PERGAMON

Water Research 37 (2003) 4119–4124

WATER RESEARCH

www.elsevier.com/locate/watres

Assessment of the surface water quality in Northern Greece

V. Simeonov[a,*], J.A. Stratis[b], C. Samara[c], G. Zachariadis[b], D. Voutsa[c], A. Anthemidis, M. Sofoniou[b], Th. Kouimtzis[c]

[a] *Faculty of Chemistry, University of Sofia "St. Kl. Okhridski", J. Bourchier Blvd. 1, Sofia 1164, Bulgaria*
[b] *Laboratory of Analytical Chemistry, Department of Chemistry, Aristotle University of Thessaloniki, Thessaloniki 540 06, Greece*
[c] *Environmental Pollution Control Laboratory, Department of Chemistry, Aristotle University of Thessaloniki, Thessaloniki 540 06, Greece*

Received 2 July 2002; accepted 16 June 2003

Abstract

The application of different multivariate statistical approaches for the interpretation of a large and complex data matrix obtained during a monitoring program of surface waters in Northern Greece is presented in this study. The dataset consists of analytical results from a 3-yr survey conducted in the major river systems (Aliakmon, Axios, Gallikos, Loudias and Strymon) as well as streams, tributaries and ditches. Twenty-seven parameters have been monitored on 25 key sampling sites on monthly basis (total of 22,350 observations). The dataset was treated using cluster analysis (CA), principal component analysis and multiple regression analysis on principal components. CA showed four different groups of similarity between the sampling sites reflecting the different physicochemical characteristics and pollution levels of the studied water systems. Six latent factors were identified as responsible for the data structure explaining 90% of the total variance of the dataset and are conditionally named organic, nutrient, physicochemical, weathering, soil-leaching and toxic-anthropogenic factors. A multivariate receptor model was also applied for source apportionment estimating the contribution of identified sources to the concentration of the physicochemical parameters. This study presents the necessity and usefulness of multivariate statistical assessment of large and complex databases in order to get better information about the quality of surface water, the design of sampling and analytical protocols and the effective pollution control/management of the surface waters.

Keywords: Water quality; River water; Statistical analysis; Nutrients; Heavy metals

1. Introduction

The quality of surface waters is a very sensitive issue. Anthropogenic influences (urban, industrial and agricultural activities, increasing consumption of water resources) as well as natural processes (changes in precipiation inputs, erosion, weathering of crustal materials) degrade surface waters and impair their use for drinking, industrial, agricultural, recreation or other purposes [1,2]. Due to spatial and temporal variations in water chemistry a monitoring program that will provide a representative and reliable estimation of the quality of surface waters is necessary. Thus, monitoring programs including frequent water samplings at many sites and determination of a large number of physicochemical parameters are usually conducted resulting in a large data matrix, which needs a complex data interpretation [3].

The application of different multivariate approaches (cluster analysis (CA), principal components analysis, source apportionment by multiple regression on principal components) for the interpretation of these complex data matrices offers a better understanding of water quality and ecological status of the studied systems, allows the identification of the possible

*Corresponding author.
E-mail address: vsimeonov@chem.uni-sofia.bg (V. Simeonov).

0043-1354/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/S0043-1354(03)00398-1

factors/sources that influence the water systems and offers a valuable tool for reliable management of water resources as well as rapid solutions on pollution problems [4–6].

A 3-yr survey (February 1997–January 2000) aiming to establish national databases on surface water quality has been performed in surface waters in Northern Greece. The procedure established by the European Community to assess pollution levels in rivers and lay down guidelines for the control of pollution and nuisance of the environment was followed [7,8]. Results on the speciation of nitrogen and phosphorus, the eutrophication status and the distribution of heavy metals—toxic elements in surface waters are presented in previous papers [9,10].

In the present article, the large data matrix obtained during the 3-yr monitoring program (22,350 observations) subjected to different multivariate statistical approaches in order to extract information about: (a) the similarities or dissimilarities between sampling sites and/or river systems, (b) the latent factors explaining the structure of the dataset, (c) the influence of possible sources (natural and anthropogenic) on the physico-chemical parameters and (d) source apportioning for the estimation of the contribution of possible sources on the concentration of determined parameters.

2. Experimental

2.1. Monitoring sites

The region of sampling covers a wide range of catchments and surface water types (rivers, streams, tributaries, ditches). The main river systems are Aliakmon, Axios, Loudias, Strymon and Gallikos. These rivers drain the major rural, agricultural, urban and industrial areas of Northern Greece and discharge into the North Aegean Sea (Fig. 1). The main pollutant loads include domestic wastewaters, agricultural runoff, animal husbandry and industrial effluents.

Aliakmon river is the longest Greek river with a total length of nearly 310 km and mean flowrate 42 $m^3 s^{-1}$. It originates from northwestern Greece and after traversing a basin of about 5600 km^2 discharges into the Thermaikos Gulf. Significant pollution loadings are discharged into the lower stream of Aliakmon river due to its confluence with Ditch-66. This ditch, which also receives water from smaller streams, is a major recipient for the effluents produced by a number of local industries, principally vegetable, fruit and juice canneries. The ditch water is expected to be highly polluted with organic matter. *Axios* river originates from FYROM and flows into Thermaikos Gulf. Of its total

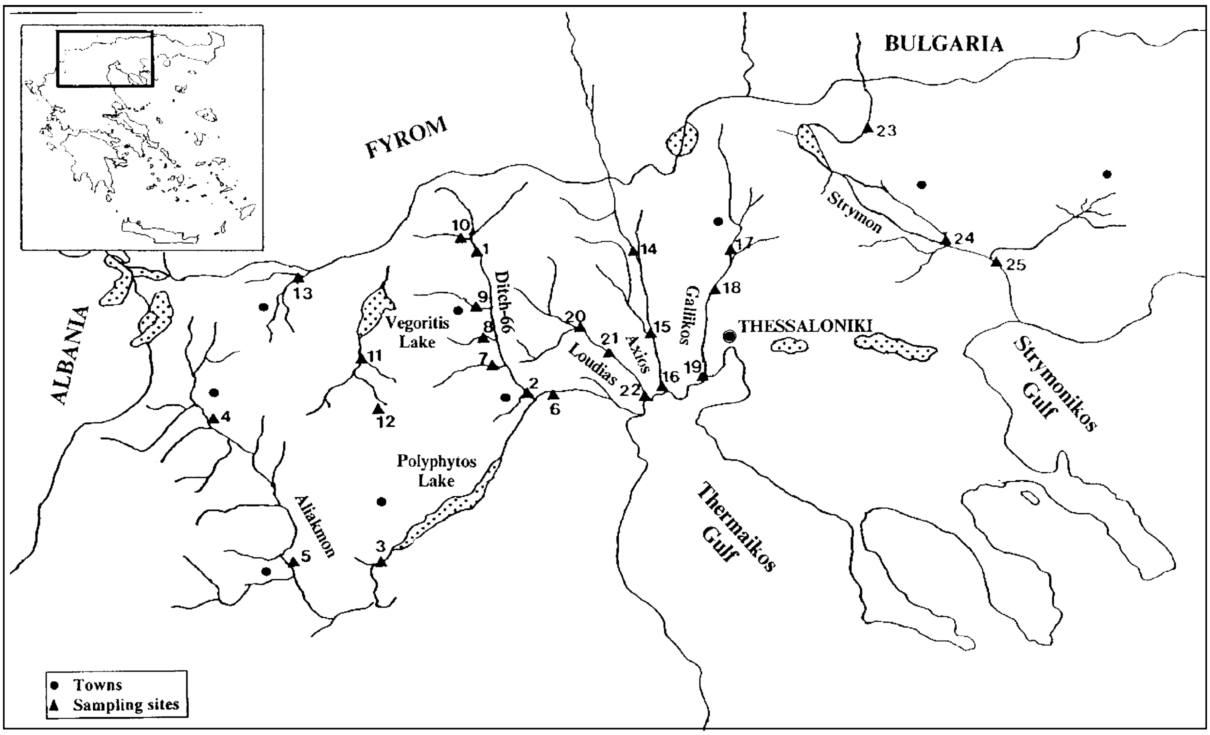

Fig. 1. Map of the sampling site locations.

length (nearly 300 km) only the lowest 75 km flow through Greece with a mean flowrate $65\,m^3\,s^{-1}$. The estuaries of the river form an extended delta, which is protected by the Ramsar convention. *Loudias* river with a length of 38 km starts from the mountainous area of northern Greece and drains an intensively cultivated area with a flow rate of about $20\,m^3\,s^{-1}$. This river was found in the past to be relatively polluted by organic substances and heavy metals released from food processing, sugar and diary industries [11]. *Strymon* river originates from Bulgaria and flows into the northern Aegean Sea. The river has a total length of 330 km, of which 115 km flow through Greece. The mean flowrate is $28\,m^3\,s^{-1}$. It is used to irrigate an extensive cultivated plain. *Gallikos* river is characterized by very low flowrate (less than $10\,m^3\,s^{-1}$ at the upper part) which periodically drops to zero in the lower part of its route, thus usually the water never reaches the mouth of the river. The river is not used for irrigation, however is supplies aquifers used for drinking purposes.

2.2. Sampling and chemical analysis

The sampling strategy was designed in such a way to cover a wide range of determinants at key sites that accurately represent the water quality of the river systems and account for tributary inputs that can have important impacts upon downstream water quality. Samples were taken from upper, mid and downstream sites. Samples were also collected from tributaries, small streams and open channels-ditches (Fig. 1).

Monthly sampling was carried to monitor changes caused by the seasonal hydrological cycle. Sampling, preservation and analytical protocols were conducted by standard methods for surface waters and have been previously reported [9,10,12]. The measured parameters include field pH, electrical conductivity (EC), dissolved oxygen (DO), total suspended solids (TSS), nitrate (NO_3^-), nitrite (NO_2^-), ammonium (NH_4^+), orthophosphate (PO_4^{3-}), chemical oxygen demand (COD), biological oxygen demand (BOD_5), organic (Kjeldahl) nitrogen (TON), acid-hydrolysable (total) phosphorus (TP). The acid-available fraction of metals and other toxic elements (Ag, As, B, Ba, Cd, Cr, Cu, Hg, Fe, Mn, Ni, Pb, Se and Zn) was also determined. The data quality was checked by careful standardization, procedural blank measurements, spiked and duplicate samples [9,10]. In Table 1 the summarized basic statistics of the dataset is presented.

Table 1
Summary basic statistics (concentration units in $mg\,L^{-1}$ or $\mu g\,L^{-1}$ for metal ions; conductivity in $\mu s\,cm^{-1}$)

Component	Mean	Stand. dev.	Minimum	Maximum
COD	12.2	11.6	4.0	94.0
BOD_5	11.4	9.3	2.0	8.0
TON	0.62	0.71	0.02	2.55
TP	0.57	0.63	0.14	1.97
PO_4^{3-}	0.22	0.26	0.06	0.53
NO_2^-	0.21	0.27	0.01	1.56
NO_3^-	0.38	0.34	0.3	10.2
NO_4^+	1.22	1.06	0.03	3.08
pH	8.1	0.26	7.7	8.6
DO	7.4	1.2	3.7	12.3
EC	421	193	126	690
TSS	17.7	14.4	6.2	45.8
B	52.3	38.1	19.0	104.0
Ba	47.2	17.1	31.0	78.0
Cu	4.2	2.4	2.0	7.0
Cr	6.5	5.5	1.0	18.0
Ni	4.1	2.9	2.0	12.0
Mn	155.4	102.3	45	291
Fe	326.6	211.9	113	833
Pb	3.4	3.1	1.0	16.0
Zn	57.2	44.8	20	157
Cd	0.26	0.19	0.1	0.6
Se	Less than 0.1	—	Less than 0.1	Less than 0.1
As	Less than 0.1	—	Less than 0.1	Less than 0.1
Hg	Less than 0.2	—	Less than 0.2	Less than 0.2
Ag	1.1	0.02	1.0	3.0

2.3. Multivariate statistical methods

Multivariate statistical methods for classification, modeling and interpretation of large datasets from environmental monitoring programs allow the reduction of the dimensionality of the data and the extraction of informations that will be helpful for the water quality assessment and the management of surface waters [13]. Thus, various environmetric approaches for the assessment of water quality are frequently employed [3–6,10,14–18]. In the present study, CA, principal component analysis (PCA) and source apportionment by multiple regression analysis on principal components (PC/MR) are employed in a dataset of almost twenty thousand values (21 parameters determined at 25 sampling sites for a period of 36 months). The elements As, Hg, Ag and Se, which exhibited values usually lower than the detection limit of the method were excluded. Missing data were completed by mean values of the neighbor data. The STATISTICA 5.0 software package was employed for data treatment.

3. Results and discussion

3.1. Site similarity

CA was applied to detect similarity groups between the sampling sites. The dataset was treated (after data scaling by *z*-transformation) by the Ward's method of linkage with squared Euclidean distance as measure of similarity. The significance of the clusters obtained was tested by the Sneath's index of disjunction [19].

Four statistically significant clusters are formed: Cluster 1 (sampling sites 12, 14–16, 20–22) corresponds to Axios and Loudias rivers, Cluster 2 (sampling sites 1–6, 11) corresponds Aliakmon river and Ditch-66, Cluster 3 (sampling sites 17–19, 23–25) corresponds to rivers Strymon and Gallikos, both used mainly for irrigation and Cluster 4 (sampling sites 7–10, 13) corresponds to streams discharged into Ditch-66. The clustering procedure reveals the groups of similar sites in a quite convincing way. These clusters include sampling sites with similar characteristics features and natural background that are affected by sources of similar type/strength.

The CA carried out using our data indicates that this approach makes possible the design of a future spatial sampling strategy in an optimal way and offers a reliable classification of surface waters in the whole region. For instance, the number of the sampling site could be optimized in such a way that for rapid quality assessment studies only representative sites from each cluster (not all monitoring sites) can be used. This reduces the number of analysis and the cost of the risk assessment procedure.

Table 2
Varimax rotated factor matrix for the whole data set[a]

Variable	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6
COD	**0.91**	0.38	−0.26	0.09	0.11	0.06
BOD_5	**0.84**	0.29	−0.18	0.10	0.12	0.08
TON	**0.73**	0.41	0.11	0.21	0.09	0.23
TP	**0.70**	0.61	0.19	0.31	0.11	0.13
PO_4^{3-}	0.67	0.58	0.22	0.29	0.34	0.08
NO_2^-	0.33	**0.88**	0.26	0.24	−0.09	0.18
NO_3^-	0.36	**0.71**	0.29	0.16	0.11	0.24
NO_4^+	0.38	**0.86**	0.14	0.16	0.21	0.26
pH	0.16	0.29	**0.84**	0.08	−0.17	0.11
DO	0.21	0.22	**0.86**	0.14	0.25	0.28
EC	0.34	0.27	**0.82**	0.07	0.11	0.16
TSS	0.42	0.28	0.33	0.01	**0.79**	0.14
B	0.23	0.19	0.07	**0.71**	0.30	0.24
Ba	0.19	0.31	0.12	**0.72**	0.28	0.26
Cu	0.28	0.23	−0.08	0.36	**0.84**	0.39
Cr	0.16	0.19	0.11	0.27	**0.72**	0.43
Ni	0.09	0.16	0.07	0.28	**0.70**	0.39
Mn	0.23	0.18	0.14	0.29	**0.74**	0.34
Fe	0.31	0.15	0.05	0.34	**0.73**	0.27
Pb	0.06	0.15	0.12	0.23	0.35	**0.78**
Zn	0.11	0.08	0.14	0.29	0.37	**0.70**
Cd	0.06	0.09	0.07	0.37	0.25	**0.77**
Expl. Var.	22.1%	19.8%	15.5%	13.0%	10.9%	8.2%

[a] Significant factor loadings are bold faced.

3.2. Data structure determination and source identification

PCA is a powerful pattern recognition technique that attempts to explain the variance of a large dataset of intercorrelated variables with a smaller set of independent variables (principal components) [20]. PCA was employed on our dataset to compare the compositional patterns between the examined water systems and to identify the factors that influence each one.

Six principal components were obtained with eigenvalues >1 summing almost 90% of the total variance in the water dataset (Table 2). The first PC, accounting for 22.1% of the total variance was correlated with COD, BOD_5, TON, TP and PO_4^{3-}. This "organic" factor may be interpreted as representing influences from point sources such as municipal and industrial effluents. The second PC, accounting for 19.8% of the total variance was correlated primarily with water soluble N-species, NO_2^-, NH_4^+ and NO_3^- and secondarily with PO_4^{3-} or TP. This "nutrient" factor represents influences from non-point sources such as agricultural runoff and atmospheric deposition. The third PC was weighted on pH, DO and EC and represents the "physicochemical" source of the variability. The fourth PC was loaded on B and Ba, probably represent "weathering" processes from rocks, deposits of marine origin, salt intrusion etc. The fifth PC was highly loaded with TSS, Cu, Cr, Ni, Mn and Fe and is likely to represent "soil leaching" processes. Finally the sixth factor was correlated with Pb, Zn and Cd and is considered as representing "anthropogenic-toxic" pollution from metal activities/industrial effluents.

3.3. Source apportionment

The source apportionment is an important environmetric approach aiming to the estimation of contribution of identified sources to the concentration of each parameter. After the determination of the number and identity of possible sources affecting surface waters by using PCA, source contributions were calculated next by using multiple regression of sample mass concentration on the absolute principal component scores (APCS). A detailed description of the modeling approach can be found elsewhere [21]. It makes it possible to apportion the component mass among various source components obtained by PCA (six PCs in our case). The PCA assumes that the total concentration of each element is made up of the sum of elemental concentrations from each identified polluting or natural source component. The approach calculates the weight of source in the total sum using multiple regression.

Table 3
Source contribution to surface waters in Northern Greece

Variables	Intcpt	Source types						
		1 Organic	2 Nutrient	3 Nutrient	4 Weathering	5 Soil-leaching	6 Toxic	r^2
COD	3.8	72.3	23.9	—	—	—	—	0.54
BOD_5	4.2	83.3	12.5	—	—	—	—	0.56
TON	—	66.8	23.8	—	5.0	—	4.4	0.64
TP	2.8	70.1	20.4	—	6.7	—	—	0.59
PO_4^{3-}	7.0	45.1	30.2	5.2	6.0	6.5	—	0.67
NO_2^-	2.5	29.1	59.9	4.0	4.5	—	—	0.49
NO_3^-	6.7	30.2	60.3	—	—	—	2.8	0.57
NH_4^+	2.4	28.0	62.6	—	—	4.0	3.0	0.62
pH	3.9	—	—	96.1	—	—	—	0.38
DO	6.2	3.8	3.2	86.8	—	—	—	0.44
EC	5.8	11.2	10.3	72.7	—	—	—	0.49
TSS	4.4	15.5	6.1	8.0	—	66.0	—	0.61
B	7.2	—	11.2	—	69.6	6.2	5.8	0.64
Ba	9.0	—	10.6	—	72.0	4.8	4.0	0.59
Cu	4.9	4.1	3.9	—	8.2	68.7	10.2	0.72
Cr	1.6	—	—	—	5.1	72.9	20.4	0.74
Ni	—	—	—	—	12.4	79.6	8.0	0.62
Mn	6.0	—	—	—	—	75.5	18.5	0.55
Fe	4.4	7.2	—	—	16.4	72.2	9.8	0.64
Pb	4.1	—	—	—	—	21.1	74.8	0.66
Zn	5.5	—	—	—	—	16.6	77.9	0.71
Cd	8.8	—	—	—	9.4	—	81.8	0.53

Intcpt: intercept of the multiple regression.

The contribution of the possible sources in each physicochemical parameter is presented in Table 3. The multiple regression exhibited relatively good adequacy between observed and predicted values for most parameters, as shown by the correlation coefficients. A significant biased part of the total concentration is present for almost all parameters. Results showed that point (municipal and industrial effluents) and nonpoint sources (agricultural runoff) are the main contributors to organic and nutrient parameters. Soil leaching appeared to be the main source of most elements whereas anthropogenic activities contribute highly to Pb, Zn and Cd concentrations. It is believed that these apportionment results could be very useful to the local authorities for the pollution control/management of the examined surface waters.

References

[1] Carpenter SR, Caraco NF, Correll DL, Howarth RW, Sharpley AN, Smith VH. Nonpoint pollution of surface waters with phosphorus and nitrogen. Ecol Appl 1998;8(3):559–68.

[2] Jarvie HP, Whitton BA, Neal C. Nitrogen and phosphorus in east coast British rivers: speciation, sources and biological significance. Sci Total Environ 1998;210–211:79–109.

[3] Chapman D. Water Quality Assessment. In: Chapman D on behalf of UNESCO, WHO and UNEP, London: Chapman & Hall; 1992. 585p.

[4] Morales MM, Martih P, Llopis A, Campos L, Sagrado J. An environmental study by factor analysis of surface seawater in the Gulf of Valencia (western Mediteranean). Anal Chim Acta 1999;394:109–17.

[5] Reghunath R, Murthy TRJ, Raghavan BR. The utility of multivariate statistical techniques in hydrogeochemical studies: an example from Karnataka, India. Water Res 2002;36:2437–42.

[6] Wunderlin DA, Diaz MP, Ame MV, Pesce SF, Hued AC, Bistoni M. Pattern recognition techniques for the evaluation of spatial and temporal variation in water quality. A case study: Suquia river basin (Cordoba Argentina). Wat Res 2001;35(12):2881–94.

[7] European Communities, Council Decision 77/795/EEC establishing a common procedure for the exchange of information on the quality of surface fresh water in the Community, 1977.

[8] European Communities, Council Decision 86/574/EEC, amending Decision 77/795/EEC establishing a common procedure for the exchange of information on the quality of surface fresh water in the Community, 1986.

[9] Voutsa D, Zachariadis G, Kouras A, Anthemidis A, Samara C. A quality study of surface waters of Macedonia, Greece. Heavy metals and other toxic elements. Proceedings of the International Conference on Instrumental methods of analysis. Modern trends and applications, Greece. Proc. September 1999; Vol. II: 83–87.

[10] Voutsa D, Manoli E, Samara C, Sofoniou M, Stratis I. A study of surface water quality in Macedonia, Greece: speciation of nitrogen and phosphorus. Water Air Soil Pollut 2001;129:1–20.

[11] Kouimtzis Th, Samara C, Voutsa D, Zachariadis G. Evaluation of chemical parameters in Aliakmon River, Northern Greece. Part I: quality characteristics and nutrients. J Environ Sci Health 1994;A29(10):2115–26.

[12] APHA, AWWA, WPCF. Standard methods for the examination of water and wastewater. 16th Edition. Washington, DC EPA; 1985.

[13] Massart DL, Kaufman L. The interpretation of analytical chemical data by the use of cluster analysis. New York: Wiley; 1983.

[14] Chen GN. Assessment of environmental water with fuzzy cluster analysis and fuzzy recognition. Anal Chim Acta 1993;271:115–24.

[15] Hagedorn C, Robinson SL, Filtz JR, Grubbs SM, Angier TA, Renaeau JRB. Determining sources of fecal pollution in a rural Virginia watershed with antibiotic resistance patterns in fecal streptococci. Appl Environ Microbiol 1999;65:5522–31.

[16] Voutsa D, Samara C, Kouimtzis Th. Groundwater quality in the major industrial area of Thessaloniki, Greece. Part 1: chemical composition—geochemical processes. Toxicol Environ Chem 1994;45:105–19.

[17] Voutsa D, Zachariadis G, Samara C, Kouimtzis Th. Evaluation of chemical parameters in Aliakmon River/Northern Greece. Part II. Dissolved and particulate heavy metals. J Environ Sci Health 1995;A30(1):1–13.

[18] Wiggins BA, Andrews RW, Conway RA, Corr CL, Dubratz EJ, Dongherty DP, Eggard JR, Knupp S, Limjoco MC, Mettenburg JM, Rinehardt JM, Sonsino J, Torrijos RE, Zimmerman ME. Use of antibiotic resistance analysis to identify nonpoint sources of fecal pollution. Appl Environ Microbiol 1999;65:3483–6.

[19] Sneath PHA. Data analysis and informatics. Amsterdam: North-Holland; 1980.

[20] Hopke PK. Receptor modeling in environmental chemistry. USA: Wiley; 1985.

[21] Thurston GD, Spengler JD. A quantitative assessment of source contributions to inhalable particulate matter pollution in Metropolitan Boston. Atmos Environ 1985;19(1):9–25.