ASSESSMENT OF GROUNDWATER QUALITY
AND HYDROGEOLOGICAL PROFILE OF KAVALA AREA,
NORTHERN GREECE

THOMAS SPANOS1, ANTOANETA ENE2,*, CHRISTINA XATZIXRISTOU1,
AGELOS PAPAIOANNOU3

1Eastern Macedonia and Thrace Institute of Technology, Department of Petroleum and Mechanical
Engineering Sciences, St. Lucas, 654 04 Kavala, Greece
2, * Dunarea de Jos University of Galati, Faculty of Sciences and Environment, Department of
Chemistry, Physics and Environment, 47 Domneasca St., 800008 Galati, Romania;
E-mail: aene@ugal.ro
3Technological Educational Institute of Larissa, Department of Medical Laboratories, 41110
Larissa, Greece

Received December 29, 2014

In the present study thirty two representative sites have been selected for the
collection of water samples from central water supplies in the region of Kavala,
Northern Greece. Water source sites were identified depending on the geographical
location. Seven physicochemical parameters (electric conductivity, chloride, sodium,
total hardness, total alkalinity, bicarbonate and calcium) were analyzed monthly, in the
period from January 2010 to December 2010. Hierarchical cluster analysis grouped the
384 cases (32 × 12) of observation into three clusters (1: coastal, 2: lowland and 3:
semi–mountainous) based on the similarities of potable water quality characteristics
specific to geological origin. The mean plots of variables (Ward’s method) allowed the
membership assignment of each parameter in the three–cluster solution. The
classification scheme obtained through cluster analysis was confirmed by ANOVA.
The mean values of the studied physicochemical parameters were found to be within
the limits given in the 98/83/EC Directive. The water samples are appropriate for
human consumption. The results of this study provide an overview of the
hydrogeological profile of water supply system for the studied area.

Key words: Groundwater quality, Prefecture of Kavala, public health, hierarchical
cluster analysis, ANOVA.

1. INTRODUCTION

Groundwater is one of the most important major natural resources necessary for
human consumption, domestic services, agriculture, industry, manufacturing and other
sectors [1]. The quality of the groundwater receivers is influenced by pollution of soil
and air, industrial and domestic waste disposal, organic components, pathogenic
microorganisms, application of fertilizers and pesticides in agriculture, etc. [2–4].
In view of the importance for public health of water for human consumption, the European Union (EU) laid down various quality standards with which potable water must comply, establishing a number of Council Directives: 98/83/EC [5], 2000/60/EU [6], 2006/118/EU [7], that refer to the protection of the water resources and public health. Careful studies on the quality of water intended for human consumption could significantly contribute to the creation of a common strategy for potable water assessment [8]. During monitoring and assessment of water quality, large data sets could be collected, which contain rich information about the behavior of the potable water supplies and the potable water characteristics [1].

Although the values of the physical and chemical parameters represent an important monitoring tool of the ecological status of a water system [9–11], the simple comparison of monitoring data with allowable values of hazardous chemicals is univariate and this approach does not correspond to reality, due to the multivariate dimension of the nature [12]. It is required to take into account many environmental parameters and interpret information based on the complex relationships between various chemicals in order to achieve a reliable pathway of sustainable development [12].

The classification, modeling and interpretation of the monitoring data is a very important step in the complete assessment of the quality of the water intended for human consumption. In the last decades, multivariate statistical methods such as cluster analysis (CA), Principal Component Analysis (PCA), Discriminant Analysis (DA), Factor Analysis (FA), have demonstrated their utility in detecting similarities in site locations and identifying temporal trends and sources of pollution in various environmental systems [1, 3, 12–21]. They proved to be an excellent exploratory tool for interpreting complex water quality data sets and for understanding spatial variations, which are useful and effective for water quality management. By the use of multivariate statistical approaches one is able to derive hidden information from the data set regarding the possible influences of the environment on the water quality and, thus, various water sources could be carefully compared according to their physicochemical characteristics [8]. Also, based on statistical methods, a territory could be divided into areas with distinct groundwater qualities, thus aiding to the management and future development of groundwater sources in that region [22].

In the present study, the geological profile of the groundwater in the region of Kavala Prefecture, Northern Greece, and the quality of groundwater of this area have been determined. The goal of this study is to apply multivariate statistical approaches to datasets from different sampling sites \( n = 32 \) in order to certify the quality of water intended for human consumption. Our intent was to find the correlations between the sampling sites and the variables obtained by physical and chemical measurements, which could be helpful to official authorities in building a fast and reliable decision model for separating sources with different water quality and optimizing the potable water monitoring plan [23].
2. EXPERIMENTAL

2.1. STUDIED AREA AND SAMPLING

Water quality monitoring data were collected monthly between January 2010 and December 2010. The selection of monitoring parameters was mainly based on the requirements of the European directives 98/83/EC [5] and 2000/60/EU [6]. The samples were collected from thirty two sampling sites and seven physical and chemical parameters have been investigated, corresponding to one year of monitoring with a monthly frequency of analytical determination for all 384 cases (32 × 12) of observation.

The location of the potable water sampling sites (1 N. Peramos, 2 N. Iraklitsa, 3 Eleftheres, 4 Eleochori, 5 Keramoti, 6 Monastiraki, 7 Chaidefto, 8 Piges, 9 Agiasma, 10 Chrisoupoli, 11 Pontolivado, 12 Gravuna, 13 Zarkadia, 14 Xerias, 15 Dialektos, 16 Krini, 17 Chrisochori, 18 Eratino, 19 Moustheni, 20 Mesia, 21 Sidirochori, 22 Domatia, 23 Mesoropi, 24 Platanotopos, 25 Folia, 26 Galipsos, 27 Podochori, 28 Kokkinochori, 29 Nikisiani, 30 Georgiani, 31 Ag. Xristoforos, 32 Paleochori) is presented in Fig. 1. They are in the geographical area of Kavala, situated between the Strimon and Nestos rivers in Northern Greece, in different geographical (and geological) locations: close to Aegean Sea coast, in mountainous regions, in lowlands or close to rivers.

Fig. 1 – The location of the sampling sites in Northern Greece.
2.2. PARAMETERS AND ANALYTICAL METHODS

The physicochemical parameters determined and analyzed in groundwater samples are the following: electric conductivity (EC, 25°C, μS/cm), total hardness (TH, mg/L CaCO₃), total alkalinity (TA, mg/L HCO₃⁻ and CO₃²⁻), chloride (Cl⁻, mg/L), bicarbonate (HCO₃⁻, mg/L), sodium (Na⁺, mg/L) and calcium (Ca²⁺, mg/L).

The methods used for the determination of electric conductivity, chloride and sodium and the conditions for sampling, sample preparation, instrument calibration, limits of detection and procedure uncertainty were standard procedures as recommended by the Council Directive 98/83/EC [5]. The methods used for the determination of total hardness, total alkalinity, bicarbonate and calcium are described in the Council Directive 80/778/EC [24]. The analytical measurements were performed in the Laboratory of Instrumental Analysis, Department of Applied Sciences, Technological Educational Institute (TEI) of Eastern Macedonia and Thrace, Kavala, Greece as described in [8]. Potentiometric methods were applied for electric conductivity using InoLab WTW measurements; flame photometry was used for determination of calcium and sodium with the aid of a PFP7 JENWAY flame photometer; chloride was measured by UV-spectroscopy using a Hitachi UV-VIS U-2000 spectrophotometer; auto-titrator METTLER TOLEDO 50X was used for the measurement of total alkalinity and bicarbonate, and for total hardness by complexometry.

The complete datasets are available on request from the authors.

2.3. STATISTICAL METHODS

The statistical approaches cluster analysis (CA) and analysis of variances (ANOVA) were used in order to classify, model and interpret the data for the water quality. The software package used in this study was SPSS 19 [25].

The aim of CA is to group objects based on their features. Essentially, CA classifies objects characterized by a set of variables so that each object is very close to each other in the same cluster [25]. Hierarchical CA is the most common approach used for object classification, which makes it possible to detect similarities or dissimilarities within a large group of objects [20, 26–28]. Basic requirements for applying CA are the following: the variables are of the same type - quantitative, frequencies or binary variables; the variables are measured on comparable scales. In order to avoid classification problems with objects described by variables of completely different size, in the preliminary step of the classification the input data matrix (n objects x m variables) is normalized to dimensionless values (by the use of autoscaling or z-transform, range scaling, logarithmic transformation) which replace the real data values, reducing them to close numbers [12, 18]. Then, a similarity measure is applied to calculate the distance between all objects of interest. Very often the Euclidean distance
(ordinary, weighted, standardized) is used as a reliable measure of similarity between the classified objects [8, 12, 13, 18]. Finally, an appropriate linkage algorithm (single, average, centroid, Ward’s linkage etc.) is used to link the objects into a group (cluster) with similar distance and to separate those located at large distances [8]. Ward’s method is the most common way of calculating the distance between two clusters, as all the “differences” of all the members of a group are measured [29].

The representation of the results of CA is usually performed by a tree-like scheme called dendrogram comprising a hierarchical structure (large groups are divided into small ones) [12]. Dendrograms can be used to assess the cohesiveness of the clusters formed and can provide information about the appropriate number of clusters to keep [25].

The One-Way ANOVA procedure produces a one-way analysis of variance for a quantitative dependent variable by a single factor (independent) variable. Analysis of variance is used to test the hypothesis that several means are equal. It displays an analysis-of-variance table which includes univariate F tests for each clustering variable and the final cluster assignment for each case, based on the Euclidean distance between the cases and the cluster centers used to classify the cases. The ANOVA table indicates which variables contribute the most to the cluster solution; variables with large mean squares errors provide the least help in differentiating between clusters [25].

3. RESULTS AND DISCUSSION

Using Hierarchical Cluster Analysis (HCA) ($z$-transformed input data, squared Euclidean distance as similarity measure, Ward’s method of linkage) for all 384 cases of observations (32 sampling sites, 7 physicochemical parameters, 1 year monitoring, monthly frequency of analytical determinations), our data were grouped into three major groups. The results as depicted in Fig. 2 (dendrogram) prove the existence of the three clusters in the investigated region. More specifically, the clusters are the following:

- **Cluster 1**: the coastal area with 6 sites (Keramoti, Chaidefto, Monastiraki, Agiasma, Piges and N. Peramos),
- **Cluster 2**: the lowland area with 15 sites (Xerias, Eratino, Chrisochori, Pontolivado, Krini, Gravuna, Zarkadia, Chrisoupoli, Dialekti, Mesia, Platanotopos, Sidirochori, Mesoropi, Moustheni and Domatia) and
- **Cluster 3**: a semi–mountainous region with 11 sites (Ag. Xristoforos, Paleochori, Nikisiani, Georgiani, Galipsos, Kokkinochori, N. Iraklitsa, Podochori, Eleftheres, Eleochori and Folia).
Fig. 2 – Hierarchical dendrogram for sites clustering (Ward linkage; 32 sites and 7 variables).

The calculated basic statistical results for the thirty two sampling sites and for the seven parameters analyzed monthly, such as mean, median, standard deviation (SD), minimum, maximum, and range (the maximum minus the minimum value), are presented in Table 1 for each cluster.

First results show that for almost all variables, the minimum and maximum values differ due to the large range of values. After defining the groups, it is advisable to measure the distance within each group. This is achieved with the Ward method that measures the distance between the centers of groups and multiplies it by a factor. It should not be ignored that this method is only for quantitative data and serves to minimize information loss that occurs when our data create clusters.
Table 1
Descriptive statistics for the groundwater physical and chemical parameters for the three clusters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cl⁻</td>
<td>124.11</td>
<td>125</td>
<td>55.41</td>
<td>67</td>
<td>218</td>
<td>151</td>
</tr>
<tr>
<td>EC</td>
<td>993.66</td>
<td>976.5</td>
<td>203.69</td>
<td>802</td>
<td>1305</td>
<td>503</td>
</tr>
<tr>
<td>Na⁺</td>
<td>180.16</td>
<td>180.5</td>
<td>24.86</td>
<td>163</td>
<td>220</td>
<td>57</td>
</tr>
<tr>
<td>TA</td>
<td>159.3</td>
<td>163.2</td>
<td>19.04</td>
<td>122.6</td>
<td>174</td>
<td>51.4</td>
</tr>
<tr>
<td>TH</td>
<td>8.56</td>
<td>4.45</td>
<td>10.52</td>
<td>3.5</td>
<td>30</td>
<td>26.5</td>
</tr>
<tr>
<td>HCO₃⁻</td>
<td>194.25</td>
<td>199.15</td>
<td>23.18</td>
<td>149.5</td>
<td>212</td>
<td>62.5</td>
</tr>
<tr>
<td>Ca²⁺</td>
<td>8.66</td>
<td>2</td>
<td>16.34</td>
<td>1</td>
<td>42</td>
<td>41</td>
</tr>
</tbody>
</table>

Cluster 2

| Cl⁻       | 19.35 | 11.08 | 19.62 | 2.1 | 56.1 | 54 |
| EC        | 488.81 | 531 | 134.47 | 314 | 680 | 366 |
| Na⁺       | 15 | 11 | 11.84 | 2 | 218 | 216 |
| TA        | 101.24 | 106 | 23.56 | 67.7 | 136 | 68.3 |
| TH        | 18.75 | 22.3 | 6.39 | 6.7 | 26.5 | 19.8 |
| HCO₃⁻     | 123.46 | 129.32 | 28.81 | 82.6 | 166 | 83.4 |
| Ca²⁺      | 37 | 37 | 7.57 | 26 | 46 | 20 |

Cluster 3

| Cl⁻       | 5.55 | 5 | 2.92 | 2 | 14 | 12 |
| EC        | 399.73 | 424 | 62.16 | 231 | 486 | 255 |
| Na⁺       | 6.86 | 4.7 | 4.57 | 2.5 | 18 | 15.5 |
| TA        | 199.3 | 210 | 33.41 | 125 | 267 | 142 |
| TH        | 82.47 | 22.2 | 80.43 | 20.6 | 232 | 211.4 |
| HCO₃⁻     | 242.9 | 256 | 40.44 | 153 | 325 | 172 |
| Ca²⁺      | 43.54 | 48 | 11.91 | 16 | 65 | 49 |

SD = standard deviation; EC is expressed in μS/cm and the other parameters in mg/L.

After the definition of the clusters, the similarity between them was measured. It is obvious from Table 1 that the value differences between the clusters are very large and thus we conclude that the clusters are well formed. Also, the standard deviation between the different clusters for the same parameter is equally great. This is a first indication of the inability to express an opinion on all of our measurements and we are therefore forced to classify them in clusters.

Completing our statistical analysis, we compare the mean values of the separate samples. Table 1 identified the existence of three clusters and this is confirmed by using ANOVA for the analysis of variables. In Table 2 we observe the variables $F$ (Mean Square Regression/Mean Square Residual) and
Sig. (Significance) which confirm the differentiation of clusters. The null hypothesis for the control of average values is to be equal between subgroups. The null hypothesis is rejected when the Sig. of $F$ is less than 0.05. It could be seen from Table 2 that all values are 0.000, with one exception (TH) in which Sig. is 0.008, which compels us to reject the null (initial) hypothesis and accept the alternative of an existing significant difference between the averages of the clusters. According to the results of Table 2 we adopt the initial hypothesis at a significance level 0.05.

Table 2

Variables $F$ & Sig. for the investigated groundwater parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$F$</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cl$^-$</td>
<td>47.381</td>
<td>0.000</td>
</tr>
<tr>
<td>EC</td>
<td>51.09</td>
<td>0.000</td>
</tr>
<tr>
<td>Na$^+$</td>
<td>473.604</td>
<td>0.000</td>
</tr>
<tr>
<td>TA</td>
<td>38.474</td>
<td>0.000</td>
</tr>
<tr>
<td>TH</td>
<td>5.782</td>
<td>0.008</td>
</tr>
<tr>
<td>HCO$_3^-$</td>
<td>38.739</td>
<td>0.000</td>
</tr>
<tr>
<td>Ca$^{2+}$</td>
<td>19.707</td>
<td>0.000</td>
</tr>
</tbody>
</table>

For a better presentation of the diversity of clusters, Fig. 3a–g shows the graphs for parameter mean plots depicting the three averages of each cluster using Ward method.

Concluding the findings of the statistical analysis of our measurements, in Table 3 we determined the cluster memberships [25], highlighting what parameters are indicative for each cluster.

Table 3

Cluster memberships

<table>
<thead>
<tr>
<th>Case</th>
<th>3 Clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cl$^-$</td>
<td>1 coastal</td>
</tr>
<tr>
<td>EC</td>
<td>2 lowland</td>
</tr>
<tr>
<td>Na$^+$</td>
<td>1 coastal</td>
</tr>
<tr>
<td>TA</td>
<td>3 semi-mountainous</td>
</tr>
<tr>
<td>TH</td>
<td>1 coastal</td>
</tr>
<tr>
<td>HCO$_3^-$</td>
<td>3 semi-mountainous</td>
</tr>
<tr>
<td>Ca$^{2+}$</td>
<td>1 coastal</td>
</tr>
</tbody>
</table>
Fig. 3 – Mean plots of water parameters (Ward method) in the three clusters (1 – coastal; 2 – lowland; 3 – semi-mountainous): a) EC; b) Cl⁻; c) Na⁺; d) TA; e) TH; f) HCO₃⁻; g) Ca²⁺.
Interestingly, for the lowlands we have found only one indicator parameter of the seven investigated parameters, e.g. electric conductivity (EC), while for the coastal areas we have found four parameters (Cl\(^-\); Na\(^+\); TH and Ca\(^{2+}\)) and for the semi-mountainous region two parameters (TA and HCO\(_3^--\)).

The comparison of mean values among the studied areas (Fig. 3a–g) shows that there is a significant difference between the three clusters. This can be attributed to the geological differences among areas and to the different use of land. However, in the studied area, the mean values of all studied physicochemical parameters in all sites (Table 1) did not exceed the acceptable limits mentioned in the 98/83/EC Directive (EC, 2500 μS/cm; Cl\(^-\), 250 mg/L; Na\(^+\), 200 mg/L) [5].

In addition, we can see in the Table 1 for the first cluster (coastal areas), influenced by the intrusion of the seawater, that the values of electric conductivity, chloride ions and sodium exhibit a great difference compared to the other two clusters. For the third cluster (semi-mountainous areas) the values of alkalinity and hardness are very high again in comparison to the other two clusters.

One finding of particular interest, resulting from this analysis, is that some areas, such as N. Iraklitsa, which on the map are located by the sea and would be expected to belong to Cluster 1 (coastal), actually belong to Cluster 3 (semi-mountainous) because their water source originates from the mountainous area and, therefore, is not influenced by the sea.

In Greece, because of the calcium carbonate constitution of the mountainous regions through which the natural waters pass, usually high values of hardness, calcium, alkalinity and conductivity are present.

The cluster analysis has provided a useful classification of the groundwater in the study area, which could be used to design an optimal future spatial monitoring network with lower costs [14]. It seems more important to organize several well-equipped and monitored sites from the obtained categories (patterns) instead of distributing many sites with similar pattern of quality [12]. Sampling at fewer existing sites, or for a reduced set of existing analytes, may allow for additional sampling at other currently unsampled sites and for currently unsampled analytes [1]. On the basis of our results, the number of monitoring sites in Kavala area, Northern Greece, could be reduced and only chosen from clusters “coastal”, “lowland” and “semi–mountainous”, using a reduced set of indicator parameters for each cluster.

4. CONCLUSIONS

Management of the water quality intended for human consumption requires not only the strict monitoring of the physicochemical parameters of the water sources and the potable water as offered by the instructions and directives of EU,
but also the involvement of environmental methods as standard procedures for a
real and low-cost assessment of the groundwater resources in a certain region.

Using multivariate statistical approaches like hierarchical cluster analysis
(HCA) one is able to extract hidden information from the multidimensional data set
about the possible influences of the environmental factors on the water quality and
various water sources could be compared.

The assessment study carried out in this work leads to important additional
information about the water sources in Northern Greece and can assist in important
scientific or political decisions. The differences in the physicochemical parameters
due to geological origin have been objectively detected by statistical analysis.
Multivariate statistical analysis first involved the application of a typical
classification approach, such as HCA, that resulted in three main clusters of
sampling sites with specific properties of groundwater. It is noteworthy that
ANOVA analysis of variance confirmed exactly the results obtained by HCA.

On the basis of our results, the number of monitoring sites in Kavala area,
Greece, could be selected only from the identified clusters “C1 – coastal”, “C2 –
lowland” and “C3 – semi–mountainous”, and the monitoring parameters should be
the following: Cl, Na+, TH and Ca2+ for C1; EC for C2; and TA and HCO3– for C3.
This approach reduces the number of analyses and the cost of the risk assessment.

Acknowledgements. This research program was completely financed by the six Municipal
Authorities (Chrisoupoli, Keramoti, Elefheres, Paggaio, Orfanou and Piereon). Special thanks are
extended to the President of T.E.I. Kavala, for providing the SPSS software license. Prof. dr.
Antoaneta Ene acknowledges the financial support of LLP Erasmus (2010–2014) at “Dunarea de Jos”
University of Galati, Romania.

REFERENCES

Rigas, Application of multivariate statistical methods for groundwater physicochemical and
biological quality assessment in the context of public health, Environ. Monit. Assess. 170, 87
(2010).
2. A. Ene, A. Bosneaga, L. Georgescu, Determination of heavy metals in soils using XRF
3. A. Ene, N. Tiga, M. Praiser, L. Moraru, Principal component analysis of physico-chemical
4. T. Spanos, A. Ene, I.B. Karadjova, Assessment of toxic elements Cu, Cr, Ni, Pb, Cd, Hg, Zn, As
and hexavalent chromium in sewage sludge from municipal wastewater treatment plants by
9. L. Opran, D. Chicea, E. Gaspar, E. Lengyel, Results of physical and chemical parameters


