Statistical to Fuzzy Approach Toward CPT Soil Classification

By Zhongjie Zhang,[1] P.E., Member, ASCE, and Mehmet T. Tumay,[2] P.E., Fellow, ASCE

Abstract: A soil engineering classification derived from the cone penetration test (CPT) involves the uncertainty of correlation between soil composition and soil mechanical behavior. This uncertainty results in overlaps of different soil types in currently used CPT soil classification systems. Accordingly, two statistical soil classification criteria, region estimation and point estimation, are suggested to address this problem. Further, a new fuzzy subset approach is introduced to develop a truly independent CPT soil engineering classification, and to establish a transition between the new fuzzy approach and conventional soil classifications by utilizing local site- and project-specific calibrations. CPT results conducted at the National Geotechnical Experimentation Site at Texas A&M University are used to demonstrate this new CPT soil engineering classification methodology.

INTRODUCTION

A soil engineering classification from the cone penetration test (CPT) relies on the physical response of soils to cone intrusion, which indicates the mechanical properties of geologic media tested, such as the strength, compressibility, and so on. Due to complicated environmental conditions, the correlation between soil composition and mechanical properties will never be a simple one-to-one correspondence, as discussed previously by the writers (Zhang and Tumay 1996). The CPT soil classification charts currently used do not provide an accurate prediction of soil types defined by compositional indices, but only serve as a guide to soil behavior type (Douglas and Olsen 1981; Campanella and Robertson 1988). Because of the way in which current CPT classification charts were developed, it is sometimes unavoidable to incorrectly identify soil types. Fig. 1 displays a CPT sounding data with its boring log profile, the corresponding Robertson and Campanella (1983a,b) soil classification, and Olsen and Mitchell (1995) soil classifications at the National Geotechnical Experimentation Site (NGES) at Texas A&M University (Simon and Briaud 1996). The somewhat questionable agreement between the carefully documented boring log and the soil types derived from current CPT soil classification charts shows disparities that would have to be resolved by additional CPT tests or costly soil borings. This inability to correctly identify the sediments, specifically in transition zones, and the absence of a quantifiable parameter in mixed/interbedded strata (i.e., clayey, silty, sandy; pockets/seams) are the obvious shortcomings. The probability that soil types could be incorrectly identified by current CPT classification methodologies requires a new soil classification approach to address this probability explicitly so that the user will not be misled. Nonconventional approaches, such as statistical and fuzzy methodologies, are therefore investigated here to meet this need and to help users to correctly interpret and use CPT soil classification.

NORMAL DISTRIBUTION OF SOIL CLASSIFICATION INDEX *U*

The nonconventional approaches presented here are based upon CPT soil engineering classification index U that results from preliminary data reduction on CPT sounding data (Zhang and Tumay 1996). The results published in the previous research indicate empirically that U has a statistical correlation with soil types defined by a soil type index (SI) presented in Table 1, and each soil type has a pronormal frequency distribution of index U. Therefore, U is treated or approximated as a random variable, with assumed normal distributions to be determined by using the results from the 1996 paper. The statistical analysis of the U data was accomplished by using STATGRAPHICS software (*Statgraphics* 1989).

The validation of the normal distribution assumption is checked first by two distribution fitting tests. Table 2 shows the results from chi-square and Kolmogorov-Smirnov one-sample tests. Dashes in the table signify that the size of the corresponding sample is too small to perform the test. These results indicate with confidence that all the soil samples follow the normal distributions. Fig. 2 exhibits all the normal density functions of the seven soil types.

REGION ESTIMATION (ZHANG 1994)

Region estimation is an approach to classify soils, similar to a conventional soil classification. In this approach, the index U-axis is divided into several regions. The soil type predicted for a given soil sample would depend on a U region within which the U value of the sample falls. Each U region, however, will correspond to not just one soil type, but several ones, each with different probabilities. In other words, it directly addresses the probability of the misidentification of soil types in situ.

The U regions are determined using the following assumptions:

1. The sample density functions presented in Fig. 2 are the real density functions of soil types.
2. All seven soil types identified in Table 1 are equally important when the determination of boundary values is considered and performed between adjacent soil types.

The boundary values $U_{i,i+1}$ for soil types i and $i+1$, as defined in Table 1, are derived from the condition

$$1 - F_i(U_{i,i+1}) = F_{i+1}(U_{i,i+1}) \tag{1}$$

where F_i and F_{i+1} = cumulative distribution functions for soil types i and $i+1$ ($i = 1, 2, \ldots, 6$). Eq. (1) gives

$$U_{i,i+1} = \frac{\mu_{i+1}\sigma_i + \mu_i\sigma_{i+1}}{\sigma_i + \sigma_{i+1}} \tag{2}$$

where μ_i, σ_i, and μ_{i+1}, σ_{i+1} = means and standard deviations for soil types i and $i+1$, respectively ($i = 1, 2, \ldots, 6$). The boundary values calculated by (2) will divide the U axis into

[1]Engr., Louisiana Dept. of Transp. and Devel., Baton Rouge, LA 70804.

[2]Assoc. Dean for Res. and Prof. of Civ. and Envir. Engrg., Coll. of Engrg., Louisiana State Univ., Baton Rouge, LA 70803.

Note. Discussion open until August 1, 1999. To extend the closing date one month, a written request must be filed with the ASCE Manager of Journals. The manuscript for this paper was submitted for review and possible publication on May 13, 1996. This paper is part of the ***Journal of Geotechnical and Geoenvironmental Engineering***, Vol. 125, No. 3, March, 1999. ©ASCE, ISSN 1090-0241/99/0003-0179–0186/$8.00 + $.50 per page. Paper No. 13267.

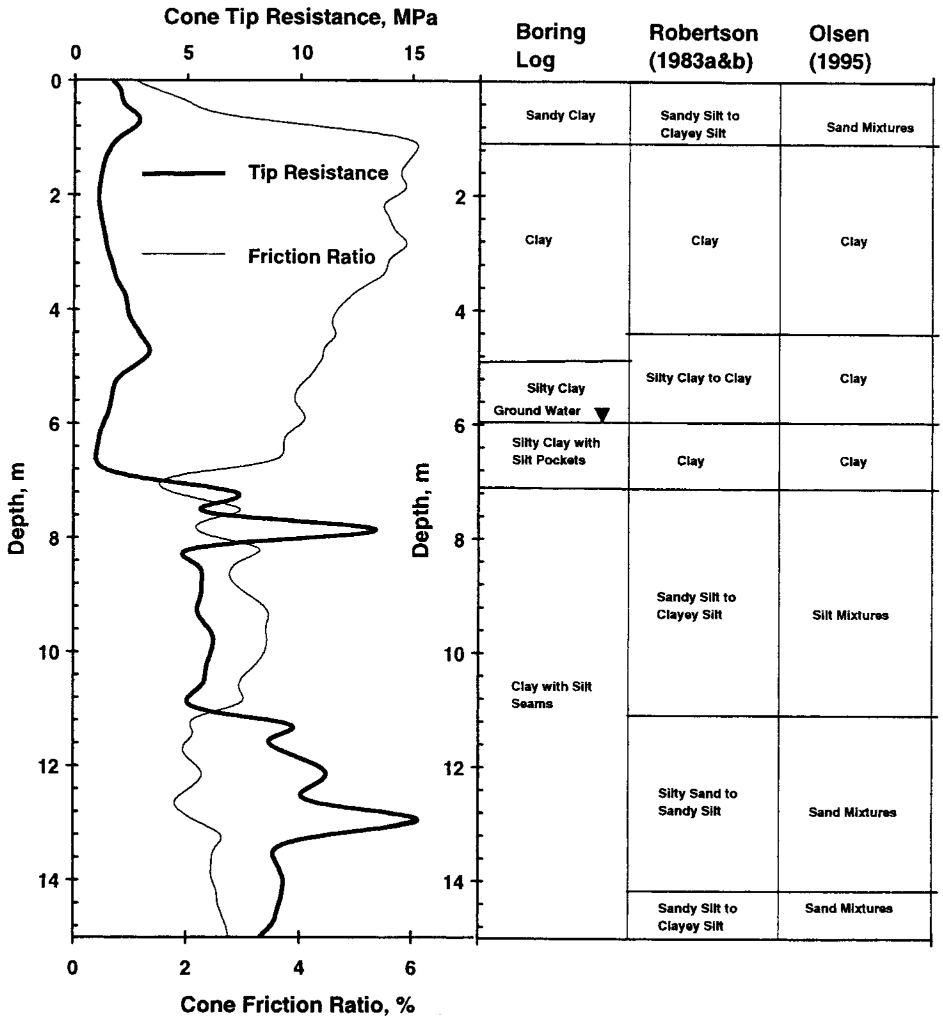

FIG. 1. CPT Data with Its Boring Log Profile and CPT Classifications for National Geotechnical Experimentation Site at Texas A&M University (NGES/Texas A&M)

TABLE 1. Soil Types from Unified Soil Classification System (*Standard* 1987)

Symbol (1)	Soil type index (*SI*) (2)	Typical names (3)
GP	1	Poorly graded gravels, gravel-sand mixtures, little or no fines
SP	2	Poorly graded sands, gravelly sands, little or no fines
SM	3	Silty sands, poorly graded sand-silt mixtures
SC	4	Clayey sands, poorly graded sand-clay mixtures
ML	5	Inorganic silts and very fine sands, rock flour, silty or clayey fine sands with slight plasticity
CL	6	Inorganic clays of low to medium plasticity, gravelly clays, sandy clays, silty clays, lean clays
CH	7	Inorganic clays of high plasticity, fat clays

seven regions, $R_1, R_2, \ldots, R_7$, which constitute the base of region estimation as shown in Table 3. Table 4 gives the probabilities $q_{i,j}$ that an i type of soil falls in a j region of U. These $q_{i,j}$ values are determined by

$$q_{i,j} = F_i(U_{j,j+1}) - F_i(U_{j-1,j}) \tag{3}$$

Here, $i, j = 1, 2, \ldots, 7$; $U_{0,1} = -\infty$; $U_{7,8} = +\infty$; and

$$\sum_{j=1}^{7} q_{i,j} = 1 \tag{4}$$

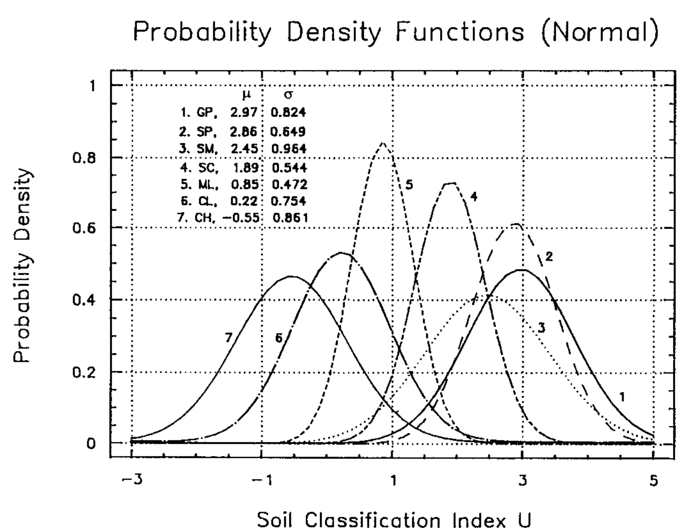

FIG. 2. Normal Density Functions for Seven Soil Types

Table 5 presents the probabilities $p_{i,j}$ with which a j region contains an i type of soil. Probability $p_{i,j}$ is determined by

$$p_{i,j} = \frac{q_{i,j}}{\sum_{i=1}^{7} q_{i,j}} \tag{5}$$

Here, $i, j = 1, 2, \ldots, 7$, too. Table 5, with the primary clas-

TABLE 2. Results of Distribution Fitting Tests

Soil type (1)	Sample size (2)	Distribution function (see Fig. 2) (3)	Mean μ (4)	Standard deviation σ (5)	Chi-Square Test: Estimate (6)	Chi-Square Test: Degrees of freedom (7)	Chi-Square Test: Significance level (8)	*K-S* Test: D_{plus} (9)	*K-S* Test: D_{mi} (10)	*K-S* Test: Significance level (11)
GP	6	F_1	2.97	0.824	—	—	—	0.116	0.174	0.993
SP	115	F_2	2.86	0.649	4.07	7	7.7*E*-1	0.035	0.065	0.708
SM	69	F_3	2.45	0.964	54.43	7	1.9*E*-9	0.167	0.173	0.032
SC	18	F_4	1.89	0.544	—	—	—	0.115	0.126	0.939
ML	17	F_5	0.85	0.472	—	—	—	0.141	0.097	0.889
CL	77	F_6	0.22	0.754	6.36	6	3.8*E*-1	0.066	0.053	0.887
CH	85	F_7	−0.55	0.861	23.90	9	4.5*E*-3	0.071	0.101	0.355

Note: D_{plus} = maximum positive deviation of empirical cumulative distribution over hypothesized cumulative distribution function; D_{mi} = maximum negative deviation of empirical cumulative distribution over hypothesized cumulative distribution function.

TABLE 3. Division of Seven Regions over *U*-Axis

Between regions (1)	Boundary value, $U_{i,i+1}$ (2)	$1 - F_i(U_{i,i+1})$ (3)
R_1 and R_2	2.91	0.471
R_2 and R_3	2.70	0.401
R_3 and R_4	2.01	0.354
R_4 and R_5	1.33	0.152
R_5 and R_6	0.61	0.306
R_6 and R_7	−0.14	0.316

TABLE 4. Probability with Which Different Soil Types Fall in Each Region

Distribution (1)	Probability $q_{i,j}$ over Regions: Region R_1 (2)	Region R_2 (3)	Region R_3 (4)	Region R_4 (5)	Region R_5 (6)	Region R_6 (7)	Region R_7 (8)
F_1	0.5285	0.0991	0.2274	0.1214	0.0214	0.0021	0.0000
F_2	0.4715	0.1274	0.2821	0.1098	0.0090	0.0002	0.0000
F_3	0.3200	0.0811	0.2452	0.2323	0.0940	0.0238	0.0036
F_4	0.0308	0.0380	0.2848	0.4944	0.1428	0.0090	0.0001
F_5	0.0000	0.0000	0.0040	0.1478	0.5421	0.2873	0.0187
F_6	0.0002	0.0003	0.0060	0.0642	0.2353	0.3781	0.3159
F_7	0.0000	0.0000	0.0009	0.0134	0.0754	0.2261	0.6841
[Subtotal]	1.3510	0.3459	1.0504	1.1833	1.1200	0.9266	1.0224

TABLE 5. Probability with Which Each Region Receives Different Soil Types

Soil type (1)	Probability $p_{i,j}$ over Regions: Region R_1 (2)	Region R_2 (3)	Region R_3 (4)	Region R_4 (5)	Region R_5 (6)	Region R_6 (7)	Region R_7 (8)
GP	0.3912	0.2865	0.2165	0.1026	0.0191	0.0023	0.0000
SP	0.3490	0.3683	0.2686	0.0928	0.0080	0.0002	0.0000
SM	0.2369	0.2345	0.2334	0.1963	0.0839	0.0257	0.0035
SC	0.0229	0.1098	0.2711	0.4178	0.1275	0.0097	0.0001
ML	0.0000	0.0000	0.0038	0.1249	0.4840	0.3100	0.0183
CL	0.0000	0.0009	0.0057	0.0542	0.2100	0.4080	0.3090
CH	0.0000	0.0000	0.0009	0.0113	0.0675	0.2441	0.6691

TABLE 6. Simplified Results from Table 5

Soil type (1)	Probability $p_{i,j}$ over Regions: Region R_1 (2)	Region R_2 (3)	Region R_3 (4)	Region R_4 (5)	Region R_5 (6)	Region R_6 (7)	Region R_7 (8)
GP, SP, SM	0.9771	0.8893	0.7185	0.3971	0.1110	0.0282	0.0035
SC, ML	0.0229	0.1098	0.2749	0.5427	0.6115	0.3197	0.0184
CL, CH	0.0000	0.0009	0.0066	0.0655	0.2775	0.6521	0.9781
GP, SP, SM, SC	1.0000	0.9991	0.9896	0.0896	0.2358	0.0379	0.0036
ML, CL, CH	0.0000	0.0009	0.0104	0.1904	0.7615	0.9621	0.9964

sification frame in Table 3, consists of the region estimation approach to identify different soil types with certain probability.

Table 5 can be rearranged by grouping soil types, as shown in Table 6. This new table indicates that sandy and gravelly soils generally fall in regions 1, 2, and 3; silty soils in regions 4 and 5; and clayey soils in regions 6 and 7. If the silty soils are further divided and merged with the sandy and clayey soils separately, the boundary value between regions 4 and 5 can reasonably be taken as the dividing point. These results are summarized and depicted in Fig. 3.

Region estimation assumes that different points of a region have exactly the same statistical property so that the whole region is treated in exactly the same way. This assumption sometimes is far from reality, since points in a region can have a big difference in probabilities corresponding to different soil types. From a theoretical point of view, the validation of the region estimation therefore is somewhat limited. This problem may be resolved by another approach called point estimation, where every point in a *U* region is treated distinctively.

POINT ESTIMATION

Point estimation is an approach to predict soil types directly by probability, where each probable *U* value is evaluated individually. The fundamental question for this approach to an-

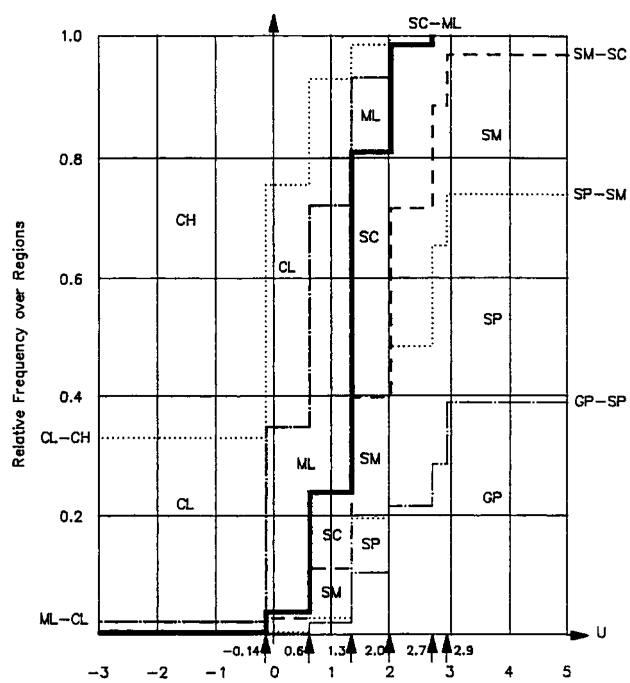

FIG. 3. Region Estimation Chart for CPT Soil Classification

swer is this: Given a specific U value, what are the probabilities that the corresponding soil sample belongs to a different soil type?

A probabilistic model of two dimensions has been derived for point estimation. One dimension is the continuous random variable of soil index U. The other is a discrete random variable representing soil types, as defined in Table 1. Both variables are defined in the soil media considered. Discussing a hypothetical case can enhance understanding of this probabilistic model.

Assume there is a box containing balls with different colors. The number of colors is N and the number of balls is infinite for each color. If a ball is taken out of the box, the probability, P(color), for a specific color selected is $1/N$; i.e., $P(\text{color}) = 1/N$.

In addition to the prior assumptions, suppose that a quantity x can be measured for each ball, and this quantity is a continuous random variable X with a specific distribution for each color. Now, if a ball is taken out of the box revealing a corresponding x value, the probability of a specific color being selected will be modified by the x value, since balls of different colors will carry x values with different probabilities. Then, what will be the modification of this x value on the probability that a ball of a specific color is selected, or $P(\text{color}|X = x) = ?$

The solution to this problem, according to the multiplication rule in probability theory (Hoel et al. 1971), will be

$$P(\text{color}|X = x) = \frac{P[(X = x) \cap (\text{color})]}{P(X = x)} \tag{6}$$

Since X is a continuous random variable, $P(X = x) = 0$ exists. So (6) needs to be changed to

$$P(\text{color}|X = x) = \lim_{\Delta x \to 0} \frac{P[(x \le X < x + \Delta x) \cap (\text{color})]}{P(x \le X < x + \Delta x)} \tag{7}$$

Now, imagine that the big box is the earth and the countless balls are the soil media. The continuous random variable X is the soil classification index U and the different colors stand for different soil types to be identified in a site investigation. Then, the solution to statistically predict a soil type based on a single U observation should be

$$P(\text{soil type}|U = u) = \lim_{\Delta u \to 0} \frac{P[(u \le U < u + \Delta u) \cap (\text{soil type})]}{P(u \le U < u + \Delta u)} \tag{8}$$

If a soil type is represented by a discrete numerical random variable SI (soil type index from Table 1), (8) will be

$$P(SI = si|U = u) = \lim_{\Delta u \to 0} \frac{P[(u \le U < u + \Delta u) \cap (SI = si)]}{P(u \le U < u + \Delta u)} \tag{9}$$

Here, SI will take values of $1, 2, \ldots, M$. M is the number of soil types in a specific soil classification system. Uppercase SI or U signifies a variable; lowercase si or u is the value of that variable.

Suppose $F(u, si)$ is a two-dimensional probability distribution function of U and SI, and $F_m(u)$ is its marginal distribution function of U. By definition, the probabilities of events in (9) can be rewritten in terms of $F(u, si)$ and $F_m(u)$. That is

$$P[(u \le U < u + \Delta u) \cap (SI = si)] = F(u + \Delta u, si) - F(u, si) - F(u + \Delta u, si - 1) + F(u, si - 1) \tag{10}$$

and

$$P(u \le U < u + \Delta u) = F_m(u + \Delta u) - F_m(u) \tag{11}$$

In probability theory, the two-dimensional distribution function $F(u, si)$ is defined as

$$F(u, si) = \sum_{y=1}^{si} \int_{-\infty}^{u} f(x, y)\, dx \tag{12}$$

or

$$F(u, si) = \int_{-\infty}^{u} \sum_{y=1}^{si} f(x, y)\, dx \tag{13}$$

and x and y are used in place of u and si to observe the rule of integration and avoid possible confusion. The corresponding marginal distribution function of soil classification index U is then

$$F_m(u) = \int_{-\infty}^{u} \sum_{si=1}^{M} f(x, si)\, dx \tag{14}$$

or

$$F_m(u) = \sum_{si=1}^{M} \int_{-\infty}^{u} f(x, si)\, dx \tag{15}$$

The marginal distribution function of SI has the form

$$F_m(si) = \sum_{y=1}^{si} \int_{-\infty}^{+\infty} f(u, y)\, du = \sum_{y=1}^{si} q(y) \tag{16}$$

with

$$q(y) = \int_{-\infty}^{+\infty} f(u, y)\, du; \quad y = 1, 2, \ldots, si \tag{17}$$

Under certain mathematical assumptions, (12) is actually equivalent to (13), and (14) is equivalent to (15). Substituting (12) into (10) and rearranging produces

$$\begin{aligned} P[(u \le U < u + \Delta u) \cap (SI = si)] &= \sum_{y=1}^{si} \left[\int_{-\infty}^{u+\Delta u} f(x, y)\, dx - \int_{-\infty}^{u} f(x, y)\, dx \right] \\ &\quad - \sum_{y=1}^{si-1} \left[\int_{-\infty}^{u+\Delta u} f(x, y)\, dx - \int_{-\infty}^{u} f(x, y)\, dx \right] \end{aligned} \tag{18}$$

Also, substituting (15) into (11) yields

$$P(u \le U < u + \Delta u) = \sum_{si=1}^{M} \int_{u}^{u+\Delta u} f(x, si)\, dx \tag{19}$$

Now, the two-dimensional density function $f(u, si)$ in the previous equations has to be determined. From the definition of $F(u, si)$, we have

$$F(u, si - 1) = P(U \le u, SI \le si - 1) \tag{20}$$

and

$$F(u, si) = P(U \le u, SI \le si) \tag{21}$$

Subtracting (20) from (21) gives

$$F(u, si) - F(u, si - 1) = P(U \le u, SI \le si) - P(U \le u, SI \le si - 1) = P[(U \le u) \cap (SI = si)] \tag{22}$$

since SI is a discrete random variable. Therefore

$$\begin{aligned} P[(U \le u) \cap (SI = si)] &= F(u, si) - F(u, si - 1) \\ &= \sum_{y=1}^{si} \int_{-\infty}^{u} f(x, y)\, dx - \sum_{y=1}^{si-1} \int_{-\infty}^{u} f(x, y)\, dx = \int_{-\infty}^{u} f(x, si)\, dx \end{aligned} \tag{23}$$

On the other hand, the multiplication rule will give

$$P[(U \leq u) \cap (SI = si)] = P(SI = si) \cdot P[(U \leq u)|(SI = si)] \qquad (24)$$

By definition

$$P[(U \leq u)|(SI = si)] = \int_{-\infty}^{u} g_{si}(x)\, dx \qquad (25)$$

where $g_{si}(x)$ in (25) represents $g_{si}(u)$, the conditional density function of U for soil type si. Also, according to (16), $P(SI = si)$ in (24) is

$$P(SI = si) = F_m(si) - F_m(si - 1) = q(si) \qquad (26)$$

It is the probability with which a soil type is met randomly in situ, and is constant for soil type si.

Now, substitute (25) and (26) into (24), and then (24) into (23); we get

$$\int_{-\infty}^{u} f(x, si)\, dx = q(si) \int_{-\infty}^{u} g_{si}(x)\, dx \qquad (27)$$

Since u can be any real value, (27) exists only if

$$f(u, si) = q(si)g_{si}(u) \qquad (28)$$

This is the two-dimensional density function needed.

Substitute (28) into (18) and (19). After some derivation, we have

$$P[(u \leq U < u + \Delta u) \cap (SI = si)] = \int_{u}^{u+\Delta u} f(x, si)\, dx = q(si) \int_{u}^{u+\Delta u} g_{si}(x)\, dx \qquad (29)$$

and

$$P(u \leq U < u + \Delta u) = \sum_{si=1}^{M} q(si) \int_{u}^{u+\Delta u} g_{si}(x)\, dx \qquad (30)$$

Therefore, the conditional probability determined by (9) will be

$$P(SI = si|U = u) = \lim_{\Delta u \to 0} \frac{q(si) \int_{u}^{u+\Delta u} g_{si}(x)\, dx}{\sum_{si=1}^{M} q(si) \int_{u}^{u+\Delta u} g_{si}(x)\, dx} \qquad (31)$$

Since Δu is an infinitesimal quantity, (31) can be rewritten as

$$P(SI = si|U = u) = \lim_{\Delta u \to 0} \frac{q(si)g_{si}(u)\Delta u}{\sum_{si=1}^{M} q(si)g_{si}(u)\Delta u} \qquad (32)$$

The limit in (32) is actually independent of Δu. Consequently, we have

$$P(SI = si|U = u) = \frac{q(si)g_{si}(u)}{\sum_{si=1}^{M} q(si)g_{si}(u)} \qquad (33)$$

This is the basic formula to perform the point estimation.

Eq. (33) requires a series of conditional density functions $g_{si}(u)$, $si = 1, 2, \ldots, M$, which are the normal distributions of the soil classification index U for different soil types, as shown in Fig. 2. These soil types are represented by the variable soil type index, SI, defined in Table 1. Also defined previously, $q(si)$ is the probability with which an si soil type is found randomly in situ. Fig. 4 shows one example of the conditional probabilities determined by (33). They are the conditional

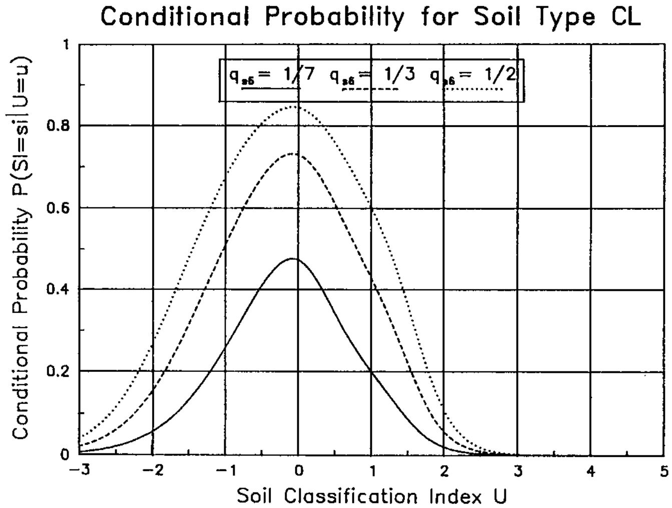

FIG. 4. CPT Point Estimation Chart for Soil Type CL

probabilities for soil type CL, assuming function $q(si = 6)$ takes values of 1/7, 1/3, and 1/2, respectively. These $q(si)$ values are only for estimation purposes. With $q(si)$ taking a value of 1/7, the seven soil types considered in the present study would generally have the same probability of being encountered in situ. In the case that soil type CL may have higher probabilities of being encountered than other soil types, the results from $q(si = 6) = 1/3$ and 1/2 will provide a range for estimation. Also, when $q(si = 6)$ is equal to one of $\beta = 1/7$, 1/3, 1/2, the $q(si)$ values for the other soil types are assumed to be $(1 - \beta)/6$ for simplicity.

FUZZY CLASSIFICATION

The results obtained so far have shown the statistical correlation between the soil engineering classification by CPT profile data and the Unified Soil Classification System (USCS). In using CPT profile data to identify soil types, we actually classify geo-media according to their mechanical behavior characteristics. Like the USCS, this kind of classification will also be useful for helping us study and understand the engineering properties of in situ soils. Thus, there is a need for a truly independent CPT soil engineering classification to be developed. In such a new system, the classification criteria will not be borrowed or inherited from the USCS, but will consider the accumulated engineering experience related to it. The fuzzy subset theory (Kaufmann 1975; Brown et al. 1985) is an ideal tool to accomplish this task. The suggested fuzzy approach will, contrary to a conventional soil engineering classification, release the constraint of soil composition and put an emphasis on the certainty: soil behavior (i.e., cone tip resistance and local friction).

Three fuzzy soil types are defined in a CPT fuzzy soil classification. They are highly probable clayey soil (HPC), highly probable mixed soil (HPM), and highly probable sandy soil (HPS). Empirically, the three membership functions of HPS, HPM, and HPC are determined based upon the data (relative frequency of U) given in Fig. 5 (Zhang and Tumay 1996). In this representation, soils have been reorganized into three groups (group 1, group 2, and group 3) that are directly related to HPS, HPM, and HPC. The density functions of the three soil groups are approximated as

$$f_s(u) = \frac{1}{0.834586\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{u - 2.6575}{0.834586}\right)^2\right] \qquad (34)$$

$$f_m(u) = \frac{1}{0.724307\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{u - 1.35}{0.724307}\right)^2\right] \qquad (35)$$

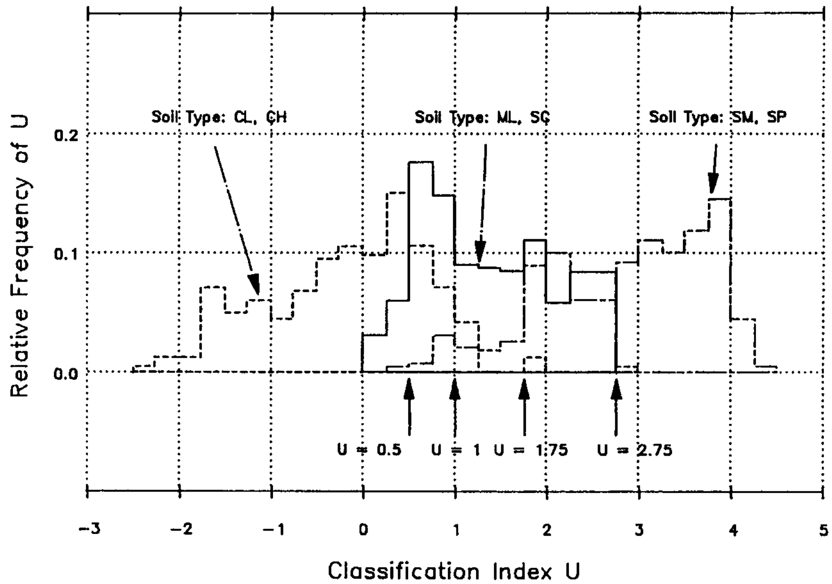

FIG. 5. Empirical CPT Soil Classification for Three General Groups: Group 1—SM and SP; Group 2—ML and SC; Group 3—CL and CH

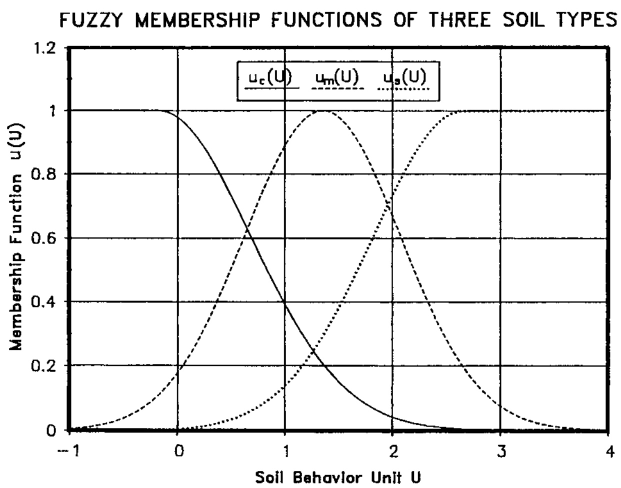

FIG. 6. Tentative CPT Fuzzy Soil Classification Chart

$$f_c(u) = \frac{1}{0.86332\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{u + 0.1775}{0.86332}\right)^2\right] \quad (36)$$

After normalization and other empirical modifications on (34), (35), and (36), the three fuzzy membership functions of HPS, HPM, and HPC are defined as (Zhang 1994)

$$\mu_s(u) = \begin{cases} 1.0 & \text{for} \quad u > 2.6575 \\ \exp\left[-\frac{1}{2}\left(\frac{u - 2.6575}{0.834586}\right)^2\right] & \text{for} \quad u \le 2.6575 \end{cases} \quad (37)$$

$$\mu_m(u) = \exp\left[-\frac{1}{2}\left(\frac{u - 1.35}{0.724307}\right)^2\right]; \quad -\infty < u < \infty \quad (38)$$

$$\mu_c(u) = \begin{cases} \exp\left[-\frac{1}{2}\left(\frac{u + 0.1775}{0.86332}\right)^2\right] & \text{for} \quad u \ge 0.1775 \\ 1.0 & \text{for} \quad u < 0.1775 \end{cases} \quad (39)$$

as shown in Fig. 6.

As shown in Fig. 6, these empirically defined functions will have a maximum value of 1 and an ''S'' or a bell-shaped curve so that they can approximately reveal the law of quantity change to quality change concerning soil composition and properties. The three fuzzy membership functions as a whole will reflect the overall perspective of soil properties. Soils in these three groups have fundamentally different engineering properties, but with no sharp boundaries between them. The changes are gradual from one soil type to another.

Three soil types HPS, HPM, and HPC with their fuzzy membership functions compose a basic CPT fuzzy soil classification with several unique characteristics and properties. First, it is a real behavior type of soil classification, where not only the classification index but also the classification criteria are based upon the measurements of soil behavior. There is no place in this classification for misidentification of soil types, i.e., the uncertainty from randomness.

Second, soil types in this classification include the information about their relations with the corresponding compositional soil types. It clearly states that these soil types correspond to their counterparts in USCS with high probabilities, but not always 100%. The emphasis is put on the certainty of CPT data: soil behavior. Since ''highly probable'' is an ambiguous term, the variation of these soil types is described by the fuzzy membership functions.

Third, this fuzzy soil classification includes an empirical summary of current knowledge about soil behavior. HPS generally has the properties of high strength, high permeability, and low compressibility, which correspond to a higher tip resistance q_c and a lower friction ratio R_f, and therefore a larger U value. HPC is supposed to have a lower strength, a lower permeability, and a higher compressibility that is usually consistent with a lower q_c and a higher R_f, and therefore a lower U value. The engineering properties of HPM lie in between HPS and HPC. Thus, its U will take a value between the values of HPS and HPC.

Fourth, this classification provides a tool to separate the description of soils in situ from the simplification of the soil state in situ. Here, the simplification means a conventional process to define the stratigraphy of a testing site. By using this new type of soil classification, the in situ soil situation can be depicted continuously, and the identification of the stratigraphy of that site can be postponed. With a condensed format, the fuzzy soil classification describes the situation of soil in situ as it is. No simplification is performed in advance. A conditional truncation on fuzzy soil types (simplification) can be conducted later as need occurs and some criteria are available. The resulting crisp soil types (CST) are expressed by a fuzzy classification with a group of α threshold values

$$CST_j = [u|\alpha_j \le \mu_i(u)]; \quad i = c, m, s \tag{40}$$

Here, i and j = subscripts representing fuzzy and crisp soil types, separately. These α values can be determined for different engineering concerns. In this way, different users can also accumulate and express their own engineering experience mathematically. Here is an example. Suppose that CPT data are used to check the liquefaction potential of a soil so that some precautionary measures can be taken. This soil could be defined, for instance, by the following criterion:

$$CST_{\text{liquefiable}} = [u|\alpha_{\text{liquefiable}} = 0.6 \le \mu_m(u)] \tag{41}$$

where $\mu_m(u)$ is given by (38).

Last but not least, the CPT fuzzy soil classification will further serve a general communication purpose. An example of this can be: Given $U = 2.0$, the corresponding soil layer belongs to (''looks like''), depending upon the aspect that needs to be emphasized,

- HPC with a degree of $\mu_c(2.0) = 0.04$
- HPM with a degree of $\mu_m(2.0) = 0.67$
- HPS with a degree of $\mu_s(2.0) = 0.73$

This kind of description is used in daily life. For instance, consider the following statement: A boy looks more like his mother (with a degree 0.7, or 70%) than his father (with a degree of 0.4, or 40%). People have the capability to understand such a statement with little confusion, although it is a fuzzy description. $U = 2.0$ is directly calculated from actual cone measurements of tip resistance, q_c and frictional resistance, f_s. Therefore, the engineering behavior of that soil layer is certain.

CPT SOIL CLASSIFICATION RESULTS

The results from using the new CPT soil engineering classification methodology, suggested in the present paper and papers published previously by the writers, will at least include the following:

1. A soil classification index U profile from CPT sounding data
2. Corresponding probability profiles to determine probable USCS (compositional) soil types
3. A fuzzy soil type index profile (the values of the fuzzy membership functions) to determine the soil types of the fuzzy CPT soil classification

If required, some other criteria for different engineering concerns can be used to produce additional profiles, and the whole process can be routinely implemented and easily computerized.

A typical application of the suggested methodology of using a statistical to fuzzy approach in soil classification by CPT is presented in Figs. 7 and 8. This example is based upon the CPT data from the NGES at Texas A&M University, shown in Fig. 1. Fig. 7 depicts a profile of related probabilities of soil types encountered based upon the three-component region estimation suggested in the present study. As the soil classification index U varies with the depth, the probabilities with which the sediments are classified as a sandy soil (GP, SP, SM), silty soil (SC, ML), or clayey soil (CL, CH) will also vary. At some locations, sandy or clayey soils can be identified clearly and reliably due to the very high probabilities. At other locations, the U values can be statistically related to sandy, silty, or clayey soils with different individual probabilities. In other words, due to the fact that different soil types can sometimes share the same U values, the exact soil types cannot be verified without additional soil investigations (i.e., soil bor-

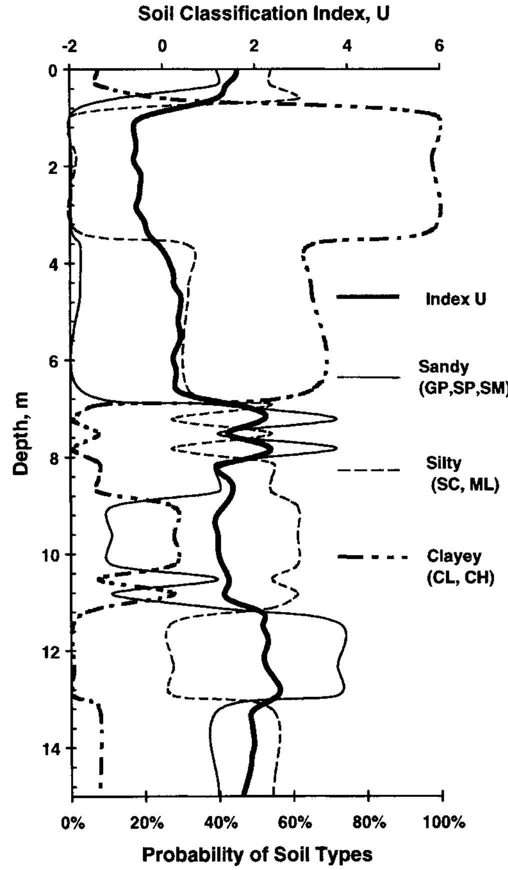

FIG. 7. Probability Profile Related to NGES/Texas A&M CPT Data

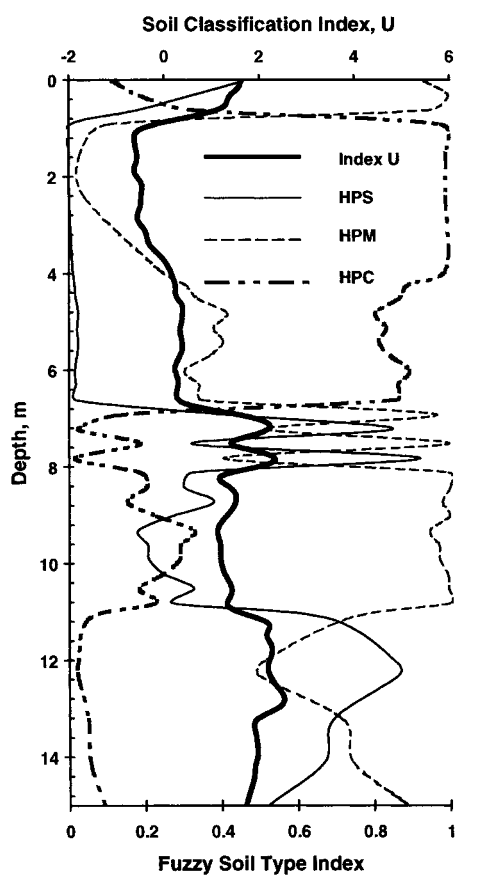

FIG. 8. Fuzzy Soil Type Index Profile Related to NGES/Texas A&M CPT Data

ings). The significance of this information is that it alerts users about the existence of the problematic sediment layering (transition zones or mixed/interbedded strata, etc.) that otherwise would be neglected.

Fig. 8 presents a profile of the fuzzy soil type indices related to the CPT data in Fig. 1. The indices describe and classify the soil media in Fig. 1 from a different viewpoint. For instance, the HPC index shows how the mechanical behaviors of the soils in the profile correspond to the mechanical behavior of a typical clayey soil. It is a kind of comparison and classification without simplifying both within and between soil types. The three variable indices provide a more comprehensive perspective of soils encountered in Fig. 1. This is a fundamental classification upon which other site- or project-specific classifications can be integrated by using (40). For example, a local soil classification corresponding to the USCS can be developed based on local experience by calibrating the α values in (40) with actual data accumulated from available conventional soil boring logs and laboratory tests.

CONCLUSIONS

The current paper presents a new methodology to classify soils by using CPT data. This methodology includes the statistical and fuzzy subset approaches, and intends to address the problem of potentially misidentifying soil types inherent in using the existing CPT soil engineering classifications. Although there is still room for improvement by further accumulation and calibration of in situ data at well-documented sites, the fundamental structure of this methodology sets up a logical connection between the site- and project-specific needs of geotechnical engineers and the reality of the complex nature of geo-materials in the field. The fact that there is no simple one-to-one correlation between soil composition and mechanical properties dictates two alternatives: (1) Establish a statistical correlation between CPT data and compositional soil types, such as the ones in USCS (a statistical approach); and (2) classify soils independently by the values of CPT data in order to evaluate the behavior of different soil types. A totally new soil classification would have the apparent disadvantage of disregarding the vast past experience related to a long time–used methodology. However, the fuzzy approach can build a bridge between the new and old by a local calibration of α threshold values obtained from actual field data (boring logs and laboratory test results). The consequent statistical and fuzzy classifications will make users aware of the existence and interpretation of problematic stratification in geo-media (transition zones or mixed/interbedded strata, etc.) that might otherwise be misinterpreted. Due to its flexible format, the fuzzy approach also adds value to fundamental classifications by integration of other site- or project-specific behavior properties necessary for versatile soil engineering predictions.

ACKNOWLEDGMENTS

This research was supported by the Louisiana Transportation Research Center (LTRC) under contract number 88-1 GT [state project number 746-13-36, Louisiana HPR number 0010(12)] and partially by the National Science Foundation, MSS-9018249. LTRC is jointly sponsored by the Louisiana State University and the Louisiana Department of Transportation and Development. The views reflected in this paper are not necessarily those of the agencies and institutions with which the writers are affiliated.

APPENDIX. REFERENCES

Brown, C. B., Chameau, J. L., Palmer, R., and Yao, J. T. P. (1985). *Proc., NSF Workshop on Civ. Engrg. Applications of Fuzzy Sets*, Purdue University, West Lafayette, Ind.

Campanella, R. G., and Robertson, P. K. (1988). ''Current status of the piezocone test.'' *Proc., Penetration Testing 1988, ISOPT-1*, De Ruiter, ed., Orlando, Fla., 93–116.

Douglas, B. J., and Olsen, R. S. (1981). ''Soil classification using the electric cone penetrometer.'' *Proc., Cone Penetration Testing and Experience*, ASCE, Reston, Va., 209–227.

Hoel, P. G., Port, S. C., and Stone, C. J. (1971). *Introduction to statistical theory*. Houghton Mifflin, Boston.

Kaufmann, A. (1975). *Introduction to the theory of fuzzy subsets, I: Fundamental theoretical elements*. Academic Press, New York.

Olsen, R. S., and Mitchell, J. K. (1995). ''CPT stress normalization and prediction of soil classification.'' *Proc., CPT '95*, Swedish Geotechnical, Linkoping, Sweden, 2, 257–262.

Robertson, P. K., and Campanella, R. G. (1983a). ''Interpretation of cone penetration tests. Part I: Sand.'' *Can. Geotech. J.*, 20(4), 718–733.

Robertson, P. K., and Campanella, R. G. (1983b). ''Interpretation of cone penetration tests. Part II: Clay.'' *Can. Geotech. J.*, 20(4), 734–745.

Simon, P. A., and Briaud, J.-L. (1996). ''The national geotechnical experimentation sites at Texas A&M University: Clay and sand.'' *NGES-TAMU-006*—December, Texas A&M University, College Station, Tex.

Standard test method for classification of soils for engineering purposes—Annual book of ASTM standards; soil and rock, building stones, Vol. 04.08(Section 4); *D2487-85*. (1987). ASTM, West Conshohocken, Pa.

Statgraphics—Version 4.0: Statistical graphics system. (1989). Statistical Graphics Corp., STSC.

Zhang, Z. (1994). ''Use of uncertainty methodology in identification and classification of soils based upon CPT,'' PhD dissertation, Louisiana State University, Baton Rouge, La.

Zhang, Z., and Tumay, M. T. (1996). ''The reliability of soil classification derived from cone penetration test.'' *ASCE Spec. Conf.—Uncertainty in the geologic environment: From theory to practice*, ASCE, Reston, Va., 383–408.