

## GIS- BASED APPLICATION FOR GEOTECHNICAL DATA MANAGING

**Tsangaratos P.<sup>1</sup>, Koumantakis I.<sup>2</sup>, Rozos D.<sup>3</sup>**

<sup>1</sup> National University of Athens, School of Mining and Metallurgical Engineering, Department of Geology, Laboratory of Engineering Geology and Hydrogeology, Iroon Polytechniou, 15780, Zografou, Greece, [ptisag@metal.ntua.gr](mailto:ptisag@metal.ntua.gr)

<sup>2</sup> National University of Athens, School of Mining and Metallurgical Engineering, Department of Geology, Laboratory of Engineering Geology and Hydrogeology, Iroon Polytechniou, 15780, Zografou, Greece, [mmgski@central.ntua.gr](mailto:mmgski@central.ntua.gr)

<sup>3</sup> National University of Athens, School of Mining and Metallurgical Engineering, Department of Geology, Laboratory of Engineering Geology and Hydrogeology, Iroon Polytechniou, 15780, Zografou, Greece, [rozos@metal.ntua.gr](mailto:rozos@metal.ntua.gr)

### Abstract

*The need to provide data management capabilities in geotechnical projects, makes data visualization in a more understanding way vital, while improvements in computer science, have created an opportunity to rethink the manner in which such data is archived and presented. Geographic Information Systems are considered nowadays as principal methods for analysis, utilizing their ability of manipulating, compiling and processing spatial data, such as geotechnical one.*

*In this paper, the development of Borehole Analysis System (BAS) a specific Graphical User Interface (GUI) application is proposed to access geotechnical data with the aim of a relational database and an open source GIS platform, embodied in the application. The BAS, is able to integrate multiple layers of gathered information and to derive additional knowledge by applying statistical and data mining algorithms with the use of spatial query tools. These can give reasonable conclusions and better representation in 2-D and 3-D environment.*

*The presented application is illustrated with an example from field practice, testifying its ability to be a useful tool for management and presentation of geological and geotechnical borehole data.*

**Key words:** Geotechnical Data, Statistical Analysis, K-Means, GIS.

### 1. Introduction

The common practice was and in some extent still is, to report geotechnical data as borehole logs and laboratory spreadsheets as a supplement of geotechnical surveys that follows the planning, and designing of any civil engineering projects. 2-D graph or map is thought to be the basic tool of communication in geological science. However in the case of a borehole log, the most important document in a geological modelling, a 3-D environment is converted to a 2-D plane consequently with the loss of the third dimension, producing uncertain patterns of objects with similar properties. Geotechnical characterization of large civil engineering projects typically requires 3-D data such as stratification of soil types, elevation of water table, soil properties at various depths producing large

volumes of data that are difficult to manage and analyze (Adams, 1994). The applied procedure followed for manipulating multiple data from different locations involves, the opening of many different achieves, overlay series of papers, maps, photos, geotechnical surveys and spreadsheets, demanding a great amount of time and considerable effort. The common practice in order to be more productive implies averaging techniques with uncertain outcomes, which certainly influence the design and planning parameters of any project (Luna and Frost, 1998).

However, from that early practice, the introduction of Information Systems and Technology in geotechnical science had significantly changed the process of collecting, retrieving, analyzing and visualizing geotechnical data (Foster and Culshaw, 1990; Laxton and Becken, 1996; Bain and Giles, 1997; Pininska and Dzedzic, 2004; Denas et al, 2005; Lan and Martin, 2007). Moreover, Geographic Information System (GIS), has emerged to be a powerful computer-based technique that integrates spatial analysis, database management, and graphic visualization capabilities (MacCarthy and Graniero, 2006; Kaufmann and Martin, 2008). Many research tools and commercial software packages as those discussed by MacCarthy and Graniero (2006), have as main objective to discover similar patterns in spatially distributed geotechnical data. Those systems although they contain powerful features for statistical analysis, visualization and data exploration, in most cases they provide strict structural requirements for data input, limiting the user's freedom to structure his own data, while other software are developed on specific platforms or as an extension of commercial software narrowing its use and interoperation with other tools.

The Borehole Analysis System (BAS), a specific Graphical User Interface application, introduced in this paper, is presented as a tool to aid in the management, querying, data mining, and visualization of geotechnical data. It is capable to integrate multiple layers of gathered information, and to derive additional information visualizing them in a more easy-to-use way. This enables the production of maps with unlimited data, introduced in the form of attributes, layers and combinations of these. It fulfils the desire to visualize spatial patterns in data that may not be so obvious when presented in tabular format, establishing links with borehole and cross sections, processing spatially the data onto 2-D and 3-D environment. Finally data mining algorithms were applied, clustering data in such a manner that areas could be identified having certain characteristics or set of characteristics and seek knowledge from their locations and distribution.

## **2. Designing and compiling the application**

The target was to develop an easy-to-use open source application, not to replace existing modelling techniques and known commercial software, but rather to serve as a geotechnical data manager assistant. In this way expert would be able to determine the geotechnical behaviour of the formations in a given research area, using interactive displays and visual correlation, but also saving time and working productive with considerable less effort.

The basic functions that it should supply the users, includes: a) a tool to import data in specific template or by structuring data manual according to users needs (Data Input Manager), b) a tool to apply simple data mining algorithms along with classical statistical analysis to identify similar physical or mechanical characteristics and draw correlations (Statistical / Data Mining Manager), c) a tool to indicate queries (Query Builder Manager) and finally d) a tool to visualize the outcomes in 2-D and 3-D environment (Visualization Manager).

The application has a graphical interface and has been developed with the use of Visual Basic 6.0 (VB6), a third – generation event driven programming language. The relational database, that is

constructed to contain geological, geotechnical and other related data, was embodied in the application by Structured Query Language (SQL) connection. It's implemented on the Microsoft Jet Engine, storing data in Microsoft Access Database, guarding an easy access, management and updating of data. Clustering algorithms (K-Means algorithm), written in VB6 code, were also embodied to source code in order to cluster data with similar characteristics providing an indication of the engineering behaviour. Finally the data were able to be manipulated and spatial analyzed by using the MapWindow GIS ActiveX Control. MapWindow project started in 1998 at Utah Water Research Lab at Utah State University as an alternative to using ESRI MapObjects LT 1.0 for custom applications. It is also defined as an open source GIS (mapping) code and set of programmable mapping components, capable for storing and spatial analysis, as well as properly modified and embodied in the applications source code (Dunsford and Ames, 2008).

### **3. Basic Functions and tools**

Running BAS is practically a three step procedure. In the first step the user provides the application with all the available data, in the second step the application creates the needed files for analysis and in the final step the user is able to query, explore and analyse but also to visualize the data in 2-D or 3-D environment, through the appropriate functions and tools.

The obtained information is derived from existing borehole logs, field and laboratory test spreadsheets, geological, hydro-geological and geotechnical surveys, available topographic, hydrological, hydro-geological, and infrastructure information, aerial photos, boring photos, etc. The main objective during data pre-processing is to convert all the available data into digital format and import them to the system, such as maps, road infrastructure, topographic contours, and points of interest. Photos from boreholes are scanned and imported in the database. It is thought to be the most important phase as the quality and quantity of data determines the outcomes of the implementation of the system. The main functions and tools embodied in the BAS are described in more details in the following paragraphs.

#### **3.1 Data Input Manager**

The application provides a Data Input Manager which archives information concerning borehole identifiers and location data (x, y, z coordinates), such as columns containing sample depths, lithological and formation descriptions, columns with physical and mechanical properties, photos and scanned information maps. While using the Data Input Wizard the user has two options: to manually collect information by linking each data with the appropriate field names, or to import data using a specific template provided by the application. The database used, is built having five primary tables (project-info, borehole-info, physical-info, mechanical-info, description-info), which contained different types of information, categorical and numerical, but also spatial and non-spatial components. The key field to link each table was the Borehole Boring \_ID (Tsagaratos, 2008). After the completion of the data input process the application creates the appropriate shapefile and adds it to the projects table as data point files. In addition, the user is able to create his own shapefiles by selecting the attributes of interest applying either Data mining Manager or Query Builder Manager.

#### **3.2 Statistical and Data Mining Manager**

By the utilization of statistical analysis the application could provide an indication of the range of variation and the distribution of geotechnical parameters in each lithostratigraphic unit, obtaining me-

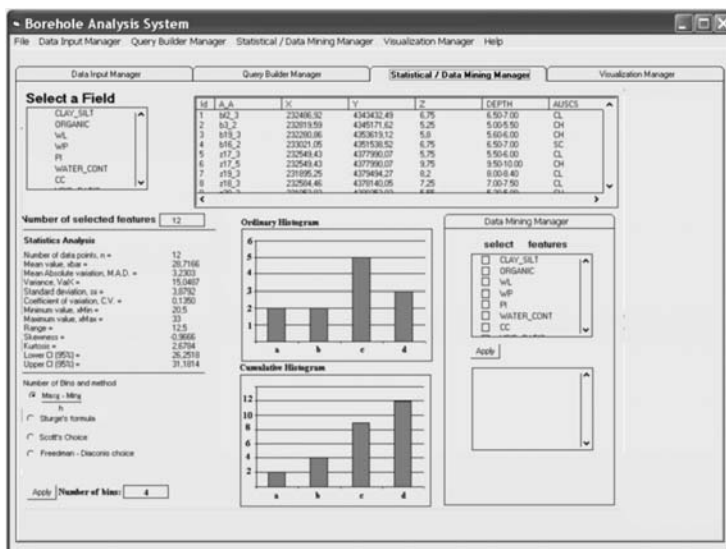


Fig. 1: Screenshot of Statistical / Data Mining Manager

dian, mode, mean, and standard deviation values. Moreover measuring asymmetry (skewness) and whether the data are peaked or flat relative to a normal distribution (kurtosis) is of valuable importance. High values of skewness and kurtosis indicate that data are not normally distributed, thus one could approximately expect that the variable in question tends not to cluster around the mean value. In addition more advanced statistical analysis could be initialized. Specifically by applying Pearson's test, the degree to which each one of the data parameters included in the database varies linearly with another could be calculated giving a more precisely knowledge of the geological environment. Moreover, Kendall's tau test (Daniel 1990), could be initialized when there is no linear relationship among the variables involved in the model. The Kendall' tau test measures the degree of concordance or discordance between two random variables, ranging also from -1 to +1, indicating respectively perfect, negative and positive correlation. The test can also be used to establish whether a random variable should be addressed as independent, if there is a significant spatial trend which would result in dependency (Jaska, 1995; Uzielli et al, 2004).

The statistical analysis and also graphical plots were used to prepare detailed reports and surveys describing the geotechnical properties of the materials, their engineering behaviour but also producing maps showing their distribution in the area of interest (Fig. 1).

The integration of data analysis and data mining techniques ultimately aims to the discovery of interesting, implicit and previously unknown knowledge hidden in the geotechnical database. The application provides an algorithm for clustering data. Clustering can be considered the most important unsupervised learning problem grouping sets of objects into clusters of similar objects. Applying these techniques on geotechnical data could be useful in identifying different lithology or formations with specific relative values and frequencies or distributions of certain attributes (Goktepe et al, 2005). In this application, K-Means algorithm (MacQueen, 1967) was utilized, which is considered as one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set (n observations) by implementing either the option that assumes a certain number of clusters (k clusters) fixed a priori,

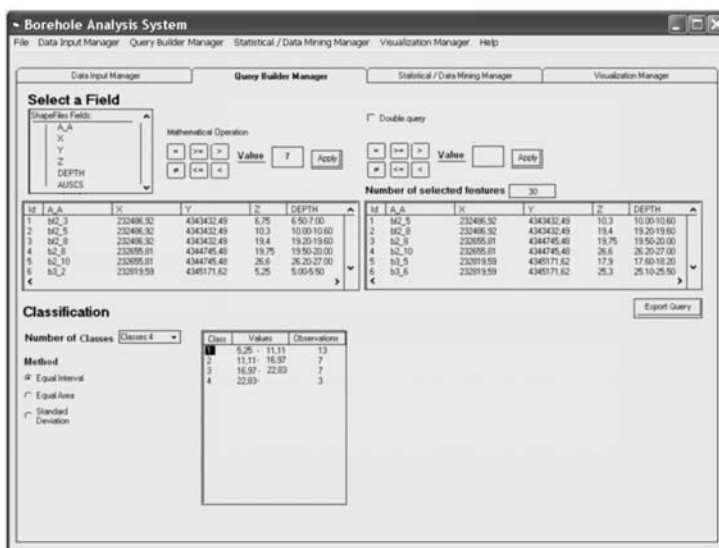


Fig. 2: Screenshot of Query Builder Manager

or the option of finding the optimum number through the data, in which each observation belongs to the cluster with the nearest mean.

### 3.3 Query Builder Manager

The Query Builder Manager along with the Statistical and Data Mining Manager are the most powerful tools in the developed application (Fig.2). The former can be helpful in identifying patterns of similar behaviour. It's used to combine categorical or numerical query terms into a single or complex query.

For example it could enable all boreholes that contain a particular formation or lie within a defined area, with relative information available at a point defined by grid coordinates, or within a defined lithostratigraphic unit. It could be also find a segment or lithostratigraphic unit with a particular set of physical or mechanical properties necessary to meet the design criteria for a specific engineering application. It's also the main tool to prepare the data for further analysis, as it builds the appropriate shapefiles.

### 3.4 Visualization Manager

The visualization tools were designed to visualize data, linking borehole photos, logs and cross-sections, physical and mechanical properties, based on a response to query parameters and data mining results or simply visualize the whole database (Fig. 3, 4). The application has also the ability of visualizing data in a 3D environment presenting the spatial distribution of the query or data mining results.

## 4. Applying the methodology and results

The database consisted of 39 boreholes data, derived from a civil engineering project located at the western part of Greece and involves the construction of the **Ionía Odos**, Ionian Highway. The total

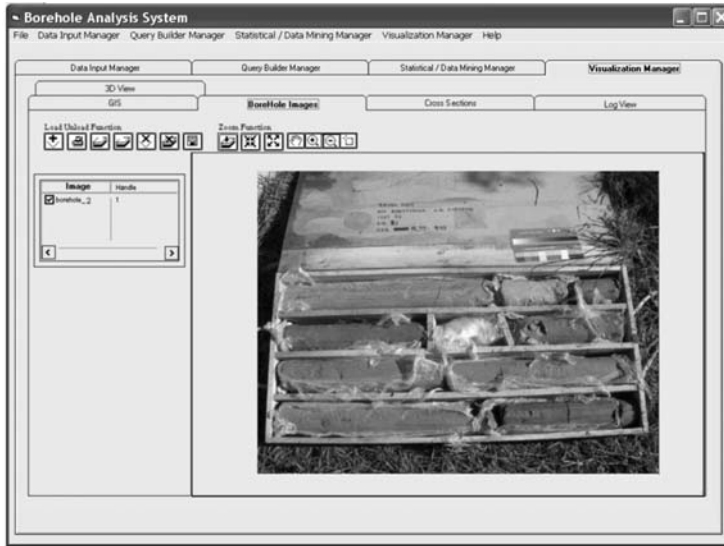


Fig. 3: Screenshot of Borehole Image / Visualization Manager

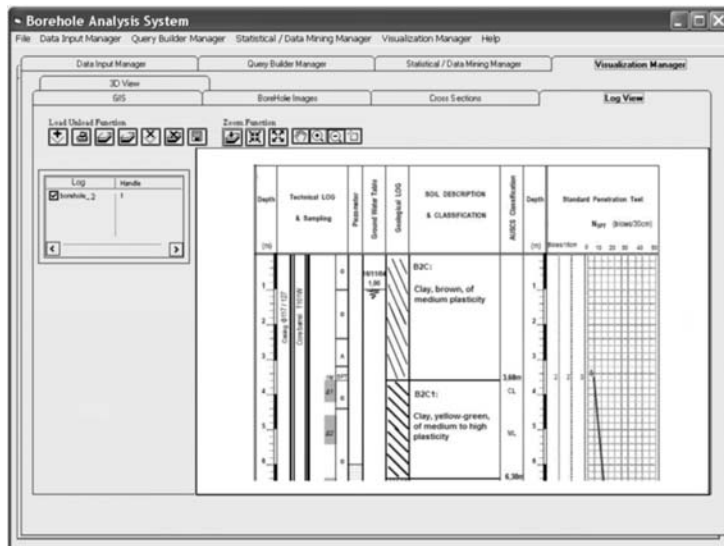
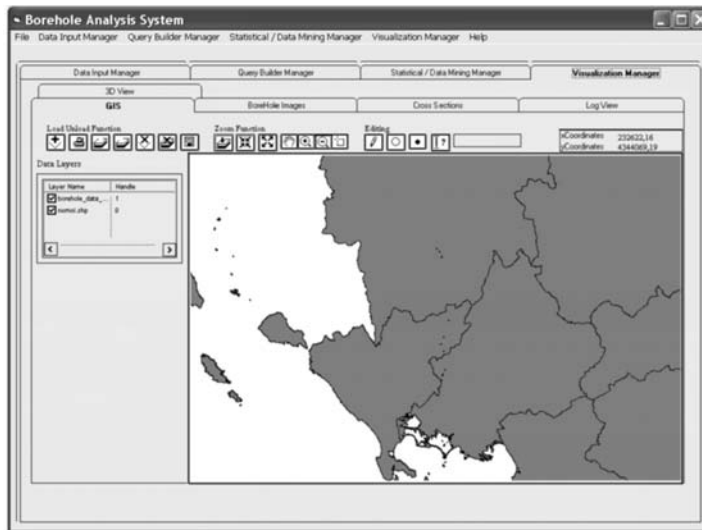


Fig. 4: Screenshot of Log View / Visualization Manager

length of the motorway will be 384 km, becoming part of the E55 highway in its full length. In the presented application we used data from the section that begins from Rio, Patras and ends at the town of Ioannina. It includes physical and mechanical properties obtained by engineering geological surveys, aerial photos, and log photos, collected and evaluated during the design phase of the project. After importing into the system the available data (1st phase), the aim was to quantify the degree of spatial variation of the physical and mechanical properties of the studied soils and thus obtain a more meaningful estimate at unsampled locations but also to provide input to reliability analyses. The main





**Fig. 5:** Screenshot of GIS View / Visualization manager

steps were: a) perform statistical analysis for the entire database, describing the variation that the geotechnical parameters have and also using the Query Builder Manager to perform statistical analysis in specific areas, b) find the correlations between the variable used in the analysis, c) determine which variables would be included in the data mining process, and finally d) interpret the results in accordance with the available geotechnical surveys, evaluate and provide additional knowledge.

The mean value and the standard deviation, considered as the best estimate and the uncertainty of the best estimate respectively. In descriptive statistics the coefficient of variation, appears to be dimensionless. In most cases, it provides a more physically meaningful measure of dispersion relative to the mean value. When comparing between data sets with different units or widely different means, the coefficient of variation is better to be used instead of the standard deviation as can be illustrated below. Table 1 shows the results of the descriptive statistical analysis.

It seems that parameter of Plasticity Index PI, a parameter of great importance in geotechnical engineering, is the one with the higher value of coefficient of variation, while parameter Plastic Limit ( $W_p$ ) the one with the lowest, having great spatial variation and varying behaviour response within the same area. The knowledge of variation for each parameter is more reliable when referred to each lithostratigraphic unit or within a specific segment of known depth. To focus our analysis in those areas the Query Builder Manager was used, following either the classification used in the log files, and applying the statistical analysis in the lithostratigraphic unit CH, or producing segments of certain depth, by applying simple methods of classification (equal intervals, equal area, and standard deviation). The lithostratigraphic unit CH refers to inorganic clay of high plasticity. It is used in the analysis as an example since it appears as the most dominate unit throughout the research area. During the first analysis, for the entire database, PI appeared with the highest value of coefficient of variation. However, it does not appear as such, within the CH lithostratigraphic unit, but also within the 10m depth segment that was produced by the Query Builder. From the above analysis it is obvious that each parameter varies differently in respect to different lithostratigraphic units, or within different segments of certain depth.

Table 1.

<b>Statistical Analysis applied to the entire database</b>						
<b>Parameters</b>	<b>W<sub>L</sub></b>	<b>W<sub>p</sub></b>	<b>P<sub>I</sub></b>	<b>W<sub>c</sub></b>	<b>C<sub>c</sub></b>	<b>V<sub>r</sub></b>
Mean Value	45,38	23,40	21,98	26,536	0,153	0,723
Variance	168,24	14,394	115,719	47,588	3,3711E-02	3,2724E-02
Standard Deviation	12,97	3,794	10,757	6,898	5,8061E-02	0,1808
Coefficient of variance	0,285	0,1620	0,4894	0,2599	0,3774	0,2501
Minimum	20,2	14,1	1,3	8,3	0,051	0,238
Maximum	66,2	32,3	36,2	42,6	0,279	1,15
Range	46	18,2	34,9,6	34,3	0,228	0,911
Skewness	-0,3429	0,1628	-0,4215	-0,4418	0,2281	-0,3349
Kurtosis	2,1118	3,1251	1,8546	3,5297	2,2985	3,5596
<b>Statistical Analysis applied to the CH lithostratigraphic unit</b>						
<b>Parameters</b>	<b>W<sub>L</sub></b>	<b>W<sub>p</sub></b>	<b>P<sub>I</sub></b>	<b>W<sub>c</sub></b>	<b>C<sub>c</sub></b>	<b>V<sub>r</sub></b>
Mean Value	57,207	25,615	31,593	29,115	0,174	0,785
Variance	25,672	7,963	12,184	3,265	3,4673E-03	1,3711E-02
Standard Deviation	5,0667	2,8218	3,4905	4,2537	5,9048E-02	0,117
Coefficient of variance	8,8568E-02	0,1101	0,1104	0,1461	0,338	0,149
Minimum	50,4	22,2	24,6	18,5	0,057	0,51
Maximum	66,2	32,3	36,2	34,9	0,279	0,955
Range	15,8	10,1	11,6	16,4	0,222	0,455
Skewness	0,23	0,796	-0,3298	-1,0269	8,047E-03	-0,7931
Kurtosis	1,7177	3,2236	0,1614	3,7637	2,5736	3,2263
<b>Statistical Analysis applied to the 10m depth segment</b>						
Mean Value	50,583	24,458	26,125	28,716	0,152	0,776
Variance	62,765	13,22447	37,53296	15,04879	3,0666E-02	1,1636E-02
Standard Deviation	7,922	3,636	6,126	3,879	5,5377E-02	0,107
Coefficient of variance	0,1566	0,1486	0,2345	0,135	0,3631	0,1389
Minimum	38,1	20,8	15,1	20,5	0,057	0,54
Maximum	66,2	32,3	35,4	33,0	0,235	0,908
Range	28,1	11,5	20,3	12,5	0,178	0,368
Skewness	0,3554	0,8821	-0,2967	-0,9666	0,1006	-0,8570
Kurtosis	2,4855	2,4761	2,0671	2,6784	2,0558	2,6982

W<sub>L</sub>: Liquid Limit, W<sub>p</sub>: Plasticity Limit, P<sub>I</sub>: Plasticity Index, W<sub>c</sub>: Water Content, C<sub>c</sub>: Compression Index, V<sub>r</sub>: Void ratio



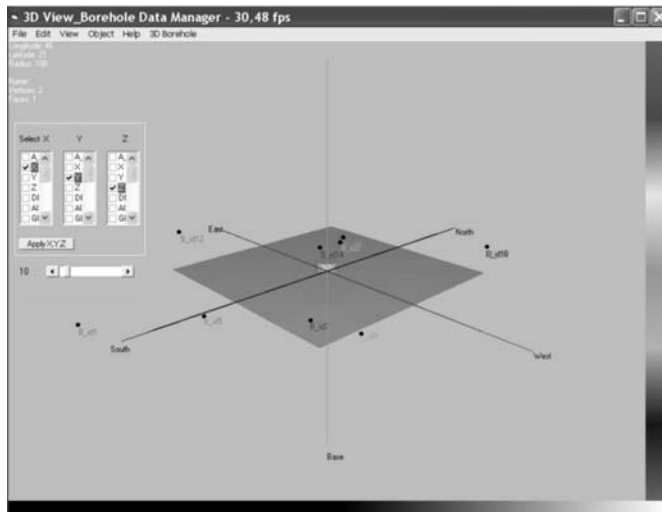


Fig. 6: Screenshot of 3D View / Visualization manager

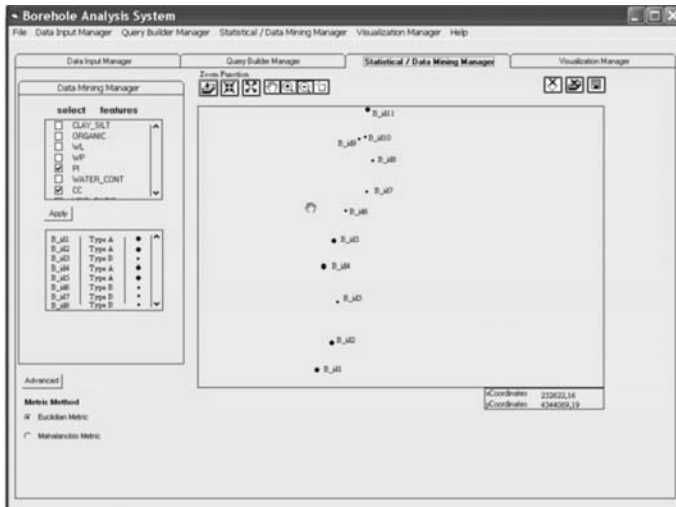


Fig. 7: Visualizing the Data Mining Manager results

By the end of each performed analysis the BAS provides the user within a second the statistical reports and exports data into word or excel format. Moreover the results were encoded as shapefiles and introduced in the Visualization Manager. There the user could produce 2D views of the spatial distribution of the borehole database, showing the values of each parameter in a formulated list-view object (Fig. 5). When ever the user needs to visualize the log photos or the log profile, he could do so just by clicking in the appropriate tab, which illustrates the image data.

The application can commute the features into 3D shapefiles, where x, y, z values are imported by the user, while the application converts those values in such a manner that the new values of x,y,z are within a space of [-1, 1]. This conversion provides a better visualization perspective of the boreholes location and the spatial variances of the geotechnical parameters (Fig.6).

The problem of classifying formations according to certain variables becomes more complicated in areas where few experimental data exist. However data mining algorithms could server as an appropriate tool, in order to perform clustering and classification (Chen et al, 1996; Kanungo et al, 2002). In our case the existing data could be used to identify areas with similar characteristics by clustering them into classes that appear minimum coefficient of variance within each class and maximum coefficient of variance among classes. The outcomes of the initialization of data mining algorithms, were in accordance with those described by the geotechnical surveys and indicate that certain areas could be identify having similar geotechnical profile, according to their Euclidean distance (Fig.7).

## 5. Conclusions

The developed application is based on structuring and storing in digital formats the various geotechnical data of different types (borehole logs, cross-section, geological maps, and in-situ and laboratory test parameters). It can also be used to answer specific enquiries concerning the variables that characterized the material in question, to aid to the production of engineering geological maps, geotechnical summaries, and assessments of engineering behaviour for the soils and rocks of an area.

It could be used as a simple querying tool to answer simple questions about geological and geotechnical characteristics in a spatial context or as an advanced investigation tool, identifying spatial patterns in their data based on any attribute through querying, data mining and visual explore them in 2-D and 3-D environment.

The application of data mining algorithms in the available geotechnical database proved to be a very powerful technique, identifying spatial patterns of similar geotechnical parameters that were much more difficult to identify following the classical procedure applied in geotechnical investigation.

The final product is for use both to engineers and geologists, providing a better understanding of engineering geological conditions, but also to planners identifying areas that may geologically constrain development.

## 6. References

- Adams M., Bosscher J., 1994. Integration of GIS and knowledge-based systems for subsurface characterization. *In ASCE Monograph on Integration of Expert Systems*, eds. M.L. Maher and I.Tommelein, ASCE.
- Ames, D.P., C. Michaelis, A. Anselmo, L. Chen, and H. Dunsford, 2008. *MapWindow GIS. Encyclopedia of GIS*. New York, Sashi Shekhar and Hui Xiong (eds). Springer, 633-634.
- Bain A., Giles A., 1997. A standard model for storage of geological map data. *Computers & Geosciences* Vol.23, No 6, 613-620.
- Chen M.-S., Han, J., Yu, P.S., 1996. Data mining: An overview from a database perspective. *IEEE Transactions on Knowledge and Data Engineering*, Vol.8, Issue 6, 866-883.
- Daniel, W. W., 1990. *Applied non parametric statistics*. Boston, 2nd edition: PWS-Kent, 635pp.
- Denas Z., Belickas J., Sliupa S., 2005. Intergrated approach and application of GIS for management of geological data. *The Current Role of Geological Mapping in Geosciences*, S. Ostaficzuk (eds), 135–150.
- Foster A., Culshaw G., 1990. The Use of Site Investigation Data for Preparation of Engineering Geological Maps and Reports for Use by Planner and Civil Engineer. *Engineering Geology*, Vol.29, 347-354.
- Goktepe B., Altun S., Sezer A., 2005. Soil clustering by fuzzy c-means algorithm, *Advances in Engi-*

- neering Software*, Vol.36, Issue 10, 691-698.
- Jaksa, M.B., 1995. The Influence of Spatial Variability on the Geotechnical Design Properties of a Stiff, Overconsolidated Clay. Ph.D. Thesis, University of Adelaide.
- Kanungo, T., Mount, D.M., Netanyahu, N.S., Piatko, C.D. Silverman, R., Wu, A.Y., 2002. An efficient k-means clustering algorithms: Analysis and implementation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.24, Issue 7, 881-892.
- Kaufmann O., Martin T., 2008. 3D geological modelling from boreholes, cross section and geological maps, application over former natural gas storages in coal mines. *Computers & Geosciences*, Vol. 34, 278-290.
- Lan H., Martin D., 2007. A digital approach for integrating geotechnical data and stability analyses. *Rock Mechanics: Meeting Society's Challenges and Demands*, E.Eberhardt, D. Stead & Morrison, T. (eds), London, Taylor and Francis Group.
- Laxton L., Becken K., 1996. The Design and implementation of a spatial database for the production of geological maps. *Computers & Geoscience*, Vol. 22, 123-733.
- Luna R., Frost D., 1998. Spatial Liquefaction Analysis System. *Journal of Computing in Civil Engineering*, ASCE, Vol. 12, No. 1, 48-56.
- McCarthy D., Graniero A., 2005. A GIS-based borehole data management and 3D visualization system. *Proceedings of the 2005 Annual Conference of the International Association for Mathematical Geology, Toronto*, 267-272.
- MacQueen B., 1967. Some Methods for classification and Analysis of Multivariate Observations. *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, University of California Press, 1:281-297.
- Pininska J., Dziedzic A., 2004. GIS application for geomechanics—a Polish example, *Bulletin of Engineering Geology and the Environment*, Vol.63, No 1, 83-87.
- Tsagaratos P., 2008. Development of a graphical user interface (GUI) application for the management of borehole data, *Geophysical Research Abstracts*, Vol. 10, EGU2008-A-06732.
- Uzielli M., Vannucchi G., Phoon K., 2004. Assessment of weak stationarity using normalized cone tip resistance. *Proceedings of the 9th ASCE Specialty Conference on Probabilistic Mechanics and Structural Reliability, Albuquerque, New Mexico, July 26–28*.