RANDOM FIELD MODELING OF CPT DATA

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

ABSTRACT: An extensive set of cone penetration tests (CPT) soundings are analyzed statistically to produce an a priori 1D stochastic soil model for use at other similar sites. The data were collected by the Norwegian Geotechnical Institute at the site of a new airport just north of Oslo, Norway, and consists of 143 CPT soundings over an area of about 18 km^2 in a reasonably homogeneous soil mass. The CPT data consist of cone tip resistance, side friction, and pore-water pressure measurements. Only the cone tip resistance is considered in this study, it being considered closest to a ''point'' property of the soil, and only the vertical variation is characterized. To perform the statistical analysis, the data sets are viewed as independent 1D realizations extracted from a statistically homogeneous 3D random field. Plots of various transformations of the data indicate that the cone tip resistance records are best represented using a fractal stochastic model corresponding to so-called fractional Brownian motion, and its parameters are estimated via maximum likelihood.

INTRODUCTION

In the reliability analysis of any geotechnical project, a certain amount of data must be gathered. Typical risk analyses, simulations, optimizations, and decision analyses all involve knowledge of the stochastic nature of the soil. Because soils are spatially variable, the mean, variance, and covariance structure of the soil are needed for any realistic probability modeling. Although the mean and variance of a soil property can usually be established reasonably accurately with only a small number of samples, a reasonably accurate estimate of the entire covariance structure typically requires a vast amount of data. As a result of limited budgets, many reliability studies make use of covariance functions reported in the literature for what are deemed to be similar sites. It is believed that this is a reasonable approach, namely, that the soil property mean and variance should be estimated from site data but that the correlation structure can be derived from the more detailed analysis of similar sites. The idea that the correlation structure might be due more to external influences, which are common across soil properties and across similar sites, allows this contention (Fenton 1999).

In an attempt to establish a correlation model for one such site, the writer visited the Norwegian Geotechnical Institute (NGI) in Oslo, Norway, to statistically analyze an extensive spatially distributed set of cone penetration tests (CPT) with measured pore-water pressures (strictly speaking, this is properly called CPTU readings, but the more common CPT notation will be kept here for simplicity and because the pore-water pressure results were not used). This paper, using the techniques presented by Fenton (1999), summarizes the findings of the analysis.

The CPT data were gathered at the site of the Oslo Main Airport, Gardermoen, located to the north of the city of Oslo, and consists of measurements of cone tip resistance q_c (kPa), side friction f_s (kPa), and pore-water pressure μ (kPa). The tests were carried out using a piezocone (ENVI memocone) where the pore pressure is measured just behind the cone. The measurements were recorded at vertical intervals of 0.02 m. In total, 143 soundings are considered, scattered over an area of about 6 km by 3 km. Fig. 1 illustrates the relative surface locations of the CPT soundings. The soundings ranged from 2 to 50 m in (usable) length.

Only the q_c readings are analyzed in this study. Of the three measurements taken by the piezocone, it is believed that the q_c value most closely represents a ''point'' property of the soil (both q_c and f_s are actually local average measures, but side friction is more so). Most soil properties are not, in fact, point properties but rather local averages over some representative volume, often a laboratory scale volume. Because the major goal of this study is to establish correlation structures for soil properties, the q_c measurements, assumed representative of laboratory scale soil properties, are considered sufficient for this purpose. The issue of measurement error is ignored at this time even though it is undoubtedly present in the CPT data. There is, however, considerable debate at this time on what distribution is best for measurement error. It is usually considered to be Gaussian white noise but evidence is now appearing that suggests that improved error models may be fractal in nature (Beran 1994). It is assumed here that the signal-to-noise ratio is high enough to permit reasonable estimates of the signal statistics without complicating the issue with an error model study. One may think of the results, at this point, as a measurement model; the relationship with real soil properties, including an error model, can be incorporated in later studies.

The site of the Main Oslo Airport is an almost horizontal plateau, at about 205 m above sea level, formed as a complex system of several glacifluvial deltas during a period of about 50–100 years in the Preboreal period, some 10,000 years ago (Watn et al. 1995). The airport site is generally dominated by sands and gravels in the first 5–15 m, with finer materials appearing as lenses, overlying silty-sands in thicknesses from 5 to 30 m, in turn overlying silty-clays with thicknesses in excess of 10 m (Gregersen 1997). Sandven and Watn (1995) reported that ''due to the relatively complex geological history, the soil conditions in the investigated area show large variations, containing soil types ranging from soft, sensitive clays to very dense morainic material.'' Gregersen indicated that the soil is slightly preconsolidated with a preconsolidation pressure of about 70 kPa. Because of the large volume of CPT data and the reasonably typical soil variability, this site is considered a good testing ground for the estimation techniques discussed in Fenton (1999) as well as for making some useful inferential statistical statements about soils. The study does not consider soil layers separately—the possible migration from soil type to soil type with depth viewed as part of the overall uncertainty being characterized. To separately account for the soil layers (and possible seams) at the Oslo Airport renders the analysis site specific and results in an unconservative estimate of the a priori uncertainty for use at other sites. By a priori is meant statistical results that may be used at a site for which little or no data have yet been collected, and yet a need

[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 1090-0241/99/0006-0486–0498/$8.00 + $.50 per page. Paper No. 18421.

J. Geotech. Geoenviron. Eng., 1999, 125(6): 486-498

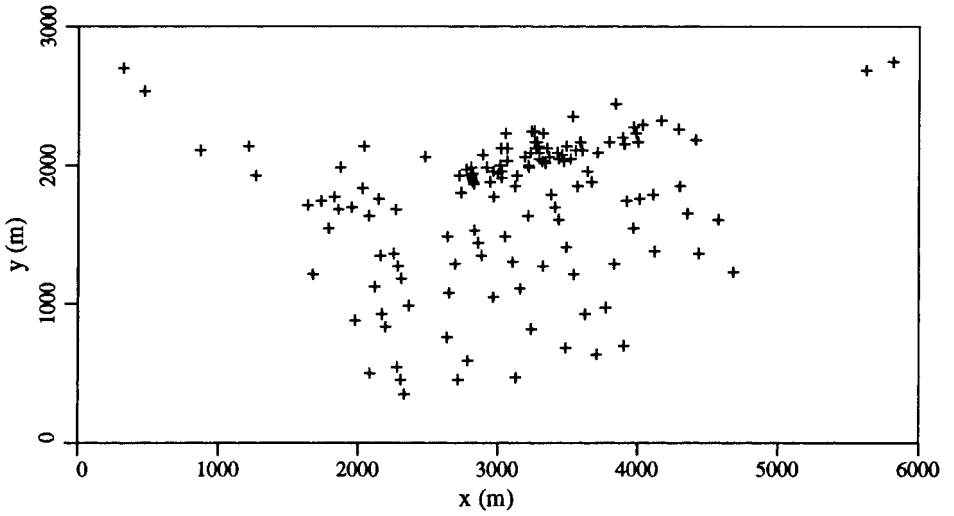

FIG. 1. Relative Locations of CPT Soundings

exists to perform a preliminary stochastic analysis, as in sampling planning, kriging analysis to ''interpolate'' from limited data, preliminary reliability assessments, etc.

Although 27 additional CPT soundings were available, they were rejected from this analysis due to one or more of the following:

- The measurements were considered to be faulty, having extended periods of zero cone tip resistance at various depths (in excess of about 5% of the total sounding). This accounted for 19 of the rejected soundings.
- The measurements were considered to be suspicious due to an excessive drift in the cone pressure calibration (in excess of 500 kPa). This accounted for five of the rejected soundings.
- The measurements were not taken at an increment of 0.02 m, yielding them inconsistent with the majority of data sets. This accounted for three of the rejected soundings.

The presence of a zero cone tip resistance well below grade was considered to be a clear indication that the CPT measurements were faulty at that point—there being no evidence or belief that the soil contained extended holes. Of the acceptable 143 CPT soundings, a few have been modified to eliminate such obvious errors and permit the use of the data set as follows:

- If only a few of the readings at intermediate depths were zeros (i.e., less than about 5% of the total sounding), the zeros were shifted to values approximately interpolated between the previous and next nonzero values, which were assumed to be valid. This shift is not believed to have a significant impact on the statistical analysis—much less in fact than the presence of the sequential zeros—and allows the data set to be included in the event that a logarithmic transform is selected. Because the zeros are clearly in error, the shift was deemed an acceptable, possibly even robust, correction, essentially downweighting the effect of the errors.
- Zeros near the beginning and/or end of a CPT record were eliminated simply by shortening the record at the beginning and/or end.
- Initial seating problems sometimes led to clearly erroneous observations in the first few centimeters of the soundings. These outliers were simply deleted from the record by slightly shortening the record.

Many of the statistical procedures discussed in this paper depend on the data having equal increments between recording points, which motivated the rejection of some of the data sets. In general, however, such data can be transformed to another increment through interpolation (linear in the simplest case, to kriging in the most sophisticated case), which presumes high correlation between data at small lags. Such a transformation was not considered warranted in this initial study because a large volume of acceptable data were readily available.

Before discussing the details of the analysis, it is important to clearly define the goals of this study. In general, statistical analyses can be separated into two areas that can be thought of as inferential and descriptive in nature. In the latter, the goal is to best describe a particular data set with a view toward interpolating within the data set. For example, this occurs when geotechnical data are obtained at the site for which a design is destined. The interpolation techniques most often used are those of regression, using an appropriate polynomial that explains most of the variability, and best linear unbiased estimation. Inference, on the other hand, is the term used here to denote the considerably less certain venture of characterizing the soil at another site for which data are not yet available. In this case trends in the data at the investigated site must be viewed with considerable caution and only explicitly described if there is a strong belief that such a trend is also to be expected at the target site. Otherwise, trends should be viewed merely as segments of a large-scale fluctuation, and this large-scale fluctuation should appear as part of the statistical characterization; that is, the model should reflect a greater uncertainty in the soil property when applied to another site.

Here, the CPT data gathered at the Main Oslo Airport site are studied primarily to obtain reasonable a priori estimates of the correlation structure of natural soils to be used at other similar sites. This is therefore an inferential statistical analysis. Although the mean and variances are obtained as part of the estimation process, these parameters tend to be specifically related to, and affected by, the soil type or property; this is particularly true of the mean. Thus, these parameters are of limited interest here and should be separately estimated at any target site. It is the correlation structure that is of primary interest, because this structure is believed to be more ''portable'' so that it can be used at other, similar sites. Although additional site investigations are required to verify such a contention, it is a reasonable first hypothesis. The use of the term ''correlation structure'' here specifically refers to the correlation coefficient function but is also intended to include any function that equivalently describes the second moment of the random process, such as the spectral density function. Because

J. Geotech. Geoenviron. Eng., 1999, 125(6): 486-498

one can be found from the other (given the variance), the functions are, strictly speaking, interchangeable.

To this end, the NGI data set provides a high-quality spatial realization, which is representative of geotechnical soil properties, from which estimates of the in situ correlation structure can be made. Lunne et al. (1997) provided a comprehensive list of relationships between CPT sounding data and engineering properties of interest. For example, both the overconsolidation ratio (OCR) and the undrained shear strength can be obtained through equations of the form $OCR = \beta_1 q_c + \beta_2$ and $s_u = \alpha_1 q_c + \alpha_2$, respectively, where the coefficients depend on cone properties, in situ stresses, pore pressures, etc. In particular, in the case of the undrained shear strength, the coefficients suggested by Lunne et al. are deterministic constants, so that the mean and variance of s_u are easily computed in terms of the mean and variance of q_c (although the variance should now include a component deriving from the above regression) and their spatial correlation structures are identical. Because statistical characterization of the correlation structure is the primary thrust of this paper, the results may be viewed as being representative of the spatial variation of this (and possibly other) engineering soil property as well.

It is assumed that the soil is spatially statistically homogeneous and that the CPT soundings represent an ensemble of largely independent realizations of the same 1D random process. That is, the random process representing the variation of q_c with depth has the same joint distribution at any horizontal location, so that various soundings are just different realizations of this process. Although it is likely that horizontal spatial dependence renders the CPT soundings somewhat dependent, rather than independent as assumed, this error is assumed not to be particularly significant over the area considered. It also is only believed to result in a slight underestimation of the variability of statistical estimates coming from this analysis. It should be noted, however, that this lateral independence assumption implies that the results are best used at sites deemed to be similar.

To carry out maximum likelihood estimation of the random field of q_c measurements, a distribution for q_c must be assumed. Because q_c is bounded below by zero and has no arbitrarily set upper limit, q_c may be assumed to be lognormally distributed. There is some supporting evidence for the lognormal distribution for a number of strictly positive soil properties (e.g., strength, elastic modulus, permeability) [see Lumb (1966), Hoeksema and Kitanidis (1985), and Sudicky (1986)], so this is often a reasonable assumption. Perhaps more importantly, and also a significant motivation for its use by many researchers, it leads to a simple model because the distribution of $\ln q_c$ is then normal. The jointly normal distribution is fully specified by only its first two moments (mean and covariance). Because this is generally all that can be estimated with any degree of confidence in geotechnical applications, the normal distribution is then at least appropriate as a ''minimal assumption'' model.

Other transformations are also available. In particular, raising q_c to some power between 0 and 1 is quite common. The disadvantage to these transformations is that they do not precisely lead to a nonnegative distribution on q_c. In fact if $X = q_c^r$ is normally distributed, then X ranges over the entire real line and for negative values of X, the inverse $q_c = X^{1/r}$ is only defined for a certain r and, of these, some are imaginary. However, if the probability of X being negative under the normal assumption is negligible (i.e., having a small coefficient of variation), these transformations can give reasonable results.

Once a suitable transformation has been selected, it is applied to all of the q_c observations, whereupon the remainder of the statistical analysis is performed under the assumption that all randomness is jointly normal. It should be emphasized, however, that all subsequent statistics refer to the transformed field, not the q_c field itself. The q_c random field must be obtained through the inverse transformation.

With these assumptions in place, the possibly transformed q_c data are studied in this paper as follows:

1. In the following section, an initial regression analysis is performed to determine if a statistically significant trend with depth is present. Because a trend with depth often has some physical basis, it may make sense to predict this trend and assume it to be valid at other sites. It should be noted that such a trend does not necessarily translate into all other soil properties (such as d_{15}, color, etc.) and so is not particularly portable. However, for the CPT data themselves, it seems reasonable to investigate this trend. The regression analysis is performed on all 143 CPT soundings, and the average trend computed in two ways: (1) By computing the slope and intercept values for each CPT site individually and then averaging over the ensemble; and (2) by performing a global regression on all 143 soundings at once. Although it is uncertain at this time which approach is superior, with respect to robustness issues, the global regression is somewhat more appealing from the point of view of yielding a single representative trend. This regression can be used to detrend the data in preparation for further statistical analyses, if the trend is deemed to exist at all target sites using the statistical analysis.
2. A correlation model is identified for the CPT data by looking at a number of sample transforms, namely, the sample correlation function, variance function, variogram, wavelet coefficient variance, and periodogram.
3. Finally, the parameters of the selected correlation model are estimated using maximum likelihood. The parameter estimator variability is derived by looking at the estimates over the ensemble of 143 CPT soundings.

REGRESSION ANALYSIS

In keeping with the philosophy of only considering physically based trends that are expected to appear at other sites, the regression analysis of the data is restricted to a simple straight line trend. It is initially assumed that a reasonable data transformation is the natural logarithm, so that the trend is expressed as

$$\ln q_c = a + bz + \varepsilon \tag{1}$$

where a = mean value of $\ln q_c$ at the surface ($z = 0$); b = slope; and ε = mean-zero residual random component. To compare data sets, all sets are ''aligned'' at the preboring depth; that is, the first CPT reading for each data set is assumed to occur at a depth of $z = \Delta z = 0.02$. As will be seen, the different preboring depths actually result in only small changes in the local CPT intercept (a values).

Fig. 2 shows a typical $\ln q_c$ recording with superimposed local and global regressions—''local'' being that determined for the current CPT sounding, and ''global'' being that determined using a global regression on all 143 data sets simultaneously. The significance of the slope was tested against the null hypothesis that $b = 0$ at the 5% level and found to be significant for 90% of the CPT soundings. However, the slope was very small and was just as often negative as positive. Thus, it is reasonable to assume that, on average, there is no substantial depth dependence to the data at this site (also justifying the data set depth alignment).

To assess the assumption that $\ln q_c$ is normally distributed, Fig. 3 plots the average value of $\ln q_c$ along with the minimum and maximum observed $\ln q_c$ at each depth z across all 143

J. Geotech. Geoenviron. Eng., 1999, 125(6): 486-498

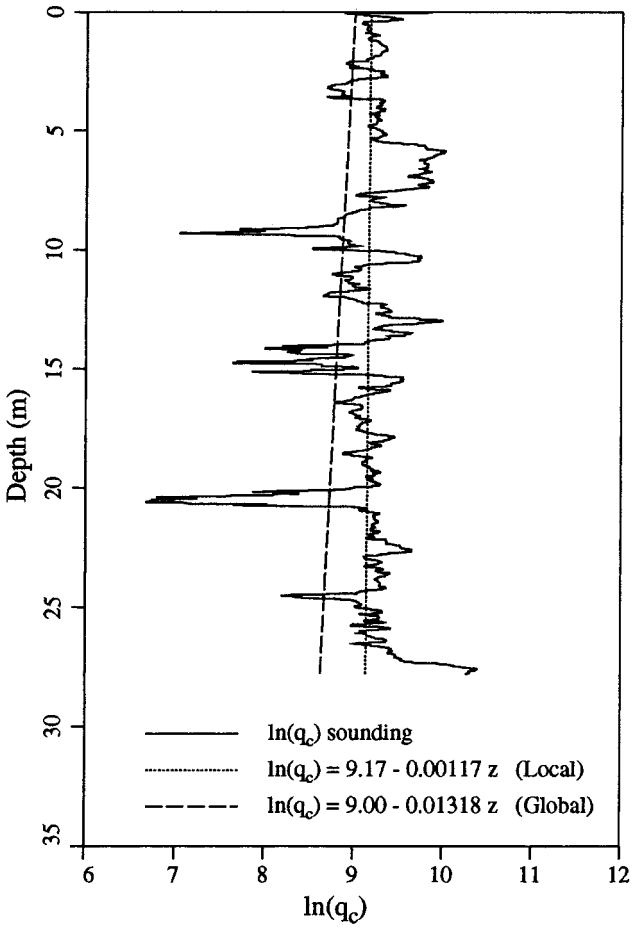

FIG. 2. Typical Variation of ln q_c versus Depth along with Local and Global Regressions

data sets. Clearly there is significant nonsymmetry in the scatter, despite the logarithmic transformation. Fig. 4 shows a histogram of the ln q_c values along with a fitted normal distribution—the skew of the data is very evident with skewness coefficient of over 40.

It is possible that Fig. 4 may also be interpreted as the superposition of two normal distributions: (1) For a weaker seam; and (2) for the more common silty-sands and clays. If the goal of the paper were to completely characterize this site, then the two materials should be treated separately. However, because the goal is to provide an a priori model that can be used to initially characterize other sites, this additional uncertainty must be incorporated into a single model and fitted as well as possible to the data. Although this approach introduces some inaccuracies into the following analysis, it is to be emphasized that the goal here is to find an approximate a priori model that must be subsequently refined at the target site as data are gathered at that site. If data are not gathered at the target site, then one would be no further ahead regardless of the level of detail in the current analysis. That is, separation of the material types in the current site analysis may actually result in an underestimation of possible variability at another site.

A transform that yields a zero skewness coefficient over all CPT data is $q_c^{0.74}$, and Fig. 5 shows the histogram and fitted normal distribution for this transformation. Unfortunately, the fitted distribution still fails to capture the apparent bimodal nature of the $q_c^{0.74}$ data. However, as only the first two moments are being characterized in this work, a more complex distribution will not be considered. Because the $q_c^{0.74}$ transformation yields a zero skewness coefficient, it was selected for use in this study over the logarithmic transformation. The approximation of a normal to the $q_c^{0.74}$ data transformation will be considered adequate for this and for reasons discussed earlier. In addition, it will be found that the estimated correlation parameters are little affected by the exact nature of the data transformation.

Although the details of the regression itself are affected by the choice of data transformation, the slopes remain small and inconsistent in sign. It is concluded, therefore, that there is no significant trend in the data that needs to be considered explicitly in an inferential study of this sort. If the regression were to be used, it must be emphasized that detrending the data significantly affects the resulting mean, variance, and correlation structures, so that these estimates must be accompanied by information regarding the regression to form a complete and usable picture of the random field model.

CHOICE OF CORRELATION MODEL

Attention is now turned to the stochastic characterization of the $q_c^{0.74}$ data. In addition to estimating the mean and variance, the spatial correlation structure must be deduced. In the writer's opinion, the large-scale geologic mixing processes leading eventually to soil formation lends support to the contention that many geotechnical properties are ''long-memory'' or fractal in nature. What these two terms mean is that there is a significant correlation between a soil's properties at two points even when the points are very widely separated. This long-range dependence is not too difficult to believe considering, for example, the erosion, transport, and weathering processes involved in the formation of a soil. The soil at a point on the surface of the earth is derived from a rock source that may be quite a distance away. For example, rivers and/or glaciers may carry soil particles many hundreds or even thousands of kilometers. Thus, a soil may have the same constitutive components as some other widely separated soil having the same original rock source. It is yet to be seen if this idea holds in both the horizontal and vertical directions—there perhaps being less reason to believe that soils separated in the vertical direction derive from the same distant locations. Nevertheless, it is worth investigating if statistical evidence supports a fractal model in the vertical direction because this possibility cannot be ruled out.

The stochastic soil model most often employed currently in practice is a short-memory or finite-scale model (as in the exponentially decaying correlation function). That is, the correlation between spatial points is assumed to fall off rapidly as the distance between the points increases, and the area under the correlation function, which is the scale of fluctuation, is finite (unlike that for the fractal model). One reason that it is important to determine whether soils are more closely fractal in nature than they are short-memory or finite-scale arises from the fact that maximum likelihood estimates of the scale of fluctuation, when sampling from fractal processes, are dependent on the size of the sampling domain. For example, as part of this research, a simulation study was carried out in which a fractal process with significant lingering correlation was generated first over a 2-m length and then again over 20 m. The maximum likelihood estimate of the scale of fluctuation was 0.36 m in the smaller domain and 3.6 m in the larger domain. This sampling domain dependence has been observed by the writer over the years as a typical feature of long-memory or fractal processes when characterized by short-memory estimators. What this means is that if soils are actually fractal in nature, then an estimate of the scale of fluctuation of, for instance, 2 m obtained from the literature must be viewed with caution because it is intrinsically tied to the size of the sampling domain. If the same researchers reporting this estimate had data over 10 times the length, they may have obtained a much larger scale of fluctuation. Similarly, if the end-user is

J. Geotech. Geoenviron. Eng., 1999, 125(6): 486-498

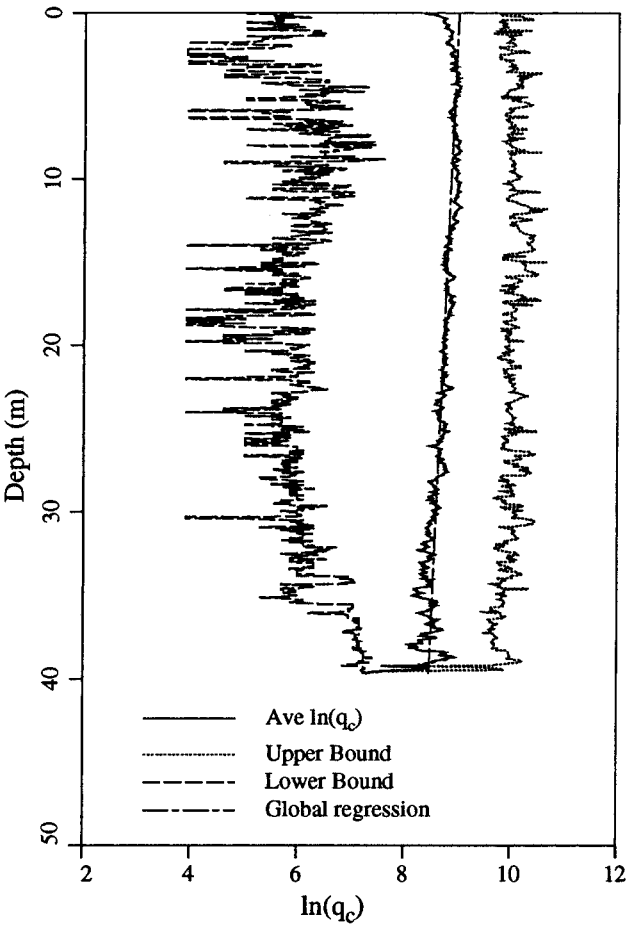

FIG. 3. Average, Minimum, and Maximum of ln q_c at Each Depth

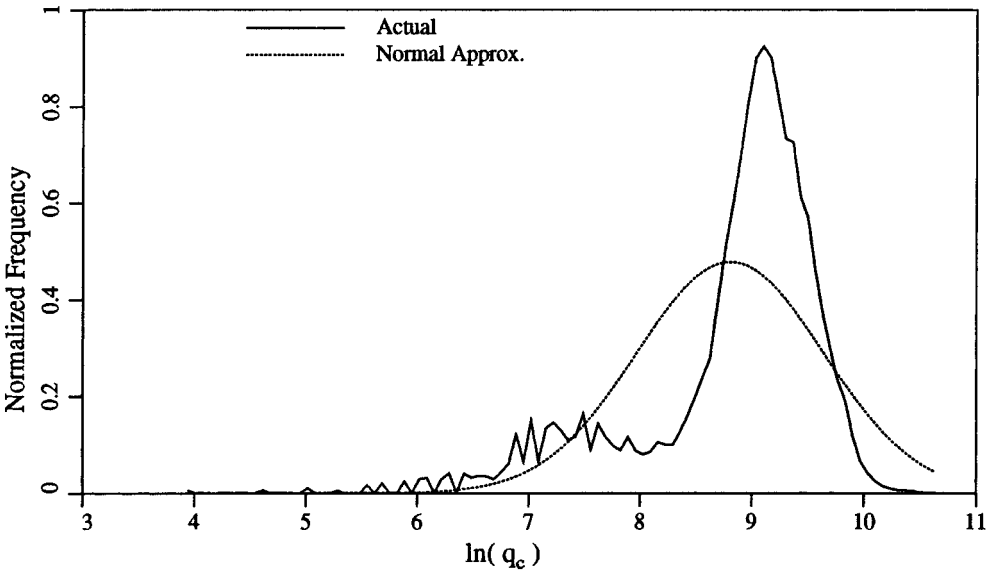

FIG. 4. Histogram of ln q_c Values with Fitted Normal Distribution

interested in characterizing a much smaller (larger) domain than reported in the literature, then he or she should use a smaller (larger) scale of fluctuation in the site model.

Thus, one motivation for the use of the fractal model, if found to be appropriate, instead of an ''equivalent'' finite-scale model where the scale is adjusted to reflect the domain size is that the main parameter of the fractal model becomes independent of the domain size. As discussed later, the dependence on the domain size is not entirely eliminated, but it now becomes explicitly part of the model, allowing a better understanding of the stochastic variation.

Although there may be other arguments in support of fractal models for soils, it is worth considering the data itself at this point. Fig. 6 shows the sample correlation function computed for the same data set shown in Fig. 2. The dashed lines are the $\pm 2n^{-1/2}$ significance bounds within which the sample correlation is commonly assumed not different from zero [see Priestley (1981)]. For simplicity, using

J. Geotech. Geoenviron. Eng., 1999, 125(6): 486-498

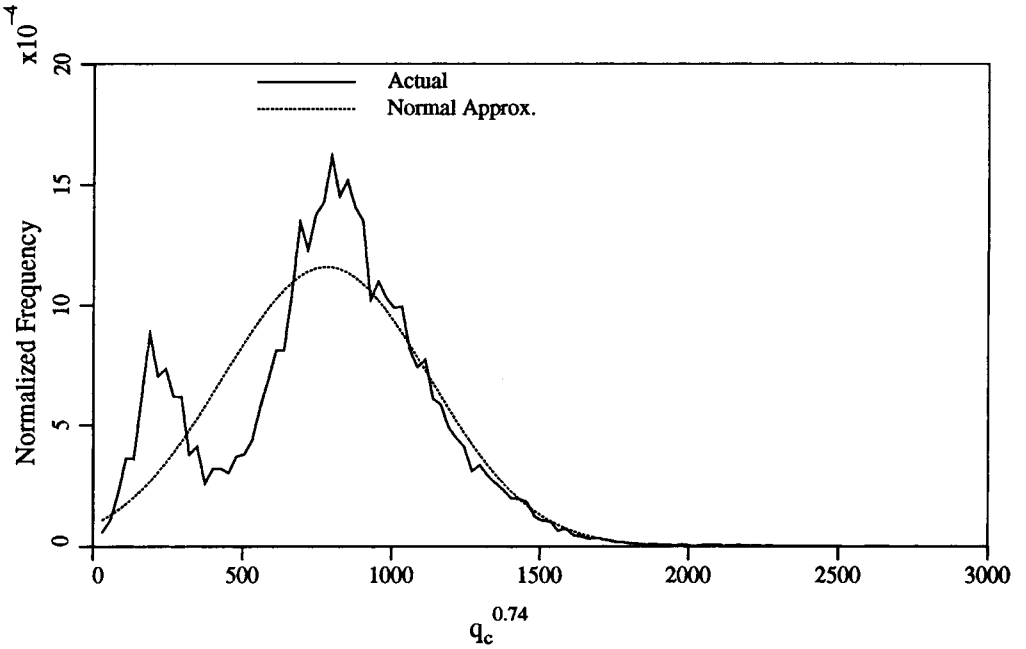

FIG. 5. Histogram of $q_c^{0.74}$ and Fitted Normal Distribution

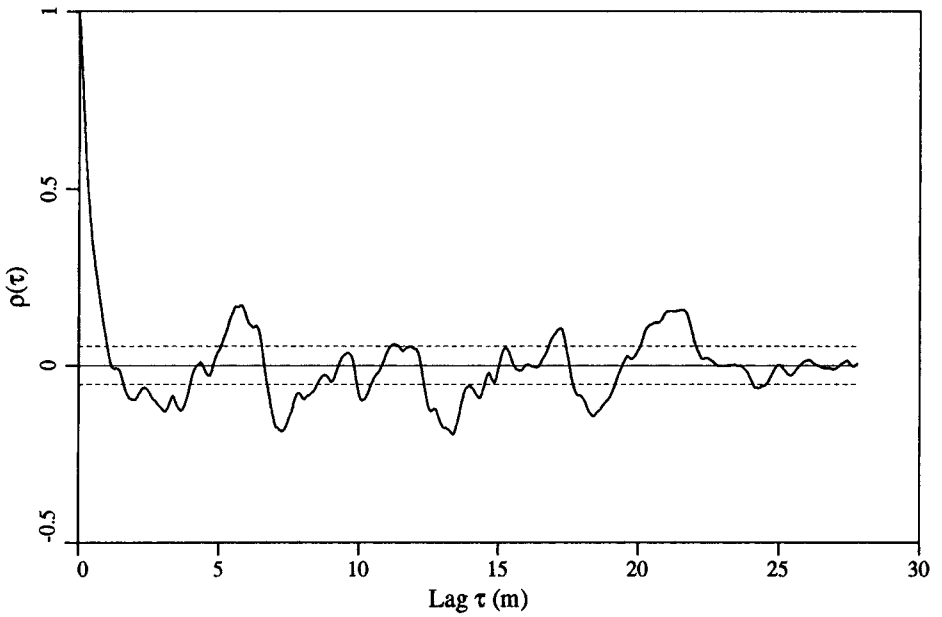

FIG. 6. Typical Sample Correlation Function Computed for $q_c^{0.74}$ Record with Length of 27.8 m (n = 1,390)

$$x_i = [q_c(z_i)]^{0.74} \tag{2}$$

at depth $z_i = i\Delta z$, $i = 1, 2, \ldots, n$, in the following, the sample covariance function is obtained from the moment estimator

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

$$j = 0, 1, \ldots, n-1 \tag{3}$$

for lag $\tau_j = j\Delta z$. The sample correlation is then $\hat{\rho}(\tau_j) = \hat{C}(\tau_j)/\hat{C}(0)$.

Fig. 6 appears to show a fairly short scale of fluctuation (which, roughly speaking, is the lag distance at and beyond which points are effectively uncorrelated), somewhere around 1 m. This does not appear to be a particularly long-memory record. This conclusion, however, may be very much in error due to the large bias of this estimator for strongly correlated fields [see Priestley (1981, Section 5.3) and Fenton (1999)].

When averaging the sample correlation functions over the ensemble of CPT soundings, a problem arises over how to deal with the differing sample lengths. Only a few records contribute to the average at lags greater than 40 m, for instance. Because the large lag behavior is vital to the determination of the large-scale behavior of the random process, it is important to have reasonably accurate estimates in these regions and to avoid basing conclusions on a small subset of the total database. Because accuracy is generally improved by increasing the level of averaging, one cannot expect good accuracy at large lags for this data set. This will also appear as a problem in interpreting the sample spectral density functions—only the longest samples contribute to the very low frequencies, and the low frequency behavior tells about the fractal (or lack of such) nature of the random process.

Because of the necessity, therefore, for attaining as much accuracy as possible at the larger scales, which requires that the estimates be averaged over as many soundings as possible,

J. Geotech. Geoenviron. Eng., 1999, 125(6): 486-498

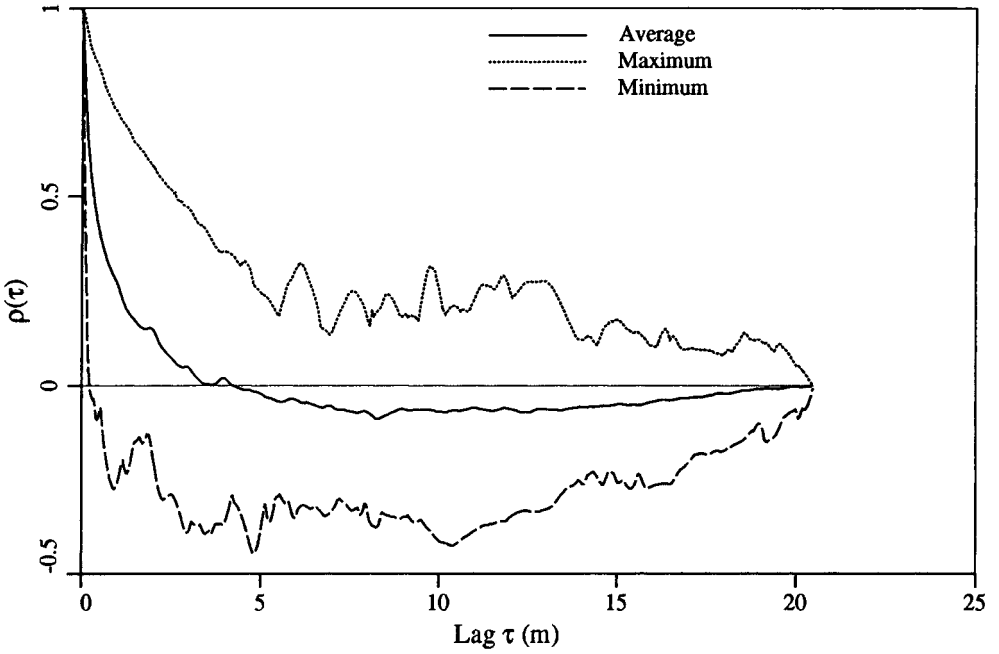

FIG. 7. Average of Sample Correlation Functions over 93 CPT Soundings along with Associated Minimum and Maximum at Each Lag

it was decided to truncate the soundings to equal length in the remainder of this section. Shorter soundings would be ignored, as well as the remainder of longer soundings. In this way, the decision about the nature of the stochastic variation can be made using estimates of high accuracy, rather than on estimates of questionable accuracy available at the very large scales (greater than about 40 m). The length selected is a tradeoff between the desire to obtain as many contributing soundings (encouraging a short truncation length) and the need to investigate larger lags to ascertain fractal behavior (encouraging longer truncation lengths). A truncation length of 20.48 m (n = 1,024) was selected. Ninety-three of the soundings have lengths (depths) greater than or equal to 20.48 m. The word ''global'' in the following will refer to results obtained by averaging over all 93 soundings.

Fig. 7 shows the average of the sample correlation functions over the 93 soundings. The global average $q_c^{0.74}$ value, across all 93 soundings is 790.9, with individual sounding averages ranging from 375 to 1,053. The global standard deviation is 279, with sounding standard deviations ranging from 141 to 551. (The averages across all 143 soundings are little changed, with an average of 782 and a standard deviation of 301.)

It appears from Fig. 7 that the scale of fluctuation, on average, is considerably larger than was indicated by Fig. 6. It may be tempting to assume that it is in the neighborhood of 4 m. However, because the sample correlation function can be highly biased, little conclusion can be drawn from these results until it is shown by other means that the scale of fluctuation is considerably <20.48 m.

The sample variogram defined by

$$\hat{V}(\tau_j) = \frac{1}{(n-j)} \sum_{i=1}^{n-j} (x_{i+j} - x_i)^2 \tag{4}$$

is averaged over the 93 soundings and plotted in Fig. 8. Because the values of $\hat{V}(\tau_j)$ are based on fewer data pairs at larger lags, they are more variable, accounting for the erratic behavior near the right tail of the figure. The plot does show a gradual increase over most of its length, indicating that the $q_c^{0.74}$ process possesses a large scale of fluctuation or may be fractal in nature.

Fig. 9 shows the average sample variance function, defined by

$$\hat{\gamma}(i) = \frac{1}{\hat{\sigma}_X^2(n-i+1)} \sum_{j=1}^{n-i+1} (X_{i,j} - \hat{\mu}_X)^2, \quad i = 1, 2, \ldots, n \tag{5}$$

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

$$X_{i,j} = \frac{1}{i} \sum_{k=j}^{j+i-1} X_k, \quad j = 1, 2, \ldots, n-i+1 \tag{6}$$

To emphasize the relationship with averaging over the physical domain, the plot is in terms of $T = i\Delta z$. Clearly, there is some serial dependence between observations because the sample variance function remains much higher than that for the ideal sample where observations are independent. Beyond this, not much can be deduced from the sample variance function because of its high bias in the presence of strong correlation [see Fenton (1999)].

The sample wavelet coefficient variance, as defined by Wornell (1996) and discussed by Fenton (1999), is computed, averaged over the soundings, and plotted in Fig. 10. Although the slope of the average coefficient variance does not remain perfectly constant at the intermediate and smaller scales (increasing m), the trend at the larger scales (small m) is a constant negative slope. Because the critical aspect of a fractal model is that correlation remains significant over very large scales, it is primarily the large-scale behavior of Fig. 10 that is of interest. Small deviations from the fractal model at small scales are not particularly important and may be due to measurement/recording errors as discussed next.

The behavior indicated by the sample spectral density function, shown in Fig. 11, reflects that of the wavelet coefficient variance. Again, fractal behavior is indicated by the constant negative slope at smaller frequencies (corresponding to longer scales), whereas some deviation from the ideal fractal model is seen at higher frequencies. Such deviations are not particularly important as the power in these higher frequencies is considerably reduced. The deviations may also be in part due to the measurement/recording resolution of the CPT q_c data that are rounded to the nearest 50 kPa. Such rounding eliminates the fine-scale (high-frequency) self-similar nature of the recording. Unfortunately, rounding is not a linear filter, and so

J. Geotech. Geoenviron. Eng., 1999, 125(6): 486-498

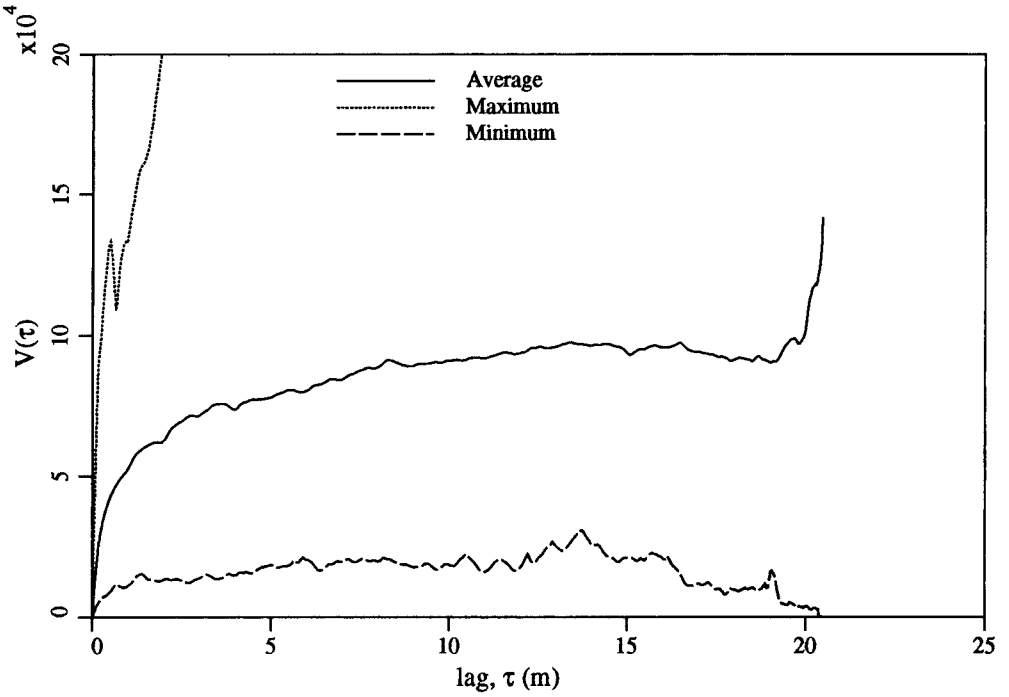

FIG. 8. Sample Variogram, Averaged over 93 CPT Soundings, along with Associated Minimum and Maximum at Each Lag

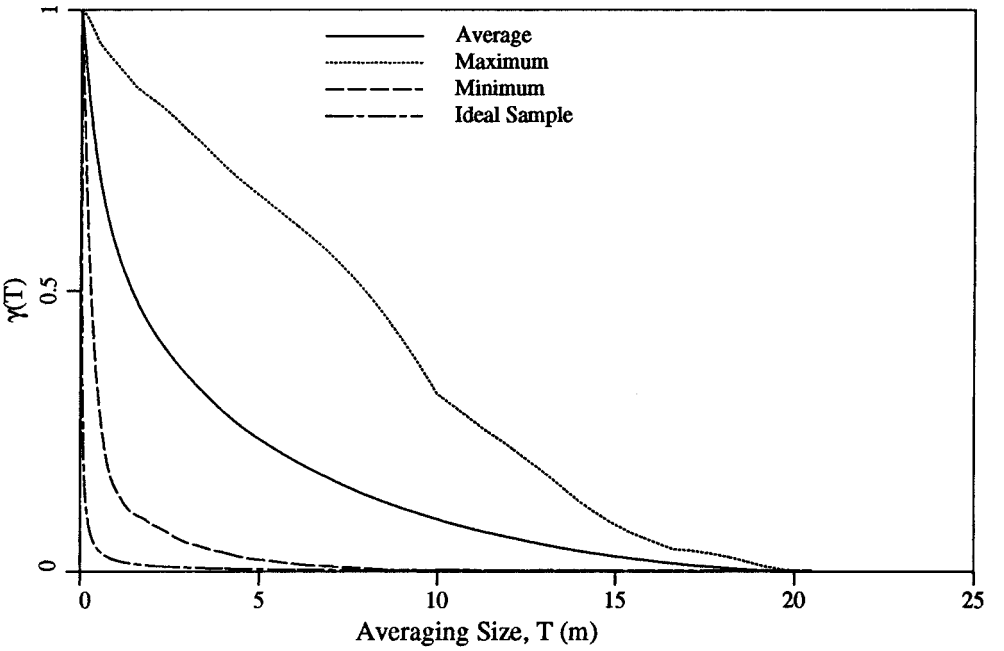

FIG. 9. Sample Variance Function, Averaged over 93 CPT Soundings, along with Associated Minimum and Maximum for Each Averaging Dimension

the exact nature of the change in the spectral density function is difficult to express. Suffice it to say that the nonlinearities in Fig. 11 may be artificially created to some extent due to the rounding. The recording does not imply that the fine-scale self-similar nature does not exist, just that it has not been observed. A basic representative fractal model would naturally include the fine-scale detail missing in the record.

Because the curves shown in both Figs. 10 and 11 retain a distinct negative slope near the origin, rather than flattening out, the long-scale (small-frequency, small-scale index) nature of the soil clearly shows a fractal behavior. On the basis of these plots, and in particular on Fig. 11, it appears that the CPT data, after transforming $q_c^{0.74}$ to yield a more symmetric distribution, are well modeled by a fractal process with spectral density $G(\omega) = G_o\omega^{-\gamma}$, where G_o is the spectral intensity, and γ is the spectral exponent controlling distribution of the power from high-to-low frequencies. If $0 \leq \gamma < 1$, then the process is stationary with infinite power in the high frequencies. Alternatively, if $\gamma \geq 1$, then the infinite power is in the low frequencies and the process ''wanders''; that is, the process is nonstationary.

The fractal nature of the CPT data is indicated by both the wavelet and the spectral density plots and suggested by the variogram. In addition, if the logarithmic transformation had been used instead of the power transform, these observations would not have changed significantly. The correlation structure is largely insensitive to the choice of transformations even though the shape of the marginal distribution is clearly affected.

J. Geotech. Geoenviron. Eng., 1999, 125(6): 486-498

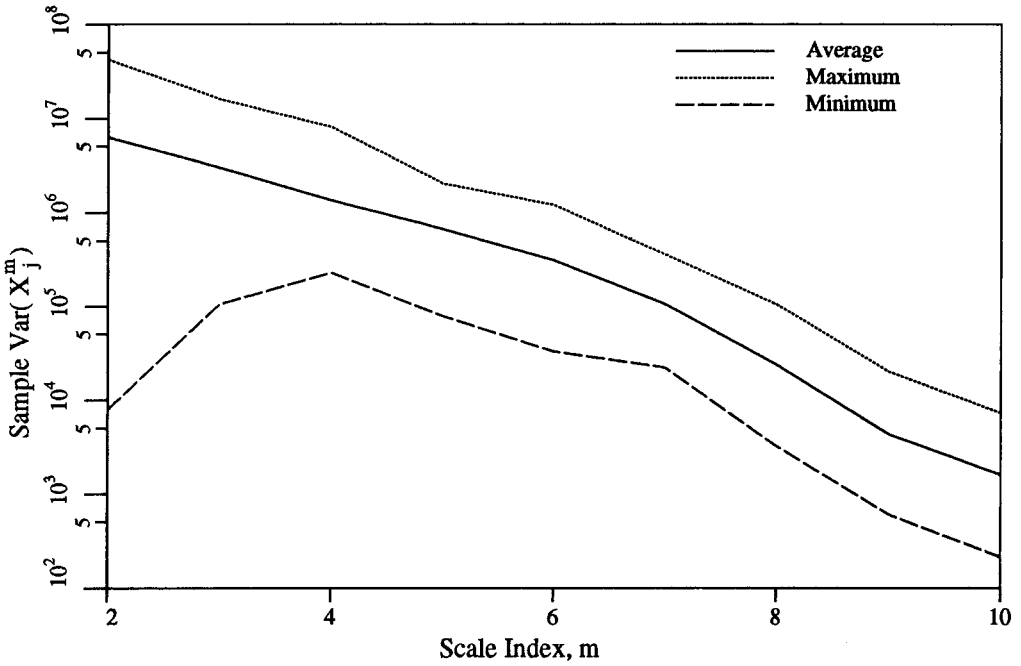

FIG. 10. Sample Wavelet Coefficient Variance, Averaged over 93 CPT Soundings, along with Associated Minimum and Maximum at Each Scale

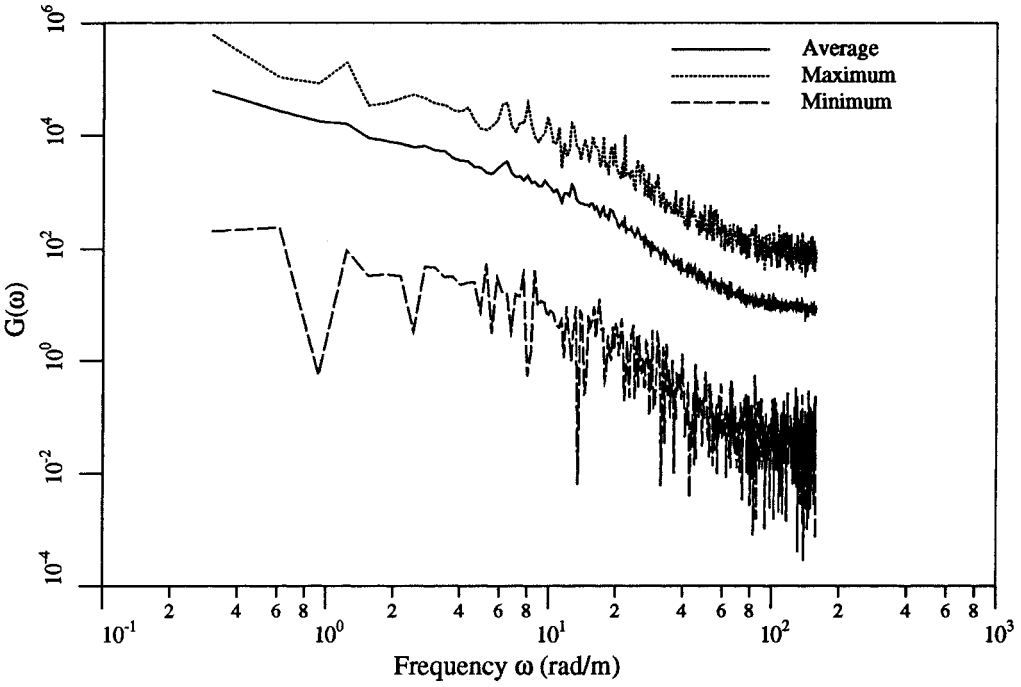

FIG. 11. Periodogram, Averaged over 93 CPT Soundings, along with Associated Minimum and Maximum at Each Frequency

ESTIMATION OF STATISTICAL PARAMETERS

Once a stochastic model has been decided upon using the tools of the previous section, the parameters of the model can be estimated using all of the data. There is no longer a necessity to maintain a high level of accuracy at the larger scales. Admittedly, the longer records will tend to yield more accurate estimates. It is expected that the variety of record lengths used here will lead to a higher estimator variability, which is at least conservative.

The spectral exponent γ can be estimated in a number of ways, of which the maximum likelihood estimates using the wavelet and periodogram will be considered here. Using the procedures outlined in Fenton (1999), a histogram of the wavelet maximum likelihood estimate of γ is shown in Fig. 12 based on the estimated γ for all 143 CPT soundings. In this case, the average estimated γ value was 1.9 ± 0.18, where the 0.18 value is the estimated standard deviation of the estimated γ value. This notation will be used throughout this section. For all the soundings, the γ estimates ranged from 1.28 to 2.30.

Fig. 13 shows a histogram of γ estimates over the 143 soundings using the periodogram maximum likelihood estimator discussed in Fenton (1999). The average estimated γ value is now 1.8 ± 0.13 with an overall range from 1.23 to 2.07. The periodogram estimates agree very well with those obtained using the wavelet basis, and they have somewhat superior precision.

J. Geotech. Geoenviron. Eng., 1999, 125(6): 486-498

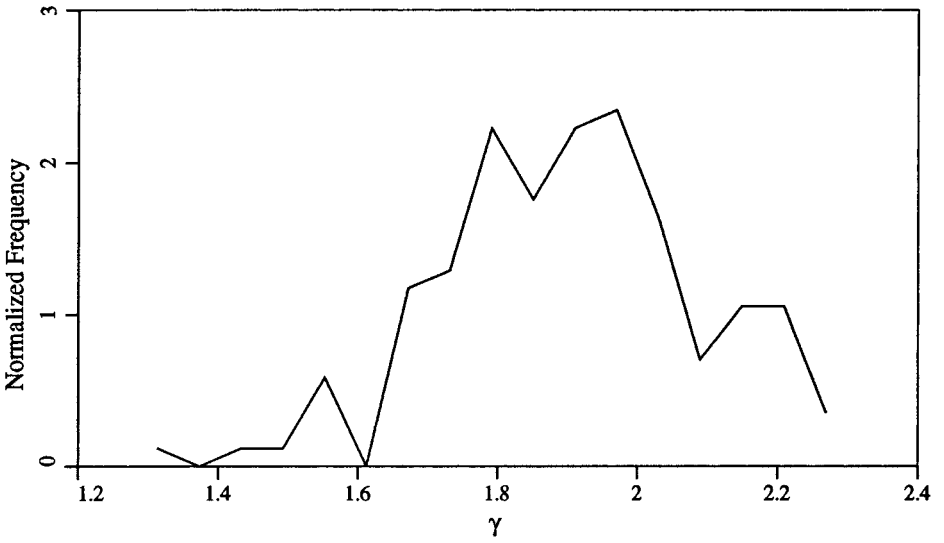

FIG. 12. Distribution of Maximum Likelihood Estimates of γ via Wavelet Basis over 143 CPT Soundings

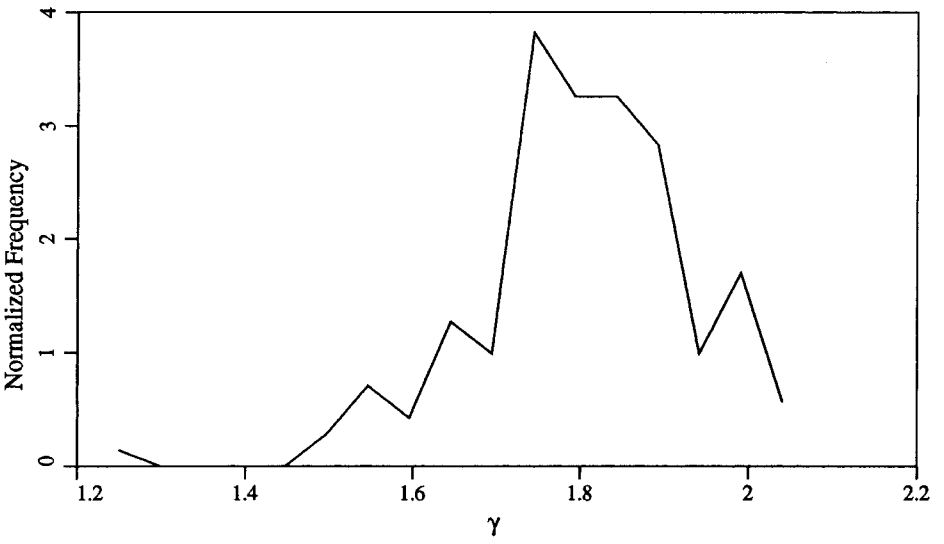

FIG. 13. Distribution of Maximum Likelihood Estimates of γ Obtained Using Periodogram over 143 Soundings

Again, it should be noted that the estimates obtained using a lognormal transformation of the original data were very nearly the same, with the wavelet and periodogram average estimates being 1.84 ± 0.20 and 1.76 ± 0.14, respectively. The estimation process does not appear to be particularly sensitive to the data transformation. In the following, the periodogram estimator will be used alone, with the wavelet estimator being viewed primarily as corroborative evidence on the estimate of γ.

The spectral exponent γ may be equivalently expressed by the fractal dimension $D_f = (5 - \gamma)/2$ or the Hurst coefficient of fractional Brownian motion, $H = (\gamma - 1)/2$. The latter coefficient can lead to some confusion, however, as the Hurst coefficient for fractional Gaussian motion, the derivative of Brownian motion, which applies when $0 \leq \gamma < 1$, is $H = (\gamma + 1)/2$.

The spectral intensity parameter G_o was found to have an average value of $1.03 \times 10^5 \pm 0.67 \times 10^5$. This parameter is related to the process variance and is considered to be site specific and also significantly affected by the type of data transform used. The result quoted here is for the $q_c^{0.74}$ transformation. Over the 143 CPT soundings, the average variance of the $q_c^{0.74}$ data was estimated (by method of moments) to be $9.1 \times 10^4 \pm 4.8 \times 10^4$ and the mean was estimated to be 782 ± 124. Armed with these parameters, which can be summarized as $\hat{\mu}_X = 782$, $\hat{G}_o = 1.03 \times 10^5$, and $\hat{\gamma} = 1.8$, one can presumably create a stochastic model for the $X = q_c^{0.74}$ process. Then letting $q_c = X^{1.35}$ recovers the desired process. There are, however, a few more details to work out. Recall that the fractal process is, strictly speaking, an infinite variance process, and for $\gamma \geq 1$ it is also nonstationary. The infinite variance is physically unrealizable, and the nonstationarity is inconvenient as it introduces an origin issue. The stochastic model must be modified to become physically useful, and for this purpose a lower frequency cutoff ω_o must be introduced. An appropriate spectral density function is of the form illustrated in Fig. 14.

The choice of lower frequency cutoff, according to the suggestion by Fenton (1999), can be taken to be $2\pi/D$, where D is the soil depth. This suggestion is perhaps particularly appropriate if the spectral intensity at the target site is unknown. However, if one or more CPT soundings are made at the target site and the spectral intensity estimated, then the cutoff frequency can be estimated by matching the area under the spectral density function shown in Fig. 14 to the method of moments estimated average process variance. According to theory, the variance can be obtained from the spectral density function according to

J. Geotech. Geoenviron. Eng., 1999, 125(6): 486-498

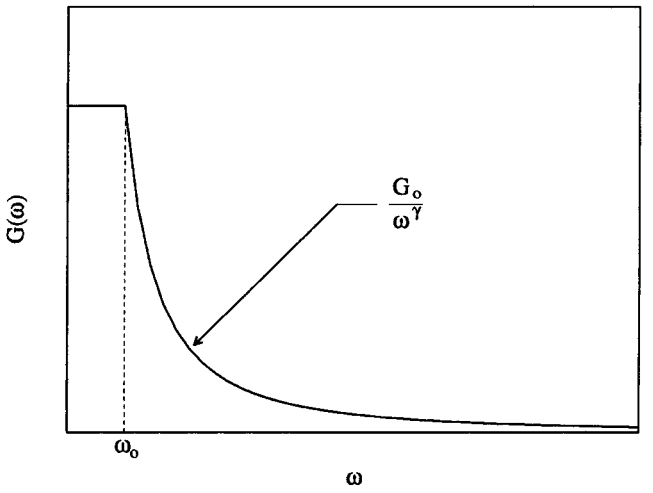

FIG. 14. Truncated Spectral Density Function for Fractal Processes with $\gamma > 1$

$$\sigma_x^2 = \int_0^\infty G(\omega)\, d\omega = \int_0^{\omega_o} \frac{G_o}{\omega_o^\gamma}\, d\omega + \int_{\omega_o}^\infty \frac{G_o}{\omega^\gamma}\, d\omega$$

$$= G_o \left(\frac{\gamma}{1-\gamma}\right) \omega_o^{1-\gamma} \tag{7}$$

for the $\gamma > 1$ case. This can be inverted to solve for the required lower cutoff frequency in terms of the estimated variance and spectral parameters

$$\omega_o = \left(\frac{\hat{G}_o \hat{\gamma}}{\hat{\sigma}_X^2(\hat{\gamma}-1)}\right)^{1/(\hat{\gamma}-1)} \tag{8}$$

Because of the fractal nature of the data, the cutoff frequency should be selected for a particular sample length D. Using the CPT sounding shown in Fig. 2 as an example, the $q_c^{0.74}$ readings are redrawn in Fig. 15(a). This particular sounding has length $D = 27.8$ m and parameter estimates $\hat{\gamma} = 1.81$, $\hat{G}_o = 3.85 \times 10^4$, and $\hat{\sigma}_x^2 = 7.77 \times 10^4$. For these values, the cutoff frequency is given by (8) to be $\omega_o = 1.13$ rads/m. This corresponds to a cutoff frequency 5 times that suggested in the absence of knowledge of the spectral intensity ($2\pi/D$). The frequency cutoff effectively controls how stationary realizations of the model appear. A smaller cutoff frequency leads to realizations with more pronounced apparent ''trends'' over the soil depth, although the trends in this case are actually random.

A realization of the resulting random process, with truncated spectral density function

$$G(\omega) = \begin{cases} \hat{G}_o/\omega_o^{\hat{\gamma}} & \text{if } 0 \le \omega < \omega_o \\ \hat{G}_o/\omega^{\hat{\gamma}} & \text{if } \omega > \omega_o \end{cases} \tag{9}$$

using the parameters found above for the CPT data in Fig. 15(a), is shown in Fig. 15(b). This realization was produced using the fast Fourier transform method (Fenton 1994). In addition to its increased high frequency content, it has very much the same statistical nature as seen in Fig. 15(a). It should be noted that it is not expected to be identical as it is merely one possible realization. It does, however, include the high-frequency content essential to the fine-scale self-similarity of a true fractal process. If such detail is not desired, it can be eliminated by placing an upper bound on the spectral density function of (9) or by passing each realization through a low-pass filter.

Using the result that the scale of fluctuation is proportional to $G(0)$ (VanMarcke 1984), an equivalent scale of fluctuation

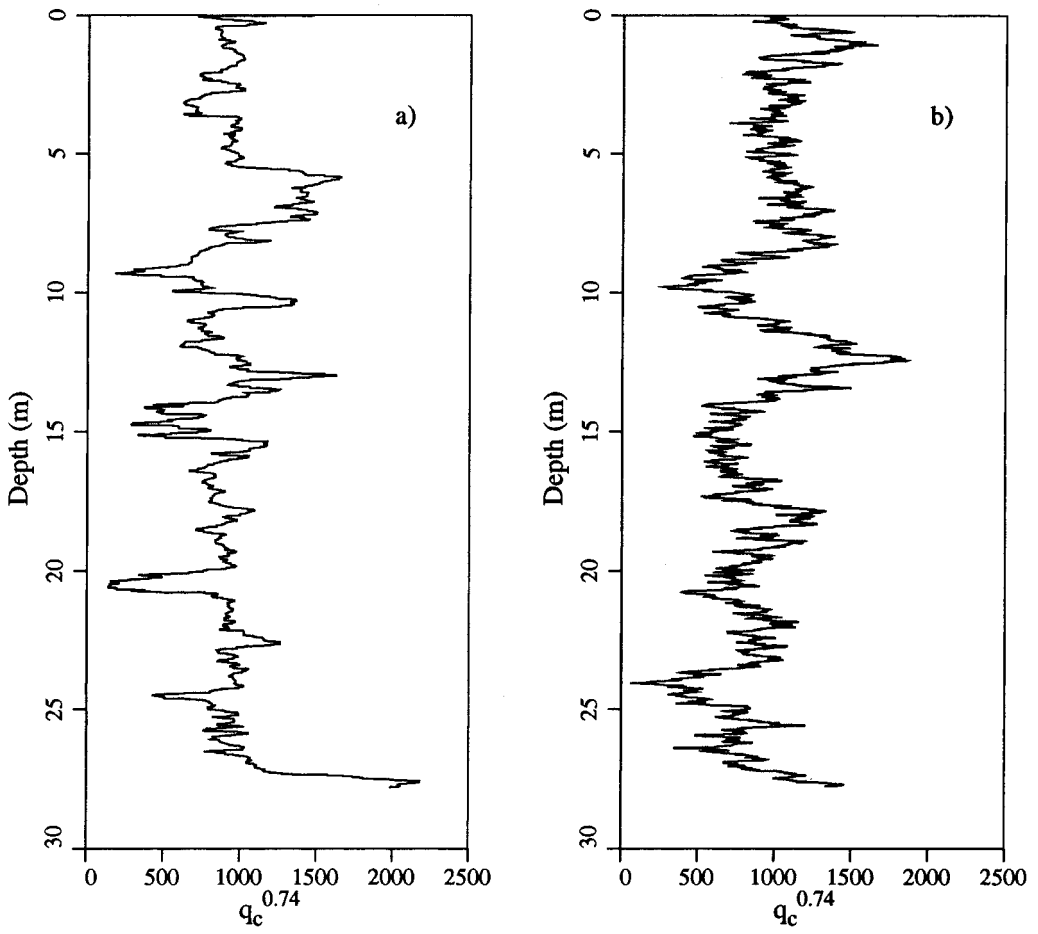

FIG. 15. (a) Observed $q_c^{0.74}$ Record from One CPT Sounding; (b) Simulated $q_c^{0.74}$ Record at Same Location Using $\gamma = 1.81$

J. Geotech. Geoenviron. Eng., 1999, 125(6): 486-498

can be associated with the lower cutoff frequency applied to the fractal model, as follows:

$$\theta^* = \frac{\pi \hat{G}_o}{\hat{\sigma}_X^2 \omega_o^{\hat{\gamma}}} \tag{10}$$

Interestingly, the equivalent scale of fluctuation for the sounding shown in Fig. 15(a) is computed to be $\theta^* = 1.24$ m, whereas its maximum likelihood estimate under the Gauss-Markov model is $\hat{\theta} = 3.3$ m. The Gauss-Markov model has a finite-scale, exponentially decaying correlation function whose decay rate is dictated by θ. Although these are quite different correlation models, the values of θ^* and $\hat{\theta}$ are reasonably similar. A brief investigation of the maximum likelihood scale estimates under the Gauss-Markov model indicates that $\hat{\theta}$ is highly variable, with an average over the 143 soundings of 2.3 ± 2.1 m, and a range from 0.22 to 13.8 m. Thus, the equivalent scale computed for the truncated fractal model is well within the range of those values estimated under the finite-scale model.

Once the cutoff frequency, G_o and γ are known, the model is fully specified, at least as far as the second moment is concerned. Unfortunately, no simple closed form expression for the covariance function corresponds to the truncated fractal spectral density function of (9), and so this quantity is most easily obtained by numerical integration using the Weiner-Khinchine relationship

$$C(\tau) = \int_0^\infty G(\omega)\cos(\omega\tau)\, d\omega \tag{11}$$

if it is desired. Because the integrand above contains the cosine function, it alternates in sign. Numerically, (11) is therefore prone to errors due to so-called catastrophic cancellation (loss of accuracy in the difference between two large numbers). The same integral can be written

$$C(\tau) = \frac{1}{\tau} \sum_{k=0}^{\infty} \int_0^{\pi/2} \Delta_G(k, u)\cos u\, du \tag{12}$$

where

$$\Delta_G(k, u) = G\left(\frac{2\pi k + u}{\tau}\right) + G\left(\frac{2\pi k + 1.5\pi + u}{\tau}\right) - G\left(\frac{2\pi k + \pi + u}{\tau}\right) - G\left(\frac{2\pi k + 0.5\pi + u}{\tau}\right) \tag{13}$$

which, still involving differences, can at least be analytically approximated when the sum of the first two terms is very similar to the sum of the second two terms. For the case where $G(\omega) = G_o/\omega^\gamma$ the following approximation is appropriate when k is large

$$\Delta_G(k, u) = \frac{1}{2}\pi^2\gamma(\gamma + 1)(2\pi k)^{-(\gamma+2)} G_o \tau^\gamma \cdot \left[1 - \frac{\gamma + 2}{6\pi k}\left(\frac{9\pi}{4} + u\right)\right] \tag{14}$$

The remaining integration in (12) can be performed as usual (e.g., by using Gaussian quadrature).

CONCLUSIONS

The primary result of this paper lies in the observation that the vertical variation of CPT q_c data appears to be fractal in nature. If CPT soundings are available at the target site for which the stochastic soil model is to be applied then all of the fractal parameters, including the spectral exponent, can be estimated at the site. In that case, the major use of this study is in establishing the basis for the use of the fractal model and an indication of the variability of the estimators.

Alternatively, if data taken at a similar target site is sufficient only to establish the mean and variance of the desired soil property, then an a priori second-moment model for the vertical soil variability would have $\gamma = 1.8$, a lower frequency cutoff equal to about $2\pi/D$, where D is the soil depth, and a spectral intensity G_o computed so that the area under the spectral density function is equal to the estimated variance [see (7)].

Clearly there is much work yet to be done on the inferential characterization of soils. For one, this study only considers a single site, and so its results are still somewhat limited. The issues related to the choice of a lower cutoff frequency need additional clarification as this parameter is still somewhat arbitrary in nature. Also needed is a good error model to distinguish between the real soil behavior and measurement error. Because such a model involves the estimation of additional parameters, an even more extensive database may be required to make confident inferences. It is believed, however, that the methodologies and reasoning set out in this paper and a companion paper (Fenton 1999) lay the groundwork for additional inferential studies on 1D variation in soil properties. In higher dimensions, fractal models will presumably still apply, so that researchers can concentrate on estimation issues.

ACKNOWLEDGMENTS

The writer would like to thank the Norwegian Geotechnical Institute (NGI) and, in particular, Suzanne Lacasse and Farrokh Nadim, for their help interpreting and using the CPT data, for their comments on the early research, as well as for their hospitality during the summer of 1997. Thanks are also due to Odd Gregersen, who paved the way behind the scenes and helped gather the site information, as well as to Tor Sager who made the CPT data available in electronic form. The writer extends his gratitude to the Research Council of Norway for their generous support and for making the project possible. After visiting NGI, 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 Natural Sciences and Engineering Research Council of Canada for their support under operating Grant OPG0105445. The use of the CPT data in this publication was permitted by NGI, the Oslo Airport Authority (Oslo Lufthavn A/S), and by the Norwegian Aviation Administration.

APPENDIX I. REFERENCES

Beran, J. (1994). "Statistics for long-memory processes." *Monographs on statistics and applied probability*, Chapman & Hall, New York.

Fenton, G. A. (1994). "Error evaluation of three random field generators." *J. Engrg. Mech.*, ASCE, 120(12), 2478–2497.

Fenton, G. A. (1999). "Estimation for stochastic soil models." *J. Geotech. and Geoenvir. Engrg.*, ASCE, 125(6), 470–485.

Gregersen, O. (1997). "Soil conditions Gardermoen." *Tech. Note to S. Lacasse*, Norwegian Geotechnical Institute, Oslo.

Hoeksema, R. J., and Kitanidis, P. K. (1985). "Analysis of the spatial structure of properties of selected aquifers." *Water Resour. Res.*, 21(4), 563–572.

Lumb, P. (1966). "The variability of natural soils." *Can. Geotech. J.*, Ottawa, 3(2), 74–97.

Lunne, T., Robertson, P. K., and Powell, J. J. M. (1997). *Cone penetration testing in geotechnical practice.* Blackie Academic and Professional, New York.

Priestley, M. B. (1981). *Spectral analysis and time series*. Vol. 1, Univariate Series, Academic, New York.

Sandven, R., and Watn, A. (1995). "Soil classification and parameter evaluation from piezocone tests: Results from the major site investigations at Oslo Main Airport, Gardermoen." *Proc.*, CPT'95, Vol. 3, 35–55.

Sudicky, E. A. (1986). "A natural gradient experiment on solute transport

in a sand aquifer: Spatial variability of hydraulic conductivity and its role in the dispersion process.'' *Water Resour. Res.*, 22(13), 2069–2083.
VanMarcke, E. H. (1984). *Random fields: Analysis and synthesis*. MIT Press, Cambridge, Mass.
Watn, A., Tuttle, K., and Sandven, R. (1995). ''Stratigraphic and sedimentological modelling at Oslo Main Airport, Gardermoen.'' *Proc., 12th Eur. Conf. of Soil Mech. and Found. Engrg.*, Vol. 6, 221–227.
Wornell, G. W. (1996). *Signal processing with fractals: A wavelet based approach*. Prentice-Hall, Englewood Cliffs, N.J.

APPENDIX II. NOTATION

The following symbols are used in this paper:

a, b, r = constants;
$\hat{C}(\tau_j)$ = estimated covariance function at discrete lag τ_j;
D = CPT sample length;
D_f = fractal dimension;
d = representative length;
f_s = side friction from CPT data (kPa);
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;
m = scale (dilation) index for wavelet basis;
n = number of observations in sample of $X(z)$;
q_c = cone resistance from CPT data (kPa);
$\hat{V}(\tau_j)$ = estimated variogram;
X_i = random value of transformed CPT data at z_i;
$X_{i,j}$ = average of X_j, X_{j+1}, . . . , X_{j+i};
X_j^m = random wavelet coefficient;
x_i = observed value of transformed CPT data at z_i;
z = depth coordinate;
z_i = discrete points along z;
γ = spectral exponent in fractal model;
$\hat{\gamma}$ = estimated spectral exponent;
$\hat{\gamma}(i)$ = sample discrete variance function;
Δz = incremental distance between observations;
$\Delta_G(k, u)$ = differencing operation on $G(\omega)$;
ε = residual random process;
θ = scale of fluctuation;
$\hat{\theta}$ = estimated scale of fluctuation;
θ^* = equivalent scale of fluctuation;
μ = pore-water pressure from CPT data (kPa);
μ_X = mean of $X(z)$;
$\hat{\mu}_X$ = estimated mean of $X(z)$;
$\hat{\rho}(\tau_j)$ = estimated correlation function at discrete lag τ_j;
σ_X^2 = variance of $X(z)$;
$\hat{\sigma}_X^2$ = estimated variance of $X(z)$;
τ = separation distance;
τ_j = discrete separation distance, $=j\Delta z$;
ω = frequency or wavenumber; and
ω_o = lower frequency cutoff.

J. Geotech. Geoenviron. Eng., 1999, 125(6): 486-498