

József Kovács, Nikolett Bodnár, and Ákos Török\*

# The application of multivariate data analysis in the interpretation of engineering geological parameters

Miocene sediments, cored along the metro line in Budapest

DOI 10.1515/geo-2016-0005

Received October 15, 2015; accepted January 03, 2016

**Abstract:** The paper presents the evaluation of engineering geological laboratory test results of core drillings along the new metro line (line 4) in Budapest by using a multivariate data analysis. A data set of 30 core drillings with a total coring length of over 1500 meters was studied. Of the eleven engineering geological parameters considered in this study, only the five most reliable (void ratio, dry bulk density, angle of internal friction, cohesion and compressive strength) representing 1260 data points were used for multivariate (cluster and discriminant) analyses. To test the results of the cluster analysis discriminant analysis was used. The results suggest that the use of multivariate analyses allows the identification of different groups of sediments even when the data sets are overlapping and contain several uncertainties. The tests also prove that the use of these methods for seemingly very scattered parameters is crucial in obtaining reliable engineering geological data for design.

**Keywords:** cluster analysis; sedimentary rocks; density; cohesion; angle of internal friction

## 1 Introduction

Multivariate analysis is an important tool in data management and it has been widely used in managing and interpreting geochemical characterisation of groundwater [1, 2] in geochemical characterization of soils science [3, 4]. Wa-

ter quality changes of surface waters [5, 6] and groundwater resources [7] were also studied recently by using this tool. The broad application also includes the ecology of lacustrine environment [8] and dry land palaeoenvironments, such as paleosols in loess [9, 10]. More recently increasing number of publications appears in hydrogeochemical and water quality research [11–13]. Thus, the use of multivariate methods requires large data sets and therefore its application in engineering geology is less common, since even for large construction projects the number of laboratory analyses usually provides only a relatively limited amount of data in statistical terms. At the same time, it has been known for some time that data analyses are useful tools in engineering geology [14]. More recent studies have demonstrated the applicability of multivariate data analyses in various fields of engineering geology such as rock engineering [15, 16], soil liquefaction [17], landslide susceptibility analyses [18–20] and even in investigating the correlation between clay mineralogy and shear strength of soils [21]. The present paper attempts to gain new insights into the problem of engineering geological data analysis by using a data set of mechanical parameters obtained from laboratory tests during the construction of a new metro line in Budapest. The data set was obtained from 30 core drillings (2041 m of cores in total) with 9554 data points. Each data point represents the result of a laboratory test of 11 different engineering geological parameters. The data was digitized, and after screening the data sources it was found that from the above mentioned nearly ten thousand data points only 252 samples had been tested for the required amount of parameters, thus allowing the use of 1260 data points for multivariate data analyses. Although a large number of data were produced during laboratory tests a careful selection of data is required to carry out cluster analyses. The main aim of this research was to demonstrate the use of multivariate data analyses in the identification of different lithotypes based on their engineering geological parameters and to classify the sediments according to their physical parameters.

**József Kovács:** Eötvös Loránd University, Department of Applied Geology

**Nikolett Bodnár:** Geological and Geophysical Institute of Hungary, Department of Environmental Geology

**\*Corresponding Author: Ákos Török:** Budapest University of Technology and Economics, Department of Geotechnics and Engineering Geology, Műegyetem rakpart 3., H-1111, Budapest, Hungary, E-mail: torokakos@mail.bme.hu

## 2 Geological setting

Budapest is characterized by morphology controlled geological setting, with low laying flats mostly covered by Miocene sediments on the eastern side of river Danube, and an elevated side with Triassic-Eocene-Oligo-Miocene sediments on the western flanks of the river [22]. The new metro line (line no 4) in Budapest can be divided into three sections on the basis of their differing geological structure: i) the first section on the Buda side is characterized by Oligocene highly consolidated clay layers, ii) the Danube crossing part, which is intersected by faults and includes karstified Triassic dolomite horsts and iii) the Pest side that has been cut into various sediments having different consistency and engineering geological properties [23]. The typical sediment of the first section belongs to the Kiscell Clay Formation. It includes thick-bedded, grey to bluish grey pyrite-rich clay with minor carbonate and mica content. The upper part of the clay is weathered and show signs of disintegration [24]. The clay forms a relatively impermeable boundary; however some faults serve as conduits. The cover beds contain alluvial sand and sandy gravel of Quaternary age. The topmost part is characterized by anthropogenic landfill. The Danube crossing part forms a typical asymmetric horst that is intersected by NW-SE faults. The tunnel was cut here into an Oligocene Clay (Tard Clay Formation). This laminated dark grey clay is in a tectonic contact with a sandy sequence that forms part of a Late Oligocene - Early Miocene sequence. The metro line on the Pest side intersects predominantly Miocene sediments, with minor amount of Oligocene deposits (Figure 1). The Miocene sequence shows a great deal of lithological variety, which is covered by Quaternary river deposits, consisting of sandy and gravelly sediments. From the riverbank to Kálvin Square clay, siltstone, sandy clay, and weakly cemented sandstone are found (Figure 1). Tuffaceous beds represent the “middle tuff horizon” [25]. Bentonitic clays are also very commonly found as widespread layers, intercalations or lenticular bodies. The area of Rákóczi Square metro station is covered by variegated siltstone, which encompasses bentonitic clays and lenticular sand bodies [25, 26]. Previous studies [27] suggest the presence of faults that intersects the siltstone layers. The groundwater table is controlled by the River Danube forming a hydrostatic system. At a distance of 2 kilometres from the river bed the influence of the river is clearly documented especially because of the high conductivity of sandy Miocene layers on the Pest side [28].

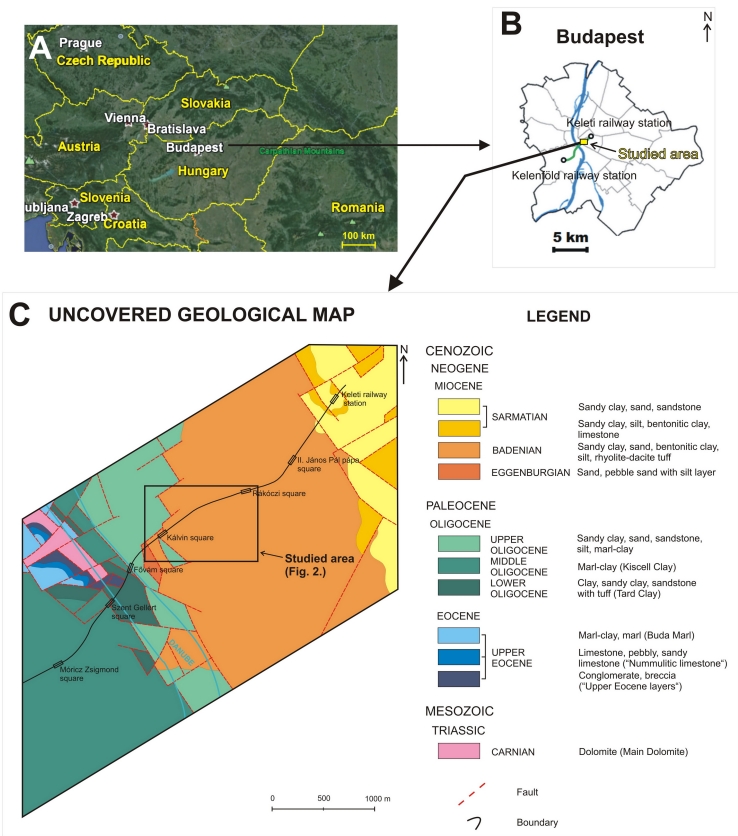
## 3 Data

The data set under investigation comprises the core description, that is the engineering geological and soil mechanical laboratory analyses of 30 cores. The boreholes were drilled in the surroundings of Kálvin and Rákóczi Squares (Figure 2). The data was available only on paper in the form of core logs and laboratory analyses providing information on the soil, the mechanical and engineering geological parameters of Miocene sediments (Figure 1). The coring depth was between 31 and 75 metres. The studied cores were selected from a set of 70 cores, the selection being based on the availability of geographical data, taking into consideration the question of whether laboratory data were measured at the same time. Some of the previous laboratory analyses had used archive data and units, and these were converted into SI units. 11 geotechnical parameters were selected in the preliminary phase of research providing 9554 data points. These parameters were the following: water content, index of plasticity, coefficient of skewness, void ratio, water saturated density, dry bulk density, angle of friction, cohesion, compressive strength, modulus of elasticity, Poisson-ratio. After the digitalization, the data base was further processed and a coherent data set comprising 1260 data points was used for multivariate analyses (Table 1). The deposits that were described in core logs as gravel were not used in the multivariate analyses since several parameters (such as index of plasticity) were not available for these sediments. The core log descriptions were reviewed and five different lithologies were considered in the present study: sand, silt, moderately swelling clay, swelling clay, and bentonite.

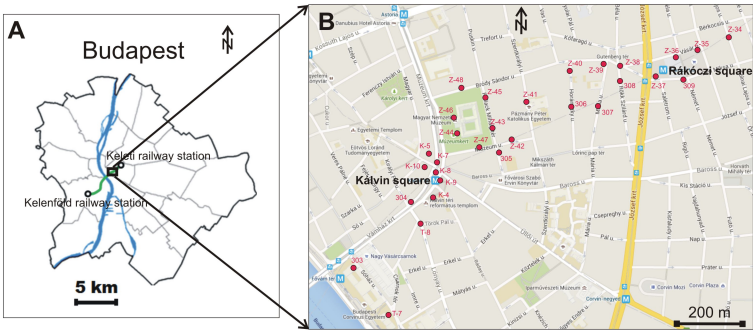
## 4 Statistical and multivariate data analyses

To evaluate the engineering geological parameters and their correlation matrix SPSS software was used. Prior to the application of multivariate data analyses, the filtering of data was necessary, since data with strong correlation parameters are not recommended for use as input variables in cluster analyses. To analyse stochastic relationships, a matrix of correlation was used. The correlation coefficient ( $R$ ) and its square, the coefficient of determination ( $R^2$ ) describe the linear connection [29]. The correlation is strong, when  $|R| \geq 0.7$  and weak, when  $|R| \leq 0.5$ .

For data evaluation, it is important that no missing data occur in the matrix. The samples with missing data



**Figure 1:** Location of Budapest in Hungary (a); the outline of Budapest with the new metro line and the study area (b); and (c) the uncovered geological map of Budapest line 4 with the studied area, marked by a square. Please note that the locations of boreholes are shown on Figure 2 (the geological map (c) is simplified after [26]).



**Figure 2:** The location of studied drillings in Budapest (drillings are marked by red circles).

**Table 1:** Data set used in this paper (see the location of drillings 300, K, T, Z on Figure 2).

number of data		drillings (code of boreholes)				summary (no. of data points)
		300	K	T	Z	
Void ratio	$e [-]$	68	58	18	108	252
Dry bulk density	$\rho_d [\text{kg}/\text{m}^3]$	68	58	18	108	252
Angle of friction	$\varphi [^\circ]$	68	58	18	108	252
Cohesion	$c [\text{kN}/\text{m}^2]$	68	58	18	108	252
Compressive strength	$\delta [\text{kPa}]$	68	58	18	108	252
Summary no. of data - set of drillings		340	290	90	540	1260

points were not used in the analyses. The extreme values were evaluated using published results [23–25, 27] and also theoretically, thus mistyped and incorrect analytical results were eliminated. Cluster analysis is a kind of multivariate data analysis that allows the reduction of dimensions and the grouping of samples into fairly “homogeneous” groups. These groups are called clusters<sup>1</sup> [30, 31]. The grouping is based on similarities and dissimilarities, and for its application a previous understanding of groups is not needed. The key tool in the cluster analysis is linkage distance, which is gradually calculated [32] and visualized in dendrograms. Based on our experience and the references cited in the text [1, 2, 12] it is necessary to clarify how many clusters or geologically justified groups there are within the data set.

The existence of the clusters was verified by using hypothesis tests and linear discriminant analysis. This was necessary, since the lack of hypothesis tests might lead to the misinterpretation of results. Discriminant analysis was performed to describe the extent to which the planes separating the groups could be distinguished. The results of discriminant analysis are shown in the percentages of the planes that separate groups [33, 34], and it provides information on the grouping of each sample. When a repeated discriminant analysis is performed, the first grouping is considered as primary and the second grouping provides a new result. These steps are repeated until there is no difference between the primary and the suggested grouping. The results are often shown in planes representing the first two discriminate functions [12].

The role of each parameter in determining the clusters was analyzed by using Wilks’  $\lambda$  distribution as it is given in Equation (1). This equation provides information on the sum of squares within the group as a ratio of the total sum of squares<sup>1</sup> [35].

$$\lambda = \frac{\sum_i \sum_j (x_{ij} - \bar{x}_i)^2}{\sum_i \sum_j (x_{ij} - \bar{x})^2}, \quad (1)$$

where  $x_{ij}$  is the „jth” element of the „ith” group and  $\bar{x}_i$  the mean of the „ith” group, while  $\bar{x}$  is the total mean.

The relationships between each group can be visualized on box-and-whiskers plots [36].

From the selected and gathered eleven parameters several set of physical parameters (e.g. index of plasticity) cannot be defined for all studied lithologies. For cluster analyses strongly correlating parameters are not appropriate, therefore stochastic analyses were made for

each lithotypes to ensure the required parameter set. After stochastic analyses only five parameters remained in the data set, including void ratio, dry bulk density, angle of internal friction, cohesion and compressive strength. 252 samples which contain 1260 data were analysed by the means of mathematical statistics.

## 5 Results and discussions

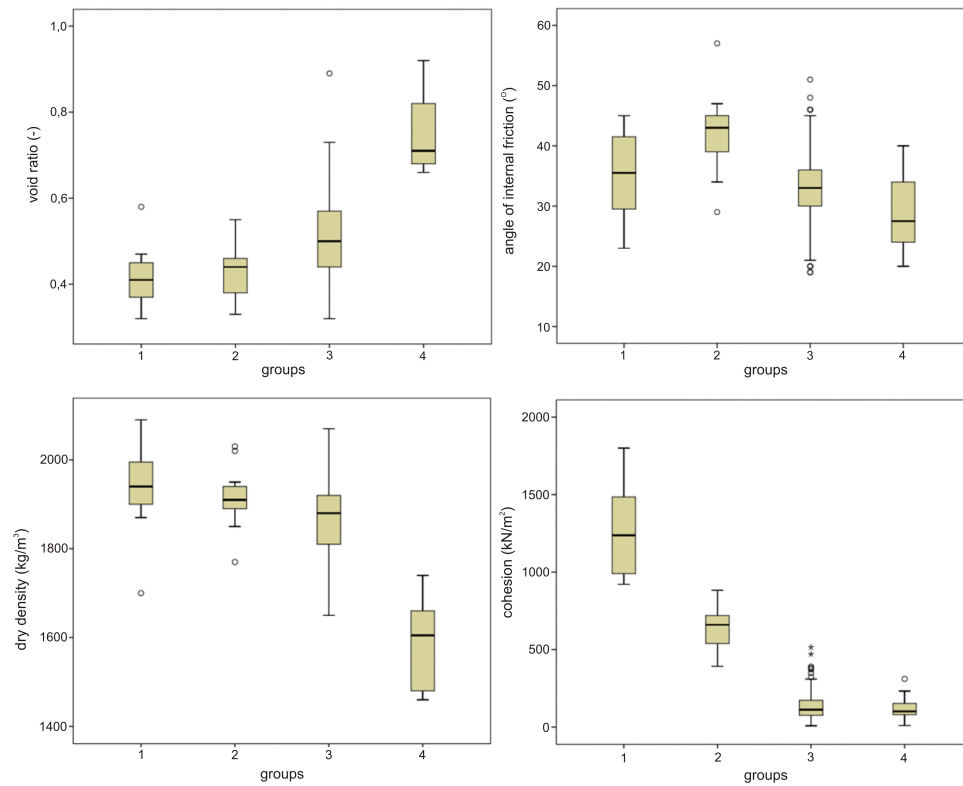
The correlation analyses of the selected five parameters indicated that there was a very strong correlation between cohesion and compressive strength, with a correlation coefficient of 0.95. It suggests that these parameters can be calculated from each other in 90% of cases. Comparison of box-and-whisker’s plots of 4 parameter analyses (first compressive strength than cohesion was excluded from the analysis) clearly indicate that these two parameters are interrelated (Figure 3 and Figure 4). The two parameters are linking parameters, since the cohesion of soft sediments can be calculated from unconfined compressive strength. It is also necessary to emphasize, that the groups on figures (Figure 3, Figure 4) do not represent uniform lithologies but rather contain samples with different lithologies.

It has been reported previously that cohesion strongly correlates with slake durability in the case of mudrocks [21]. The same study also suggests that cohesion is a key parameter in assessing mudrock properties. Cohesion and compressive strength would be negligible in mathematical terms, since they can be calculated from each other, but from an engineering geological point of view, they represent important information. Therefore these were also used in the multivariate analyses. Three different sets of parameters were studied: one with 5 parameters, and two with 4 parameters (first compressive strength than cohesion was excluded from the analysis).

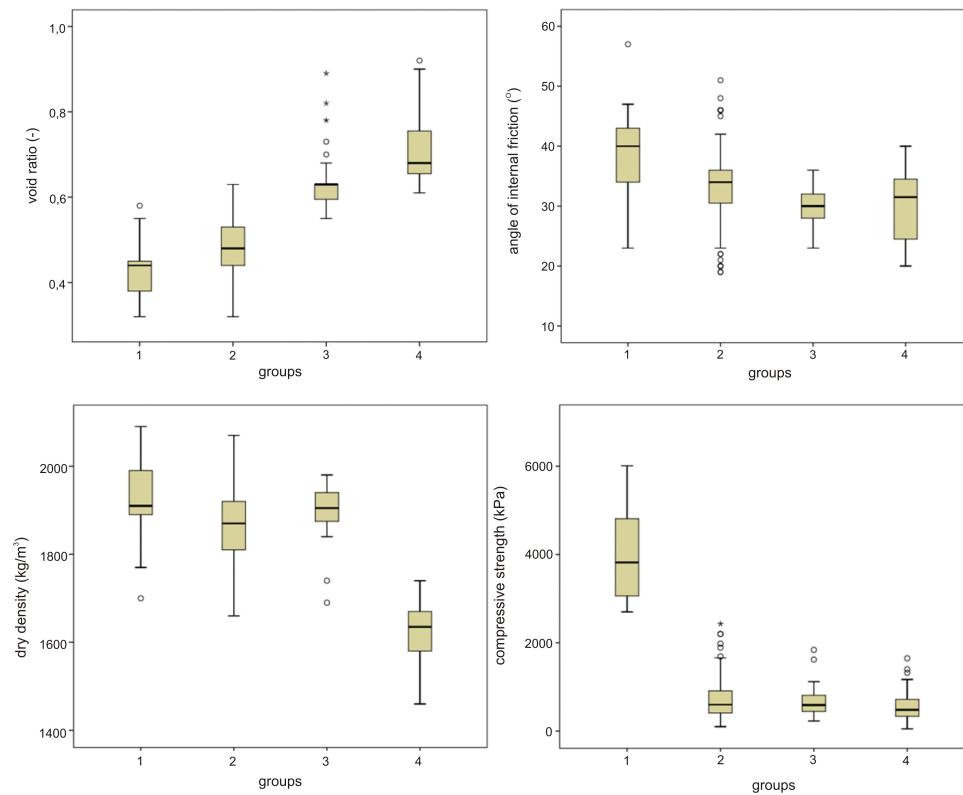
Based on the results of cluster analyses the samples were grouped into 4 clusters in each case (5 parameters and two times 4 parameters analyses).

The groupings were verified using discriminant analysis, which indicated that the obtained clusters may be considered to have a verification of 90.8% when 5 parameters were used. The centroids of the groups are very distinct even in 2D (Figure 5). The fourth iteration step of the discriminant analysis allowed a 100% distinction between the groups. The linear discriminant analysis also verified the existence of four distinct groups when 4 parameters were used.

<sup>1</sup> IBM Introduction to statistical analysis with PASW Statistics. IBM Company, Chicago IL, USA, 2010.

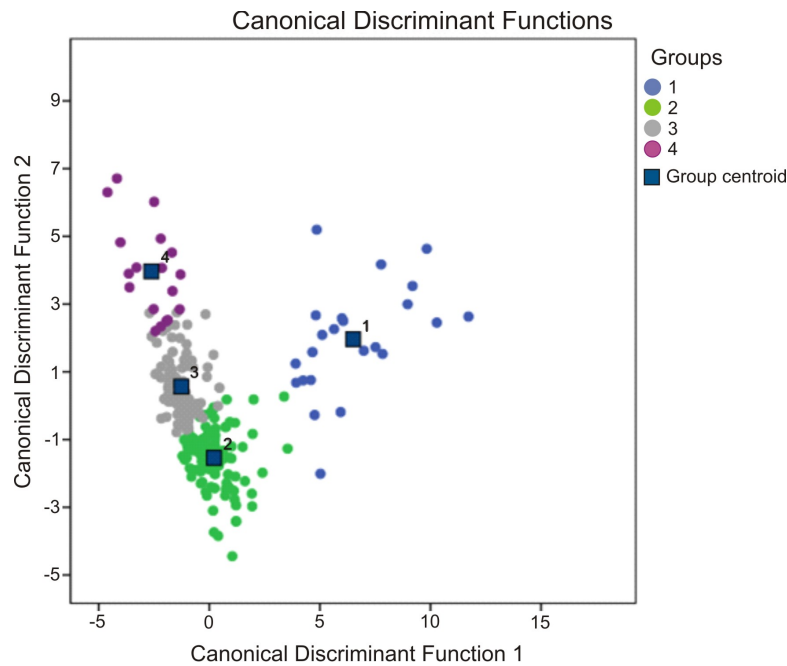


**Figure 3:** Box and whisker's plots of four parameter analyses: (a) void ratio; (b) dry density, (c) angle of internal friction, and (d) cohesion.



**Figure 4:** Box-and-whisker's plot of four parameter analyses: (a) void ratio; (b) dry density, (c) angle of internal friction, and (d) compressive strength.





**Figure 5:** Discriminant results projected onto the 2D surface of the first two discriminant functions (5 parameters: void ratio, dry density, angle of internal friction, cohesion and compressive strength).

When cohesion or compressive strength was not considered in the cluster analysis, the results of the cluster analysis were different. The grouping was modified less when compressive strength was not included (Figure 6), since in the 5 parameter analysis the first group overlapped 90% with the first and second group of the 4 parameter analysis. The second and third groups are coherent with the united third group. The fourth groups of the analysis are equal.

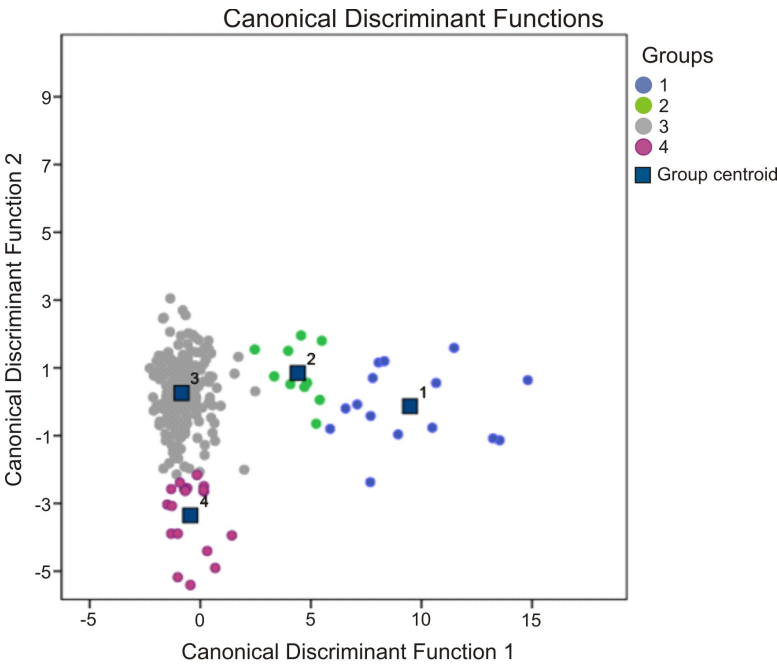
Cluster analysis performed without using cohesion shows a very different picture with significant data scattering (Figure 7). The first group of data in the 5 parameter analysis cross-comply 90% with the first group, but for all other groups the data set distribution changed significantly.

In cluster analysis it is suggested that one of the two strongly correlated parameters not be considered; however from an engineering geological point of view this study showed that the cluster analysis with 4 parameters without cohesion or without compressive strength gave very different results. As a consequence, the coherent use of these strongly correlating parameters is required to obtain reliable results. This is in good agreement with the findings of [21].

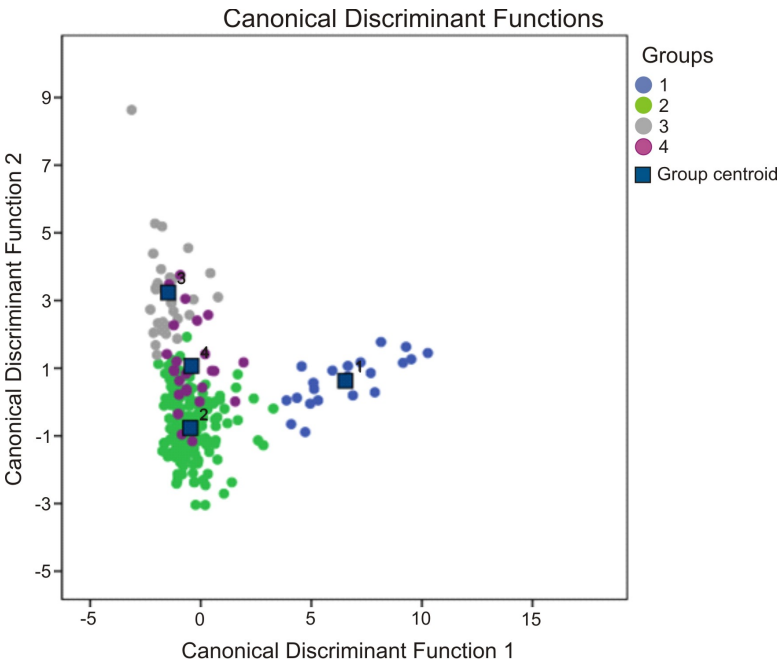
Wilks'  $\lambda$  statistics indicates which parameter has the greatest influence on cluster formation. According to our analyses compressive strength and cohesion has the greatest influence on the grouping while angle of internal friction

is the least influencing factor for all three scenarios (that is the 5 parameter and the two 4 parameter analyses) (Table 2). Void ratio and dry bulk density have a moderate degree of influence on cluster formation. The angle of internal friction was also found to be less important parameter when an intensive data set of riverbank soils was studied [37].

The discriminant analysis did not allow the differentiation of the various lithological categories of core logs; a significant overlap was found for the same lithologies. The original groups of lithotypes were correctly identified only in 35-45% of cases (Figure 8). As a consequence, sediments that were described as sand in the core logs might have properties associated with swelling clay, or alternatively the opposite may also occur. The overlaps are also related to the fact that the physical parameters of sediments are strongly controlled by material properties and micro-fabric, such as lamination, orientation of clay layers, and syn-sedimentary deformation structures. The importance of clay mineralogy and the clay content as a control function of the physical properties such as density, Atterberg consistency limits and compressive strength of mudrocks were also emphasized by [38]. Carbonate content can also have a strong effect on the strength and plasticity of hard soils-soft rocks [39]. Our study provides a fresh example of the fact that lithological variations only partly determine the strength parameters and cohesion, angle of friction can be highly variable and display major changes even



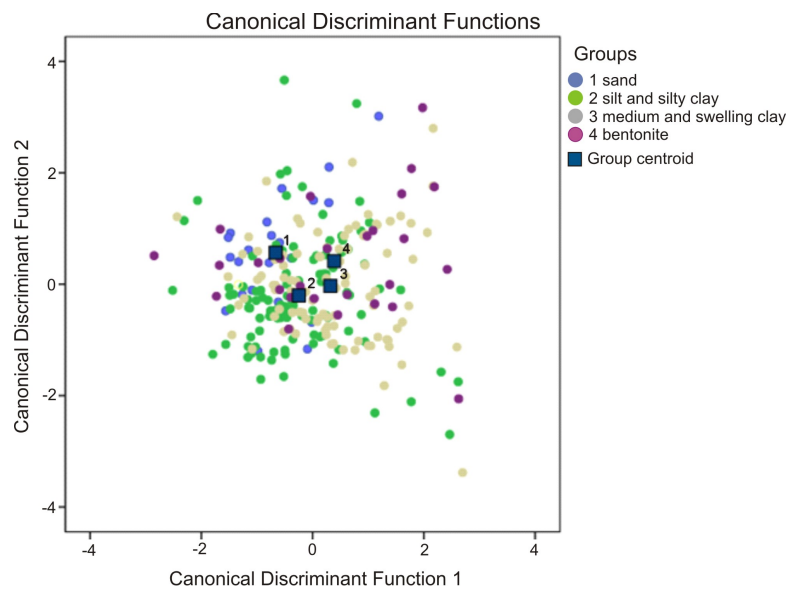
**Figure 6:** Discriminant analysis results projected onto the 2D surface of the first two discriminant functions (4 parameters: void ratio, dry density, angle of internal friction, cohesion).



**Figure 7:** Discriminant analysis results projected onto the 2D surface of the first two discriminant functions (4 parameters: void ratio, dry density, angle of internal friction, and compressive strength).

**Table 2:** Wilks'  $\lambda$  values for 5 parameters and 4 parameters (alternately, either compressive strength or cohesion was not considered for the 4 parameter analyses).

Parameter	5 parameters	4 parameters (no compressive strength)	4 parameters (no cohesion)
compressive strength	0.228	-	0.223
cohesion	0.232	0.139	-
void ratio	0.324	0.578	0.405
dry bulk density	0.438	0.564	0.513
angle of friction	0.77	0.836	0.86



**Figure 8:** Discriminant analysis results obtained from lithological descriptions of core logs (5 parameters: void ratio, dry density, angle of internal friction, cohesion and compressive strength).



when only a minor amount of clay is found in the sample [21]. The other factors that might influence the parameters are micro-fractures and cementation. Minor amounts of clay minerals in sand or silt can increase the compressive strength to such an extent that the clay containing sediment has a higher strength than pure sand or silt [39]. In the current study, no clay content data was available and the mineralogy of the samples was not listed in the core logs or laboratory analyses. The results of discriminant analysis clearly demonstrate that the lithological descriptions can only be used conscientiously with strong reservation. Accordingly, it is not possible to predict the physical properties of a given strata based on the lithological description of these core logs.

## 6 Conclusions

A large data set of engineering geological parameters obtained from laboratory tests of core drillings over a relative small area represent very heterogeneous rock types with various parameters. Of the 11 available geotechnical parameters and 9554 data points only five parameters remained in the data set after filtration suggesting that archive data sources may often be problematic to deal with. Of the available parameters a coherent data set containing 1260 data points was used for multivariate analyses representing five engineering geological index properties such as void ratio, dry bulk density, angle of friction, cohesion and compressive strength. In the study five different lithologies were considered: sand, silt, moderately swelling clay, swelling clay, and bentonite, but the discriminant analysis did not allow the differentiation of the various lithological categories of core logs; a significant overlap was found for the same lithologies. This indicates that minor differences in lithology such as clay content or carbonate cementation can cause major discrepancies in physical parameters. The Wilks'  $\lambda$  distribution analysis suggests that compressive strength and cohesion have the highest influence on the grouping, while internal friction angle has the lowest influence on data point distribution in clusters. In correlation analyses a very strong correlation between cohesion and compressive strength was found with a correlation coefficient of 0.95. From the point of view of mathematical considerations these parameters are strongly related, but our study suggests that from an engineering geological point of view they represent important information and thus it is suggested that both parameters be used in multivariate analyses.

**Acknowledgement:** We appreciate the support provided by the colleagues of the Department of Environmental Geology, of the Geological and Geophysical Institute of Hungary as well as to the colleagues from the Budapest University of Technology and Economics. The presentation of the research has been supported in the framework of the project 'Talent care and cultivation in the scientific workshops of BME' project by the grant TÁMOP - 4.2.2.B-10/1-2010-0009.

## References

- [1] Cloutier V., Lefebvre R., Therrien R., Savard M.M., Multivariate data analysis of geochemical data as indicative of the hydro-geochemical evolution of groundwater in a sedimentary rock aquifer system, *J. Hydrol.*, 2008, 353(3-4), 294-313.
- [2] Belkhir L., Boudoukha A., Mouni L., Baouz T., Application of multivariate statistical methods and inverse geochemical modeling for characterization of groundwater - A case study: Ain Azel plain (Algeria). *Geoderma*, 2010, 159(3-4), 390-398.
- [3] Draw L.J., Grunsky E.C., Sutphin D.M., Woodruff L.G., Multivariate analysis of the geochemistry and mineralogy of soils along two continental-scale transects in North America. *Sci. Total. Environ.*, 2010, 409(1), 218-227.
- [4] Bradák B., Thamó-Bozsó E., Kovács J., Márton E., Csillag G., Horváth E., Characteristics of Pleistocene climate cycles identified in Cérna Valley loess-paleosol section (Vértesszicsa, Hungary). *Quatern. Int.*, 2011, 234(1), 86-97.
- [5] Hatvani I.G., Kovács J., Kovács I.S., Jakusch P., Korponai J., Analysis of long-term water quality changes in the Kis-Balaton Water Protection System with time series-, cluster analysis and Wilks' lambda distribution. *Ecol. Eng.*, 2011, 37(4), 629-635.
- [6] Magyar N., Hatvani I. G., Kovácsné Székely I., Herzig A., Dinka M., Kovács J., Application of multivariate statistical methods in determining spatial changes in water quality in the Austrian part of Neusiedler See. *Ecol. Eng.*, 2013, 55, 82-92.
- [7] Ujevic Bosnjak M., Capak K., Jazbec A., Casiot C., Sipos L., Poljak V., Dadic Z., Hydrochemical characterization of arsenic contaminated alluvial aquifers in Eastern Croatia using multivariate statistical techniques and arsenic risk assessment. *Sci. Total. Environ.*, 2012, 420, 100-110.
- [8] Dalu T., Richoux N.B., Froneman P.W., Using multivariate analysis and stable isotopes to assess the effects of substrate type on phytobenthos communities. *Journal of the International Society of Limnology, Inland Waters*, 2014, 4(4), 397-412.
- [9] Bradák B., Kiss K., Barta G., Varga Gy., Szeberényi J., Józsa S., Novothny Á., Kovács J., Markó A., Mészáros E., Szalai Z., Different paleoenvironments of Late Pleistocene age identified in Verőce outcrop, Hungary. *Quatern. Int.*, 2014, 319, 119-136.
- [10] Bradák B., Kovács J., Quaternary surface processes indicated by the magnetic fabric of undisturbed, reworked and fine-layered loess in Hungary. *Quatern. Int.*, 2014, 319, 76-87.
- [11] Hatvani I.G., Clement A., Kovács J., Kovács I.S., Korponai J., Assessing water-quality data: The relationship between the water quality amelioration of Lake Balaton and the construction of its mitigation wetland. *J. Great Lakes Res.*, 2014, 40(1), 115-125.

- [12] Kovács J., Kovács S., Magyar N., Tanos P., Hatvani I.G., Anda A., Classification into homogeneous groups using combined cluster and discriminant analysis. *Environ. Modell. Softw.*, 2014, 57, 52-59.
- [13] Matiatos I., Alexopoulos A., Godelitsas A., Multivariate data analysis of the hydrogeochemical and isotopic composition of the groundwater resources in northeastern Peloponnesus (Greece). *Sci. Total. Environ.*, 2014, 476-477, 577-590.
- [14] Muspratt M.A., Numerical statistics in engineering geology. *Eng. Geol.*, 1972, 6, 67-78.
- [15] Ulusay R., Türeli K., Ider M.H., Prediction of engineering properties of a selected litharenite sandstone from its petrographic characteristics using correlation and multivariate statistical techniques. *Eng. Geol.*, 1994, 38(1-2), 135-157.
- [16] Rigopoulos I., Tsikouras B., Pomonis P., Hatzipanagiotou K., Determination of the interrelations between the engineering parameters of construction aggregates from ophiolite complexes of Greece using factors analysis. *Constr. Build. Mater.*, 2013, 49, 747-757.
- [17] Zhang W.G., Goh A.T.C., Multivariate adaptive regression splines for analysis of geotechnical engineering systems. *Comput. Geotech.*, 2013, 48, 82-95.
- [18] Nandi A., Shakoor A., A GIS-based landslide susceptibility evaluation using bivariate and multivariate statistical analyses. *Eng. Geol.*, 2009, 110, 11-20.
- [19] Shicker R., Moon V., Comparison of bivariate and multivariate statistical approaches in landslide susceptibility mapping at a regional scale. *Geomorphology*, 2012, 161-162, 40-57.
- [20] Ramesh V., Anbazhagan S., Landslide susceptibility mapping along Kolli hills Ghat road section (India) using frequency ratio, relative effect and fuzzy logic models. *Environ. Earth. Sci.*, 2015, 73, 8009-8021.
- [21] Hajdarwish A., Shakoor A., Wells N.A., Investigating statistical relationships among clay mineralogy, index engineering properties, and shear strength parameters of mudrocks. *Eng. Geol.*, 2013, 159, 45-58.
- [22] Fodor L., Magyarai Á., Fogarasi A., Palotás K., Tercier szerkezetfejlődés és késő paleogén üledékképződés a Budai-hegységben. A Budai-vonal új értelmezése. [Tertiary tectonics and Late Palaeogene sedimentation in the Buda Hills, Hungary. A new interpretation of the Buda Line.]. *Földtani Közlöny*, 1994, 124(2), 129-305. (in Hungarian)
- [23] Raincsákné Kosáry Zs., A Budapest 4. sz. Metróvonal és környezetének földtani viszonyai [Geological setting of Budapest Metro Line 4 and its surrounding]. *Földtani Kutatás*, 2000, 37(2), 4-19. (in Hungarian)
- [24] Geovil Kft., Budapest 4. metróvonal, I. szakasz, Összefoglaló mérnökgeológiai, hidrogeológiai és geotechnikai szakvélemény, „A” kötet, Természetföldrajzi és földtani adottságok a nyomvonal mentén. Manuscript, Szentendre, Geovil Kft., 2005.
- [25] Bubics I., A budapesti metróépítés földtani eredményei. *Mérnökgeológiai Szemle*, 1978, 21, 5-87.
- [26] Raincsákné Kosáry Zs., Hermann V., Ollrád A., Végh H., A Dél-Buda - Rákospalota irányú 4. sz. metró-vonal földtani szakvélemény; Duna alatti átvezetési szakasz. Magyar Állami Földtani Intézet, Budapest, Report, 1998.
- [27] Bodnár N., Kovács J., Török Á., Using of Multivariate data analysis in Eng. Geol. at the Pest Side of the Metro Line 4 in Budapest, Hungary. - In: G. Lollino, D. Giordan, K. Thuro, C. Carranza-Torres, F. Wu, P. Marinos, C. Delgado (Eds.) *Engineering Geology for Society and Territory - Volume 6: Applied Geology for Major Engineering Projects*. Cham: Springer International Publishing, 2015, 851-854.
- [28] Schafarzik F., Szökevény hévforrások a Gellérthegy tövében. *Földtani Közlöny*, 1920, 3, 79-158.
- [29] Miller R.L., Kahn J.S., *Statistical Analysis is the Geological Sciences*. Wiley, New York, 1962.
- [30] Anderberg M.R., *Cluster analysis for applications*. Academic Press, New York, 1973, 359.
- [31] Stockburger D.V., *Multivariate Statistics: Concepts, Models and Applications*. Missouri State University, 1998.
- [32] Gross D.S., Atlas R., Rzeszutarski J., Turetsky E., Christensen J., Benzaid S., Olson J., Smith T., Steinberg L., Sulman J., Ritz A., Anderson B., Nelson C., Musicant D.R., Chen L., Snyder D.C., Schauer J.J., *Environmental chemistry through intelligent atmospheric data analysis*. *Environ. Modell. Softw.*, 2010, 25(6), 760-769.
- [33] Duda R.O., Hart P.E., Stork D.G., *Pattern Classification*. Wiley InterScience, New York, 2000.
- [34] McLachlan G., *Discriminant Analysis and Statistical Pattern Recognition*. Wiley InterScience, New York, 2004.
- [35] Kovács J., Tanos P., Korponai J., Kovácsné Székely I., Gondár K., Gondár-Sóregi K., Hatvani I.G., Analysis of Water Quality Data for Scientists. In: V. Kostas, V. Dimitra (Eds.), *Water Quality Monitoring and Assessment*. InTech Open Access Publisher, Rijeka, 2012, 65-94.
- [36] Norusis M.J., *SPSS for Windows Professional Statistics Release 6.0*. SPSS Inc., Englewood Cliffs, Prentice Hall, 1993.
- [37] Wu X.Z., Trivariate analysis of soil ranking - correlated characteristics and its application to probabilistic stability assessments in geotechnical engineering problems. *Soils Found.*, 2013, 53(4), 540-556.
- [38] Masoud A.A., Geotechnical evaluation of the alluvial soils for urban land management zonation in Gharbiya governorate, Egypt. *J. Afr. Earth Sci.*, 2015, 101, 360-374.
- [39] Ilia I., Rozos D., Perraki T., Tsangaratos P., Geotechnical and mineralogical properties of weak rocks from Central Greece. *Cent. Eur. J. Geosci.* 1(4), 2009, 431-442, DOI: 10.2478/v10085-009-0029-0.