ESTIMATION FOR STOCHASTIC SOIL MODELS

By Gordon A. Fenton,[1] Associate Member, ASCE

ABSTRACT: Although considerable theory exists for the probabilistic treatment of soils, the ability to identify the nature of spatial stochastic soil variation is almost nonexistent. We all know that we could excavate an entire site and there would be no doubt about the soil properties. However, there would no longer be anything to rest our structure on, and so we must live with uncertainty and attempt to quantify it rationally. Twenty years ago the mean and variance was sufficient. Clients are now demanding full reliability studies, requiring more sophisticated models, so that engineers are becoming interested in rational soil correlation structures. Knowing that soil properties are spatially correlated, what is a reasonable correlation model? Are soils best represented using fractal models or finite-scale models? What is the difference? How can this question be answered? Once a model has been decided upon, how can its parameters be estimated? These are questions that this paper addresses by looking at a number of tools that aid in selecting appropriate stochastic models. These tools include the sample covariance, spectral density, variance function, variogram, and wavelet variance functions. Common models, corresponding to finite scale and fractal models, are investigated, and estimation techniques are discussed.

INTRODUCTION

The reliability assessment of geotechnical projects has been receiving increased attention from regulatory bodies in recent years. To provide a rational reliability analysis of a geotechnical system, there is a need for realistic random soil models that can then be used to assess probabilities relating to the design. Unfortunately, little research on the nature of soil spatial variability is available, and this renders reliability analyses using spatial variability suspect. In an attempt to remedy this situation, this paper lays out the theory and discusses the analysis and estimation tools needed to analyze spatially distributed soil data statistically. Because of the complexity of the problem, the concentration herein is purely on the 1D case. That is, the overall goal is to establish reasonable models for soil variability along a line. To achieve this goal, existing tools and estimators need to be critically reviewed to assess their performance for both large and small geotechnical data sets.

In general, statistical analyses can be separated into two areas that can be thought of as descriptive and inferential in nature. In the former, the goal is to best describe a particular data set with a view toward interpolating within the data set. For example, this commonly occurs when geotechnical data are obtained at a site for which a design is destined. The descriptive techniques most often used are those of regression, using an appropriate polynomial that explains most of the variability, or best linear unbiased estimation (BLUE). Regression is purely geometry and observation based, whereas BLUE also incorporates the covariance structure between the data. Thus, the BLUE techniques require an a priori estimate of the covariance function governing the soil's spatial variability; this is often obtained by inference from other sites because it generally requires a very large data set to estimate reliably.

In general, inference occurs whenever one estimates properties at any unobserved spatial location. Here, the word inference will be taken to mean the estimation of stochastic model parameters that allow one to make probabilistic statements about an entire site for which data are limited or not available. This may be necessary, for example, in preliminary designs, designs involving a future state, or designs where a large site is to be characterized on the basis of a small test region. This paper focuses on inferential statistics for several reasons: (1) Descriptive statistics are already reasonably well established and understood; (2) a priori knowledge of the second-moment (covariance) structure of soil properties is essential for BLUE estimators and Bayesian updating; and (3) site investigations are often not complete enough to even begin a spatial covariance estimation with any accuracy at all. Thus, most reliability based designs will benefit from a database of second-order soil statistics.

Consider a site from which a reasonably dense set of soil samples has been gathered. The goal is to make statements using this data set about the stochastic nature of the soil at a different, although presumed similar, site. The collected data are a sample extracted over some sampling domain of extent D from a continuously varying soil property field. An example may be seen in Fig. 1, where the solid line could represent the known, sampled, undrained shear strength of the soil, and the dashed lines represent unknown shear strengths outside the sampling domain.

Clearly, this sample exhibits a strong spatial trend and would classically be represented by an equation of the form

$$S_u(z) = m(z) + \varepsilon(z) \tag{1}$$

where $m(z)$ = deterministic function giving the mean soil property at z; and $\varepsilon(z)$ = random residual. If the goal were purely descriptive, then $m(z)$ would likely be selected to allow optimally accurate (minimum variance) interpolation of S_u between observations. This generally involves letting $m(z)$ be a polynomial trend in z with coefficients selected to render ε mean zero with small variance.

However, if the data shown in Fig. 1 are to be used to characterize another site, then the trend must be viewed with considerable caution. In particular, one must ask if a similar trend is expected to be seen at the site being characterized, and, if so, where the trend origin is to be located. In some cases, where soil properties vary predictably with depth, the answer to this question is affirmative. For example, undrained shear strength is commonly thought to increase with depth (but not always). In cases where the same trend is not likely to reappear at the target site, then removal of the trend from the data and dealing with just the residual $\varepsilon(z)$ have the following implications:

- Typically, the covariance structure of $\varepsilon(z)$ is drastically different from that of $S_u(z)$—it shows more spatial independence and has reduced variance.
- The reintroduction of the trend to predict the deterministic

[1]Assoc. Prof., Dept. of Engrg. Math., Dalhousie Univ., Halifax, NS, Canada B3J 2X4.

Note. Discussion open until November 1, 1999. Separate discussions should be submitted for the individual papers in this symposium. To extend the closing date one month, a written request must be filed with the ASCE Manager of Journals. The manuscript for this paper was submitted for review and possible publication on May 27, 1998. This paper is part of the ***Journal of Geotechnical and Geoenvironmental Engineering***, Vol. 125, No. 6, June, 1999. ©ASCE, ISSN 1090-0241/99/0006-0470–0485/$8.00 + $.50 per page. Paper No. 18420.

J. Geotech. Geoenviron. Eng., 1999, 125(6): 470-485

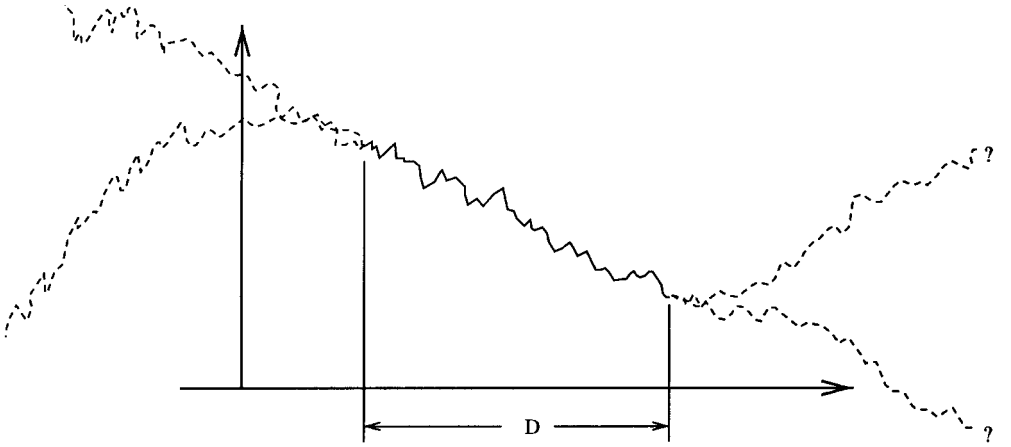

FIG. 1. Soil Sample on Finite-Sampling Domain

part of the soil properties at the target site may be grossly in error.

- The use of only the residual process $\varepsilon(z)$ at the target site will considerably underestimate the soil variability—the reported statistics will be unconservative. In fact, the more variability accounted for by $m(z)$, the less variable is $\varepsilon(z)$.

From these considerations, it is easily seen that trends that are not likely to reappear, that is, trends that are not physically (or empirically) based and predictable, must not be removed prior to performing an inferential statistical analysis. The trend itself is part of the uncertainty to be characterized and removing it leads to unconservative reported statistics.

It should be pointed out at this time that one of the reasons for ''detrending'' the data is precisely to render the residual process largely spatially independent. This is desirable because virtually all classical statistics are based on the idea that samples are composed of independent and identically distributed observations. Alternatively, when observations are dependent, the distributions of the estimators become very difficult to establish. This is compounded by the fact that the actual dependence structure is unknown. Only a limited number of asymptotic results are available to provide insight into the spatially dependent problem (Beran 1994), and simulation techniques are proving very useful in this regard (Cressie 1993).

Another issue to be considered is the level of information available at the target site. Generally, a design does not proceed in the complete absence of site information. The ideal case involves gathering enough data to allow the main characteristics, for instance, the mean and variance, of the soil property to be established with reasonable confidence. Then, inferred statistics regarding the spatial correlation (where correlation means correlation coefficient throughout this paper) structure can be used to complete the uncertainty picture and allow a reasonable reliability analysis or internal linear estimation. Under this reasoning, this paper will concentrate on statistics relating to the spatial correlation structure of a soil. Although the mean and variance will be obtained along the way as part of the estimation process, these results tend to be specifically related to, and affected by, the soil type; this is true particularly of the mean. The correlation structure is believed to be more related to the formation process of the soil; that is, the correlation between soil properties at two disjoint points will be related to where the materials making up the soil at the two points originated and to the common weathering processes experienced at the two points (i.e., geological deposition processes, etc.). Thus, the major factors influencing a soil's correlation structure can be thought of as being ''external'' (i.e., related to transport and weathering rather than to chemical and mechanical details of the soil particles themselves), which is common to most soil properties and types.

Although it is undoubtedly true that many exceptions to the idea of an externally created correlation structure exist, the idea nevertheless gives a reasonable starting point for this type of investigation. It allows the correlation structure derived from a random field of a specific soil property to be used without change as a reasonable a priori correlation structure for other soil properties of interest and for other similar sites (although changes in the mean and possibly variance may, of course, be necessary).

Fundamental to the following statistical analysis is the assumption that the soil is spatially statistically homogeneous. This means that the mean, covariance, correlation structure, and higher-order moments are independent of position (and thus are the same from any reference origin). In the general case, isotropy is not usually assumed. That is, the vertical and horizontal correlation structures are allowed to be quite different. However, this is not an issue in this study because only the 1D case is considered.

The assumption of spatial homogeneity does not imply that the process is relatively uniform over any finite domain. It allows apparent trends in the mean, variance, and higher-order moments as long as those trends are just part of a larger-scale random fluctuation; that is, the mean, variance, etc. need only be constant over infinite space, not when viewed over a finite distance. (Fig. 1 may be viewed as an example of a homogeneous random field that appears nonstationary when viewed locally.) Thus, this assumption does not preclude large-scale variations, such as those often found in natural soils, although the statistics relating to the large-scale fluctuations are generally harder to estimate reliably from a finite-sampling domain.

The assumption of spatial homogeneity does, however, seem to imply that the site over which the measurements are taken is fairly uniform in geological makeup (or soil type). Again, this assumption relates to the level of uncertainty about the site for which the random model is aimed. Even changing geological units may be viewed as simply part of the overall randomness or uncertainty, which is to be characterized by the random model. The more that is known about a site, the less random the site model should be. However, the initial model that is used before significant amounts of data are explicitly gathered should be consistent with the level of uncertainty at the target site at the time the model is applied. Bayesian updating can be used to improve a prior model under additional site data.

With these thoughts in place, an appropriate inferential analysis proceeds as follows:

1. An initial regression analysis may be performed to determine if a statistically significant spatial trend is pres-

J. Geotech. Geoenviron. Eng., 1999, 125(6): 470-485

ent. Because a trend with, for example, depth may have some physical basis and may be expected to occur identically at other sites, it may make sense to predict this trend and assume it to hold at the target site. If so, the remainder of the analysis is performed on the detrended data, $\varepsilon(\mathbf{x}) = S_u(\mathbf{x}) - m(\mathbf{x})$, for example, and the trend and residual statistics must both be reported for use at the target site because they are intimately linked and cannot be considered separately. Using just the residual statistics leads to a stochastic model that is likely to be grossly in error.
2. Establish the second-moment behavior of the data set over space. Here, interest may specifically focus on whether the soil is best modeled by a finite-scale stochastic model having limited spatial correlation, or by a fractal model having significant lingering correlation over very large distances. These terms are discussed in more detail later.
3. For a selected spatial correlation function, estimate any required parameters from the data set.

By and large, probably the most common spatial correlation model in use currently is the 1D Markov model that has an exponentially decaying correlation function

$$\rho(\tau) = \exp\left(-\frac{2|\tau|}{\theta}\right) \tag{2}$$

where $\rho(\tau)$ = correlation coefficient between two points separated by a lag distance τ. In this model the parameter θ is a distance called the ''scale of fluctuation.'' It may be loosely interpreted as the separation distance beyond which soil properties are largely uncorrelated. VanMarcke (1984) defined θ to be equal to the area under the correlation function. Eq. (2) is considered to be a finite-scale model because the correlation dies out very rapidly for separation distances $\tau > \theta$; the area under this function, in particular, is finite. Such models are also called short-memory (Beran 1994). Other common finite-scale models are the Gaussian model, $\rho(\tau) = \exp\{-\pi(\tau/\theta)^2\}$, and the spherical model, $\rho(\tau) = 1 - 1.5|\tau/\theta| + 0.5|\tau/\theta|^3$, ($|\tau| \leq \theta$; $\rho(\tau) = 0$ if $|\tau| > \theta$).

An alternative model, which is rapidly gaining acceptance in a wide variety of applications, is the fractal model, also known as statistically self-similar, long-memory, and $1/f$ noise. This model has an infinite scale of fluctuation, and correlations remain significant over very large distances. An example of such a process is shown in Fig. 2. It should be noted that the samples remain statistically similar, regardless of viewing resolution, under suitable scaling of the vertical axis. Such processes are often described by the (one-sided) spectral density function

$$G(\omega) = \frac{G_o}{\omega^\gamma} \tag{3}$$

in which the parameter γ controls how the spectral power is partitioned from the low-to-high frequencies, and G_o can be viewed as a spectral intensity (white noise intensity when $\gamma = 0$). In particular, the case where $0 \leq \gamma < 1$ corresponds to infinite high frequency power and results in a stationary random process called fractional Gaussian noise (Mandelbrot 1968). When $\gamma > 1$, the spectral density falls off more rapidly at high frequencies, but grows more rapidly at low frequencies so that the infinite power is now in the low frequencies. This then corresponds to a nonstationary random process called fractional Brownian motion. Both cases are infinite-variance processes that are physically unrealizable. Their spectral densities must be truncated in some fashion to render them stationary with finite variance.

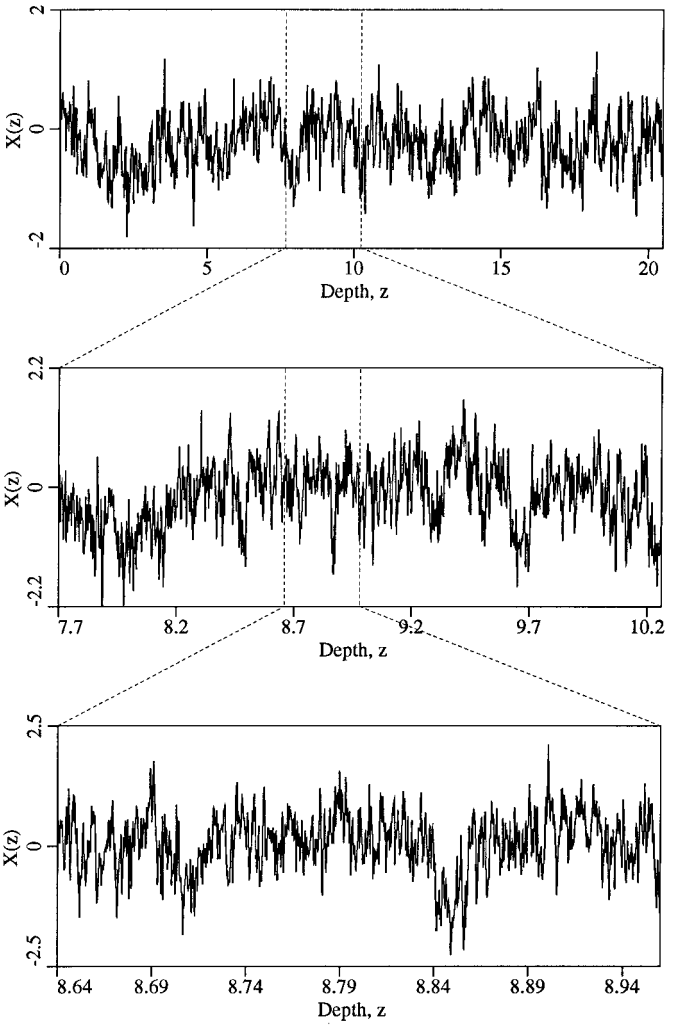

FIG. 2. Realization of Self-Similar Fractional Gaussian Noise ($H = 0.95$) Seen at Various Resolutions

Probably the best way to envisage the spectral density interpretation of a random process is to think of the random process as being composed of a number of sinusoids, each with random amplitude (power). The fractal model indicates that these random processes are made up of high amplitude, long wavelength (low frequency) sinusoids added to successively less powerful, short wavelength sinusoids. The long wavelength components provide for what are seen as trends when viewed over a finite interval. As one ''zooms'' out and views progressively more of the random process, even longer wavelength (scale) sinusoids become apparent. Conversely, as one zooms in, the short wavelength components dominate the (local) picture. This is the nature of self-similarity attributed to fractal processes—realizations of the process look the same (statistically) at any viewing scale.

By locally averaging the fractional Gaussian noise ($0 < \gamma < 1$) process over the small distance δ, Mandelbrot (1968) renders fractional Gaussian noise physically realizable with finite variance and correlation function

$$\rho(\tau) = \frac{1}{2\delta^{2H}}\left[|\tau + \delta|^{2H} - 2|\tau|^{2H} + |\tau - \delta|^{2H}\right] \tag{4}$$

where $H = (1/2)(\gamma + 1)$ is called the Hurst or self-similarity coefficient with $1/2 \leq H < 1$. The case $H = 1/2$ gives white noise, whereas $H = 1$ corresponds to perfect correlation [all $X(z) = X$, in the stationary case]. Strictly speaking, the process variance σ_X^2 is fully determined by G_o, H, and δ as $\sigma_X^2 = G_o\delta^{2H-2}\Gamma(1 - 2H)\cos(\pi H)/(\pi H)$, which goes to infinity as the

J. Geotech. Geoenviron. Eng., 1999, 125(6): 470-485

local averaging distance δ goes to zero, as expected. The local averaging is effectively a low pass filter, damping out high frequency contributions, so that Mandelbrot's approach essentially truncates the spectral density function at the high end.

Self-similarity for fractional Gaussian noise is expressed by saying that the process $X(z)$ has the same distribution as the scaled process $a^{1-H}X(az)$, for some $a > 0$. Alternatively, self-similarity for fractional Brownian motion means that $X(z)$ has the same distribution as $a^{-H}X(az)$, where the different exponent on a is due to the fact that fractional Gaussian noise is the derivative of fractional Brownian motion. Fig. 2 shows a realization of fractional Gaussian noise with $H = 0.95$ produced using the local average subdivision method (Fenton 1990). The uppermost plot is of length $n = 65{,}536$. Each plot in Fig. 2 zooms in by a factor of $a = 8$, so that each lower plots has its vertical axis stretched by a factor of $8^{0.05} = 1.11$ to appear statistically similar to the next higher plot.

In the following sections, the applicability of these models to characterize soil properties will be considered. In particular, the Second-Order Structural Analysis section takes a qualitative look at a number of tools that reveal something about the second-moment behavior of a 1D random process. The intent of that section is to evaluate these tools with respect to their ability to discern between finite scale and fractal behavior. In the Estimation of First- and Second-Order Statistical Parameters section, various maximum likelihood approaches to the estimation of the parameters for the finite-scale and fractal models are given. Finally, the results are summarized in the Conclusions with a view toward their use in developing a priori soil statistics from a large geotechnical database.

SECOND-ORDER STRUCTURAL ANALYSIS

Attention is now turned to the stochastic characterization of the soil data itself. Aside from estimating the mean and variance, the spatial correlation structure must be deduced. In the following it is assumed that the data, x_i, $i = 1, 2, \ldots, n$, are collected at a sequence of equispaced points along a line and that the best stochastic model along that line is to be found. Note that the x_i may be some suitable transformation of the actual data derived from the samples. In the following, x_i is an observation of the random process $X_i = X(z_i)$, where z is an index (commonly depth) and $z_i = (i - 1)\Delta z$, $i = 1, 2, \ldots, n$. A variety of tools will be considered in this section and their ability to identify the most appropriate stochastic model for X_i will be discussed. In particular, interest focuses on whether the process $X(z)$ is finite scaled or fractal in nature. The performance of the various tools in answering this question will be evaluated via simulation employing 2,000 realizations of finite-scale (Markov) [(2)] and fractal [(4)] processes. Each simulation is 20.48 m in length with $\Delta z = 0.02$, so that $n = 1{,}024$, and realizations are produced via covariance matrix decomposition (Fenton 1994)—a method that follows from a Choleski decomposition. Unfortunately, large covariance matrices are often nearly singular and so are difficult to decompose correctly. Because the covariance matrix for a 1D equispaced random field is symmetric and Toeplitz (i.e., the entire matrix is known if only the first column is known—all elements along each diagonal are equal), the decomposition is done using the numerically more accurate Levinson-Durbin algorithm [see Marple (1987) and Brockwell and Davis (1987)].

Sample Correlation Function

The classical sample average of x_i is computed as

$$\hat{\mu}_x = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{5}$$

and the sample variance as

$$\hat{\sigma}_x^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \hat{\mu}_x)^2 \tag{6}$$

This estimator is biased because it is not divided by $n - 1$ as is usually seen. (A biased estimator is one whose expected value is not equal to the parameter it purports to estimate.) The use of a biased estimator here is for three reasons:

1. The expected error variance is (slightly) smaller than that for the biased case.
2. The biased estimator, when estimating covariances, leads to a tractable nonnegative definite covariance matrix.
3. It is currently the most popular variance estimator in time series analysis (Priestley 1981).

The main reason for its popularity is probably due to its nonnegative definiteness. The covariance $C(\tau)$ between $X(z)$ and $X(z + \tau)$ is estimated, using a biased estimator for reasons discussed above, as

$$\hat{C}(\tau_j) = \frac{1}{n}\sum_{i=1}^{n-j+1} (x_i - \hat{\mu}_x)(x_{i+j-1} - \hat{\mu}_x), \quad j = 1, 2, \ldots, n \tag{7}$$

where the lag $\tau_j = (j - 1)\Delta z$. Notice that $\hat{C}(0)$ is the same as the estimated variance $\hat{\sigma}_x^2$. The sample correlation is obtained by normalizing

$$\hat{\rho}(\tau_j) = \frac{\hat{C}(\tau_j)}{\hat{C}(0)} \tag{8}$$

One of the major difficulties with the sample correlation function resides in the fact that it is heavily dependent on the estimated mean $\hat{\mu}_x$. When the soil shows significant long-scale dependence, characterized by long-scale fluctuations (see, e.g., Fig. 1), then $\hat{\mu}_x$ is almost always a poor estimate of the true mean. In fact, it is not too difficult to show that although the mean estimator [(5)] is unbiased, its variance is given by

$$\mathrm{Var}[\hat{\mu}_x] = \left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \rho(\tau_{i-j})\right]\sigma_x^2 = \gamma_n\sigma_x^2 \simeq \gamma(D)\sigma_x^2 \tag{9}$$

where $D = (n - 1)\Delta z$ = sampling domain size (interpreted as the region defined by n equisized ''cells,'' each of width Δz centered on an observation); $\gamma(D)$ = variance function (VanMarcke 1984) that gives the variance reduction due to averaging over the length D

$$\gamma(D) = \frac{1}{D^2}\int_0^D\int_0^D \rho(\tau - s)\, d\tau\, ds = \frac{2}{D^2}\int_0^D (T - \tau)\rho(\tau)\, d\tau \tag{10}$$

The discrete approximation to the variance function, denoted γ_n in (9), approaches $\gamma(D)$ as n becomes large. For highly correlated soil samples (over the sampling domain), $\gamma(D)$ remains close to 1.0, so that $\hat{\mu}_x$ remains highly variable, almost as variable as $X(z)$ itself. Notice that the variance of $\hat{\mu}_x$ is unknown because it depends on the unknown correlation structure of the process.

In addition, it can be shown that $\hat{C}(\tau_j)$ is biased according to the following (VanMarcke 1984):

$$\mathrm{E}[\hat{C}(\tau_j)] \simeq \sigma_x^2\left(\frac{n - j + 1}{n}\right)[\rho(\tau_j) - \gamma(D)] \tag{11}$$

where, again, the approximation improves as n increases. From this it can be seen that

$$\mathrm{E}[\hat{\rho}(\tau_j)] \simeq \frac{\mathrm{E}[\hat{C}(\tau_j)]}{\mathrm{E}[\hat{C}(0)]} \simeq \left(\frac{n - j + 1}{n}\right)\left(\frac{\rho(\tau_j) - \gamma(D)}{1 - \gamma(D)}\right) \tag{12}$$

using a first-order approximation. For soil samples that show considerable serial correlation, $\gamma(D)$ may remain close to 1,

J. Geotech. Geoenviron. Eng., 1999, 125(6): 470-485

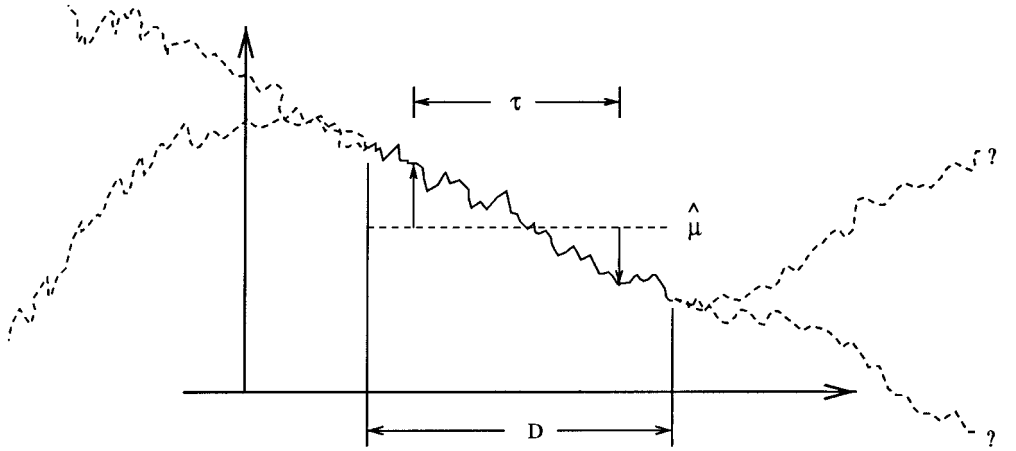

FIG. 3. Covariance Estimator on Strongly Dependent Finite Sample

and generally the term $\rho(\tau_j) - \gamma(D)$ will become negative for all $j^* \leq j \leq n$ for some j^* less than n. What this means is that the estimator $\hat{\rho}(\tau)$ will typically dip below zero even when the field is actually highly positively correlated.

Another way of looking at this problem is as follows. Consider the sample shown in Fig. 1 and assume that the apparent trend is in fact just part of a long-scale fluctuation. Clearly, Fig. 1 is then a process with a large scale of fluctuation, compared with D. The local average $\hat{\mu}_X$ is estimated and shown by the dashed line in Fig. 3. Now, for any τ greater than about half of the sampling domain, the product of the deviations from $\hat{\mu}_X$ in (7) will be negative. This means that the sample correlation function will decrease rapidly and become negative somewhere before $\tau = (1/2)D$ even though the true correlation function may remain much closer to 1.0 throughout the sample.

It should be noted that if the sample does in fact come from a short-scale process, with $\theta << D$, the variability of (9) and the bias of (12) largely disappear because $\gamma(D) \simeq 0$. This means that the sample correlation function is a good estimator of short-scale processes as long as $\theta << D$. However, if the process does in fact have long-scale dependence, then the correlation function cannot identify this and in fact continues to illustrate short-scale behavior. In essence, the estimator is analogous to a self-fulfilling prophecy: It always appears to justify its own assumptions.

Fig. 4 illustrates the situation graphically using simulations from finite-scale and fractal processes. The dashed lines show the maximum and minimum correlations observed at each lag over the 2,000 realizations. The finite-scale ($\theta = 3$) simulation shows reasonable agreement between $\hat{\rho}(\tau)$ and the true correlation because $\theta << D \simeq 20$. However, for the fractal process ($H = 0.95$) there is a very large discrepancy between the estimated average and true correlation functions. Clearly, the sample correlation function fails to provide any useful information about large-scale or fractal processes.

Sample Semivariogram

The semivariogram [half of the variogram, as defined by Matheron (1962)] gives essentially the same information as the correlation function because, for stationary processes, they are related according to

$$V(\tau_j) = \frac{1}{2} \mathrm{E}[(X_{i+j} - X_i)^2] = \sigma_X^2(1 - \rho(\tau)) \tag{13}$$

The sample semivariogram is defined by

$$\hat{V}(\tau_j) = \frac{1}{2(n-j)} \sum_{i=1}^{n-j} (x_{i+j} - x_i)^2, \quad j = 0, 1, \ldots, n-1 \tag{14}$$

The major difference between $\hat{V}(\tau_j)$ and $\hat{\rho}(\tau_j)$ is that the semivariogram does not depend on $\hat{\mu}_X$. This is a clear advantage because many of the problems of the correlation function relate to this dependence. In fact, it is easily shown that the semivariogram is an unbiased estimator with $\mathrm{E}[\hat{V}(\tau_j)] = (1/2)\mathrm{E}[(X_{i+j} - X_i)^2]$. Fig. 5 shows how this estimator behaves for finite-scale and fractal processes. Notice that the finite-scale semivariogram rapidly increases to its limiting value (the variance) and then flattens out whereas the fractal process leads to a semivariogram that continues to increase gradually throughout. This behavior can indicate the underlying process type and allow identification of a suitable correlation model. Noteworthy, however, is the very wide range between the observed minimum and maximum (the maximum going off the plot but having maximum values in the range from 5 to 10 in both cases). The high variability in the semivariogram may hinder its use in discerning between model types unless sufficient averaging can be performed.

The semivariogram finds its primary use in mining geostatistics applications [see, e.g., Journel and Huijbregts (1978)]. Cressie (1993) discussed some of its distributional characteristics along with robust estimation issues, but little is known about the distribution of the semivariogram when $X(z)$ is spatially dependent. Without the estimator distribution, the semivariogram cannot easily be used to test rigorously between competing model types (as in fractal versus finite scale), nor can it be used to fit model parameters using the maximum likelihood method. The latter is one of the goals of this paper. For these reasons, the semivariogram will not be pursued further in this paper, although indications are that such a pursuit may be warranted.

Sample Variance Function

The variance function measures the decrease in the variance of an average as an increasing number of sequential random variables are included in the average. If the local average of a random process X_D is defined by

$$X_D = \frac{1}{D} \int_0^D X(z)\, dz \tag{15}$$

then the variance of X_D is just $\gamma(D)\sigma_X^2$. In the discrete case, which will be used here, this becomes

$$\bar{X} = \hat{\mu}_X = \frac{1}{n} \sum_{i=1}^{n} X(z_i) = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{16}$$

where $\mathrm{Var}[\bar{X}] = \gamma_n \sigma_X^2$; and γ_n, defined by (9), is the discrete approximation of $\gamma(D)$ (see Sample Correlation Function sub-

J. Geotech. Geoenviron. Eng., 1999, 125(6): 470-485

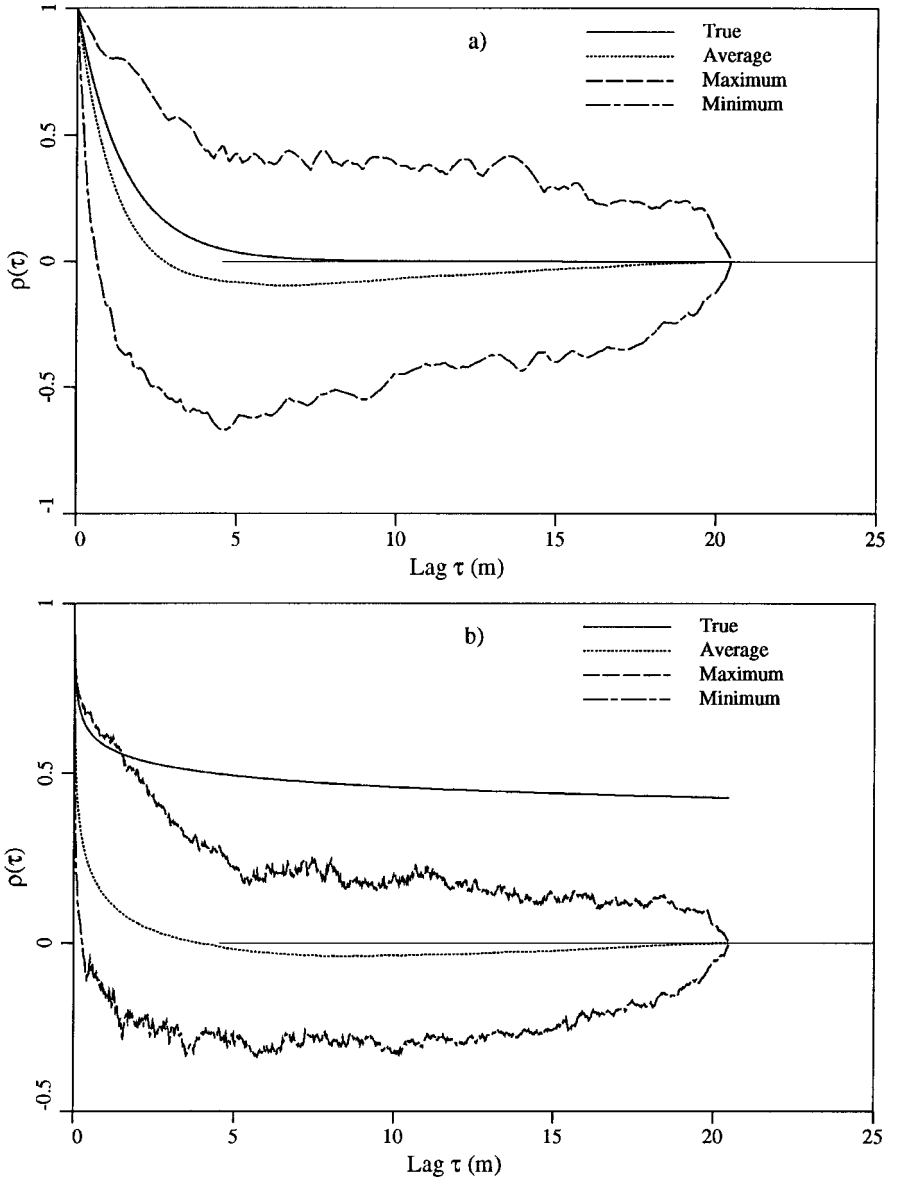

FIG. 4. Sample Correlation Functions for Models Averaged Over 2,000 Realizations: (a) Finite Scale (θ = 3); (b) Fractal (*H* = 0.95)

section). If the X_i values are independent and identically distributed then $\gamma_n = 1/n$, whereas if $X_1 = X_2 = \cdots = X_n$, then the values of X are completely correlated and $\gamma_n = 1$ so that averaging does not lead to any variance reduction. In general, for correlation functions that remain nonnegative, $1/n \leq \gamma_n \leq 1$.

Conceptually, the rate at which the variance of an average decreases with averaging size states the spatial correlation structure. In fact these are equivalent because in the 1D (continuous case)

$$\gamma(D) = \frac{2}{D^2}\int_0^D (D - \tau)\rho(\tau)\, d\tau \Leftrightarrow \rho(\tau) = \frac{1}{2}\frac{\partial^2}{\partial\tau^2}[\tau^2\gamma(\tau)] \quad (17)$$

Given a sequence of n equispaced observations over a sampling domain of size $D = n\Delta z$ the sample (discrete) variance function is estimated to be

$$\hat{\gamma}_i = \frac{1}{\hat{\sigma}_X^2(n - i + 1)}\sum_{j=1}^{n-i+1}(X_{i,j} - \hat{\mu}_X)^2, \quad i = 1, 2, \ldots, n \quad (18)$$

where $X_{i,j}$ = local average as follows:

$$X_{i,j} = \frac{1}{i}\sum_{k=j}^{j+i-1} X_k, \quad j = 1, 2, \ldots n - i + 1 \quad (19)$$

It should be noted that $\hat{\gamma}_1 = 1$ as the sum of (18) is the same as that used to find $\hat{\sigma}_X^2$ when $i = 1$. Also, when $i = n$, the sample variance function $\hat{\gamma}_n = 0$ because $X_{n,j} = X_{n,1} = \hat{\mu}_X$. Thus, the sample variance function always connects the points $\gamma_1 = 1$ and $\gamma_n = 0$.

Unfortunately, the sample variance function is biased, and its bias depends on the degree of correlation between observations. Specifically, it can be shown that

$$E[\hat{\gamma}_i] \simeq \frac{\gamma_i - \gamma_n}{1 - \gamma_n} \quad (20)$$

using a first-order approximation. This becomes unbiased as $n \to \infty$ only if $D = (n - 1)\Delta z \to \infty$ and $\gamma(D) \to 0$ as well. In other words, we need both the averaging region to grow large and the correlation function to decrease sufficiently rapidly within the averaging region for the sample variance function to become unbiased.

Figs. 6(a and b) show sample variance functions averaged over 2,000 simulations of finite-scale and fractal random processes, respectively. There is very little difference between the estimated variance function in the two plots, despite the fact that they come from quite different processes. Clearly, the estimate of the variance function in the fractal case is highly

J. Geotech. Geoenviron. Eng., 1999, 125(6): 470-485

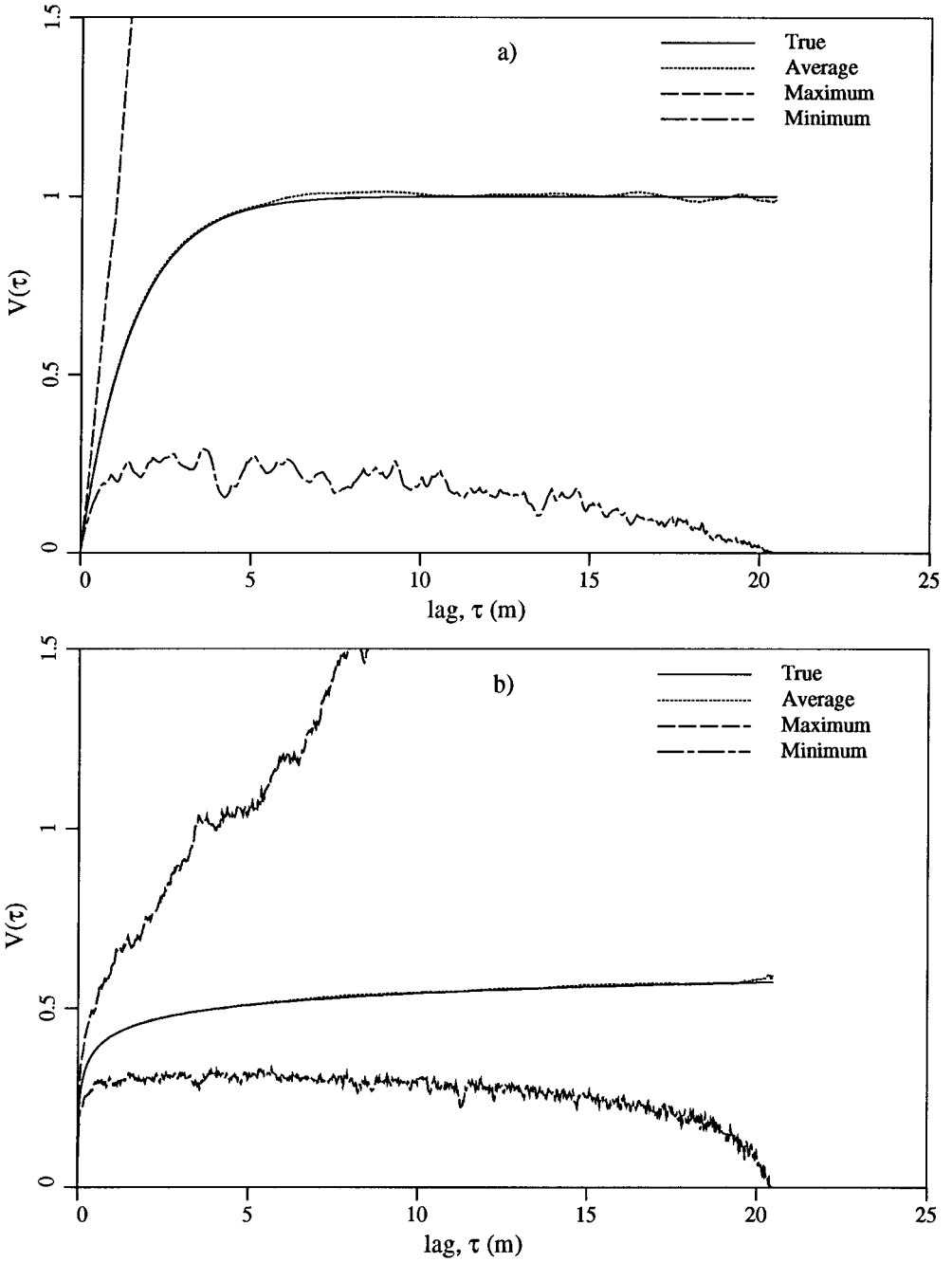

FIG. 5. Variograms for Models Averaged Over 2,000 Realizations: (a) Finite Scale (θ = 3); (b) Fractal (*H* = 0.95)

biased. Thus, the variance function plot does not appear to be a good identification tool and is really only a useful estimate of second-moment behavior for the finite-scale process with $\theta << D$. In the plots, $T = i\Delta z$ for $i = 1, 2, \ldots, n$.

Wavelet Coefficient Variance

The wavelet basis has attracted much attention in recent years in areas of signal analysis, image compression, and among other things, fractal process modeling (Wornell 1996). It can basically be viewed as an alternative to Fourier decomposition except that sinusoids are placed by ''wavelets'' that act only over a limited domain. 1D wavelets are usually defined as translations along the real axis and dilations (scalings) of a ''mother wavelet,'' as in

$$\psi_j^m(z) = 2^{m/2}\psi(2^m z - j) \tag{21}$$

where m and j = dilation and translation indices, respectively. The appeal to using wavelets to model fractal processes is that they are self-similar in nature [i.e., all wavelets look the same when viewed at the appropriate scale (which, in the above definition, is some power of 2)]. The random process $X(z)$ is then expressed as a linear combination of various scalings, translations, and dilations of a common ''shape.'' Specifically

$$X(z) = \sum_m \sum_j X_j^m \psi_j^m(z) \tag{22}$$

If the wavelets are suitably selected to be orthonormal, then the coefficients can be found through the inversion

$$X_j^m = \int_{-\infty}^{\infty} X(z)\psi_j^m(z)\, dz \tag{23}$$

for which highly efficient numerical solution algorithms exist. The details of the wavelet decomposition will not be discussed in this paper. [The interested reader should see, for example, Strang and Nguyen (1996).]

A theorem by Wornell (1996) states that, under reasonably

J. Geotech. Geoenviron. Eng., 1999, 125(6): 470-485

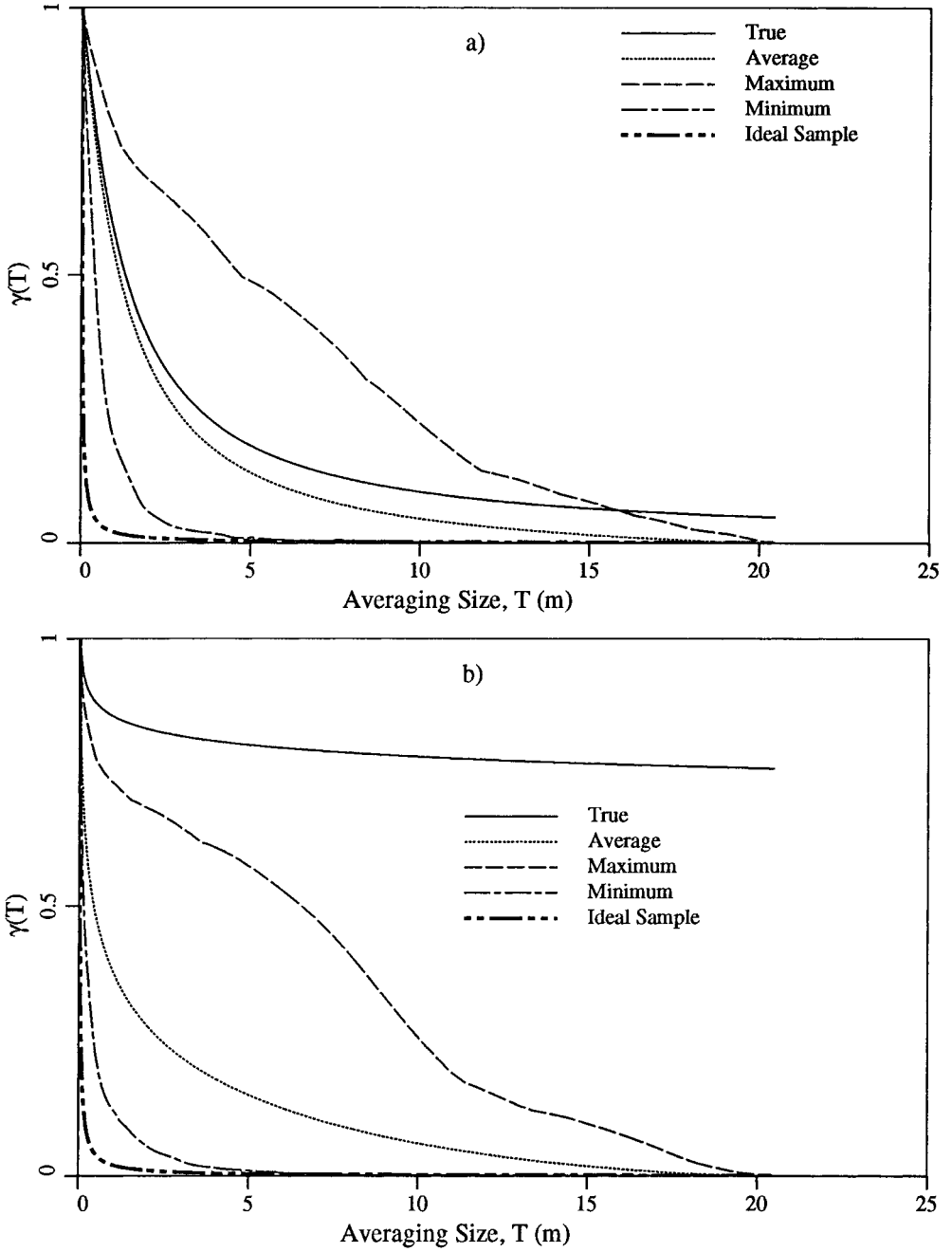

FIG. 6. Sample Variance Functions for Models Averaged Over 2,000 Realizations: (a) Finite Scale (θ = 1); (b) Fractal (*H* = 0.98)

general conditions, if the coefficients X_j^m are mutually uncorrelated, zero-mean random variables with variances

$$\sigma_m^2 = \mathrm{Var}[X_j^m] = \sigma^2 2^{-\gamma m} \tag{24}$$

then $X(z)$ obtained through (22) will have a spectrum that is very nearly fractal. Furthermore, Wornell made theoretical and simulation based arguments showing that the converse is also approximately true, namely, that if $X(z)$ is fractal with spectral density proportional to $\omega^{-\gamma}$, then the coefficients X_j^m will be approximately uncorrelated with variance given by (24). If this is the case, then a plot of $\ln(\mathrm{Var}[X_j^m])$ versus the scale index m will be a straight line.

Using a fifth-order Daubechies wavelet basis, Fig. 7 shows a plot of the estimated wavelet coefficient variances $\hat{\sigma}_m^2$, where

$$\hat{\sigma}_m^2 = \frac{1}{2^{m-1}} \sum_{j=1}^{2^{m-1}} (x_j^m)^2 \tag{25}$$

against the scale index m for the finite-scale and fractal simulation cases. The fractal simulations yield a straight line, as expected, whereas the finite-scale simulations show a slight flattening of the variance at lower values of m (larger scales). The lowest value of $m = 1$ is not plotted because the variance of this estimate is very large, and it appears to suffer from the same bias as estimates of the spectral density function at $\omega = 0$ (as discussed later). On the basis of Fig. 7, it appears that the wavelet coefficient variance plot may have some potential in identifying an appropriate stochastic model, although the difference in the plots is quite small. Confident conclusions will require a large data set.

If it turns out that $X(z)$ is fractal and Gaussian, then the coefficients X_j^m are also Gaussian and (largely) uncorrelated as discussed above. This means that a maximum likelihood estimation can be performed to evaluate the spectral exponent γ by looking at the likelihood of the computed set of coefficients x_j^m.

J. Geotech. Geoenviron. Eng., 1999, 125(6): 470-485

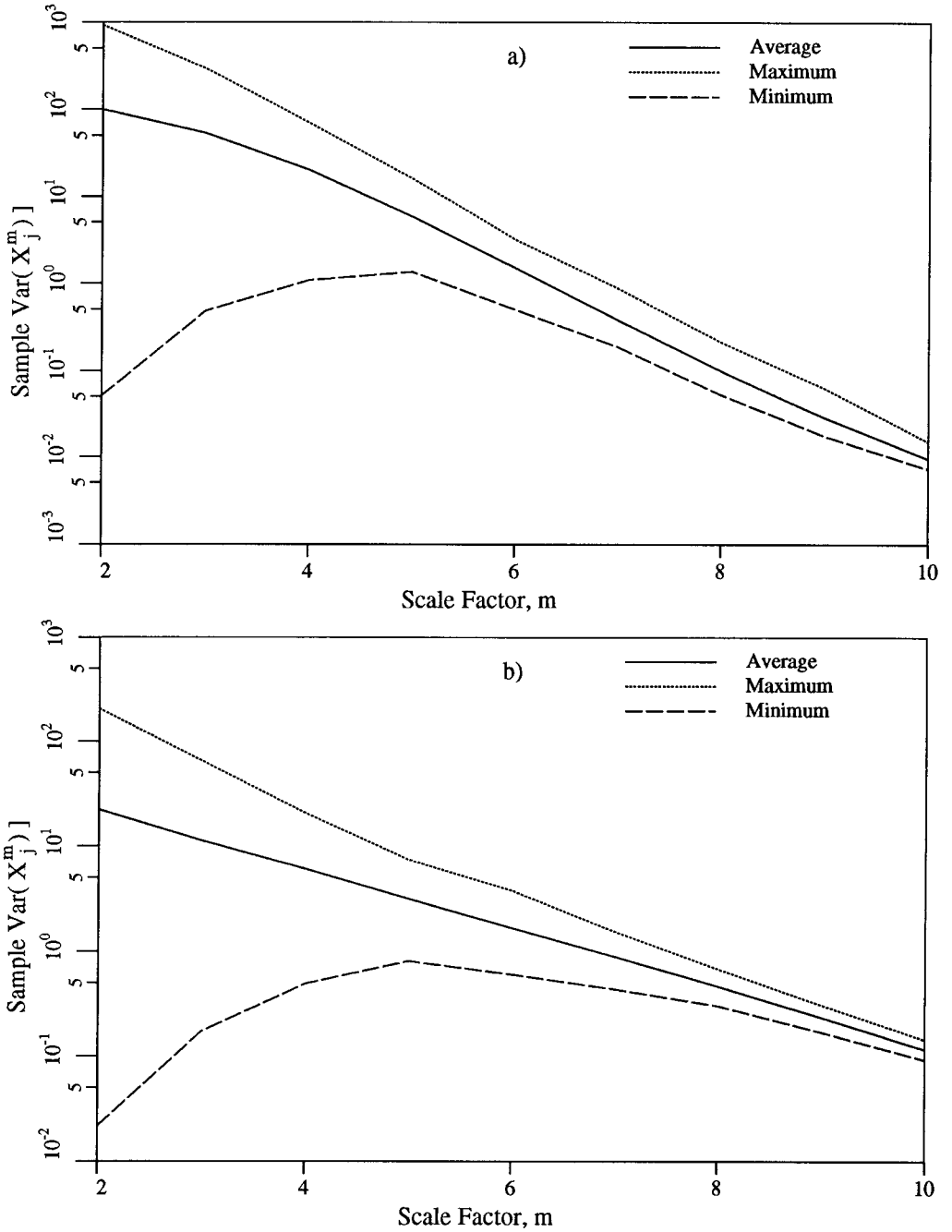

FIG. 7. Sample Wavelet Coefficient Variance Functions for Models Averaged Over 2,000 Realizations: (a) Finite Scale (θ = 3); (b) Fractal (*H* = 0.95)

Sample Spectral Density Function

The sample spectral density function, referred to here also as the periodogram despite its slightly nonstandard form, is obtained by first computing the Fourier transform of the data

$$\chi(\omega_j) = \frac{1}{n}\sum_{k=0}^{n-1} X(t_k)e^{-i\omega_j k} \tag{26}$$

at each Fourier frequency $\omega_j = 2\pi j/D$, $j = 0, 1, \ldots, (n-1)/2$. This is efficiently achieved using the fast Fourier transform. The periodogram is then given by the squared magnitude of the complex Fourier coefficients according to

$$\hat{G}(\omega_j) = \frac{D}{\pi}|\chi(\omega_j)|^2 \tag{27}$$

where $D = n\Delta z$. For stationary processes with finite variance, the periodogram estimates as defined here are independent and exponentially distributed with means equal to the true one-sided spectral density $G(\omega_j)$ [see Beran (1994)]. VanMarcke (1984) showed that the periodogram itself has a nonzero scale of fluctuation when $D = n\Delta z$ is finite, equal to $2\pi/D$. This suggests the presence of serial correlation between periodogram estimators. However, because the periodogram estimates at Fourier frequencies are separated by $2\pi/D$, they are, therefore, approximately independent, according to the physical interpretation of the scale of fluctuation distance. The independence and distribution have also been shown by Yajima (1989) to hold for both fractal and finite-scale processes. Armed with this distribution on the periodogram estimates, one can perform maximum likelihood estimation as well as (conceptually) hypothesis tests. If the periodogram is smoothed using some sort of smoothing window, as discussed by Priestley (1981), the smoothing may lead to loss of independence between estimates at sequential Fourier frequencies so that likelihood ap-

J. Geotech. Geoenviron. Eng., 1999, 125(6): 470-485

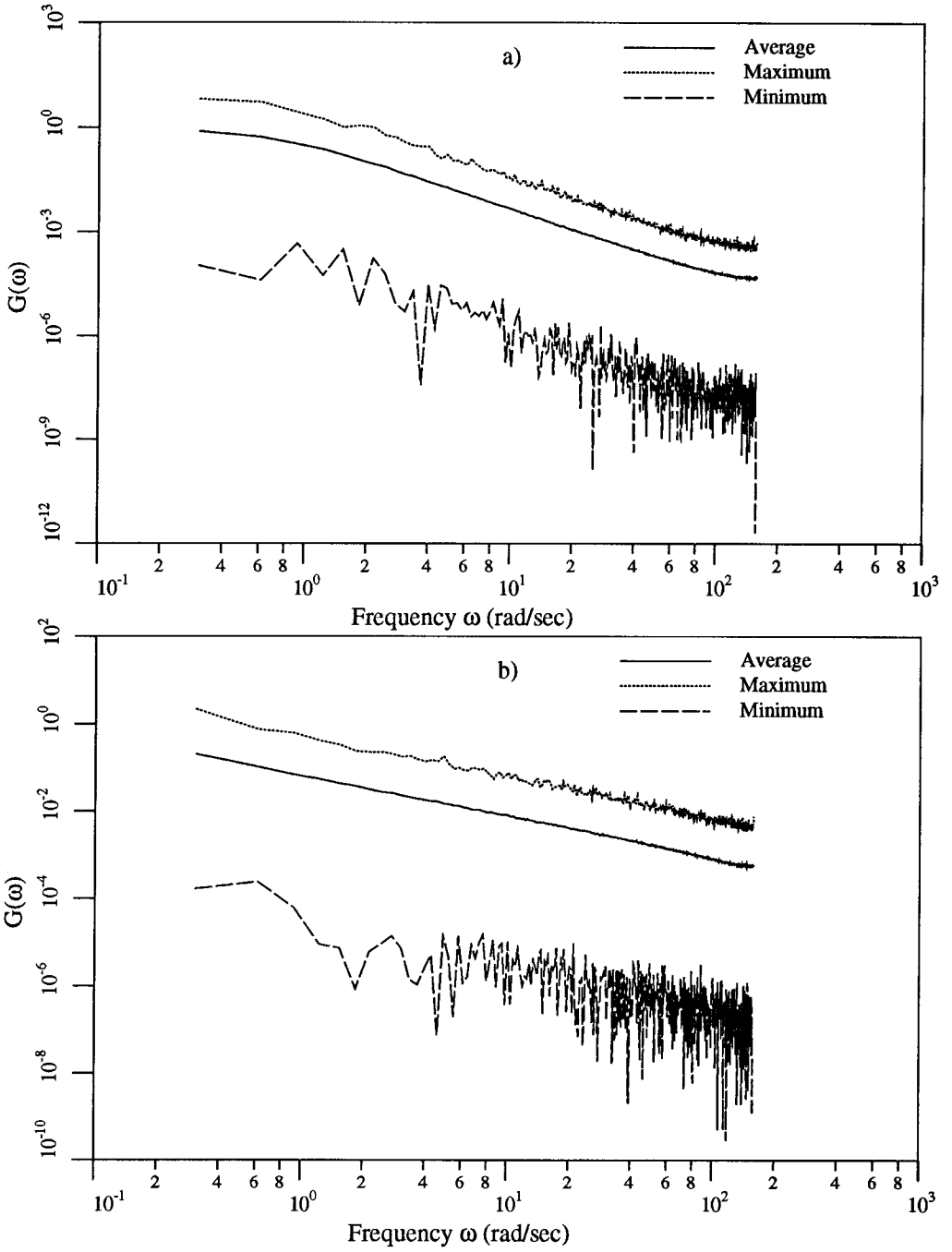

FIG. 8. Periodograms for Models Averaged Over 2,000 Realizations: (a) Finite Scale (θ = 3); and (b) Fractal (*H* = 0.95)

proaches become complicated. In this sense, it is best to smooth the periodogram (which is notoriously rough) by averaging over an ensemble of periodogram estimates taken from a sequence of realizations of the random process, where available.

It should be noted that the periodogram estimate at $\omega = 0$ is not a good estimator of $G(0)$, and so it should not be included in the periodogram plot. In fact, the periodogram estimate at $\omega = 0$ is biased with $\mathrm{E}[\hat{G}(0)] = G(0) + n\mu_X^2/(2\pi)$ (Brockwell and Davis 1987). Recalling that μ_X is unknown, and its estimate highly variable when strong correlation exists, the estimate $\hat{G}(0)$ should not be trusted. In addition, its distribution is no longer a simple exponential.

The easiest way to determine whether the data are fractal in nature is to look directly at a plot of the periodogram. Fractal processes have spectral density functions of the form $G(\omega) \propto \omega^{-\gamma}$ for $\gamma > 0$. Thus, $\ln G(\omega) = c - \gamma \ln \omega$, for some constant c, so that a log-log plot of the sample spectral density function of a fractal process will be a straight line with a slope of $-\gamma$. Fig. 8 illustrates how the periodogram behaves when averaged over both fractal and finite-scale simulations. The periodogram is a straight line with a negative slope in the fractal case and becomes more flattened at the origin in the finite-scale case, as was observed for the wavelet variance plot. Again, the difference is only slight, so that a fairly large data set is required to decide on a model with any degree of confidence.

ESTIMATION OF FIRST- AND SECOND-ORDER STATISTICAL PARAMETERS

Upon deciding on whether a finite-scale or fractal model is more appropriate in representing the soil data, the next step is to estimate the pertinent parameters. In the case of the finite-scale model, the parameter of interest is the scale of fluctuation

J. Geotech. Geoenviron. Eng., 1999, 125(6): 470-485

θ. For fractal models, the parameter of interest is the spectral exponent γ, or, equivalently for $0 \leq \gamma < 1$, the self-similarity parameter $H = (1/2)(\gamma + 1)$.

Finite-Scale Model

If the process is deemed to be finite-scale in nature, then a variety of techniques are available to estimate θ, some of which are as follows:

- Directly compute the area under the sample correlation function. This is a nonparametric approach, although it assumes that the scale is finite and that the correlation function is monotonic. The area is usually taken to be the area up to when the function first becomes negative (the scale of fluctuation is not well defined for oscillatory correlation functions—other parameters may be more appropriate if this is the case). It should also be noted that correlation estimates lying within the band $\pm 2n^{-1/2}$ are commonly deemed to be not significantly different than zero [see Priestley (1981, p. 340) and Brockwell and Davis (1987, Chapter 7)].
- Use regression to fit a correlation function to $\hat{\rho}(\tau)$ or a semivariogram to $\hat{V}(\tau)$. For certain assumed correlation or semivariogram functions, this regression may be nonlinear in θ.
- If the sampling domain D is deemed to be much larger than the scale of fluctuation, then the scale can be estimated from the variance function using an iterative technique such as that suggested by VanMarcke (1984, p. 337).
- Assuming a joint distribution for $X(t_j)$ with corresponding correlation function model, estimate unknown parameters (μ_X, σ_X^2, and correlation function parameters) using maximum likelihood in the space domain.
- Using the established results regarding the joint distribution for periodogram estimates at the set of Fourier frequencies, an assumed spectral density function can be "fit" to the periodogram using maximum likelihood.

Because of the reasonably high bias in the sample correlation (or covariance) function estimates, even for finite-scale processes, using the sample correlation directly to estimate θ will not be pursued here. The variance function techniques have not been found by the writer to be particularly advantageous over, for example, the maximum likelihood approaches and are also prone to error due to their high bias in long-scale or fractal cases. Here, the maximum likelihood estimator in the space domain will be discussed briefly. The maximum likelihood estimator in the frequency domain will be considered later.

It will be assumed that the data are normally distributed, or have been transformed from their raw state into something that is at least approximately normally distributed. For example, many soil properties are commonly modeled using the lognormal distribution, often primarily because this distribution is strictly nonnegative. To convert lognormally distributed data to a normal distribution, it is sufficient merely to take the natural logarithm of the data prior to further statistical analysis. It should be noted that the normal model is commonly used for at least two reasons: (1) It is analytically tractable in many ways; and (2) it is completely defined through knowledge of the first two moments, namely the mean and covariance structure. Because other distributions commonly require higher moments and because higher moments are generally quite difficult to estimate accurately, particularly in the case of geotechnical samples that are typically limited in size, the use of other distributions is often difficult to justify. The normal assumption can be thought of as a minimum knowledge assumption that succinctly expresses the first two moments of the random field, where it is hoped that even these can be estimated with some confidence.

Because the data are assumed to be jointly normally distributed, the space domain maximum likelihood estimators are obtained by maximizing the likelihood of observing the spatial data under the assumed joint distribution. The likelihood of observing the sequence of observations $\mathbf{x}^T = \{x_1, x_2, \ldots, x_n\}$ (superscript T denotes the vector or matrix transpose) given the distributional parameters $\boldsymbol{\phi}^T = \{\mu_X, \sigma_X^2, \theta\}$ is

$$L(\mathbf{x}|\boldsymbol{\phi}) = \frac{1}{(2\pi)^{n/2}|\mathbf{C}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T\mathbf{C}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right\} \quad (28)$$

where $\mathbf{C}$ = covariance matrix between the observation; $C_{ij} = \mathrm{E}[(X_i - \mu_i)(X_j - \mu_j)]$; $|\mathbf{C}|$ = determinant of $\mathbf{C}$; and $\boldsymbol{\mu}$ = vector of means corresponding to each observation location. In the following, the data are assumed to be modeled by a stationary random field, so that the mean is spatially constant and $\boldsymbol{\mu} = \mu_X\mathbf{1}$, where $\mathbf{1}$ is a vector with all elements equal to 1. Again, if this assumption is not deemed warranted, a deterministic trend in the mean and variance can be removed from the data prior to the following statistical analysis through the transformation

$$x'(z) = \frac{x(z) - m(z)}{s(z)}$$

where $m(z)$ and $s(z)$ = deterministic spatial trends in the mean and standard deviation, respectively, possibly obtained by regression analysis of the data. Recall, however, that this is generally only warranted if the same trends are expected at the target site.

Also due to the stationarity assumption, the covariance matrix can be written in terms of the correlation matrix $\boldsymbol{\rho}$ as

$$\mathbf{C} = \sigma_X^2\boldsymbol{\rho} \quad (29)$$

where $\boldsymbol{\rho}$ = function only of the unknown correlation function parameter θ. If the correlation function has more than one parameter, then θ is treated as a vector of unknown parameters and the maximum likelihood will generally be found via a gradient or grid search in these parameters. With (29), the likelihood function of (28) can be written as

$$L(\mathbf{x}|\boldsymbol{\phi}) = \frac{1}{(2\pi\sigma_X^2)^{n/2}|\boldsymbol{\rho}|^{1/2}} \exp\left\{-\frac{(\mathbf{x} - \boldsymbol{\mu})^T\boldsymbol{\rho}^{-1}(\mathbf{x} - \boldsymbol{\mu})}{2\sigma_X^2}\right\} \quad (30)$$

Because the likelihood function is strictly nonnegative, maximizing $L(\mathbf{x}|\boldsymbol{\phi})$ is equivalent to maximizing its logarithm, which, ignoring constants, is given by

$$\mathcal{L}(\mathbf{x}|\boldsymbol{\phi}) = -\frac{n}{2}\ln\sigma_X^2 - \frac{1}{2}\ln|\boldsymbol{\rho}| - \frac{(\mathbf{x} - \boldsymbol{\mu})^T\boldsymbol{\rho}^{-1}(\mathbf{x} - \boldsymbol{\mu})}{2\sigma_X^2} \quad (31)$$

The maximum of (31) can in principle be found by differentiating with respect to each unknown parameter, μ_X, σ_X^2, and θ, in turn and setting the results to zero. This gives three equations in three unknowns. The partial derivative of $\mathcal{L}$ with respect to μ_X, when set equal to zero, leads to the following estimator for the mean

$$\hat{\mu}_X = \frac{\mathbf{1}^T\boldsymbol{\rho}^{-1}\mathbf{x}}{\mathbf{1}^T\boldsymbol{\rho}^{-1}\mathbf{1}} \quad (32)$$

Because this estimator still involves the unknown correlation matrix, it should be viewed as the value of μ_X that maximizes the likelihood function for a given value of the correlation parameter θ. If the two vectors $\mathbf{r}$ and $\mathbf{s}$ are solutions of the two systems of equations

$$\boldsymbol{\rho}\mathbf{r} = \mathbf{x}; \quad \boldsymbol{\rho}\mathbf{s} = \mathbf{1} \quad (33a,b)$$

J. Geotech. Geoenviron. Eng., 1999, 125(6): 470-485

then the estimator for the mean can be written as

$$\hat{\mu}_X = \frac{\mathbf{1}^T\mathbf{r}}{\mathbf{1}^T\mathbf{s}} \tag{34}$$

It should be noted that this estimator is generally not very different from the usual estimator obtained by simply averaging the observations (Beran 1994). It should also be noted that $\mathbf{1}^T\mathbf{r} = \mathbf{x}^T\mathbf{s}$, so that only $\mathbf{s}$ needs to be found to compute $\hat{\mu}_X$. However, $\mathbf{r}$ will be needed in (35), so it should be found anyhow.

The partial derivative of $\mathcal{L}$ with respect to σ_X^2, when set equal to zero, leads to the following estimator for σ_X^2

$$\hat{\sigma}_X^2 = \frac{1}{n}(\mathbf{x} - \hat{\mu}_X\mathbf{1})^T\mathbf{r} \tag{35}$$

which is also implicitly dependent on the correlation function parameter through $\hat{\mu}_X$ and $\mathbf{r}$.

Thus, both the mean and variance estimators can be expressed in terms of the unknown parameter θ. Using these results, the maximization problem simplifies to finding the maximum of

$$\mathcal{L}(\mathbf{x}|\boldsymbol{\phi}) = -\frac{n}{2}\ln\hat{\sigma}_X^2 - \frac{1}{2}\ln|\boldsymbol{\rho}| \tag{36}$$

where the last term in (31) became simply $n/2$ and was dropped from (36) because it does not affect the location of the maximum.

In principle, (36) need now only be differentiated with respect to θ, and the result set to zero to yield the optimal estimate $\hat{\theta}$ and subsequently $\hat{\sigma}_X^2$ and $\hat{\mu}_X$. Unfortunately, this involves differentiating the determinant of the correlation matrix, and closed form solutions do not always exist. Solution of the maximum likelihood estimators may therefore proceed by iteration as follows:

1. Guess at an initial value for θ.
2. Compute the corresponding correlation matrix elements $\rho_{ij} = \rho(|z_i - z_j|)$, which in the current equispaced 1D case is both symmetric and Toeplitz (i.e., elements along each diagonal are equal).
3. Solve (33) for vectors $\mathbf{r}$ and $\mathbf{s}$.
4. Solve for the determinant of $\boldsymbol{\rho}$ (because this is often vanishing, it is usually better to compute the log-determinant directly to avoid numerical underflow).
5. Compute the mean and variance estimates using (35) and (34).
6. Compute the log-likelihood value $\mathcal{L}$ using (36).
7. Guess at a new value for θ and repeat Steps 2–7 until the global maximum value of $\mathcal{L}$ is found.

Guesses for θ can be arrived at simply by stepping discretely through a likely range (and the speed of modern computers make this a reasonable approach), increasing the resolution in the region of located maxima. Alternatively, more sophisticated techniques may be employed that look also at the magnitude of the likelihood in previous guesses. One advantage to the brute force approach of stepping along at predefined increments is that it is more likely to find the global maximum in the event that multiple local maxima are present. With the speed of modern computers, this approach has been found to be acceptably fast for a low number of unknown parameters, for instance, less than four, and where bounds on the parameters are approximately known.

For large samples, the correlation matrix $\boldsymbol{\rho}$ can become nearly singular, so that numerical calculations become unstable. In the 1D case, the Durbin-Levinson recursion, taking full advantage of the Toeplitz character of $\boldsymbol{\rho}$, yields a faster and more accurate decomposition and allows the direct computation of the log-determinant as part of the solution [see, e.g., Marple (1987)].

One finite-scale model that has a particularly simple maximum likelihood formulation is the jointly normal distribution with the Markov correlation function given by (2). When observations are equispaced, the correlation matrix has a simple closed form determinant and a tridiagonal inverse

$$|\boldsymbol{\rho}| = (1 - q^2)^{n-1} \tag{37}$$

$$\boldsymbol{\rho}^{-1} = \left(\frac{1}{1-q^2}\right) \cdot \begin{bmatrix} 1 & -q & 0 & 0 & \cdots & 0 & 0 \\ -q & 1+q^2 & -q & 0 & \cdots & 0 & 0 \\ 0 & -q & 1+q^2 & -q & \cdots & 0 & 0 \\ 0 & 0 & -q & 1+q^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1+q^2 & -q \\ 0 & 0 & 0 & 0 & \cdots & -q & 1 \end{bmatrix} \tag{38}$$

where $q = \exp\{-2\Delta z/\theta\}$ for observations spaced Δz apart. Using these results, the maximum likelihood estimation of q reduces to finding the root of the cubic equation

$$f(q) = b_0 + b_1 q + b_2 q^2 + b_3 q^3 = 0 \tag{39}$$

on the interval $q \in (0, 1)$, where

$$b_0 = nR_1; \quad b_1 = -(R_0 + nR_0'); \quad b_2 = -(n-2)R_1 \tag{40a–c}$$

$$b_3 = (n-1)R_0'; \quad R_0 = \sum_{i=1}^{n}(x_i - \hat{\mu}_X)^2 \tag{40d,e}$$

$$R_0' = R_0 - (x_1 - \hat{\mu}_X)^2 - (x_n - \hat{\mu}_X)^2 \tag{40f}$$

$$R_1 = \sum_{i=1}^{n-1}(x_i - \hat{\mu}_X)(x_{i+1} - \hat{\mu}_X) \tag{40g}$$

For given q, the corresponding maximum likelihood estimates of μ_X and σ_X^2 are

$$\hat{\mu}_X = \frac{Q_n - q(Q_n + Q_n') + q^2 Q_n'}{n - 2q(n-1) + q^2(n-2)} \tag{41a}$$

$$\hat{\sigma}_X^2 = \frac{R_0 - 2qR_1 + q^2 R_0'}{n(1-q^2)} \tag{41b}$$

where

$$Q_n = \sum_{i=1}^{n} x_i; \quad Q_n' = Q_n - x_1 - x_n \tag{42a,b}$$

According to Anderson (1971), (39) will have one root between 0 and 1 (for positive R_1, which is n times the lag 1 covariance and should be positive under this model) and two roots outside the interval $(-1, 1)$. The root of interest is the one lying between 0 and 1 and it can be efficiently found using Newton-Raphson iterations with starting point $q = R_1/R_0'$, as long as that starting point lies within (0, 1) (if not, use starting point $q = 0.5$).

Because the coefficients of the cubic depend on $\hat{\mu}_X$, which in turn depends on q, the procedure actually involves a global iteration outside the Newton-Raphson root finding iterations. However, $\hat{\mu}_X$ changes only slightly with changing q, so global convergence is rapid if it is bothered with at all. Once the root q of (39) has been determined, the maximum likelihood estimate of the scale of fluctuation is determined from

J. Geotech. Geoenviron. Eng., 1999, 125(6): 470-485

$$\hat{\theta} = -\frac{2\Delta z}{\ln q} \tag{43}$$

In general, estimates of the variances of the maximum likelihood estimates derived above are also desirable. One of the features of the maximum likelihood approach is that asymptotic bounds on the covariance matrix $\mathbf{C}_{\hat{\theta}}$ between the estimators $\hat{\boldsymbol{\theta}}^T = \{\hat{\mu}_X, \hat{\sigma}_X^2, \hat{\theta}]$ can be found. This covariance matrix is called the Cramer-Rao bound, and the bound has been shown to hold asymptotically for both finite-scale and fractal processes (Dahlhaus 1989; Beran 1994), in both cases for both n and the domain size going to infinity. If we let $\boldsymbol{\theta} = \{\mu_X, \sigma_X^2, \theta\}$ be the vector of unknown parameters and define the vector

$$\underset{\sim}{\mathcal{L}}' = \frac{\partial}{\partial \theta_j}\mathcal{L} \tag{44}$$

where $\mathcal{L}$ is the log-likelihood function defined by (31), then the matrix $\mathbf{C}_{\hat{\theta}}$ is given by the inverse of the Fisher information matrix

$$\mathbf{C}_{\hat{\theta}}^{-1} = \mathrm{E}[[\underset{\sim}{\mathcal{L}}'][\underset{\sim}{\mathcal{L}}']^T] \tag{45}$$

where the superscript T denotes the transpose and the expectation is overall possible values of $\mathbf{X}$ using its joint distribution [see (28)] with parameters $\hat{\boldsymbol{\theta}}$. The above expectation is generally computed numerically because it is quite complex analytically. For the Gauss-Markov model the vector $\underset{\sim}{\mathcal{L}}'$ is given by

$$\underset{\sim}{\mathcal{L}}' = \begin{cases} \dfrac{1}{\sigma_X^2}\mathbf{1}^T\boldsymbol{\rho}^{-1}(\mathbf{X}-\boldsymbol{\mu}) \\ \dfrac{1}{2\sigma_X^4}(\mathbf{X}-\boldsymbol{\mu})^T\boldsymbol{\rho}^{-1}(\mathbf{X}-\boldsymbol{\mu}) - \dfrac{n}{2\sigma_X^2} \\ \dfrac{2\Delta z(n-1)q^2}{\theta^2(1-q^2)} - \dfrac{1}{2\sigma_X^2}(\mathbf{X}-\boldsymbol{\mu})^T\mathbf{R}(\mathbf{X}-\boldsymbol{\mu}) \end{cases} \tag{46}$$

where $\mathbf{R}$ is the partial derivative of $\boldsymbol{\rho}^{-1}$ with respect to the scale of fluctuation θ.

Fractal Model

The use of a fractal model is considerably more delicate than that of the finite-scale model. This is because the fractal model, with $G(\omega) \propto \omega^{-\gamma}$, has infinite variance. When $0 \leq \gamma < 1$, the infinite variance contribution comes from the high frequencies so that the process is stationary but physically unrealizable. Alternatively, when $\gamma > 1$, the infinite variance comes from the low frequencies that yield a nonstationary (fractional Brownian motion) process. In the latter case, the infinite variance basically arises from the gradual meandering of the process over increasingly large distances as one looks over increasingly large scales. Though nonstationarity is an interesting mathematical concept, it is not particularly useful nor practical in soil characterization. It does, however, emphasize the dependence of the overall soil variation on the size of the region considered. This explicit emphasis on domain size is an important feature of the fractal model.

To render the fractal model physically useful for the case when $0 \leq \gamma < 1$, Mandelbrot (1968) introduced a distance δ over which the fractal process is averaged to smooth out the high frequencies and eliminate the high frequency (infinite) variance contribution. The resulting correlation function is given by (4). Unfortunately, the rather arbitrary nature of δ renders Mandelbrot's model of questionable practical value, particularly from an estimation point of view. If δ is treated as known, then one finds that the parameter H can be estimated to be any value desired simply by manipulating the size of δ. Alternatively, if both δ and H are estimated simultaneously via maximum likelihood in the space domain (see previous subsection), then one finds that the likelihood surface has many local maxima making it difficult to find the global maximum. Even when it has been found with reasonable confidence, it is the writer's experience that the global maximum tends to correspond to unreasonably large values of δ corresponding to over-averaging and thus is far too smooth a process. Why this is so is yet to be determined.

A better approach to the fractal model is to employ the spectral representation, with $G(\omega) = G_o/\omega^\gamma$, and apply an upper frequency cutoff, in the event that $0 \leq \hat{\gamma} < 1$, or a lower frequency cutoff in the event that $\hat{\gamma} > 1$. Both of these approaches render the process stationary and having finite variance. When $\gamma = 1$, the infinite variance contribution appears at both ends of the spectrum and both an upper and lower cutoff are needed. The appropriate cutoff frequencies should be selected on the basis of the following:

- A minimum descriptive scale, in the case $0 \leq \hat{\gamma} < 1$, below which details of the process are of no interest. For example, if the random process is intended to be soil permeability, then the minimum scale of interest might correspond to the laboratory sample scale d at which permeability tests are carried out. The upper frequency cutoff might then be selected such that this laboratory scale corresponds to, for instance, one wavelength, $\omega_u = 2\pi/d$.
- For the case $\hat{\gamma} > 1$, the lower bound cutoff frequency must be selected on the basis of the dimension of the site under consideration. Because the local mean will be almost certainly estimated by collecting some observations at the site, one can eliminate frequencies with wavelengths that are large compared with the site dimension. This issue is more delicate than that of an upper frequency cutoff discussed above because there is no natural lower frequency bound corresponding to a certain finite scale (whereas there is an upper bound corresponding to a certain finite-sampling resolution). If the frequency bound is made to be too high, then the resulting process may be missing the apparent long-scale trends seen in the original data set. As a tentative recommendation, the writer suggests using a lower cutoff frequency equal to the least nonzero Fourier frequency $\omega_o = 2\pi/D$, where D is the site dimension in the direction of the model.

The parameter γ is perhaps best estimated directly in the frequency domain via maximum likelihood. There are at least two possible approaches, but here only the wavelet and periodogram maximum likelihood estimators will be discussed.

In the case of the wavelet basis representation, the approach is as follows [from Wornell (1996)]. For an observed Gaussian process, $x(t_i)$, $i = 1, 2, \ldots, n$, the wavelet coefficients x_j^m, $m = 1, 2, \ldots, M$, $j = 1, 2, \ldots, 2^{m-1}$, where $n = 2^M$, can be obtained via (23) (preferably using an efficient wavelet decomposition algorithm). Because the input is assumed to be Gaussian, the wavelet coefficients will also be Gaussian. Wornell has shown that they are mean zero and largely independent if $X(z)$ comes from a fractal process so that the likelihood of obtaining the set of coefficients x_j^m is given by

$$L(\mathbf{x};\gamma) = \prod_{m=1}^{M}\prod_{j=1}^{2^{m-1}} \frac{1}{\sqrt{2\pi\sigma_m^2}} \exp\left\{-\frac{(x_j^m)^2}{2\sigma_m^2}\right\} \tag{47}$$

where $\sigma_m^2 = \sigma^2 2^{-\gamma m}$ = model variance for some unknown intensity σ^2. The log-likelihood function is thus

$$\mathcal{L} = -\frac{1}{2}\sum_{m=1}^{M}(2^{m-1})\ln(\sigma_m^2) - \frac{1}{2}\sum_{m=1}^{M}\left(\frac{1}{\sigma_m^2}\right)\sum_{j=1}^{2^{m-1}}(x_j^m)^2 \tag{48}$$

J. Geotech. Geoenviron. Eng., 1999, 125(6): 470-485

(discarding constant terms) that must be maximized with respect to σ^2 and γ. [See Wornell (1996, Section 4.3) for details on a relatively efficient algorithm to maximize $\mathcal{L}$.]

An alternative approach to estimating γ is via the periodogram. In the writer's opinion, the periodogram approach is somewhat superior to that of the wavelet because the two methods have virtually the same ability to discern between finite-scale and fractal processes, and the periodogram has a vast array of available theoretical results dealing with its use and interpretation. Although the wavelet basis may warrant further detailed study, its use in geotechnical analysis seems unnecessarily complicated at this time.

Because the periodogram has been shown to consist of approximately independent exponentially distributed estimates at the Fourier frequencies for a wide variety of random processes, including fractal, it leads easily to a maximum likelihood estimator for γ. In terms of the one-sided spectral density function, $G(\omega)$, the fractal process is defined by

$$G(\omega) = \frac{G_o}{\omega^\gamma}, \quad 0 \le \omega < \infty \tag{49}$$

The likelihood of seeing the periodogram estimates $\hat{G}_j = \hat{G}(\omega_j)$, $j = 1, 2, \ldots, k$, where $k = (n - 1)/2$, and $\omega_j = 2\pi j/D$ is just

$$L(\hat{\mathbf{G}}; \boldsymbol{\theta}) = \prod_{j=1}^{k} \left(\frac{\omega_j^\gamma}{G_o}\right) \exp\left\{-\frac{\omega_j^\gamma}{G_o}\hat{G}_j\right\} \tag{50}$$

and its logarithm is

$$\mathcal{L} = -k \ln G_o + \gamma \sum_{j=1}^{k} \ln \omega_j - \frac{1}{G_o}\sum_{j=1}^{k} \omega_j^\gamma \hat{G}_j \tag{51}$$

which is maximized with respect to G_o and γ. The spectral intensity parameter G_o is not necessarily of primary interest because it may have to be adjusted anyhow to ensure that the area under $G(\omega)$ is equal to $\hat{\sigma}_x^2$ after the cutoff frequency discussed above is employed (or, alternatively, the cutoff frequency adjusted for fixed G_o).

Differentiating (51) with respect to G_o and setting the result to zero yields

$$\hat{G}_o = \frac{1}{k}\sum_{j=1}^{k} \hat{G}_j \omega_j^\gamma \tag{52}$$

In turn, differentiating (51) with respect to γ and setting the result to zero leads to the following root finding problem in γ

$$\frac{\sum_{j=1}^{k} \hat{G}_j \omega_j^\gamma \ln \omega_j}{\sum_{j=1}^{k} \hat{G}_j \omega_j^\gamma} - \sum_{j=1}^{k} \ln \omega_j = 0 \tag{53}$$

For almost all common processes, $0 \le \gamma \le 3$, so that (53) can be solved efficiently via bisection.

The Fisher information matrix, and thus the covariance matrix between the unknown parameters G_o and γ, is especially simple to compute for the periodogram maximum likelihood estimator because, asymptotically, the estimator $\hat{G}_o$ is independent of $\hat{\gamma}$. The estimated variances of each estimated parameter can be found from

$$\sigma^2_{\hat{G}_o} \simeq \frac{\hat{G}_o^2}{\left(\sum_{i=1}^{k} \omega_i^{1-\hat{\gamma}} - k\right)^2 - \sum_{i=1}^{k} \omega_i^{2(1-\hat{\gamma})}} \tag{54}$$

$$\sigma^2_{\hat{\gamma}} \simeq \frac{1}{\left(\sum_{i=1}^{k} \ln \omega_i - k\right)^2 + k} \tag{55}$$

However, it should be pointed out that these variances are only achieved asymptotically as the sample length increases to infinity. In practice, this is not possible, so that estimates of γ and G_o will show much greater variability from sample to sample than suggested by the above bounds. What this means is that the uncertainty in the estimates of γ and G_o are best obtained via simulation or by considering a large number of sampling domains.

CONCLUSIONS

When attempting to identify which of a suite of stochastic models is best suited to represent a soil property, a variety of data transforms are available. Most commonly, these are the sample correlation or covariance function, the semivariogram, the sample variance function, the sample wavelet coefficient variance function, and the periodogram. In trying to determine whether the soil property best follows a finite-scale model or a fractal 1/*f*-type noise, the periodogram, wavelet variance, and semivariogram plots were found to be the most discriminating. In this sense, the periodogram is perhaps the most preferable due to the fact it has been extensively studied, particularly in the context of time-series analysis, and because it has a nice physical interpretation relating to the distribution of power to various component frequencies.

It is recognized that these data transforms have been evaluated by averaging over an ensemble of realizations. In many real situations only one or a few data sets will be available. Judging from the minimum/maximum curves shown on the plots of Figs. 4–8, it may be difficult to assess the true nature of the process if little or no averaging can be performed. This is, unfortunately, a fundamental issue in all statistical analyses—confidence in the results decreases as the number of independent observations decreases. All that can really be said about the tools discussed above (i.e., the periodogram) is that on average it shows a straight line with negative slope for fractal processes and flattens out at the origin for finite-scale processes. Because the periodogram ordinates are exponentially distributed, and the ability to distinguish between process types depends on just the first few (low frequency) ordinates, the use of only a single sample set may not lead to a firm conclusion. For this, special large-scale investigations, yielding a large number of soil samples, may be necessary.

The sample correlation or covariance functions are acceptable measures of second moment structure when the scale of fluctuation of the process is small relative to the sampling domain, implying that many of the observations in the sample are effectively independent. However, these sample functions become severely biased when the sample shows strong dependence, preventing them from being useful to discern between finite-scale and fractal type processes. Because the level of dependence is generally not known a priori, inferences based on the sample covariance and correlation functions are not generally reliable. Likewise, the sample variance function is heavily biased in the presence of strong dependence, rendering its use questionable unless the soil property is known to be finite scale with $\theta << D$.

Once a class of stochastic models has been determined using the periodogram, the periodogram can again be used to estimate the parameters of an assumed spectral density function via maximum likelihood. This method can be applied to either finite-scale or fractal processes, requiring only an assumption on the functional form of the spectral density function. The maximum likelihood approach is preferred over other estimation techniques such as regression, because of the many available results dealing with the distribution of maximum likelihood estimates [see, e.g., Beran (1994) and DeGroot and Baecher (1993)].

If the resulting class of models is deemed to be fractal with

J. Geotech. Geoenviron. Eng., 1999, 125(6): 470-485

$0 \leq \hat{\gamma} < 1$, then the Mandelbrot model of (4) can also be fitted using maximum likelihood in the space domain (preferably with δ taken to be some assumed small averaging length, below which details of the random soil property process are of no interest). For $\gamma > 1$, the fitted spectral density function $G(\omega) = G_o/\omega^\gamma$ is still of limited use because it corresponds to infinite variance. It must be truncated at some appropriate upper or lower bound (depending on whether γ is below or above 1.0) to render the model physically useful. The choice of truncation point needs additional investigation although some rough guidelines were suggested above.

In the finite-scale case, usually only a single parameter, the scale of fluctuation, needs to be estimated to provide a completely usable stationary stochastic model. However, indications in a companion paper (Fenton 1999) are that soil properties are fractal in nature, exhibiting significant correlations over very large distances. This proposition is reasonable if one thinks about the formation processes leading to soil deposits—the transport of soil particles by water, ice, or air, often takes place over hundreds if not thousands of kilometers. There may, however, still be a place for finite-scale models in soil models. The major strength of the fractal model lies in its emphasis on the relationship between the soil variability and the size of the domain being considered. However, once a site has been established, there may be little difference between a properly selected finite-scale model and the real fractal model over the finite domain. The relationship between such an ''effective'' finite-scale model and the true but finite-domain fractal model can be readily established via simulation and is the subject of ongoing research.

ACKNOWLEDGMENTS

The writer would like to thank the Norwegian Geotechnical Institute (NGI) and, in particular, Suzanne Lacasse, Farrokh Nadim, and Odd Gregerson for their comments on the early research as well as for their hospitality during the summer of 1997. The writer extends his gratitude to the Research Council of Norway for their generous support and for making the project possible. Over the 10 months following the NGI visit, further analysis work was carried out at Princeton University, and thanks are due to Erik VanMarcke and the Department of Civil Engineering and Operations Research at Princeton for the use of their facilities and for their support in many ways. Finally, the writer would like to thank the National Sciences and Engineering Research Council of Canada for their support under operating Grant OPG0105445.

APPENDIX I. REFERENCES

Anderson, T. W. (1971). ''The statistical analysis of time series.'' *Probability and mathematical statistics*, Wiley, New York.
Beran, J. (1994). ''Statistics for long-memory processes.'' *Monographs on statistics and applied probability*, Chapman & Hall, New York.
Brockwell, P. J., and Davis, R. A. (1987). *Time series: Theory and methods.* Springer, New York.
Cressie, N. A. C. (1993). *Statistics for spatial data*, 2nd Ed., Wiley, New York.
Dahlhaus, R. (1989). ''Efficient parameter estimation for self-similar processes.'' *Ann. Statist.*, 17, 1749–1766.
DeGroot, D. J., and Baecher, G. B. (1993). ''Estimating autocovariance of in-situ soil properties.'' *J. Geotech. Engrg.*, ASCE, 119(1), 147–166.
Fenton, G. A. (1994). ''Error evaluation of three random field generators,'' *J. Engrg. Mech.*, ASCE, 120(12), 2478–2497.
Fenton, G. A. (1999). ''Random field modeling of CPT data.'' *J. Geotech. and Geoenvir. Engrg.*, ASCE, 125(6), 486–498.
Fenton, G. A., and VanMarcke, E. H. (1990). ''Simulation of random fields via local average subdivision,'' *J. Engrg. Mech.*, ASCE, 116(8), 1733–1749.
Journel, A. G., and Huijbregts, Ch. J. (1978). *Mining geostatistics*. Academic, New York.
Marple, S. L., Jr. (1987). *Digital spectral analysis.* Prentice-Hall, Englewood Cliffs, N.J.
Mandelbrot, B. B., and Ness, J. W. (1968). ''Fractional Brownian motions, fractional noises and applications.'' *SIAM Rev.*, 10(4), 422–437.
Matheron, G. (1962). ''Traite de Geostatistique Appliquee, Tome I,'' *Memoires du Bureau de Recherches Geologiques et Minieres*, 14, Editions Technip, Paris (in French).
Priestley, M. B. (1981). *Spectral analysis and time series.* Vol. 1, Univariate Series, Academic, New York.
Strang, G., and Nguyen, T. (1996). *Wavelets and filter banks.* Wellesley-Cambridge, New York.
VanMarcke, E. H. (1984). *Random fields: Analysis and synthesis*, MIT Press, Cambridge, Mass.
Wornell, G. W. (1996). *Signal processing with fractals: A wavelet based approach.* Prentice-Hall, Englewood Cliffs, N.J.
Yajima, Y. (1989). ''A central limit theorem of Fourier transforms of strongly dependent stationary processes.'' *J. Time Ser. Anal.*, 10, 375–383.

APPENDIX II. NOTATION

The following symbols are used in this paper:

a = constant;
b_i = coefficients in Gauss-Markov maximum likelihood problem;
$\mathbf{C}$ = covariance matrix;
C_{ij} = element of covariance matrix;
$\hat{C}(\tau_j)$ = estimated covariance function at discrete lag τ_j;
D = sampling domain length or depth;
d = representative length;
$f(q)$ = cubic function in Gauss-Markov maximum likelihood problem;
G_o = spectral intensity parameter;
$\hat{G}_o$ = estimated spectral intensity parameter;
$G(\omega)$ = one-sided spectral density function;
$\hat{G}(\omega)$ = sample one-sided spectral density function;
H = Hurst or self-similarity coefficient;
i = index, also $\sqrt{-1}$ when appearing in exponential;
$L(\cdot)$ = likelihood function;
$\mathcal{L}(\cdot)$ = log-likelihood function;
m = scale (dilation) index for wavelet basis;
$m(z)$ = mean trend varying with spatial position;
n = number of observations in sample of $X(z)$;
Q_n, Q'_n = coefficients in Gauss-Markov maximum likelihood problem;
q = root in Gauss-Markov maximum likelihood problem;
R_0, R'_0, R_1 = coefficients in Gauss-Markov maximum likelihood problem;
$\mathbf{r}$ = vector used in space domain maximum likelihood estimator;
$\mathbf{s}$ = vector used in space domain maximum likelihood estimator;
$s(z)$ = standard deviation trend varying with spatial position;
T = averaging dimension;
$V(\tau_j)$ = variogram;
$\hat{V}(\tau_j)$ = estimated variogram;
$\bar{X}$ = average of X_i across sample of length n;
X_i = random value of $X(z)$ at z_i;
$X_{i,j}$ = average of $X_j, X_{j+1}, \ldots, X_{j+i}$;
X_j^m = random wavelet coefficient;
X_T = random local average of $X(z)$ over length T;
$X(z)$ = random 1D process;
$\mathbf{x}$ = vector of observations in sample;
x_i = observed value of $X(z)$ at z_i;
x_j^m = computed wavelet coefficient from transform;
$x'(z)$ = mean and variance standardized random process;
$Y(z)$ = random 1D process;
z = coordinate, commonly with depth;
z_i = discrete points along 1D random process;
γ = spectral exponent in fractal model;
$\hat{\gamma}$ = estimated spectral exponent;
$\gamma(D)$ = variance function;
$\hat{\gamma}(D)$ = sample variance function;
Δz = incremental distance between observations;
δ = averaging dimension in Mandelbrot's fractal model;

J. Geotech. Geoenviron. Eng., 1999, 125(6): 470-485

$\varepsilon(z)$ = residual random process;
θ = scale of fluctuation;
$\hat{\theta}$ = estimated scale of fluctuation;
$\boldsymbol{\theta}$ = vector of unknown parameters to be estimated;
$\hat{\boldsymbol{\theta}}$ = vector of estimated parameters;
$\boldsymbol{\mu}$ = vector of mean values associated with set of spatial points;
$\boldsymbol{\mu}_j$ = element of $\boldsymbol{\mu}$;
μ_X = mean of $X(z)$;
$\hat{\mu}_X$ = estimated mean of $X(z)$;
$\boldsymbol{\rho}$ = correlation coefficient matrix;
$\rho_{i,j}$ = element of correlation coefficient matrix;
$\rho(\tau)$ = correlation coefficient between two points separated by τ;
$\hat{\rho}(\tau_j)$ = estimated correlation function at discrete lag τ_j;
σ^2 = wavelet coefficient intensity parameter;
σ_m^2 = wavelet coefficient variance at scale m;
σ_X^2 = variance of $X(z)$;
$\hat{\sigma}_X^2$ = estimated variance of $X(z)$;
τ = separation distance;
τ_j = discrete separation distance, $=j\Delta z$;
$\chi(\omega_j)$ = complex Fourier coefficient;
$\psi_j^m(z)$ = daughter wavelet;
$\psi(z)$ = mother wavelet;
ω = frequency or wavenumber;
ω_o = lower frequency cutoff;
ω_u = upper frequency cutoff; and
$\mathbf{1}$ = vector with all elements equal to 1.

J. Geotech. Geoenviron. Eng., 1999, 125(6): 470-485