Geotechnical Infrastructure Asset Management, 2009. EIT-JSCE Joint International Symposium 2009, September 7-8, 2009, Imperial Queen's Park Hotel, Bangkok, Thailand, Organized by EIT, JSCE, Kyoto University, AIT

SPATIAL DATA ANALYSIS BY USING GEOSTATISTICS EVALUATE THE UNDRAINED SHEAR STRENGTH OF SOFT BANGKOK CLAY

L. Prompong[1] and S. Soralump[2]

Doctoral student[1], Assistant Professor[2], Department of Civil Engineering, Faculty of Engineering, Kasetsart University, Bangkok, Thailand,
nong37016@hotmail.com[1], fengsus@ku.ac.th[2]

Abstract

Geostatistics is a set of method which evaluates the unknown quantities of one or more variables such as the undrained shear strength which is distributed in space. The model of the spatial structure of the undrained shear strength of soft Bangkok clay represented by using variogram model. The variogram is a function of distance vector , h , equal to the mean of $0.5 * [Z(X+h) - Z(X)]^2$ for all pairs of points Z (X) and Z (X + h) at distance , h , from each other. Estimating the undrained shear strength of soft Bangkok clay should be fitted with an appropriated variogram model in order to minimize the sum of square fitting errors. This research has collected the bore holes data over Bangkok region and to estimate the undrained shear strength. First , stratifies soft clay layer by 1-2 m interval and classifies into 3 groups as very soft , soft and medium to soft which have been associated to the value of undrained shear strength 0-1.25 , 1.25 - 2.25 and 2.25 - 5 ton/m^2 respectively. Secondly , each specific layer , for a single variable the basic statistics include histogram , the range of value , the mean , the variance , the standard deviation and the coefficient of variation. Since the sample variance can be written as an average of square differences between pairs of samples, which mean that sample variance equals the mean of the variogram values. The variogram will be bounded and incorporated with a sill (or sample variance) which is stabilized when it reaches some distance call the range. On the other hand , the range gives the distance from which correlation vanishes , and is the limit of influence of the samples. The research procedure is emphasized on the experimental variogram as a set of trying out directional variograms fitted with a model and appropriated boundary , which given the minimum errors. Kriging method which is well known in the term of geostatistics requires the best fit of variogram model to predict the optimal value and also probability of that value at any given point. As the optimal solution , the undrained shear strength has been estimated , that is , the appropriated value is furthermore reliable to the design of geotechnical engineer.

Keyword : Geostatistics , Variogram model , Sill , Range , Undrained shear strength , Kriging

Introduction

From history background has proved that Bangkok subsoil profile has been governed by soft clay in depth between 0 - 20 m., since it used to be the sedimentation of marine clay. As mention, such a soil characteristic the geotechnical engineer has been faced many problems involving soil data properties and their spatial variation. The uncertainties may be in the form of a lack of information about the subsoil profile or a large scatter in the soil test results and so on (Gordon A. fenton,1997). This research has collected the soil data properties and test results more than three thousand boreholes all over Bangkok and surrounding area (fig. 1). These soil data will be analyzed by mean of statistics approach and common statistics involving of sample mean, sample variance, sample standard deviation, coefficient of variation (COV.), correlation, distribution types, and probability density function(pdf.). Further more, in term of geostatistics and field modeling, the relationship among the soil properties both in horizon and vertical direction is deeply concern and this refers to the average trend, spatial averaging, autocovariance function, and scale of fluctuation.

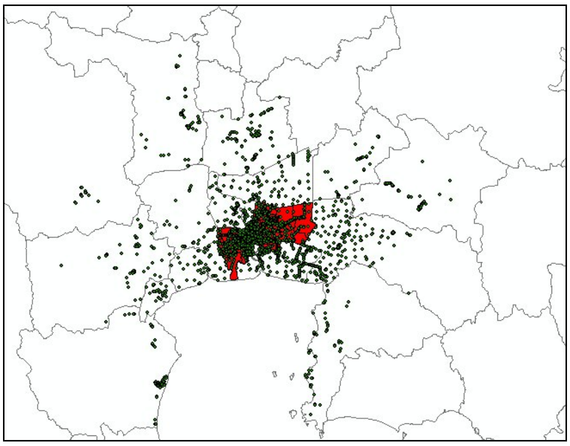


Fig.1 Position of bore holes over Bangkok and surrounding Area

Statistical treatment of soil data properties

Statistical analysis is restricted in the number of samples and can be approached by analytical mean(average) if a large number of sample is available. So statistical treatment of soil data properties will be applied to any single bore hole. This paper strictly concern the data obtained from field test of undrain shear strength of a soft Bangkok clay, since most of the soil data properties are continuous random variable, therefore the term of statistical treatment define as follow

Estimating the mean

The sample mean

$$\hat{\mu}_X = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Estimating the variance

The sample variance

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

The sample standard deviation(σ)

σ = square root of the sample variance

The sample coefficient of variance COV.(δ)

$$\hat{\delta}_X = \frac{\hat{\sigma}_X}{\left|\hat{\mu}_X\right|}$$

The sample correlation coefficient(ρ)

$$\hat{\rho}_{XY} = \frac{\sum_{i=1}^{n}[\{x_i - \hat{\mu}_X)(y_i - \hat{\mu}_Y)]}{\sqrt{\sum_{i=1}^{n}(x_i - \hat{\mu}_X)^2 \sum_{j=1}^{n}(y_i - \hat{\mu}_y)^2}}$$

The normal distribution(Gaussian)

In general, soil properties data may be assumed to be normally distribution. The undrain shear strength are fit well and tend toward normal distribution which are common in nature. Normal distribution has a symmetrical distribution with bell-shape curve and its tail decay in an exponential manner. There is a 68-percent chance within + 1 standard deviation from mean value, 95-percent chance within +2σ and 99.7-percent chance within +3σ

The probability density function(PDF.)

Continuous random variables, (undrain shear strength Su in this paper), there is an infinite number of possible values within the sample space which any value is greater than Zero. PDF describes its probability distribution and can define the probability of that variable value within a very small interval, thus, this probability is proportional to the PDF. which denote as f(x). For normal distribution

$$f_x(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}(\frac{x-\mu}{\sigma})^2}$$

$$-\infty < x < \infty$$

The cumulative distribution function(CDF.)

CDF. is the area under the PDF. describes the probability that the variable takes on a value less than or equal to a given value which denote as F(x)

$$F(x) = P[X < x]$$

Procedures

Exploratory Data Analysis

The preliminary exploration of the data should always be done before other formal analysis , by displaying them as histograms and scatter diagrams, and compute summary statistics. From any suspected observations data that are very different from their neighbors which might be true outliers, or errors of measurement , or recording mistakes they should be examined closely consideration. If the data are not approximately normal, they will be experimented with transformation to make them are normal and fitting theoretical distributions from the estimated parameters of the distribution to the histogram. If the histogram appears erratic another way of examining the data for normality is to compute the cumulative distribution and plot it against the normal probability on normal probability paper. This paper has an scaled that a normal cumulative distribution appears as a straight line. A strong deviation from the line indicates non-normality. The positions of the sampling points will be plotted and data should also be examined for trend .

Geostatistics methods

Spatial variability

The value of random variable under field conditions generally varies in time and space, and the ordinary statistical methods fail to explain this variation. Under such conditions, geostatistical methods are used to assess the variability. In a field, the values of a parameter, $Z(u)$, may exist in random or stationary form. Random variable is one of which varies with location within the same region. Spatial variability of field measured properties is commonly described with autocorrelation or semi-variogram.

Variogram

variogram represents the variation between pairs of measurement as function of separated distance(fig. 2, a, b,c), define as

$$2\gamma(h) = E\{[Z(u) - Z(u+h)]^2\}$$

$$2\gamma(h) = \frac{1}{N(h)} \sum_{N(h)} [Z(u) - Z(u+h)]^2$$

For semi-variogram

$$\gamma(h) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} [Z(u_\alpha + h) - Z(u_\alpha)]^2$$

where r(h) = semivariogram for lag distance h

N(h) = number of pairs for lag distance h

The variogram for lag distance h is defined as the average square different of values separated approximately by h , and lag distance should coincide with data spacing, considerably, the variogram is only valid for a distance one half of the field site.

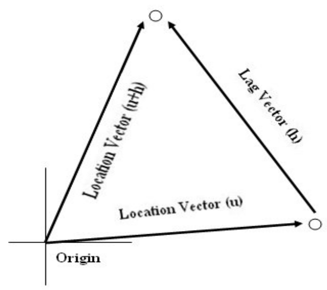


fig. 2(a) Lag distance h

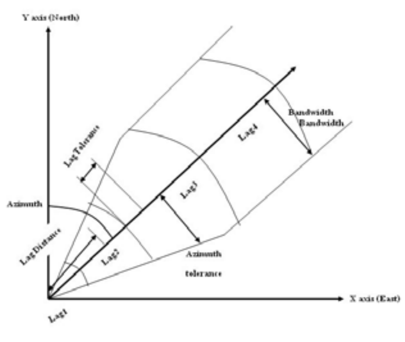


fig. 2(b)

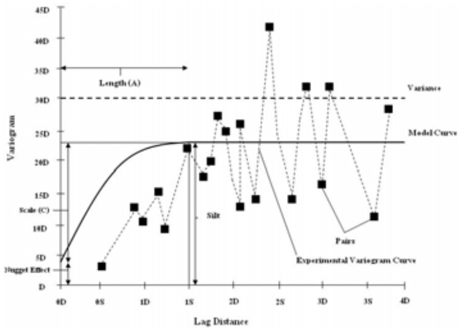


fig. 2(c) Variogram structure

Covariance Function

Covariance function and variogram are two converse concept. Variogram is model of spatial variability, but covariance function is model of spatial dependency, both are as function of separating distance. The covariance function will show decreasing of correlation as distance increase, conversely, variogram will show increasing of variability as distance increase. Covariance function define as

$$C(h) = \frac{1}{n(h)} \sum_{\alpha=1}^{N(h)} Z(u_\alpha) \cdot Z(u_\alpha + h) - m_0 \cdot m_{+h}$$

Where m_0 and m_{+h} are the means of the tail and head values:

$$m_o = \frac{1}{N(h)} \sum_{\alpha=1}^{N(h)} Z(u_\alpha) \quad \text{and}$$

$$m_{+h} = \frac{1}{N(h)} \sum_{\alpha=1}^{N(h)} Z(u_\alpha + h)$$

and correlation define as

$$\rho(h) = \frac{C(h)}{\sqrt{\sigma_0 \cdot \sigma_{+h}}}$$

σ_0 and σ_{+h} are the corresponding standard deviations:

$$\sigma_0 = \frac{1}{N(h)} \sum_{\alpha=1} [Z(u_\alpha) - m_0]^2$$

$$\sigma_{+h} = \frac{1}{N(h)} \sum_{\alpha=1}^{N(h)} [Z(u_\alpha + h) - m_{+h}]^2$$

Results

The experimental variogram of soil strength parameter (Su) at depth 3.5,5,6.5,8,9.5 m. from natural ground line, respectively, have been treated. For instance the histogram of undrained shear strength data at depth 3.5 m has been shown in fig.3. As we perform mathematical transformations to the data set to create a Gaussian (or normal) distribution from a non-Gaussian raw data set, a Gaussian distribution is desired because kriging assumes that the data are normally distribution. To check how the data are distributed, using the Normal QQ plot. The QQ plot shows the linear relationship between log(Su) and the standard normal distribution. In this case, the undrained shear strength data at depth 3.5 m follows a log normal distribution(fig.4) but in the other depth they follow either normal distribution or log normal distribution.

The trend analysis of the undrained shear strength at depth 3.5 m provides a regression on the XZ and YZ planes (fig.5) and this allow us to visualize the data and to observe any large-scale trends which may be removed prior to estimation. Practically, using kriging to predict only the residuals after the trend is removed and thus will optimize the kriging model.

Geostatistics analysis of the undrain shear strength , to this paper, at depth 3.5 ,5 , 6.5, 8, 9.5 m the kriging surfaces (Su) have shown in fig. 15 implied that they have significantly variation both in horizon and vertical direction and further study is to find out the horizon thickness of soil profile over the Bangkok area.

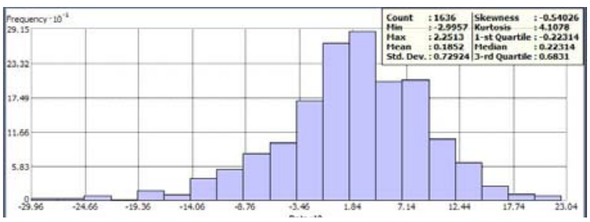


Fig.3 Histogram (Su) at depth 3.50 m

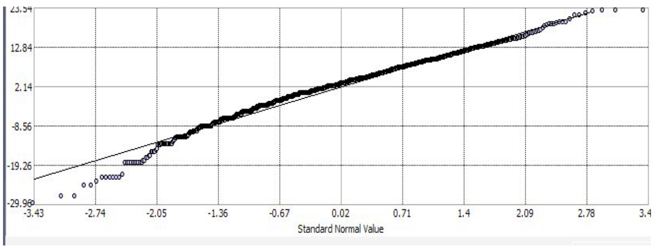


Fig.4 QQ-plot (Su) at depth 3.50 m

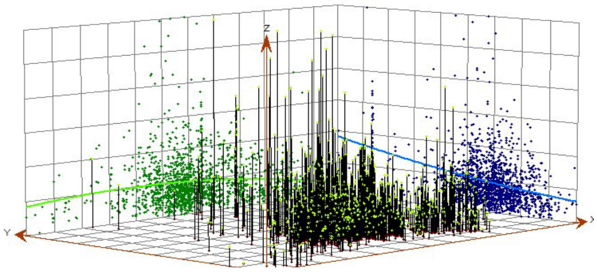


Fig. 5 Trend in XZ and YZ planes (SU) at depth 3.50 m

The results of geostatistics analysis such as variogram model, covariance model, prediction, error and cross validation model checking are shown in fig.(6,7,8,9,10,11,12,13,14)

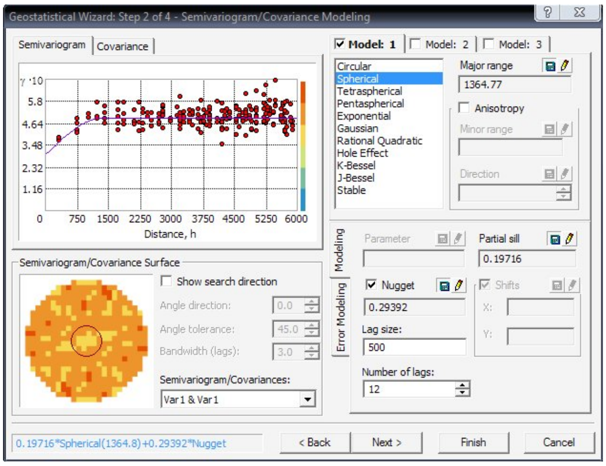


Fig. 6 Variogram model (Su) at depth 3.50 m

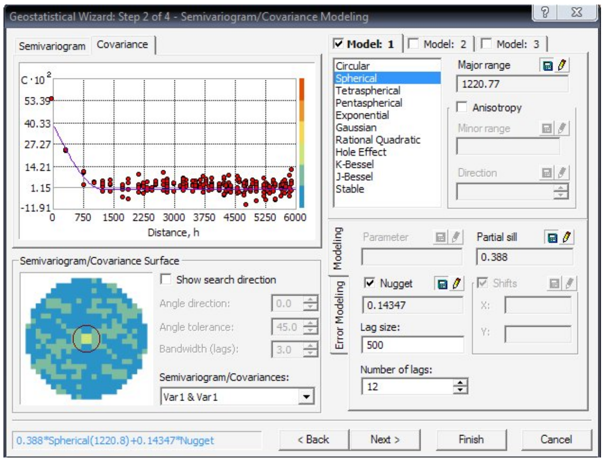


Fig. 7 Covariance model (Su) at depth 3.50 m

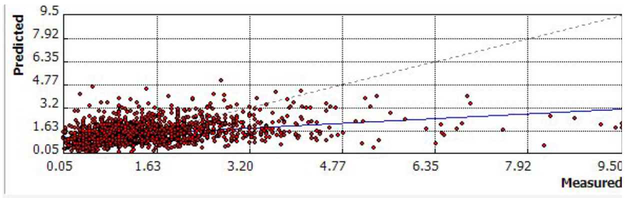


Fig. 8 Measured and Predicted value of Su.

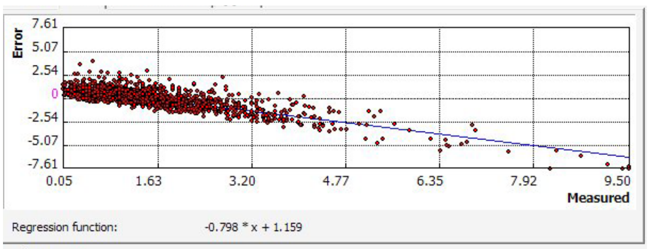


Fig. 9 Prediction errors (Su) at depth 3.50 m

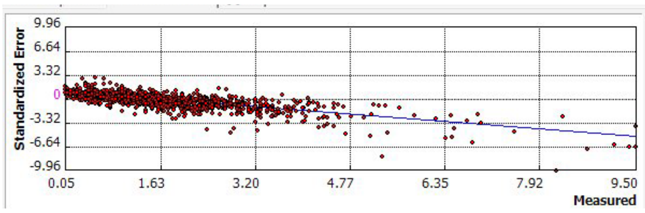


Fig. 10 Standardized error (Su) at depth 3.50 m

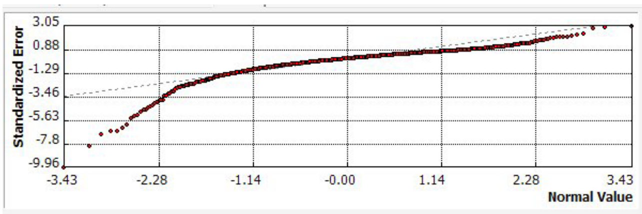
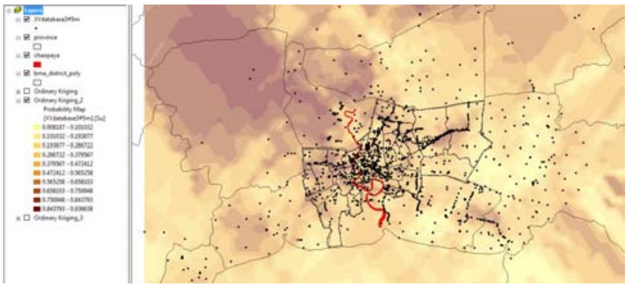
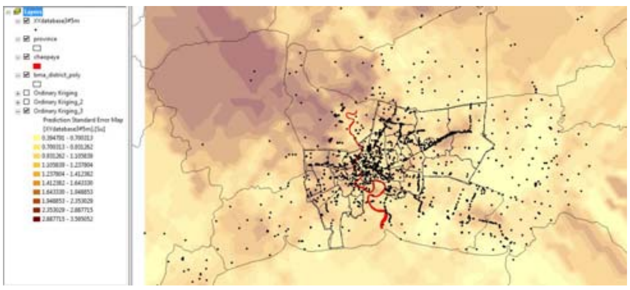


Fig. 11 QQ-plot error (Su) at depth 3.50 m

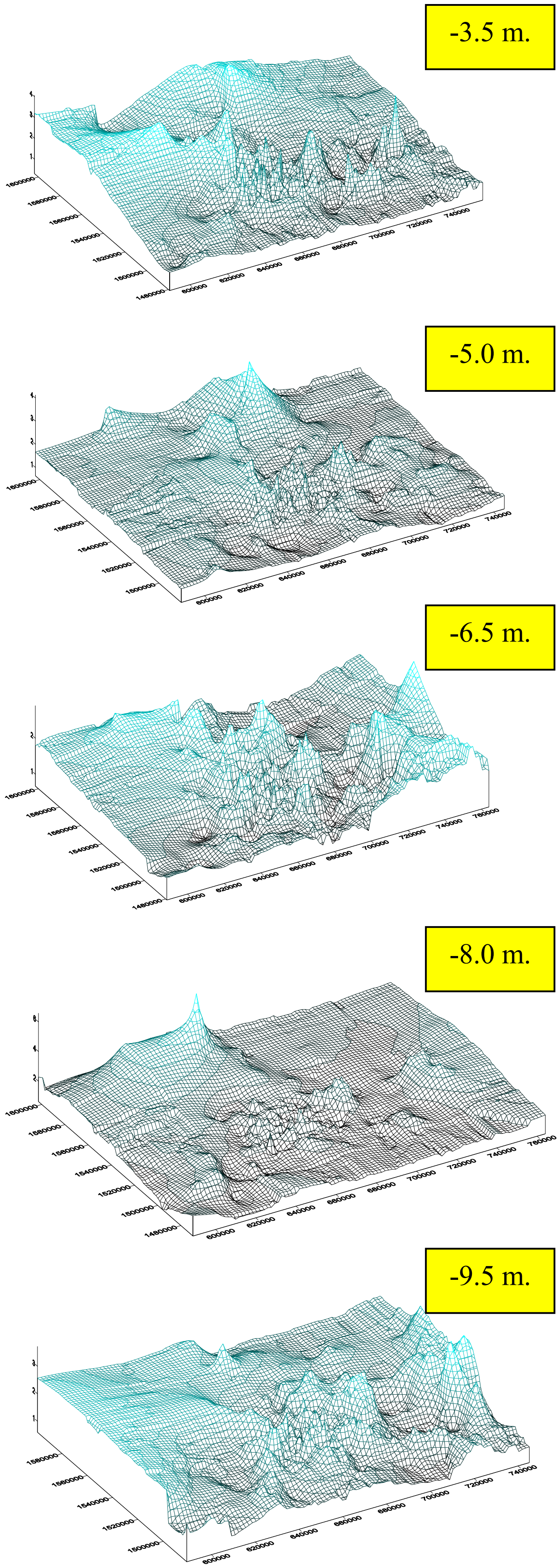
<u>Fig.</u> 12 Prediction Map (Su) at depth 3.50 m



<u>Fig. 13</u> Probability Map (Su) at depth 3.50 m



<u>Fig. 14</u> Prediction Standard Error Map (Su) at depth 3.50 m



<u>Fig. 15</u>   Kriging Surfaces for undrained shear strength at depth (3.5, 5, 6.5, 8, 9.5 m)

REFERENCES

1. Spatial data analysis theory and practice, Robert Haining,Cambridge University Press, 2003
2.Geostatistical Software Library and User's Guide,Clayton V. Deutsch and Andre G. Journel, Oxford University Press, 1992
3.Practical Geostatistics, Isobel Clark, 2001
4.Interoperable and Distributed Processing in Gis, Andrej Vckovski, Taylor & Francis 1798-1998, 1998
5.Geostatistical Glossary and Multilingual Dictionary, Oxord University Press, 1991
6.Geologic and Mine Modelling using Techbase and Lynx, Martin Lloyd Smith, A.A.Balkema/Rotterdam/Brookfield, 1999
7.Numerical Analysis for the Geological Sciences, James R. Carr, Prentice Hall, 1995
8.Reliability and Statistics in Geotechnical Engineering, Gregory B. Baecher and John T. Christian, 2003
9.Scaling Methods in Soil Physics, Yakov Pachepsky & David E. Radcliffe & H. Magdi Selim, CRC Press, 2003