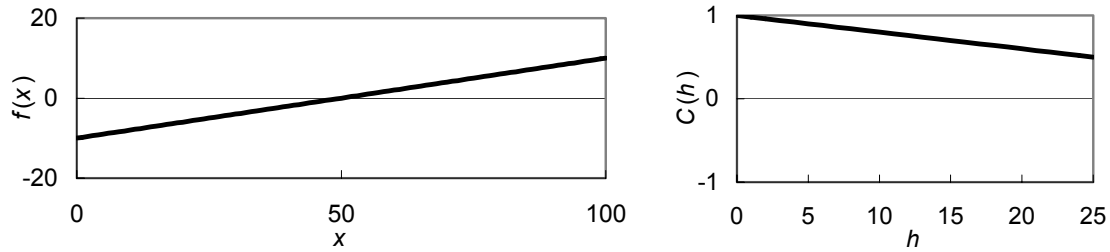


Fig. 5.5 A linear increasing time series with its autocorrelation function

Figure 5.6 shows a sinusoidal time series with a wavelength of 20 units. Therefore, the autocorrelation function will exhibit sinusoidal correlation across the expected wavelength. The autocorrelation function has a value of 1 at zero lag distance, a value of –1 at one-half wavelength (at which point the lagged function is equal to the negative of the original function), and a value of 1 again at a lag distance of one wavelength (at which point the lagged function is identical to the original function).

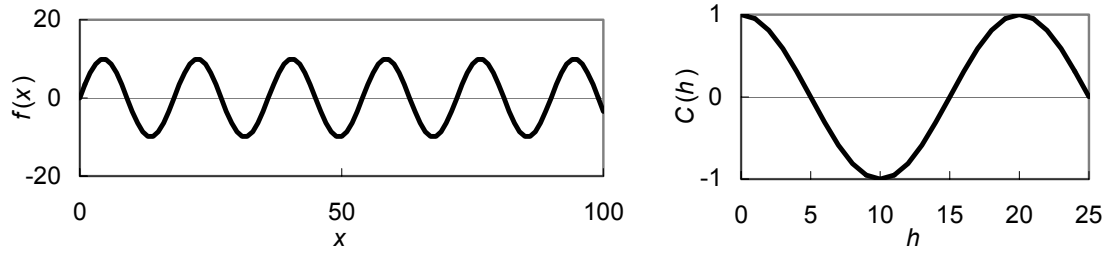


Fig. 5.6 An idealized time series and autocorrelogram consisting of a sine wave with wavelength of 20 units

In reality, perfectly sinusoidal data are rarely encountered. The addition of random noise to a sinusoidal function keeps the function from being strictly periodic, and therefore prevents the autocorrelation function from ever reaching a value of 1 or –1 at non-zero lag distances. Fig-

ure 5.7 shows a sinusoidal function to which random noise has been added. The sinusoidal nature of the original function can be seen, and its effect is still reflected in the autocorrelation function.

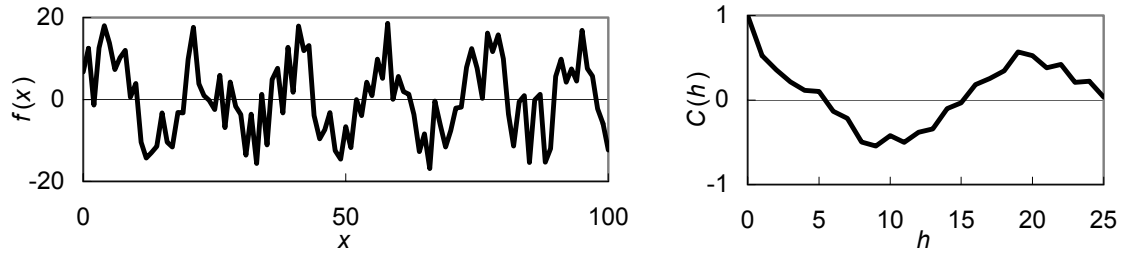


Fig. 5.7 An idealized time series consisting of a sine wave plus random "noise" and its corresponding autocorrelogram

Another function could include a linear trend in addition to sinusoidal and random components. Figure 5.8 shows a time series that has linear, sinusoidal, and random components, and its corresponding autocorrelogram. The autocorrelogram shows more rapid decay in autocorrelation than would have occurred without the linear trend.

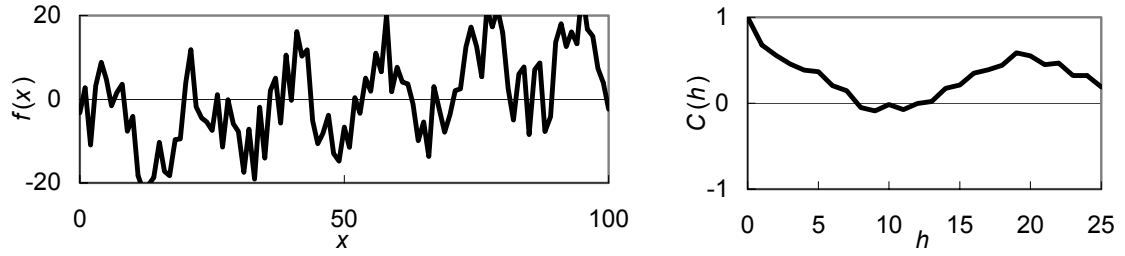


Fig. 5.8 A linear increasing, sinusoidal time series with noise and its corresponding autocorrelogram

A completely random function would be expected to show no autocorrelation. Figure 5.9 shows a sequence of random numbers in the form of a time series that exhibits no autocorrelation. The autocorrelogram can be seen to decay very quickly to near-zero autocorrelation values, even at short lag distances.

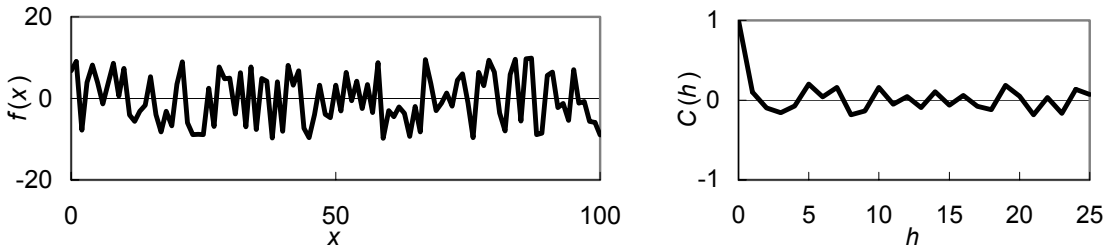


Fig. 5.9 A random times series and corresponding autocorrelogram

Using the example CPT data shown in Figure 5.1, Figure 5.10 presents an autocorrelogram of the tip resistance. The plot begins at a maximum value at zero lag, drops and then rises at a lag of about 3 m. This corresponds to a positive correlation at that distance and is likely the vertical distance between interbedded sand and silt layers in the CPT profile.

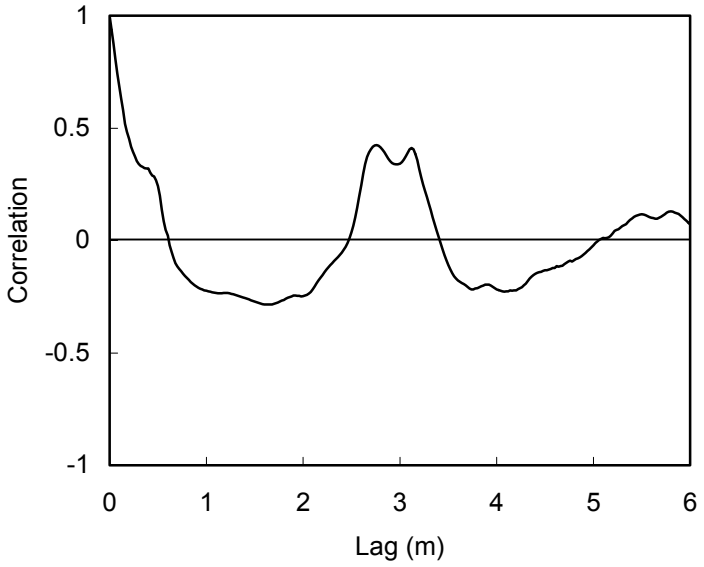


Fig. 5.10 Example autocorrelogram for CPT data shown in Figure 5.1

Many specific standardized expressions are proposed for the autocorrelation function. Four are given in Table 5.1 where each is characterized by a parameter (*a* through *d*, respectively). The relationship between the scale of fluctuation and the autocorrelation function parameters is also given in Table 5.1.

Table 5.1 Autocorrelation functions and corresponding scale of fluctuation (after Vanmarcke, 1977)

Correlation Function	Scale of Fluctuation
$e^{-\lvert h\rvert/a}$	$2a$
$e^{-(h/b)^2}$	$\sqrt{\pi}b$
$e^{-\lvert h\rvert/c}\cos(h/C)$	C
$e^{-\lvert h\rvert/d}\left[1+\frac{\lvert h\rvert}{d}\right]$	$4d$

5.3.2 Variograms

The *variogram* function is a more appropriate way to describe spatial relations, as it is not related to the sample mean or variance, and allows the use of standardized spatial models. The variogram function is a measure of the degree of spatial dependence between samples along a specific support. The variogram function is defined as the variance between data at a particular lag distance, $Z(X_1) - Z(X_2)$:

$$2\gamma(X_i, X_{i+h}) = 2\gamma(h) = E\left\{\left[X_i - X_{i+h}\right]^2\right\} \tag{5.6}$$

with the function $\gamma(h)$ called the *semivariogram* function (due to the 2). The reader should be aware that conflict exists between variogram terminologies found in the literature. While $\gamma(h)$ is defined as the semivariogram function, many authors also refer to it as the "variogram function." This summary will use the term "semivariogram function" to describe $\gamma(h)$. Computationally, the semivariogram is given by:

$$\gamma(h) = \frac{1}{2N(h)}\sum_{i=1}^{N(h)}[X_i - X_{i+h})]^2 \tag{5.7}$$

where $N(h)$ is the number of data pairs separated by h. The contrasting autocovariance function is:

$$C(h) = E\left\{\left[X_i - \mu_{X_i}\right]\left[X_{i+h} - \mu_{X_{i+h}}\right]\right\}. \tag{5.8}$$

Therefore, the relationship between the variogram function and the autocovariance is:

$$\gamma(h) = \sigma_0^{\,2} - C(h) \tag{5.9}$$

where σ_0^2 is the variance at h = 0, or V(X) as discussed above. The relationship between $\gamma(h)$, $C(h)$ and σ_0^2 is shown graphically in Figure 5.11.

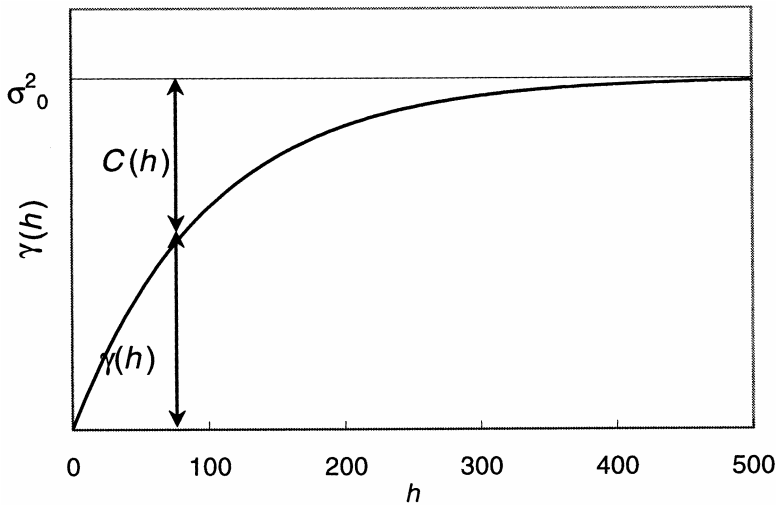


Fig. 5.11 Relationship between variogram and autocovariance

Consider an illustrative simple data set (e.g., SPT values) containing $(N_1)_{60}$ resistance values equally spaced at depth intervals of one meter. To compute the semivariogram, the trend must first be removed and the semivariogram calculated suing the residuals. The original data and trend are shown in Figure 5.12.

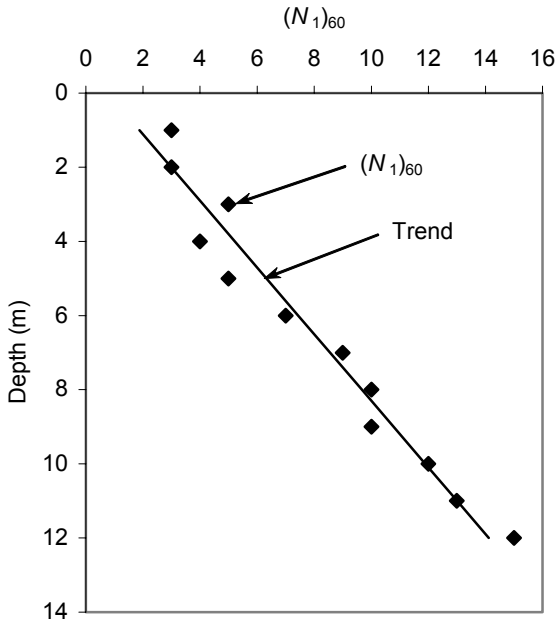


Fig. 5.12 Illustrative SPT data set

There are 12 data points, which produce 11 pairs of points, hence $N(h) = 11$.

$$\gamma(0) = \frac{1}{2(11)}\left[(X_1 - X_1)^2 + (X_2 - X_2)^2 + \ldots + (X_6 - X_6)^2 + (X_{12} - X_{12})^2\right] = 0$$

The corresponding variogram calculation for a lag of 1 m (h = 1) is as follows:

$$\gamma(1) = \frac{1}{2(N)}\left[(X_1 - X_2)^2 + (X_2 - X_3)^2 + \ldots + (X_{10} - X_{11})^2 + (X_{11} - X_{12})^2\right] = 0.496 .$$

Similarly, for a lag of $h = 2$, $N(2) = 10$, and the calculation becomes:

$$\gamma(2) = \frac{1}{2(N)}\left[(X_1 - X_3)^2 + (X_2 - X_4)^2 + \ldots + (X_9 - X_{11})^2 + (X_{10} - X_{12})^2\right] = 0.680$$

where X_i are the residuals after removal of the trend. The plotted semivariogram values are presented in Figure 5.13.

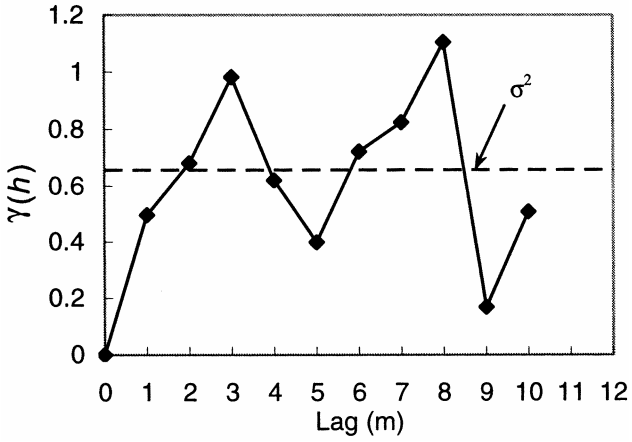


Fig. 5.13 Semivariogram for SPT data shown in Fig. 5.12

Most geotechnical sample data are not uniformly spaced; therefore a method to compute variogram and autocorrelation functions for irregularly spaced data is needed. The most common method is to calculate separation distances (lags) and the value difference $[X_i - X_{i+h}]$ for the calculated lags. The lags and differences are then sorted into ranges. The sample pairs corresponding to a given range are then used to compute the variogram function for that range, using the center of the range for h. For example, if a particular range were considered between lag distances of 3 m and 5 m, then the computed variogram function value would be plotted at 4 m. Ranges of lags are usually referred to as "classes." Other methods use weighting functions (e.g., Journel and Huijbregts, 1978).

5.3.3 Stationarity

Given that regionalized variables are spatially continuous, it is not possible to know their values everywhere. Instead, the values are known through samples at specific locations; therefore, the covariance and variogram functions depend simultaneously on two supports, for example, X_1 and X_2. Furthermore, these functions depend only on the distance between the two supports. This means that for a pair of random variables, the correlation that exists between the data values does not depend on their positions in the support, but on the distance that separates them (Journel and Huijbregts, 1978). To that end, it is important to define the four basic conditions regarding spatial stationarity:

1. Strict Stationarity — A random variable has strict stationarity when the CDF (of the stochastic process) is invariant over all shifts or changes in location. This means that the two random variables (x_i and x_{i+h}) have the same distribution regardless of the distance between them.
2. Second-Order Stationarity — The mean of the process is constant and does not depend on the location, *and* the covariance exists and depends only on the separation distance, h, for each pair of random variables [Z(x), $Z(x+h)$]. Other than assuming the existence of the covariance, an *a priori* variance is also assumed to exist [$V(x) = C(0)$]:

   $$E\{Z(x)\} = \mu \text{, and} \tag{5.10}$$

   $$C(h) = E\{[Z(x) - \mu][Z(x+h) - \mu]\}. \tag{5.11}$$

3. Intrinsic Hypothesis — The mean of the process is constant and does not depend on location *and* the increment of $V(x) - V(x+h)$ has finite variance that does not depend on x. Thus:

   $$V[Z(x) - Z(x+h)] = E\{[Z(x) - Z(x+h)]^2\} = 2\gamma(h). \tag{5.12}$$

   Therefore, second-order stationarity satisfies the intrinsic hypothesis, but the converse is not true.
4. Quasi-stationarity — Quasi- stationarity assumes second-order stationarity of the intrinsic hypothesis to be applied to a local region only, i.e., $|h| \le b$, where b is a limiting distance that represents the neighborhood of estimation or simulation.

5.3.4 Properties of the Variogram Function

Given the basic definitions and assumptions regarding spatial analysis, it is appropriate to summarize the properties of the variogram function. The variogram (as well as the spatial variance) is actually two-sided, with a negative part that is a mirror image of the positive part (resulting from reversal of the lag sequence). In practice, the negative part is not considered in spatial analysis of geotechnical soil properties; however it must be considered when applying these concepts to Fourier transforms of time series.

In general, $\gamma(h)$ increases from its initial value at $\gamma(0)$ as h increases. For continuous random variables, the variogram function levels off and becomes stable about a limiting value called the *sill*, which is generally at, or near, the variance of the stochastic process. The separation distance where the variogram function approaches the sill is called the *range of influence*, designated a or h_r. Where the variogram approaches zero autocorrelation ($\gamma(\infty) = C(0)$) at lags greater than the range of influence is called a "transition phenomenon." The *autocovariance distance*, r_0, is the distance at which the spatial variance has decayed by 1/e (37%). In practice, the initial value of the sill is taken as the sample variance, but as stated earlier, can be greater for actual data. Variograms that have a sill and a range of influence are called "transition variograms" and represent stochastic processes that are not only intrinsic but second-order stationary. This is shown in Figure 5.14 in that the covariance depends only on the separation distance, and that the process has a finite variance that does not depend on the increment between lags.

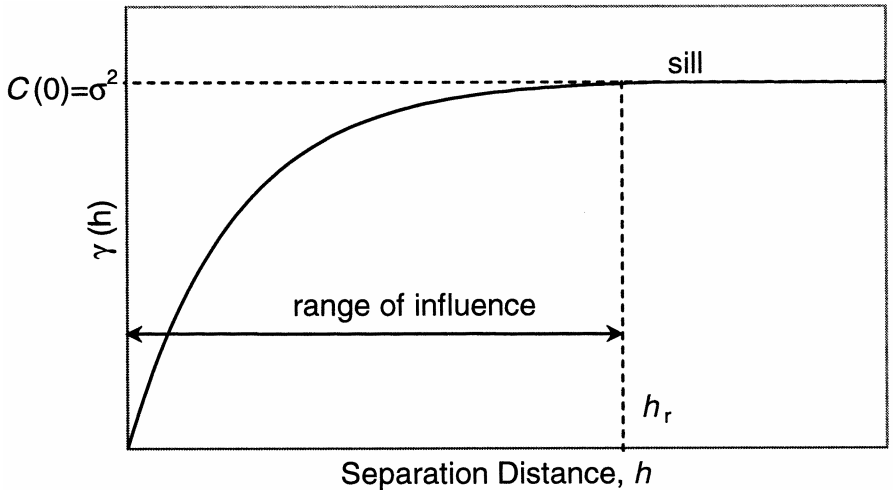


Fig. 5.14 Properties of the variogram function

In some cases, the behavior of the variogram near the origin ($h = 0$) can have a discontinuity. The discontinuity near the origin is called the *nugget effect*, and can result from (a) the variability between two values taken at two very close points, (b) differences in sample values that are not measurable at the scale of data collection, i.e., small-scale effects or (c) be partially attributed to measurement errors. The nugget effect produces an apparent intercept at zero separation distance, i.e., $\gamma(0) = a^2$.

5.3.5 Standard Variogram Models

It is common to use standard variogram models to characterize spatial variability of geotechnical soil properties. Variogram models typically use functions that are positive definite. Commonly used models are categorized into transition models with and without sills. A special class uses models that do not monotonically increase.

5.3.5.1 Transition Models with a Sill

One class of transition models uses relatively simple functions to describe a smooth variation of the variogram function between zero (or a^2 if the nugget effect is present) and the sill. Typical transition models with a sill include simple linear, spherical, exponential, and Gaussian models.

A *simple linear* model increases linearly from zero lag to the sill. Its semivariogram function is given by:

$$\gamma(h) = \begin{cases} a^2 + (\sigma^2 - a^2)\dfrac{h}{h_r}, \; for \; 0 \le h \le h_r \\ \sigma^2, \; h \ge h_r \end{cases} \tag{5.13}$$

where a^2 is the nugget, h_r is the range, and σ^2 is the sill.

A *modified linear* model includes a normal transition at the origin and increases linearly from zero lag to the sill but. Its semivariogram function is given by:

$$\gamma(h) = \begin{cases} a^2\left(1 - e^{-\frac{h^2}{h_{\min}}}\right) + (\sigma^2 - a^2)\dfrac{h}{h_r}, \; for \; 0 \le h \le h_r \\ \sigma^2, \; h \ge h_r \end{cases} \tag{5.14}$$

where h_{min} is the lag at which the transition takes place.

A *spherical* model increases linearly from the origin (zero lag) and includes a normal transition at the range of influence. Its semivariogram function is given by:

$$\gamma(h) = \begin{cases} a^2 + (\sigma^2 - a^2)\left[\frac{3}{2}\frac{h}{h_r} - \frac{1}{2}\frac{h^3}{{h_r}^3}\right], \text{ for } 0 \le h \le h_r \\ \sigma^2, h \ge h_r. \end{cases} \quad (5.15)$$

The corresponding scale of fluctuation is $\frac{3}{4}h_r$.

A *modified spherical* model includes a normal transition at the origin and a normal transition at the range of influence. Its semivariogram function is given by:

$$\gamma(h) = \begin{cases} a^2\left(1 - e^{-\frac{h^2}{h_{\min}}}\right) + (\sigma^2 - a^2)\left[\frac{3}{2}\frac{h}{h_r} - \frac{1}{2}\frac{h^3}{{h_r}^3}\right], \text{ for } 0 \le h \le h_r \\ \sigma^2, h \ge h_r \ . \end{cases} \quad (5.16)$$

An *exponential* model approaches the sill asymptotically. Its semivariogram function is given by:

$$\gamma(h) = \left\{ a^2 + (\sigma^2 - a^2)\left[1 - e^{-\frac{h}{c}}\right], \text{ for } h \ge 0 \right. \quad (5.17)$$

where c is a constant related to an "effective," range as this model never theoretically reaches the sill. For geotechnical purposes, the effective range, h_r' is the lag at which the exponential variogram function reaches 95% of the sill, or where $h = 3c = h_r'$ (given that $1\text{-}e^{-3} = 0.95$). Therefore, $c = h_r'/3$. The corresponding scale of fluctuation is $2h_r$

A *Gaussian* model is parabolic at the origin. Its semivariogram function is given by:

$$\gamma(h) = \left\{ a^2 + (\sigma^2 - a^2)\left[1 - e^{-\frac{h^2}{c^2}}\right], \text{ for } h \ge 0 \right. \quad (5.18)$$

where the effective range is at $h = c\sqrt{3} = h_r'$, therefore $c = h_r'/\sqrt{3}$. The corresponding scale of fluctuation is $\sqrt{\pi}h_r$. These models are shown in graphical form in Figure 5.15.

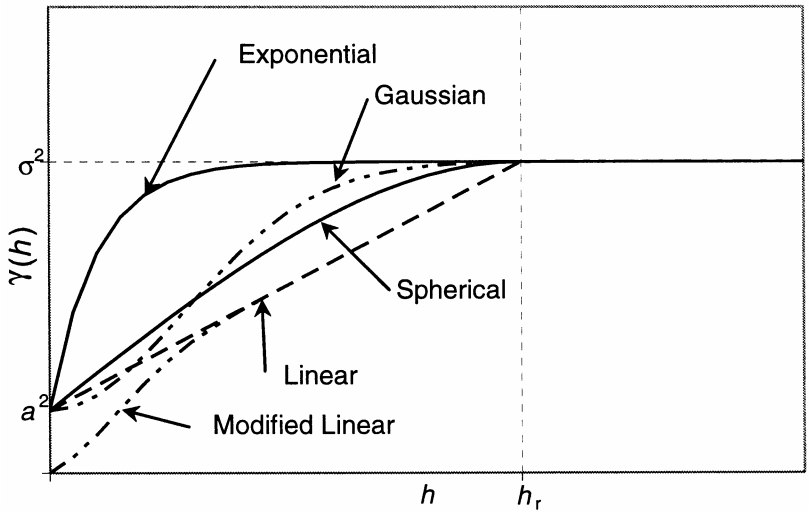


Fig. 5.15 Transition variogram models with sills

5.3.6 Nested Structures

When applied to soil properties, the variability between X_i and X_{i+h} (characterized by the variogram) is due to different causes at different scales:

- at the level of support (e.g., at a separation distance of almost zero), variability is primarily due to measurement (sampling and measurement errors);
- at the small visible level or observable level (e.g., $h < 1$ cm), variability can result from transition from one particle to the next, or from one pore to the next, etc.;
- at the local level (e.g., h <100 m to 500 m), geologic heterogeneities (lithologic changes, lens or anomalies, etc.) produce additional variability;
- at the regional level (e.g., h <1000 km), there are also large-scale variabilities.

From a practical standpoint, all these variabilities are never observed simultaneously, but can certainly occur simultaneously. These scale-dependent sources of variability are contained within the spatial system as nested structures. Nested structures can be represented as the sum of several variograms at different separation distances as shown in Figure 5.16.

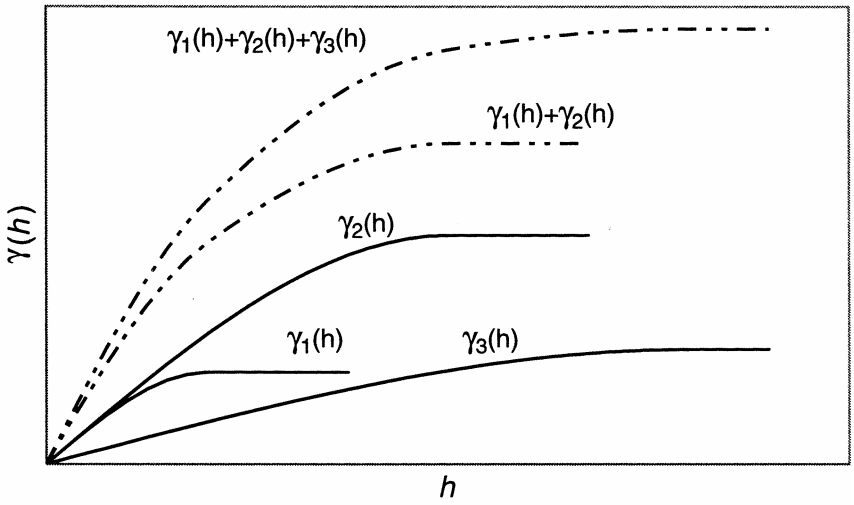


Fig. 5.16 Idealized nested structures

5.3.7 Spatial Anisotropy

A spatial phenomenon is said to be anisotropic if its variability is not the same in every direction. For example, a preferred direction to hydraulic conductivity is commonly observed in soil deposited by sedimentation. Differences in grain size distribution and other properties such as permeability would be more pronounced perpendicular to the bedding planes. Anisotropies can be handled by reducing the isotropy by a linear transformation of the rectangular coordinates in the case of *geometric anisotropy*. Geometric anisotropy is present if *n*-directional variograms can be represented by *n*-transition-type models of the same type with the same sill, with varying ranges. This is similar to the way permeabilities are considered in two dimensions.

A more commonly used method for anisotropic conditions is to treat the anisotropy as a series of zones, called *zonal anisotropy*. In this method, the variable will exhibit different sills in different directions while the range remains constant; therefore a nested structure is used in which each component will have its own anisotropy. A model of zonal anisotropy built from nested structures having their own anisotropies is flexible enough to be used for most types of experimental anisotropy.

Anisotropies can be accounted for by using anisotropic variograms. To detect anisotropies, variograms are typically calculated in different directions and a rose diagram (a diagram of ranges in the different directions) is plotted. The anisotropy ratio is the ratio between the smallest and largest range (these directions are assumed to be approximately perpendicular to

each other and are termed the major and minor axes). An anisotropy ratio of one would denote an isotropic variogram (same variogram for both the major and minor axes).

6 Spatial Variability of Soil Properties

The spatial variability of soil properties is highly dependent on soil type or the method of soil deposition or formation. As a result, good geologic characterization is an important part of any effort toward modeling spatial variability. Measured soil properties can also be influenced by method of measurement, measurement errors, etc. It is important to realize that the values of the parameters used to model spatial variability of geotechnical soil properties will be highly site specific.

Spatial variability in the values of soil properties is a major contributor to spatial analysis in geotechnical engineering. Nevertheless, relatively little quantitative data on the spatial variability of soil properties have been reported. This chapter summarizes values of scale of fluctuation, and the parameters used to define variogram models for various soil properties from sources that have been reported in the literature.

6.1 SCALE OF FLUCTUATION

The scale of fluctuation provides a measure of how rapidly a parameter varies with position about a trend. Several studies have reported scale of fluctuation values for laboratory- and field-measured properties. Table 6.1 presents a tabulated scale of fluctuation data for some laboratory-measured properties.

Table 6.1 Scale of fluctuation values for some laboratory-tested properties (after Phoon and Kulhawy (1999)

Fluctuation Direction	Property	Soil Type	No. of Studies	Scale of Fluctuation (m)	
				Range	Mean
Vertical	s_u (kN/m^2)	Clay	5	0.8 – 6.1	2.5
	w_n (%)	Clay, loam	3	1.6 – 12.7	5.7
	LL (%)	Clay, loam	2	1.6 – 8.7	5.2
	$\bar{\gamma}$ (kN/m^3)	Clay	1	*	1.6
	γ (kN/m^3)	Clay, loam	2	2.4 – 7.9	5.2
Horizontal	w_n (%)	Clay	1	*	170.0
* Not reported.					

Scales of fluctuation have also been computed for insitu test data. Phoon and Kulhawy (1996) suggest the values of scale of fluctuation for SPT, CPT, and VST parameters presented in Table 6.2.

Table 6.2 Scale of fluctuation values for insitu testing

Test Type	Fluctuation Direction	Property	Soil Type	No. of Studies	Scale of Fluctuation (m)	
					Range	Mean
SPT	Vertical	N	*	1	2.4	*
CPT	Vertical	q_c	Sand, Clay	7	0.1 – 2.2	0.9
		q_T	Clay	10	0.2 – 0.5	0.3
	Horizontal	q_c	Sand, Clay	11	3.0 – 80.0	47.9
		q_T	Clay	2	23.0 – 66.0	44.5
VST	Vertical	s_u	Clay	6	2.0 – 6.2	3.8
	Horizontal	s_u	Clay	3	46.0 – 60.0	50.7

The amount of information on the scale of fluctuation is limited in comparison to the COV of inherent variability. Therefore, the actual values of scale of fluctuation used in geotechnical engineering analyses should be carefully chosen. Note that for both laboratory- and field-measured data, the scale of fluctuation in the horizontal direction is much higher than in the vertical direction. Also note that the scale of fluctuation for the CPT is much lower than the scale of fluctuation for SPT, very likely because the sampling interval of the CPT is less than that of the SPT.

6.2 VARIOGRAM MODEL, NUGGET, AND SILL

A number of researchers have investigated the spatial variability of soil properties, and interpreted those data in terms of standard variogram models and identified representative values of the variogram model parameters. Such studies, however, require detailed subsurface investigations with closely spaced data (in at least one direction) in order to accurately resolve the spatial variability. Data of this type are most conveniently (and economically) acquired using CPT; hence the majority of the available data is based on CPT parameters. Table 6.3 provides typical values for the variogram model, nugget, and sill for CPT parameters in different soil conditions.

Table 6.3 Tabulated Values for variogram model, nugget, and sill for various CPT parameters (after Hegazy et al., 1996)

Soil Property	Soil Type	Direction	Variogram Model	Nugget (atm^2)	Sill (atm^2)
CPT Tip Resistance	Sandy fill	Vertical	Spherical	0 – 7	2.8 – 127
	Sandy clay	Vertical	Spherical	0	14 – 2000
	Clayey sand to silty sand	Vertical	Exponential/ Spherical	0	1940 – 3312
	Clays	Vertical	Exponential/ Spherical	0 – 4	0.6 – 21.6
	Sandy fill	Non-directional	Exponential	0.00	70
	Sandy clay	Non-directional	Spherical	0.00	4200
	Clayey sand to silty sand	Non-directional	Spherical	0.00	1700
	Clays	Non-directional	Exponential	4.50	27
CPT Sleeve Friction	Sandy fill	Vertical	Exponential/ Spherical	0.00 – 0.03	0.03 – 0.13
	Sandy clay	Vertical	Exponential/ Spherical	0	0.03 – 0.80
	Clayey sand to silty sand	Vertical	Spherical	0.00 – 0.05	0.13 – 0.83
	Clays	Vertical	Exponential/ Spherical	0	0.00 – 0.26
	Sandy fill	Non-directional	Exponential	0.00	0.12
	Sandy clay	Non-directional	Spherical	0.00	0.85
	Clayey sand to silty sand	Non-directional	Exponential	0.00	0.47
	Clays	Non-directional	Exponential	0.00	0.25
CPT Pore Pressure	Sandy fill	Vertical	Exponential/ Spherical	0	0.00 – 0.12
	Sandy clay	Vertical	Spherical/ Gaussian	0	0.02 – 0.37
	Clayey sand to silty sand	Vertical	Exponential/ Spherical	0	0.03 – 7.17
	Sandy fill	Non-directional	Exponential	0.00	0.05
	Sandy clay	Non-directional	Spherical	0.00	1.16
	Clayey sand to silty sand	Non-directional	Exponential	0.00	0.10

6.3 RANGE/AUTOCOVARIANCE DISTANCE

Data on the distances over which various soil properties remain correlated have also been reported in the literature. By fitting such data to standard variogram models, values of the model parameters that describe the variation of autocorrelation with distance can be determined. Tables 6.4 and 6.5 present typical values of the range and autocovariance distance.

Table 6.4 Tabulated values of range and autocovariance distance for SPT and CPT parameters

Soil Property	Soil Type	Direction	Range (a) or Autocovariance Distance (r_0) (m)*	Note
SPT N Value	Dune sand	Horizontal	$r_0 = 20$	1
	Alluvial sand	Horizontal	$r_0 = 17$	
CPT Resistance	Offshore soils	Horizontal	$r_0 = 30$	2
	Offshore soils	Horizontal	$r_0 = 14$-38	
	Silty clay	Horizontal	$r_0 = 5$-12	
	Clean sand	Vertical	$r_0 = 3$	
	Mexico clay	Vertical	$r_0 = 1$	
	Clay	Vertical	$r_0 = 1$	
	Sensitive Clay	Vertical	$r_0 = 2$	
	Silty Clay	Vertical	$r_0 = 1$	
CPT Tip Resistance	North Sea clay	Horizontal	$r_0 = 30$	1
	Copper tailings	Vertical	$r_0 = 0.5$	
	Clean sand	Vertical	$r_0 = 1.6$	
	North Sea	Horizontal	$r_0 = 14 – 38$	
	Sensitive Clay	Vertical	$a = 2$	
	Sandy fill	Vertical	$a = 0.27 – 0.94$	3
	Sandy clay	Vertical	$a = 0.30 – 1.22$	
	Clayey sand to silty sand	Vertical	$a = 1.83 – 2.90$	
	Clays	Vertical	$a = 0.70 – 2.65$	
	Sandy fill	Non-directional	$a_1 = 1.07$, $a_3 = 0.57$[(a)]	
	Sandy clay	Non-directional	$a_1 = 0.98$, $a_3 = 0.69$[(b)]	
	Clayey sand to silty sand	Non-directional	$a_1 = 3.05$, $a_3 = 2.32$[(c)]	
	Clays	Non-directional	$a_1 = 3.05$, $a_3 = 1.32$[(d)]	
CPT Sleeve Friction	Sandy fill	Vertical	$a = 0.61 – 0.82$	3
	Sandy clay	Vertical	$a = 0.34 – 1.77$	
	Clayey sand to silty sand	Vertical	$a = 1.37 – 3.05$	
	Clays	Vertical	$a = 0.46 – 4.42$	
	Sandy fill	Non-directional	$a_1 = 1.83$, $a_3 = 0.74$[(a)]	
	Sandy clay	Non-directional	$a_1 = 1.22$, $a_3 = 1.20$[(b)]	
	Clayey sand to silty sand	Non-directional	$a_1 = 3.66$, $a_3 = 2.36$[(c)]	
	Clays	Non-directional	$a_1 = 4.57$, $a_3 = 2.39$[(d)]	
	Sensitive Clay	Vertical	$a = 2$	
CPT Pore Pressure	Sandy fill	Vertical	$a = 0.46 – 1.68$	3
	Sandy clay	Vertical	$a = 0.37 – 1.37$	
	Clayey sand to silty sand	Vertical	$a = 2.59 – 3.66$	
	Sandy fill	Non-directional	$a_1 = 1.52$, $a_3 = 1.04$	
	Sandy clay	Non-directional	$a_1 = 1.22$, $a_3 = 0.81$	
	Clayey sand to silty sand	Non-directional	$a_1 = 3.96$, $a_3 = 2.16$	
	Sensitive Clay	Vertical	$a = 2$	1

* a_1 = major range, a_3 = minor range.
(a) 312 points.
(b) 126 points.
(c) 450 points.
(d) 636 points.
Notes:
(1) DeGroot (1996)
(2) Lacasse and Nadim (1996)
(3) Hegazy, et al. (1996)

Table 6.5 Tabulated values of autocovariance distance for additional soil parameters (after DeGroot, 1996)

Soil Property	Soil Type	Direction	Autocovariance Distance (r_0) (m)
DMT P_0	Varved clay	Vertical	$r_0 = 1.0$
FVT Undrained Shear Strength	Clay	Vertical	$r_0 = 1 – 3$
	Sensitive clay	Vertical	$r_0 = 1$
	Sensitive clay	Horizontal	$r_0 = 23$
	Sensitive clay	Vertical	$a = 2$
Laboratory Undrained Shear Strength	Chicago clay	Vertical	$r_0 \sim 0.5$ (Unconfined Compression)
	Offshore sites	Vertical	$r_0 = 0.3 – 3.6$ (Triaxial and DDS)
Hydraulic Conductivity	Salt dome	Horizontal	$r_0 = 1.5$ km
	Compacted clay	Horizontal	$r_0 = 0.5 – 2$
	Sand aquifer	Horizontal	$r_0 = 1 – 2.5$

7 Estimation and Simulation

Previous chapters presented techniques for characterizing the spatial correlation of regionalized variables such as soil properties. In cases where soil property data are available at specific locations, it is frequently desirable to want to know the value of that property at a different location. *Estimation* techniques can be used to determine the desired value. In other cases, discrete data may not be available, but the CDF and autocorrelation function for the data are known or can be assumed. In such cases, *simulation* techniques can be used to generate spatial variable values of the property of interest.

7.1 ESTIMATION

If measurements have been made at random locations and the form of the autocorrelation is known, it is generally possible to estimate the value of the random variable at other locations through a process called *kriging*. Kriging is a weighted, moving average interpolation (or extrapolation) procedure that minimizes the estimated variance of the interpolated (extrapolated) value with the weighted average of its neighbors. The weighting factors and the variance are computed using the variogram model. Since the autocorrelation is related to distance, the weights depend on the spatial arrangement of the locations of interest. Significantly, kriging provides an unbiased estimator of the mean characteristic of the random field.

Local estimation is used to find the best estimator over a limited domain of the regionalized variable, i.e., at small lags compared to the dimensions at which the homogeneous property is measured. Global estimation considers lags larger than the limits of quasi-stationarity and therefore reduces the estimation to computation of general statistics such as a mean and variance.

The goal of kriging is to predict the average value of X within a block, B, over a volume, area, or point. Simple kriging assumes that the covariance is second-order stationary, and that

the mean and variance of the data being considered are constant. The true average value of B is $\mu = \overline{Z}_B = \frac{1}{B}\int_B Z(x)dx$, but an approximation of the true average can be estimated from:

$$\overline{\hat{Z}_B} = \sum_{i=1}^{n} a_i Z(x_i) \tag{7.1}$$

where a_i are weights applied to the respective data values, $Z(x_i)$, such that $\sum_{i=1}^{n} a_i = 1.0$. The best linear estimate will minimize the expected squared error between the true and approximated average value of the block. The expected squared error, Q, becomes:

$$Q = E\left\{\left[\overline{Z_B} - \overline{\hat{Z}_B}\right]^2\right\} \tag{7.2}$$

noting that:

$$E\left(\overline{\hat{Z}_B}\right) = \sum_{i=1}^{n} a_i E\left[Z(x_i)\right]. \tag{7.3}$$

For an unbiased estimator:

$$E\left(\overline{\hat{Z}_B}\right) = E\left[Z(x_i)\right] \tag{7.4}$$

therefore $\sum_{i=1}^{n} a_i = 1.0$. Therefore, minimizing the squared error can be accomplished by finding a set of a_i values for which $\frac{\partial Q}{\partial a_i} = 0$. By using the method of Lagrangian multipliers, the solution to this partial differential equation can be obtained from the following set of simultaneous equations:

$$\sum_{j=1}^{n} a_j C_{ij} + \lambda = C_{Bi}, \text{ and} \tag{7.5}$$

$$\sum_{i=1}^{n} a_i = 1.0 \tag{7.6}$$

where $C_{ij} = C(h)$, the autocovariance between the data values, λ is a Lagrangian multiplier, $C_{Bi} = C(h)$ between the block B and all data values, and a_i are the kriging weights. In matrix form, the equations are:

$$\underbrace{\begin{bmatrix} C_{11} & C_{12} & C_{13} & \cdots & C_{1n} & 1 \\ C_{21} & C_{22} & C_{23} & \cdots & C_{2n} & 1 \\ \vdots & & & & & \\ C_{n1} & C_{n2} & C_{n3} & \cdots & C_{nn} & 1 \\ 1 & 1 & 1 & \cdots & 1 & 0 \end{bmatrix}}_{C} \underbrace{\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \\ \lambda \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} C_{B1} \\ C_{B2} \\ \vdots \\ C_{Bn} \\ 1 \end{bmatrix}}_{C_{Bi}} \tag{7.7}$$

Therefore, the solution, which yields the weighting factors, a_i, is given by:

$$\hat{A} = C^{-1} C_{Bi} \,. \tag{7.8}$$

The kriging variance, an estimate of the estimation variance, can be obtained as:

$$\sigma_K^2 = \sigma_B^2 - \sum_{i=1}^{n} \hat{a}_i C_{Bi} - \hat{\lambda} \tag{7.9}$$

or in matrix form:

$$\sigma_K^2 = \sigma_B^2 - \left[\hat{A}^T \right] C_{Bi} \,. \tag{7.10}$$

The information needed as input to kriging estimation includes the sample values and their spatial coordinates, locations of blocks or points to be estimated, and an estimated variogram function. Computationally, the process consists of the following steps:

(1) Entry in the data file of data points and selection of samples that influence the block of interest,

(2) Computation of covariances between the selected samples,

(3) Computation of covariances between the selected samples and the estimation point,

(4) Assembly of the kriging equations,

(5) Solution of the kriging equations to obtain the weights,

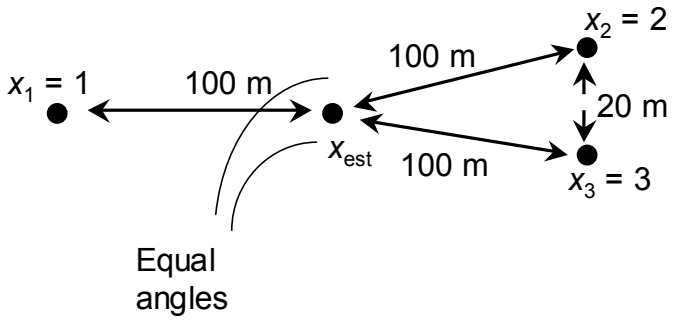
(6) Computation of estimated values using computed weights, and

(7) Calculation of kriging variance.

Many forms of kriging exist, and more detained explanations may be found in Journel and Huijbregts (1978), Davis (1986), and Carr (1995). The simplest form has been presented here for illustration. Figure 7.1 shows an example of kriging where the known data are at equal

distances from the location of where estimation is desired. In this example, all known points are separated from x_i by 100 m. However, in Cases (a) and (b), points 2 and 3 are separated by 20 m and 174 m, respectively. The use of kriging discriminates the spatial structure of the known data resulting in an estimated value, x_{est}, of 1.767 for Case (a) and a value of 2.000 for Case (b).

Case (a)



Case (b)

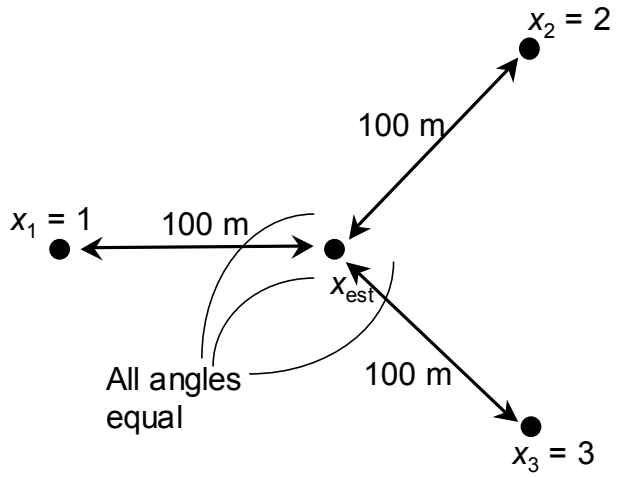


Fig. 7.1 Hypothetical kriging Cases (a) and (b)

7.2 SIMULATION

A regionalized variable, $Z(X)$, can be interpreted as one realization of a random function Z(x) that can be characterized by a distribution function and variogram model. The goal of simulation is to create other realizations of the random function and retain only those realizations that meet the requirements set forth, such as the experimental values in a known data set or a pdf. Two simulation techniques are discussed in detail in this section: random fields and Monte Carlo methods.

It is important to understand the difference between estimation and simulation. Unbiased estimators and properly simulated data should have the same values at the experimental data locations and/or should also have the same or specified dispersion (variability in a particular direction) characteristics. The goal of estimation is to provide an unbiased estimate of the actual data with minimal variance across the support. However, this does not mean that the estimators will reproduce the true variability of the property. This is why estimation typically has a smoothing effect on the real data. Conversely, the goal of simulation is to produce a set of data with the same two moments (mean and covariance) and spatial variation (variogram) as the known distribution.

7.2.1 Random Fields

Simulation of random fields can be useful to quantify uncertainty in probabilistic terms. This simulation technique generates a number of realizations of random fields that describes the data at hand. The data from each realization can be used to solve the deterministic problem of interest. From the set of solutions obtained, inferences of the design problem can then be made in a probabilistic manner (Hasofer, 1993). The optimum design can then be made by ascertaining probabilities or by balancing safety or cost. Simulation of random fields is a broad subject in and of itself. The materials presented in this section are a brief summary of the work by Vanmarcke (1983), Yamazaki and Shinozuka (1988), and Hasofer (1993).

7.2.1.1 Gaussian versus Non-Gaussian Fields

The assumption of Gaussian behavior greatly simplifies stochastic problems. Most simulation methods produce fields which are either approximately or exactly Gaussian. However, the distributions of many geotechnical soil properties may be far from Gaussian. To determine whether or not to use Gaussian or non-Gaussian random fields, one must look at the specified probabilistic structure of the random field. A Gaussian field requires only that the mean and covariance function at each pair of points be specified. Conversely, for a general, non-Gaussian field, a full specification requires that all *n*-dimensional joint distribution functions (a function that defines how all distributions are dependent on each other) at *n* arbitrary points in the field be given. This is generally impractical so portions of the distributions of the field are adjusted to conform to the required specification.

7.2.1.2 Methods of Random Field Generation

The two most common types of random field simulation are (a) discretization in the time domain and (b) discretization in the spatial domain. Because geotechnical soil properties are inherently spatial, the following discussion will be limited to the spatial domain and, further, to two spatial dimensions. Spatial models are extensions of one-dimensional, time series models and will yield discrete realizations.

The simplest model is the moving average (MA) model given by the formula:

$$X_{ij} = \sum_{r=-n}^{n} \sum_{s=-m}^{m} b_{rs} U_{i+r,j+s} \tag{7.11}$$

where *n* and *m* are the field dimensions, *r* and *s* are dimensions of the moving average, and $U_{i,j}$ are independent standard normal variables. The optimum choice for the weighting function, b_{rs}, is given by:

$$b_{rs} = \frac{1}{4\pi^2} \int_{-\pi}^{+\pi} \int_{-\pi}^{+\pi} \sqrt{S_D} \exp\{i(r\lambda_1 + s\lambda_2)\} d\lambda_1 d\lambda_2 \tag{7.12}$$

where λ_1 and λ_2 are wavelengths, $S_D(\lambda_1, \lambda_2) = \sum_r \sum_s R_{rs} \exp\{i(r\lambda_1 + s\lambda_2)\}$ is the spectral density function of the discrete field X_{ij}, and $R_{rs} = E\lfloor X_{i+r,j+s} X_{ij} \rfloor$. The numbers *n* and *m* are chosen so that the b_{rs} values are negligible outside the summation range. Autoregressive (AR) models are also used, as well as combined autoregressive, moving average (ARMA) models.

Generation of non-Gaussian, random fields is accomplished by generating a Gaussian field and then transforming it into a non-Gaussian field by a mapping technique on the probability distribution function. However, the spectral density of the mapped field does not match the target spectral density because the transformation is nonlinear. Therefore, an iterative algorithm is required to match the target spectral density. Such an iterative method is outlined in Yamazaki and Shinozuka (1988).

7.2.2 Monte Carlo Methods

Simulation by Monte Carlo methods (MCM) solves problems by generating suitable random numbers (or pseudo-random numbers) and observing that fraction of the numbers that obey some property or properties. The methods are useful for obtaining numerical solutions to problems that are too complicated to solve analytically. Given a function $Y = g\,(Z(x_1, x_2, \ldots\ldots x_n))$, where $x_1, x_2, \ldots x_n$ are random variables with known statistics, the MCM can be used to find the statistics of Y. The MCM is a simple but versatile computational procedure that is well suited for numerical analysis. In general, the implementation of the method (Yang, et al., 1993) involves:

- Selection of a model that will produce a deterministic solution to a problem of interest.
- Decisions regarding which input parameters are to be modeled probabilistically and the representation of their variabilities in terms of probability distributions.
- Repeated estimation of input parameters that fit the appropriate probability distributions and are consistent with the known or estimated correlation between input parameters.
- Repeated determination of output using the deterministic model.
- Determination of the pdf of the computed output, Y. The pdf may be approximated by a histogram from which useful statistics of Y can be computed.

For example, in probabilistic slope stability analyses, a critical slip surface can first be determined based on the mean value of the input parameters using appropriate deterministic (limit equilibrium or finite element) analyses. Probabilistic analysis can then be performed, taking into consideration the variability of the input parameters on the critical slip surface. In most cases, a distribution of the variability of the soil input properties is assumed with user-specified parameters, such as mean values and standard deviations. During each Monte Carlo trial, the input parameters are updated based on a normalized random number. The factors of safety are

then recomputed based on the updated soil input parameters. The result is a mean and standard deviation of the factor of safety. The probability distribution function is then obtained from the normal curve. The number of Monte Carlo trials in an analysis of this type will be dependent on the number of variable input parameters and the expected probability of failure. In general, the number of required trials increases as the number of variable input increases or the expected probability of failure becomes smaller. It is not unusual to do thousands of trials in order to achieve an acceptable level of confidence in a Monte Carlo probabilistic slope stability analysis (Mostyn and Li, 1993).

7.2.2.1 Random Number Generation

Fundamental to the Monte Carlo method are the randomly generated input parameters that are used in the deterministic model. In general, this is accomplished using a random number generation function, with random numbers generated during each analysis run. To ensure that new random numbers are generated for each run, the random number function is "seeded" with the current time of a computer clock.

The random numbers generated from the function are *uniformly distributed* with values between 0.0 and 1.0. In order to use the generated random number in calculations of other distribution, the random number needs to be parameterized to the distribution of interest. For example, to generate a normally distributed random number, the uniform random number is transformed to a normally distributed one. For the normal distribution, the normalization process is done using the transformation equation suggested by Box and Muller, (1958):

$$N = \sqrt{(-2\ln r_1)}\sin(2\pi r_2) \tag{7.13}$$

where N = normalized random number, r_1= uniform random number 1, and r_2= uniform random number 2.

The transformation equation requires the generation of two uniform random numbers to generate a parameterized random number with a standard normal distribution (mean value of 0 and standard derivation of 1 as discussed in Section 3.6.2).

7.2.2.2 Estimation of Input Parameters

At the beginning of each Monte Carlo trial, all uncertain input parameters are re-evaluated based on the specified distribution. For ease of computation, the mean value, μ, the standard deviation, σ, and the normalized random number, N, are used. The general equation for updating the parameters is simply:

$$P = \mu + N\sigma \tag{7.14}$$

where P is the new trial value of any of the parameters specified with a standard deviation assuming a normal distribution. For example, consider undrained shear strength, s_u, with a specified mean value of 2200 psf and a standard deviation of 250 psf. In a particular Monte Carlo trial, if the normalized random number is –2.0, the trial s_u will be 1700 psf.

It is important to consider the parameter of interest relative to common probability distributions. For example, Lumb (1966) has shown that the tangent of the friction angle (tanϕ) conforms better than the friction angle itself to the normal probability distribution function. Therefore, it may be appropriate to use the tangent of the friction angles in the estimation of all trial friction angles.

7.2.2.3 Adjustments for Correlation

When soil properties are not independent, the normalized random number should be adjusted to consider the effect of correlation. The following equation is used in the adjustment for two correlated properties (Lumb, 1970):

$$N_a = N_1\rho + \left(1 - |\rho|\right)N_2 \tag{7.15}$$

where N_a = adjusted parameterized random number for the second property, ρ = correlation coefficient between the first and second properties, N_1 = normalized random number for the first property (assuming independence), and N_2 = normalized random number for the second property (assuming independence). For the adjustment of the more than two correlated properties, the covariances from the covariance matrix can be used for the adjustment.

7.2.2.4 Statistical Analysis

Assuming that the trial model is normally distributed, statistical analysis can be conducted to determine the mean, standard deviation, probability density function, and the cumulative distribution function of the model output. The equations used in the statistical analysis are summarized as follows (Lapin, 1983):

$$\text{Mean:}\quad \mu = \frac{1}{n}\sum_{i=0}^{n} Y_i \tag{7.16}$$

$$\text{Standard deviation:}\quad \sigma = \frac{1}{n}\sum_{i=o}^{n}\left(Y_i - \mu\right)^2 \tag{7.17}$$

$$\text{Probability density function:}\quad f\left(F\right) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(Y-\mu)^2}{2\sigma^2}} \quad -\infty < F < \infty \tag{7.18}$$

$$\text{Cumulative distribution function:}\quad F\left(F\right) = P\left[X \le Y\right] = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{Y} e^{-\frac{(Y-\mu)^2}{2\sigma^2}} dy \tag{7.19}$$

where Y_i = the output from the trial model, N = number of trial runs, and F = output parameter of interest. Note that these equations are derived from those presented in Chapter 3.

7.2.2.5 Number of Monte Carlo Trials

Probabilistic analysis using the MCM involves many trial runs. The more trial runs used in an analysis, the more accurate the statistics will be. The number of required Monte Carlo trials is dependent on the desired level of confidence in the solution as well as the number of variables being considered (Harr, 1987), and can be estimated from:

$$N_{mc} = \left[\frac{d^2}{4\left(1-\varepsilon\right)^2}\right]^m \tag{7.20}$$

where N_{mc} = number of Monte Carlo trials, d = the standard normal deviate corresponding to the level of confidence, ε = the desired level of confidence (0 to 100%) expressed in decimal form; and m = number of variables. This relationship is shown graphically in Figure 7.2.

The number of Monte Carlo trials increases geometrically with the level of confidence and the number of variables. For example, if the desired level of confidence is 90%, the normal standard deviate will be 1.64, the number of Monte Carlo trials will be 67 for one variable, 4,521

for two variables, and 304,007 for three variables. Theoretically, for a 100% level of confidence, an infinite number of trials would be required.

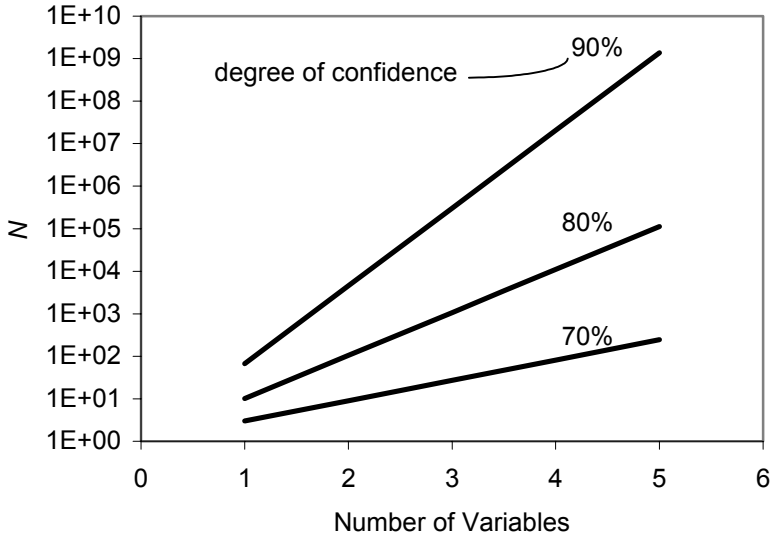


Fig. 7.2 Number of required Monte Carlo trials

For practical purposes, the number of Monte Carlo trials is usually in the order of thousands. This may not correspond to a high level of confidence when multiple variables are being considered; however, the statistics computed from the MCM are typically not very sensitive to the number of trials after a few thousands trials.

7.2.2.6 Reducing the Number of Monte Carlo Trials

For geotechnical earthquake engineering in which a nonlinear, dynamic finite element analysis may be required for each realization, minimizing the number of Monte Carlo trials can be critically important. Techniques such as the Latin hypercube and importance sampling can be used to reduce, in some cases substantially, the number of Monte Carlo trials.

Latin Hypercube

Simulation can be used to study how the outcome of a model, *g(Z)*, depends on the distribution of a random variable, Z, with known distribution, F_Z. This can be done using simple random sampling (Monte Carlo simulation) by randomly drawing values from F_Z. However, if the function *g(Z)* is computationally expensive to evaluate, other sampling strategies may be more efficient. One method is the use of stratified sampling. This method can characterize the population equally well as simple random sampling with a smaller sample size. Stratified sampling works as follows:

(1) The distribution of Z is divided into m segments;

(2) The distribution of n samples over these segments is proportional to the probabilities of Z falling in the segments; and

(3) Each sample is drawn from within its segment by simple random sampling.

Maximum stratification takes place when the number of segments (strata), m, equals the number of data, n, and when Z has probability m^{-1} of falling in each segment, as this is the most efficient.

When the model depends on two variables, $h(Z_1, Z_2)$, then values of both Z_1 and Z_2 can be drawn by simple random sampling or by stratified random sampling. The most efficient sample is maximally stratified for both Z_1 and Z_2 simultaneously. This means that a sample of size n of pairs (z_1, z_2) is marginally stratified for both Z_1 and Z_2. Such a sample is called a Latin hypercube sample and is outlined in McKay et al. (1979). The model need not depend only two variables only.

For the spatially distributed case, multiple Gaussian simulations can be created. The simulations are independent in the sense that they are a random sample from the ensemble of all realizations that were possible under the specified model (Pebesma, 2001). At a specific location, x_0, the subsequently realized values of the variable $Z(x_0)$, $(z^1(x_0),\ldots,z^n(x_0))$, are a simple random sample from the distribution of $Z(x_0)$. When a stratified sample of $Z(x_0)$ is desired, then this could be drawn when $F_{Z(x_o)}$ was known. However, this is less trivial when $Z^i(x_0)$ and $Z^i(x')$ have to obey the specified spatial correlation. A procedure for obtaining a Latin hypercube sample in this context (multiple, spatially correlated variables) is presented by Stein (1987).

Importance Sampling

Monte Carlo calculations can be carried out using sets of random points sampled from any arbitrary probability distribution. The choice of distribution obviously makes a difference to the efficiency of the method. In most cases, Monte Carlo calculations carried out using uniform probability distributions give very poor estimates of high-dimensional integrals and are not a useful method of approximation. Metropolis et. al. (1953) introduced a new algorithm for sampling points from a given probability function. This algorithm enabled the incorporation of "importance sampling" into Monte Carlo integration. Instead of choosing points from a uniform distribution, they are now chosen from a distribution which concentrates the points where the function being integrated is large. Sampling from a non-uniform distribution for this function should therefore be more efficient than doing a crude Monte Carlo calculation without importance sampling.

7.2.3 Other Methods

The turning bands method (TBM) is based on the theory of random fields. The basic concept of TBM is to transform a multidimensional simulation into the sum of a series of equivalent unidimensional simulations. The purpose is to preserve the statistics of the true field, particularly the variogram of the stationary random field. In the TBM, the simulated field is assumed to have second-order stationarity and to be isotropic. Techniques are available to transform the uniform or normal distribution at each point to other distributions. The basic TBM uses realizations of the random function generated on lines rotating in space to produce a three-dimensional realization.

A more advanced approach has been recently used, called the **"**Intrinsic Random Function of Order *k"* (IRF-*k*), to accommodate the varying nature of the trend in a regionalized soil variable (McBratney et al., 1991). The term *k* represents the order of polynomial trends where *k* = 0 means constant drift, and the IRF-*k* is equivalent to the ordinary kriging system of equations. Therefore, if *k* = 1, linear drift exists; *k* = 2 yields quadratic drift. But when the deterministic relationships are with some known or readily available and inexpensive covariates or other easy-to-measure soil variables, co-kriging has played a major role in efficiently predicting the target soil variable (Odeh et al., 1995).

7.3 AVAILABLE TOOLS

There are many currently existing software tools to complete many types of the statistical and spatial analyses presented in this report. This chapter provides several sources of open source software that exist in the public domain.

7.3.1 StatLib Website http://lib.stat.cmu.edu

StatLib is a WWW clearinghouse for distributing statistical software, datasets, and information by electronic mail, FTP, and WWW hosted by the statistics department at Carnegie Mellon University. Software can be downloaded in Fortran, C++, Matlab, etc. The site can also be used to search for statistical functions at other sites.

7.3.2 GSLIB Website http://www.gslib.com

GSLIB is an acronym for Geostatistical Software LIBrary. The GSLIB website serves to point researchers and practitioners to the public domain GSLIB programs for geostatistical problem solving. It also informs users of commercial supplements to GSLIB and announces training and support.

7.3.3 AI-GEOSTATS Website http://www.ai-geostats.org

AI-GEOSTATS is the central server for GIS and spatial statistics on the Internet administered by the University of Lausanne in Switzerland. The purpose of the site is to provide information on available open source and commercial software commonly used in spatial data analysis and spatial statistics. The site also administers a discussion mailing list.

7.3.4 Gstat Software http://www.gstat.org/

"Gstat" is a computer program for geostatistical modeling, prediction, and simulation. Gstat works on many platforms and is open source software (Pebesma and Wesseling, 1998). The

domain, **www.gstat.org,** is hosted by the Department of Geography at Utrecht University, the Netherlands. The software can be downloaded for free.

7.4 EXAMPLES

PEER's PBEE framework accounts for uncertainties at all stages of the design/evaluation process. These uncertainties include those associated with earthquake ground motions, with the response of soils and foundations to ground motions, with the response of structures and their contents to soil and foundation motions, and with the assignment of economic losses to various levels of structural and nonstructural performance.

A number of sophisticated deterministic procedures for analysis of soil and soil-structure systems have been developed by PEER researchers. Implementation of these models within the PEER framework requires an understanding of their performance, and of the variability of the responses they predict when used with uncertain soil parameters. The various procedures described in the earlier chapters of this report will allow PEER researchers to study the influence of uncertain and spatially variable soil parameters on performance.

These examples illustrate how deterministic models can be used in conjunction with simulation models to investigate the uncertainty in output associated with uncertain and spatially variable soil properties.

7.4.1 One-Dimensional Lateral Spreading Analysis

The performance of structures during earthquakes is strongly influenced by the seismic response of the underlying soil. Performance evaluations, therefore, should account for the spatial variability of the soil. In many cases, site response can be reasonably modeled as a one-dimensional problem, i.e., as a series of horizontal soil layers subjected to vertically propagating shear waves. To model the important problem of lateral spreading, initial (static) shear stresses can be imposed on the soil profile. Such shear stresses will, under reasonably symmetric ground motions, cause permanent deformations to develop in liquefied soils.

7.4.1.1 Soil Profile

This example is based on a site along the waterfront of Seattle, Washington. Soils at the site consist of loose natural and hydraulically placed fill deposits to a depth of 9 m, with static groundwater at about 3 m in depth (Fig. 7.3). A sloping ground surface of 3% was used to impose initial shear stresses. Prior SPT explorations showed a mean $(N_1)_{60}$ of 10 and a *COV* of 40%.

7.4.1.2 Method of Analysis

One-dimensional site response analyses were performed using the computer program WAVE, a one-dimensional, nonlinear, effective stress-based site response analysis program. Originally developed by Horne (1996), it has been extended with a new constitutive model, *UWsand*, which captures important elements of liquefiable sand behavior in a manner that can easily be calibrated.

7.4.1.3 Deterministic Analysis

In typical engineering practice, it is common to perform analyses deterministically using average soil properties. For this example, a deterministic site response analysis, performed using the mean value of $(N_1)_{60}$, produced the permanent displacement profile shown in Figure 7.3. The deterministic analysis predicts that permanent displacements of slightly more than 30 cm would develop in the loose, saturated sand above the till. Most of this displacement results from straining of the lower half of the loose, saturated sand.

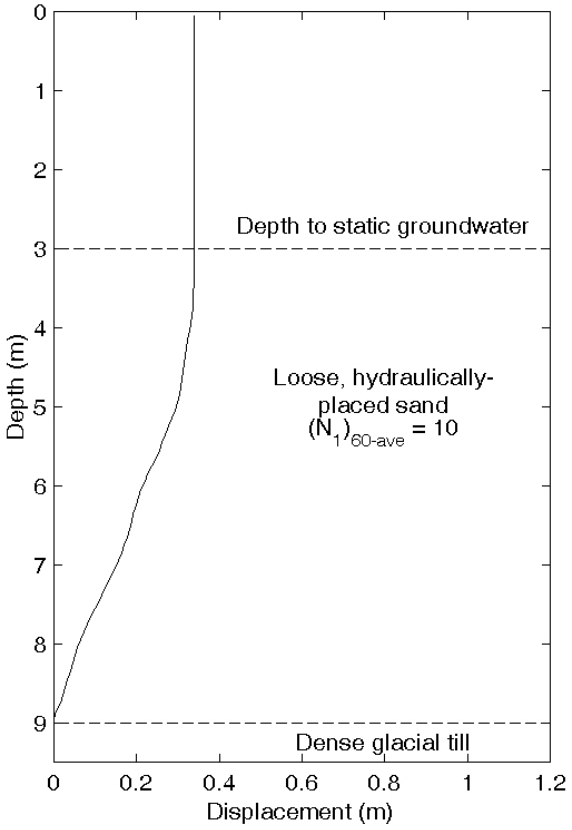


Fig. 7.3 Displacement profile of example soil profile

7.4.1.4 Simulation

Based on the statistics determined from the available field data, 100 one-dimensional Gaussian random fields of SPT profiles were generated similarly to Section 7.2.1 and outlined in Yamazaki (1988). A mean of 10, COV of 40%, and a Gaussian spectral density function (standard deviation of 1 and a correlation decay coefficient equal to 1 for the spectral density function) were used to generate the profiles. These profiles are shown in Figure 7.4. A histogram plot of the simulated $(N_1)_{60}$ values shown in Figure 7.5 (a) confirms that the SPT resistances were normally distributed. A plot of an individual variogram for a particular profile in Figure 7.5(b) shows the spatial distribution of $(N_1)_{60}$.

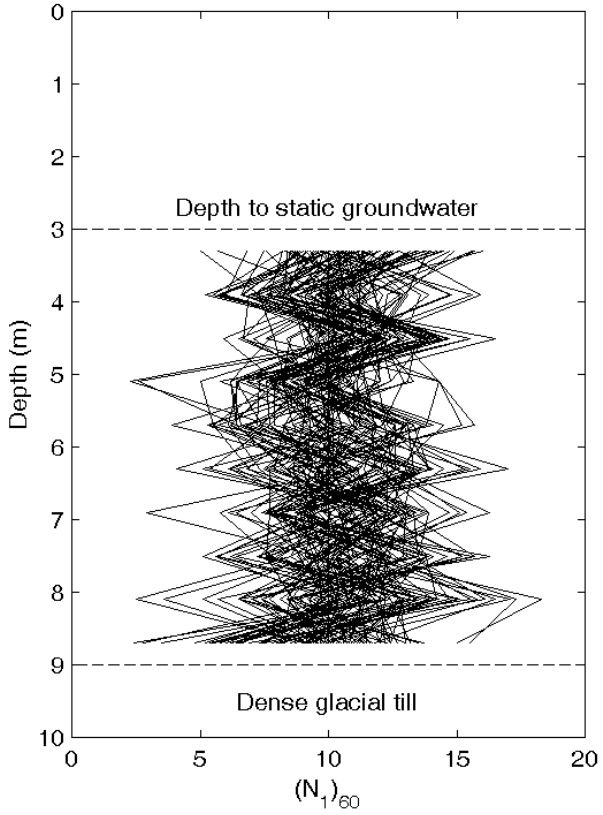
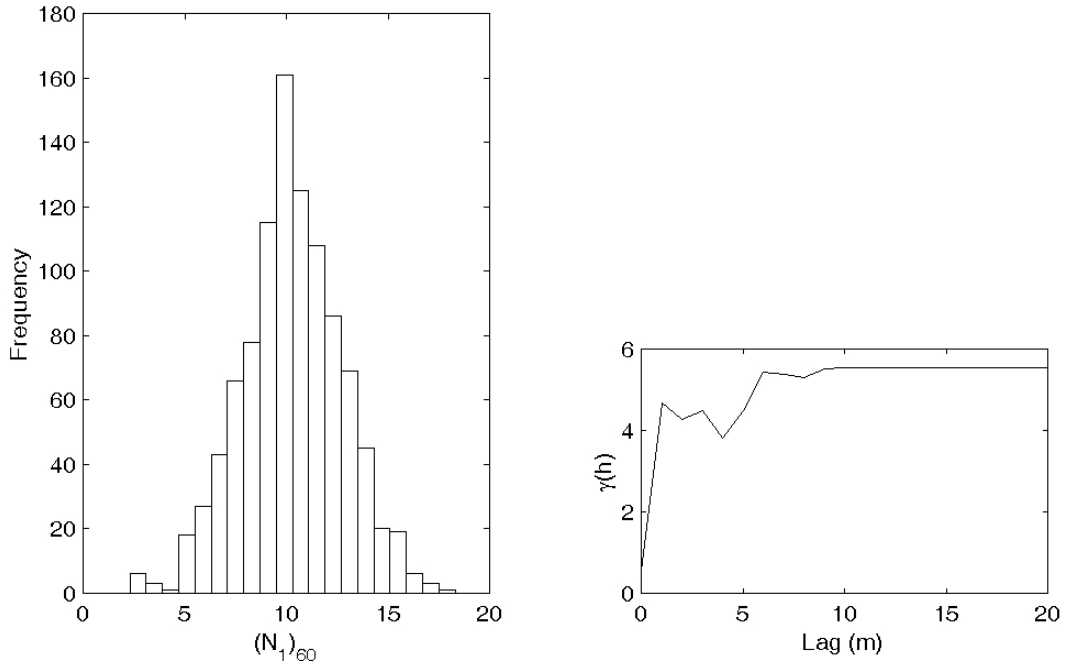


Fig. 7.4 Variation of $(N_1)_{60}$ with depth for 100 simulated soil profiles



(a) **(b)**

Fig. 7.5 (a) Histogram of all $(N_1)_{60}$ data and (b) variogram of individual $(N_1)_{60}$ profile

7.4.1.5 Monte Carlo Analyses

The 100 soil profiles were then analyzed using WAVE. These analyses produced 100 lateral ground displacement profiles for end-of-shaking conditions. Figure 7.6(a) shows the permanent displacements predicted by all of the site response analyses.

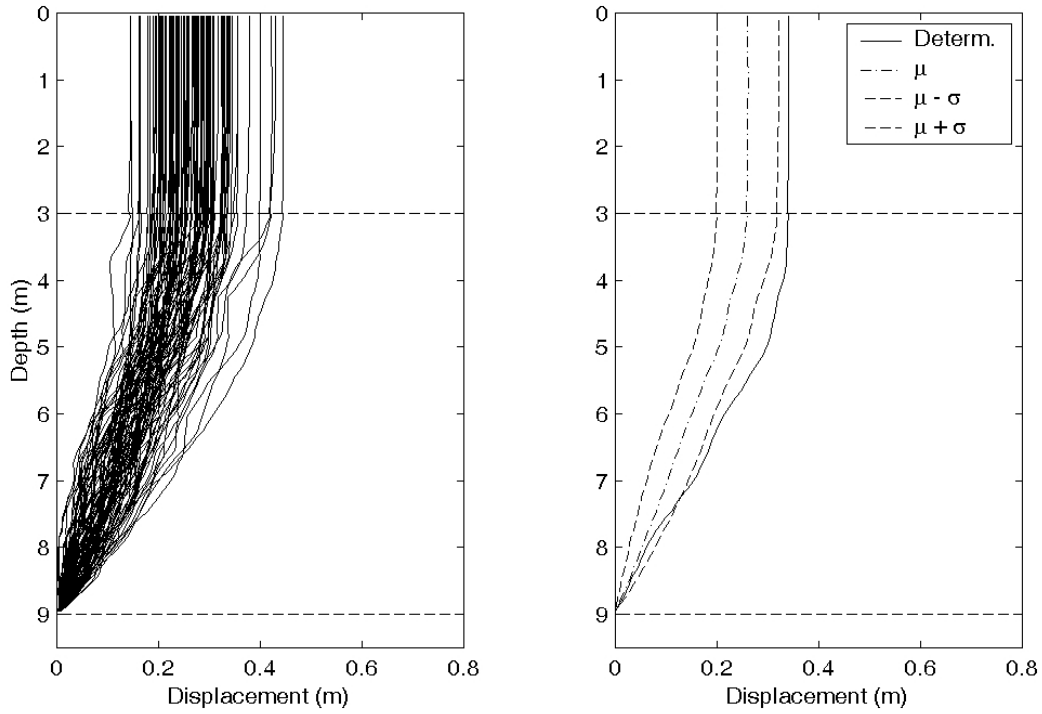


Fig. 7.6 Results of site response analysis: (a) profiles of permanent lateral displacement for all analyses and (b) comparison of deterministic profile with μ and μ ± σ profiles

Figure 7.6(b) shows that the displacement of the deterministic profile (based on $(N_1)_{60}$ = 10 with depth) is significantly greater than the average of the displacement profiles from the Monte Carlo analyses.

The results of the Monte Carlo analyses, however, allow evaluation of conditions such as the probability that lateral displacement will exceed 0.5 m at a depth of 3 m. This probability can be determined by constructing a cumulative distribution function for lateral displacement from the histogram of lateral displacement at 3 m. The CDF for the example problem is shown in Figure 7.7. Based on the CDF, the probability that the permanent lateral displacement at 3 m

depth will exceed 0.5 m is 41%. The probability that displacement is greater than the deterministic value of lateral displacement at a depth of 3 m (0.607 m) is about 7%.

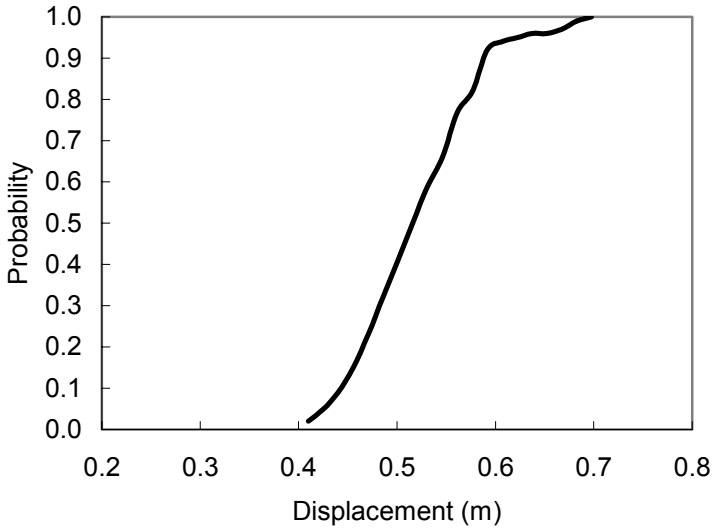


Fig. 7.7 CDF for lateral displacement at a depth of 3 m

7.4.2 Two-Dimensional Analysis

In many cases one-dimensional models are not sufficient for characterization of the response of soil-structure systems. Often, spatially varying boundary or loading conditions dictate the need for two- and three-dimensional numerical simulations.

7.4.2.1 Soil Profile

The example presented in this section represents a hypothetical footing subjected to oscillatory loading at the edge of a vertical cut in cohesive soil. The geometry and loading conditions were selected to clearly show the effect of spatial variability of soil properties on the response of the system. They do not represent any specific or real case. The geometry, shown in Figure 7.8, consists of a 10 ft vertical cut in a 20 ft thick soil deposit resulting in a 20 ft high, unsupported slope. Vertical boundaries are placed 20 ft from the cut. Soil conditions at the site consist of cohesive material with mean undrained shear strength, s_u, of 2200 psf. A COV of 30% was selected based on data collected from the literature (Table 4.3).

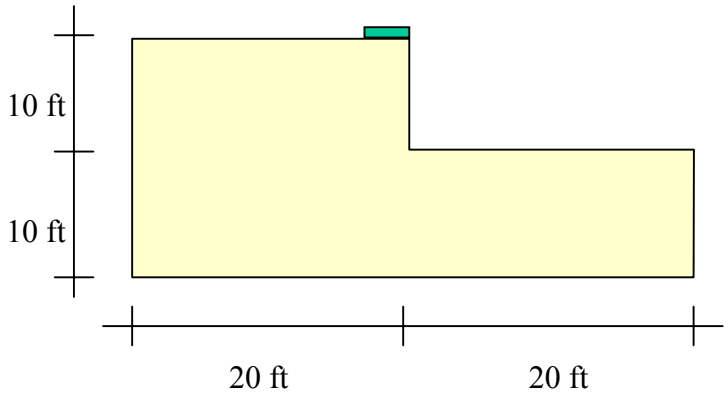


Fig. 7.8 Two-dimensional slope geometry

7.4.2.2 Method of Analysis

Two-dimensional numerical analyses were performed using the finite element tool OpenSees. OpenSees (Open System for Earthquake Engineering Simulation) is a PEER-developed software framework for simulating the seismic response of structural and geotechnical systems. OpenSees is intended to serve as the computational platform for research in performance-based earthquake engineering at PEER. OpenSees is open source, i.e., the source code is freely available to anyone who wishes to use it.

Linear equation solvers, time integration schemes, and solution algorithms are the core of the OpenSees computational framework. The components of a solution strategy are interchangeable, allowing analysts to find sets suited to their particular problem. Separate from the computational and problem solving classes, are classes for building models (termed Model-Builder classes). The latest release of OpenSees includes a general model-builder for creating two- and three-dimensional frame and continuum models using TCL. Several elements, material constitutive models, and loading conditions are available. More information on OpenSees can be found at **http://opensees.berkeley.edu.**

The example presented in this section uses 2-D quadrilateral elements for the geometry, and a J2 elasto-plastic model to characterize the material behavior. J2 plasticity characterizes fairly well the undrained shear strength of cohesive soils. The version used in this example does not consider kinematic hardening and therefore underestimates undrained strength during dynamic loading. However, its current implementation in OpenSees is very robust and produces re-

sults that are sufficient for the purposes of this example. Figure 7.9 shows the mesh, boundary conditions, and loading for the hypothetical cut slope problem.

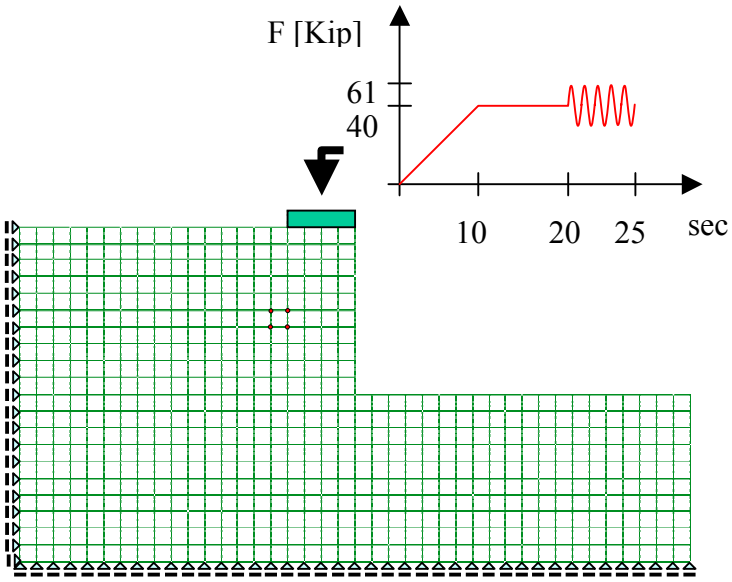


Fig. 7.9 Two-dimensional OpenSees finite element mesh and loading condition

The loading condition consists of a vertical static load of 40 kips applied using a ramp function, and a superimposed harmonic load with an amplitude of 42 Kips and a frequency of 1.0 Hz.

7.4.2.3 Deterministic Analysis

For this example, a deterministic dynamic analysis, performed using the mean value of c_u = 2200 psf produced the permanent displacement time history shown in Figure 7.10. The deterministic analysis predicts a maximum displacement of slightly more than 0.2 ft would develop. Most of this displacement results from straining of the upper 10 ft of soil beneath the footing.

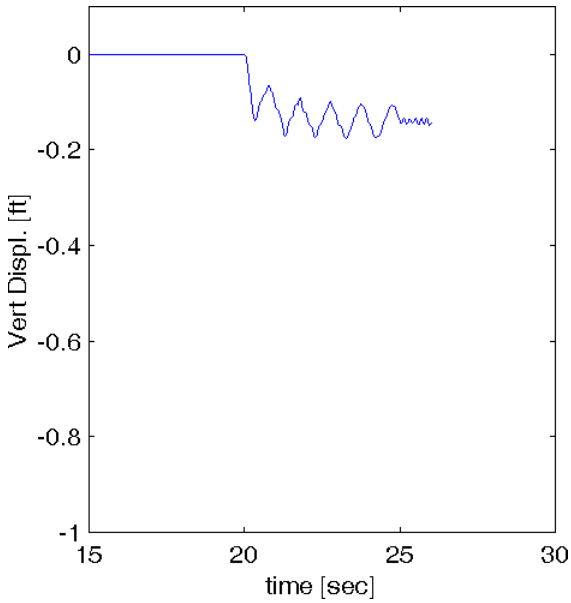


Fig. 7.10 Displacement time history of top of the footing — homogeneous soil undrained shear strength

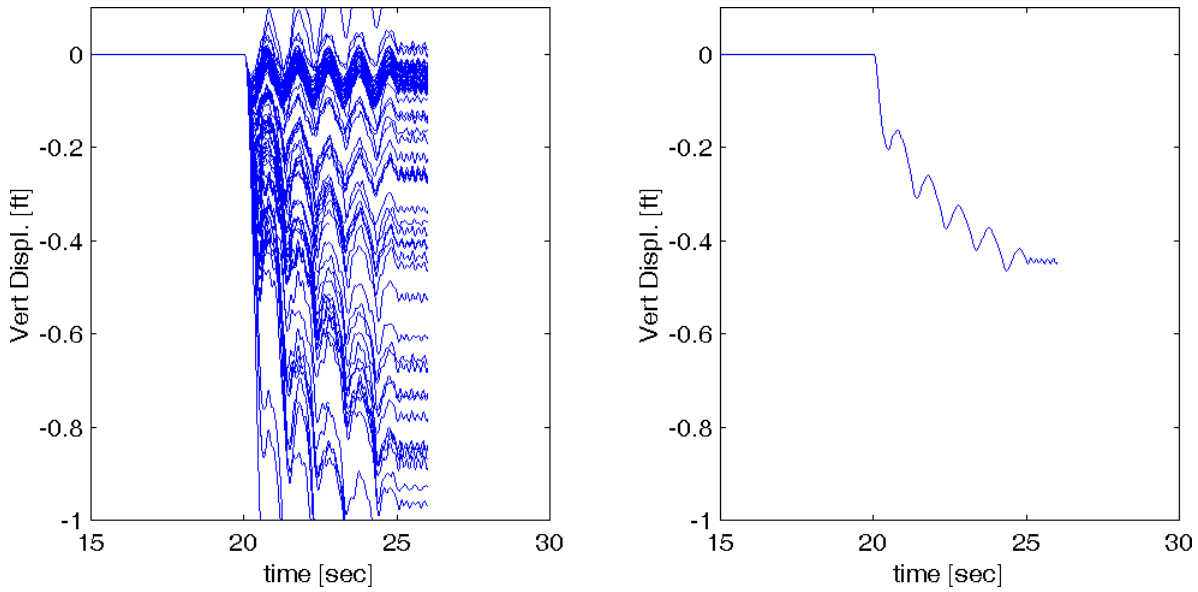
7.4.2.4 Simulation

Based on the statistics determined from the available field data, 100 two-dimensional Gaussian random fields of undrained shear strength were generated in a manner similar to that in Section 7.2.1 and outlined in Yamazaki (1988). A mean of 2200 psf, a COV of 30%, and a Gaussian spectral density function (standard deviation of 1.0 and a correlation decay coefficient equal to 4.0 for the spectral density function) were used to generate the soil profiles. Six of these soil profiles are shown in Figures 7.13(a) and 7.14. The histogram plots of the simulated undrained shear strength, s_u, shown in Figures 7.13(b) and 7.14, confirm that the undrained shear strength was normally distributed for each case. A minimum value of s_u = 1000 psf was used to avoid numerical instabilities. This is clearly reflected in the histograms presented in Figures 7.13(b) and 7.14.

7.4.2.5 Monte Carlo Analyses

The 100 soil profiles were analyzed using OpenSees. These analyses produced 100 vertical displacement time histories for the footing at the top of the open cut. Figure 7.11(a) shows the displacement time histories induced by the oscillatory loading component as predicted by all of the numerical analyses. Figure 7.11(b) shows the time history of the mean displacement of the foot-

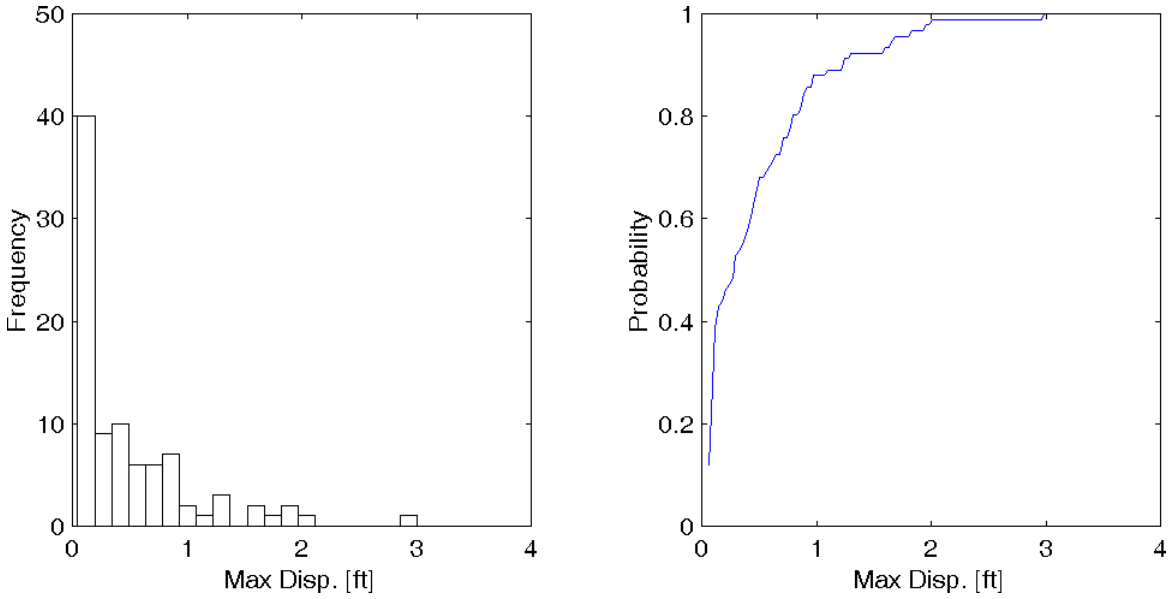
ing at the top of the cut. Comparison of Figure 7.10 and 7.11(b) shows that the displacement time history of the deterministic analysis (based on s_u = 2200 psf) is significantly less that the average of the displacement time histories from the Monte Carlo analyses.



(a) (b)

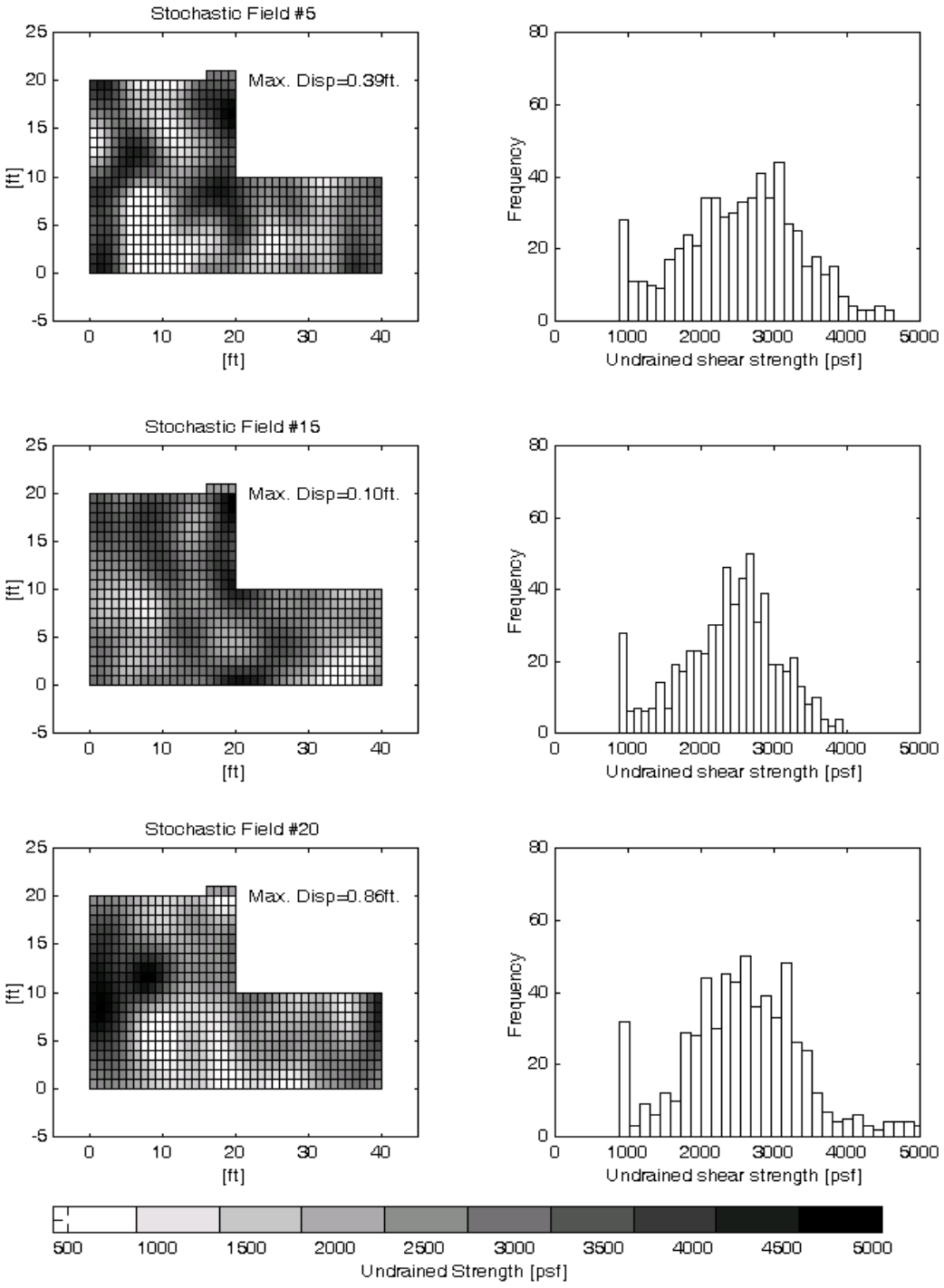
Fig. 7.11 Displacement of top of the footing (a) displacement time histories corresponding to all stochastic fields; (b) average time history of displacements

The results of the Monte Carlo analyses, however, allow estimation of the probability distribution of footing displacement. Therefore, quantities such as the probability that maximum vertical displacements of the top of the cut slope will exceed 6 in. (0.5ft) can be determined. The CDF of footing displacement for the example problem, which corresponds to the histogram of maximum displacements shown in Figures 7.12(a), is shown in Figure 7.12(b). Based on this CDF, the probability that the maximum vertical displacement of the top of the cut slope will exceed 6.0 in. is 32%. The probability that the maximum vertical displacement is greater than the deterministic value of the maximum vertical displacement (0.2 ft) is about 53%.



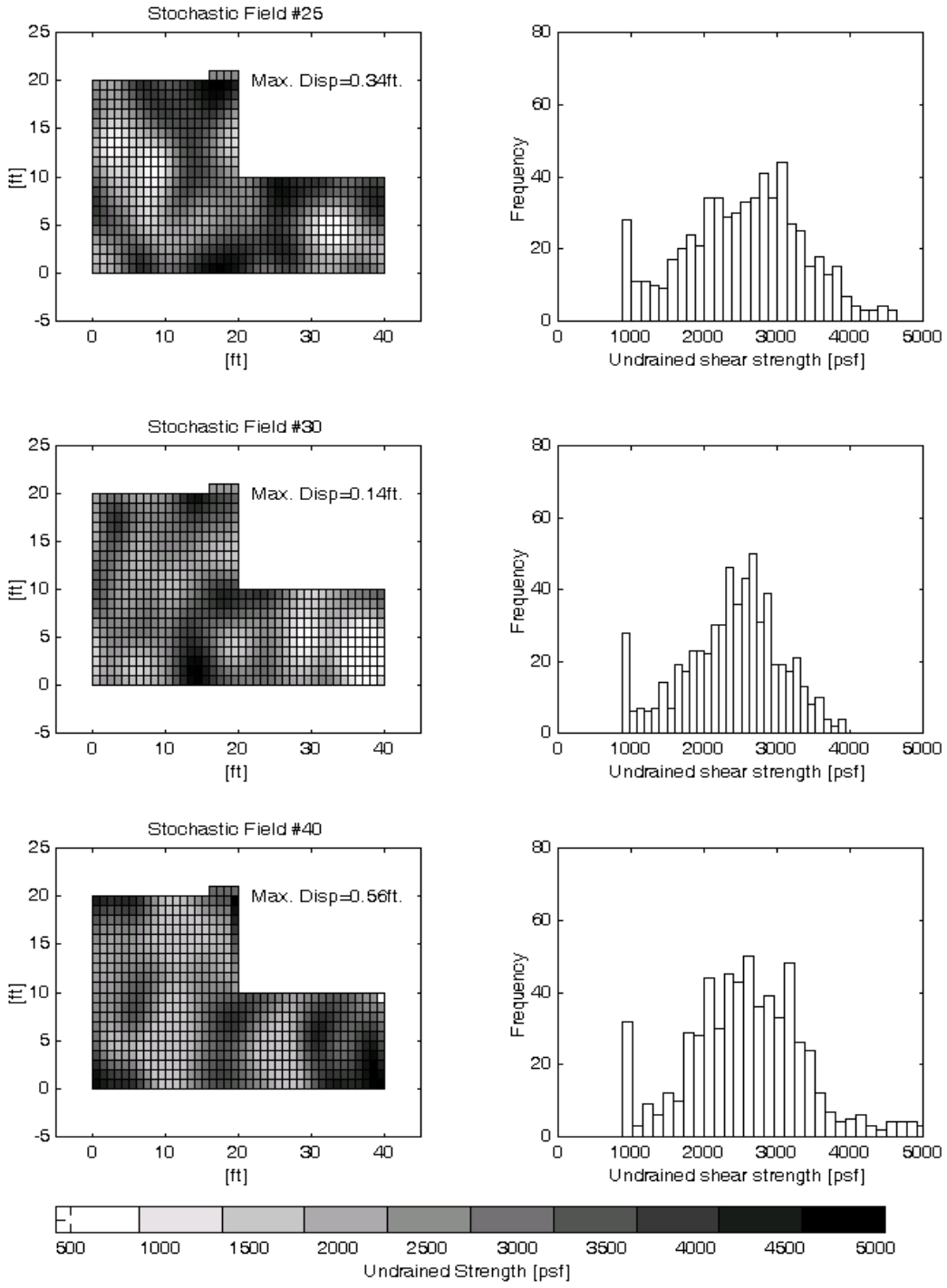
(a) (b)

Fig. 7.12 (a) Histogram of maximum displacements for the footing at the top of the cut slope; (b) cumulative distribution function (CDF) of maximum displacements



(a) **(b)**

Fig. 7.13 (a) Spatial variation of undrained shear strength, s_u; (b) cumulative histogram of undrained shear strength



(a) (b)

Fig. 7.14 (a) Spatial variation of undrained shear strength, s_u; (b) cumulative histogram of undrained shear strength

8 Summary

Prediction of the performance of structures during earthquakes requires accurate modeling of the geotechnical components of soil-structure systems. Geotechnical performance is strongly dependent on the properties of the soil beneath and adjacent to the structure of interest. These soil properties can be described using deterministic and probabilistic models. Deterministic models typically use a simple discrete descriptor for the parameter of interest. Probabilistic models account for uncertainty in soil properties by describing those properties by using discrete statistical descriptors or probability distribution functions.

This report is intended to provide PEER researchers with background information and available data on the uncertainty and spatial variability of soil properties. Such data will be useful in the evaluation of the effects of geotechnical uncertainty and variability on the performance of structures during earthquakes. It is anticipated that the information presented in this report will be used with deterministic analyses being developed by PEER researchers, both within and outside the OpenSees framework, to investigate the sensitivity of performance to geotechnical uncertainty.

The report describes sources and types of uncertainty in geotechnical engineering practice, and introduces the basic concepts and terminology of the theory or probability. Statistical parameters and the probability distributions most commonly used to describe geotechnical parameters are reviewed. The report then presents tabulated statistical parameters for soil properties that have been reported in the literature; both laboratory- and field-measured parameters describing moisture-density, plasticity, strength, and compressibility characteristics are presented. The theory of regionalized variables, including concepts of autocorrelation, variograms, and stationarity are presented, along with tabulated values of parameters describing spatial variability that have been reported in the literature. Finally, the procedures and tools for estimation and simulation of spatially variable soil properties are presented.

References

Benjamin, J.R., and Cornell, C.A. (1970). *Probability, Statistics, and Decisions for Civil Engineers*, McGraw-Hill, New York.

Benson, C.H., Daniel, D.E., and Boutwell, G.P. (1999). "Field performance of compacted clay liners," *Journal of Geotechnical and Geoenvironmental Engineering.,* ASCE, 125(5), pp. 390-403.

Boulanger, R., (1999). "Void Redistribution in sand following earthquake loading," *Physics and Mechanics of Soil Liquefaction*, Lade & Yamamuro (eds.), Rotterdam.

Box, G.E.P., and Muller, M.E. (1958). "A note on the generation of random normal deviates," *The Annals of Mathematical Statistics*, American Statistical Association, USA., Vol. 29, pp. 610-613.

Burington, R. S., and May, D. C., Jr. (1970). *Handbook of Probability and Statistics*, McGraw-Hill, New York.

Chiasson, P., Lafleur, J., Soulie, M., and Lay, K. T. (1995). "Characterizing spatial variability of a clay by geostatistics," *Canadian Geotechnical Journal*, Vol. 32, pp. 1-10.

Christian, J.T., and Baecher, G.B. (2001). Discussion of "Factor of safety and reliability in geotechnical engineering," *Journal of Geotechnical and Environmental Engineering*, Vol. 127, No. 28, pp. 700-703.

Christian, J.T., Ladd, C.C., and Baecher, G.B. (1994). "Reliability applied to slope stability analysis," *Journal of Geotechnical Engineering*, Vol. 120, No. 12, pp. 2180-2207.

Darendian, M.B. (2001). "Development of a new family of normalized modulus reduction and damping curves," Ph.D. Dissertation, The University of Texas at Austin, 362 pp.

Das, B. M., Puri, V. K., and Prakash, S. (1999). "Liquefaction of silty soils," Earthquake Geotechnical Engineering, Pinto (ed.), Rotterdam.

Davis, J. C. (1986). *Geostatistics and Data Analysis in Geology*, John Wiley & Sons.

DeGroot, D. J. (1996). "Analyzing spatial variability of insitu soil properties (invited paper)." *Uncertainty in the Geologic Environment, From Theory to Practice*, Proceeding of Uncertainty '96, Geotechnical Special Publication No. 58.

DeGroot, D. J., and Baecher, G. B., (1993). "Estimating autocovariance of insitu soil properties," *Journal of Geotechnical Engineering*, Vol. 119, No. 1, pp. 147-166.

Duncan, J. M. (2000). "Factors of safety and reliability in geotechnical engineering," *Journal of Geotechnical and Geoenvironmental Engineering*, Vol. 126, No. 4, pp. 307-316.

Elgamal, A. W., Parra, E., Yang, Z., Dobry, R., and Zeghal, M. (1999). "Liquefaction constitutive model," *Physics and Mechanics of Soil Liquefaction*, Lade & Yamamuro (eds.), Rotterdam.

Elgamal, A., Yang, Z., Parra, E., and Dobry, R. (1999). "Modeling of liquefaction-induced shear deformation," *Physics and Mechanics of Soil Liquefaction*, Lade & Yamamuro (eds.), Rotterdam.

Elgamel, A. W., Dobry, R., Parra, E., and Yang, Z. (1998). "Soil dilation and shear deformations during liquefaction," Proceedings, 4th International Conference on Case Studies in Geotechnical Engineering, St. Louis, Missouri.

Fardis, M. N., and Zeneziano, D. (1982). "Probabilistic analysis of deposit liquefaction," *Journal of Geotechnical Engineering Division*, Vol. 108, No. GT3, pp. 395-417.

Fenton, G. A., and Vanmarcke, E. H. (1998). "Spatial variation in liquefaction risk," *Geotechnique*, Vol. 48, No. 6, pp. 819-831.

Frost, J. D., Carroll, D. P., and Rockaway, T. D. (1998). "Spatial liquefaction analysis," *Spatial Analysis in Soil Dynamics and Earthquake Engineering*, Geotechnical Special Publication No. 67, ASCE, pp. 70-86.

Griffiths, D. V., and Fenton, G. A. (1997). "Three-dimensional seepage through spatially random soil," *Journal of Geotechnical Engineering and Geoenvironmental Engineering*, Vol. 123, No. 2, pp. 153-160.

Griffiths, D. V., and Fenton, G. A. (1993). "Seepage beneath water retaining structures founded on spatially random soil," *Geotechnique*, Vol. 43, No. 4, pp. 577-587.

Grivas, D.A. (1981). "How Reliable are the Present Slope Failure Prediction Methods?," Proceedings of the Tenth International Conference of Soil Mechanics and Foundation Engineering, Stockholm, Sweden, Vol. 3, pp.427-430.

Harr, M.E. (1987). *Reliability-Based Design in Civil Engineering*. McGraw-Hill Book Company.

Harr, M.E. (1977) *Mechanics of Particulate Media*, McGraw-Hill.

Hasofer, A. M. (1993). "Simulation of random fields." *Probabilistic Methods in Geotechnical Engineering*, Li & Lo (eds.), Balkema, Rotterdam.

Hegazy, A.H., Mayne, P.M., and Rouhani, S. (1996). "Geostatistical assessment of spatial variability in piezocone tests," *Uncertainty in the Geologic Environment, From Theory to Practice*, Proceeding of Uncertainty '96, Geotechnical Special Publication No. 58.

Horne, J.C., (1996) "Effects of liquefaction-induced lateral spreading on pile foundations," Ph.D. Dissertation, University of Washington.

Issaks, E. H., and Srivastava, R. M. (1989). *An Introduction to Applied Geostatistics*, Oxford University Press.

Lacasse and Nadim (1996) "Uncertainties in characterizing soil properties," *Uncertainty in the Geologic Environment, From Theory to Practice*, Proceeding of Uncertainty '96, Geotechnical Special Publication No. 58.

Journel, A. G., and Huijbregts, C. J. (1978). *Mining Geostatistics*, Academic Press.

Kelly, D. G., (1994). *Introduction to Probability*, Macmillan Publishing Company.

Kulhawy, F.H. (1992). "On the evaluation of static soil properties," ASCE Geotechnical Special Publication No. 31., pp. 95-115.

Kulhawy, F.H., and Phoon, K. (1996). "Engineering judgment in the evolution from deterministic to reliability-based foundation design (plenary)," *Uncertainty in the Geologic Environment, From Theory to Practice*, Proceeding of Uncertainty '96, Geotechnical Special Publication No. 58.

Kulhawy, F., and Trautman, C.H. (1996), "Estimation of insitu test uncertainty," in *Uncertainty in the Geologic Environment: From Theory to Practice*, Proceeding of Uncertainty '96, ASCE Geotechnical Special Publication No. 58.

Kutter, B. L., and Balakrishnam, A. (1999). "Visualization of soil behavior from dynamic centrifuge model tests," Earthquake Geotechnical Engineering, Pinto (ed.), Rotterdam.

Lacasse, S., and Nadim, F. (1996). "Uncertainties in characterizing soil properties (plenary)," *Uncertainty in the Geologic Environment, From Theory to Practice*, Proceeding of Uncertainty '96, Geotechnical Special Publication No. 58.

Lapin, L.L. (1983). *Probability and Statistics for Modern Engineering*. PWS Publishers. pp. 624.

Lumb, P. (1966). *The Variability of Natural Soils* . Canadian Geotechnical Journal, Vol. 3, No. 2, pp. 74-97.

Lumb, P. 1970. "Safety factors and the probability distribution of soil strength," *Canadian Geotechnical Journal*, Vol. 7, No. 3, pp. 225-242.

Mantoglou, A., and Wilson, J. L. (1982). "The turning bands method for simulation of random fields using line generation by a spectral method," *Water Resources Research*, Vol. 18, No. 5, pp. 1379-1394.

Mendenhall, W., and Beaver, R. (1991). *Introduction to Probability and Statistics*, PWS-Kent Publishing Company.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., and Teller, E. (1953). Journal o f Chemical Physics, Volume 21, pp. 1087.

McBratney A.B., Hart, A.B. and McGarry, D. (1991). "The use of region partitioning to improve the representation of geostatistically mapped soil attributes," *Journal of Soil Science*, Vol. 42, pp. 513-532.

McKay, M.D., Beckman, R.J., and Conover, W.J. (1979). "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code." Technometrics, 21(2), pp.239-245.

McKenna, S. A., Cromer, M. V., Rautman, C. A., and Zelinshi, W. P. (1996). "Addressing uncertainty in rock properties through geostatistics simulation," *Uncertainty in the Geologic Environment, From Theory to Practice*, Proceeding of Uncertainty '96, Geotechnical Special Publication No. 58.

Mostyn, G.R., and Li, K.S. (1993). *Probabilistic Slope Stability Analysis - State-of-Play*, Proceedings of the Conference on Probabilistic Methods in Geotechnical Engineering, Canberra, Australia. pp. 281-290.

Odeh, I. O. A., McBratney, A. B., and Chittleborough, D. J. (1995). "Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging," *Geoderma* Vol. 67, pp. 215-225.

Ott, L., and Mednehall, W. (1994). *Understanding Statistics*, Duxbury Press, Belmont, California.

Paloheimo, E., and Hannus, M. (1974). "Structural design based on weighted fractals." *Journal of Structural Division*, ASCE, Volume 100, No. 7, pp. 1367-1378.

Pardo-Iquzquiza, E., and Dowd, P. A. (1998). "The second-order stationary universal kriging model revisited," *Mathematical Geology*, Vol. 30, No. 4, pp. 347379.

Pebesma, E.J. (2001). *GSTAT User's Manual*, Department of Physical Geography, The Netherlands.

Pebesma, E.J., and Wesseling, C.G. (1998). "Gstat: a program for geostatistical modeling, prediction and simulation," *Computers & Geosciences* Vol. 24, No. 1.

Phoon, K., and Kulhawy, F.H. (1999). "Characterization of geotechnical variability," *Canadian Geotechnical Journal*, Vol. 36, pp. 625-639.

Phoon, K., and Kulhawy, F.H. (1999). "Evaluation of geotechnical property variability," *Canadian Geotechnical Journal*, Vol. 36, pp. 612-624.

Popescu, R., Deodatis, G., and Prevost, J. H. (1998). "Simulation of homogeneous non-Gaussian stochastic vector fields," *Probabilistic Engineering Mechanics* Vol. 13, No. 1, pp. 1-13.

Popescu, R., Prevost, J. H., and Deodatis, G. (1997). "Effects of spatial variability on soil liquefaction: some design recommendations," *Geotechnique* Vol. 47, No. 5, pp. 1019-1036.

Popescu, R., Prevost, J. H., and Deodatis, G. (1998). "Characteristic percentile of soil strength for dynamic analyses," Proceedings, Geotechnical Earthquake Engineering and Soil Dynamics III, Seattle, Washington, pp. 1461-1471.

Popescu, R., Prevost, J. H., and Deodatis, G. (1998). "Spatial variability of soil properties: two case studies," Proceedings, Geotechnical Earthquake Engineering and Soil Dynamics III, Seattle, Washington, pp. 568-579.

Prevost, J. H., Popescu, R. (1996). "Constitutive relations for soil materials," *Electronic Journal of Geotechnical Engineering*.

Shannon and Wilson, Inc., and Wolff, T.F. (1994) "Probability models for geotechnical aspects of navigation structures," report to the St. Louis District, U.S. Corps of Engineers.

Sinowski, W., Scheinost, A. C., and Auerswald, K. (1997). "Regionalization of soil water retention curves in a highly variable soilscape, II. Comparison of regionalization procedures using a pedotransfer function," *Geoderma*, 78, pp. 145-159.

Stein, M.L., (1987). "Large sample properties of simulations using Latin hypercube sampling." Technometrics, 29(2). pp. 143-151.

Tan, C.P. Donald, I.B. and Melchers, R.E. (1993). *Probabilistic Slope Stability Analysis - State-of-Play*, Proceedings of the Conference on Probabilistic Methods in Geotechnical Engineering, Canberra, Australia. pp. 89-110.

United States Army Corps of Engineers (USACE) (1999). "Appendix A – An overview of probabilistic analysis for geotechnical engineering problems," Engineering Technical Letter (ETL) 1110-2-556.

United States Army Corps of Engineers (USACE) (1995). "Appendix B – Introduction to probability and reliability in geotechnical engineering," Engineering Technical Letter (ETL) 1110-2-547.

Vallejo, L.E. (1996). "Evaluating the variability of engineering properties of soil deposits using fractals," *Uncertainty in the Geologic Environment, From Theory to Practice*, Proceeding of Uncertainty '96, Geotechnical Special Publication No. 58.

Vanmarcke, E. (1983). *Simulation of Random Fields*, The MIT Press, Cambridge, Massachusetts.

Vanmarcke, E. H. (1977). "Probabilistic modeling of soil profiles," *Journal of the Geotechnical Engineering Division of the ASCE* Vol. 103, GT11, pp. 1227-1246.

Wickremesinghe, D., and Campanella, R.G. (1993). "Scale of fluctuation as a descriptor of soil variability," *Probabilistic Methods in Geotechnical Engineering*, Li & Lo (eds.), Balkema, Rotterdam.

Wolff, T.F., Hassen, A., Khan, R., Ur-Rasul, I., and Miller, M. (1995), "Geotechnical reliability of dam and levee embankments," Report for the U.S. Army Corps of Engineers, Experiment Station, by Michigan State University, September 1995.

Wolff, T.F., Demsky, E.C., Schauer, J., and Perry, E. (1996). "Reliability assessment of dike and levee embankments," *Uncertainty in the Geologic Environment, From Theory to Practice*, Proceeding of Uncertainty '96, Geotechnical Special Publication No. 58.

Yamazaki, F., Shinozuka, M. (1988). "Digital generation of non-Gaussian stochastic fields," *Journal of Engineering Mechanics*, Vol. 114, No. 7, 1183-1197.

Yang, Fredlund, D.J., and Stolte, (1993) – General Monte Carlo Procedure.

Yoshida, N., Ito, K. (1999). "Liquefaction of improved ground at Part Island, Japan, during the 1995 Hyogoken-nanbu earthquake," Earthquake Geotechnical Engineering, Pinto (ed.), Rotterdam.

Youd, T. L. (1999). "Physics and mechanics of liquefaction from field records and experience," Physics and Mechanics of Soil Liquefaction, Lade & Yamamuro (eds.), Rotterdam.

Young, C.T., and Doucette, J.P. (1996). "Sand variability from ground penetrating radar," *Uncertainty in the Geologic Environment, From Theory to Practice*, Proceeding of Uncertainty '96, Geotechnical Special Publication No. 58.

PEER REPORTS

PEER reports are available from the National Information Service for Earthquake Engineering (NISEE). To order PEER reports, please contact the Pacific Earthquake Engineering Research Center, 1301 South 46th Street, Richmond, California 94804-4698. Tel.: (510) 231-9468; Fax: (510) 231-9461.

PEER 2002/16 *Estimation of Uncertainty in Geotechnical Properties for Performance-Based Earthquake Engineering.* Allen L. Jones, Steven L. Kramer, and Pedro Arduino. December 2002.

PEER 2002/14 *Inelastic Seismic Response of Extended Pile Shaft Supported Bridge Structures.* T.C. Hutchinson, R.W. Boulanger, Y.H. Chai, and I.M. Idriss. December 2002.

PEER 2002/13 *Probabilistic Models and Fragility Estimates for Bridge Components and Systems.* Paolo Gardoni, Armen Der Kiureghian, and Khalid M. Mosalam. June 2002.

PEER 2002/12 *Effects of Fault Dip and Slip Rake on Near-Source Ground Motions: Why Chi-Chi Was a Relatively Mild M7.6 Earthquake.* Brad T. Aagaard, John F. Hall, and Thomas H. Heaton. December 2002.

PEER 2002/11 *Analytical and Experimental Study of Fiber-Reinforced Strip Isolators*. James M. Kelly and Shakhzod M. Takhirov. September 2002.

PEER 2002/10 *Centrifuge Modeling of Settlement and Lateral Spreading with Comparisons to Numerical Analyses***.** Sivapalan Gajan and Bruce L. Kutter. January 2003.

PEER 2002/09 *Documentation and Analysis of Field Case Histories of Seismic Compression during the 1994 Northridge, California, Earthquake.* Jonathan P. Stewart, Patrick M. Smith, Daniel H. Whang, and Jonathan D. Bray. October 2002.

PEER 2002/08 *Component Testing, Stability Analysis and Characterization of Buckling-Restrained Unbonded Braces™.* Cameron Black, Nicos Makris, and Ian Aiken. September 2002.

PEER 2002/07 *Seismic Performance of Pile-Wharf Connections.* Charles W. Roeder, Robert Graff, Jennifer Soderstrom, and Jun Han Yoo. December 2001.

PEER 2002/06 *The Use of Benefit-Cost Analysis for Evaluation of Performance-Based Earthquake Engineering Decisions*. Richard O. Zerbe and Anthony Falit-Baiamonte. September 2001.

PEER 2002/05 *Guidelines, Specifications, and Seismic Performance Characterization of Nonstructural Building Components and Equipment*. André Filiatrault, Constantin Christopoulos, and Christopher Stearns. September 2001.

PEER 2002/03 *Investigation of Sensitivity of Building Loss Estimates to Major Uncertain Variables for the Van Nuys Testbed.* Keith A. Porter, James L. Beck, and Rustem V. Shaikhutdinov. August 2002.

PEER 2002/02 *The Third U.S.-Japan Workshop on Performance-Based Earthquake Engineering Methodology for Reinforced Concrete Building Structures*. July 2002.

PEER 2002/01 *Nonstructural Loss Estimation: The UC Berkeley Case Study. Mary C. Comerio and John C. Stallmeyer.* December 2001.

PEER 2001/16 *Statistics of SDF-System Estimate of Roof Displacement for Pushover Analysis of Buildings.* Anil K. Chopra, Rakesh K. Goel, and Chatpan Chintanapakdee. December 2001.

PEER 2001/15 *Damage to Bridges during the 2001 Nisqually Earthquake.* R. Tyler Ranf, Marc O. Eberhard, and Michael P. Berry. November 2001.

PEER 2001/14 *Rocking Response of Equipment Anchored to a Base Foundation.* Nicos Makris and Cameron J. Black. September 2001.

PEER 2001/13 *Modeling Soil Liquefaction Hazards for Performance-Based Earthquake Engineering.* Steven L. Kramer and Ahmed-W. Elgamal. February 2001.

PEER 2001/12 *Development of Geotechnical Capabilities in OpenSees.* Boris Jeremic. September 2001.

PEER 2001/11 *Analytical and Experimental Study of Fiber-Reinforced Elastomeric Isolators*. James M. Kelly and Shakhzod M. Takhirov. September 2001.

PEER 2001/10 *Amplification Factors for Spectral Acceleration in Active Regions.* Jonathan P. Stewart, Andrew H. Liu, Yoojoong Choi, and Mehmet B. Baturay. December 2001.

PEER 2001/09 *Ground Motion Evaluation Procedures for Performance-Based Design.* Jonathan P. Stewart, Shyh-Jeng Chiou, Jonathan D. Bray, Robert W. Graves, Paul G. Somerville, and Norman A. Abrahamson. September 2001.

PEER 2001/08 *Experimental and Computational Evaluation of Reinforced Concrete Bridge Beam-Column Connections for Seismic Performance*. Clay J. Naito, Jack P. Moehle, and Khalid M. Mosalam. November 2001.

PEER 2001/07 *The Rocking Spectrum and the Shortcomings of Design Guidelines.* Nicos Makris and Dimitrios Konstantinidis. August 2001.

PEER 2001/06 *Development of an Electrical Substation Equipment Performance Database for Evaluation of Equipment Fragilities*. Thalia Agnanos. April 1999.

PEER 2001/05 *Stiffness Analysis of Fiber-Reinforced Elastomeric Isolators.* Hsiang-Chuan Tsai and James M. Kelly. May 2001.

PEER 2001/04 *Organizational and Societal Considerations for Performance-Based Earthquake Engineering.* Peter J. May. April 2001.

PEER 2001/03 *A Modal Pushover Analysis Procedure to Estimate Seismic Demands for Buildings: Theory and Preliminary Evaluation.* Anil K. Chopra and Rakesh K. Goel. January 2001.

PEER 2001/02 *Seismic Response Analysis of Highway Overcrossings Including Soil-Structure Interaction.* Jian Zhang and Nicos Makris. March 2001.

PEER 2001/01 *Experimental Study of Large Seismic Steel Beam-to-Column Connections.* Egor P. Popov and Shakhzod M. Takhirov. November 2000.

PEER 2000/10 *The Second U.S.-Japan Workshop on Performance-Based Earthquake Engineering Methodology for Reinforced Concrete Building Structures.* March 2000.

PEER 2000/09 *Structural Engineering Reconnaissance of the August 17, 1999 Earthquake: Kocaeli (Izmit), Turkey.* Halil Sezen, Kenneth J. Elwood, Andrew S. Whittaker, Khalid Mosalam, John J. Wallace, and John F. Stanton. December 2000.

PEER 2000/08 *Behavior of Reinforced Concrete Bridge Columns Having Varying Aspect Ratios and Varying Lengths of Confinement.* Anthony J. Calderone, Dawn E. Lehman, and Jack P. Moehle. January 2001.

PEER 2000/07 *Cover-Plate and Flange-Plate Reinforced Steel Moment-Resisting Connections.* Taejin Kim, Andrew S. Whittaker, Amir S. Gilani, Vitelmo V. Bertero, and Shakhzod M. Takhirov. September 2000.

PEER 2000/06 *Seismic Evaluation and Analysis of 230-kV Disconnect Switches.* Amir S. J. Gilani, Andrew S. Whittaker, Gregory L. Fenves, Chun-Hao Chen, Henry Ho, and Eric Fujisaki. July 2000.

PEER 2000/05 *Performance-Based Evaluation of Exterior Reinforced Concrete Building Joints for Seismic Excitation.* Chandra Clyde, Chris P. Pantelides, and Lawrence D. Reaveley. July 2000.

PEER 2000/04 *An Evaluation of Seismic Energy Demand: An Attenuation Approach.* Chung-Che Chou and Chia-Ming Uang. July 1999.

PEER 2000/03 *Framing Earthquake Retrofitting Decisions: The Case of Hillside Homes in Los Angeles.* Detlof von Winterfeldt, Nels Roselund, and Alicia Kitsuse. March 2000.

PEER 2000/02 *U.S.-Japan Workshop on the Effects of Near-Field Earthquake Shaking.* Andrew Whittaker, ed. July 2000.

PEER 2000/01 *Further Studies on Seismic Interaction in Interconnected Electrical Substation Equipment.* Armen Der Kiureghian, Kee-Jeung Hong, and Jerome L. Sackman. November 1999.

PEER 1999/14 *Seismic Evaluation and Retrofit of 230-kV Porcelain Transformer Bushings*. Amir S. Gilani, Andrew S. Whittaker, Gregory L. Fenves, and Eric Fujisaki. December 1999.

PEER 1999/13 *Building Vulnerability Studies: Modeling and Evaluation of Tilt-up and Steel Reinforced Concrete Buildings.* John W. Wallace, Jonathan P. Stewart, and Andrew S. Whittaker, editors. December 1999.

PEER 1999/12 *Rehabilitation of Nonductile RC Frame Building Using Encasement Plates and Energy-Dissipating Devices.* Mehrdad Sasani, Vitelmo V. Bertero, James C. Anderson. December 1999.

PEER 1999/11 *Performance Evaluation Database for Concrete Bridge Components and Systems under Simulated Seismic Loads.* Yael D. Hose and Frieder Seible. November 1999.

PEER 1999/10 *U.S.-Japan Workshop on Performance-Based Earthquake Engineering Methodology for Reinforced Concrete Building Structures.* December 1999.

PEER 1999/09 *Performance Improvement of Long Period Building Structures Subjected to Severe Pulse-Type Ground Motions*. James C. Anderson, Vitelmo V. Bertero, and Raul Bertero. October 1999.

PEER 1999/08 *Envelopes for Seismic Response Vectors.* Charles Menun and Armen Der Kiureghian. July 1999.

PEER 1999/07 *Documentation of Strengths and Weaknesses of Current Computer Analysis Methods for Seismic Performance of Reinforced Concrete Members.* William F. Cofer. November 1999.

PEER 1999/06 *Rocking Response and Overturning of Anchored Equipment under Seismic Excitations.* Nicos Makris and Jian Zhang. November 1999.

PEER 1999/05 *Seismic Evaluation of 550 kV Porcelain Transformer Bushings.* Amir S. Gilani, Andrew S. Whittaker, Gregory L. Fenves, and Eric Fujisaki. October 1999.

PEER 1999/04 *Adoption and Enforcement of Earthquake Risk-Reduction Measures.* Peter J. May, Raymond J. Burby, T. Jens Feeley, and Robert Wood.

PEER 1999/03 *Task 3 Characterization of Site Response General Site Categories*. Adrian Rodriguez-Marek, Jonathan D. Bray, and Norman Abrahamson. February 1999.

PEER 1999/02 *Capacity-Demand-Diagram Methods for Estimating Seismic Deformation of Inelastic Structures: SDF Systems*. Anil K. Chopra and Rakesh Goel. April 1999.

PEER 1999/01 *Interaction in Interconnected Electrical Substation Equipment Subjected to Earthquake Ground Motions.* Armen Der Kiureghian, Jerome L. Sackman, and Kee-Jeung Hong. February 1999.

PEER 1998/08 *Behavior and Failure Analysis of a Multiple-Frame Highway Bridge in the 1994 Northridge Earthquake.* Gregory L. Fenves and Michael Ellery. December 1998.

PEER 1998/07 *Empirical Evaluation of Inertial Soil-Structure Interaction Effects.* Jonathan P. Stewart, Raymond B. Seed, and Gregory L. Fenves. November 1998.

PEER 1998/06 *Effect of Damping Mechanisms on the Response of Seismic Isolated Structures.* Nicos Makris and Shih-Po Chang. November 1998.

PEER 1998/05 *Rocking Response and Overturning of Equipment under Horizontal Pulse-Type Motions.* Nicos Makris and Yiannis Roussos. October 1998.

PEER 1998/04 *Pacific Earthquake Engineering Research Invitational Workshop Proceedings, May 14–15, 1998: Defining the Links between Planning, Policy Analysis, Economics and Earthquake Engineering.* Mary Comerio and Peter Gordon. September 1998.

PEER 1998/03 *Repair/Upgrade Procedures for Welded Beam to Column Connections.* James C. Anderson and Xiaojing Duan. May 1998.

PEER 1998/02 *Seismic Evaluation of 196 kV Porcelain Transformer Bushings.* Amir S. Gilani, Juan W. Chavez, Gregory L. Fenves, and Andrew S. Whittaker. May 1998.

PEER 1998/01 *Seismic Performance of Well-Confined Concrete Bridge Columns.* Dawn E. Lehman and Jack P. Moehle. December 2000.