Proceedings of the XVII ECSMGE-2019
Geotechnical Engineering foundation of the future
ISBN 978-9935-9436-1-3

doi: 10.32075/17ECSMGE-2019-0244

XVII European Conference
ECSMGE 2019
Reykjavik
on Soil Mechanics
and Geotechnical Engineering

Developing correlations between the soil fines content and CPT results using neural networks

Développer des corrélations entre la teneur en fines de sol et les résultats de CPT en utilisant des réseaux de neurones

M. S. Kovacevic
University of Zagreb, Faculty of Civil Engineering, Zagreb, Croatia

K.G. Gavin
Delft University of Technology, Faculty of Civil Engineering and Geosciences, Delft, Netherlands

C. Reale
Delft University of Technology, Faculty of Civil Engineering and Geosciences, Delft, Netherlands

L. Libric
University of Zagreb, Faculty of Civil Engineering, Zagreb, Croatia

D. Juric Kacunic
University of Zagreb, Faculty of Civil Engineering, Zagreb, Croatia

ABSTRACT: Knowledge of the fines content is necessary for all soil classification systems and an important factor in the evaluation of soil strength in liquefaction and seismic settlement analysis. This paper presents the application of cone penetration test, CPT data for estimating the soil fines content. The correlation can be used either as a first estimate of fines content (for example in the offshore environment) or to provide statistical information on the variation of fines content within a given area of interest (e.g. for a regional liquefaction study). The paper shows how field and laboratory test data were used with a neural network to correlate the CPT results and the fines content. Data from five site investigation locations across Northern Croatia were utilised. Verification of the approach is performed using field and lab test data from the Veliki vrh landslide.

RÉSUMÉ: La connaissance de la teneur en fines est nécessaire pour tous les systèmes de classification des sols et constitue un facteur important dans l'évaluation de la résistance des sols lors de l'analyse de la liquéfaction et du tassement sismique. Cet article présente l'application du test de pénétration au cône et des données CPT pour l'estimation de la teneur en particules fines du sol. La corrélation peut être utilisée soit comme première estimation de la teneur en fines (par exemple dans l'environnement offshore), soit pour fournir des informations statistiques sur la variation de la teneur en fines dans une zone d'intérêt donnée (par exemple, pour une étude régionale sur la liquéfaction). Le document montre comment les données de tests sur le terrain et en laboratoire ont été utilisées avec un réseau de neurones pour corréler les résultats du CPT et le contenu en fines. Les données provenant de cinq sites de recherche sur le nord de la Croatie ont été utilisées. La vérification de l'approche est effectuée à l'aide des données de test de terrain et de laboratoire du glissement de terrain Veliki vrh..

Keywords: CPT; fines content; correlation; neural network

1 INTRODUCTION

The Cone Penetration Test, CPT is a simple, fast and cost-effective in-situ test that provides continuous data over the depth of penetration of an instrumented cone into the subsurface. The CPT penetrates at a constant rate, with a continuous measurement of the cone resistance at the cone head (q_c), sleeve friction (f_s) and the pore pressure (u_2) which represents a sum of the in-situ equilibrium pore pressure (u_0) and the excess pore pressure (Δu). Using the three measured parameters (q_c, f_s and u_2), procedures have been established for determining the soil profile, soil identification and classification and the determination of mechanical, flow and consolidation characteristics of the soil. Many correlations (Robertson, 2009; Mayne, 2014; Librić et al., 2017, Kovačević et al., 2018) have been developed over recent years indirectly relating CPTs to various geotechnical parameters.

This paper investigates the use of both statistical regression and a machine learning technique, artificial neural networks (ANN), for developing CPT based correlations between the soil beviour type index (I_c) and fines content (FC) as a percentage of fine particles in the soil. Knowledge of the fines content is necessary for all soil classification systems and an important factor in the evaluation of soil strength in liquefaction and seismic settlement analysis.

The soil behviour index I_c, is determined directly from CPT measurements as:

$$I_c = \sqrt{(3.47 - logQ_{tn})^2 + (logF_r + 1.22)^2} \qquad (1)$$

where Q_{tn} and F_r are normalised cone resistance and normalised friction ratio, calculated as:

$$Q_{tn} = \frac{q_t - \sigma_{v0}}{p_a} \cdot \left(\frac{p_a}{\sigma'_{v0}}\right)^n \qquad (2)$$

$$F_r = \frac{f_s}{q_t - \sigma_{v0}} \cdot 100\% \qquad (3)$$

where: σ_{v0} is total vertical stress in the ground, σ'_{v0}' is effective vertical stress in the ground, p_a is atmospheric pressure (100 kPa) and n is stress exponent dependent on soil type and stress level, with possible values between 0 and 1 and calculated as:

$$n = 0.381 \cdot I_c + 0.05 \cdot \left(\frac{\sigma'_{v0}}{p_a}\right) - 0.15 \qquad (4)$$

The soil behaviour type index I_c as defined in Equation (1) represents a series of radii of concentric circles in the soil classification charts, which present soil types according to the correlation of normalised cone resistance Q_{tn} and normalised friction ratio F_r. Contours of identical I_c in the Q_{tn} - F_r chart represent limits between different soil types. Jefferies and Davis (1993) suggest using I_c to modify the empirical correlations that vary depending on soil type. Robertson (2009) points that this is an extremely powerful concept, proposing that I_c be used for creating statistical correlations whenever possible.

One of the advantages of using the behaviour index I_c is that it is not highly sensitive sleeve friction f_s measurements that tend to high higher variability than q_c. , Rather its value is largely dependent on the value of the corrected tip resistance q_t, which has a significantly higher measuring precision. It can be shown that changing the sleeve friction f_s by ±50% in general results in a change in the soil behaviour type index I_c by less than ±10%. For soft soils belonging to the bottom part of the Q_{tn} - F_r chart, I_c is practically insensitive to changes in f_s (Robertson, 2009).

The research sites considered in this paper are geographically distributed over the area of northern Croatia. A summary database of 216 pairs of laboratory testing and CPT results was created. Verification of new correlations and developed neural network using the database for northern Croatia was carried out on the example of Veliki vrh landslide located in the same region and which was not used in the initial development of the models.

2 EXISTING CORRELATIONS

Several correlation exist between soil behaviour type index (I_c) and percentage of fines content (FC). In order to evaluate the soil liquefaction potential Robertson and Wride (1998) proposed the following correlation:

$$I_c \leq 1.26,\ FC = 0 \tag{5a}$$

$$1.26 \leq I_c \leq 3.50,\ FC = 1.75 \cdot I_c^{3.25} + 3.7 \tag{5b}$$

$$I_c > 3.50,\ FC = 100 \tag{5c}$$

$$1.64 \leq I_c \leq 2.36 \text{ and } F_r < 0.5,\ FC = 5 \tag{5d}$$

For the same purpose Idriss & Boulanger (2008) proposed the correlation:

$$FC = 2.80 \cdot I_c^{2.60} \tag{6}$$

They concluded that general correlations between I_c and FC developed across a broad range of sites and geologic settings are poor and have large scatter. They suggest that this variability can be greatly reduced by collecting, and calibrating against, site-specific data.

In order to develop the correlation beetwen I_c and FC for site-specific data, Yi (2014) used 124 samples of laboratory measured fines contents from a total of 11 sites located near the southern edge of the San Bernardino Valley in Southern California. All of these sites geologically consist of very young to young sandy, late Holocene age alluvial deposits with low plasticity. They suggested the following relationship:

$$I_c \leq 1.31,\ FC = 0 \tag{7a}$$

$$1.31 \leq I_c < 2.50,$$
$$FC = 42.0 \cdot I_c - 55 + 10 \cdot sin\left(\left(\frac{I_c - 2.5}{1.19}\right) \cdot \pi\right) \tag{7b}$$

$$2.50 \leq I_c \leq 3.10,\ FC = 83.3 \cdot I_c - 158.3 \tag{7c}$$

$$I_c > 3.10,\ FC = 100 \tag{7d}$$

$$1.31 \leq I_c \leq 2.36 \text{ and } F_r < 0.6,\ FC = 5F_r \tag{7e}$$

For the purpose of developing probabilistic CPT based soil classification models, Cetin and Ozan (2009) used 484 pairs of CPT / laboratory results from seven different databases located around the world. They develop the following correlation:

$$FC = \frac{R - 238.50}{1.75} \cdot 100 \tag{8}$$

where R is a parameter similar to I_c:

$$R = \sqrt{\left(log q_{t,1,net} - 233.52\right)^2 + (log F_r + 55.42)^2} \tag{9}$$

where $q_{t,1,net}$ is the normalized net cone tip resistance and is defined as:

$$q_{t,1,net} = \frac{q_t - \sigma_{v0}}{\left(\frac{\sigma'_{v0}}{p_a}\right)^c} \tag{10}$$

and c is stress exponent dependent on soil type and stress level, with possible values between 0.25 and 1 and calculated as:

$$c = \frac{R - 272.38}{2.81} \tag{11}$$

3 ARTIFICIAL NEURAL NETWORKS

Artificial neural networks are a form of computational intelligence (Rosenblatt, 1958) developed to mimic how the human brain interprets information and solves problems. Interconnected neural elements share information in order to establish how different variables within a system interact, in order to emulate its behaviour. As new information becomes available the system is able to reinterpret its learned behaviour and update as appropriate. Neural networks are commonly used for regression, classification and prediction tasks.

Every neuron is connected to every other neuron and each connection receives a weighting. These weightings control how sensitive the system response is to a variable.

When there are a number of known sets of inputs and outputs these weightings can be optimised to map the system inputs onto the system outputs.

Technically any function can be used for this process however if backpropagation is going to be used the function should be continuously differentiable. This study used the Bayesian Regularisation backpropagation algorithm for training. This is known as neural network training and needs to be carried out before the neural network can be used for regression, classification or prediction. Training continues until all data has been exhausted or predictions match outputs within a certain preordained tolerance.

Typically neural networks are organised into an input layer – where the inputs are feed into the system, a hidden layer(s) – where the weightings between the different parameters are generated and an output layer where the system output is generated, see Figure 1. While the number of input and output nodes is dictated by the underlying engineering problem in question. The number of hidden neurons needed is much more subjective and should be specifically investigated for a given problem. This study uses three hidden layers.

If there are more hidden neurons than appropriate then the system will be slow to converge and will risk being overtrained, while if the converse is true the network will be too general to consistently deal with unseen datasets.

This study utilises a multilayer feed-forward neural network in conjunction with a sigmoid activation function for hidden neurons and a linear activation function for output neurons. A feed-forward neural network, means that information isn‘t recursive instead it can only move in one direction from input to hidden to output.

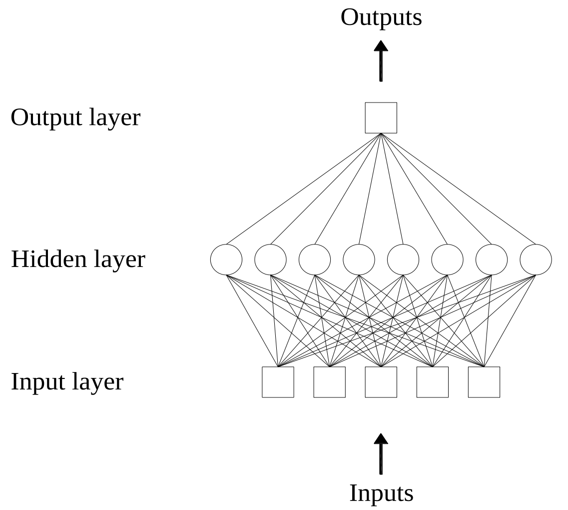

Figure 1. General schematic of a feed-forward artificial neural network

When training is completed the ANN needs to be validated to ensure the system is performing as expected, this should be carried out with a new dataset the ANN has not previously been exposed to during training. During validation the model is only given access to inputs. If the ANN behaves as expected and predicts the outputs correctly following this, then it can be said to model the system accurately. Provided enough input and output data has been provided during training, an ANN model should be able to determine the significance each individual parameter has on the outcome.

The ANN described in this study takes normalised cone resistance Q_{tn} and normalised friction ratio F_r as inputs, and uses them to predict fines content FC as an output.

4 DESCRIPTION OF TEST SITES

Data from five test sites located in Northern Croatia were used to train, validate and test the model. In total 216 pairs of CPT/ Laboratory test pairs were collated from the test sites. Data from a sixth site Veliki vrh was used as a validation dataset. A short overview of each test site is given below.

4.1 Biđ-Bosut Irrigation canal

Construction of the 14 km long irrigation canal in Biđ-Bosut Field, is the 1st phase in the construction of the multi-purpose Danube-Sava canal. The geotechnical investigation performed between chainage km 0+600 to km 4+800 included: 15 No. co-located boreholes and CPT tests, Each probe was taken to a depth of 12 m deep, with an average distance between investigation points of 300m. At each location soil samples were taken for classification testing and 75 pairs of laboratory testing and CPT results were available .

4.2 Ilok port

The town of Ilok is located on the Rhine-Main-Danuber river system, which connects the North and Black Seas. The work described herein was performed as part of the Danube-Sava canal project that would connect the Danube and the Adriatic regions. The geotechnical investigation comprised a total of 9 boreholes, made, with continuous coring to the maximum depth of 30 m, dynamic (SPT) and static (CPTU) testing, geophysical testing using seismic refraction, multichannel analysis of surface waves (MASW), seismic static cone penetration test (SCPT), together with laboratory tests. The database consists of 36 pairs of laboratory testing and CPT results.

4.3 Krsišće landslide

Krsišće landslide is located on the southern slopes of the Medvednica mountain, in the Markuševec area, at an altitude of approximately 300 meters. The investigation described herein relates to a potentially unstable section located to the west of, the Krišišće road,. A total of 5 boreholes were made, with continuous coring to the maximum depth of 8 m, dynamic (SPT) and static (CPTU) testing, together with laboratory tests. The database consists of 20 pairs of laboratory testing and CPT results.

4.4 Mirogoj landslide

The Mirogoj landslide is located on the southern slopes of the Medvednica mountain, at the Mirogoj cemetery. The incline of the part of the slope where the landslide initiated is between 20° and 25°. As part of the conducted geotechnical investigation work, a total of 5 boreholes were made, with continuous coring to the maximum depth of 8 m, dynamic (SPT) and static (CPTU) testing, together with laboratory tests. The database consists of 25 pairs of laboratory testing and CPT results.

4.5 Krematorij landslide

The Krematorij landslide is located east of the Kameniti stol street, in the Gornji grad - Medveščak area, on the southern, more indented slopes of the Medvednica mountain. The incline of the part of the slope where instabilities or landslides have been identified is between 10° and 30°. As part of the conducted geotechnical investigation work, a total of 5 boreholes were made, with continuous coring to the maximum depth of 12 m, dynamic (SPT) and static (CPTU) testing, together with laboratory tests. The database consists of 60 pairs of laboratory testing and CPT results.

4.6 Verification site: Veliki vrh landslide

The Veliki vrh landslide is located on southern slopes of Medvednica mountain, in the area between Čućerje and Vugrovec, at an altitude between 205 and 225 metres. To the southeast of the Veliki vrh street (house no. 242), an unstable slope was noticed some time ago. Newer research has shown that a new part of the unstable slope had appeared next to and below house no. 242 of the Veliki vrh street. As part of the conducted research work, a total of 4 boreholes were made, with continuous coring to the maximum depth of 12 m, dynamic (SPT) and static (CPTU) testing, together with laboratory tests. The database consists of 19 pairs of laboratory testing and CPT results.

5 NEW CORRELATION

Using the data obtained from the five test sites, this paper proposes a model similar in formulation to that proposed by Robertson and Wride (1998). The best fit relationship found in this study is compared to Robertson and Wride's relationship in Figure 2. Both equations have very similar regression values when applied to the dataset, however, as can be seen from Figure 2 Robertson and Wride's equation significantly underpredicts fines content magnitude but captures the relative increase reasonably well. The relationship from this paper effectively increases the magnitude of Robertson and Wrides relationship to more closely approximate reality.

$$1.40 \leq I_c \leq 3.42,\ FC = 17.45 \cdot I_c^{1.662} - 35.42 \quad (12)$$

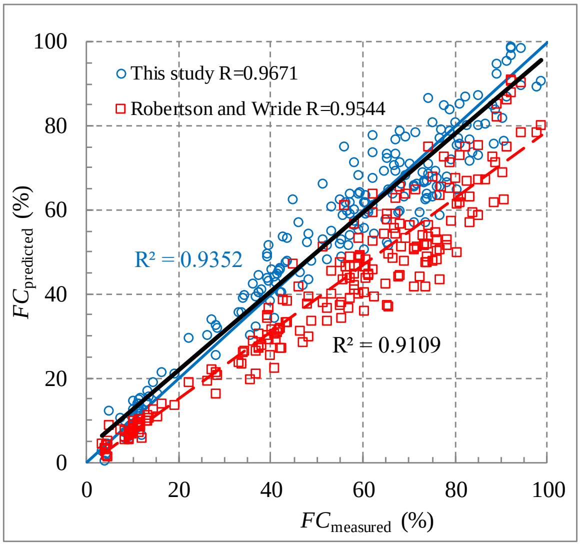

Figure 2. The statistical correlation developed in this study, with Robertson and Wride (1998) for comparison

6 ANN RESULTS AND DISCUSSION

The model development dataset which comprised of cone resistance Q_{tn} and normalised friction ratio F_r as inputs and fines content FC as an output was split randomly into the following proportions 70% for training, 15% for testing, and 15% for validation. For training, the ANN had access to both inputs and outputs allowing it to learn the sensitivity of each variable and understand each parameters effect on the system response. The next 15% was used as a test set, during the testing process only the inputs were supplied to the model. At the end of the testing phase, the neural network performed a system recalibration on itself so that system inputs could be more accurately mapped onto system outputs based on the test results. Following completion of the testing phase the final 15%, or the validation set, was sent to the neural network. Only inputs are sent in the validation phase, thus allowing the direct comparison of outputs from the validation set to actual measured values. Provided a good correlation has been achieved the neural weightings are saved and the entire data set is subsequently inputted blind. The resultant outputs are compared to actual outputs, see Figure 3.

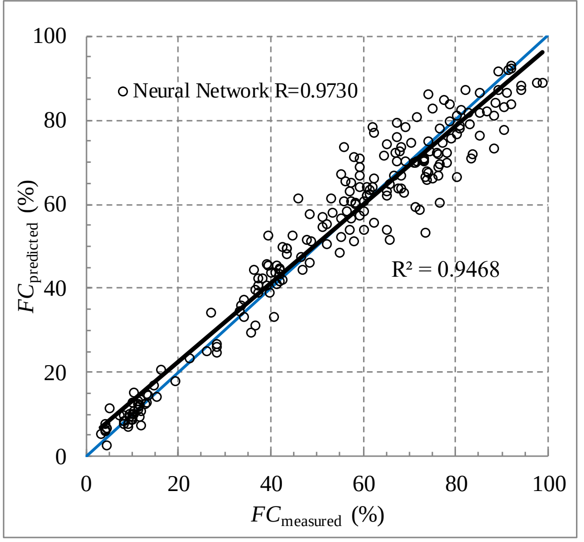

Figure 3. ANN predicted fines content five test sites used in model development

A regression coefficient of 0.9468 was achieved for the entire dataset, with a correlation coefficient of 0.973. As can be seen from Figure 3, there is very little data scatter, and importantly no extreme outliers. Therefore while a misclassification could occur, an extreme difference between predicted fines content and measured fines content is unlikely.

To ensure the model was working correctly input data from an additional site within the same geographic region, Veliki vrh was supplied to the model. This data which can be seen in Table 1, consisted of 19 pairs of CPT and laboratory unit weight and fines content results.

Table 1. Fines content, unit weight and CPT results from Veliki vrh

No	*z* [m]	q_t [MPa]	f_s [kPa]	γ_t [kN/m^3]	*FC* [%]
1	2.20	1.63	116.00	19.05	65.72
2	2.80	0.74	62.00	18.20	84.15
3	3.30	0.60	38.00	17.61	83.25
4	5.00	3.38	172.00	19.72	59.72
5	2.10	1.55	93.00	18.84	61.83
6	2.80	0.79	72.00	18.72	85.16
7	3.60	0.62	55.00	18.52	89.25
8	4.40	1.29	69.00	18.76	79.22
9	5.60	1.12	90.00	18.73	90.25
10	6.60	1.51	101.00	18.90	83.88
11	7.20	1.58	96.00	18.90	81.00
12	1.60	0.81	54.00	17.43	68.12
13	2.20	0.62	26.00	17.18	71.45
14	3.80	4.69	188.00	20.46	46.02
15	5.80	2.24	104.00	18.62	65.58
16	1.80	2.04	122.00	19.11	49.32
17	2.20	1.99	84.00	18.79	48.25
18	3.10	4.99	193.00	19.27	44.15
19	3.70	6.59	203.00	19.99	37.26

An extremely good R^2 of 0.9732 was obtained for this external verification with a correlation coefficient of 0.9865. The predicted fines content versus measured fines content is shown in Figure 4. The statistical approach proposed earlier in Equation 14 performed equally well on the unseen dataset, Veliki vrh, achieving an R^2 of 0.9765. Both are shown in Figure 4, giving very similar results.

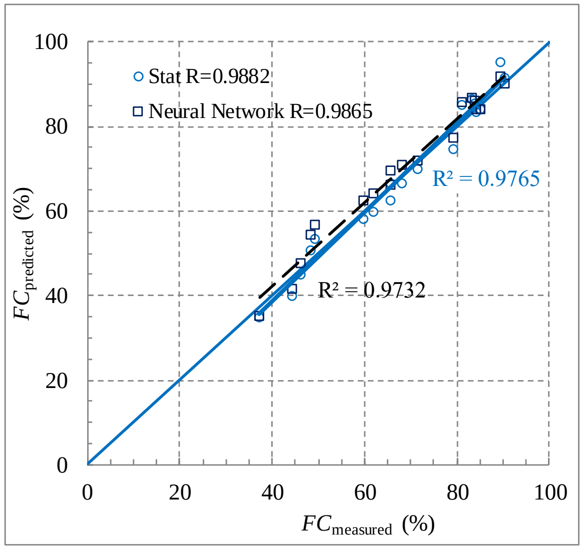

Figure 4. Predicted fines content for Veliki Vri using closed loop ANN and regression approach

7 CONCLUSION

This paper presents two approaches, regression and neural network, for automatically calculating fines content using CPT measurements as inputs. Both approaches could easily be performed automatically onsite as the CPT is ongoing, thus allowing for an extremely fast interpretation of fines content. This would reduce the quantity of laboratory tests needed per site thus saving time and money. An additional benefit of such an approach is that any laboratory tests that are carried out can then combined with their respective CPT soundings become additional data entries for both the regression and ANN models, thus improving their future accuracy. In this way, the models can continue to evolve over time, gradually increasing in both accuracy and precision.

The approaches were developed using 216 pairs of CPT/laboratory fines content tests from five different locations across Northern Croatia. An entirely separate sixth site Veliki vri was used as an external verification measure for the saved neural networks. The models performed extremely well on both the initial dataset and the subsequent verification dataset.

Unfortunately, ANN-based models have some drawbacks, of particular concern is the black box nature of the results, which makes proof of concept hard to verify, while also making their standalone implementation a risky process for the engineer involved. The authors think that much of this can be mitigated by testing a small number of samples from every site in the laboratory for local verification. Thus allowing the training database to continue to grow in size over time making incorrect classifications less likely to occur. Over time reducing the cost, time, and labour involved.

This study confirms the functional link between CPT results, and soil fines content. The developed neural network and regression models performed admirably for a wide range of soil types closely predicting fines content between 3 and 99 %. The close prediction between the neural networks and the regression model is a testament to the accuracy of power regression models for predicting soil unit weights and further validates their use in everyday design situations, given their simplicity and transparency.

8 ACKNOWLEDGEMENTS

The authors acknowledge funding obtained from the EU H2020 project GoSAFE RAIL (Grant No: 730817).

9 REFERENCES

Author, A.N. 2004. Article title, *Journal Title* **20**, 1012–1030.

Cetin, K.O., Ozan, C., 2009. CPT - Based Probabilistic Soil Characterization and Classification, *Journal of Geotechnical and Geoenvironmental Engineering*, **135**, 84–107.

Idriss, I. M., Boulanger, R. W. 2008. *Soil Liquefaction During Earthquake*, Earthquake Engineering Research Institute, EERI Publication MNO-12.

Jefferies, M.G., Davies, M.P. 1993. Use of CPTU to estimate equivalent SPT N_{60}, *Geotechnical Testing Journal*, ASTM, **16**, 458–468.

Kovačević, M.S., Gavin, G.K., Reale, C., Librić, L. 2018. The use of neural networks to develop CPT correlations for soils in northern Croatia, *4rd International Symposium on Cone Penetration Testing (CPT`18)*, 387–382.

Librić, L., Jurić-Kaćunić, D., Kovačević, M.S. 2017. Application of cone penetration test (CPT) results for soil classification, *Građevinar*, **69**, 11–20.

Mayne, P. 2014. Interpretation of geotechnical parameters from seismic piezocone tests, *3rd International Symposium on Cone Penetration Testing (CPT`14)*, 47–73.

Robertson, P. 2009. Interpretation of cone penetration tests—a unified approach, *Canadian Geotechnical Journal*, **46**, 1337–1355.

Robertson, P.K., Wride, C.E. 1998. Evaluating cyclic liquefaction potential using the cone penetration test, *Canadian Geotechnical Journal*, **35**, 442–459.

Rosenblatt, F. 1958. The perceptron: A probabilistic model for information storage and organization in the brain, *Psychological review*, **65**, p. 386.

Yi, F. 2014. Estimating soil fines contents from CPT data, *3rd International Symposium on Cone Penetration Testing (CPT`14)*, 949–955.