Cone penetration test for site characterization

La caractérisation d'un site à l'aide du pénétromètre

R.G.CAMPANELLA, Department of Civil Engineering, University of British Columbia, Vancouver, B.C., Canada
D.S.WICKREMESINGHE, Department of Civil Engineering, University of British Columbia, Vancouver, B.C., Canada
H.J.ECHEZURIA, Intevep, S.A., Caracas, Venezuela

SYNOPSIS: Several statistical methods have been applied to Cone Penetrometer Test (CPT) data in order to better characterize a soil stratum. While trend analysis is used to better characterize soil property variation and identify sublayers within a soil stratum, the traditional method of identifying layers using the friction ratio method is compared to a method based on the scale of fluctuation of a sublayer. The random measurement error in data from various test methods has been obtained using two different techniques and a comparison has been made for those obtained from these methods. A method of two dimensional soil property interpolation considering correlation between different points is presented and applied to some data obtained from the CPT. The need for the consideration of the variability of a soil profile prior to determining the optimum sampling spacing is also highlighted.

1 INTRODUCTION

The natural variability of the soil, the limitation of available data, soil disturbance while testing or sampling and measurement errors, all contribute to the uncertainty of soil property evaluation. If every point in the ground could be tested, soil properties could be known at all intended locations. However, in practice this is not feasible and the need arises to treat this variation as random. In this regard statistics and probabilistic methods become a very appropriate tool in characterizing such variations.

This presentation deals with the Cone Penetrometer Test (CPT), the soundings of which are primarily used to identify soil stratigraphy. All data analyzed herein have been either obtained from the McDonald Farm in situ research site of the University of British Columbia or from its newer site adjacent to the Arthur Laing Bridge, both situated at the Vancouver International Airport on Sea Island. The data have been acquired at 2.5 cm intervals using a cone with a base area of 10 sq.cm, penetrating at 2 cm per second. Trend analysis to characterize different types of layering, the concept of the scale of fluctuation and methods of eliminating random error will be discussed in the initial sections of the paper together with a method of determining the optimum number of samples required to identify a layer of soil, at a certain confidence level and precision. The last part of the paper will discuss the interpolation of soil property values in two dimensions considering the correlation between points. The importance of the autocorrelation function and the variogram function will be highlighted with an application to a two dimensional interpolation problem.

2 TREND ANALYSIS

The main purpose of the CPT is to identify different types of soil layers in a stratum. Soil properties are highly depth dependent, and in most profiles a significant depth dependency is observed as can be seen from Fig. 1. The breaks in the trend will also indicate the different kinds of layering and a closer examination of the cone bearing log will indicate the approximate layer start and layer end depths of the sublayers within the entire profile. At present the friction ratio is often used to identify different layers (Fig. 2). However, the friction ratio did not identify the trends shown in Fig. 1. For example, the change in layer trend at 10 m depth is not identified in Fig. 2. Once the sublayer boundaries are decided, linear regression is performed on the data to obtain the trend line. The correlation coefficient is also determined to assess the efficiency of the fit. It has been found by the authors that the simple statistical parameter, the coefficient of variation, is a very good indicator for the above purpose of identifying different types of layering. The coefficient of variation is the ratio between the standard deviation and the mean, and its variation with depth is given in Fig. 3 which illustrates that the different layers identified possess different degrees of variation, the average values of which are tabulated in Table 1. The trend lines of the different layers are illustrated in Fig. 1.

The importance of trend analysis in two and three dimensional analysis will be discussed later. Most dimensional statistical methods deal with stationary data (data with no trend) which is referred to as homogeneous data in two and three dimensional analysis. Non-stationary data with which geotechnical engineers deal with regularly, can be transformed to stationary data by removing the trend as follows:

$$\text{RESIDUAL} = \text{DATA} - \text{TREND} \qquad (1)$$

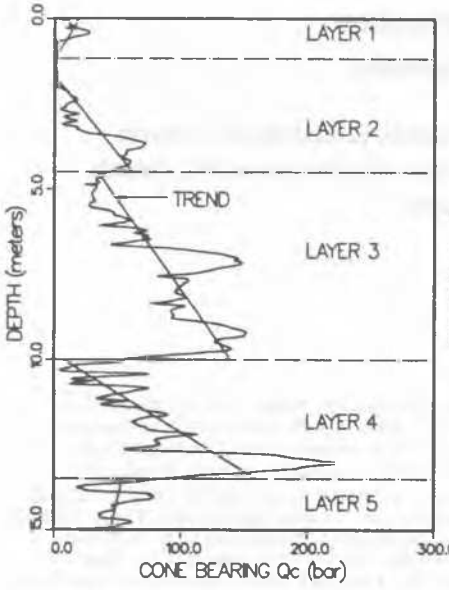

Figure 1. Cone bearing and trend lines of different sublayers, McDonald Farm site

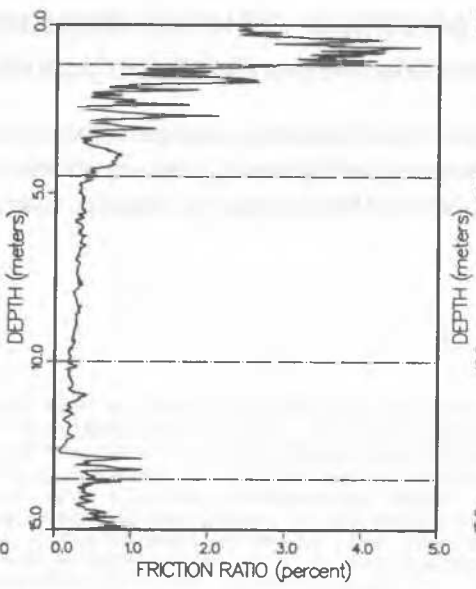

Figure 2. Variation of friction ratio with depth, McDonald Farm site

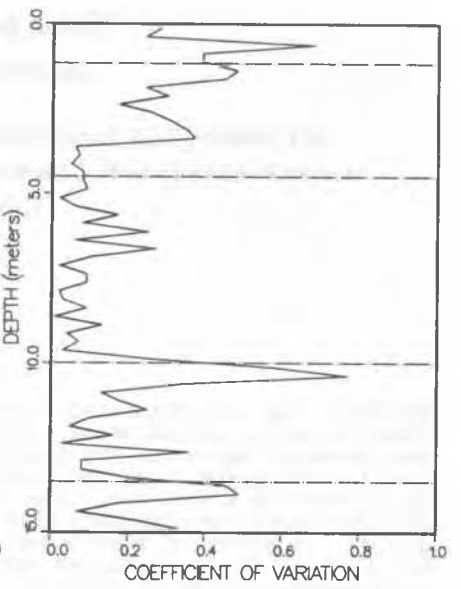

Figure 3. Variation of coefficient of variation with depth, McDonald Farm site

Table 1. Variation of scale of fluctuation and average coefficient of variation for profile in Fig. 1

Parameter	Layer 1	Layer 2	Layer 3	Layer 4
Scale of fluctuation, δ (cm)	20	60	72	40
Average coefficient of variation	.21	.17	.09	.19

The deterministic component, the trend, may be obtained by some form of a least squares regression technique. The appropriate trend for any profile provides a fit that gives the best correlation coefficient and the least variance. The residual is the stationary component which is used in correlation analysis, for various purposes such as interpolation. The final estimated value in such a procedure is the sum of the regressed trend term and the correlated residual term.

3 SCALE OF FLUCTUATION

In order to describe a soil profile completely, the scale of fluctuation, δ, is required in addition to the mean and standard deviation. The scale of fluctuation gives an indication of the degree of variability of a profile. A highly variable profile will have a low δ while a slowly varying profile will result in a high δ. The scale of fluctuation of any stratum is inversely proportional to the coefficient of variation, as indicated in Table 1. The scale of fluctuation is also referred to as the distance of perfect correlation since it is the distance within which the soil property shows relatively strong correlation from point to point. When two different test methods are being compared, it is recommended that the sampling distance be less than δ, so that comparison is being done in a region of perfect correlation. The opposite is true when sampling is performed using the same equipment, where for optimum sampling benefit, a spacing greater than δ is advisable.

The scale of fluctuation δ is determined by means of the variance function Γ^2 (Vanmarke, 1977), and is defined as follows;

$$\Gamma^2 = (\sigma_z/\sigma)^2 \qquad (2)$$

where σ is the standard deviation of the entire layer and σ_z is the standard deviation of spatial averages of sublayers of different thickness z. A detailed explanation of the method obtaining Γ^2 is given in Campanella et al (1984). When z is very large the value $\Gamma^2 \cdot z$ will approach the value of the scale of fluctuation (Vanmarke, 1977); i.e., for very large lag distance, z, the scale of fluctuation, δ, equals $\Gamma^2 \cdot z$.

The lag distance expressed as z in Fig. 4 is the separation distance between data points. Soil properties are correlated between points and this correlation is expected to decrease with increasing lag distance. This occurs because soil properties for closely spaced data would exhibit more similarities than for more widely spaced data.

The maximum values of the curves (Fig. 4) for the different layers of Fig. 1, are the respective scales of fluctuation for the different layers, and are also tabulated in Table 1. It can be observed from Table 1 that the soil that is most variable is layer 1, as given by the average coefficient of variation of 0.21. In keeping with the above argument it is the layer with the lowest δ of 20 cm. Similarly, layer 3 has the smallest coefficient of variation (0.09) and the highest δ of 71 cm. Therefore, for an efficient testing program

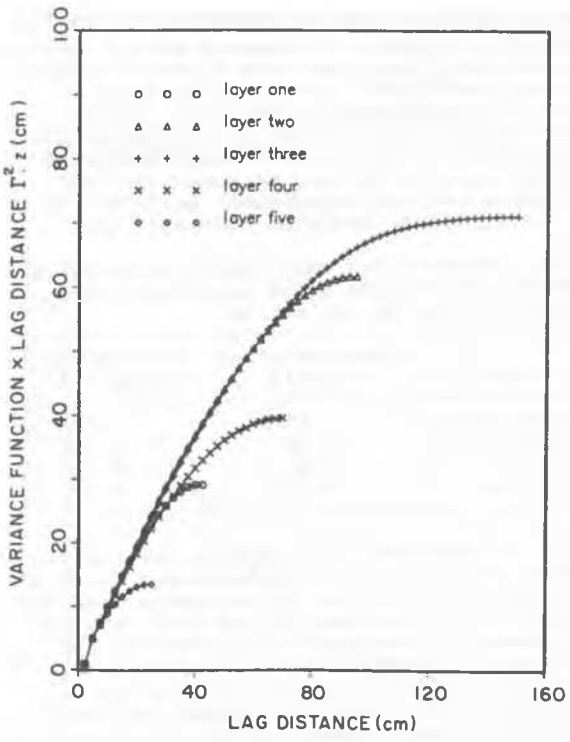

Figure 4. Scale of fluctuation of the different sublayers

using the same test equipment, sampling can be performed at spacings as much as 72 cm in layer 3, while in layer 1 it drops down to 20 cm, solely due to the fact of its high variability.

All values of δ so far discussed and appearing in Table 1 have been derived for the cone bearing value. Table 2 has these values compared to the δ values for sleeve friction. Since the sleeve fricton at a particular depth

Table 2. Comparison of the scale of fluctuation obtained for cone bearing and sleeve friction for profile in Fig. 1

Scale of Fluctuation, δ	Layer 1	Layer 2	Layer 3	Layer 4
δ for bearing (cm)	20	60	72	40
δ for friction (cm)	20	53	64	33

is an average value over a finite length one might expect a much higher δ (due to a lower variability caused by the averaging effect) for sleeve friction than for cone bearing. However, this will be true only if the cone bearing measures the bearing at a point. Results in Table 2 indicate that the δ values for the two soil parameters are approximately equal and thus suggesting that the cone bearing too is indicative of an averaged value over a region which is almost the same as that of friction, and does not indicate the bearing at a particular point.

4 OPTIMUM SAMPLE SPACING

The CPT at UBC performs data logging at a depth interval of 2.5 cm. For logging purposes the spacing is ideal, because it provides almost a continuous profile. However, if a soil parameter is to be estimated from a profile the spacing may need to be adjusted, depending on the soil variability and the required confidence level of the estimate. It is assumed that any layer is fully characterized when the average value obtained from the data for that layer is within 10 percent (Δ) of the actual average which is unknown. The tolerance allowed is therefore 10%. In other words, the precision is 90%.

The sample size (n) needed to estimate the mean to the above precision, with a confidence level of $(1 - \gamma)$, can be shown to be given by,

$$n = (V t^{\gamma}_{n-1})^2 / \Delta^2 \qquad (3)$$

where V is the coefficient of variation and t^{γ}_{n-1} is the student's t-variate with (n - 1) degrees of freedom. It is seen from the above expression that the number of samples is dependent on the variability of the soil and the level of confidence and precision required for the estimate. Table 3 has tabulated some values for the profile in Fig. 1, to show the above relationship. The results in Table 3

Table 3. Spacing of samples for a given precision and confidence level for data in Fig. 1

Layer (m)	Tolerance %	Conf. level %	n	Spacing (cm)
4.50 - 5.0	±10	90	5	12.5
		95	7	8.3
Coefficient of variation		99	11	5.0
= 0.11	±5	90	15	3.5
(low variability)		95	20	2.5
		99	34	1.5
10.0 - 10.5	±10	90	13	4.2
		95	18	2.9
Coefficient of variation		99	30	1.7
= 0.20	±5	90	43	1.2
(high variability)		95	64	0.8
		99	110	0.5

show that for a tolerance of ±10% and a confidence level of 95%, the sample spacing required in the soil of high variability is 2.9 cm while for a soil of low variability it is 8.3 cm. However, if a higher confidence level of 99% is required in the more variable soil, spacing should be decreased to 1.7 cm. Similarly, if the engineer requires to reduce the tolerance by half, in order to increase the precision of the estimate, the sample spacing will have to be reduced to 0.5 cm, for the same confidence level of 99%. However, for the less variable soil, the sample spacing required for 99% confidence level and a precision of ±5% is 1.5 cm. This clearly illustrates the three factors which contribute to the selection of sample spacing in a soil stratum; namely, soil variability, precision of the estimate and confidence level of the estimate.

5 ERRORS IN TESTING AND DATA SCATTER

The scatter in geotechnical data is obtained from three sources; actual variability of soil properties, random measurement error and bias. The bias is a systematic error introduced by systematic influences in testing. Measurement errors with non zero mean are considered as biases. While there are suggested ways of removing the random error, the bias can be determined only in relative terms, that is in comparison with a result obtained more accurately using a better test method. For example, if the undrained strength of a clay has been determined both from the triaxial test in the laboratory and in the field from the vane test, the bias in the vane test can be determined relative to the laboratory values, or vice versa. In the absence of such results the only error that can be removed from data scatter is the random error. Random measurement error or "noise" is assumed to be independent from point to point, whereas actual soil properties are not. The covariance, C(h), which is used for this purpose is defined below.

The covariance at lag h, (C(h)), is given by,

$$C(h) = (N - h)^{-1} \sum_{i=1}^{N-h} (Q_i - Q_{av})\ (Q_{i+h} - Q_{av}) \qquad (4)$$

where Q_i is the measured soil property value at a point 'i', Q_{av} is the mean of the data and N is the total number of data. If the associated error at point 'i' is E_i, the actual value at that point $(Q_c)_i$ is given by:

$$(Q_c)_i = Q_i + E_i \qquad (5)$$

Considering covariances (C) of the above equation, it can be shown that,

$$C(Q_c) = C(Q) + C(E) \qquad (6)$$

Since the random error E is independent from point to point, C(E) will have a value not equal to zero only at zero distance, while the soil property variation will have a maximum value of C(Q) at zero distance and a slowly decaying function, with increasing distance. This is because soil properties at points closer to each other show a stronger correlation than points further apart. The random measurement error would therefore be the difference between the variance of the data (covariance at zero separation distance) and the value of the decaying covariance function, at the point where it meets the ordinate. When the covariance function is divided by the variance of the data the autocorrelation function results. This is the normalized version of the covariance function and has a maximum value of unity and therefore the resulting random error using the above technique will be expressed as a percentage. This method was first introduced by Beacher (1978) and has been performed on different test methods.

The random measurement error for the CPT was very low compared to that of the field vane (Table 4). The random measurement errors of various in situ test methods have also been obtained using Box-Jenkins (1976) methods of Time Series analysis (Wu et al, 1986) using the statistical package SAS. Although the results from this analysis were approximate, due to the assumptions employed, it compared well with the values obtained using the more rigorous method previously described. A low random error reflects the efficiency of the test method since it is directly related to the degree of disturbance during testing. This statistical technique can also be used to ascertain the effectiveness of any improvement performed on a test. Table 4 has tabulated values of the

Table 4. Random measurement error obtained for different test methods using autocorrelation analysis and time series method

Test method	Autocorrelation analysis	Time Series method (SAS)
Cone Penetration	5.0%	5.1%
Flat Dilatometer	5.8%	5.5%
Field Vane	36.0%	38.7%
Dynamic Cone	6.1%	8.2%

results for different test methods for the two different techniques. The measurement noise of 36% obtained for the vane is extremely high and indicates inconsistencies in the test results likely caused by sand and silt lenses at the site tested. Figures 5 and 6 illustrate typical variations of the autocorrelation function with lag distance obtained for the CPT and the field vane, respectively.

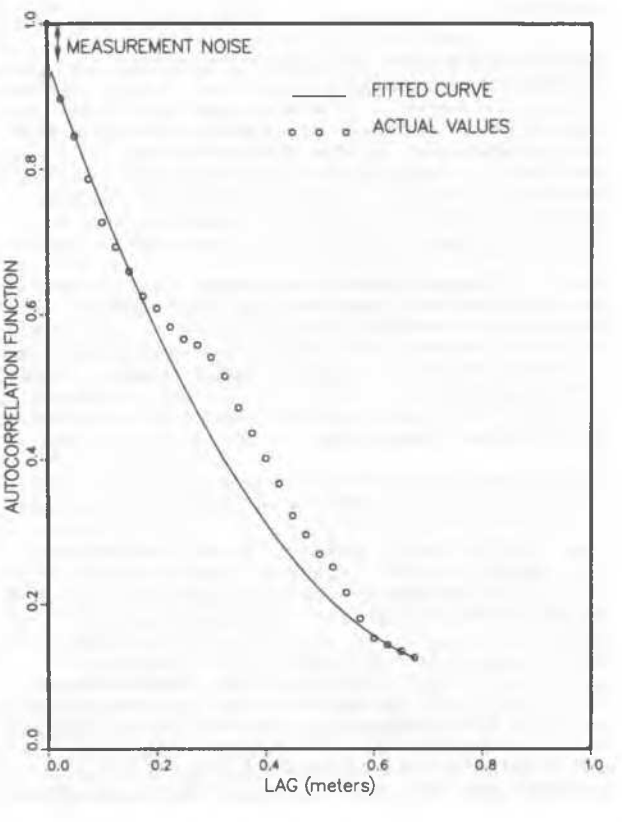

Figure 5. Autocorrelation function of CPT data

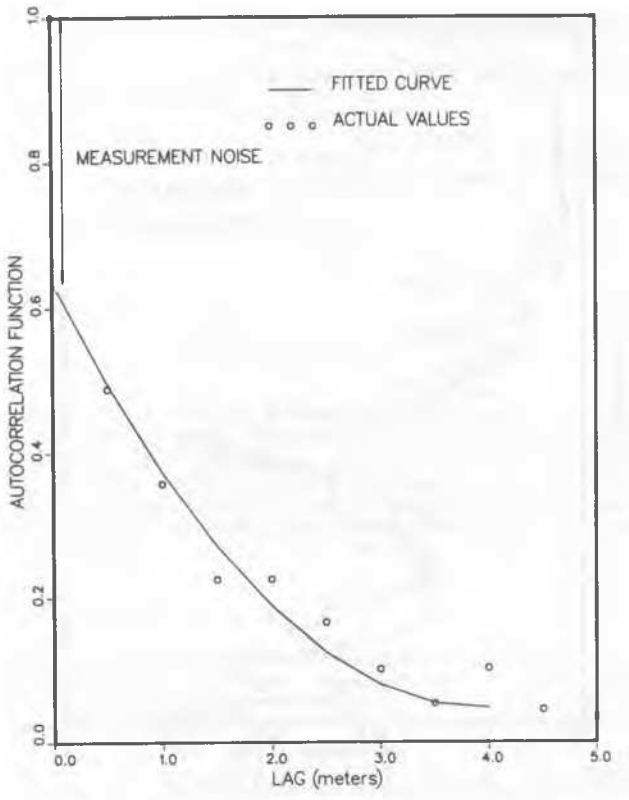

Figure 6. Autocorrelation function of vane data

6 INTERPOLATION OF SOIL PROPERTY VALUES CONSIDERING CORRELATION

As already expressed in an earlier section, the limitation of data availability at a site results in interpolating soil property values to obtain values at untested locations. Traditionally, geotechnical engineers are very conservative and would typically use very low bounds of the soil parameters in design and analysis. If the independence of soil property values between points is assumed, then interpolation methods such as least squares regression, distance weighting and simple weighting functions can be used to obtain values at untested points. For example, the mean of a data set can be obtained from a simple weighting function where the weights are all equal to the reciprocal of the number of data. An obvious shortcoming of these simplified methods, is that redundant information is not discriminated against. That is, a cluster of n data points located very close to each other, will each get the same weight as a single data point if they are located at the same distance from the point to be estimated. Therefore, when n is very large, the estimation will almost totally depend on the cluster of data points, completely neglecting the effect of the isolated single data point. This is actually a hypothetical extreme case, but clearly exemplifies the shortcoming of such methods. In contrast, methods which account for correlation overcome this drawback.

Soil property values, situated closer to each other are expected to be related more to each other, compared to points which are separated, wider apart. This relationship between data points is expressed by a correlation function, either in the form of the autocorrelation function ρ or the variogram function, γ, which are defined below.

The autocorrelation at a lag h, $\rho(h)$ is defined as:

$$\rho(h) = \frac{N}{N-h} \sum_{i=1}^{N-h} (Q_i - Q_{av})(Q_{i+h} - Q_{av}) / (Q_i - Q_{av})^2 \quad (7)$$

where Q_i are the data points, Q_{av} the mean of the data and N is the total number of data.

The variogram function at lag h, $\gamma(h)$, is defined as:

$$\gamma(h) = \sum_{i=1}^{N-h} (Q_i - Q_{i+h})^2 / N-h \quad (8)$$

To develop an autocorrelation function which closely resembles the actual process in the field, in an analytical form, a sizable data set will be required. However, in geotechnical engineering, this requirement will rarely be satisfied in a typical project, except for large projects like a site investigation for an offshore oil platform, an earth dam or a very high risk project. Therefore, it is very important that the test locations be chosen in such a way as to optimize the information that could be derived from the investigation.

The approach of separating the trend (if a trend exists) from the observed data, and performing interpolations on the correlated residuals has been carried out on CPT data sets obtained from the McDonald Farm site. Seven CPT's were performed at 5 meter spacings (Fig. 7). The profiles were divided into

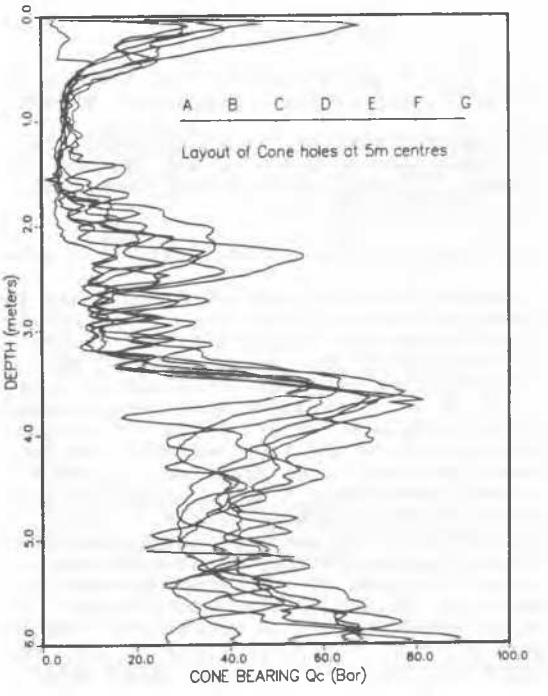

Figure 7. Variation of the cone bearing profile across site

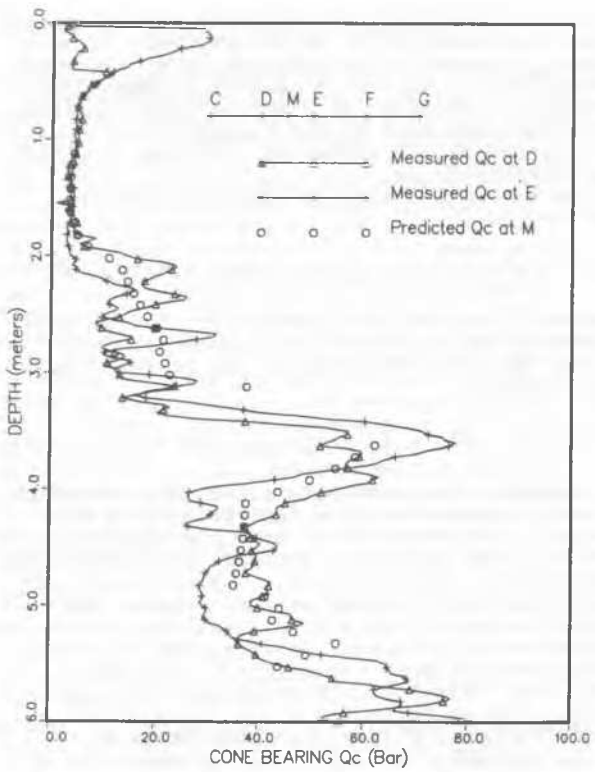

Figure 8. Interpolated profile at M

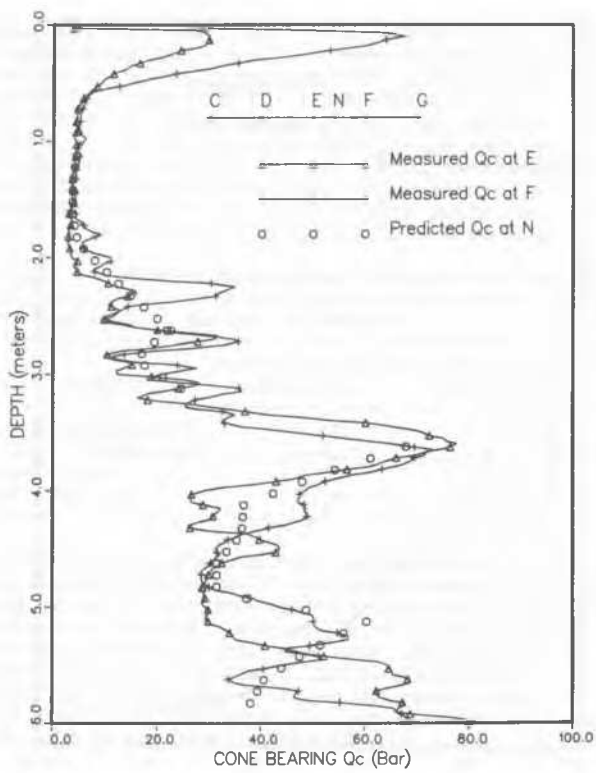

Figure 9. Interpolated profile at N

layers and where applicable, the trend was removed. The trend was fitted by regression with an equation of the form,

$$Q = A + By^2 + Cxy \qquad (9)$$

where A, B and C are constants, and x and y are the horizontal and vertical distances, respectively.

Several models were attempted to model the field autocorrelation function and the one with the closest fit was an exponential sinusoidal function of the form,

$$\rho(x,y) = \exp[-(x/q+y/r)]\cdot\mathrm{Cos}(x/q+y/r) \qquad (10)$$

where q and r are constants and x and y are the horizontal and vertical distances respectively. Fig. 7 also shows the variability across the site at McDonald's Farm, in the form of the seven cone bearing profiles, obtained at locations A, B, C, D, E, F and G which were spaced at 5 meter intervals. Figs. 8 and 9 illustrate the interpolated values together with the two immediately adjacent cone profiles. Point M is located exactly between D and E while point N is located 2 meters from E towards F. The interpolated values clearly indicate the effect of correlation because the values at M are clearly not the mean of the data at D and E. The same also applies to the interpolated profile at N. It should be emphasized that the autocorrelation or variogram function must be determined for the trend removed data, otherwise the interpolation would result in significant error.

7 ACKNOWLEDGEMENT

The financial support of the Natural Sciences and Engineering Research Council, Canada is gratefully appreciated.

8 REFERENCES

Box, G. & Jenkins, G.M. (1978). Time series analysis. Forecasting and Control. Holden-Day, San Francisco, CA.

Baecher, G.C. (1978). Analysing exploration strategies. Site Characterization, ed. C. H. Dowding, NSF-ASCE.

Campanella, R.G. & Wickremesinghe, D.S. (1978). Statistical treatment of cone penetrometer test data. ICASP5, Vancouver, B.C., Canada.

Vanmarke, E.H. (1977). Probabilistic modeling of soil profiles. ASCE, 103, GT11.

Wu, T.H. & El-Jandali, A. (1985). Use of time series in geotechnical data analysis. Geotechnical Testing Journal, ASTM, 8(4).