Usage of the graph clustering algorithm to the recognition of geotechnical layers

SIMON RABARIJOELY[1], PIOTR BILSKI[2], TOMASZ FALKOWSKI[1]
[1]Department of Geotechnical Engineering, Warsaw University of Life Sciences – SGGW
[2]Department of Informatics, Warsaw University of Life Sciences – SGGW

Abstract: *Usage of the graph clustering algorithm to the recognition of geotechnical layers.* The aim of the paper is to present the approach to the application of the graph clustering algorithm to the recognition of geotechnical layers from the dilatometer tests. Results of the measurements obtained from the DMT test in the test site (subsoil of one of the buildings in the Warsaw University of Life Sciences campus) were analyzed by the clustering algorithm which was able to extract the separate groups of the measurements, representing identical soil type. This method is parameterized, so its verification by the geotechnical experts was necessary to determine the optimal parameter values. They lead to the determination of the soil types as close to the actual situation, as possible. Also, the output of the algorithm was analyzed by the geotechnical experts to identify and label the extracted soil types.

Key words: soil layer, artificial intelligence methods, DMT test.

INTRODUCTION

A basic activity in the geotechnical engineering design is a preparation of the documentation concerning the soil profiles of the analyzed area. The standard dilatometer (DMT) test is one of the most widely used method in the in-situ investigation. The soil type in DMT measurements is determined using several statistical correlations based on practical experience. As the powerful and handful computational methods emerge, their application in the geotechnical research may improve and facilitate soil investigation procedure. One of the most often proposed methods for the task are the artificial neural networks (ANN). They are now used to predict and categorize the soil types using information extracted from the DMT measurements. Usage of the ANN is advised for different geotechnical problems. Toll (1996) and Shahin et al. (2001) presented the state-of the art report on the different applications (liquefaction prediction, soil classification, compaction, to the bearing capacity, settlement analysis etc.) of ANN in geotechnical engineering. ANNs were also used for site characterization, based on SPT results (Itani and Najjar, 2000; Das and Basudhar, 2004) and CPT (Juang et al., 2001).

Although the most popular, ANNs are only one of the possible methods that can be used in geotechnical investigations.

The first task during the soil investigation using DMT tests is determination of the different soil categories in the analyzed profile. To find out, how many soil types exist in the test site, and how can they be distinguished, an unsupervised learning method must be used. Besides ANNs there are methods, which should be tested in such a task. One of them is the graph clustering algorithm. It allows

to group measurements taken during the DMT tests. Every group would represent one soil type, which must be identified by the geotechnical expert. Graph clustering algorithm has a simpler structure than the ANN and also works much faster, therefore it was selected for the experiment. The interpretation and DMT test results, are presented in the geological conditions of the Warsaw area, the description of the graph clustering algorithm and example of its application are presented in the paper.

INTERPRETATION OF DILATOMETER TEST AND TEST RESULTS

The Marchetti dilatometer (Fig. 1) is a pressure measuring equipment; it is terminated with the flat steel blade and elastic membrane attached to it. The dilatometer test is based on the subsoil penetration by the flat blade. In the selected depths the measurements of the working gas pressure on the probe's membrane are undertaken. During the research the two parameter measurements are possible: pressure A (taken at the first contact of the soil with the membrane) and pressure B (taken after pushing the membrane with the air transported by the internal pneumatic tubing – Fig. 1). They are used to calculate further two corrected pressures: p_0, p_1 (Briaud 1991, Lutenegger 1988, Marchetti 1980) as follous:

$$p_0 = 1.05 \, (A - Z_M + \Delta A) + \\ - 0.05 \, (B - Z_M - \Delta B) \qquad (1)$$

$$p_1 = B - Z_M - \Delta B \qquad (2)$$

where ΔA, ΔB are the corrections determined by membrane calibration and Z_M the gage zero offset (gage reading when vented to atmospheric pressure). In the graph clustering algorithm, only the pressures A and B are analyzed.

The first parameter, p_0 is the gas pressure received in the first phase of the movement of the membrane, as a result of the contact with the surrounding soil. The second parameter – p_1, is the corrected value of the gas pressure obtained at the additional inclination of the membrane centre towards the soil at about 1 mm. All these pressures with the effective vertical stress σ'_{vo} and water pores pressure u_0 are used to estimate the dilatometer indexes (Marchetti, 1980) – see equations (3) to (5).

The test site in the Warsaw University of Life Sciences campus was located at the subsoil of one of the buildings. The measured parameters are the pressures p_0 and p_1, which, after conditioning, are registered by the device as the pressures A and B, respectively (Fig. 2). Based on these measurements, the soil indexes (3), (4), (5) are calculated and presented in Figure 3.

DESCRIPTION OF THE TEST SITE

The Warsaw University of Life Sciences campus in Ursynów is founded on the upland of morainal origin. To the east of the investigated area upland drops with the steep slope to the higher (one of the three Pleistocene terrace), so--called Praga terrace of the Vistula valley (Sarnacka, 1976). The basis of the valley mostly consists of the Quaternary sediments as well as Tertiary, lacustrine Pliocene 'motley clay'. Those sediments
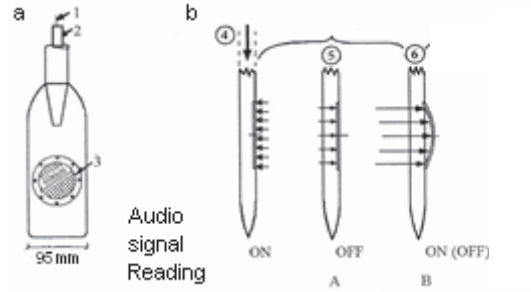
Material index $I_D = \dfrac{p_1 - p_o}{p_o - u_o}$ (3)

Horizontal earth pressure index $K_D = \dfrac{p_o - u_o}{\sigma'_{vo}}$ (4)

FIGURE 1. Marchetti dilatometer: (a) flat blade: 1 – electric wire, 2 – pneumatic tubing, 3 – steel membrane, (b) test stages: 4 – pushing, 5 – pressure p_o, 6 – pressure p_1,

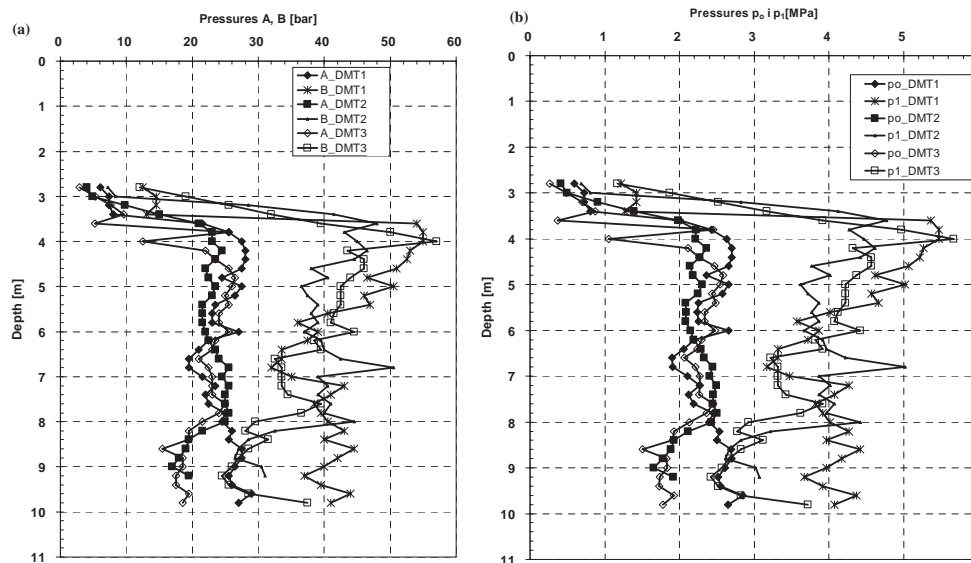Dilatometer modulus $E_D = 34.7 \cdot (p_1 - p_o)$ (5)

FIGURE 2. Profiles of pressures from dilatometer test in the subsoil of Warsaw University of Life Sciences campus building – values of the (a) and A, B (b) p_0, p_1 (Geotechnical documentation... 2001)

form in the Vistula valley a glaciotectonic culmination. According to Różycki and Sujkowski (1936), in the escapement zone of the upland the Quaternary basement surface forms glaciotectonically uplifted protrusion. Pliocene 'motley clays' crop out in the escarpment northward the analyzed area.

Above the Pliocene deposits there are preglacial (Eopleistocene) alluvial deposits consist of quartz gravels, sands and silts with lydites.
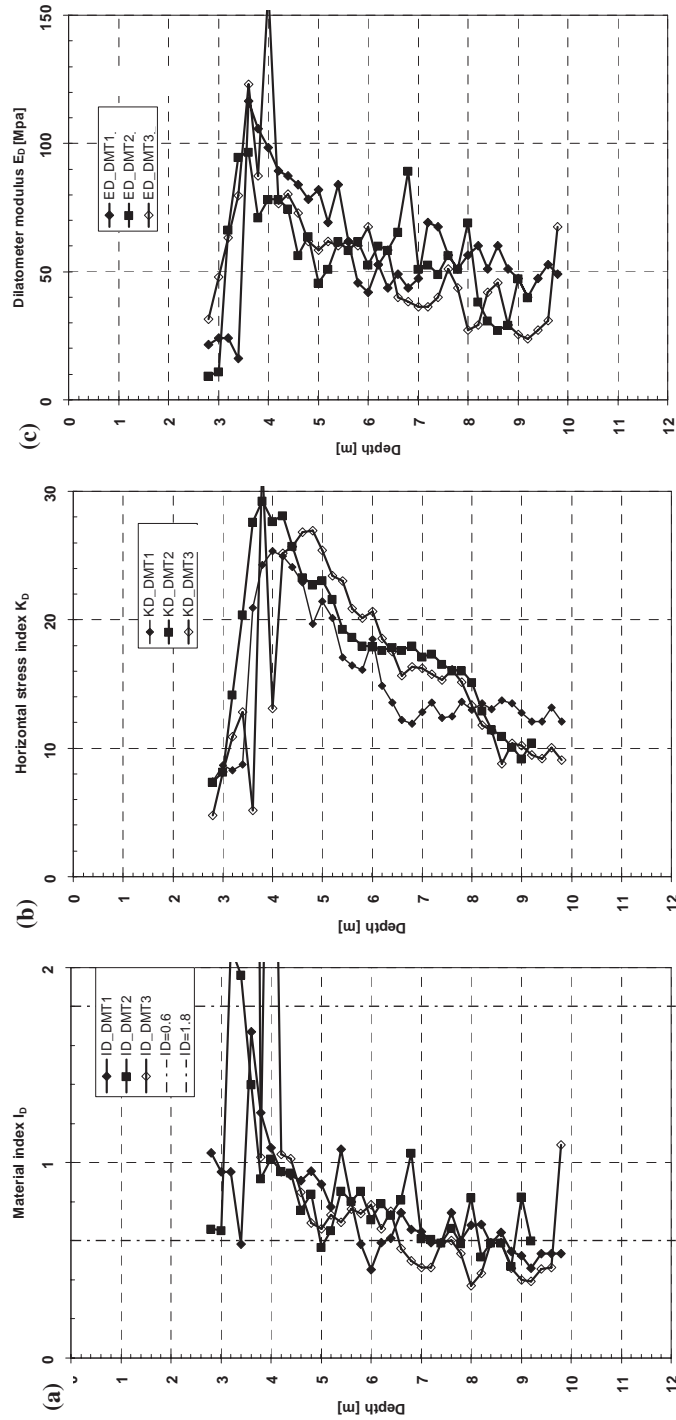
FIGURE 3. Profiles of indexes from dilatometer test in the subsoil of Warsaw University of Life Sciences campus building – values of I_D (a), K_D (b), and E_D (c) (Geotechnical documentation... 2001)

The oldest glacial formations and the subsequent Cromerian Interglacial (Sarnacka, 1992) are not present in the profile of the Quaternary sediments in the Ursynów area. In the Warsaw region they appear among the severe depressions of the top of the Pliocene formation (Sarnacka, 1992; Morawski, 1980). In the Ursynów upland, above the preglacial deposits, or directly on the lacustrine Pliocene clays there is a series of the South-Polish Glaciation – boulder clays, glaciofluvial sediments and ice dam deposits, which form non-contiguous layers. The lack of continuity is of river erosion origin. Over the sediments of the South-Polish Glaciacion exists the 15 to 25 meters thick formation of the river-originated sediments from the Mazovian Interglacial. It consists mostly of sands and gravels.

The top of these sediments is at the depth from 10 to 20 meters below the ground level (Sarnacka, 1992). In the Ursynów region such compositions form the continuous, distinct layer, consisting mostly of fine sands.

Above the interglacial sediments in the Warsaw area there are series of the Middle-Polish Glaciacions. The deepest layer of these formation – embedded loams and silts was deposited in the ice dam lake so-called "basin of Pilica" (Różycki, 1972). On the top of the glacio-lacustrine deposits there are fluvioglacial sands and gravels testify the transgression of the Odra glacier (the maximal in extend of the Middle Polish Glaciations). Those deposits crops out in the escarpment zone of the Służewski Stream valley, in the vicinity of the analyses area (Sarnacka, 1992). Above the fluvioglacial sediments continuous

layer of the melt-out formations from the Odra glaciation – grey boulder clay is being found. The thickness of the clay layer (gQpO, Fig. 4, Fig. 5) is
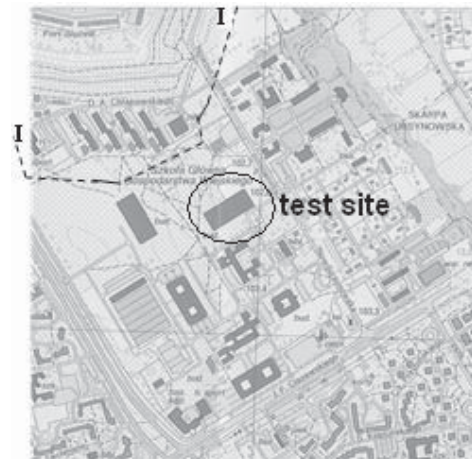

FIGURE 4. Situation plan of the WULS – SGGW campus

from several to tens meters (Sarnacka 1992). In some places Odra boulder clay is covered by fluvioglacial sands and gravel or/and ice dam silts. Above them or directly on the boulder clay (as is in the case of the Warsaw University of Life Sciencescampus) the melt-out till from the Warta glaciation (the last glaciation on this area) is being found (gQpW, Fig. 4, Fig. 5). It has the form of brown sandy boulder clay with thickness of about 2 to 5 meters. The lower part of this layer is presumably a melted-out sediment (*melt-out till*, Boulton, 1970). Above it there is non-contiguous complex of sediments representing presumably the formations of *flow till* (as in Boulton 1970): sandy clays, containing thin layers of sands and silts (gQpW). The out-crops of the Warta flow-tills as well as fluvioglacial sands and glaciolacustrine silts flows on the
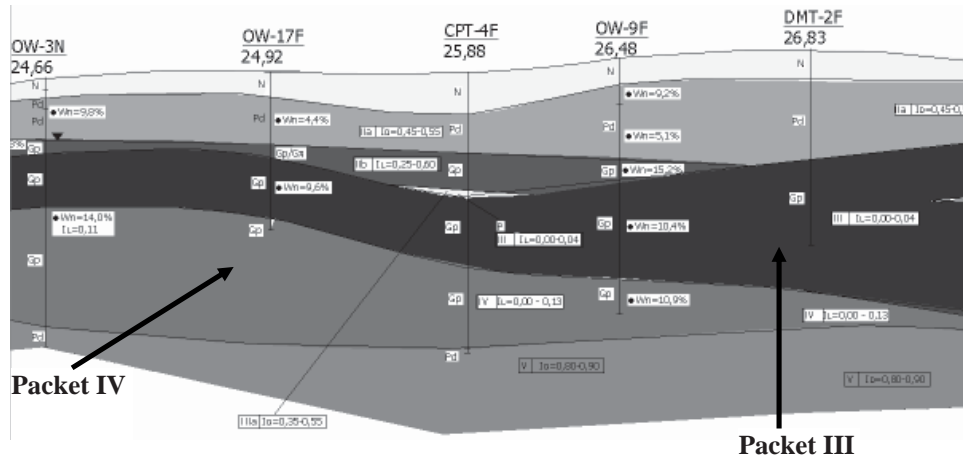
FIGURE 5. Geotechnical conditions of the basis of the WULS – SGGW campus – cross section I-I

surface of the Ursynów upland testify the areal kind of deglaciation.

Glacial sediments in the surface layer were in strongly transformed antropogenically many places. There are embankments of the various thickness, consisting largely of sandy clays, clayey sands, silt sands and silts with the different admixture of organic matters.

The most important water-bearing levels of the investigated area are Mazovian interglacial river sands. Coefficients of permeability in this layer (defined in the basis of the probationary pumping are about $1 \cdot 10^{-4}$ m/s (Wiencław et al., 1996). The general direction of the underground waters flow is from the southeast to the northwest. The hydraulic gradient of water table in this layer is from $1 \cdot 10^{-3}$ to $2 \cdot 10^{-3}$ (Wiencław et al., 1996).

Considering the physical and mechanical ground properties in the campus test site, there are five geological packets. The first packet consists of the fluvioglacial sediments of the Warta glaciation (fgQpW): medium-grained and fine sands, with $I_D = (0.35 \div 0.55)$ and clayey sands with sandy clays and silts for which $I_L = (0.15 \div 0.20)$. The second pack consists of the ice-dam sediment of the Warta glaciation (bQpW): medium grained and fine sands with medium density index $I_D = (0.3 \div 0.5)$ and sandy clays and clayey sands for which $I_L = (0.0 \div 0.2)$ or $I_L = (0.25 \div 0.54)$. The third pack forms brown glacial boulder clay of the Warta glaciation (gQpW), sandy clays for which $I_L = (0 \div 0.11)$. In the fourth pack there are grey glacial boulder clay of the Odra glaciation (gQpO), sandy clays with boulders, for which $I_L = (0 \div 0.12)$. The last pack determines the river deposits of the Mazovian Interglacial (fQpM): small and medium-grained sands, in the ceiling of the very condensed layer with $I_D = (0.8 \div 0.9)$.

APPLICATION OF THE CLUSTERING METHOD TO THE RECOGNITION OF GEOTECHNICAL LAYERS

The DMT measurements were used to perform categorization of the soil

types. The measurement data formed a learning data set for the graph clustering algorithm (Bilski, 2007), which is one of the applicable self organization methods that can be used here. The other possible solutions are neural networks – currently considered theoretically (Shahin et al., 2005) (concurrent learning or Kohonen networks). The method used in this case treats every set of measurements (further called "experiment") made at the particular depth of the soil as the node in the graph. The edges connect the similar nodes, i.e. the nodes which measurements' values are close to each other. This way the connected nodes form a cluster, which is interpreted as the particular type of the soil. It must then be identified by the human expert. The similarity graph can be stored in two ways: as the matrix or the list. Since the former requires more memory and makes the algorithm slower, the latter was selected for the implementation of the method. Every experiment from the data set has its own similarity list.

The learning data set for the clustering method has the form of the matrix with dimensions of $n \times m$, where n is the number of rows – measurement sets (experiments), and m the number of the measured parameters of the single experiment, in this case the preprocessed pressures A and A (in bars) – see Section 2. Additional two columns store information about the depth of the experiment and its ordinal number (identifier). The former is used to mark the ground category by the human operator, while the latter is used in the similarity list constructions. The exemplary fragment of the learning data set is presented in Table 1.

TABLE 1. Fragment of the learning data set for the clustering algorithm

Ident.	Depth [m]	Pressure A [bar]	Pressure B [bar]
1	0.3	0.20	1.10
2	0.9	0.20	1.10
3	1.7	0.24	1.01
4	2.3	0.35	1.13

Example 1:

For four exemplary experiments the similarity lists would have the form as presented in Figure 6.
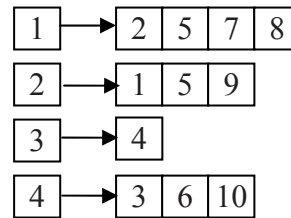

FIGURE 6. Example of the similarity lists for the number of measurements

Experiments labeled with number "1" are similar to the experiments labeled as "2", "5", "7", and "8". Note that because there is the similarity between the experiments labeled as "2" and "1", the cluster would contain the experiments numbered with "1", "2", "5", "7", "8", and "9". The second cluster covers the experiments labeled with "3", "4", "6", and "10".

The algorithm of the cluster generation consists of the following steps:

1. For every measured parameter calculate similarity factors of the experiments, expressing, how close are the measurements to each other, considering the particular parameter. The similarity factor considers the difference between the particular parameter values for two experiments. It is then

compared to the threshold value θ. The results of the tests are the experiments similar to each other considering the particular parameter.

2. If the similarity lists are empty (the first parameter is considered), add the similar examples to the list of the particular measurement sets. Otherwise, leave in the list only these experiments similar to the experiment considering particular parameter, which were there before. This way after the whole process only the similar experiments considering every parameter are left in the list.

3. Mark the experiments belonging to the same group with a number. This way the clusters are ready to be described by the human operator.
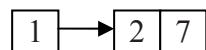
Example 2:

If the first pressure in the DMT measurements learning data set from Example 1 result in the following similarity list for the first experiment:

$$1 \longrightarrow \boxed{2 \mid 5 \mid 7 \mid 8}$$
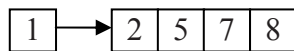
and the second pressure results in the following experiments similar to the one labeled as "1":

$$1 \longrightarrow \boxed{2 \mid 3 \mid 7 \mid 10}$$
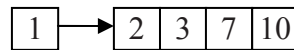
then the resulting similarity list is as follows:

$$1 \longrightarrow \boxed{2 \mid 7}$$

After the last iteration of the clustering algorithm the similarity lists contain the minimum number of the experiments, which form the clusters, denoting the particular ground category.

This last step of the algorithm must ensure that all the similar experiments (from all the generated lists) are covered by the same category.

The main problem of the algorithm is the selection of the threshold value. It is set individually for each parameter, as it represents another quantity. The first approach is the selection of the threshold as the percentage value for the dynamic range of the parameter value (the difference between the maximum and the minimum values of the measured parameter). For example, the initial value of the threshold θ used in the described algorithm was set to the 5 percent of the dynamic range for both pressure measurements.

The additional step in the algorithm, run optionally, is the de-noising of the measurement data. Because the measurements are taken with the finite accuracy and sometimes are result of the randomly encountered deviations of the ground layers (such as rocks), elimination of such disturbances can facilitate the clustering operation and minimize the number of the developed subsoil categories. In the described analysis, a median filter was tested. Every sample of the original signal is replaced by the value in the middle of the ordered neighbor values (called mask). This way the noise is eliminated from the signal. The size of the mask determines the range of the de-noising and affects the quality of the procedure. The exemplary results of the denoising for the median filter with the mask of length 5 are in Figure 6.
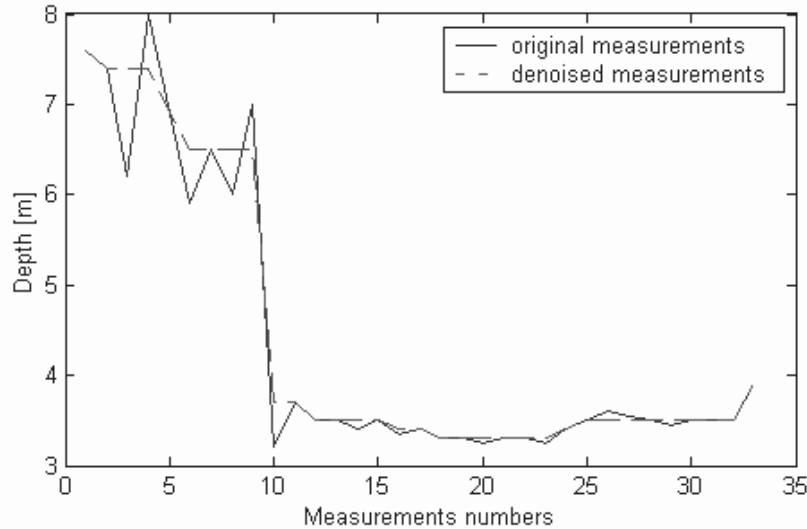
FIGURE 6. Example of the DMT measurements denoising

The algorithm was applied to the different DMT tests in the various sites within the campus of WULS area.. Its main purpose was to determine, which are the optimal values of the threshold θ, allowing to correctly assign the soil types to the measurements. Firstly, the data sets were analyzed by the algorithm. It resulted in the scheme illustrating the profile of the soil. Then, the human expert performed analysis of the profile, deciding, for which value of the threshold clustering is the closest to the actual soil profile. Finally, the soil types within the profile were described. Results for the exemplary soil test site and particular values of the threshold are in Figure 7.

The results show that the profile closest to the actual situation is in Figure 7a, where subsoil classes were identified as brown boulder clay and grey boulder clay. For too small threshold values the number of the ground classes is too large, while for too large θ – too small. Selection of the proper threshold is the

difficult problem, affecting the clustering effectiveness. As for every test site, the θ value will be different, there is the need to develop methodology of the automatic threshold selection, based on the characteristics of the data set obtained from the ground measurements.

CONCLUSIONS

A method of the DMT-based soil investigation was examined and on its base the usage of the artificial intelligence and graph clustering algorithm. The parameterized algorithm for the grouping of the soil categories was introduced, using the different values of the dilatometer measurements (A and B pressures). The influence of the parameters of the clustering algorithm on the accuracy of the categorization was verified by the human expert. Also, based on the clustering method, the subsequent soil types were identified.
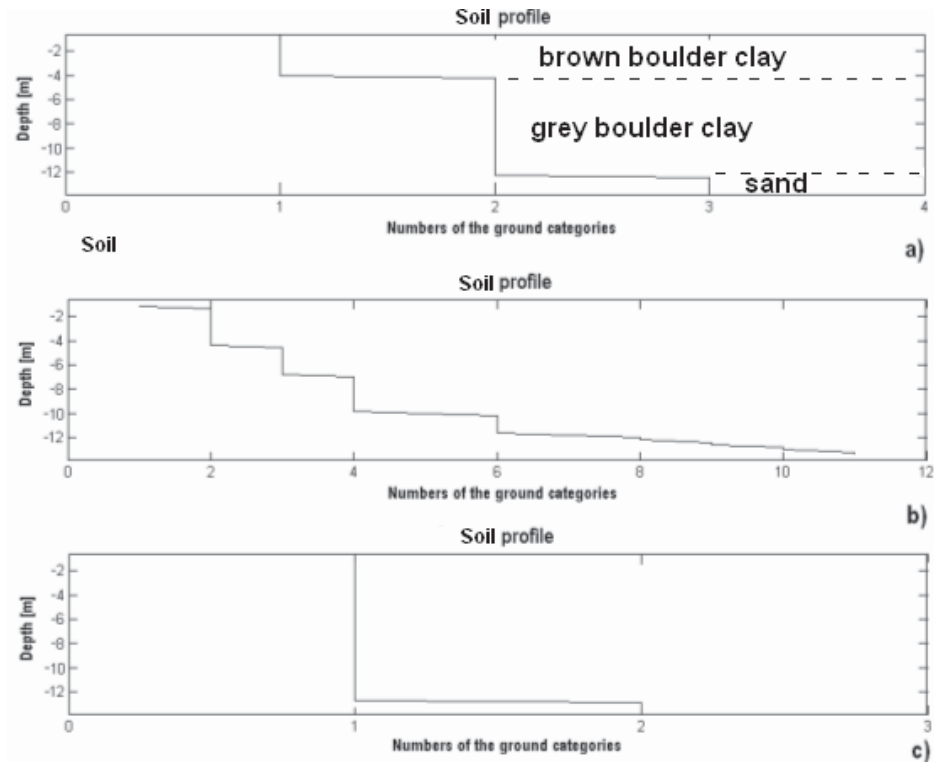
FIGURE 7. Results of the soil clustering for the threshold values: θ=0.05 (a), θ=0.04 (b), θ=0.06 (c)

Such cooperation between the algorithm designers and geotechnical experts will result in the soil classification data base, containing the required parameters required to distinguish the particular soil types. Therefore, further analysis of the DMT tests of soil investigation supported by the computer algorithms to extract knowledge from the measured data is required. Also, other algorithms for grouping the measurements in order to determine the soil profile in the test site, should be examined.

REFERENCES

BILSKI P. 2007: Automated Diagnostic System Using Graph Clustering Algorithm and Fuzzy Logic Method, ECCTD 2007, Sevilla, Spain, Aug. 26––30, pp. 779–782.

BOULTON G.S. 1970: On the deposition of subglacial and melt-out tills at the margins of certain Svalbard glaciers. Journal of Glaciology 9 (56): 231–245.

BRIAUD J., MIRAN J. 1991: The flat dilatometer test, TX, 77843-3136 USA for The Federal Higway Administration.

DAS S.K., and BASUDHAR P.K. 2004: "Site characterization of an alluvial site using artificial neural network," International workshop on Risk Assessment in Site Characterization and Geotechnical Design, Banglore.

GEOTECHNICAL DOCUMENTATION concerning soil and water conditions in site destined for two laboratorial and educational buildings of Warsaw University of Life Sciences campus. Departament of Geotechnical engineering, Warsaw Uni-

versity of Life Sciences campus, 2001 (in Polish).

ITANI O.M., NAJJAR Y.M. 2000: "Three-dimensional modeling of spatial properties via artificial neural network," Transport Research Record, No 1709 , pp. 50–59.

JUANG, C.H., JIANG T., and CHRISTOPHER R.A. 2001: "Three-dimensional site characterisation: neural network approach," Geotechnique, Vol. 51 No 9, pp. 799–809.

LUTENEGGER A. 1988: Current status of the Marchetti dilatometer test. Proc. Int. Sym. on Penetration Testing ISOPT-1. Orlando, 1: 137–155.

MARCHETTI S. 1980: In situ tests by flat dilatometer. J. Geotech. Eng. Div., ASCE, 106, GT3: 299–321.

MARCHETTI S. CRAPPS D. 1981: Flat dilatometer manual, Internal report of GPE.

MORAWSKI W. 1980: Explanations to the Detailed Geological Map of Poland 1:50 000, Warsaw West sheet, (in Polish), Polish Geological Institute, Warsaw

RÓŻYCKI S.Z., SUJKOWSKI Z. 1936: Geological profiles in Warsaw. City Department in Warsaw (in Polish).

RÓŻYCKI S.Z. 1972: Pleistocene of the central Poland on the tertiary background. PWN. Warsaw (in Polish).

SARNACKA Z. 1976: Explanations to the Detailed Geological Map of Poland 1:50 000. Piaseczno sheet; Geological Institute. Warsaw (in Polish).

SARNACKA Z. 1992: Stratigraphy of the quaternary sediments in Warsaw and neighborhood. PIG Works nr 138. pp. 36 (in Polish).

SHAHIN M.A., JAKSA M.B., MAIER H.R. 2005: Neural network based stochastic design charts for settlement prediction, Can. Geotech. Jour., pp.110–120.

SHAHIN M.A., JAKSA M.B., and MAIER H.R. 2001: "Artificial neural network applications in geotechnical Engineering," Australian Geomechanics, Vol. 36, No 1, p. 49–62.

TOLL D.G. 1996: "Artificial Intelligence Applications in Geotechnical Engineering," Electronic Journal of Geotechnical Engineering, Vol. 1.

WIENCŁAW E., SIWIEC T., GRUNWALD P., MORAWSKI D. 1996: Quality of the underground waters in the Quality of the underground waters in the Scientific-Research Water Supply Station of the Warsaw University of Life Sciencesin Ursynów ; Przegląd Naukowy Wydziału Melioracji i Inżynierii Środowiska. z. 9, 39–50 (in Polish).

Streszczenie: W artykule przedstawiono zastosowanie opartej na algorytmie clusteringu grafowego do rozpoznania warstw gruntu na podstawie badań dylatometrycznych. Wyniki pomiarów uzyskiwane dla podłoża jednego z budynków na terenie kampusu SGGW zostały przeanalizowane przez algorytm, który wyodrębnił grupy pomiarów charakterystyczne dla określonych rodzajów gruntów. Przeprowadzona analiza umożliwiała określenie optymalnych wartości parametrów pozwalających pogrupować pomiary i wyodrębnić najbliższe rodzaje gruntów. Ponadto, wyniki działania algorytmu zostały przeanalizowane przez geotechników w celu weryfikacji identyfikacji poszczególnych rodzajów gruntów wskazanych przez metodę.

Authors' addresses:
Simon Rabarijoely, Tomasz Falkowski
Katedra Geoinżynierii
Wydział Inżynierii i Kształtowania Środowiska
SGGW
02-776 Warszawa, ul. Nowoursynowska 159, Poland
e-mail: simon_rabarijoely@sggw.pl
tomasz_falkowski@sggw.pl

Piotr Bilski
Katedra Informatyki
SGGW
02-776 Warszawa, ul. Nowoursynowska 159, Poland
e-mail: foxtrot@sgk.info.pl