Statistical Characterization of Cone Penetration Test Variability for Ibis Hotel Soil

Anto Budi Listyawan and Renaningsih *

Abstract: Geotechnical variability is a complex attribute that results from many disparate sources of uncertainties. It is strongly dependent on the properties of the soil beneath and adjacent to the structure of interest. Probabilistic models began more realistic design compare to the old deterministic design as it can describe and take account of soil variability. Although the deterministic approach is widely used, it is well known, that almost all natural soils are spatially variable in their properties and rarely homogenous. This paper focuses on the preliminary analysis to prepare the probabilistic analysis of Pile Foundation design by characterizing the tip resistance dan sleeve friction for 6 CPTs data taken from Ibis Hotel Surakarta. It involves an extensive analysis to perform the best-fit distribution of pointwise variability of tip resistance and sleeve friction using computer program written in MATLAB and FORTRAN. Finally, the point statistics (i.e. mean, standard deviation, and coefficient of variation) across the site were derived together with the interpretation of the possibility of the existence of different materials. The results show that, there is no objection to the hypothesis of normality in the chi-square analysis, although the best fit distribution for each profile or 6 profiles which collected at once are varying (i.e.normal, log-normal, gamma, beta. When all tip resistance data are collected at once, the mean and standard deviation is 42.02 kg/cm2 and 40 kg/cm2 respectively. The mean and standard deviation of all sleeve friction data is 1.02 kg/cm2 and 0.8 kg/cm2 respectively. The coefficient of variation of tip resistance and sleeve friction tend to be skewed as its value is high (i.e. 0.95 and 0.78 respectivel).

Keywords: chi-square, deterministic, probabilistic, tip resistance, variability.

I. Introduction

Traditionally factor of safety (F) is attained by the assumption that a single value factor of safety can represent a homogeneous soil property. However, even in a so-called “uniform” deposit, spatial variation does exist and therefore

* Lecturer, Civil Engineering Dept. Muhammadiyah University of Surakarta

assigning material properties is one of the main sources of uncertainty. The combination of variability in material properties and inadequate number of observations leads to a very high degree of uncertainty in the values used for design. In addition, many codes in Europe and USA, has considered the reality of spatial variation in soil properties, and statistical methods have been proposed as a means of accounting for the effect of spatial variability. The reality of complex, varying geological formations lead to the representation of soil properties as a function of a statistical distribution, e.g. normal, lognormal or exponential. Traditional probabilistic analysis of structures are often based on pointwise variability, which often can be approximated by a normal distribution with two statistical parameters, namely, the mean (μ) and standard deviation (σ).

In 1950s and 1960an, Freudenthal published some papers on risk and reliability concept and followed by many researchers in Civil Engineering (Baecher et al. 2003). This theory has been developed on the geotechnical construction in 1970s. In that time, many offshore projects, mining, slope stability analysis and some giant geotechnical projects applied the reliability concept. Furthermore, the collapse of some giant projects (i.e. Teton Dam and many artificial islands in Canada) have been brought to a new paradigm of geotechnical design from the old method to the risk and reliability method. Lumb (1966) examined the using of the normal distribution for numerous soil properties in terms of consistency, compressibility and soil strength variations. He concluded that most of the soil properties were normally distributed except the coefficient of consolidation. Lumb (1970) subsequently investigated the use of beta and normal distribution for the soil strength by chi-square method and it had been shown that the soil strength distribution can be approximated quite closely by a beta distribution, but in some cases showed that the central portions can also be approximated almost as closely by the more familiar normal distribution. Moreover, Lee *et al*. (1983) noted that the normal or log-normal distributions are adequate for the large majority of geotechnical data, unless the extreme values of a parameter are of critical interest.

A. Statistics

Statistics deals with the collection and the analysis of data to solve real problems. What makes the discipline of statistics useful and unique is that it is concerned with the process of getting data and understanding problems in the presence of variability (Walpole *et al*. 1998). Data sets are often large, so that data must be organized, summarized, and displayed before any interpretation can be attempted. Graphical displays, such as plots and diagrams, are especially useful to uncover unknown features in the data (Hogg & Ledolter, 1987). For large data sets it is better to construct a frequency distribution and display the results in form of histogram. Frequently, however, scientists also want to summarize the information numerically and obtain a few statistics that characterized the data set, in particular, the location and variability measurements. Location measures in a data set are designed to provide the analysis of some quantitative measure of where the data centre is in a sample. One obvious and very useful measure is the sample mean. The mean is simply a numerical average. Suppose that the data in a sample are x_1, $x_2, x_3, \ldots, x_n$. The mean is

$$\bar{x} = \mu = \sum_{i=1}^{n} \frac{x_i}{n} = \frac{x_1 + x_2 + \ldots + x_n}{n} \tag{1}$$

The simple measure of variability is the range of the data (x_{max}-x_{min}) which tells about the extent of the variability of such data. The sample measure of spread that is used most often is the sample standard deviation,

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

Since μ and σ are expressed in the same units, the coefficient of variation is independent on the unit measurement and is expressed by

$$V = \frac{\sigma}{\mu} \tag{3}$$

B. Probability Density Function

The concept of probability distribution always deals with discrete and continuous distribution. A discrete distribution assumes each of its values with a certain probability that often represented in the form of bar chart or probability histogram. A continuous distribution has a probability of zero of assuming exactly any of its values and consequently, its probability cannot be given in tabular form. It can be stated as a formula that would necessarily be a function of numerical values of a continuous variable X and as such will be represented by the functional notation $f(x)$. In dealing with continuous variables, $f(x)$ is usually called the probability density function. The most important continuous probability distribution in the entire field of statistics is the normal distribution. Its graph, called the normal curve, is the bell shaped curve. The normal distribution is often referred to as the Gaussian distribution. The density function of a normal variable X, with mean μ and standard deviation σ, is

$$n(x;\mu;\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(1/2)[x-\mu)/\sigma]^2} \tag{4}$$

Although the normal distribution can be used to solve many problems in engineering and science, there are still numerous situations that require different types of density functions, such as the gamma distribution. Its name was derived from the well-known gamma function (Walpole *et al.* 1998). The probability density function of the gamma distribution can be expressed in terms of the gamma function:

$$f(x;k;\theta) = x^{k-1} \frac{e^{-x/\theta}}{\theta^k \Gamma(k)} \text{ for } x > 0 \tag{5}$$

Alternatively, the gamma distribution can be parameterized in terms of a shape parameter α = *k* and an inverse scale parameter $\beta = 1 / \theta$, called a rate parameter:

$$g(x;\alpha;\beta) = x^{\alpha-1} \frac{\beta^\alpha e^{-\beta x}}{\Gamma(\alpha)} \text{ for } x > 0 \tag{6}$$

In probability theory and statistics, the beta distribution is a continuous probability distribution with the probability density function (pdf) defined on the interval [0,1]:

$$f(x;\alpha;\beta) = x^{\alpha-1}(1-x)^{\beta-1} \frac{1}{B(\alpha,\beta)} \tag{7}$$

In which α and β are parameters that must be greater than zero and *B* is the beta function (Farrington *et al.* 1999).

The Log-normal distribution is used for wide variety of applications. The distribution applies in cases where a natural log transformation results in normal distribution (Limpert *et al.* 2001). The log-normal distribution has probability density function (pdf):

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-[\ln(x)-\mu]^2/(2\sigma^2)} \text{ for } x > 0 \tag{8}$$

C. Chi Square Goodness-of-fit Test

One of the testing methods of statistical hypotheses of a data set that has a specified theoretical distribution is chi square test. The test is based on how good a fit between the frequency of occurrence of observations in an observed sample and the expected frequencies obtained from the hypothesized distribution (Walpole *et al.* 1998). A goodness-of-fit test between observed and expected frequencies is based on the quantity:

$$\chi^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i} \tag{9}$$

χ^2 is a value of a random variable in which sampling distribution is approximated very closely by the chi-square distribution with $v = k - 1$ degrees of freedom. The symbols o_i and e_i represent the observed and expected frequencies, respectively, for the ith cell. If the observed frequencies are close to the corresponding expected frequencies, the χ^2-value will be small, indicating a good fit. If the observed frequencies differ considerably from the expected frequencies, the χ^2-value will be large and the fit is poor.

D. Available Raw CPT Data

The cone penetration test (CPT) is becoming increasingly more popular as an in-situ test for site investigation and geotechnical design. As a logging tool this technique is unequalled with respect to the delineation of stratigraphy and the continuous rapid measurement of parameter like bearing and friction (Robertson & Campanella. 1983). Be in opposition to the use of SPT, the CPT has an advantage that provides a continuous data record together with excellent repeatability and accuracy at a relatively low cost. Some experiences around the world, CPT have confirmed a repeatability of tip resistance that is better than $\pm$2% (Jefferies *et al*. 1988a). According to Robertson (1986), the CPT is perfect for investigating loose deposits, since the pushing force is small; hence, this test has become a major asset in evaluating the liquefaction potential of soils. The CPT data from the Ibis Hotel soil were obtained from 6 locations to the depth up to 20 m.

II. Technical Work Preparation

A. Pre-processing of CPT Data

As 6 CPT data has been taken, isolated unrepresentative values of CPT data are likely appearing as any other type of field test, which might cause misinterpretation of the results. These unrepresentative data exist as an effect of the soil disturbance by the cone penetration, or some errors in measuring equipment, or the material type variation. Data filtering might be carried out to overcome these uncertainties. Filtering is a process applied to a region where data do not represent the true variation of the profile; such data are therefore eliminated.

B. Post-processing of CPT Data

A very time consuming and complicated process is needed to analyze the CPT data statistically. In this present investigation, the authors use a ready Fortran 90 computer program written by Wong (2004), which has been modified by Gitman (2006) by adding the Matlab computer program to represent the probability density function and performing the best fit distribution for each profile. The authors have also developed some Fortran 90 codes to tackle the statistical parameters which are varying with depth. These programs make the possibility of investigating several CPT profiles within a reasonably short period of time. This chapter highlights the main steps in examining the field data, and Figure 1 gives a summary flowchart for all steps.

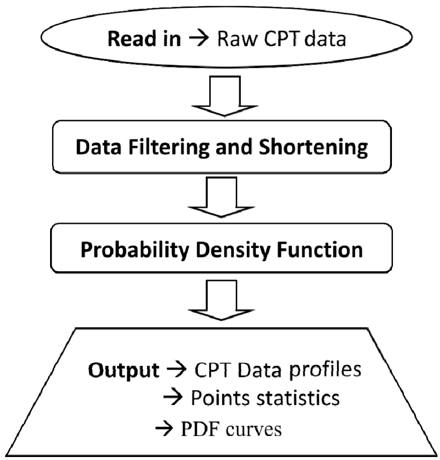


Fig. 1. Summary flowchart of CPT evaluation process.

In this current study, the four types of distribution (i.e. normal, gamma, beta and log-normal) have been analyzed to get the best fit distribution for both tip resistance and state parameter. The probability density function is defined by subdividing the range of data values into classes of equal width as following formula:

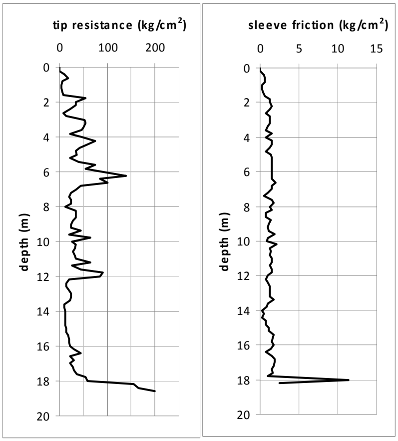
$$(PDF)_i = \left(\frac{f_i}{\sum_{i=1}^{n_c} f_i} \right) \left(\frac{1}{bw} \right) \tag{10}$$

Where f_i is the sampling class frequency for the class reference number i, n_c is the total number of classes and bw is the class width. In the automated program, 6 classes have been used for all profiles. Thus, each profile has a different class width as the range of data is differ from one profile to another. By considering the four types of distribution, the best fit distribution has been calculated using the chi square goodness-fit method. The best fit distribution is the distribution that has the lowest value of chi square. For every CPT profile, the point statistics (i.e. the mean, μ, and standard deviation, σ) are calculated.

III. Result and Discussion

A. General Representation of CPT Evaluation Data

6 CPT data of Ibis Hotel Soil have been evaluated in the current investigation. Typical sheets representation of the CPT data can be seen in Figure 2. For each profile, a typical sheet comprises: (a) the raw CPT profiles; (b) a summary of point statistics for tip resistance and sleeve friction; (c) the best fit distribution for tip resistance and sleeve friction. In general, the tip resistance of Ibis Hotel Soil starts with the lower value up to 3m depth in the order of 14-25 kg/cm^2 and then increases significantly in the depth of 4-6m.

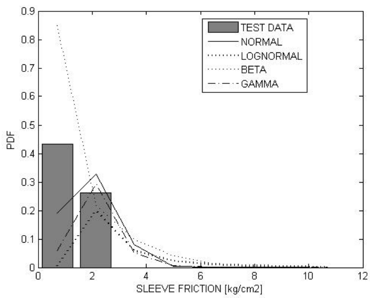


a) CPT profile

Point Statistics	*qc*
μ (kg/cm^2)	37.91
σ (kg/cm^2)	30
v	0.91

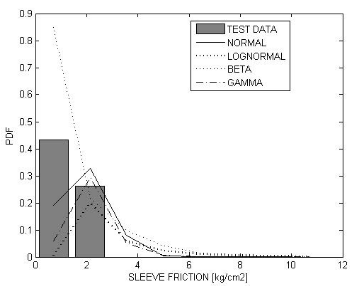
Point Statistics	*fs*
μ (kg/cm^2)	1.27
σ (kg/cm^2)	1.0
v	0.91

b) Point statistics



Normal = 0
Lognormal = 0
β = 1
Γ = 0

Total = beta
Mean = 37.9149
Std = 3e+001



Normal = 0
Lognormal = 0
β = 1
Γ = 0

Total = beta
Mean = 1.2691
Std = 1

c) Best fit distribution

Fig. 2. Point 2 sheet CPT evaluation.

B. Probability Density Function

In general, the probability distribution function of tip resistance and sleeve friction shows a different type of distribution in terms of the number of peak arises. The mono-modal, bimodal or even multi-modal distribution can be identified across the profiles as evaluated in a natural deposit (Listyawan, 2006). Indeed, it is observed across the profiles in Ibis Hotel Soil that the mono-modal, bimodal or even multi-modal distribution of tip resistance and sleeve friction can not be identified. It is probably due to the lack number of data in each profile arisen.

C. Best fit distribution

A different best fit distribution arises for each profile, as tabulated in Table I and Table II.

Table I. Best Fit Distribution For TipResistance

Profile number	Best fit distribution
1	Log-normal
2	Beta
3	Gamma
4	Gamma
5	Log-normal
6	Gamma

Table II. Best Fit Distribution For Sleeve Friction

Profile number	Best fit distribution
1	Normal
2	Beta
3	Beta
4	Normal
5	Gamma
6	Normal

For tip resistance, one profile exhibits beta best fit distribution. The number of profile having best fir distribution of log-normal and gamma are 2 and 3 profiles respectively. The log-normal result seems to be consistent with some previous studies (i.e. Fenton 1999 and Lumb 1966) which stated that there were some supporting evidence for the log-normal distribution for a number of strictly positive soil properties (e.g. strength, elastic modulus, permeability). Furthermore, Fenton & Vanmarcke (2003) also assumed that tip resistance of NGES data follows the log-normal distribution. For sleeve friction, 3 profiles exhibit normal best fit distribution. Beta best fit distribution arises twice (i.e. profile No. 2 and 3). Gamma best fit distribution shows only once in the profile No. 5. There is no evidence the log-normal distribution can be represented for sleeve friction. It can be explained from Figure 3 and 4, for all data (i.e. 6 profiles were collected at once), the best distribution for tip resistance and sleeve friction are gamma and beta distribution respectively.

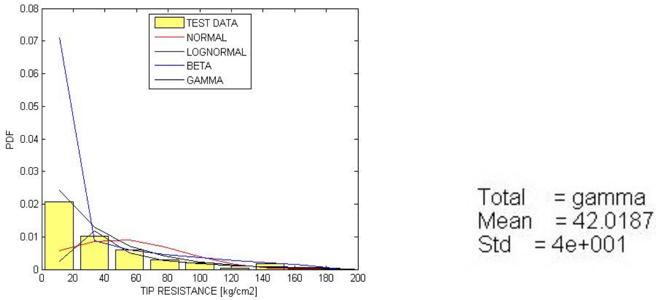


Fig. 3. Distribution of tip resistance for all profiles

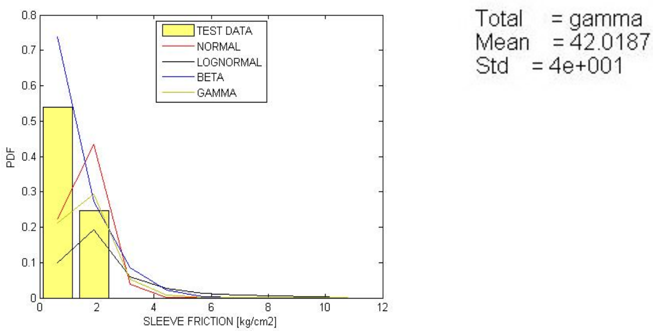


Fig. 4. Distribution of sleeve friction for all profiles.

The chi-square values of the normal, log-normal, beta and gamma distribution for both tip resistance and sleeve friction are tabulated in Table III. It can be observed that the χ^2 value of normal distribution for both tip resistance and sleeve friction are 0.050 and 0.563 respectively, whereas less than $\chi^2_{\upsilon,\alpha} = 7.815$ (for the significance level of 0.95). It means that there is no objection to the hypothesis of normality.

Table III. Chi-square Value of Tip Resistance and Sleeve Friction (All Profiles)

Distribution	χ^2 - value	
	tip resistance	sleeve friction
Normal	0.050	0.563
Log-normal	0.130	2.110
Beta	0.040	0.167 **(Best)**
Gamma	0.003 **(Best)**	0.577

D. Points Statistics Determination

Using a fully automated computer algorithm, the statistics for tip resistance and sleeve friction have been derived for all CPT profiles. The average values of tip resistance for both tip resistance and sleeve friction statistics are summarized in Tables IV. The average mean, standard deviation, and coefficient of variation (V_{qc}) of tip resistance are 42.02 kg/cm^2, 40 kg/cm^2 and 0.95, respectively. The average mean, standard deviation, and coefficient of variation (V_{fs}) of sleeve friction are 1.02 kg/cm^2, 0.8 kg/cm^2 and 0.78, respectively. The high value of coefficient of variation gives more evident that the distribution of tip resistance tends to be skewed. The coefficient of variation of tip resistance (V_{qc} = 0.95) is not consistent with Phoon & Kulhawy (1999) who proposed the range of V_{qc} for tip resistance being equal to 0.01-0.81.

Table IV. Average statistics for all profiles

Property	Tip Resistance	Sleeve Friction
μ_{qc} (kg/cm^2)	42.02	1.02
σ_{qc} (kg/cm^2)	40	0.8
V_{qc}	0.95	0.78

IV. Summary and Conclusion

Reality estimates of the variability of Ibis Hotel Soil are needed for the next development and application of reliability-based design. An extensive analysis of statistical characterization of 6 CPTs data of Ibis Hotel Soil was conducted in this present study. The different materials can not be summarized as the tip resistance and sleeve friction distribution did not clearly exhibit predominantly bimodal or even multi modal distributions. The chi-square analysis shows that the normal distribution was still satisfactory for describing tip resistance and sleeve friction although it can reasonably represented by log-normal distribution since the values are always strictly positive and a relatively high degree of variability (V_{qc}) which has been identified

V. Acknowledgment

The author expresses the sincere thanks to Indra Maiya, R.S.S for his work on the original version of this document. Gratitude is extended to all Geotechnical Engineering Lecturers in the Civil Engineering Department of Muhammadiyah University of Surakarta for their strongly support and assistance.

VI. References

Baecher, G.B. and Christian, J.T. (2003). *Reliability and Stratistics in peotechnical Engineering*, 1st ed. John Wiley & Sons Ltd, England.

Farrington, P.A., Nembhard, H.B., Sturrock, D.T. and Evans, G.W. (1999). Defining a beta distribution function for construction simulation. *Proceedings of the 1999 Winter Simulation Conference,* 1010-1015.

Fenton, G.A. (1999). Random field modeling of CPT data. ASCE, Journal of Geotechnical and Geoenvironmental Engineerng, 125, 486-498.

Fenton, G.A. and Vanmarcke, E.H. (2003). Random field characterization of NGES data. Geotechnical Special Publications, 121, 61-78.

Hogg, V. and Ledolter, J. (1987). *Applied Statistics for Engineers and Physical Scientists*, 2nd ed. New York

Jefferies, M.G., Rogers, B.T., Griffin, K.M. and Been, K. (1988a). Characterization of sandfills with the cone penetration test. *ICE Proceedings of the Geotechnology Conference*, Birmingham, 199-202.

Lee, I.K., White, W. and Ingles, O.G. (1983). Geotechnical Engineering. Pitman publishing, Ltd, England.

Limpert, E., Stahel, W.A. and Abbt, M. (2001). Log-normal Distributions across the Sciences: Keys and Clues. Bioscience, 51, 341-352.

Lumb, P. (1966). The variability of natural soils. Canadian Geotechnical Journal, 3, 74-97.

Lumb, P. (1970). Safety factors and the probability distribution of soil strength. Canadian Geotechnical Journal, 7, 225-242.

Onisiphorou, C. (2000). Stochastic analysis of saturated soils using finite elements. *PhD. Thesis*, University of Manchester, UK.

Phoon, K.K. and Kulhawy, F.H. (1999). Characterization of geotechnical variability. Canadian Geotechnical Journal, 36, 612-624.

Robertson, P.K. (1986). In-situ testing and its application to foundation engineering. *Canadian Geotechnical Journal*, 23, 573-594.

Robertson, P.K. and Campanella, R.G. (1983). Interpretation of cone penetration tests: Part I: Sand. *Canadian Geotechnical Journal*, 20, 718-733.

Tappin, R.G.R., Duivendijk, J.V. and Haque, M. (1998). The design and construction of Jamuna bridge, Bangladesh. *Proceeding of Institution of Civil Engineering*, 126, 150-162.

Walpole, R.E., Myers, R.H. and Myers, S.L. (1998). *Probability and Statistics for Engineers and Scientists*, 6th ed. New Jersey

Wong, S.Y. (2004). Stochastic characterization and reliability of saturated soils. *PhD. Thesis*, University of Manchester, UK.

Yoshimine, M., Robertson, P.K. and Wride, C.E. (1999). Undrained shear strength of clean sands to trigger flow liquefaction. *Canadian Geotechnical Journal*, 36, 891-906.

VII. Biographies

Anto Budi Listyawan was born in Klaten Indonesia, on June 29th, 1977. He graduated from the Gadjah Mada University, and finish his Master degree at The University of Manchester,UK.

He has been employed as lecturer in Civil Engineering Department, Muhammadiyah University of Surakarta for 9 years. He joined the group of Geotechnical Engineering Researcher in this Institution. His special field of interest is evaluating the soil variability and reliability.

Renaningsih was born in Semarang Indonesia, on September 24th, 1963. She graduated from the Diponegoro University, and finish her Master degree at Gadjah Mada University.

She has been employed as lecturer in Civil Engineering Department, Muhammadiyah University of Surakarta for 15 years. She joined the group of Geotechnical Engineering Researcher in this Institution. Her special field of interest is evaluating soil stabilization.