Multivariate Statistical Interpretation of Physical, Chemical and Microbiological Data of Potable Water in the Context of Public Health

A. PAPAIOANNOU[1*], K. KAKAVAS[1], P. PLAGERAS[1], A. MINAS[1], Z. ROUPA[2], A.G. PALIATSOS[3], P.T. NASTOS[4] and A. MINAS[1]

[1] Department of Medical Laboratories, Education & Technological Institute of Larissa, 41110 Larissa, GREECE
[2] Ministry of Education, Technical University Directorate, Athens, Greece
[3] General Department of Mathematics, Technological Education Institute of Piraeus, GREECE
[4] Department of Geography and Climatology, University of Athens, Panepistimioupolis, 157 84 Athens, Greece

Abstract: In the present study three representative areas, lowland (LL), mountainous (M) and coastal (C), have been selected for the collection of water samples, in the Prefectures of Larissa, Trikala, Karditsa and Magnesia, Thessaly, central Greece. Three physical: (electric conductivity (EC, 25^0C, μS/cm), pH, total hardness (TH, mg/L $CaCO_3$)), fourteen chemical (mg/L): (chloride (Cl^-), nitrate (NO_3^-), nitrite (NO_2^-), bicarbonate (HCO_3^-), carbonate (CO_3^{-2}), potassium (K^+), sodium (Na^+), ammonium (NH_4^+), calcium (Ca^{+2}), magnesium (Mg^{+2}), zing (Zn), iron Fe), manganese (Mn) and lead (Pb)) and eight microbiological: (total bacteriological flora in 22^0C and in 36^0C (TBF 22^0C and TBF 36^0C), *total coliforms*, *Escherichia Coli*, *Staphylococcus Aureus, Enterococci*, *Clostridium Perfrigens, Pseudomonas Aeruginosa*) parameters have been determined and analyzed for one-year period (2006). The mean values of the studied physical and chemical parameters, except trace elements, were found to be within the limits mentioned in the 98/83/EEC Directive. The analysis of microbiological variables shows that a big number of the studied water samples are inappropriate for human consumption because of the presence of microbes.

Key Words: Potable water, public health, physicochemical parameters, microbiological parameters, Thessaly, Greece.

1 Introduction

European Union has established a number of directives which refer to the protection of the water resources and public health [1-6]. The intention of the Drinking Water Directive (DWD) is to protect the health consumers' in the European Union, to make sure the water is wholesome and clean (free of unacceptable taste, odor, color) and has a pleasant appearance. To make sure drinking water is indeed healthy, clean and tasty everywhere in the E.U., the DWD sets standards for the most common substances (so-called parameters) that can be found in drinking water [7-9].

Our research interests focus on the assessment and monitoring the quality of water (groundwater, drinking water, surface water). Moreover, a research program (Arximedes II) for the internal and external control of measurements of potable water is applying in the region of Thessaly.

In previous work it was presented the quality of potable water in Thessaly for the period 2004-05 [10-12]. In this study is presented a multivariate statistical interpretation of physical, chemical and microbiological parameters, to assessment and monitoring the quality of potable water in Thessaly, for the year 2006. For this, physical: (electric conductivity (EC, 25^0C, μS/cm), pH, total hardness (TH, mg/L $CaCO_3$)), chemical (mg/L): (chloride (Cl^-), nitrate (NO_3^-), nitrite (NO_2^-), bicarbonate (HCO_3^-), carbonate (CO_3^{-2}), potassium (K^+), sodium (Na^+), ammonium (NH_4^+), calcium (Ca^{+2}), magnesium (Mg^{+2}), zing (Zn), iron Fe), manganese (Mn) and lead (Pb)) and microbiological: (total bacteriological flora in 22^0C and in 36^0C (TBF 22^0C and TBF 36^0C), *total coliforms*, *E. Coli*, *Staphylococcus Aureus,* *Enterococci*, *Clostridium Perfrigens, Pseudomonas Aeruginosa*) parameters have been determined and analyzed.

2 Materials and Methods

Sixteen sample sites in the Prefectures of Larissa, Trikala, Karditsa and Magnesia (central Greece), were selected for the collection of water samples. From each site samples were collected every month, for one-year period (2006).

The involved analytical methods were standard procedures, as recommended by the European directive 98/83. Potentiometric methods were applied for EC and pH measurements; atomic absorption spectrometry was used for the determination of calcium, magnesium, potassium and sodium; chloride was measured by UV-spectroscopy, complexometry was applied for total hardness determination; trace elements were measured by atomic absorption spectrophotometer; finally, absorption spectrophotometry was the analytical method for sulphate, nitrate and nitrite analysis.

The operating parameters for working elements were set according to the recommendations of the manufacturers. The conditions for sampling, sample preparation, the calibration, the limits of detection and the procedure uncertainty are described in details elsewhere [8, 13].

Chemicals used for the analysis were of analytical reagent grade and all solutions were prepared using HPLC purity water.

All this data analysis was performed with the Statistical Package for Social Sciences (SPSS 12.0, SPSS Inc.) and STATISTICA 7.0 software for Windows.

3 Results and Discussion

3.1 Statistical Analysis

Basic statistics and correlation calculations were carried out in order to give initial information about the clinical data. Unless otherwise indicated, the characteristics of the cases were described as mean and standard deviation. Tests for significance of observed mean differences were performed using the Student's t-test.

Cluster analysis (CA) was used for multivariate statistical modeling of the input data [14]. The main goal of the hierarchical agglomerative cluster analysis is to spontaneously classify the data into groups of similarity (clusters) searching objects in the n- dimensional space located in closest neighborhood and to separate a stable cluster from other clusters. Usually, the sampling sites are considered as objects for classification, each one determined by a set of variables (chemical concentrations). It is also possible to search for links between the variables turned to objects of classification.

3.2 Experimental

3.2.1 Physicochemical Study

We studied the distribution patterns of some physical and chemical parameters commonly assayed in water chemistry laboratories.

Table 1 presents the statistical characteristics of all the examined physical parameters.

Table 1. Descriptive statistics of physical variables for all studied sample sites

Variable	Mean	Min	Max	Std. Dev.
EC	504.3	8.0	2330.0	323.8
pH	7.45	6.04	8.67	0.54
TH	268,5	27,5	1645.0	187.9

Allowable limits: EC (2500 μS.cm^{-1}), pH (6.5 – 9.5) and TH (mg/L $CaCO_3$).

From the data of table 1 it can be concluded that the mean values of all studied physical variables are found in the permissible limits that the directive 98/83 puts. The low values of EC and pH are characteristic for spring waters that differ in quality from the water of the drillings.

Table 2. Descriptive statistics of chemical variables for all studied sample sites

Variable	Mean	Min	Max	Std. Dev.
Cl^-	35,9	4,7	152,6	19,3
NO_3^-	8,1	0,0	111,2	15,2
NO_2^-	0,010	0,000	0,400	0,042
HCO_3^-	285	2	1982	230
CO_3^{-2}	3,39	0,00	33,000	6,8
K^+	1,04	0,00	11,80	1,55
Na^+	22,0	0,3	215,5	31,6
NH_4^+	0,105	0,000	3,020	0,268
Ca^{+2}	58,9	0,9	581,0	53,7
Mg^{+2}	20,60	0,04	137,30	18,90
Zn	0.601	0.000	1.400	0.364
Fe	0,364	0,000	10,500	0,811
Mn	0.993	0,000	15.730	2.355
Pb	0,012	0,010	0,017	0,001

Allowable limits: (mg/L): Cl^- (250), NO_3^- (50), NO_2^- (0.10), K^+ (12), Na^+ (200), NH_4^+ (0.50), Mg^{+2} (50), Fe (0.200), Mn (0.050) and Pb (0.010).

From the data of table 2 it can be concluded that the mean values of all studied chemical parameters are found in the permissible limits that the directive 98/83 puts, except trace elements. Data from Table 2 show that the pollution of drinking water by nitrates in all studied sites is not significant. None studied site presents high concentrations of nitrite, too. The concentrations of NH_4^+ had exceeded the allowable limits in some cases. The concentrations of Fe, Mn and Pb had exceeded the allowable limits that 98/83 directive puts.

The statistical characteristics of all the examined physical and chemical parameters in each sample site are presented in Table 3.

Table 3. Descriptive statistics of physical and chemical variables per studied sample site

Variable	Site LL Mean±SD	Site M Mean±SD	Site C Mean±SD
EC	636,4±261.2	311,0±175.6	449,3±378.9
pH	7.73±0.45	7.52±0.51	7.45±0.54
TH	306.9±100.7	178.4±90.3	270.0±262.0
Cl^{-}	33.2±9.0	26.6±11.7	42.8±26.3
NO_3^{-}	15.9±20.8	4.1±6.6	2.0±3.0
NO_2^{-}	0.021±0.062	0.005±0.018	0.001±0.002
HCO_3^{-}	314.8±122.8	169.1±105.6	305.9±322.1
CO_3^{-2}	5.9±9.4	3.0±3.6	1.0±2.5
K^{+}	1.63±2.06	0.96±1.36	0.49±0.45
Na^{+}	42.7±40.3	11.6±9.9	5.6±4.9
Ca^{+2}	63.0±23.6	37.7±21.1	64.2±78.4
Mg^{+2}	26.7±20.8	11.1±11.4	18.6±17.5
NH_4^{+}	0.119±0.178	0.066±0.102	0.109±0.376
Zn	0.741±0.168	0.574±0.378	0.470±0.448
Fe	0.307±0.212	0.272±0.146	0.463±1.250
Mn	1.570±3.152	0.955±2.454	0.425±0.624
Pb	0.011±0.001	0.010±0.001	0.012±0.002

Electrical conductivity mean values are relatively low and progressively increase towards the lowland areas (Table 3). The high values of EC in LL-site can be attributed to the big amount of total dissolved salts in groundwater, and the reason for that can be the increasing fertilization. Referring to the comparison of the mean values of EC among the areas, there is a significant difference between LL-M, LL-C and M-C sites. This can be attributed to different geological constitution of each area and to different use of land. Anthropogenic activities: agricultural, cattle farming, industrial and residential uses of territory induce an increase of salinity.

The pH means are greater than 7, and their spatial progressive increase is C-M-LL, so the greatest mean of pH occurs in LL-site. The mean values of pH are greater than 7.00 in all study areas (Table 3), indicating the slight alkaline character of groundwater. There are significant differences of pH values between LL – M and M – C sites, and this can be attributed to urbanization and wastes from municipalities and industries in LL area, animal wastes and agricultural activities in M area, tourism development and agricultural activities in C area.

The means of total hardness are medium and their spatial progressive increase is M – C – LL, so the greatest mean of total hardness occurs in LL-site (Table 3). In addition, the mean values of total hardness in all areas did not exceed the acceptable limits. The comparison of mean values of TH among the studied areas shows that there is a significant difference between LL-M and M-C sites. This can be attributed to geological differences among areas and to different use of land.

Studying the results, referring to chloride concentrations, the highest mean values are observed in the C area and they are attributed to the pollution of groundwater. However, the chloride concentration values remain up to now in the acceptable limits. High chloride concentration values are also observed in LL-site due to intensive human activities. The lowest values are observed in the M-site due to increasing agricultural and animal husbandry activities.

The nitrate pollution is important and must be assessed. The lowest mean nitrate concentration occurs in C-site and the maximum in LL-site. Especially in the LL-site there are some cases that the values exceed the acceptable limit (Table 2). The results of the comparison reveal that in areas where there is higher population density and intensive human activities the nitrate pollution increases and differs significantly from areas where these parameters do not exceed (mountainous and coastal sites). The LL-site is characterized by intense industrial and agricultural activity and the high nitrate concentration in water is related to wastes and over-fertilization that pollute the groundwater of the lowland areas.

From the results of the determination of chemical parameters can be inferred that the concentration of nitrite is stable in all studied sites. In some cases there are measurements with values exceeding the acceptable limit but these are individual cases and are attributed to a sort of local pollution. In general, the mean values are stable among the areas and it is obvious from the results of the comparison since there is no significant difference among the areas.

The mean values of concentrations of HCO_3^{-}, CO_3^{-2}, K^{+}, Na^{+}, Ca^{+2} and Mg^{+2} are also found in the permissible limits that the directive 98/83 puts. The concentrations of NH_4^{+} had exceeded the allowable limits in some cases. The presence of ammonium in drinking water declares the possibility of microbe's presence. The situation in this case is judged worrying and demands further investigation with parallel control of chlorination of water. The high concentration of Fe, Mn and Pb is another factor which contributes to the deterioration of the quality of potable water in the studied sites.

Consequently, there were differences among the physicochemical variables of the three studied sites. More specific, EC , TH and pH values were found bigger in LL-site that to M and C sites. Additionally, water in LL-site had bigger concentration of nitrates, nitrites, ammonium, Zn and Mn, when compared to the other sites. All the above differences concerning the physicochemical variables between the studied sites show the degree of deterioration of waters' quality in the three sites. Water in LL-site is found more deteriorated that to the other sites and this due to urbanization and the intensive agricultural activities in this site. Same picture show the quality of water in C-site but in this case another factor, the tourism, except agricultural activities is added to the reasons of deterioration of waters' quality.

3.2.2 Microbiological Study

A total number of 192 water samples have been collected from the same mentioned sample sites and eight microbiological parameters have been determined every month and analyzed for the same period. The mean, min and max values of each determined microbiological parameter is given in Table 4.

Table 4. Descriptive statistics of microbiological variables (mean (min – max)

Variable	LL-site	M-site	C-site
TBF 22^{0C}	101.3 (0-1384)	73.5 (0-486)	54.5 (0-550)
TBF 36^{0C}	102.6 (0-3880)	10.5 (0-170)	30.7 (0-290)
Total Coliforms	22.5 (0-361)	10.9 (0-109)	23.2 (0-630)
E. Coli	1.7 (0-66)	4.7 (0-107)	1.2 (0-44)
St. Aureus	0	0	0
Enterococci	1.5 (0-88)	2.2 (0-40)	7.1 (0-410)
Cl. Perfrigens	0	0.05 (0-1)	0.05 (0-1)
Ps. Aeruginosa	0.1 (0-3)	2.4 (0-15)	0.7 (0-10)

Allowable limits: (cfu/ml): TBF 22^0C (100), TBF 36^0C (20), (cfu/100ml): ***Total Coliforms*** **(0),** ***E. Coli*** **(0),** ***Staphylococcus Aureus*** **(0),** ***Enterococci*** **(0),** ***Cl. Perfrigen*** **(0)** ***and Ps. Aeruginosa*** **(0).**

The microbiological data (Table 4) show that a big number of water samples are inappropriate for human consumption because of the presence of microbes in all the studied sites. More specific, a big number of water samples were found positive in *total coliforms*, *E. coli*, *Enterococci* and *Ps. Aeruginosa* in all studied sites. Water samples of M-site and C-site are found positive in *Cl. Pergrigens* and none water samples were found positive in *St. Aureus*. The mean value of TBF in 22^0C and 36^0C had exceeded the allowable limits that the directive 98/83 puts in all studied sites, too.

3.2.3 Cluster Analysis of Physicochemical and Microbiological Data

In Figure 1, the hierarchical dendrogram for the clustering of the determined physical and chemical parameters for all the studied sites is plotted (Ward's method of linkage, Squared Euclidean distance as similarity measure, standardization of the input data). For clustering altogether 17 physical and chemical parameters were chosen (indicated in Table 1). It could be concluded that the parameters are principally separated into two big clusters each of them divided additionally into subclusters as follows:

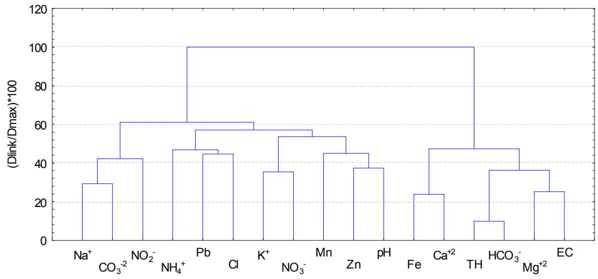

Figure 1. Hierarchical dendrogram of physicochemical variables in all sample sites

<u>**Cluster 1**</u> (seven parameters are included):

Na^+, CO_3^{-2}, NO_2^-, NH_4^+, Pb^{+2}, Cl^-, K^+, NO_3^-, Mn, Zn, pH.

Subcluster 1: Na^+, CO_3^{-2} and NO_2^-.
Subcluster 2: NH_4^+, Pb^{+2} and Cl^-.
Subcluster 3: K^+ and NO_3^-.
Subcluster 3: Mn, Zn and pH.

<u>**Cluster 2**</u> (eleven parameters are included):

Fe, Ca^{+2}, TH, HCO_3^-, Mg^{+2} and EC.

Subcluster 1: Fe and Ca^{+2}.
Subcluster 2: TH and HCO_3^-.
Subcluster 3: Mg^{+2} and EC.

In Figure 2, the hierarchical dendrogram for the clustering of the determined physical and chemical parameters for LL-site is plotted (Ward's method of linkage, Euclidean distance as similarity measure, standardization of the input data). For clustering altogether 17 physical and chemical parameters were chosen (indicated in Table 1).

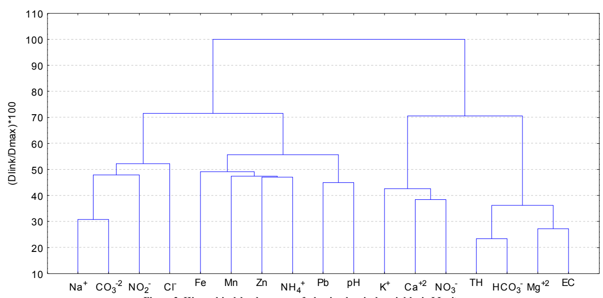

Figure 2. Hierarchical dendrogram of physicochemical variables in LL-site

In Figure 3, the hierarchical dendrogram for the clustering of the determined physical and chemical parameters for M-site is plotted (Ward's method of linkage, Euclidean distance as similarity measure, standardization of the input data). For clustering altogether 17 physical and chemical parameters were chosen (indicated in Table 1).

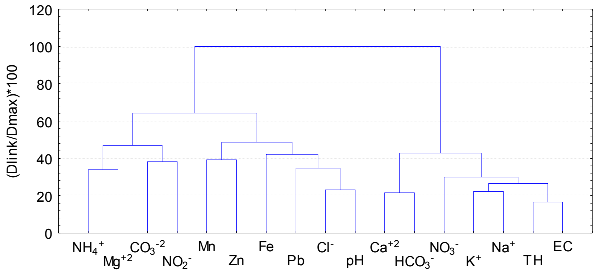

Figure 3. Hierarchical dendrogram of physicochemical variables in M-site

In Figure 4, the hierarchical dendrogram for the clustering of the determined physical and chemical parameters for C-site is plotted (Ward's method of linkage, Euclidean distance as similarity measure, standardization of the input data). For clustering altogether 17 physical and chemical parameters were chosen (indicated in Table 1).

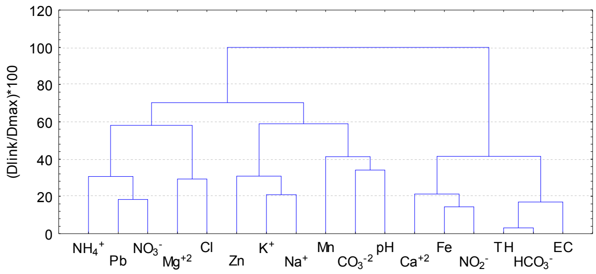

Figure 4. Hierarchical dendrogram of physicochemical variables in C-site

The hierarchical dendrograms of physicochemical parameters for each site shows small differences in clusters of variables. It could be concluded that the parameters are principally separated into two big clusters in all studied sites. The main clusters of variables in each sample site are presented in Table 5.

Table 5. Clusters of variables in each studied sample site

Site	Cluster A	Cluster B
LL	Na^+, CO_3^{-2}, NO_2^-, Cl^-, NH_4^+, Fe, Mn, Zn, Pb and pH	TH, EC, K^+, Ca^{+2}, Mg^{+2}, HCO_3^- and NO_3^-
M	CO_3^{-2}, NO_2^-, Cl^-, NH_4^+, Mg^{+2}, Fe, Mn, Zn, Pb and pH	Na^+, TH, EC, K^+, Ca^{+2}, HCO_3^- and NO_3^-
C	Na^+, CO_3^{-2}, Cl^-, NO_3^-, NH_4^+,K^+, Mg^{+2}, Mn, Zn, Pb and pH	EC, TH, NO_2^-, Ca^{+2}, HCO_3^-, and Fe,

From Table 5 it can be concluded that there are some stable clusters in all studied sites like (EC, TH and Ca^{+2}); (pH, Zn, Mn and Pb); and (Cl^-, NH_4^+). Variables like NO_3^-, NO_2^-, Cl^-, K^+, Na^+, Mg^{+2} and Fe are clustered by a different way into groups of similarity depended from sample site. Another information that we can notice is (Table 5) the stronger relation between the group of physical parameters (TH and EC) with parameters like Ca^{+2} and HCO_3^- that to pH.

In Figure 5, the hierarchical dendrogram for the clustering of the microbiological parameters for all the studied sites is plotted (Ward's method of linkage, Euclidean distance as similarity measure, standardization of the input data).

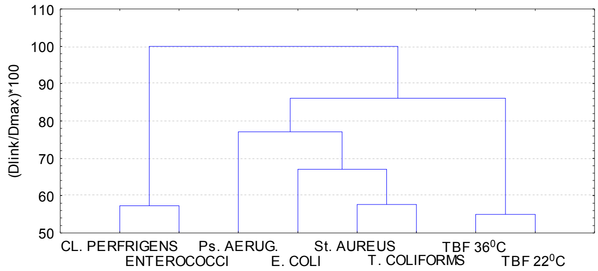

Figure 5. Hierarchical dendrogram of microbiological variables for all studied sample sites

For clustering altogether eight microbiological parameters were chosen (indicated in Table 3). It could be concluded that the parameters are principally separated into two big clusters each of them divided additionally into subclusters as follows:

Cluster 1 (two parameters are included):
Cl. Perfrigen and Enterococci.
Cluster 2 (seven parameters are included):
TBF 22^{0}C, TBF 36^{0}C, *Ps. Aeruginosa, E. Coli, Staphylococcus Aureus, and Total Coliforms.*
Subcluster 1: TBF 22^{0}C and TBF 36^{0}C.
Subcluster 2: Ps. Aeruginosa, E. Coli, Staphylococcus Aureus, Total Coliforms.

This intelligent data analysis gives an idea how the single physical, chemical and microbiological parameters should be compared and related to each other if the water sample is treated as depending on all parameters values simultaneously, not on single ones. For instance (Fig. 1), within a group of water there is a stronger relation between the group of physical parameters (TH and EC) with parameters like Fe, Ca^{+2}, HCO_3^- and Mg^{+2} that to Na^+, CO_3^{-2}, NO_2^-, Mn, K^+, NO_3^-, Pb^{+2}, NH_4^+, pH, Cl^- and Zn. Similarly (Fig. 2), there is a stronger relation between the group of parameters concerning the totally bacteriological flora (TBF 22^{0}C, TBF 36^{0}C) with variables like *Ps. Aeruginosa, E. Coli, Staphylococcus Aureu and, Total Coliforms* that to *Cl. Perfrigens and Enterococci*.

4 Conclusions

It is well established that the pollution of the environment and groundwater affects the quality of potable water. Additionally, the quality of drinking water is "strongly" correlated with the protection of public health.

The results of the present study show that the problems that have appeared in the quality of drinking water in the studied sites for the year 2006 can be mainly attributed to the microbiological pollution.

The intelligent data analysis on the drinking water physicochemical and microbiological parameters dataset has shown that when a complex system (in this particular case – drinking water quality) is considered as multivariate one, the information about the system substantially increases. For instance, the application of a typical classification approach as cluster analysis proved that two major groups of similarity between all 17 physicochemical parameters are formed. All this results support an idea that probably a general water quality indicator could be constructed taking into account the existing classification groups in the list of physicochemical or microbiological parameters.

Monitoring the quality of drinking water must be an ongoing process, so every change in pollutants should be recognized at an early stage. As a result, appropriate actions could be taken by official authorities to avoid the expansion of pollution and protect public health.

5 Acknowledgements

This study is co-funded by the European Social fund and National Resource – EPEAEK II (ARCHIMEDES II).

6 *References*

[1] Shearer L.A., Goldsmith J.R., Young C. Methemoglobin levels in infants, in an area with high nitrate water supply, Amer. J. of Public Health, 1972, 62, 1174-80.

[2] Shuval H.I. and Gruener N. Epidemiological and toxicological aspects of nitrates and nitrites in the environment. Amer. J. of Public Health, 1972, 62, 1045-51.

[3] Fattal B., Gutman-Bass N., Agursky T., Shuval HI. Evaluation of health risk associated with drinking water quality in agricultural communities. Water Sci. Technol. 1988, 20, 409-415.

[4] Zmirou D, Ferley JP, Collin JF, Charrel M, Berlin J. A follow-up study of gastro-intestinal diseases related to bacteriologically substandard drinking water. Am J Public Health 1987, 77, 582–584.

[5] Raina PS, Pollari FL, Teare GF, Goss MJ, Barry DA, Wilson JB. The relationship between *E. coli* indicator bacteria in well-water and gastrointestinal illnesses in rural families. Can J Public Health 1999, 90, 172–175.

[6] Pruss A. Review of epidemiological studies on health effects from exposure to recreational water. Int J Epidemiol. 1998, 27, 1–9.

[7] Strebel, O., Duynisveld, W.H.M. and Bottcher, J. Nitrate Pollution of groundwater in Western Europe. Agr. Ecosyst. Environ. 1989, 26, 189-214.

[8] EU Directive 98/83/EC, Official Journal of the European Communities. No. L 31/1.

[9] EU Directive 2000/60/EC, Official Journal of the European Communities. No. L 297/1.

[10] Papaioannou A., Plageras P., Dovriki E., Kakavas K., Nastos and Paliatsos A. Quality Control of Drinking Water and Public Health. WSEAS Transactions on Environment and Development 2006, 2, 845-850.

[11] Papaioannou A., Kakavas K., Dovriki E., Plageras P., Noulas A., Minas A., Krikelis V. and A. G. Paliatsos A.G. Quality Control of Groundwater of Region Thessaly (Greece). Fresenious Environmental Bulletin 2006, 15, 1015-1022.

[12] Papaioannou A., Plageras P., Dovriki E., Minas A., Krikelis V., Nastos P.Th., Kakavas K. and Paliatsos A.G. Groundwaters' quality and location of productive activities in the region of Thessaly (Greece). Desalination 2007, 213, 209-217.

[13] Standard Methods for the Examination of Water and Wastewater. 16th edition American American Public Health Association, New York. APHA, (1985).

[14] Massart D.L. and Kaufman L. (1983) "The Interpretation of Chemical Data by the Use of Cluster Analysis", J. Wiley, New York.