Spatial variability of SPT data using ordinary and disjunctive kriging

Pijush Samui, T.G. Sitharam
Department of Civil Engineering, Indian institute of Science, India

ABSTRACT: The purpose of this study is to develop a geostastistical model based on ordinary and disjunctive kriging technique to estimate spatial variability of SPT (N) data in the three dimensional subsurface of Bangalore. The database consists of 766 boreholes spread over a 220 sq km area, with several N values in each of them. The analysis has been done for corrected SPT (N_c) value. Ordinary kriging produces linear estimator whereas, disjunctive kriging produces nonlinear estimator. The knowledge of the semivariogram of the SPT data is used in the kriging theory to estimate the values at points in the subsurface of Bangalore where field measurements are not available. The capability of disjunctive kriging to be a nonlinear estimator and an estimator of the conditional probability is explored. A cross validation (Q1 and Q2) analysis is also done for the developed ordinary and disjunctive kriging model. For the data sets used in this study, disjunctive kriging has shown to be a better estimator than ordinary kriging in terms of reduced kriging variance and the comparison between an estimated and actual value.

1 INTRODUCTION

Information about spatial variability of soil properties is of great value in reliability analysis of geotechnical facilities and panning and optimization of soil exploration and testing. Also, it helps in selecting suitable construction procedures for soil engineering structure. In the literature, spatial variability of soil properties is generally studied by combining statistical analysis of site specific data with insights from random field theory (Vanmarcke 1977; Vanmarcke, 1983; Phoon and Kulhawy 1999; Uzielli et al, 2005). The standard penetration test is popularly used for field testing to characterize the subsurface soil profiles and field SPT value (N) derived from the test is used to determine the bearing capacity, settlement, liquefaction potential. N values are also correlated to many soil properties such as shear wave velocity, angle of internal friction, cone tip resistance, etc.

In this study, ordinary kriging and disjunctive kriging has been adopted to evaluate the spatial and depth variability of corrected N (N_c) in the three dimensional (3D) subsurface of Bangalore. The kriging method was developed during the 1960s and 1970s and has been acknowledged as a good spatial interpolator (Matheron, 1963; Isaaks et al, 1989; Davis, 2002). A major advantage of kriging is that it is more flexible than other interpolation methods such as inverse-distance weighting, deterministic splines and Thiessen polygons. The weights are not selected arbitrarily but are based on how a function varies in space. Recently, much attention has been given to linear kriging methods in the analysis of spatially dependent phenomena. These include simple, ordinary, and universal kriging. Generally, the linear kriging estimator is not the best possible estimator. The minimum variance unbiased estimator of a random variable Y in terms of random variable $Y, X_1, X_2, \cdots X_n$ is the conditional expectation of Y given $Y, X_1, X_2, \cdots X_n$. The knowledge of the joint density of $Y, X_1, X_2, \cdots X_n$ is required to determine the conditional ecpectation of Y. But, the joint density of $Y, X_1, X_2, \cdots X_n$ is difficult to obtain in practice. The conditional expectation is a function of n variables, that is,

$$Y^* = g(X_1, X_2, \cdots X_n) \quad (1)$$

It is also possible to relax the requirement that the joint density of (n+1) variables be known and define another nonlinear estimator

$$Y^{*}_{DK} = \sum_{i=1}^{n} f_i\left(X_i\right) \tag{2}$$

Where each f_i is a function of one X variable only. This is the disjunctive kriging(DK) estimator and requires only the bivariate densities be known.
The linear kriging estimator has a special form

$$Y^{*}_{K} = \sum_{i=1}^{n} \lambda_i X_i \tag{3}$$

In terms of estimating the value of a random variable at an unsampled location, equation (2) generally is a better estimator than equitation (3) in the sense of reduced kriging variance.
The paper has the following aims:

1. To investigate the feasibility of ordinary kriging and disjunctive kriging model for modelling the spatial variability of N_c in the 3D subsurface of Bangalore.
2. To estimate the variance of predicted data for ordinary kriging and disjunctive kriging model.
3. To determine conditional probability that the value of a N_c at a location is above a known cutoff level (N_c=30) for disjunctive kriging.
4. To develop a new type of cross-validation analysis for ordinary kriging and disjunctive kriging model.
5. To compare the performance of developed ordinary kriging and disjunctive kriging model.

2 GEOTECHNICAL DATA

A large amount of geotechnical data consisting of 766 boreholes has been collated along with index and engineering properties of subsoil layers at different locations in Bangalore (see Fig.1). Geotechnical data was evaluated for geotechnical investigations of several major projects in Bangalore. In total, 766 borelogs information have been entered into the data-base using a GIS with ARCINFO package. The sub-surface three dimensional (3D) GIS model of Bangalore has been developed with a scale of 1:20000. Fig.1 depicts a 1kmX1km grid within the corporate boundary, along with other boundaries, ring roads and borehole locations. It may be noted that an average of about two boreholes are available within each grid. Geotechnical data was collated from archives of Torsteel Research Foundation (India) and the Indian Institute of Science; these data were collected as part of several major projects in Bangalore during the years 1995-2003. The data in the model are on average to a depth of 30m below the ground level. The borelogs contain information about depth, density of the soil, fines content and N values and depth of ground water table.

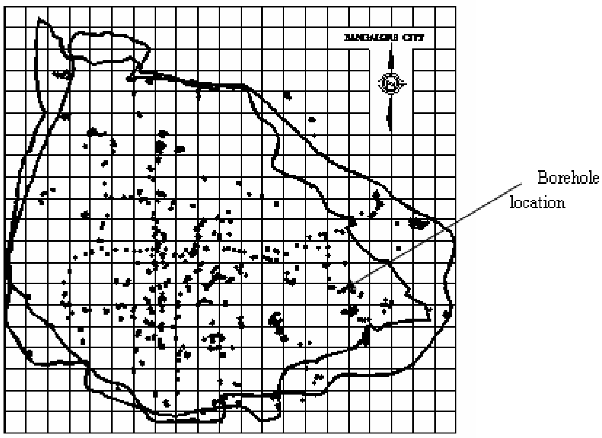

Fig.1 Borehole location in Bangalore Map (scale: 1:20000).

3 METHODOLOGY

In this study, two models (ordinary and disjunctive) have been adopted for prediction of N in the subsurface of Bangalore. Brief description of two models developed for our study is given below:

3.1 Ordinary kriging

Ordinary kriging is a geostatistical approach to modelling. Instead of weighting nearby data points by some power of their inverted distance, Ordinary kriging relies on the spatial correlation structure of the data to determine the weighting values. This is a more rigorous approach to modeling, as correlation between data points determines the estimated value at an unsampled point. In ordinary kriging, the most important thing is semivariogram. The semivariogram model used in this analysis is the spherical model. Spherical model converges to the sill value more quickly than the exponential model. A vertical anisotropy factor has been introduced to the semivariogram model for vertical dimension. Because of multiple layers of different properties, vertical variations are more significant compare to horizontal variation. The anisotropy is taken into account to avoid the distortion of the semivariogram relationship due to the large fluctuation over small vertical distance. A horizontal/vertical anisotropy factor has been used to weight the influence of horizontal samples more than the influence of vertical samples during prediction. Anisotropy factor depends on the soil type. Based on previous studies of other researchers on the spatial correlation in soil deposits (DeGroot, 1996; Phoon and Kulhawy, 1996), a factor of 20 has been chosen. In the literature, the range for horizontal to vertical anisotropy used for soil is 10-40 (DeGroot, 1996; Phoon and Kulhawy, 1996). a factor of 20 means that the known data values located vertically from a prediction point influence the prediction of the same data points located 20 times the distance horizontally from the prediction point. Once the model of semivariogram is constructed, the weights are computed for kriging. The details of ordinary kriging are given by Journel & Huijbregts,(1978).

3.2 Disjunctive kriging

Disjunctive kriging represents a form of nonlinear kriging(i.e., results in a nonlinear estimator) which in general offers an improvement over linear kriging methods, yet doesn't require knowledge of the n=1 joint probability distributions necessary for the conditional expectation. The disjunctive kriging estimator is made up of a sum of nonlinear functions and it can be written as (Yates, Warrick, and Myers, 1985a)

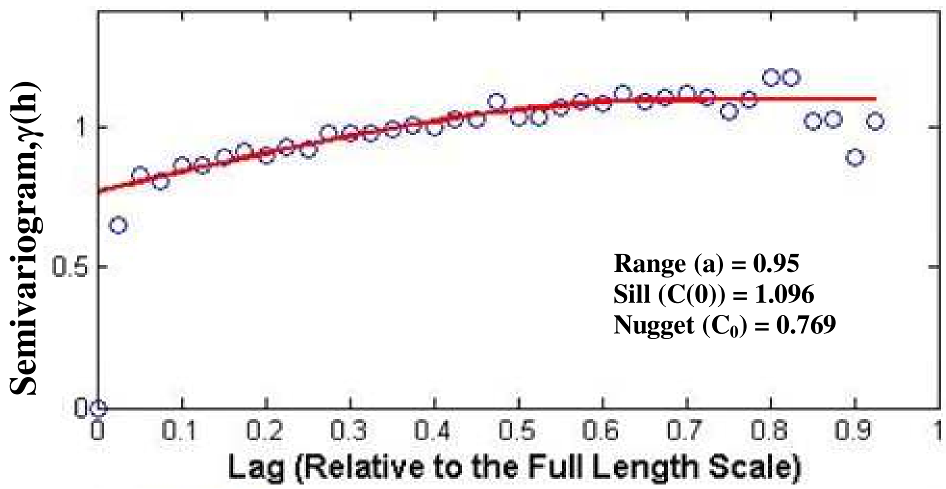

Fig.2 Semivariogram model for ordinary kriging

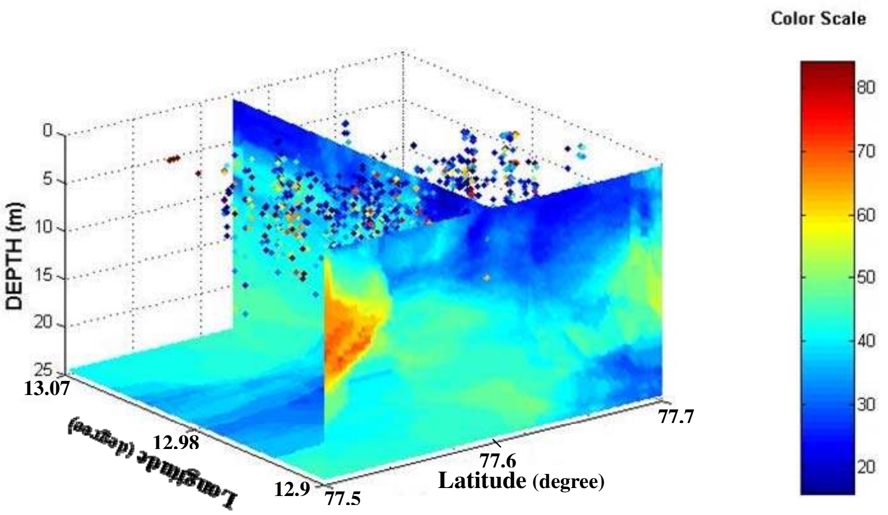

Fig.3 Kriged surface of N_c using ordinary kriging

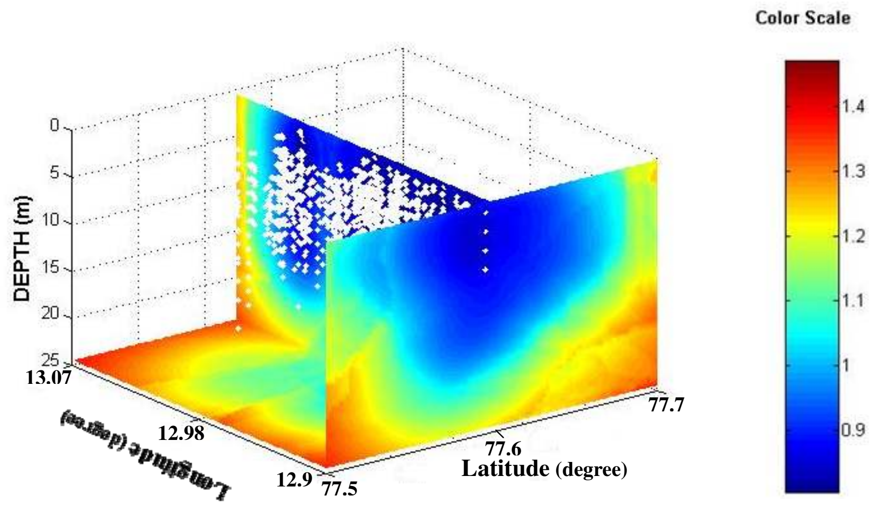

Fig.4 Variance map for ordinary kriging

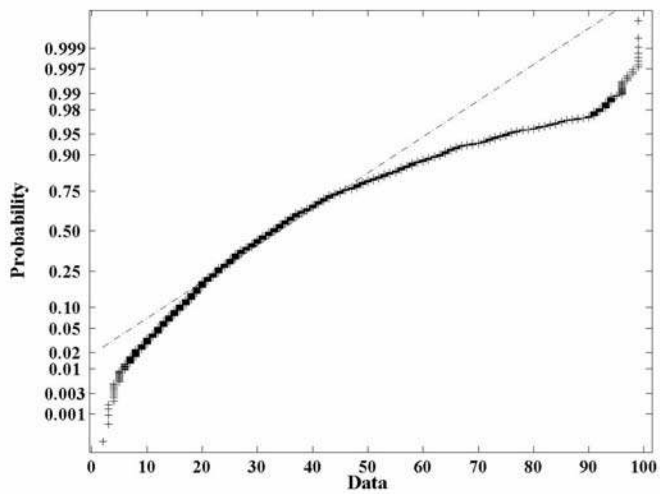

Fig.5 Test for normality

$$Z^*_{DK} = \sum_{i=1}^{n} f_i\left[Y(x_i)\right] = \sum_{k=0}^{\alpha}\sum_{i=1}^{n} f_{ik} H_k\left[Y(x_i)\right] \quad (4)$$

Where x is a vector in 3D space which represents the coordinates of the sample location, n is the number of transformed sample values, $Y(x_i)$, used in the estimation process, f_i is a function to be determined and expressed on the right hand side of equation (4) as a series of Hermite polynomials of order k where the f_{ik}'s are the coefficients of the Hermite expansion. Now, a transform function, $\phi[Y(x)]$, is necessary to transform Z(x) to a random function with a standard normal distribution. The transformation relationship, $(\phi(x))$, is written as a linear combination of Hermite polynomials (Abramowitz and Stegun, 1970)

$$Z(x_i) = \varphi\left[Y(x_i)\right] = \sum_{k=0}^{\alpha} C_k H_k\left[Y(x_i)\right] \quad (5)$$

The coefficients,C_k, are determined by orthogonality and numerical integration as follows

$$C_k = \frac{1}{\left(k!\sqrt{2\pi}\right)} \sum_{j=1}^{J} \varphi(y_j) w_j H_k(y_j) \exp\left[-y_j^2/2\right] \quad (6)$$

Where the y_j s are the abscissas and the w_j's are the weight factors for Hermite integration. Values of y_j and w_j are given by Abramowitz and Stegun(1970,p.924).

The disjunctive kriging estimator is defined as follows (Journael and Huijbregts, 1978; Rendu, 1980 and Yates, Warrick, and Myers, 1985a for further details)

$$Z^*_{DK}(x_0) = \sum_{k=o}^{K} C_k H_k\left[Y(x_0)\right] \quad (7)$$

Where the coefficients Ck are calculated from the sample value distribution as indicated earlier. In practice, only a finite number K of coefficients, Ck will be calculated (k=1, 2,…, K).

Hk[Y(x0)] in equation is estimated by

$$H^*_k\left[Y(x_0)\right] = \sum_{i=1}^{n} b_{ik} H_k\left[Y(x_i)\right] \quad (8)$$

Where the bik' are obtained by solving the following system of n linear equations

$$\sum_{i=1}^{n} b_{ik}\left(\rho_{ij}\right)^k = \left(p_{oj}\right)^k;\ j=1, 2,\ldots, n \quad (9)$$

Where ρij is the spatial correlation function for a separation distance, x_i-x_j. The correlation function also can be written as: $\rho_{ij} = \frac{C(x_i - x_j)}{C(0)}$, where $C(x_i - x_j)$ is the spatial covariance function or for second-order stationary conditions $\rho_{ij} = 1 - \frac{\gamma(x_i - x_j)}{C(0)}$, where γ is the semivariogram..

The disjunctive kriging variance on the estimation is

$$\sigma_{DK} = \sum_{k=1}^{K} k! C_k^2\left[1 - \sum_{i=1}^{n} b_{ik}\left(\rho_{oi}\right)^k\right] \quad (10)$$

When estimating N_c data at a particular location in 3D subsurface of Bangalore, estimating a conditional probability that the value of a N_c data is above a specified cutoff value(z_c) is possible because disjunctive krigiing estimator is nonlinear (Matheron, 1976; Journel and Huijbregts, 1978; Kim, Myers and Kundsen, 1977). A number of publications can be found in the literature which preset the procedure of the determination of conditional probability (Yates er al, 1986; Carr et al, 1986). The conditional probability is given by the following equation (Journel and Huijbregts, 1978; Yates, Warrick, and Myers, 1985a).

$$P^*\left[Y(x_0)\right] = 1 - G(y_c) + g(y_c) \times \sum_{k=1}^{K} H_{k-1}(y_c) \times \frac{H_k\left[Y(x_0)\right]}{k!} \quad (11)$$

Where, y_c is the associated transformed cutoff value of z_c , $g(y_c)$ and $G(y_c)$ are the Gaussian density and cumulative distribution functions, respectively. The conditional probability density function, Pdf*(u), is determined by taking the derivative of equation (12) with respect to y_c.

$$\mathrm{Pdf}^*(u) = g(u)\left\{1 + \sum_{k=1}^{K} \frac{H_k(u) H_k^*[Y(x_0)]}{k!}\right\} \tag{12}$$

4 CROSS-VALIDATION OF THE MODEL

Cross-validation of the model has been carried out before it is used for predictions. In practice, model validation based on statistical tests is dependent on the residuals. The detailed description of method of residuals in the case of kriging is given by Kitanidis (1997) and it is detailed as below:

It has been assumed that the n measurements are available at a time, in a given sequence and the kriging estimate of z at the second point (x_2) from the first measurement (x_1) is calculated. So, one can write $\hat{z}_2$=z(x_1) and $\sigma_2^2 = 2\gamma(x_1 - x_2)$. Where, $\hat{z}_2$ is the kriged value at the point x_2. The actual error (δ_2) $= z(x_2) - \hat{z}_2$ is normalized by the standard error (σ_2) and this normalized value of the error is given by:

$$\varepsilon_2 = \frac{\delta_2}{\sigma_2} \tag{13}$$

For the k-th measurement location, the actual error (δ_k) and normalized error (ε_k) can be written as respectively:

$$\delta_k = z(x_k) - \hat{z}_k \text{, for k=2,.....,n} \tag{14}$$

$$\varepsilon_k = \frac{\delta_k}{\sigma_k} \text{, for k=2,...,n} \tag{15}$$

Q1 (mean of the residual ε_k) and Q2 (variance of ε_k) are used to check the statistical distribution of the residuals between the observed data and the kriged values at the original observation location by using the kriging parameters and semivariogram model parameters. Q1 is the mean of the residual ε_k and it is written as:

$$Q1 = \frac{1}{n-1} \sum_{k=2}^{n} \varepsilon_k \tag{16}$$

Under the null hypothesis, Q1 is normally distributed with mean 0 and variance $\frac{1}{n-1}$. The probability density function (pdf) of Q1 is:

$$f(Q1) = \frac{1}{\sqrt{\frac{2\pi}{(n-1)}}} \exp\left(-\frac{Q1^2}{\frac{2}{(n-1)}}\right) \tag{17}$$

Where, n is the number of data. If the experimental value of Q1 turns out to be acceptable close to zero then this test gives no reason to question the validity of the model. The Q2 is the variance of ε_k and it is written as:

$$Q2 = \frac{1}{n-1} \sum_{k=2}^{n} \varepsilon_k^2 \tag{18}$$

(Q2)*(n-1) approximately follows the chi-square distribution with parameter (n-1). Where, n is the number of data points. The mean and variance of Q2 are 1 and $\frac{2}{n-1}$ respectively. The pdf of Q2 is given by the following equation:

$$f(Q2)=\frac{(n-1)^{\frac{n-1}{2}}\,Q2^{\frac{n-3}{2}}\exp\left(-\frac{(n-1)Q2}{2}\right)}{2^{\frac{n-1}{2}}\,\Gamma\left(\frac{n-1}{2}\right)} \tag{19}$$

Where, Γ is the gamma function. The experimental value of Q2 should be close to one.

5 RESULTS AND DISCUSSIONS

In case of ordinary kriging, the semivariogram of N_c obtained from the experimental values is shown in Fig.2. The spherical model has been plotted in Fig.2 and gives a reasonable fit to the values obtained. In the semivariogram, “relative to the full length scale” means normalized lag distance $\left(\frac{h}{a}\right)$. Where, h= lag distance and a= range of semivariogram. The range, sill and nugget of the semivariogram are 0.95, 1.096 and 0.769 respectively. The estimation of N_c has been done by using developed model of semivariogram (shown in Fig.2). The variability of the N_c data with spatial coordinate as well as depth of the Bangalore is represented by Fig.3. In Fig.3, the actual Nc data is also plotted and represented by the point marks. Fig.4 gives the associated estimation variance for ordinary kriging.

The sample mean, variance, skew, and kurtosis for N_c data were: 36.43, 367.27, 1.05 and 4.02 respectively. From the Fig.5, it is clear that the data set are not normally distributed. To calculate the C_k, all 2722 sample values have been used. C_k for k=0 to 9 are: 35.7350,-18.0881, 3.1662, 0.0768,-0.3176, 0.1048, 0.0046, 0.0803, 0.0001, and 0.0065 respectively. From these C_k’s, estimates of the mean and variance for the sample distribution has been obtained and are 35.7350 and 398.9866. The mean and variance of actual dataset are 36.4364 and 367.2766 respectively. The difference between actual mean and calculated mean from C_k’s is 1.92% and for variance, this difference is 8.63%.The semivariogram has been calculated using same anisotropy factor as used in ordinary kriging. A spherical model has been fitted to the semivariogram with parameters: 0.701 for nugget, 1.128 for sill and 0.95 for range. The autocorrelation function, which is used in Equation(9), has been calculated by using $\rho_{ij}=1-\frac{\gamma(x_i-x_j)}{C(0)}$. The estimation of N_c has been done by using developed model of semivariogram (shown in Fig.6). The variability of the N_c data with spatial coordinate as well as depth of the Bangalore is represented by Fig.7. In Fig.7, the actual N_c data is also plotted and represented by the point marks. The variance map for estimated N_c data has been given in Fig.8. The disjunctive kriging estimator has been used to determine the conditional probability the unknown value is greater than a specified cutoff value. Fig.9 shows the probability that the N_c is above 30. This information is unavailable typically when using ordinary kriging. Disjunctive kriging has also used to estimate the conditional probability density function and the cumulative probability distribution. In Fig.10 the cumulative probability distribution is with respect to the cutoff value whereas in Fig.11 the probability density is plotted as function of the transformed value Y. From this Fig.it is possible to obtain an indication of how the samples combine together to from the estimate as well as obtaining the probability level of an occurrence given a specified cutoff values.

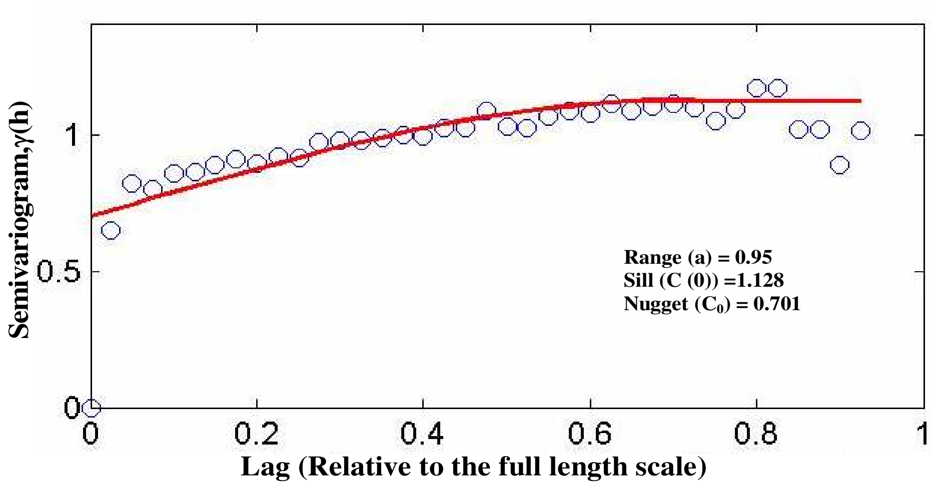

Fig.6 Semivariogram model for disjunctive kriging

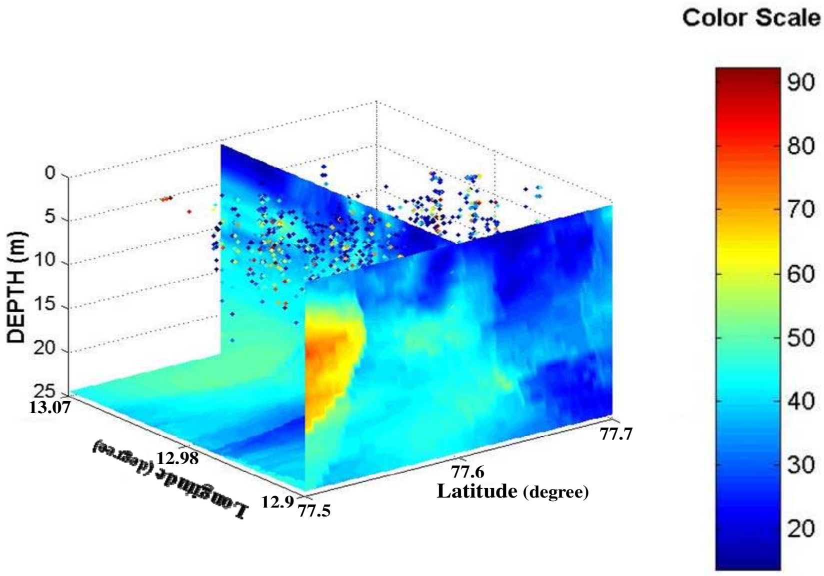

Fig.7 Kriged surface of N_c using disjunctive kriging

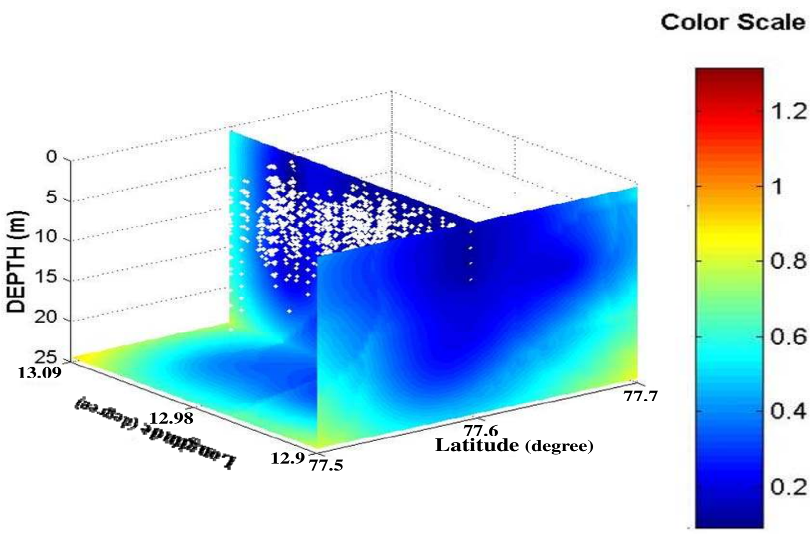

Fig.8 Variance map for disjunctive kriging

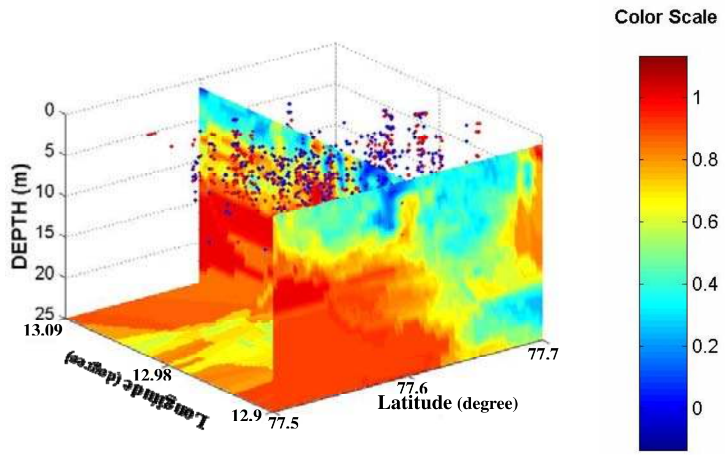

Fig.9 The map of the conditional probability that the N_c is above 30

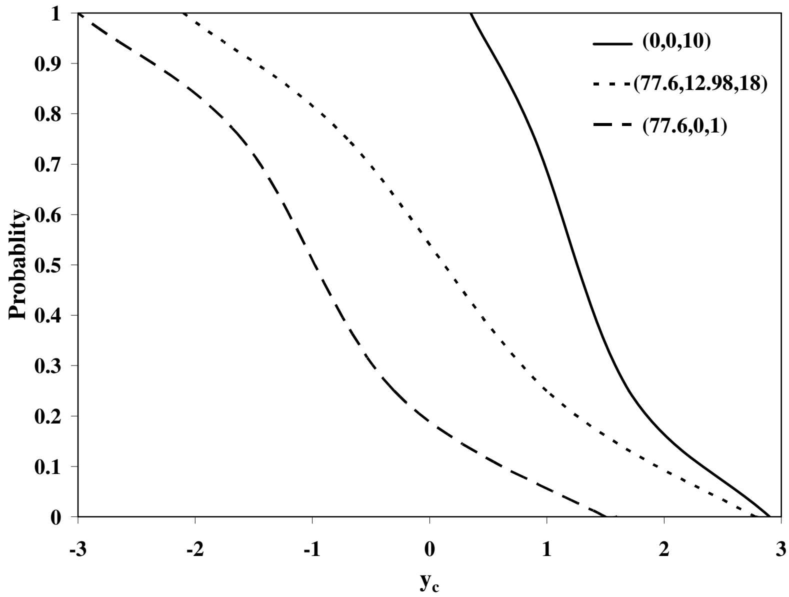

Fig.10 Cumulative probability distribution as function of cutoff value, y_c

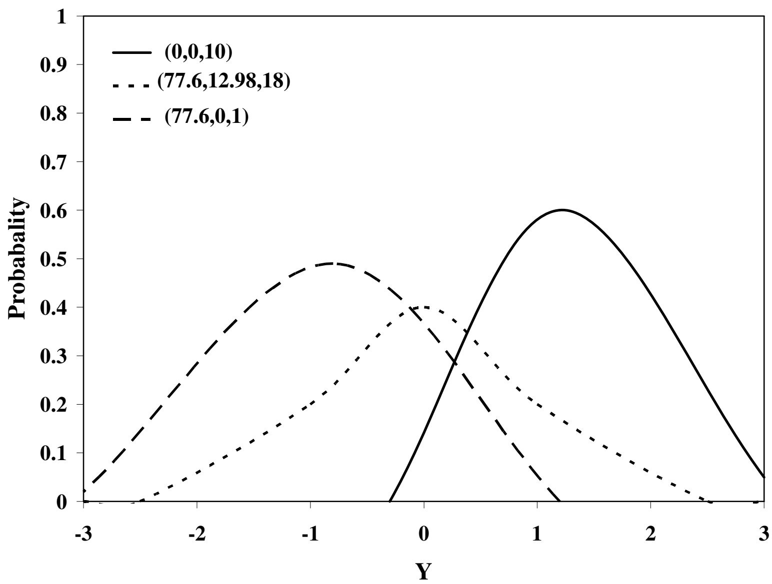

Fig.11 Probability density with respect to transformed variable, Y

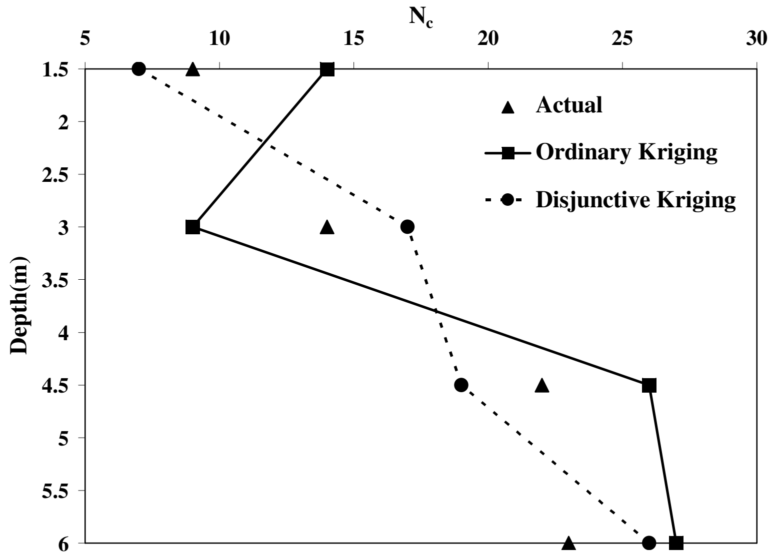

Fig.12 Comparison of actual N_c measured with predicted at BH 1253-1

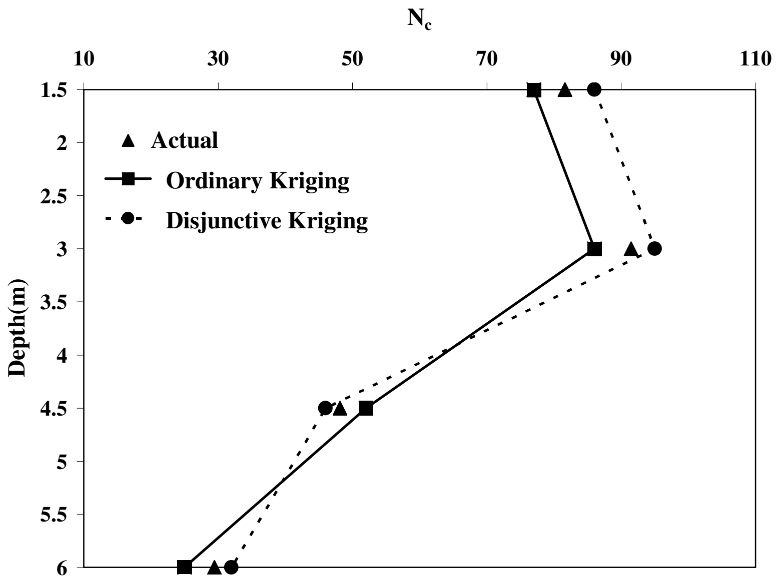

Fig.13 Comparison of actual N_c measured with predicted at BH 937-4

Kriging maps (Figures 3 and 7) provide a qualitative means for showing the difference between the ordinary kriging and simple kriging models. From Figures 4 and 8, it has been seen that the variance increases with increasing distance between estimated points and the actual point. The overall pattern of Figures 4 and 8 give an indication of where adequate or inadequate sampling occurred. In the Figures 4 and 8, it is clear that the variance of the estimated data from ordinary kriging analysis is always greater than that estimated data from disjunctive kriging analysis.

The values of Q1 and Q2 for both kriging are given in Table 2. For both the models, the values of Q1 and Q2 are close to 0 and 1 respectively. The cross-validation indicates that the developed ordinary kriging and simple kriging models estimate reasonably the N_c data in the 3D subsurface of Bangalore. However, the disjunctive kriging model seems to predict better than the ordinary kriging model.

Table 1 Result from cross validation analysis.

Model	Q1	Q2
Ordinary kriging	-0.018	0.987
Disjunctive kriging	-0.010	1.006

Two additional boreholes have been chosen in Bangalore for verifying the model results as a function of depth. For boreholes BH-1253-1, BH-937-4, Nc values have been predicted by ordinary as well as disjunctive kriging models with depth. Figures 12 and 13 show the Nc profiles with depth. It can be seen from the Fig.12 and13 that the disjunctive kriging model has given better prediction of the actual Nc data than the ordinary kriging model.

6 CONCLUSIONS

The spatial variability of $N_{corrected}$ in Bangalore has been modelled by ordinary kriging and disjunctive kriging. Cross-validation analysis has been also done for the developed ordinary and disjunctive kriging model. For the data sets used in this paper, disjunctive kriging has been shown to be a better estimator than ordinary kriging in terms of reduced kriging variance and comparision between an estimated and actual value. This result is expected since theoretically a nonlinear estimator should be equal to or better than a nonlinear estimator. Disjunctive kriging has been also explored to determine the conditional probability that the value is above an arbitrary cutoff level. This is an important result since it offers one geostatistical method whereby quantitative information is available to aid in management decisions. The geostatistical model provides valuable results that can be used for seismic hazard analysis, site response and liquefaction studies for the development of microzonation maps. The predicted 'N_c' values from the developed model can also be used to estimate the subsurface information, allowable bearing pressure of soils and elastic modulus of soils.

ACKNOWLEDGEMENTS

Authors thank Seismology division, Department of Science and Technology, Government of India for funding the project titled "Geotechnical site characterization of greater Bangalore region". Ref no. DST/23(315)/SU/2001 dated October 2003.

REFERENCES

Abramowitz, M., and Stegun, A.,(1965). Handbook of mathematical functions: Dover Publications Inc., New York, 1046p.

Carr, J.R., Deng, D.E., and Glass, C.E.(1986). An Application of Disjunctive Kriging for Earthquake Ground Motion Estimation. Mathematical Geology, Vol. 18, No. 2,

Degroot, D.J. (1996). Analyzing spatial variability of in situ soil properties."In: ASCE proceedings of uncertainty'96, uncertainty in the geologic environment: from theory to practice, ASCE geotechnical special publications no.58 p.210-38.

Isaaks, E.H., and Srivastava R.M. (1989). An Introduction to Applied Geostatics: Oxford University Press, New York.

Journel, A.G. & Huijbregts, C.J. (1978). Mining Geostatistics. New York: Academic Press.

Kim, Y.C., Myers, D.E., and Knudsen, H.P.,(1977). Advanced geostatistics in ore reserve estimation and mine planning(practitioner's guide): Report to U.S. Energy Research and Development Administration, Subcontract No. 76-003-E, Phase II, 154p.

Kitanidis, P.K. (1997). Introduction to Geostatistics: Applications in hydrogeology. Cambridge University Press, 86-95.

Matheron, G. (1963).Principles of geostatistics: Economic Geology. Volume58, 1246-1266.

Matheron, G.,(1976), A simple substitute for conditional expectation: the disjunctive kriging: Proceedings NATOASI "Gesostat 75", D. Reidel Publishing Co., Dordrecht, Netherlands, P. 221-236.

Phoon, K.K, and Kulhawy, F.H. (1996). On quantifying inherent soil variability. In: Proceeding of uncertainty '96, uncertainty in the geologic environment: from theory to practice, ASCE geotechnical special publication no.58 p.326-40.

Phoon, K.K, and Kulhawy, F.H. (1996). On quantifying inherent soil variability. In: Proceeding of uncertainty '96, uncertainty in the geologic environment: from theory to practice, ASCE geotechnical special publication no.58 p.326-40.

Rendu, J.M.,(1980). Disjunctive kriging: A simplified theory: Jour. Math. Geology, V. 12, No. 4, p.306-321.

Uzielli, M., Vannucchi, G. and phonn, K.K. (2005). Random filed chracterisation of strees-normalised cone penetration testing parameters. Geotechnique 55, No. 1, 3-20.

Vanmarcke, E.H. (1977). Probabilistic Modeling of soil profiles. J. Geotech. Engrg., ASCE, 102(11).1247-1265.

Vanmarcke, E.H. (1983). Random fields: Analysis and synthesis. The MIT Press, Cambridge, Mass.

Yates, S.R., A.W. Warrick, and D.E. Myers,(1986). Disjunctive kriging, 2, Examples, Water Rsor. Res., 22.

Yates, S.R., Warrick, A.W., and Myers, D.E.(1985a). Disjunctive kriging. I. Overview of estimation and conditional probability: Water Resource research, V.22, No.5 p-615-621.