ASSESMENT OF GROUNDWATER QUALITY CHARACTERISTICS FROM THE UPPER PART OF GALLIKOS RIVER BASIN (N. GREECE) USING FACTOR AND CLUSTER ANALYSIS METHODS

MATTAS C.*, PAPACHRISTOU M. and SOULIOS G.

Aristotle University of Thessaloniki, Faculty of Sciences, School of Geology,
Laboratory of Engineering Geology and Hydrogeology,
cmattas@geo.auth.gr, mariap@geo.auth.gr, gsoulios@geo.auth.gr

Keywords: Gallikos river basin, Factor analysis, Cluster analysis, Groundwater degradation.

Abstract

This paper presents the statistical analysis of groundwater hydrochemical data from the upper part of Gallikos River basin (Macedonia, Northern Greece), in order to assess the water quality characteristics and comprehend the hydrochemical-hydrogeological conditions of this area. The two methods used were the factor and cluster analyses. The samples were classified according to their characteristics and relevance, whereas the relation and importance of various parameters and variables were studied. Finally, all these processes were quantified. The data were collected during the wet period (April) 2004. The results of the statistical analysis were correlated to the hydrogeological pattern and the hydrologic conditions of the studied area. Due to their influence and control on the hydrogeological and hydrological regime, the specific geological, lithostratigraphical and tectonic conditions, as well as the anthropogenic activities, are also discussed in the paper.

1. Introduction

The Gallikos River Basin is located in the northern part of Greece (Macedonia Province) and covers 1050 Km^2, discharging into the Thermaikos Gulf, North Aegean Sea. This study focuses on the upper part of the basin, between the watershed and the village of Nea Philadelphia. The studied area covers 868 Km^2 and constitutes the boundary between the Prefectures of Kilkis and Thessaloniki (Figure 1). It has a pear like shape and its mean altitude has been calculated at 357.7 m. The basin is formed of metamorphic rocks, carbonates, Neogene sediments and Quaternary deposits. The hydrographic network is very well developed, with the main branch being approximately 45.5 Km long within the study area boundaries. The population is 40,207 inhabitants (2011 census), a significant percentage of which is employed in the agricultural sector.

The groundwater resources are mainly located within the carbonate rocks and the Quaternary deposits. The groundwater samples were collected from boreholes along the river course and its tributaries and were analyzed for major ions, nitrates and heavy metals. In situ measurements of pH, electrical conductivity and water temperature were also performed. The hydrochemical analyses showed water quality deterioration.

During the last decades the basin has been influenced by anthropogenic activities, which is considered to be the cause of the continuous water resource degradation. The major driving forces that affect the studied area are urbanization, intensive agriculture, industrial activities and the regional development strategy (Mattas 2014).

2. Regional setting

2.1 Geological setting

The upper part of Gallikos River Basin belongs to the Circum Rhodope and Serbo-Macedonian zones (Figure 1). The crystalline basement of the area consists of pre-alpine formations, mainly Paleozoic gneisses, schists, amphibolites and quartzites (Mercier 1966; Kockel *et al.* 1979; Ioannides *et al.* 1990). Carbonate rocks (limestones) of Triassic and Jurassic age outcrop at the central-southern part of the basin (Meladiotis 1984). The Quaternary formations consist of lacustrine, terrestrial, fluviatile and fluviatile-torrential sediments, with sands, clays and conglomerates. The tectonic setting of the area is very complex (Mountrakis 1985). An inversion of the layers at the western boundary of the Serbomacedonian and Circum Rhodope contact is observed, which results in the tectonic placement of older Serbomacedonian formations (Vertiskos series) over the younger Circum Rhodope sediments. Due to the activity of the reverse faults, many irregular tectonic contacts have been created.

2.2 Hydrogeological setting

The geological formations of the studied area present different hydrogeological behavior, according to their composition and the tectonic impact. The Quaternary deposits and the limestones are permeable, while the gneisses, schists, amphibolites and quartzites of the basement are described as impermeable.

The main aquifer is developed at the plain part of the basin, inside the Quaternary deposits (Mattas 2009). Most of the boreholes have been drilled along the river. The water level is near the ground surface. The discharge rate is high for small drilling depths (15-20 m). The aquifers of the Quaternary deposits are recharged by the river and the precipitation. The limestone outcrops cover the 2.5% of the area; however their hydrogeological importance is high. Their mean thickness is 80 m. The depth of the boreholes drilled in the limestones is usually more than 150 m, penetrating also the surface sediments or the crystalline bedrock (Mattas 2011). The available data from the drilling exploration and piezometric measurements showed that, due to the intensive tectonic activity and the creation of secondary porosity, low potential aquifers have been developed inside the impermeable formations.

The impermeable geologic formations cover 57.5% of the basin, with the gneisses prevailing. The available data acquired from the borehole lithological sections show that the thickness of the impermeable system is at least 230 m. In most cases, the depth of the boreholes is over 110m. The several boreholes drilled in these formations did not encounter any aquifers. The relatively small thickness of the Quaternary deposits, their intersects with metamorphic and carbonate rocks and the layer inversion at the western boundary of Serbomacedonian Massif and the Circum Rhodope Zone, in combination with the lepioid tectonic structure, indicate hydraulic communication between the aquifers of the different formations in the area. The investigation of the piezometric surface shows that the groundwater flow direction follows the main tributaries of the hydrographic network, towards the south.

3. Materials and Methods

3.1 Data collection, sampling methodology

The data used in this paper were collected in April 2004. This period represents the highest ground and surface water level in the year. The sampling points were chosen in order to have an adequate spatial distribution. In situ measurements of pH, Electrical Conductivity (E.C.), and water temperature were also performed. The samples were analyzed in the Laboratory of the Land Reclamation Institute (National Agricultural Research Foundation, Thessaloniki, Greece) and the following parameters were determined: Ca, Mg, Na, K, Cl, NO_3, NO_2, SO_4, B, HCO_3, Fe, Mn, Cu, Pb, Cd, Zn.

The following methods were used for the determination of the ions:

- Ca^{+2}, Mg^{+2}, Cl^-, HCO_3^-, CO_3^{-2}, were determined with titration
- NO_3^-, SO_4^{-2}, NO_2^-, B^-, were determined with the use of spectrophotometer (UV-1204 SHIMADZU)
- Na^+, K^+, were determined with the use of of flamephotometer (SHERWOOD FLAME PARAMETER 410)
- total Fe, Mn^{+2}, Zn^{+2}, were determined with the use of atomic absorption spectrometer (PERKIN ELMER PRECISELY AANALYST 200).

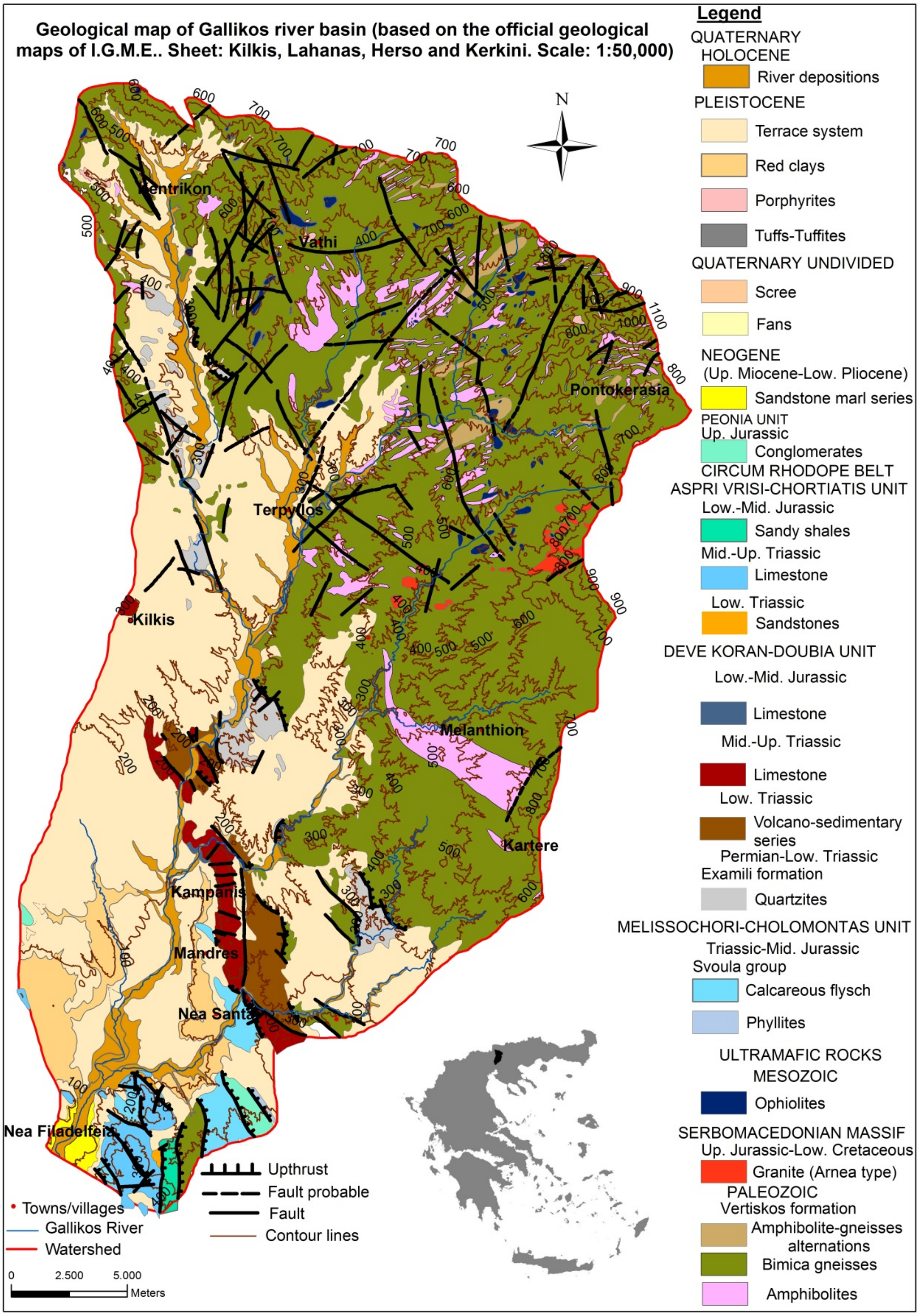

Figure 1. Geological map of Gallikos river basin.

3.2 *Groundwater quality*

The general quality characteristics of the groundwater samples are depicted in Table 1. The majority of the analyzed samples (49 out of 66) shows pH values greater than 7, while 14 sample out of 66 show values greater than 7.5, indicating a slightly alkaline type. The highest values are recorded in the north-east part of the basin, where crystalline rocks (gneisses, schists) prevail. As expected, in the area of sediments and carbonates the pH values are lower, due to the fact that these formations are more acidic. However, high values are recorded at the exit of the basin, coinciding with high Electrical Conductivity (EC) values. The EC varies between 120 and 2920 μS/cm. The higher values are probably attributed to pollution from anthropogenic activities. The ions that show unusually high values, which cannot be justified by hydrogeological criteria, are sodium and chlorides. Sodium varies from 12.3 to 310 mg/L and chlorides from 17 to 798.8 mg/L.

According to Soulios (2006), the mean value of sodium concentration in typical groundwater is usually 50 mg/L. Within the study area, 29 out of 66 samples have values greater than 50 mg/L. In addition, the sodium concentration in 4 samples is higher than 200 mg/L, which constitutes the maximum acceptable value for drinking water, while 6 samples have chloride values above 250 mg/L, which is the corresponding maximum acceptable value for chlorides. Since there is no evidence for seawater existence in the area, the most probable pollution source is the industrial activities in the area.

The values of calcium range from 12.2 to 292 mg/L. Magnesium ranges from 11 to 104 mg/L. Bicarbonate values are between 146.4 and 744.2 mg/L. The higher values of the aforementioned ions are recorded near the areas with carbonate formations, due to the dissolution of these rocks. Potassium ranges from 1 to 123 mg/L and sulphates from 0.8 to 202. Finally, nitrates range from 0 to 497 mg/L. The presence of nitrates implies pollution from fertilizers, usually occurring in cultivated areas, or from the discharge of untreated domestic effluent, mainly from septic tanks, as a result of a sewage network absence in the small settlements within the study area.

The analysis of heavy metals and minor ions showed high concentration values. In some samples they exceeded the maximum acceptable values for drinking water, regarding iron, manganese and zinc. For the other ions, the values were either below the maximum acceptable values for drinking water or below the detection limit of the applied analysis methodology.

According to the Piper diagram (1944), the majority of groundwater samples is classified in two types (Figure 2): Ca(Mg)-Cl(SO_4) and Ca(Mg)-HCO_3. A small number of samples is identified as K(Na)-Cl(SO_4) and K(Na)-HCO_3, indicating the effect of anthropogenic activities.

Table 1. Summary statistics of the groundwater samples quality characteristics

Parameter	Min value	Max value	Mean value	Standard Deviation
EC (μS/cm)	120	2920	1019	543
pH (mg/L)	6.2	9.8	7.3	0.5
Cl (mg/L)	17.8	798.8	122.2	140.9
NO_3 (mg/L)	0	497	35.9	69.7
SO_4 (mg/L)	0.8	202	63.9	49.8
HCO_3 (mg/L)	146.4	744.2	399.2	124.5
K (mg/L)	1	123	14.3	16.6
Na (mg/L)	12.3	310	63.6	58.9
Mg (mg/L)	11	104	40.8	20.7
Ca (mg/L)	12.2	292	96	52.5

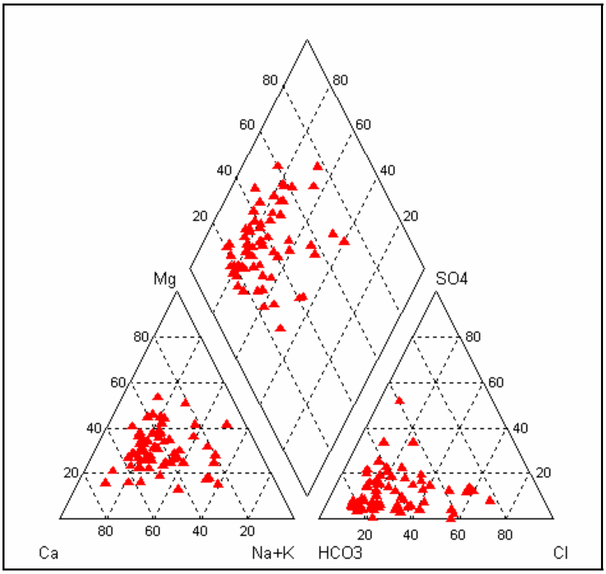

Figure 2. Piper diagram.

3.3 Statistical analysis

The R-factor analysis is a data reduction method. A large number of interdependent variables can be described by a small number of factors (latent variable) that represent to a significant degree the initial variables (Voudouris *et al*. 1997; Voudouris *et al*. 2000). This method contributes to the simplification of complex problems that are described by many parameters, without losing information (Qian *et al.* 1994). The factors that describe the interdependent variables are functionally independent and they behave in the same way with the statistically significant variables that participate in the factors (Anazawa and Ohmori 2005; Brown 1998). For the simplification of the procedure a varimax rotation is conducted in order to reduce the number of factors that describe the total variance (Mathes *et al.* 2006; Mencio and Mas-Pla 2008). Usually, the factors that are finally selected are those that show eigenvalue higher than one (Cattell 1978). The number of the factors that will result could be indicative of the reasons that cause the data variation (Sandow *et al.* 2008). Each factor describes a specific hydrogeological process, depending on the participating variables. The factor scores can be calculated for each sampling point and reveal the effect of the factor at the specific location. They can be used to draw contour maps that depict the spatial distribution of the factors. Negative scores represent areas that are not affected by the process that the factor describes. On the contrary, positive or near zero scores, represent areas that are affected by the factor.

The cluster analysis aims at grouping the hydrochemical parameters and the water samples according to their similarities (Vega *et al.* 1998; Davis 2002). For the purposes of this study the K-means cluster methodology was implemented, according to which a large number of samples are required (Norusis 1993; Coakes-Steed 1999) and the number of the clusters is predetermined. The distance of the variable from the center of the cluster is calculated by an algorithm and the variable is classified at the nearest one.

The data used in this study were derived from 66 groundwater samples. The parameters chosen for the implementation of the factor and cluster analyses are the following: Ca^{+2}, Mg^{+2}, Na^{+}, K^{+}, Cl^{-}, NO_3^{-}, SO_4^{-2}, HCO_3^{-}, pH and Electrical Conductivity.

High values of heavy metals were recorded in the samples that originated from boreholes drilled in the crystalline rocks at the northern and eastern part of the basin. However, the number of the samples was inadequate for the statistical analysis.

4. Results

4.1 Evaluation of R-Factor analysis results

The R-Factor analysis resulted in three (3) factors (Figure 3) that interpret 74.75% of the total variance (Table 2). Factor 1 interprets 45.23% of the variance. Chlorides, Sodium and Electrical Conductivity participate in this factor. These variables are associated to the impact of the anthropogenic activities on groundwater. As it is shown on the map of Figure 4, this factor has positive value at the southern part of the basin, close to its exit, where many industries operate (Figure 5). According to Mattas (2009) high values of the two variables (Cl, Na) were recorded. Factor 2 interprets 15.84% of the variance. Calcium, Magnesium and bicarbonates are the variables that participate in this factor. These variables represent the procedures of carbonate rocks dissolution by water. The spatial distribution of the factor shows positive value near the area where the carbonate rocks outcrop and it does not seem to be affected from this procedure at the eastern part of the area where the gneisses prevail (Figure 4).

Finally, Factor 3 interprets 13.67% of the variance. Potassium, nitrates and pH participate in this factor and are associated to the use of fertilizers (Sandow *et al.* 2007). The nitrates could also be attributed to the infiltration of effluents from septic tanks in the aquifers, due to the lack of treatment plants in the majority of the villages. The Potassium could be explained by the dissolution of K-feldspars (Rosen and Sarah 1998) in the gneisses. As it is shown on the map of Figure 4, Factor 3 has positive values mainly at the cultivated lands and at the part of the basin that gneisses prevail.

Table 2. Factor analysis results for the hydrochemical data of April 2004

Component	Initial Eigenvalues				Factors		
	Total	% of Variance	Cumulative%		1	2	3
1	4.663	46.631	46.631	Cl	0.95	0.166	0.026
2	2.036	20.363	66.994	NO_3	0.054	0.579	0.709
3	1.092	10.916	77.91	SO_4	0.489	0.377	-0.221
4	0.733	7.33	85.24	HCO_3	0.137	0.796	0.086
5	0.498	4.983	90.223	K	-0.227	0.282	0.807
6	0.458	4.579	94.802	Na	0.921	0.076	0.023
7	0.255	2.554	97.356	Mg	0.561	0.650	0.071
8	0.193	1.925	99.281	Ca	0.527	0.702	0.124
9	0.049	0.492	99.773	EC	0.830	0.496	0.092
10	0.023	0.227	100	pH	0.228	-0.245	0.842

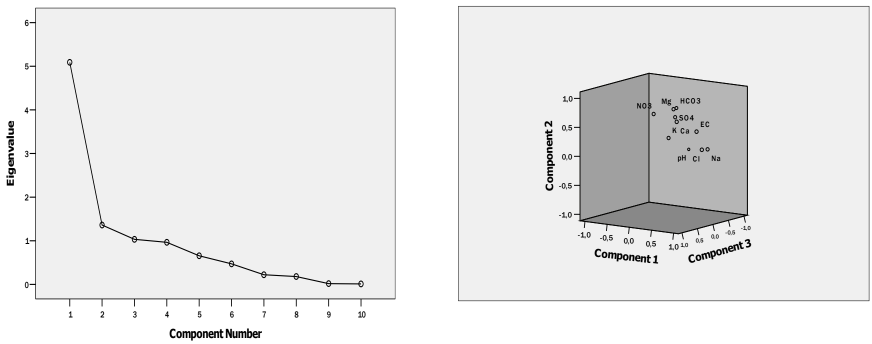

Figure 3. Eigenvalues and factors diagrams for the hydrochemical data of April 2004.

4.2 *Evaluation of K-means analysis results*

The Euclidean distance was used as a measure of the similarity between the pairs of samples for the classification in groups. The samples used for the application of the K-means method were classified in two different clusters. The final cluster centers of the two groups have been calculated by Mattas *et al.* in previous work (2005) for the plain (south) part of the basin for 36 samples and they are presented in Table 3. In the present work the cluster analysis was conducted for the 66 samples that were taken for the entire basin. As it is shown in Figure 4, fifty nine (59) samples are classified in Cluster 1 and seven (7) in Cluster 2. The samples of

Table 3. Final Cluster centers of the two hydrochemical groups

	Group 1	Group 2
EC	1026.42	2222.50
Ca^{2+}	91.33	177.67
Mg^{2+}	37.48	62.17
Na^{+}	58.78	203.5
K^{+}	13.67	5.45
HCO_3^-	424.64	433.10
SO_4^{2-}	49.79	118.90
NO_3^-	38.63	38.52
Cl^-	108.11	496.05

Cluster 2 occur in highly affected by anthropogenic activities areas, where high Nitrate, Sodium or Chloride concentrations are recorded. All the samples (of Cluster 2) show high Electrical Conductivity values that range between 1610 and 2420 μS/cm. The Chloride concentrations range from 156 to 544 mg/L. Chlorides are indicative of the anthropogenic pressures, since it is a relatively inert element and its origin could not be attributed to the lithologic composition of the geologic formations. High values of Nitrates (maximum value 497 mg/L) and/or Sodium (from 72 to 310 mg/L) are recorded in some of the samples.

5. Conclusions-Discussion

The Factor and Cluster Analyses proved very useful for the purposes of this study, i.e. the identification and grouping of samples according to the prevailing procedure that determine their chemical composition or according to their similarities. The implementation of these methods showed that groundwater resources of the basin have been affected by anthropogenic activities. High concentrations of Sodium and Chlorides are associated to the industrial activities of the area. The high nitrates are attributed to the agricultural activities (mainly the use of fertilizers) and to the pollution from the septic tanks leaks. The dissolution of carbonate rocks is an important natural geological procedure that enriches groundwater in Calcium, Magnesium and bicarbonates. The dissolution of K-feldspars increases the potassium concentration in the aquifers that are developed inside the crystalline rocks of the basement. Shifting from the current cultivations to less demanding, regarding the use of fertilizers, is necessary. The constant and strict implementation of the Code for Good Agricultural Practice (COGAP) is also essential. The industrial units of the area should be equipped with effluent treatment units. In addition, integrated treatment plants must be constructed and operate to the villages in order to eliminate the use of septic tanks. The occasional discharge of untreated waste in the torrents, or directly in the aquifers (through the boreholes) must immediately stop. Last but not least, severe punishment measures must be imposed by the authorities to those who do not comply with the regulations.

6. Acknowledgments

The authors would like to thank the Greek Ministry of Development, since the data used in this paper were collected by Dr C. Mattas in the framework of a research project funded by the Directorate General for Research and Technology that belongs to the aforementioned Ministry.

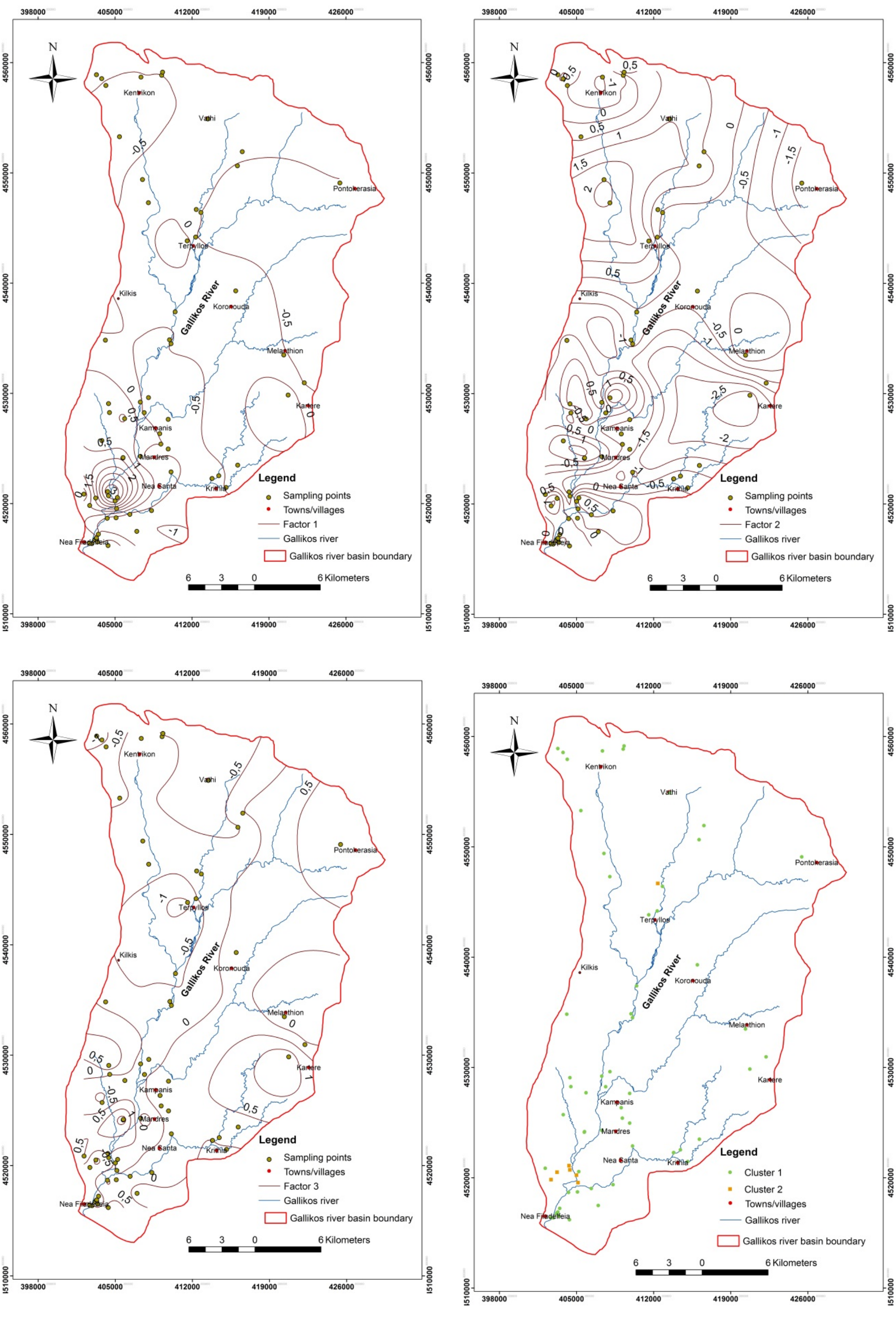

Figure 4. Spatial distribution of factors and clusters.

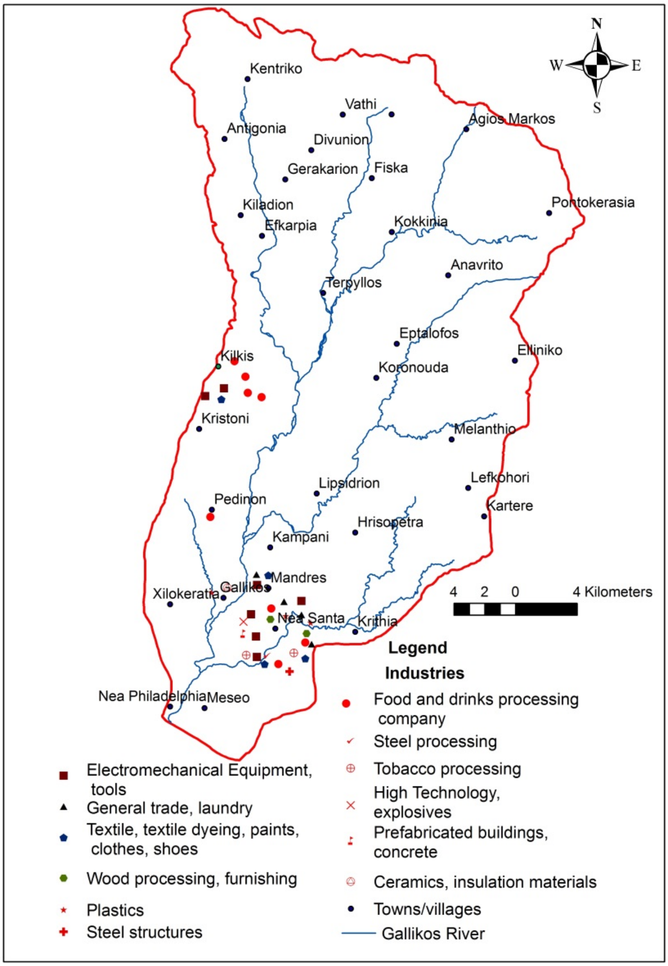

Figure 5. Spatial distribution of industries (illustrated by the authors based on data available in http://www.anki.gr/DE04721D.el.aspx. Last accessed, August 2014).

7. References

Anazawa, K., Ohmori, H. (2005). The hydrochemistry of surface waters in Andesitic Volcanic area, Norikura volcano, central Japan. Chemosphere, 59: 605-615.

Brown, C.E. (1998). Applied Multivariate Statistics in Geohydrology and Related Sciences. Springer, New York.

Cattell, R.B. (1978). The scientific use of factor analysis in behavioral and life sciences. New York, Plenum Press.

Coakes, S. J., Steed, L.G. (1999). SPSS Analysis without Anguish, Versions 7.0, 7.5, 8.0 for Windows. John Wiley & Sons: Milton, Australia.

Davis, J. (2002). Statistics and data analysis in geology. Wiley, New York.

Ioannides, A., Kelepertzis, K. (1990). Geological map of Greece, Herso sheet. Scale 1:50,000. IGMR, Athens, Greece.

Kockel, F., Ioannides, K. (1979). Geological map of Greece, Kilkis sheet. Scale 1:50,000. IGME, Athens, Greece.

Kockel, F., Mollat, H., Antoniades, P., Ioannides, K. (1979). Geological map of Greece, Lachanas sheet. Scale 1:50,000. IGMR, Athens, Greece.

Mattas, C., Soulios, G., Dimopoulos, G., Diamantis, J., Panagopoulos, A., Voudouris, K. (2005). Groundwater quality in Gallikos basin, Prefecture of Kilkis. Proc. of 7th Hellenic Conference on Hydrogeology (eds. Stournaras G.), Athens, 311-320.

Mattas, C. (2009). Hydrogeological research in Gallikos River basin. PhD Thesis, School of Geology, Aristotle University of Thessaloniki, Greece.

Mattas, C., Soulios G. (2011). Hydrogeological conditions of the upper part of Gallikos River basin. Advances in the Research of Aquatic Environment, v.1, Springer.

Mattas, C., Voudouris, K., Panagopoulos, A. (2014). Integrated Groundwater Resources Management Using the DPSIR Approach in a GIS Environment: A Case Study from the Gallikos River Basin, North Greece. Water, 6: 1043-1068; doi:10.3390/w6041043.

Mathes, E.S., Rasmussen, C.T. (2006). Combining multivariate statistical analysis with geographic information systems mapping: a tool for delineating groundwater contamination. Hydrogeology Journal, 14: 1493-1507.

Meladiotis, I. (1984). Geological research of the eastern part of Thessaloniki-Giannitsa plain and especially the area among Axios and Gallikos rivers where exploitable aquifers are developed. PhD, School of Civil Engineering, A.U.Th.

Mencio, A., Mas-Pla, J. (2008). Assessment by multivariate analysis of groundwater-surface water interactions in urbanized Mediterranean streams. Journal of Hydrology, 252: 255-266.

Mercier, J. (1966). Stude geologique des zones internes des Hellenides en Macedoine Centrale (Grece). Ann. Geol. Des Pays Hell., 20: 1-596.

Mountrakis, D. (1985). Geology of Greece. University Studio Press. Thessaloniki.

Norusis, M.J. (1993). SPSS for Windows: Professional Statistics, Release 6.0. SPSS Inc., pp, 385.

Qian, G., Gabor, G., Gupta, R.P. (1994). Principal components selection by the criterion of the minimum mean difference of complexity. Journal of Multivariate analysis, 49: 55-75.

Rosen, M., Sarah, J. (1998). Controls on the chemical composition of groundwater from alluvial aquifers in the Wanaka and Wakatipu basins, Central Otago, New Zealand. Hydrogeology Journal, 6(2): 264-281.

Sandow, M.Y., Duke, O., Bruce, B.Y. (2008). Hydrochemical evaluation of the Voltaian basin: the Afram Plains area, Ghana. Journal of Environmental Management, 53: 1213-1223.

Piper, A.M. (1944). A graphic procedure in the geochemical interpretation of water analyses. American Geophysical union, Transactions, 25: 914-923.

Sandow, M.Y., Duke, O., Bruce, B.Y. (2008). A multivariate statistical analysis of surface water chemistry data-The Ankobra Basin, Ghana. Journal of Environmental Management, 86: 80-87.

Soulios, G. (2006). General Hydrogeology , Volume IV. University Studio Press, Thessaloniki.

Vega, M., Pardo, R., Barrado, E., Deban, L. (1998). Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. Water Research, 32: 3581-3592.

Voudouris, K., Lambrakis, N., Papatheodorou, G., Daskalaki, P. (1997): An application of Factor Analysis for the study of the hydrogeological conditions in Plio-Pleistocene aquifers of NW Achaia (NW Peloponnesus, Greece). Mathematical Geology, 29(1): 43-59.

Voudouris, K., Panagopoulos, A., Koumantakis, J. (2000). Multivariate Statistical analysis in the assessment of Hydrochemistry of the Northern Korinthia Prefecture alluvial aquifer system (Peloponnesus, Greece). Natural Resources Research, 9(2): 135-146.