Novel methodologies for soil characterization from CPT data

P.U. Kurup
University of Massachusetts Lowell, Lowell, MA, USA

E.P. Griffin
Shaw Environmental & Infrastructure, Stoughton, MA, USA

M.T. Tumay
Louisiana State University, Baton Rouge, LA, USA; Bogazici University, Istanbul, Turkey

ABSTRACT: This paper presents results and comparisons of soil characterization from cone penetration test (CPT) data using some of the traditional and nontraditional soil classification methods. Recently, a general regression neural network (GRNN) model was developed for predicting soil composition (percent sand-, silt-, and clay-sized particles) from CPT data. CPT data, together with grain size distribution results of soil samples retrieved from adjacent SPT boreholes from several sites in Taiwan following the Chi-Chi earthquake, were used to train and test the network. The trained GRNN was validated with previously unseen data and the model predictions were compared with the reference particle size distributions and the soil behavior type CPT soil classification method proposed by Robertson. The results were also compared to a nontraditional statistical soil classification approach, termed probabilistic region estimation that was proposed by Zhang and Tumay to estimate the probability of sand, silt, and clay in soils.

1 INTRODUCTION

The conventional method for determining soil stratigraphy is by laboratory classification of samples retrieved from boreholes. If a continuous, or nearly continuous, subsurface profile is desired, the CPT provides time and cost savings over traditional methods of sampling and testing. A number of methods exist to infer, or predict, soil type from CPT and piezocone penetration test (PCPT or CPTu) data. These methods include classical soil classification charts based on extensive experience (Douglas and Olsen 1981; Senneset and Janbu 1985; Robertson et al. 1986; Robertson 1990), and nontraditional approaches based on statistics, probability and fuzzy subset theory (Zhang and Tumay 1999, 2000). The main challenge of using the CPT or PCPT for soil profiling is that samples are not retrieved for laboratory testing. Soil type, or soil behavior type, must therefore be inferred from the information collected during a sounding and, as a result, numerous soil classification charts based on CPT and PCPT data have been developed. Most of the CPT charts classify soils according to measured values of cone resistance (q_c) and friction ratio (R_f), defined as the ratio of f_s to q_c expressed as a percentage, and all show the same trend: sandy soils tend to have higher values of q_c and lower values of R_f, while clayey soils tend to have lower values of q_c and higher values of R_f. With the advent of the piezocone, several soil classification charts have been proposed based on values of q_c corrected for pore pressure effects (termed q_T) and u_2 (Senneset and Janbu 1985), and values of q_T, u_2, and R_f (Robertson et al. 1986).

As noted by Robertson et al. (1986), a recognized problem associated with these aforementioned soil classification charts is that "soils can gradually change in their apparent classification as cone penetration increases in depth." This is because measured cone resistance, sleeve friction, and pore pressure all tend to increase with increasing overburden pressure. Because most of the data used to develop the soil classification charts were obtained from soundings less than 30 m in depth, some errors may arise when using CPT or PCPT data obtained at significantly greater depths. Classification charts based on normalized PCPT measurements would, however, be able to account for the effects of overburden stress. Robertson (1990), therefore, proposed modified soil behavior type classification charts using normalized values of cone resistance (Q_t), and friction ratio (F_r), which are defined as:

$$Q_t = \frac{q_t - \sigma_{vo}}{\sigma'_{vo}} \tag{1}$$

$$F_r = \frac{f_s}{q_T - \sigma_{vo}} \times 100\% \tag{2}$$

where σ_{vo} is the total overburden pressure and σ'_{vo} is the effective vertical pressure.

To overcome the inherent uncertainty in correlating soil composition to mechanical behavior, Zhang and Tumay (1999) proposed a nontraditional soil classification approach based on statistics and probability. The statistical method, termed probabilistic region estimation, is similar to conventional soil classification methods in that it establishes a relationship between CPT data (q_c and f_s) and soil composition, and essentially estimates the probability of sand, silt, and clay in soils being investigated. At some depths, soil type can be clearly identified as a result of having a very high probability (e.g., greater than 80%), while at other depths, the probabilities of two or three soil types are so similar that the exact type, if desired, must be verified by means of a boring (Zhang and Tumay 1999). A computer program *Soil-CPT* capable of CPT-based probabilistic/fuzzy classifications (Zhang and Tumay 1999), and also includes others by Schmertmann (1978), Douglas and Olsen (1981), Robertson et al (1986) is developed for ease of use in performing this classification procedure (Tumay et al 2008). Latest Version 4.0 is available at: http://www.ltrc.lsu.edu/downloads.html

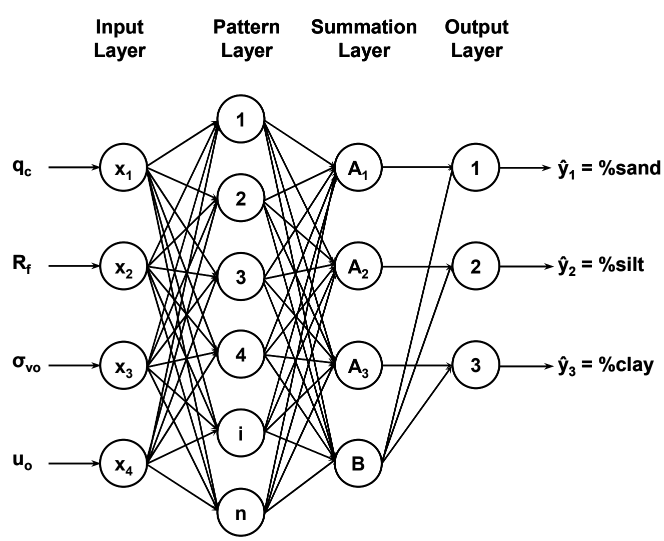
In addition to their statistical approach, Zhang and Tumay (1999) proposed a CPT soil classification method based on fuzzy subset theory with an emphasis on soil behavior (q_c and f_s) rather than soil composition. The CPT fuzzy soil classification defines three soil types: highly probable sandy soil (HPS), characterized by high strength, high permeability, and low compressibility; highly probable clayey soil (HPC), characterized by low strength, low permeability, and high compressibility; and highly probable mixed soil (HPM), whose characteristics fall in between those of HPS and HPC (Zhang and Tumay 1999). Thus, although the soils in these three groups can exhibit vastly different soil behavior, the boundaries between the soil types are "fuzzy" and, as a result, the changes from one soil type to another are gradual.

2 THE GENERAL REGRESSION NEURAL NETWORK

2.1 *The GRNN Algorithm*

The GRNN, developed by Specht (1991), is a multilayer feed-forward (i.e., signals propagate only in a forward direction) neural network which performs general regression analysis directly from sample data for the purpose of prediction. The GRNN architecture for predicting soil composition from CPT data (Figure 1) consists of four fully connected layers, one layer each of input neurons, pattern neurons, summation neurons, and output neurons (Kurup and Griffin 2006). The input layer contains one neuron for each input variable (q_c, R_f, σ_{vo}, u_o), the pattern layer contains one neuron for each training case, the number of both the "A" summation and output neurons is equal to the number of output variables (% sand, % silt, % clay), and there is always only one "B" summation neuron. The connection weights between the input and pattern neurons equal the values of the input parameters in each training case, the connection weights between the pattern and "A" summation neurons equal the values of the target parameters for each training case, and the connection weights between the pattern neurons and "B" summation neuron all equal 1. The neurons in the two hidden (pattern and summation) layers process the incoming signals (outputs of the neurons in the previous layer) by means of these connection weights and a summation and activation function. When the transformed data reaches the output layer, the desired network outputs are produced.

The GRNN basically implements the formula presented in Figure 1. In the formula m is the number of input variables, n is the number of training cases, x_i is the value of the i^{th} variable of the given testing case, x_{ij} is the value of the i^{th} variable of the j^{th} training case, y_j is the value of the target variable of the j^{th} training case, σ is the smoothing parameter which determines how closely the function implemented by the GRNN fits the training data, and $\hat{y}$ is the estimated target value corresponding to the given testing case.



$$\hat{y} = \frac{\sum_{j=1}^{n} y_j \exp\left(\frac{-\sum_{i=1}^{m}(x_i - x_{ij})^2}{2\sigma^2}\right)}{\sum_{j=1}^{n} \exp\left(\frac{-\sum_{i=1}^{m}(x_i - x_{ij})^2}{2\sigma^2}\right)}$$

Figure 1 The GRNN architecture and the governing formula

3 SOIL CLASSIFICATION AND PREDICTION OF SOIL COMPOSITION

3.1 *Background of Study Areas*

An earthquake with a main shock moment magnitude (M_w) of 7.6 struck central Taiwan on September 21, 1999. The epicenter was located near Chi-Chi, and the inland areas which suffered the most incidents of liquefaction include the towns of Yuanlin, Dachun, and Wufeng and the city of Nantou. The geologic environment of these locations is generally characterized by Holocene alluvium sediments with shallow groundwater (Juang et al. 2002). Based on data from the SPT boring logs, the subsurface deposits in the study areas are comprised of alternating layers of lean clay, silty clay, silt, sandy silt, and silty sand.

Shortly after the Chi-Chi earthquake, Moh and Associates conducted an extensive field investigation in these areas. In situ testing included the standard penetration test (SPT), the CPT, and shear-wave velocity measurements (V_s). A total of 80 borings were advanced and 70 CPT soundings were performed at these sites. As part of the boring program, all retrieved split spoon samples were subjected to a full suite of laboratory index tests to determine water content, unit weight, specific gravity of solids, liquid limit, plastic limit, and particle size distribution by means of sieve and hydrometer. The CPT logs include q_c (without corrections for pore pressure effects), f_s, and R_f (Juang 2002).

3.2 *Training and Testing the GRNN Model*

The data obtained from the CPT soundings and adjacent SPT borings advanced in the four study areas were used to train and test the GRNN. A total of 12 CPT-SPT pairs from the four sites fit this criterion, including five from Yuanlin, one from Dachun, three from Nantou, and three from Wufeng. Based on reported survey coordinates, most pairs were less than 6 m apart, with several being co-located.

The four input variables chosen were q_c, R_f, σ_{vo}, and u_o, since soil classification from CPT data has traditionally been based on various combinations of these parameters. The target variables selected were the soil constituents in terms of percent sand (grain size > 75 μm), percent silt-sized particles (75 μm ≥ grain size ≥ 5 μm), and percent clay-sized particles (grain size < 5μm). A total of 142 cases were compiled which then had to be divided into training and testing sets. All maximum and minimum target values were included in the training set, as the GRNN does not have the ability to extrapolate. Since soil types were fairly consistent at all the sites, it was decided to use certain sites exclusively for training and certain sites exclusively for testing, rather than take a percentage of all available data to accomplish these tasks. Thus, the data compiled from Yuanlin and Wufeng were used for training, and the data obtained from Dachun and Nantou were used for testing. The training set, therefore, consisted of 100 cases, while the testing set consisted of 42 cases which represent approximately 30% of the available data.

Prior to training the GRNN model, all input and output data were scaled so that they fell within the range of -1 to 1. Training then consisted of repeatedly presenting the data to the network to determine the optimal value of the smoothing parameter σ. To find the optimal σ, 10% of the data was removed from the training set for use in tuning the GRNN. The network was then trained with the remainder of the training data, and subsequently tested using this "tuning data." The soil percentages predicted by the GRNN model were compared to the known, or expected, values by determining the root mean square errors (RMSEs). The value of σ was then gradually changed until minimum RMSEs were achieved. Because this process resulted in different optimum σ values for different soil constituents (ranging from 0.18 to 0.29), a σ which gave favorable results for all constituents was chosen (0.22). After the optimum σ was found, the GRNN was tested with the testing data and the results compared to the reference soil composition values to determine the network's success.

3.3 *Results of the Soil Classification by the Various Methods*

The soil profiles based on the GRNN-predicted soil compositions are compared with the particle size distribution results (Juang 2002) for each of the four CPT soundings used for testing in Figures 2-5. Also shown in the figures are the probability of soil type (sandy, silty, or clayey) based on probabilistic region estimation (Zhang and Tumay 1999), the CPT-based soil behavior type classification (Robertson 1990), and the actual soil classification based on the USCS. The USCS soil types and Robertson's soil behavior types are given in Table 1. Because these alternative approaches to CPT soil classification (Robertson's method and Zhang and Tumay's technique) do not, unlike the GRNN model, estimate actual soil composition, it is not possible to quantitatively compare the predictions. And although the probabilistic region estimation results cannot be evaluated relative to the reference grain size distributions, the estimated soil profiles do have a similar appearance to the actual soil composition (as can be seen in Figures 2-5). It is also worth mentioning that the pore pressure data u_2 was not obtained/available for Chi-Chi earthquake database to calculate the corrected cone tip resistance q_t. Errors in classification may arise, however, for soft, fine-grained soils where q_c is typically small and u_2 can be quite large. These soil behavior types are located in the lower portion of the soil classification chart (Robertson 1990) where normalized cone resistance is less than about 10.

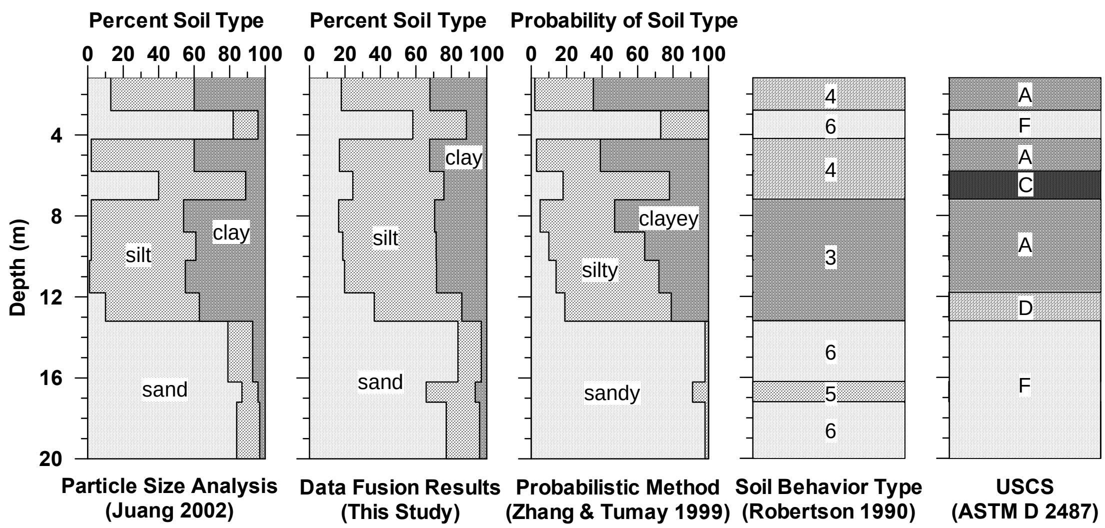


Figure 2 Comparison of classification methods at Dachun (sounding YL-C2)

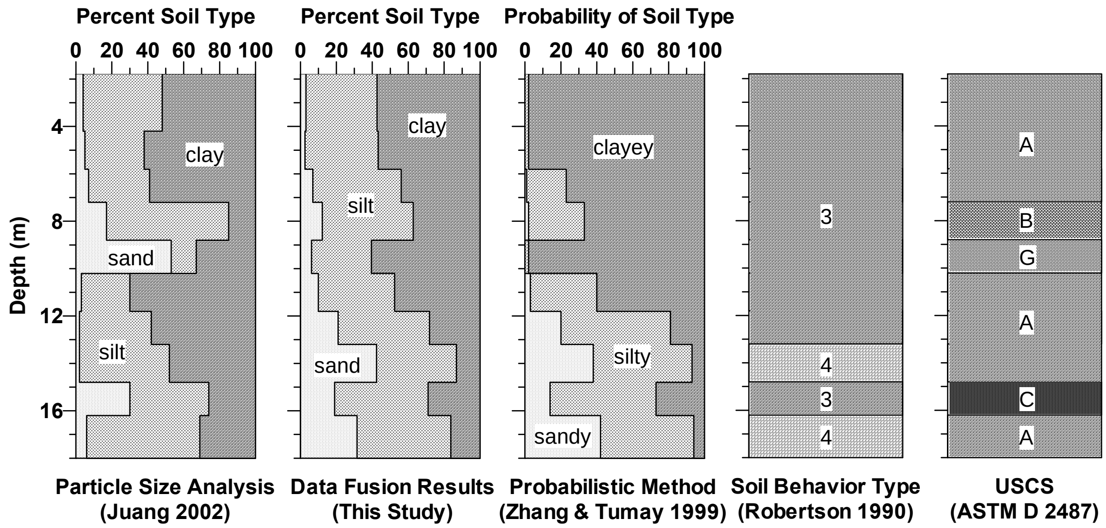


Figure 3 Comparison of classification methods at Nantou (sounding NT-C16)

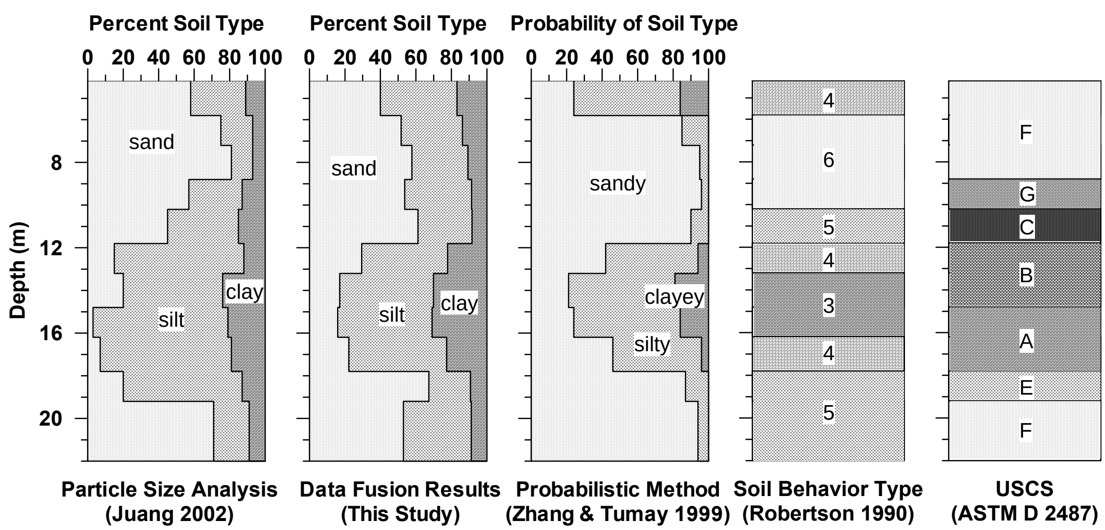


Figure 4 Comparison of classification methods at Nantou (sounding NT-C7)

4 DISCUSSION

Because soil classification is qualitative by nature, as the boundaries between soil types are set rather arbitrarily and are dependent on the classification system used, the prediction methods must be qualitatively, rather than quantitatively, evaluated. The profiles of soil composition predicted by the GRNN generally compare very well with the actual particle size distribution profiles. Overall, the neural network had an 86% success rate (defined as the ratio of correct predictions to the total number of testing cases) at classifying soils as either coarse-grained or fine-grained. The probabilistic region estimation method (Zhang and Tumay 1999) is effective in estimating the probability of soil type (sandy, silty, and clayey soils), with the probabilities actually comparing quite well with both the actual and the GRNN-predicted soil composition. Robertson's (1990) classification technique, which estimates soil behavior type rather than actual soil type based on the USCS, compares acceptably with the actual soil classification and is generally consistent with the results predicted by both the GRNN and the probabilistic region estimation method.

Site heterogeneity is expected to contribute to inconsistent (noisy) training data and may influence the performance of the soil classification methods (especially the GRNN model). Although the CPT and SPT locations may be adjacent to, or near, each other, there may be discrepancies between the CPT data recorded at a certain sounding depth and the composition (grain size distribution) of a sample collected from the same depth at a nearby borehole location. In other words, the soil sample retrieved from the borehole may be inconsistent with the CPT data obtained from the nearby sounding.

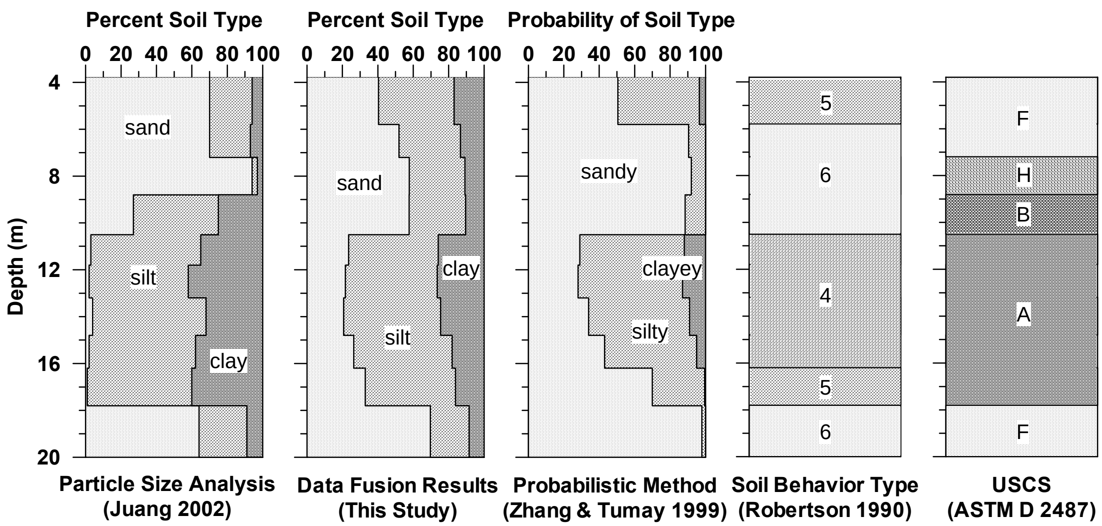


Figure 5 Comparison of classification methods at Nantou (sounding NT-C15)

Table 1. Soil Type and Soil Behavior Type Identifications

ID	Soil Type (USCS)	ID	Soil Behavior Type (Robertson 1990)
A	lean clay	1	sensitive fine-grained
B	lean clay with sand	2	Peats
C	sandy lean clay	3	clay to silty clay
D	silty clay	4	clayey silt to silty clay
E	silt with sand	5	silty sand to sandy silt
F	silty sand	6	clean sand to silty sand
G	clayey sand	7	gravelly sand to sand
H	poorly graded sand with silt	8	very stiff sand to clayey sand*
		9	very stiff fine-grained *

* Heavily overconsolidated or cemented

5 CONCLUSIONS

This study has compared three methods for soil classification and predicting soil composition, and therefore general soil type, from CPT data. The profiles of soil composition predicted by the GRNN model generally compare very well with the actual particle size distribution profiles. Overall, the neural network had an 86% success rate at classifying soils as coarse-grained or fine-grained. The probabilistic region estimation method (Zhang and Tumay 1999) is effective in estimating the probability of soil type and compares well with the results predicted by the GRNN. The soil behavior types estimated by Robertson's classification technique (1990) are also consistent with the predictions of the neural network and Zhang and Tumay's approach as well as the laboratory data.

6 ACKNOWLEDGEMENT

The authors appreciate the financial support of the U.S. National Science Foundation under Grant No. CMS-0409594. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors.

7 REFERENCES

Douglas, B. & Olsen, R.S. 1981. "Soil Classification Using Electric Cone Penetrometer, *Proc.Cone Penetration Testing and Experience,* ASCE, New York, NY, pp.209-227.

Juang, C.H. 2002. "Soil liquefaction in the 1999 Chi-Chi, Taiwan earthquake, in-situ test data." <http://www.ces.clemson.edu/chichi/TW-LIQ/In-situ-Test.htm> (Dec. 3, 2004).

Juang, C.H., Yang, H, Yuan, H., Lee, D.H. & Andrus, R.D. 2002. "CPT-Based Liquefaction Triggering Analysis with an Emphasis on the 1999 Chi-Chi, Taiwan, Earthquake." *Proc., NSF-Tubitak Turkey-Taiwan Earthquakes Grantee Workshop*, Antalya, Turkey, <http://ccee.oregonstate.edu/workshops/tubitak/reports.htm> (Jan. 12, 2005).

Kurup, P.U. & Griffin, E.P. 2006. "Prediction of Soil Composition from CPT Data Using the General Regression Neural Network," *ASCE Journal of Computing in Civil Engineering*, 20(4), pp. 281-289.

Robertson, P.K. 1990. "Soil classification using the cone penetration test." *Canadian Geotechnical Journal*, 27(1), 151-158.

Robertson, P.K., Campanella, R.G., Gillespie, D. & Greig, J. 1986. "Use of piezometer cone data." *Proc., Use of In Situ Tests in Geotechnical Engineering*, ASCE, New York, NY, 1263-1280.

Senneset, K. & Janbu, N. (1985). "Shear strength parameters obtained from static cone penetration tests." *Symposium on strength testing of marine sediments: laboratory and in-situ measurements*, ASTM, Philadelphia, PA, 41-54.

Specht, D.F. 1991. "A general regression neural network." *IEEE Transactions on Neural Networks*, 2(6), 568-576.

Tumay, M.T., Abufarsakh, M.Y., Zhang, Z., 2008, "From Theory to Implementation of a CPT-Based Probabilistic and Fuzzy Soil Classification," ASCE GSP No. 180, pp. 259-276.

Zhang, Z. & Tumay, M.T. 1999. "Statistical to fuzzy approach toward CPT soil classification." *Journal of Geotechnical and Geoenvironmental Engineering*, 125(3), 179-186. (Latest Version 4.0 Computer Code available at: http://www.ltrc.lsu.edu/downloads.html

Zhang, Z. & Tumay, M.T. 2000. Closure to "Statistical to fuzzy approach toward CPT soil classification." *Journal of Geotechnical and Geoenvironmental Engineering*, 126(6), 579-580.