

Investigation on Probability Density Function for Cone Penetration Test Data

*Maryam-Sadat Seyedein¹⁾, Reza Jamshidi Chenari²⁾, Abolfazl Eslami³⁾

^{1), 2)} *The University of Guilan, Rasht, Iran*

³⁾ *Amirkabir University of Technology, Tehran, Iran*

¹⁾ mseyedein@gmail.com

ABSTRACT

The unique nature of geo-materials that their properties vary from one location to another, even within a few distances, makes it distinct from other engineering materials. This variation of soil properties is termed spatial variability and cone penetration test is one of most useful in-situ site investigation techniques which can be used to determine soil spatial variability.

While using cone penetration test to investigate the soil properties, statistical analyses of cone tip resistance data become necessary and it is important to know the maximum likelihood estimation of the q_c measurements representing a random field. In this regard it is required to assume a distribution for q_c data measurements. Different types of distribution functions for q_c data are found in literature. Some researchers use log-normal distribution while some others prefer to raise q_c data to some power between 0 and 1 and some others use normal distribution traditionally. Each of these procedures has their own benefits and drawbacks.

In this study, results of cone penetration tests from 102 cases in different sites are collected and most representative probability distribution function for original and residual cone tip resistance data are sought. Considering different probability distribution function, the relevant probability distribution function (PDF) for both original and residual cone tip resistance data are evaluated and appropriate distribution is suggested.

Key words: Spatial variability, Cone tip resistance, Statistical analysis.

1. INTRODUCTION

Uncertainty is a central aspect of geotechnical engineering. The sources of uncertainty for the estimation of soil properties include inherent soil variability, measurement error, and uncertainty in models to infer soil properties from in situ measurements (DeGroot & Baecher, 1993, Phoon & Kulhawy, 1999a). Furthermore, when the response or the failure mode of a geotechnical structure is of interest,

¹ M.S Student

² Assistant Professor

³ Associate Professor

additional uncertainties exist due to the mechanics of the problem and to the calculation models employed (Rackwitz, 2000).

Uncertainties due to inherent soil variability and to measurement error have been traditionally addressed using the theory of random fields (Vanmarcke, 1983). The estimation of ranges of variability of random fields of soil properties at different scales has received wide attention (Phoon & Kulhawy, 1999a, Cho, et al., 2004), and several correlation models for characterization of spatial variability of soil properties have been proposed as well (Fenton, 1999a, Rackwitz, 2000, Elkateb, et al., 2003).

In general, inference occurs whenever one estimates properties at any unobserved spatial location. Here, the word inference will be taken to mean the estimation of stochastic model parameters which allow one to make probabilistic statements about an entire site for which data is limited or not available. This may be necessary, for example, in preliminary designs, in designs involving a future state, or in designs where a large site is to be characterized on the basis of a small test region (Fenton, 1999a).

Current study focuses on the evaluation of the probability distribution nature of residual CPT profiles in order to identify the best representative Probability Density Function (PDF) for CPT data. Cone penetration profiles acquired from different sources worldwide are adopted to fit different PDFs to the residual CPT profiles and seek for the most representative distribution.

2. PROBABILITY DENSITY FUNCTION

In order to carry out maximum likelihood estimation of the random field of CPT measurements, a distribution for cone resistance must be assumed. Since q_c is bounded below by zero and has no arbitrarily set upper limit, q_c may be assumed to be lognormal distributed. There is some supporting evidence for the lognormal distribution for a number of strictly positive soil properties (strength, elastic modulus, permeability, for example 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, is that it leads to a simple model since 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). In that 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.

The lognormal distribution has been widely employed to model variations of the Young's modulus of soils, see, e.g. (Paice et al., 1996, Nour et al., 2002, Fenton et al., 2002, Fenton et al., 2005). Arguments in favor of the lognormal distribution are that it has zero probability of generating negative values of Young's modulus and that it is mathematically convenient.

Other transformations are also available. In particular, rising q_c to some power between 0 and 1 is quite common. The disadvantage to these transformations is that they do not nicely lead to a non-negative 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 $q_c = X^{\frac{1}{r}}$ is only defined for certain r and, of these, some are imaginary.

However, if the probability of X being negative under the normal assumption is negligible (having small coefficient of variation), these transformations can give reasonable results (Fenton, 1999b).

Fenton (1999b) conducted research to investigate the appropriate power on CPT data collected by Norwegian Geotechnical Institute (NGI) at the site of a new airport just north of Oslo, Norway. He concluded that a function with the power of 0.74 provides the best fit for the investigated data.

Fenton (2002) also studied the settlement under a footing founded on a spatially random elastic modulus field of finite depth overlying bedrock by assuming that E is distributed lognormal. He concluded that the estimation of probabilities associated with settlement involves little more than referring to a standard normal distribution table. He also emphasized that the differential settlement is a more complicated distribution than that of settlement itself.

However, the use of the lognormal distribution to model the random field of Young's modulus in finite element computations of foundation settlement has been generally accepted (but not always) without further investigating whether it is the most adequate type of distribution to model the variability of soil's properties in each particular case, or what would be the effects of employing a different type of distribution. Zhou et al. (1999) studied the effects of uncertainties on the design of prefabricated vertical drains, and they found that the type of statistical distribution employed to model the coefficient of radial consolidation (they considered the lognormal, gamma, and Weibull distributions) has a "considerable influence" on the consolidation results. Similarly, Popescu et al. (2005) illustrated the importance of this aspect of uncertainty characterization in bearing capacity analyses, and they showed that the coefficient of variation and the marginal probability distribution of the soils shear strength "are the two most important parameters in reducing the bearing capacity (in an average sense) and producing substantial differential settlements in heterogeneous soils".

Jimenez and Sitar (2009) investigated the effects of using different types of statistical distribution including lognormal, gamma and beta to characterize the variability of Young's modulus of soils in random finite element analyses of shallow foundation settlement. Their results showed that a random field of Young's modulus with gamma distribution can produce settlements that are (in an average sense) up to more than 600% higher than settlements computed with a random field of Young's modulus with lognormal distribution. Similarly, settlements computed with the beta distribution can be (in an average sense) more than 200% higher than settlements computed with the lognormal distribution.

Jamshidi et al. (2012) statistically analyzed 8 sets of CPT profiles sourced from Urmia Lake located in west Iran, and evaluated different probability density function containing normal, lognormal and power. They proposed power distribution with the exponent averaging at 0.75 as a best distribution function.

However, the current research covers a more comprehensive database by taking advantage of 102 of CPT profiles in order to entrant the most representative probability distribution for CPT profiles.

3. DATABASE

A database of case histories from the results of 102 CPT soundings is utilized herewith. The cases were obtained from 36 sources reporting data from 40 sites in 13 countries. Table 1 summarizes the main case record data as to cases name, site location, sampling interval, total depth and soil profiles.

The majority of the case records are from the United States. The soils at the sites consist of sediments of clay (soft clay, stiff clay, silty clay, sandy clay), silt (clayey silt, sandy silt) and sand (clayey sand, silty sand, gravelly sand). About 80% of the CPT cases included in the data are obtained by electrical cone and 20% by mechanical cone. Most of the CPT measurements are at a vertical spacing of 300 mm or smaller. A typical CPTu profile taken from one of the records is shown in Figure 1 (Eslami and Fellenius, 1997).

Table 1. Case record summary

No.	Case name	Site Location	Sampling Interval (cm)	Total Depth (m)	Soil Profile
1	SEATWCP	U.S.A.	20	18	Sand
2	POLA1CH	U.S.A.	30.5	28.964	Silt,Sand
3	POLA2CH	U.S.A.	30.5	35.685	Silt,Sand
4	FHWACP	U.S.A.	20	12	Sand
5	OKLACOCP	U.S.A.	20	22	Sand,silty clay(shale)
6	OKLASTCP	U.S.A.	20	22	Sand,silty clay(shale)
7	ALABACP	U.S.A.	30	39	Sand,silty clay, sand
8	UHUC1	U.S.A.	15.25	15.25	Clay, sandy clay
9	UHUC11	U.S.A.	15.25	15.25	Clay, sandy clay
10	UHUT1	U.S.A.	15.25	15.25	Clay, sandy clay
11	UHUT11	U.S.A.	15.25	15.25	Clay, sandy clay
12	PRLCP	Puerto Rico	5	35.75	Peat,sand, clay
13	PRSCP	Puerto Rico	5	33.05	Peat,sand, soft clay
14	GIT1CP	U.S.A.	20	21	Fill, silty sand
15	UFL22CP	U.S.A.	100	19.506	Sand
16	UFL52CP	U.S.A.	61	16.463	Sand
17	UFL53CP	U.S.A.	61	25	Sand, silt
18	LSUA1CP	U.S.A.	15.25	12.2	Sand, clay
19	LSUB12CP	U.S.A.	30.5	42.395	Silt, silty clay, sand, silt
20	LSUH1	U.S.A.	30.5	35.38	Clay, sand, clay
21	LSUN11CP	U.S.A.	30.5	45.75	Silty clay, silty sand
22	LSUN15CP	U.S.A.	30.5	45.75	Silty clay, silty sand
23	LSUN28CP	U.S.A.	30.5	43.615	Clay, silty sand, clay
24	LSUN215P	U.S.A.	30.5	43.615	Clay, silty sand, clay
25	LSUN216P	U.S.A.	30.5	45.835	Clay, silty sand, clay
26	LSUR24CP	U.S.A.	30.5	24.4	Fill, sandy clay
27	LSUR30CP	U.S.A.	30.5	24.4	Fill, sandy clay

No.	Case name	Site Location	Sampling Interval (cm)	Total Depth (m)	Soil Profile
28	QBSACH	Canada	30	9.4	Sensitive Clay
29	NWUHCP	U.S.A.	3,3,9,7,...	18.302	Sand,clay
30	NWUPCP	U.S.A.	3,3,9,7,...	18.302	Sand,clay
31	SPBCP	Brazil	50	22	Silty sand
32	USPB1CP	Brazil	20	16	Clay and silt, silty sand
33	USPB2CP	Brazil	20	16	Clay and silt, silty sand
34	CWDNC	U.K.	12.5	12	Stiff clay, till, silty sand
35	CWDND	U.K.	12.5	12	Stiff clay, till, silty sand
36	CWDNE	U.K.	12.5	12	Stiff clay, till, silty sand
37	CWDNF	U.K.	12.5	12	Stiff clay, till, silty sand
38	CWDNG	U.K.	12.5	12	Stiff clay, till, silty sand
39	CWDNH	U.K.	12.5	12	Stiff clay, till, silty sand
40	CWDNI	U.K.	12.5	12	Stiff clay, till, silty sand
41	CWDNJ	U.K.	12.5	12	Stiff clay, till, silty sand
42	CWDNK	U.K.	12.5	12	Stiff clay, till, silty sand
43	N&SBI42	U.S.A.	20	24.6	Fill, sand, dense sand
44	N&SBI43	U.S.A.	20	24.6	Sand, dense sand
45	N&SBI44	U.S.A.	20	24.6	Sand
46	N&SBI215	U.S.A.	20	32	Sand, silty sand
47	N&SBI316	U.S.A.	20	24.4	Fill, silty sand
48	N&SBI348	U.S.A.	20	23.6	Sand, dense sand
49	N&SJC1CP	U.S.A.	10	12.5	Sand, clay
50	N&SWB1	U.S.A.	10	13.9	Silty sand
51	N&SWB2	U.S.A.	10	13.5	Silty sand
52	MILANOCP	Italy	25	12	Clay, silty sand, clay
53	PT361	Italy	25	46	Silty sand, clay, sand
54	PTSER	Italy	25	40	Sandy silt, dense sand
55	A&N1CP	Australia	20	18	Dense sand, limestone
56	A&N2CP	Australia	20	17	Sand
57	A&N3CP	Australia	20	16	Silt, sand, dense sand
58	NETH2CP	Netherlands	20	17	Fill, clay, silty sand
59	YOG1CP	Yugoslavia	20	14.6	Clay
60	YOG2CP	Yugoslavia	20	17	Clay
61	TWNT4CP	Taiwan	5	46.2	Sand,Clay,Sand
62	TWNT5CP	Taiwan	5	38.3	Sand,Clay,Sand
63	TWNT6CP	Taiwan	5	51.1	Sand,clay,Sand
64	JPNOT1CP	Japan	10	14.8	Sand,clay(soft rock)
65	JPNOT2CP	Japan	10	14.8	Stiff clay(soft rock)
66	JPNOT3CP	Japan	10	14.8	Stiff clay(soft rock)
67	BGHD1CP	Iraq	20	20	Uniform Sand
68	BGHD2CP	Iraq	20	20.8	Uniform Sand
69	ONSYA1	Norway	25	20	Soft clay
70	ONSYB1	Norway	25	20	Soft clay
71	KALO14A	Belgium	20	17	Peat, clay, sand

No.	Case name	Site Location	Sampling Interval (cm)	Total Depth (m)	Soil Profile
72	KALO14B	Belgium	20	17	Peat, clay, sand
73	LSTDA7	Norway	25	24.75	Soft clay
74	LSTDA8	Norway	25	24.75	Soft clay
75	LSTDB2	Norway	25	24.75	Soft clay
76	PNTRA5	U.K.	25	35	Stiff clay, silt, clay
77	PNTRA6	U.K.	25	35	Stiff clay, silt, clay
78	MUMBBCP	U.S.A.	30	15	Fill, till
79	KP1CP	Belgium	20	17	Soft soil, dense sand
80	UBC2CP	Canada	2.5	35.93	Soft clay, sand
81	UBC3CP	Canada	2.5	35.93	Soft clay, sand
82	UBC5CP	Canada	2.5	40	Soft clay, sand, silt
83	UBCACP	Canada	2.5	78.82	Soft clay, sand, silt
84	MP1CP	France	20	18	Soft clay, stiff clay
85	LTN484CP	U.S.A.	20	11.4	Silty sand, clay, sand
86	LTN741CP	U.S.A.	30.5	31.72	Stiff clay, sand
87	LTN742CP	U.S.A.	30.5	29.28	Stiff clay, sand
88	LTN930CP	U.S.A.	30.5	11.895	Silty sand, sand
89	LTN938CP	U.S.A.	30.5	13.725	Silt, gravelly sand
90	L&D12CP	U.S.A.	20	19	Sand
91	L&D13ACP	U.S.A.	20	19	Sand
92	L&D16CP	U.S.A.	20	19	Sand
93	L&D21CP	U.S.A.	20	22	Sand
94	L&D31CP	U.S.A.	20	19	Sand
95	L&D32CP	U.S.A.	20	19	Sand
96	L&D34CP	U.S.A.	20	19	Sand
97	L&D35CP	U.S.A.	20	19	Sand
98	L&D37CP	U.S.A.	20	19	Sand
99	L&D38CP	U.S.A.	20	19	Sand
100	L&D314CP	U.S.A.	20	19	Sand
101	L&D315CP	U.S.A.	20	19	Sand
102	L&D316CP	U.S.A.	20	19	Sand

4. DETERMINISTIC VARIATION

It should be pointed out at this time that one of the reasons for detrending the data is precisely to render the residual process largely spatially independent. This is desirable because virtually all classical statistics are based on the idea that samples are composed of independent and identically distributed observations. Alternatively, when observations are dependent, the distributions of the estimators become very difficult to establish. This is compounded by the fact that the actual dependence structure is unknown. Only a limited number of asymptotic results are available to provide insight into the spatially dependent problem (Beran, 1994). Simulation techniques are proving very useful in this regard (Cressie, 1993).

Soil properties are highly depth dependent and hence CPT parameters such as cone bearing, sleeve friction and pore pressure exhibit significant trends with depth. If the trends are not significant, data can be stationarized by standardizing. In the presence of significant trends, these will have to be removed, prior to evaluating the distribution type. A pre-requisite to this of course is the identification of statistically distinct layers (Wickremesinghe and Campanella, 1993).

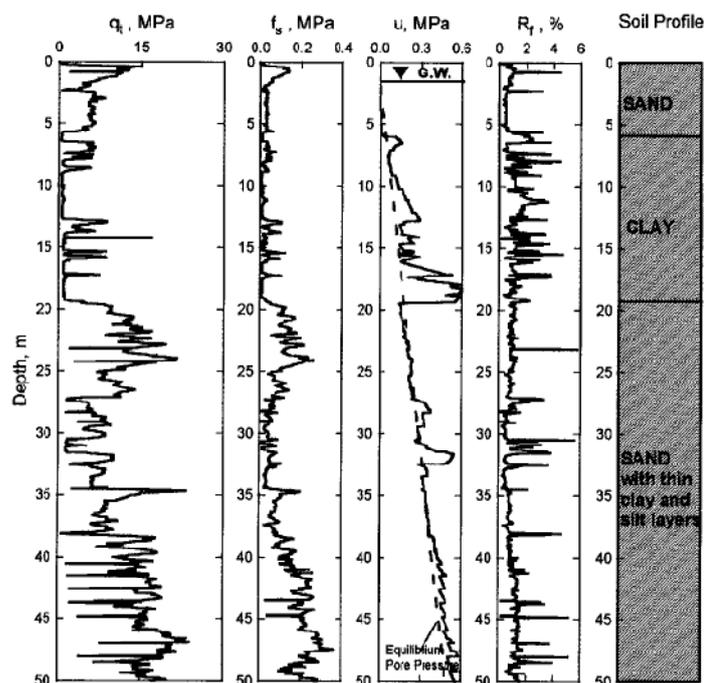


Figure 1. Typical piezocone profile (Eslami & Fellenius, 1997)

As most trends in in-situ measurements derive from overburden vertical stress, investigating the spatial variability of normalized measurements may increase the generality of the analysis, as the influence of site-specific effects such as stress history is possibly reduced (Uzielli, 2004).

Bearing and friction exhibit linear trends while it is usual for the pore pressure profile to possess a curvilinear trend. If a linear trend is used to stationerize a pore pressure profile, the resulting residuals will not be stationary, due to the curvilinear effect of the trend not being removed. This will give rise to an incorrect assessment of statistical parameters. The decision as to what type of trend removal is necessary could be taken by inspecting the residuals. If a linear trend removal results in a stationary set of residual, curvilinear trend analysis is unnecessary. The curvilinear trend would always be more suitable (Wickremesinghe and Campanella, 1993).

5. DISTRIBUTION STRUCTURE

In probability and statistics, a random variable or stochastic variable is a variable whose value is subject to variations due to chance. As opposed to other mathematical variables, a random variable conceptually does not have a single fixed value rather it can take on a set of possible different values, each with an associated probability.

Random variables can be classified as either discrete or as continuous. The mathematical function describing the possible values of a random variable and their associated probabilities is known as a probability distribution. The realizations of a random variable, i.e. the results of randomly choosing values according to the variable's probability distribution are called random varieties.

A probability density function (PDF), or density of a continuous random variable, is a function that describes the relative likelihood for this random variable to take on a given value. The probability for the random variable to fall within a particular region is given by the integral of this variable's density over the region. The probability density function is nonnegative everywhere, and its integral over the entire space is equal to one.

Many probability distributions are so important in theory or applications that they have been given specific names and some of them are: normal distribution, lognormal distribution, power distribution, gamma distribution, beta distribution, chi distribution, Chi-squared distribution, etc.

In order to find a distribution which fits best to the random variables, there are several tests, which have been utilized to seek for the most representative probability distribution scheme for original and residual CPT profiles. Table 2 provides different categories of CPT profiles and the rested distribution schemes utilized for each category.

Table 2. Different distribution schemes fitted to original and residual profiles

Type of Distribution	Original Profile	Residual Profile
Normal Distribution	✓	✓
Lognormal Distribution	✓	×
Power Distribution	✓	✓

5.1 Goodness-of-fit test

The goodness-of-fit of a statistical model describes how well it fits a set of observations. Measures of goodness-of-fit typically summarize the discrepancy between observed values and the values expected under the model in question.

In assessing whether a given distribution is suited to a data-set, the following tests and their underlying measures of fit can be used:

- Kolmogorov-Smirnov test
- Cramer-Von Mises criterion
- Anderson-Darling test
- Chi-Square test

In this research Chi-square distribution is used for testing the goodness-of-fit of a set of CPT data to a specific probability distribution. In this method, the actual frequencies in a class are compared to the frequencies that theoretically would be expected to occur if the data followed the specific probability distribution of interest.

Several steps are required to carry out a Chi-square goodness-of-fit test. First, determine the specific probability distribution to be fitted to the CPT data. Second, hypothesize or estimate the values of each parameter of the selected probability distribution (such as the mean). Next, determine the theoretical probability in each class using the selected probability distribution. Finally, use the Chi-square test statistic, Equation (1), to test whether the selected distribution is a good fit to the data.

$$X_{k-p-1}^2 = \sum_k \frac{(f_o - f_e)^2}{f_e} \tag{1}$$

Where:

f_o = observed frequency

f_e = theoretical or expected frequency

k = number of categories or classes remaining after combining classes

p = number of parameters estimated from the data

5.2 RMS procedure

In mathematics, the root mean square, known as the quadratic mean, is a statistical measure of the magnitude of a varying quantity. It is especially useful when varieties are positive and negative, e.g., sinusoids. RMS is used in various fields, including civil engineering.

It can be calculated for a series of discrete values or for a continuously varying function. The name comes from the fact that it is the square root of the mean of the squares of the values. It is a special case of the generalized mean with the exponent $p = 2$. If p is a non-zero real number, we can define the generalized mean with exponent p (or power mean with exponent p) of the positive real numbers x_1, x_2, \dots, x_n as:

$$M_p(x_1, x_2, \dots, x_n) = \left(\frac{1}{n} \sum_{i=1}^n x_i^p \right)^{1/p} \tag{2}$$

The RMS value of a set of values (or a continuous-time waveform) is the square root of the arithmetic mean (average) of the squares of the original values (or the square of the function that defines the continuous waveform).

In the case of a set of n values $\{x_1, x_2, \dots, x_n\}$, the RMS value is given by:

$$x_{rms} = \sqrt{\frac{1}{n}(x_1^2 + x_2^2 + \dots + x_n^2)} \quad (3)$$

5.3 Correlation scheme

A correlation coefficient measures the strength and direction of a linear association between two variables. It ranges from -1 to +1. The closer the absolute value is to 1, the stronger the relationship. A correlation of zero indicates that there is no linear relationship between the variables. The coefficient can be either negative or positive.

Pearson product-moment correlation coefficient, also known as r , R , or Pearson's r , is a measure of the strength and direction of the linear relationship between two variables that is defined in terms of the (sample) covariance of the variables divided by their (sample) standard deviations.

In statistics, the Pearson product-moment correlation coefficient (sometimes referred to as the PPMCC or PCC, or Pearson's r , and is typically denoted by r) is a measure of the correlation (linear dependence) between two variables X and Y , giving a value between +1 and -1 inclusive. It is widely used in the sciences as a measure of the strength of linear dependence between two variables. It was developed by Karl Pearson from a similar but slightly different idea introduced by Francis Galton in the 1880s.

Pearson's correlation coefficient between two variables is defined as the covariance of the two variables divided by the product of their standard deviations.

Pearson's correlation coefficient when applied to a population is commonly represented by the Greek letter ρ (rho) and may be referred to as the *population correlation coefficient* or the *population Pearson correlation coefficient*. The formula for ρ is:

$$\rho_{X,Y} = \frac{COV(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y} \quad (4)$$

6. RESULTS AND DISCUSSION

Table 3 illustrate, the range of P-values acquired from Chi-square test conducted with different PDF assumptions. Minimum, maximum, average value and the coefficient of variation (COV) of P-values for different assumptions are provided in the table. It is observed that for both original and residual q_c profiles power distribution fits best to the data, in comparison to other assumed distributions. However it is noteworthy that the acquired P-values are around 0.9 in average sense. This means that normal distribution is still valid for both original and residual CPT profiles.

Table 3. Summary of goodness-of-fit test on CPT data

Parameter	Goodness-of-fit Test on						
	Original q_c				Residual q_c		
	Power	P-Value			Power	P-Value	
		Power Dist.	Normal Dist.	Lognormal Dist.		Power Dist.	Normal Dist.
Min	0.01	4.8E-10	3.5E-12	5.7E-14	0.23	4.8E-10	9.1E-13
Max	0.99	2.3E-02	2.1E-02	6.5E-08	0.99	1.6E-01	5.7E-03
Average	0.9	1.2E-03	1.4E-03	9.8E-10	0.88	4.0E-03	2.70E-04
COV %	24.53	300.76	258.62	741.49	23.69	544.57	291.5

Cross confirmed the findings of goodness-of-fit test by conducting RMS technique so as to find the most representative probability distribution function for different kinds of CPT profiles. Table 4 provides the statistical parameters for RMS values obtained for different distribution schemes. It is observed that power distribution assumption renders lower RMS values confirming that power distribution with the power values averaging at 0.86 and 0.91 for original and residual CPT profiles, respectively is expected to be the best representation distribution for CPT profiles.

Table 4. Summary of RMS technique on CPT data

Parameter	RMS Technique on						
	Original q_c				Residual q_c		
	Power	RMS Coefficient			Power	RMS Coefficient	
		Power Dist.	Normal Dist.	Lognormal Dist.		Power Dist.	Normal Dist.
Min	0.43	0.003	0.002	0.034	0.65	0.003	0.003
Max	0.99	0.575	0.569	1.718	0.99	1.348	1.317
Average	0.86	0.048	0.060	0.128	0.91	0.067	0.078
COV %	19.03	187.36	157.9	142.18	11.98	289.89	268.17

Figure 2 illustrates the cumulative distribution function (CDF) for the acquired power values for 102 cases. Superimposed on the same graph is the CDF for fitted normal distribution. At least visually it is confirmed that the power values mainly focus around mean values. This can easily be confirmed by noting to the COV of power values provided in table 4.

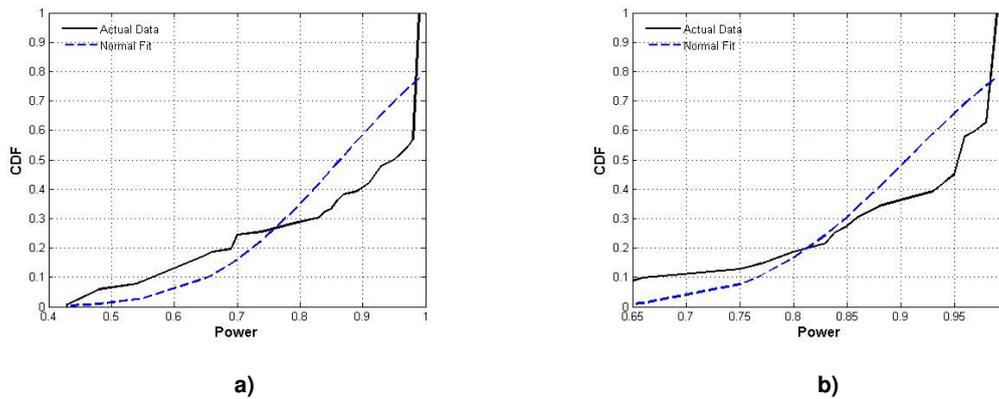


Figure 2. Probability distribution of power values acquired from RMS technique; a) Original q_c ; b) Residual q_c

Correlation coefficient was explained to be an index of probability distribution function regression. Table 5 provides information on correlation coefficient values calculated for different PDF assumptions on original and residual CPT profiles. It is concluded that power distribution is again confirmed to be the most suitable distribution function for different CPT profiles. Power values are 0.81 and 0.9 for original and residual CPT profiles respectively in average sense.

Table 5. Summary of correlation coefficient technique on CPT data

Parameter	Correlation Coefficient Technique on						
	Original q_c				Residual q_c		
	Power	Correlation Coefficient			Power	Correlation Coefficient	
		Power Dist.	Normal Dist.	Lognormal Dist.		Power Dist.	Normal Dist.
Min	0.25	0.376	0.154	-0.487	0.58	0.451	0.451
Max	0.99	0.993	0.993	0.866	0.99	0.996	0.997
Average	0.81	0.894	0.789	0.304	0.9	0.912	0.879
COV %	25.95	14.3	23.95	96.56	12.08	9.49	12.91

A sample PDF fitting is illustrated in figure 3 for case number 46 recorded in USA. All three introduced techniques namely Chi-square test, RMS and correlation tests are adopted herewith with original and residual CPT profiles evaluated separately in left and right hand side respectively. It is visually seen that both power and normal distribution are reasonable for both original and residual q_c profiles.

In fact, this is of the cases, out of all 102 cases given in table 1; a Chi-square goodness-of-fit test yields a P-values of 3.1×10^{-4} , 2.3×10^{-3} and 1.6×10^{-11} , for power, normal and lognormal PDF assumption respecting large P-values, up to 1.0, supports the assumed hypothesis, so that this small value suggests that the data does not follow either of the tested as distributions.

They perhaps correctly indicate the true distribution is not exactly as hypothesized, but say little about the reasonableness of the assumed distribution.

As can be seen from figure 3, the power and also normal distribution certainly appear reasonable.

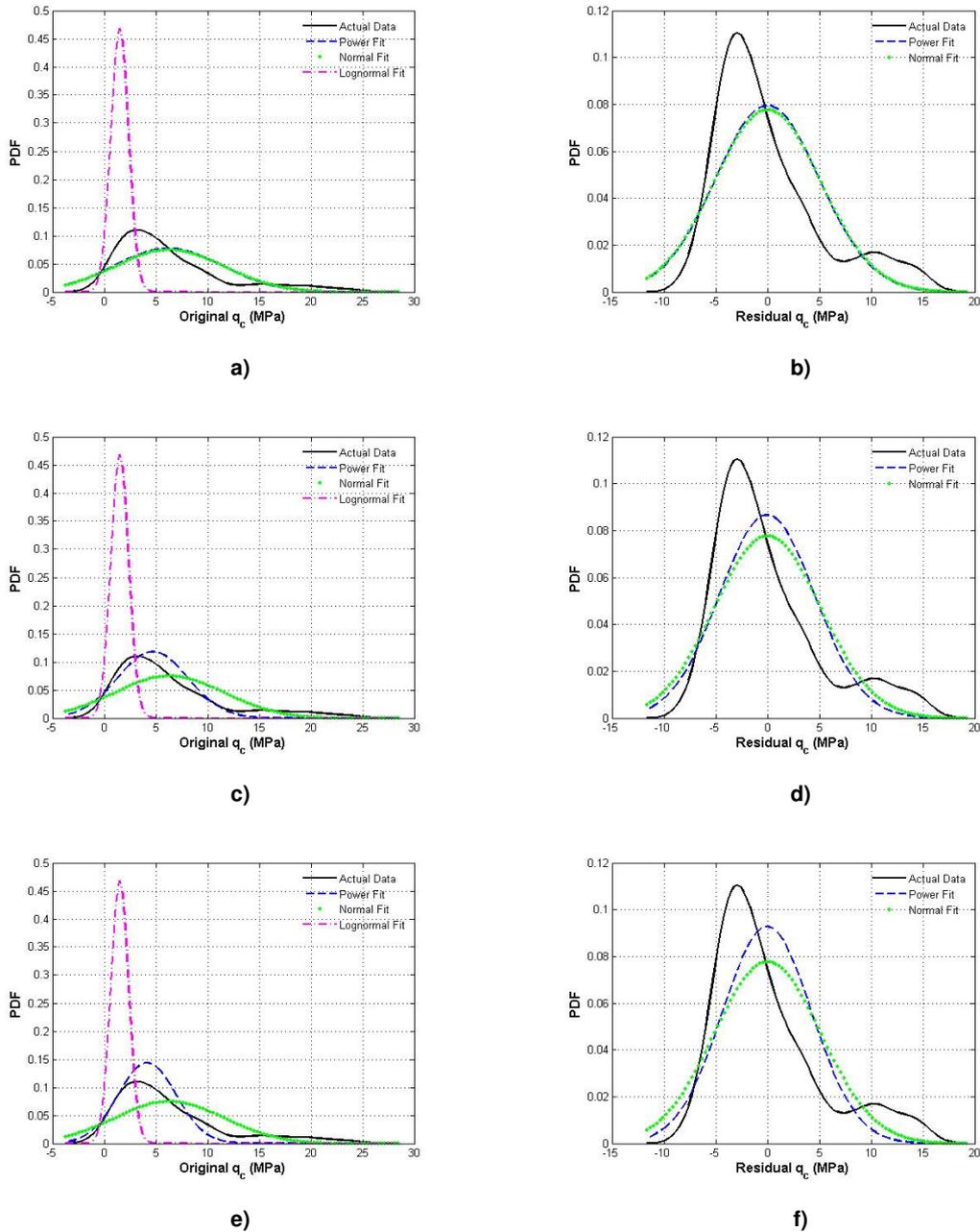


Figure 3. Comparing different PDFs fitted to case # 46; a) GOF on Original q_c ; b) GOF on Residual q_c ; c) RMS on Original q_c ; d) RMS on Residual q_c ; e) CC on Original q_c ; f) CC on Residual q_c

7. CONCLUSION

Soil inherent heterogeneity reflected in the form of deterministic and stochastic variation of in-situ soil properties is believed to play an important role in geotechnical behavior assisting in justification of abnormalities and unexpected behavior in some cases.

CPT data are thought to be one of the most suitable and representative profile of geotechnical parameters variation. This paper tried to study the stochastic properties of original and detrended (residual) CPT profiles by focusing on probability distribution function in specific.

The main points and conclusions of this article are as below:

- 1) Power distribution represents the probability distribution of most CPT profiles fairly accurate.
- 2) The estimated power value for both original and residual CPT profiles was found to be around 0.9. This indeed means that normal distribution which is traditionally used for engineering parameters is still valid for geotechnical parameter due to the closeness of power values to unity. However it should be noted that truncated normal distribution should be used in order to avoid realization of negative values.
- 3) Goodness-of-fit, RMS and correlation techniques were shown to be efficient in probability distribution detection for CPT data.

8. REFERENCES

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

Chenari, R. J., Seyedein, M. S., Faraji, S., & Kenarsari, A. E. (2012). Investigation on inherent variability of soil properties from cone penetration test. *accepted for publication in proceeding of 4th International Conference on Geotechniqueal and Geophysical Site Characterization (ISC'4)*. Porto de Galinhas, Brazil.

Cho, G. C., Lee, J. S., & Santamarina, J. C. (2004). Spatial variability in soils: High resolution assessment with electrical needle probe. *Journal of Geotechniqueal & Geoenvironmental Engineering* (130), 843-850.

Christian, J. T. (2004). Geotechniqueal engineering reliability: How well do we know what we are doing? *Journal of Geotechniqueal & Geoenvironmental Engineering* (130), 985-1003.

Cressie, N. A. (1993). *Statistics for spatial data* (2nd ed.). New York, NY: Wiley & Sons.

DeGroot, D. J., & Baecher, G. B. (1993). Estimating autocovariance of in-situ soil properties. *Journal of Geotechniqueal & Geoenvironmental Engineering* (119), 147-166.

Elkateb, T., Chalaturnyk, R., & Robertson, P. K. (2003). An overview of soil heterogeneity: Quantification and implications on geotechniqueal field problems. *Canadian Geotechniqueal Journal* (40), 1-15.

Eslami, A., & Fellenius, B. H. (1997). Pile capacity by direct CPT and CPTu methods applied to 102 case histories. *Canadian Geotechniqueal Journal* (34(6)), 880-898.

Fenton, G. A. (1999a). Estimation for stochastic soil models. *Journal of Geotechniqueal & Geoenvironmental Engineering* (125(6)), 470-485.

Fenton, G. A. (1999b). Random field modeling of CPT data. *Journal of Geotechniqueal & Geoenvironmental Engineering* (125(6)), 486-498.

Fenton, G. A., & Griffiths, D. V. (2002). Probabilistic foundation settlement on a spatially random soil. *Journal of Geotechniqueal & Geoenvironmental Engineering* (128(5)), 381-390.

Fenton, G. A., & Griffiths, D. V. (2005). Three-dimensional probabilistic foundation settlement. *Journal of Geotechniqueal and Geoenvironmental Engineering* (131), 232-239.

Fenton, G. A., Paice, G. M., & Griffiths, D. V. (1996). Probabilistic analysis of foundation settlement. *Proceeding of the ASCE uncertainty 96 conference, Uncertainty in the geological environment: From theory to practice*, (pp. 651-665). Mudison, Wisconsin.

Hoeksema, R. J., & Kitanidis, P. K. (1985). Analysis of the spatial structure of properties of selected aquifers. *Water Resources Research* (21(4)), 563-572.

Jimenez, R., & Sitar, N. (2009). The importance of distribution types on finite element analyses of foundation settlement. *Computers and Geotechniques* (36), 474-483.

Lumb, P. (1966). The variability of natural soils. *Canadian Geotechniqueal Journal* (3(2)), 74-97.

Noure, A., Slimani, A., & Laouami, N. (2002). Foundation settlement statistics via finite element analysis. *Computers and Geotechniques Journal* (29), 641-672.

Paice, G. M., Griffiths, D. V., & Fenton, G. A. (1996). Finite element modeling of settlements on spatially random soil. *Journal of Geotechniqueal and Geoenvironmental Engineering* (122(9)), 777-779.

Phoon, K. K., & Kulhawy, F. H. (1999a). Characterization of geotechniqueal variability. *Canadian Geotechniqueal Journal* (36), 612-624.

Phoon, K. K., & Kulhawy, F. H. (1999b). Evaluation of geotechniqueal property variability. *Canadian Geotechniqueal Journal* (36), 625-639.

Popescu, R., Deodatis, G., & Nobahar, A. (2005). Effects of random heterogeneity of soil properties on bearing capacity. *Probabilistic Engineering Mechanics* (20), 324-341.

Rackwitz, R. (2000). Reviewing probabilistic soils modeling. *Computers and Geotechniques Journal* (26(3-4)), 199-223.

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 Resources Research* (22(13)), 2069-2083.

Uzielli, M. (2004). *Variability of stress-normalized CPT measurements and application to seismic liquefaction initiation assessment*. PhD dissertation, University of Florence, Department of civil engineering, Florence, Italy.

Uzielli, M., Vannucchi, G., & Phoon, K. K. (2005). Investigation of correlation structures and weak stationarity using the CPT soil behavior classification index. *Proceedings of the 9th International Conference on Structural Safety and Reliability*. Rome.

Vanmarcke, E. (1983). *Random fields: Analysis and synthesis*. Cambridge, Massachusetts: MIT Press.

Wickremesinghe, D., & Campanella, R. G. (1993). Scale of fluctuation as a descriptor of soil variability. *Probabilistic Methods in Geotechniqueal Engineering* , 233-239.

Zhou, W., Hong, H. P., & Shang, J. Q. (1999). Probabilistic design method of prefabricated vertical drains for soil improvement. *Journal of Geotechniqueal and Geoenvironmental Engineering* (125), 659-664.