Asian Journal of Chemistry Vol. 22, No. 4 (2010), 2991-3000

Heavy Metal and Agricultural Toxics Monitoring in Garasou River in Iran for Water Quality Assessment¶

E. FATAEI*, S.M. MONAVARI†, A.H. HASANI†, A.R. KARBASI† and S.A. MIRBAGHERI†
Department of Environmental Management, Graduate School of the Environment and Energy, Islamic Azad University-Science and Research Branch, Tehran, Iran
FAX:(98)(21)44865002; Tel: (98)(914)3549400; Email: ebfataei@gmail.com

In this study water quality in Garasou river, which is located in north west of Iran in Ardabil province was assessed through multivariate statistical techniques including cluster analysis (CA), principal component analysis (PCA) and factor analysis (FA). During a year toxic and heavy metals were sampled in 11 stations. Based on the findings from cluster analysis, principle component analysis and factor analysis the stations were divided into three groups of highly polluted (HP), moderately polluted (MP) and less polluted (LP) stations. Station 10 fell in HP group. Station 11, 8, 7, 4 and 2 fell in MP group. The rest of the stations fell in LP group. Stations that were at the same or similar level of pollution were put in the same group. The amount of pollutants resulting from the analysis of principal components showed that the three first components accounted for 81 % of differences altogether, the first component accounting for 41.6, the second for 22.5 and the third for 16.9 % of differences, respectively. The value of KMO coefficient confirms the classification. This, in addition, confirmed that the model resulting from multi-linear regression analysis of the main component is a good indicator of each source or factor's loading in the distribution of pollution in the river.

Key Words: Cluster analysis, Principal component analysis, Water quality, Pollutant sources, Surface water resources.

INTRODUCTION

One of the methods in water resources assessment, environmental analyses and qualitative variables control to design qualitative monitoring management programs for rivers, is the use of multivariate statistical techniques which has been prevalent in recent years[1,2]. Recently, the use of principal component analysis (PCA) and factor analysis (FA) has become common in the analysis of water quality for reduction of the number of variables and better interpretation of the findings[3]. Some studies, too, have been carried out using principal component analysis (PCA), principal factor analysis (PFA), cluster analysis (CA) and DA methods to control qualitative

¶Part of Ph.D. Thesis.
†Science & Research Branch, Islamic Azad University (IAU), Tehran, Iran.

variables and monitoring of sampling stations. As examples we can refer to determining the quality of surface waters in Turkey[4] and also to the assessment of temporal and spatial fluctuations in the quality of water in Gomti river in India[5,6], Daliao river in China[2] and Fuji river in Japan[7]. The quality of water in Daliao river in China was assessed by Zhang *et al.*[2] using multivariate techniques. This study was carried out to determine the quality of water, agricultural toxic pollutants and heavy metals in the Gharasou river by using multivariate statistical techniques.

EXPERIMENTAL

Gharasou river is one of the main branches of the Aras river in the west side of Caspian sea, in Ardabil Province, northwest of Iran. This river originates from altitudes of Sabalan and Baghro mountains and after joining other streams in Ardabil plain, exits the plain in Samian hydrometric station. This river has three hydrologic units and is a prennial river with a length of 255 kilometers and average slope of 5.7 % and is considered as one of the Aras river's sub-rivers which itself is one the rivers in Caspian sea basin[8]. In this research the hydrologic unit of Ardabil plain of 4003 square kilometer was studied. The average water yield of this river[9], calculated in long term, in Samian station is about 228 m^3. Since this river passes through three urban (Ardabil, Nir and Sarein) and some rural areas, vast farmlands and some already established or under construction manufacturing units it is quite naturally exposed to pollution. For the decrease in the river's water yield on the one hand and because of the ever-increasing amount of water consumption and urban, industrial and agricultural sewage discharges on the other hand, the quality of water in the river is endangered. Since, Ardabil is an agricultural center and is in the process of development, thus constant monitoring of water quality in the river is necessary.

In determining the locations of study, pollution sources like: agricultural areas, residential and industrial areas, geological structure of land, main and subsidiary branches of the river and the ease of access were considered. The geographical position of stations is shown in Fig. 1.

Eleven sampling sites were chosen. The sampling process went on during a year from September 2007 until September 2008. The analyses were carried out based on the instructions introduced by EPA standard method[10]. The methods and tools used are given in Table-1.

TABLE-1
WATER QUALITY PARAMETERS, UNITS AND METHODS OF ANALYSIS

Parameters	Units	Analytical methods
Manganese	$\mu\ gL^{-1}$	ICP-OES
Iron	$\mu\ gL^{-1}$	ICP-OES
Aluminum	$\mu\ gL^{-1}$	ICP-OES
Cadmium	$\mu\ gL^{-1}$	ICP-OES
Copper	$\mu\ gL^{-1}$	ICP-OES
2,4-dichlorophenoxy acetic acid	$\mu\ gL^{-1}$	Gas chromatography
Fozalon	$\mu\ gL^{-1}$	Gas chromatography
Diazinon	$\mu\ gL^{-1}$	Gas chromatography

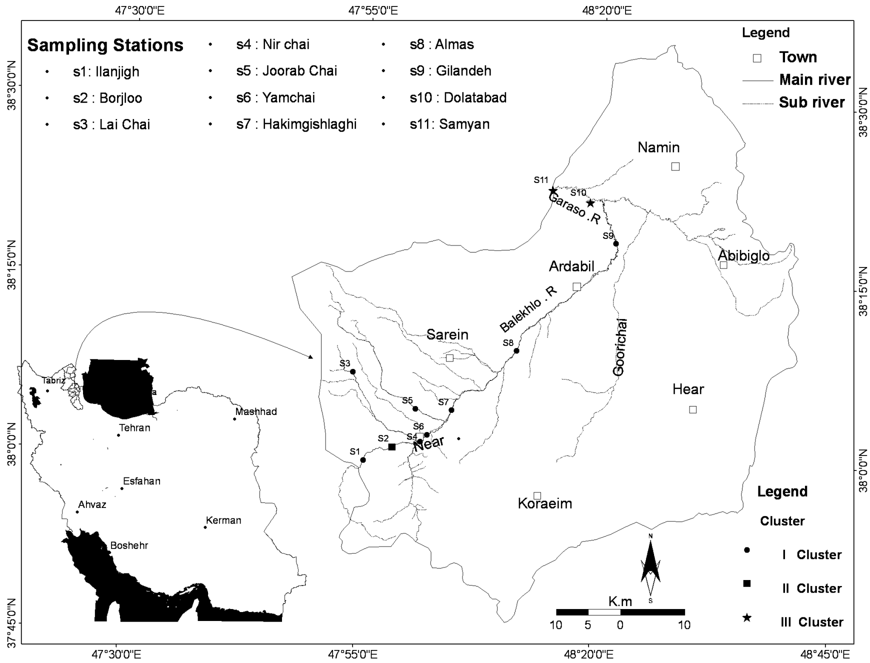

Fig. 1. Map of study area and surface water quality monitoring stations (listed 1-11) in Gharasou river basin of Ardabil, Iran

Sampling was performed by using glass bottles. Sampling containers were first washed with detergents and water and then with nitric acid and distilled water. Before sending the samples to the laboratory, initial steps like determining the temperature of water and air, fixing samples with needed chemicals and labeling (including: station specifications, sampling time and air condition) were done. Samples were sent to the laboratory and kept in a refrigerator after initial steps.

Sampling and analysis was done upon during a year on agricultural toxics and heavy metals in 11 stations, including Mn, Cu, Fe, Cd, Al, Diazinon, Fozalon and 2,4-D were controlled periodically in two periods of dry and wet seasons.

To analyze the data statistical methods like ANOVA, correlation, cluster analysis (CA), principle components analysis (PCA) and factor analysis (FA). All the mathematical and statistical calculations were done by Excell$_{2007}$, SPSS$_{16}$ and MINITAB$_{15}$.

The Kolmogorov-Smirnov (K-S) statistics were used to test the goodness of-fit of the data to log-normal distribution. The same procedure was applied to find the goodness of-fit for the assessment of principle components and factorial analysis using KMO and Bartlet test.

Multivariate statistical methods

Cluster analysis: Cluster analysis is a group of multivariate techniques whose primary purpose is to assemble objects based on the characteristics they possess. Cluster analysis classifies objects, so that each object is similar to the others in the cluster with respect to a predetermined selection criterion. The resulting clusters of objects should then exhibit high internal (within-cluster) homogeneity and high external (between clusters) heterogeneity. Hierarchical agglomerative clustering is the most common approach, which provides intuitive similarity relationships between any one sample and the entire data set and is typically illustrated by a dendrogram[11]. There are two types of cluster analysis *i.e.*, based on distance[12] and based on models[13].

Presently, methods based on distance are more frequently used. These methods themselves are divided into two groups: ordinal and chance models. Ordinal models are used more frequently compared to chance models. In this method, in the first stage of grouping, the number of parameters is equal to the number groups and each group includes only one parameter. In later stages the more similar groups are put together. Then these groups themselves join other similar groups. Finally, all the parameters are put in only one group[14]. There different grouping methods like, Unweighted paired group method using arithmetic averages (UPGMA)[15], Ward's minimum variance (WMV), single linkage (SL) and complete linkage (CL). In UPGMA method the similarities and differences between parameters and related groups are equal to the similarities or differences between parameters in the group. The distance in different groups is calculated between pairs of parameters. This is while in Ward's method[16], grouping is done based on intra-group minimum and inter-group maximum variance[17]. In this study, UPGMA method was used.

Principal component analysis/factor analysis: Principal component analysis is designed to transform the original variables into new, uncorrelated variables (axes), called the principal components, which are linear combinations of the original variables. The new axes lie along the directions of maximum variance. Principal component analysis provides an objective way of finding indices of this type[18]. Principal component provides information on the most meaningful parameters, which describes a whole data set affording data reduction with minimum loss of original information[19]. The principal component (PC) can be expressed as:

$$z_{ij} = a_{i1}x_{1j} + a_{i2}x_{2j} + a_{i3}x_{3j} + ... + a_{im}x_{mj} \quad (1)$$

where z is the component score, a is the component loading, x the measured value of variable, i is the component number, j the sample number and m the total number of variables.

Factor analysis follows principal component analysis. The main purpose of FA is to reduce the contribution of less significant variables to simplify even more of the data structure coming from PCA. This purpose can be achieved by rotating the axis defined by PCA, according to well established rules and constructing new variables, also called varifactors (VF). Principal component is a linear combination of observable water quality variables, whereas VF can include unobservable, hypothetical, latent variables[14,19]. Principal component analysis of the normalized variables was performed to extract significant PCs and to further reduce the contribution of variables with minor significance. These PCs were subjected[5,6,20-22] to varimax rotation (raw) generating VFs. As a result, a small number of factors will usually account for approximately the same amount of informations as do the much larger set of original observations. The FA can be expressed as:

$$z_{ji} = a_{f_1} f_{i_1} + a_{f_2} f_{i_2} + a_{f_3} f_{i_3} + \ldots + a_{fm} f_{mi} + e_{fi} \quad (2)$$

where z is the measured variable, a is the factor loading, f is the factor score, e the residual term accounting for errors or other source of variation, i the sample number and m the total number of factors.

Despite the fact that the efficacy of PCA and PFA methods is proved in the previous researches in controlling qualitative variables and surface water monitoring stations, but these studies are not thoroughgoing. Therefore, it is necessary to test the use of these methods and the findings through using KMO or Brtlet factor[6,7,23]. This shortcoming is not taken into account in the previous researches[3]. KMO is the criterion of goodness-of-fit and sampling which shows the appropriacy of variances. The range of KMO varies between 0 and 1. This factor is calculated by simple correlation coefficient and partial correlation[3] as shown in eqn. 3.

$$KMO = \frac{\sum_{i=1}^{p}\sum_{j=1}^{p} r_{ij}^2}{\sum_{i=1}^{p}\sum_{j=1}^{p} r_{ij}^2 + \sum_{i=1}^{p}\sum_{j=1}^{p} a_{ij}^2} \quad (3)$$

where, r_{ij} is the simple correlation coefficient between i, j and a_{ij} is the partial correlation coefficient between i and j provided the variables are constant.

With regard to eqn. 3, high values of KMO, necessitate smaller value for partial correlation coefficient and shows the accuracy of calculations by PCA and PFA. A bigger value than 0.5 shows the applicability of both methods in analyzing the data. Higher values of this statistic (values close to 1) indicate that factor analysis and principle component analysis are acceptable to a great extent[6,7].

RESULTS AND DISCUSSION

The description of the results of sampling and analyses for measured parameters is given in Table-2.

TABLE-2
DESCRIPTIVE STATISTICS OF WATER QUALITY VARIABLES

Parameters	Mean	Standard deviation	Range	
			Minimum	Maximum
Mn	34.32	24.38	15.00	97.50
Fe	333.30	150.50	96.00	557.50
Al	121.90	41.50	72.00	196.00
Cd	6.18	5.73	0.00	16.00
Cu	10.23	11.33	0.00	28.00
2,4-D	1.63	0.51	0.65	2.60
Fozolon	0.55	0.34	0.15	1.35
Diazinon	0.35	0.26	0.05	1.05

According to the K-S test, all the variables are log-normally distributed with 95 % or higher confidences.

Results of cluster analysis: To classify the water quality in sampling stations and to determine the sources of pollution, CA with UPGMA method, using Euclidean distance based on the standardized mean of measured parameters, was used. The stations were divided into three groups by dendogram cross-section, based on the farthest Euclidean distance inspired by Laurie *et al.*[24]. Fig. 2 represents cluster analysis dendrogram based on the measured parameters.

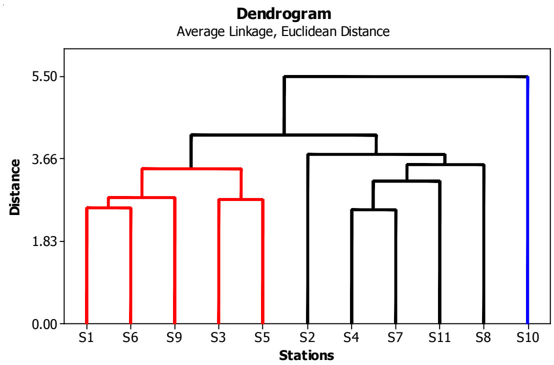

Fig. 2. Cluster analysis dendrogram of the sampling sites for surface water quality in Gharasou river basin

The dendrogram of Fig. 2 shows that station 10 has the highest pollution level. This station is distinguished from other stations concerning the level of pollution and has the biggest distance from other stations. After that is the second group with moderate pollution, which is related to stations 2, 4, 7, 8 and 11. Other stations are among the less polluted stations. The grouping resulting from CA is in agreement with grouping based on principle components analysis. The results of one-way ANOVA confirms the existence of meaningful differences among resulting groups concerning most parameters understudy with probability level of 5 and 1 %. Within-group assessments showed that these parameters were not meaningfully different within groups. This is while there were meaningful differences among clusters concerning most studied parameters. Therefore, the difference between the groups shows the difference between the pollutants.

Results of principle components analysis and factor analysis: The results of PCA including the values of particular vectors, eigen values, the values of relative and cumulative variances of principle components parameters are given in Table-3. The results showed that the three first components accounted for 81 % of the difference among stations. The two first components with eigenvalues of 3.74 and 2.02 together accounted for 64.1 % of the total qualitative differences among the stations. The first component accounted for 41.6 and the second for 22.5 %, respectively. The other components played a less important role in the qualitative changes of water among stations.

TABLE-3
LOADINGS OF EXPERIMENTAL VARIABLES (8) ON THE FIRST TWO ROTATED PRINCIPAL COMPONENTS AND FACTOR ANALYSIS*

Variables	PCA		FA	
	PC_1	PC_2	F_1	F_2
Mn	0.080	**0.583**	0.0910	**0.8750**
Fe	**0.429**	0.206	**0.8160**	0.4180
Al	0.354	**0.434**	0.6450	**0.7420**
Cd	0.224	-0.426	0.5610	-0.384
Cu	0.355	-0.206	0.6820	-0.101
2,4-D	0.355	-0.053	0.6300	-0.170
Fozalon	0.308	-0.260	0.5990	-0.417
Diazinon	**0.468**	-0.243	**0.9240**	-0.279
Eigenvalue	3.74	2.02	3.4800	1.930
% Total variance	41.6	22.5	43.480	24.09
Cumulative % variance	41.6	64.1	43.480	67.57

Bold and Italic values indicate strong and moderate loadings, respectively.

The comparison of the parameters' coefficients (especial vectors) for the first and second components shows that the first component has a major loading on the changes of model. This shows that the substantial differences among stations are mainly due to the amount of diazinon and iron. In this component the parameter manganese had a minor role in the differences among stations. Station 10 had the highest amount of PC1, which shows that the quality of water in this station is affected by the waste from agricultural drainage. Since, this station is located in a plain with big farmlands, it is affected by the polluting materials from agricultural activities which make it distinguished by having the biggest difference from other stations concerning water quality. Despite the fact that this station is located next to the industrial zone, the amount of affecting parameters which differentiate these stations shows that pollution in this station is not due to the waste from industrial units, especially Ardabil's second industrial town, which discharges a lot of heavy metals. This is an indication of the efficient treatment of waste water by these units. Therefore, the quality of water in this station is mainly affected by the drainage from agricultural activities. It is concluded that any station having high amounts of PC1 is mainly affected by waste from agricultural activities.

In the second component Mn and Fe with high positive coefficients and Cd, Cu, 2,4-D, fozalon and diazinon with negative coefficients showed different reactions compared to other parameters. The quality of water in stations 2, 4, 7, 8 and 11 which are in the other group, indicates that these stations are affected by the positive coefficients of both first and second parameters. The reason for this is the entrance of waste water from warm mineral waters, fish farming pools, swage treatment system and industrial slaughter house.

The rest of the stations including 1, 3, 5, 6 and 9 fell in the third group. These stations show that they are affected by the negative coefficients of both first and second variables. These stations due to the lack of vast farmlands and industrials units in their vicinity are classified in a group with a low level pollution. Among these stations, stations 1, 3 and 5 have the least amount of pollution. The reason for this is their location which is over the sub-rivers, making the main river themselves. This grouping shows that as we go further away from the headspring the amount of pollutants which enter the river increase and consequently decrease the water quality.

Fig. 3 shows the values for PC1 and PC2 along with the grouping of stations based on the investigated parameters. As we did in cluster analysis, these stations, based on the investigated parameters, by calculating values for PC1 and PC2, can be classified into three groups of high pollution (station 10), moderate pollution (stations 2, 4, 7, 8, 11) and low pollution (other stations).

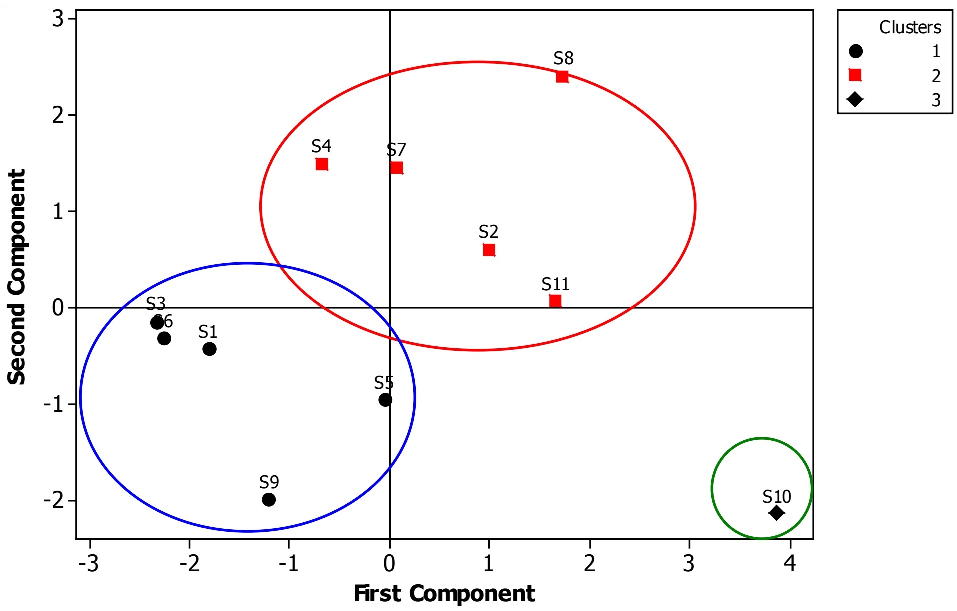

Fig. 3. Dispersion of investigated stations based on the first two principal components in the Gharasou river basin

The coefficients for rotating varimax matrix (loading) in Table-3 show that the results from FA confirm the findings from PCA. In FA, too, the two first factors account for 67.57 % of data variance. Therefore, it can be concluded that these two factors have been responsible for the great part of the differences between stations in this research. With regard to Fig. 4, that represents the results of FA, eigenvalues for factors 1 and 2 exceed 1 and are equal to 3.47 and 1.92, respectively. Except for factor 3, other factors have coefficients smaller than 1. To determine the most important parameters justifying the variance, eignvalues bigger than 1 were used.

Since the value calculated for KMO was 0.56, it confirmed the use of PFA and PCA in this study.

Fig. 4. Scree plot of the eigenvalues

Conclusion

In this study to assess the quality of water and to determine the sources of pollution in the Gharasou river basin in Ardabil plain, multivariate statistical procedures including CA, PCA and FA were used. The two dimensional representation of the first two principle components confirmed the groupings resulting from CA and FA. This way, the stations understudies were differentiated like CA. The results show the effect of different pollutant factors in the environment on the quality of water. According to the findings of this research, these methods can be used, with high confidence, in the management of environmental monitoring of surface water resources. The findings are in accordance with the findings in the Tahtali river in Turkey[4] and in the Daliao river in China[2].

Using CA method, 11 sampling stations were divided into three clusters with similar qualitative features. The results obtained from groupings, like the findings in Fuji river in Japan[7], in the Gomti river in India[5] and in the Daliao river in China[2] showed that the number of sampling stations and associated monitoring costs can be reduced without missing much information. PCA and FA in determining the parameters and sources effective in the change of water quality helped dividing the river into three different areas. The obtained KMO coefficient, too, showed that this grouping is reliable. The first two principle components of PCA and the first two factors of FA showed that the main parameters responsible for the changes in the quality of water are heavy metals (existent in the region's soil) including Fe and Mn and Al and Diazinon released in the environment because of agricultural activities. Therefore we can conclude that: (i) Agricultural drainage, waste from warm mineral waters and waste from industrial activities are the main sources for the water quality deterioration. (ii) Using multivariate statistical techniques is useful in assessing water quality, determining the amount of pollutants, determining pollution sources and making relevant data available concerning the water quality, designing water quality monitoring network and overall management of water quality.

REFERENCES

1. A. Astel, M. Biziuk, A. Przyjazny and J. Namiesnik, *Water Res*., **40**, 1706 (2006).
2. Y. Zhang, F. Guo and W. Meng, Water Quality Assessment and Source Identification of Daliao River Basin Using Multivariate Statistical Methods, Environmental Monitoring and Assessment, DOI 10.1007/S10661-008-0300-z (2008).
3. Y. Ouyang, *Water Res*., **39**, 2621 (2005).
4. Hu. Boyacioglu and H. Boyacioglu, *Turkey Evniron. Geol*., **54**, 275 (2007).
5. K.P. Singh, A. Malik, D. Mohan and S. Sinha, *Anal. Chim. Acta*, **538**, 355 (2005).
6. K.P. Singh, A. Malik, D. Mohan and S. Sinha, *Water Res*., **38**, 3980 (2004).
7. S. Shrestha and F. Kazama, *Environ. Mode. Soft*., **22**, 464 (2007).
8. Ardabil Regional Water, Ardabil Dam Irrigation and Drainage Network Design, Consulting Engineers Band Ab (2004).
9. Ardabil Regional Water, Environmental Impact Assessment of Sabalan Dam and Network, Nepta Consultative Engineering Corporation (2005).
10. Standard Methods for the Examination of Water and Wastewater, American Public Health Association/American Water Works Association/Water Environment Federation, Washington DC, USA, edn. 21 (2005).
11. J.E. McKenna Jr., *Environ. Mod. Soft.*, **18**, 205 (2003).
12. I.T. Jolliffe, Principal Component Analysis, Springer-Verlag, p. 271 (1986).
13. S.A. Mohammadi and B.M. Prasanna, *Crop Sci.*, **43**, 1235 (2003).
14. M. Vega, R. Pardo, E. Barrado and L. Deban, *Water Res.*, **32**, 3581 (1998).
15. A.L. Panchen, Classification, Evolution and the Nature of Biology, Cambridge University Press, Cambridge, England (1992).
16. J.H. Ward Jr., *J. Am. Statist. Assoc*., **58**, 236 (1963).
17. M. Otto, in eds.: R. Kellner, J.M. Mermet, M. Otto and H.M. Widmer, Multivariate Methods, Analytical Chemistry, Wiley-VCH, Weinheim (1998).
18. C. Sarbu and H.F. Pop, *Talanta*, **65**, 1215 (2005).
19. B. Helena, R. Pardo, M. Vega, E. Barrado, J.M. Fernandez and L. Fernandez, *Water Res*., **34**, 807 (2000).
20. G. Brumelis, L. Lampina, O. Nikodemus and G. Tabors, *Environ. Mod. Soft.*. **15**, 755 (2000).
21. S.A. Abdul-Wahab, C.S. Bakheit and S.M. Al-Alawi, *Environ. Mod. Soft.*, **20**, 1263 (2005).
22. D. Love, D. Hallbauer, A. Amos and R. Hranova, *Phy. Chem. Earth*, **29**, 1135 (2004).
23. D.A. Wunderlin, M. Diaz, M.M.V. AME, S.F. Pesce, A.C. Hued and M. Bistoni, *Water Res*., **35**, 2881 (2001).
24. L. Kelly and B.F.J. Manly, Multivariate Statistical Methods: A Primer, Chapman & Hall/Crc, p. 214 (2005).

(*Received*: 18 June 2009; *Accepted*: 26 December 2009) AJC-8229