Environmental Engineering and Management Journal

March 2019, Vol.18, No. 3, 719-726
http://www.eemj.icpm.tuiasi.ro/; http://www.eemj.eu

"Gheorghe Asachi" Technical University of Iasi, Romania

APPLICATION OF MULTIVARIATE STATISTICAL TECHNIQUES IN WATER QUALITY ASSESSMENT OF DANUBE RIVER, ROMANIA

Alhassan H. Ismail[1,2,*], Diana Robescu[1]

[1]University Politehnica of Bucharest, Faculty of Power Engineering, Department of Hydraulics and Environmental Engineering, Romania
[2]Middle Technical University, Institute of Technology-Baghdad, Baghdad, Iraq

Abstract

Multivariate statistical methods such as factor analysis (FA) and cluster analysis (CA) were applied to identify the main factors and pollution sources affecting on the water quality of Danube River and to evaluate spatial and temporal similarities or dissimilarities among the sampling sites and monitoring periods. Water quality of 14 parameters has been considered in four sampling stations during 1 year period. The results of factor analysis specified three factors representing 64.369% of the total variance in water quality. The main differences are related to domestic wastewater, industrial discharge, and agriculture activities from agricultural area. Hierarchical spatial CA revealed two different groups of similarities between the sampling sites, reflecting different pollution levels in the river water quality. Hierarchical temporal CA showed that the temporal variations in the river water are not relying on the local climate.

Keywords: cluster analysis, Danube River, Drobeta-Turnu Severin, factor analysis, river water

Received: January, 2016; Revised final: July, 2016; Accepted: August, 2016; Published in final edited form: March, 2019

1. Introduction

Danube River is one of the largest rivers in Europe. It originates from the Black Forest in Germany. The total length of the river is about 2860 km, with a catchment area of 817,000 km^2 (Garnier et al., 2002). The river has a mean annual discharge of 5630 m^3/s (Gan et al., 2012). The river passes through nine states and five capitals and about 2.5 million inhabitants contributing to extensive water use and pollution (Kirschner et al., 2009).

Many studies have been conducted on the water quality of the Danube river and management of river basin using different tools and techniques (Antanasijevic et al., 2014; Bajo et al., 2014; Derx et al., 2010; Gan et al., 2012; Garnier et al., 2002; Janauer et al., 2010; Kirschner et al., 2009; Matenco and Andriessen, 2013; Popa et al., 2018; Popescu et al., 2015; Szolgayova et al., 2014; Vargha et al., 2005).

The evaluation and monitoring of Danube river composition is necessary in order to control the quality of the water. Periodic assessment at regular interval of the river water through measurement of huge numbers of variables is a time-consuming, toilsome and expensive process (Das et al., 2010). In contrast, multivariate statistical techniques can be utilized to simplify this issue. Multivariate statistical methods such as factor and cluster analyses can provide better explanation of the relationship between the large numbers of variables in surface water (Chapagain et al., 2010). These methods have numerous advantages such as provide reasonable interpretation of complex data sets, possibility of considering multiple variables simultaneously and the ability to identify the potential factors/sources that affect the water system being studied (Boyacioglu and Boyacioglu, 2007; Ismail et al., 2014). In addition, these techniques help to avoid misinterpretation of environmental monitoring data.

*Author to whom all correspondence should be addressed: hassan19851988@yahoo.com; Phone: +40 732059715

Therefore, factor and cluster analyses may be helpful in looking for solutions to water pollution challenges and can be an effective tool for water resources management (Simeonov et al., 2004; Singh et al., 2004; Wunderlin et al., 2001).

The present study aimed at the utilization of these techniques to identify the major factors affecting water quality of Danube River in Drobeta-Turnu Severin city and to get information about the spatial and temporal variations among the sampling sites and monitoring periods. This may provide a proper basis for management of Danube River in the study region.

2. Materials and methods

2.1. Study area and water quality data

Previous studies reported that the river can be classified into three main segments:

- the upper Danube course with a length of 1060 km,
- the middle Danube course with a length of 725 km,
- the lower Danube course with a length of 1075 km.

The lower Danube course represents Romania's natural border with its neighbor countries (Gâştescu and Ţuchiu, 2012). In the lower course, the river is flowing through Baziaş and Gura Văii passing the Iron Gate I. The Iron Gate I was constructed in 1971 and considered as the largest dam and reservoir system on the basis of volume, area and hydropower potential among numerous impoundments on the Danube and the tributaries (Teodoru and Wehrli, 2005). Drobeta-Turnu Severin is a city in Mehedinți County of Romania and administers three villages: Gura Văii, Dudașu Schelei, and Schela Cladovei (Fig. 1).

This study covers 13 km of the Danube River starting at Gura Văii, 2 km downstream of Iron Gate I, and extends to Drobeta-Turnu Severin city. The study area is characterized by lacking proper sewage collection and treatment facilities, in addition to the effluent discharges from industrial areas in the region (Andrița, 2012; Muntean and Morariu, 2014). In addition, agricultural practices from Serbian part use fertilizers and pesticides which are being discharged to the river through surface runoff (Gavrilescu, 2011).

In the present paper, data sets of 14 parameters obtained during 1 year (2008) were subjected to multivariate statistical methods (factor and cluster analyses). The data was collected at monthly intervals in four points namely, Gura Văii (SS1) which is about 2 km downstream of Iron Gate I, Dudașu Schelei (SS2), Schela Cladovei (SS3), Drobeta-Turnu Severin (SS4) (Fig. 1).

Grab sampling procedure was adopted for the analysis of various water quality parameters as recommended by standard methods (APHA, 1998). The water samples were collected in polypropylene bottles. Only for BOD estimation, samples were collected in BOD bottles. Table 1 shows the water quality parameters, alongside some of the abbreviations, units and techniques used for the water samples analysis. The descriptive statistics of water quality data are shown in the Table 2.

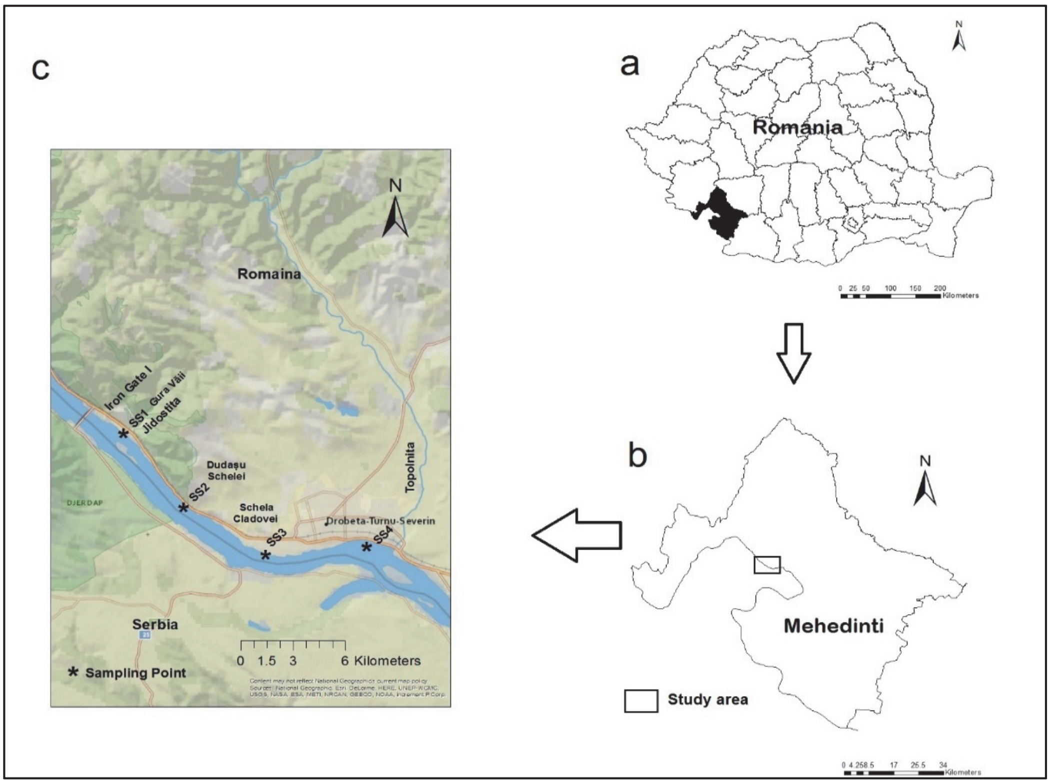

Fig. 1. Map of the study area: a) Romanian Counties, b) Mehedinți County and c) Sampling locations

Table 1. Water quality parameters, abbreviations, units and analytical methods used for Danube River

Parameter	*Abbreviation*	*Units*	*Instrument/techniques used*
Biochemical oxygen demand	BOD	mg/L	Winkler azide method
Dissolved oxygen	DO	mg/L	Winkler azide method
Ammonium	NH_4	mg/L	Spectrophotometric
Nitrate nitrogen	NO_3	mg/L	Spectrophotometric
Total phosphorous	TP	mg/L	Spectrophotometric
Water temperature	WT	˚C	Mercury thermometer
pH	pH	-	pH-meter
Discharge	Q	m^3/sec	Current meter
Total suspended solids	TSS	mg/L	Gravimetric
Cadmium	Cd	µg/L	Flame atomic absorption spectrophotometer
Cooper	Cu	µg/L	Flame atomic absorption spectrophotometer
Chromium	Cr	µg/L	Flame atomic absorption spectrophotometer
Nickel	Ni	µg/L	Flame atomic absorption spectrophotometer
Lead	Pb	µg/L	Flame atomic absorption spectrophotometer

Table 2. Descriptive statistical summary of Danube river water quality and quantity data in four sampling stations during 12 months in the year of 2008, n = 48

Parameters	*Minimum*	*Maximum*	*Mean*	*Std. Deviation*
Dissolved oxygen	5.61	12.69	9.12	2.09
Biochemical oxygen demand	1.15	2.37	1.66	0.30
Ammonium	0.087	0.522	0.19	0.07
Nitrate nitrogen	0.195	3.614	1.98	0.88
Total phosphorous	0.132	1.44	0.43	0.27
Water Temperature	4	27	15.72	7.27
pH	7.1	7.7	7.43	0.15
Discharge	2650	8760	5643	1700
Total suspended solids	21	34	26.44	2.93
Cadmium	0.11	0.44	0.28	0.07
Cooper	1.5	4	2.16	0.71
Chromium	1.4	3.5	1.97	0.53
Nickel	1.1	1.8	1.43	0.19
Lead	0.7	1.9	1.37	0.34

2.2. Data treatment and multivariate analysis

The normality distribution of each variable was checked using the Shapiro–Wilk (W) test prior to using multivariate statistical methods (Chapagain et al., 2010). Since, the factor and cluster analyses require water quality variables to fit to the normal distribution. The Shapiro–Wilk (W) test demonstrated that the variables: BOD, pH, Q, TSS, Cd, Ni and Pb were normally distributed, whereas DO, TP, NH_4, NO_3, WT, Cu and Cr were not normally distributed. Consequently, the original data of non-normal distribution variables were transformed in the form $x' = \log_{10}(x)$ (Zhou et al. 2007). After log-transformation and according to the Shapiro–Wilk (W) test, only DO and TP were normalized and thus, other variables: NH_4, NO_3, WT, Cu and Cr were excluded. Thereafter, in order to avoid misclassification due to the wide variations in data dimensionality, the original data of variables: BOD, pH, Q, TSS, Cd, Ni, Pb and the log-transformed variables: DO and TP were also standardized by setting the mean equal to 0 and the variance equal to one (Singh et al., 2004).

Factor analysis (FA) was applied to identify the most important variables influencing the water quality of the river in Drobeta-Turnu Severin. FA were conducted via three stages, (i) generating the correlation matrix for all variables, (ii) extracting the initial set of factors using principal component analysis (PCA) method, and (iii) rotating the extracted factors by Varimax rotation (Ismail et al., 2015). In factor analysis, the basic concept is given by (Eq. 1) (Altın et al., 2009):

$$H_j = b_{j1}f_1 + b_{j2}f_2 + \dots + b_{jn}f_n + E_{ij};\ j = 1,2,\dots,p \tag{1}$$

where, H = measured value, f = factor score, b = factor loading, E = residual term accounting for errors or other sources of variation, i = number of the sample, j = number of the variable, and n = total factors number.

Factor analysis was performed using principal component analysis method on the normalized data sets and Varimax rotation was made on the factor loading matrix to infer the principal parameters. Eigenvalues of 1.0 or greater are considered significant and eigenvalues less than 1 have been eliminated (Shrestha and Kazama, 2007). The Scree plot was used to identify the retained factors in order to comprehend the underlying data structure and shown in Fig. 2 (Altın et al., 2009).

Cluster analysis (CA) was also utilized to evaluate the spatial and temporal variations among the sampling sites and monitoring periods. In this paper, hierarchical agglomerative clustering was made using Ward's method of linkage, and squares Euclidean distance method was used for determining similarity distance, and the results of CA illustrated by a dendrogram (Palma et al., 2010; Varol and Sen, 2009; Wunderlin et al., 2001). All the statistical computations were performed using Microsoft Office – Excel 2007 spreadsheet and SPSS 18 for Windows.

3. Results and discussion

3.1. Factor analysis (FA)

Factor loadings with values of >0.75, 0.75–0.50 and 0.50–0.30 classified as strong, moderate and weak, respectively (Liu et al., 2003). As shown in the scree plot (Fig. 2), it can be noticed that 3 major factors have eigenvalues greater than one and explain 64.369% of the total variance in each water quality datasets.

Table 3 shows the output results of factor analysis (FA). Factor 1 explained 26.11% of the total variance and has strong positive correlation with TSS, negative moderate correlation with Q and BOD, and positive moderate correlation with Cd and Pb. The high loading factor of TSS is likely due to the high discharge of the river. Despite, geology-related factors such as Na, K, Ca and Mg were not considered in this study, it is clear that the high loading factor of TSS and the negative moderate correlation with Q indicate the hydro-geochemical variables.

The negative sign of discharge in this factor may indicate that the dilution processes of dissolved minerals increase with discharge (Varol and Sen, 2009). The other variables loading represent the contribution of organic pollution from domestic wastewater and non-point source pollution. The second factor (Factor 2) explained 26.10% of the total variance which has strong positive correlation with TP, strong negative correlation with DO and moderate positive correlation with Cd, Ni and Pb.

This factor can be interpreted as representing influences from industrial discharge and agriculture activities from agriculture area. A Group of industries are located upstream and downstream of the Drobeta-Turnu Severin city and agricultural practices from Serbian part that use fertilizers and pesticides which being discharged to the river through surface runoff (Gavrilescu, 2011). The third factor (Factor 3) is contributed by pH which explains 12.16% of the total variance. It can be observed that Factor 3 regulates the dissolution of cations and anions in which this factor reflects the acidity–alkalinity scale in river water (Ismail et al., 2015).

3.2. Cluster analysis (CA)

Classification of sampling site and monitoring periods was performed using of cluster analysis on z-standardized dataset. The output result of cluster analysis was shown as a dendrogram in which it provides a useful graphical tool determining the number of clusters which describe underlying process that lead to spatial and temporal variation (Boyacioglu and Boyacioglu, 2007).

Spatial similarity and site grouping

Spatial CA presented in a dendrogram (Fig. 3), where all the four sampling sites on the river were grouped into two statistically significant clusters. Group A consisted of SS1 and SS2 and group B consisted of SS3 and SS4. In Fig. 3 the distance between two samples corresponds to the similarity and dissimilarity between two samples, i.e. greater will be the distance, lesser will be the similarity. Group A (SS1 and SS2) corresponds to relatively less polluted sites. In group A, stations are situated upstream of Drobeta-Turnu Severin city. These stations receive pollution from nonpoint sources, i.e., mostly from agricultural activities. Group B (SS3 and SS4) corresponds to relatively moderate polluted sites in comparison with Group A.

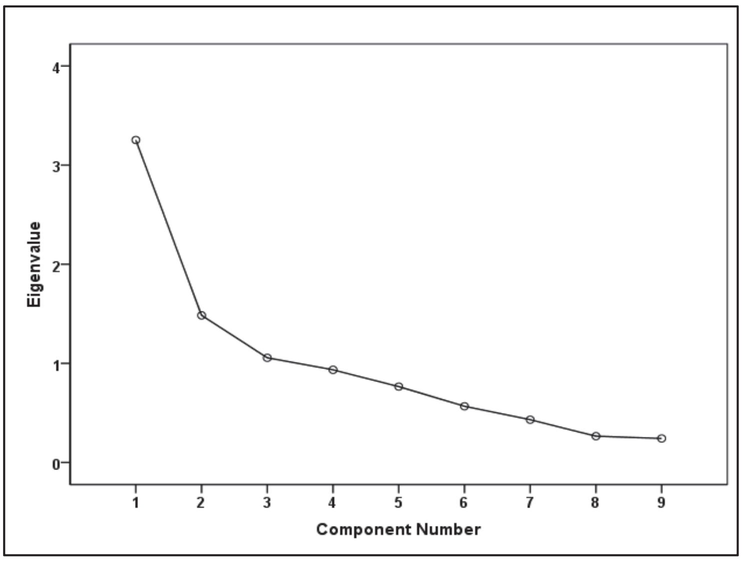

Fig. 2. Scree plot for the water quality dataset

Table 3. Rotated factor loading matrix and total variance explained (Varimax rotation)

Parameters	*Factor 1*	*Factor 2*	*Factor 3*
DO	0.088	-0.821**	-0.140
BOD	-0.657*	-0.173	0.019
pH	0.006	0.014	0.960**
Q	-0.572*	0.348	-0.103
TP	0.129	0.748**	-0.013
TSS	0.806**	0.063	-0.085
Cd	0.612*	0.573*	-0.166
Ni	0.251	0.575*	-0.326
Pb	0.693*	0.549*	0.011
Eigenvalue	2.350	2.349	1.094
% Total variance	26.11	26.10	12.16
Cumulative % variance	26.11	52.21	64.37

**Indicate moderate correlation with the factor loading; and **Indicate strong correlation with the factor loading*

In group B, stations are located on the industrial area of Drobeta-Turnu Severin city. These stations may receive pollution from point and nonpoint sources, i.e., domestic and industrial effluent, and surface runoff from the city.

Temporal similarity and period grouping

Temporal CA has been shown in Fig.4 as a dendrogram, grouping the 12 months into two major Groups (A and B). Group A included only two months (April and July). Group B comprised two significantly different subgroups (Group B1 and B2) which agglomerated remaining months. Group B1 included January, February, March, May, June, August, October, and December, whereas, Group B2 included September and November. However, if 12 months had been empirically divided into spring (March to May), summer (June to August), autumn (September to November), and winter (December to February), a mistake in grouping could have been made (Varol and Sen, 2009). The temporal variation in the Danube river water quality was not strictly determined by local climate (spring, summer, autumn, winter). Furthermore, it was observed that the discharge is the main factor influencing temporal variation among other parameters in the Danube River.

The discharge fluctuation in the region may affects the water quality in the river. Kurunc et al., (2005) studied the effect of seasonal discharge fluctuation on the water quality variables in Yeşilırmak River. They stated that there is a negative relationship between discharge and some water quality variables. Teodoru and Wehrli, (2005) has reported a time series flow rate at Drobeta–Turnu Severin station from 1960 to 1988 and the average outflow at the Iron Gate dams for the year 2001 (Fig. 5). Obviously, it can be noticed that the discharge values were roughly in the range 4000 – 8000 m^3/sec. Moreover, the discharge values in this study ranging from 2650-8760 m^3/sec with a mean value 5643 m^3/sec. Therefore, the water quality of the river in this region is highly affected by the discharge.

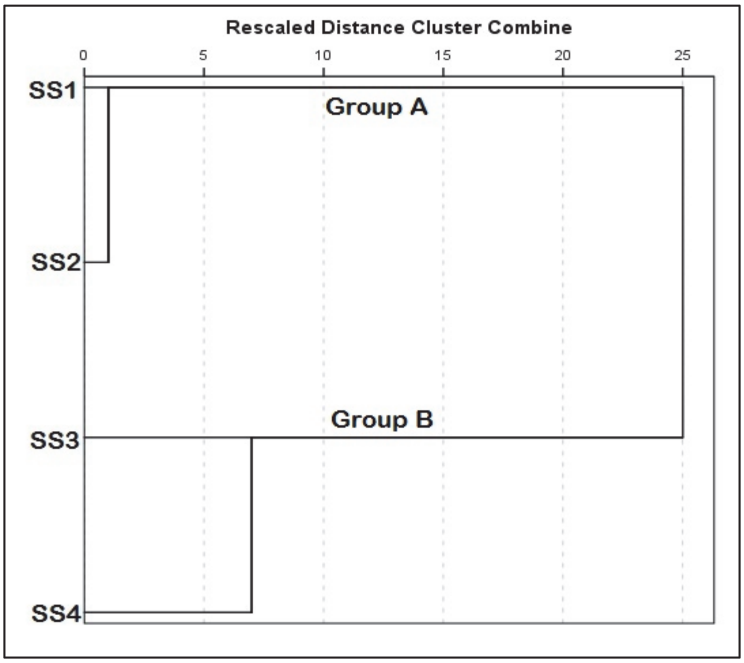

Fig. 3. Dendrogram showing spatial clustering of monitoring sites

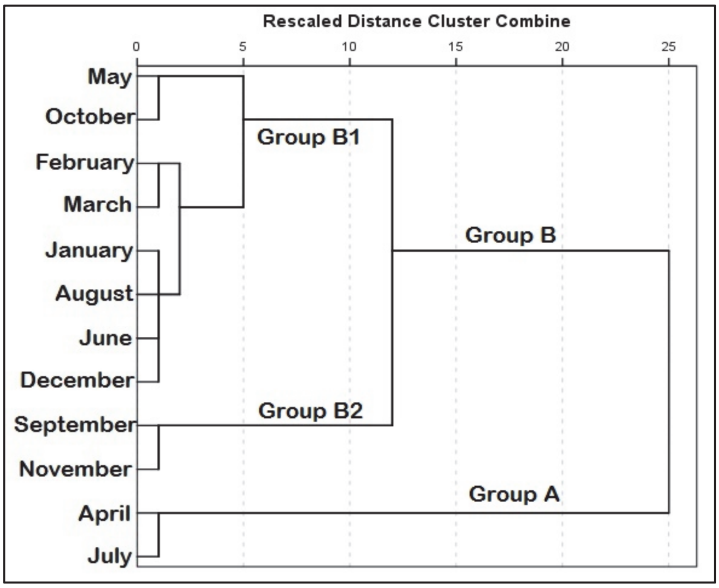

Fig. 4. Dendrogram showing spatial clustering of monitoring periods

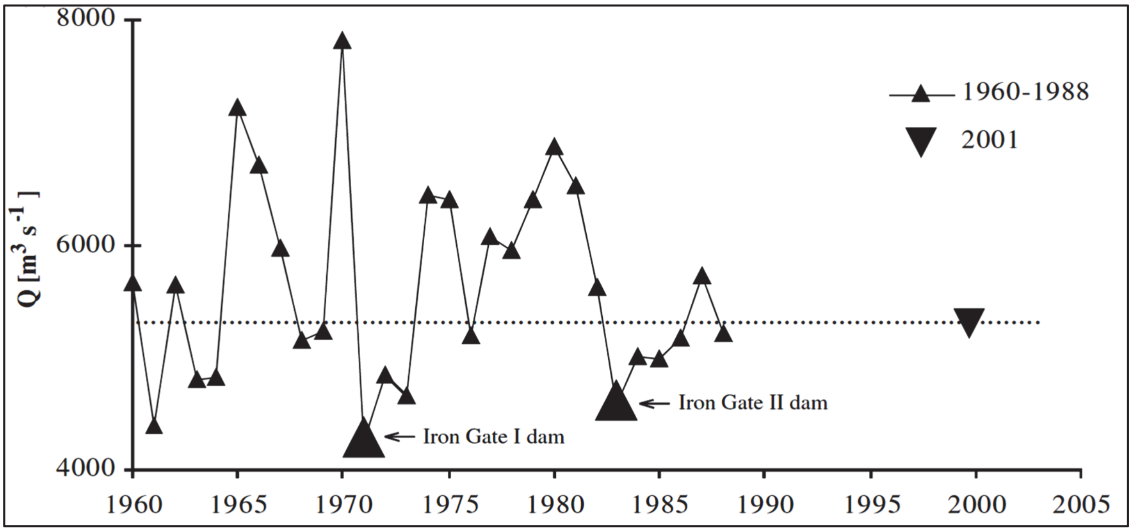

Fig. 5. Time series flow rate at Drobeta–Turnu Severin station from 1960 to 1988 and the average outflow at the Iron Gate dams for the year 2001 (adopted from Teodoru and Wehrli, 2005)

4. Conclusions

Multivariate statistical methods (FA and CA) were applied on a water quality dataset of Danube River in Drobeta-Turnu Severin for a period of 1 year (2008). The output results of factor analysis identified 3 factors responsible for the data structure, explaining 64.369% of total variance in the dataset. The first, second and third factors explained almost 26%, 26% and 12% of the total variance, respectively.

The main factors/sources influencing the quality of the river are the hydro-geochemical variables, organic pollution from domestic wastewater, industrial discharge and non-point source pollution such as agriculture activities from agriculture area and surface runoff.

Spatial CA grouped the four sampling sites on the river into two statistically significant clusters. First (SS1 and SS2) corresponds to relatively less polluted sites whereas, the second (SS3 and SS4) corresponds to relatively moderate polluted sites. It was concluded that the study area is highly affected by polution load from domestic and industrial discharges in compariosn with polution load form the agriculture area. Temporal CA grouped the 12 months into two major groups.

The results demonstrate that the temporal variation in the Danube river water quality was not strictly determined by local climate (spring, summer, autumn, winter). It was observed that the discharge (Q) is the main factor influencing temporal variation among other parameters in the Danube River.

Multivariate methods are believed to be an effective tool for proper management of water resources and providing a good explanation to understand complex dataset of surface water quality.

Acknowledgement

This paper is part of a PhD research of the first author and the authors would like to thank Professor Dan ROBESCU who comments and suggestions helped us improve the final version of this manuscript.

References

Altın A., Filiz Z., Iscen C.F., (2009), Assessment of seasonal variations of surface water quality characteristics for Porsuk Stream, *Environmental Monitoring and Assessment*, **158**, 51–65.

Andriţa V., (2012), *Water Quality in the Urban Ecosystem of Drobeta – Turnu Severin*, Int. Conf. on Water Resources and Wetlands, Gâştescu P., Lewis W., Breţcan P. (Eds.), Transversal Publishing House, 659-662.

Antanasijevic D., Pocajt V., Grujic A.P., Ristic M., (2014), Modelling of dissolved oxygen in the Danube River using artificial neural networks and Monte Carlo Simulation uncertainty analysis, *Journal of Hydrology*, **519**, 1895–1907.

APHA AWWA WEF, (1998), *Standard methods for examination of water and wastewater*, 20th Edition, American Public Health Association, Washington.

Bajo M., Ferrarin C., Dinu I., Umgiesser G., Stanica A., (2014), The water circulation near the Danube Delta and the Romanian coast modelled with finite elements, *Continental Shelf Research*, **78**, 62–74.

Boyacioglu H., Boyacioglu H., (2007), Surface water quality assessment by environmetric methods, *Environmental Monitoring and Assessment*, **131**, 371–376.

Chapagain S.K., Pandey V.P., Shrestha S., Nakamura T., Kazama F., (2010), Assessment of deep groundwater quality in Kathmandu Valley using Multivariate Statistical Techniques, *Water, Air, & Soil Pollution*, **210**, 277–288

Das M., Kumar A., Mohapatra M., Muduli S.D., (2010), Evaluation of drinking quality of groundwater through multivariate techniques in urban area, *Environmental Monitoring and Assessment*, **166**, 149–157.

Derx J., Blaschke A.P., Blöschl G., (2010), Three-dimensional flow patterns at the river–aquifer interface — a case study at the Danube, *Advances in Water Resources*, **33**, 1375–1387.

Gan T.Y., Zunic F., Kuo C., Strobl T., (2012), Flood mapping of Danube River at Romania using single and multi-date ERS2-SAR images, *International Journal of Applied Earth Observation and Geoinformation*, **18**, 69–81.

Gâştescu P., Ţuchiu E., (2012), The Danube River in the lower sector in two hydrological hypostases – high and low waters, *Revista Riscuri si Catastrofe*, **XI**, 165-182, On line at: http://riscurisicatastrofe.reviste.ubbcluj.ro/Volume/XI_Nr_10_1_2012/PDF/Gastescu.pdf

Garnier J., Billena G., Hannonb E., Fonbonnec S., Videninac Y., and Soulie M., (2002), Modelling the transfer and retention of nutrients in the drainage network of the Danube river, estuarine, *Coastal and Shelf Science*, **54**, 285–308.

Gavrilescu E., (2011), Study of heavy metals existing in the Danube waters in Turnu Severin – Bechet section, *South Western Journal of Horticulture, Biology and Environment*, **2**, 47-55.

Ismail A.H., Abed B.S., Abdul-Qader S., (2014), Application of multivariate statistical techniques in the surface water quality assessment of Tigris River at Baghdad stretch, *Journal of Babylon University*, **22**, 450- 462.

Ismail A.H., Muntasir A.H., Channo R.J., (2015), Groundwater quality assessment in urban area of Baghdad, Iraq, using multivariate statistical techniques, *Engineering and Technology Journal*, **33**, 463-476.

Janauer G.A., Mumm U., Schmidt B., (2010), Aquatic macrophytes and water current velocity in the Danube River, *Ecological Engineering*, **36**, 1138–1145.

Kirschner A.K.T., Kavka G.G., Velimirov, Mach R.l., Sommer R., Farnleitner A.H., (2009), Microbiological water quality along the Danube River: Integrating data from two whole-river surveys and a transnational monitoring network, *Water Research*, **43**, 3673 – 3684.

Kurunc A., Yurekli K., Ozturk F., (2005), Effect of discharge fluctuation on water quality variables from the Yeşilırmak River, *Tarim Bilimleri Dergisi-Journal of Agricultural Sciences*, **11**, 189-195.

Liu C.W., Lin K.H., Kuo Y.M., (2003), Application of factor analysis in the assessment of groundwater quality in a Blackfoot disease area in Taiwan, *The Science of the Total Environment*, **313**, 77–89.

Matenco L., Andriessen P., (2013), Quantifying the mass transfer from mountain ranges to deposition in sedimentary basins: Source to sink studies in the Danube Basin–Black Sea system, *Global and Planetary Change*, **103**, 1-18.

Muntean M.D, Morariu L., (2014), The influence of slag and ash deposit used by Drobeta -Turnu Severin power plant concerning groundwater in the area, *Journal of Young Scientist*, **2**, 91-98.

Palma P., Alvarenga P., Palma V.L., Fernandes R.M., Soares A.M.V.M., Barbosa I.R., (2010), Assessment of anthropogenic sources of water pollution using multivariate statistical techniques: a case study of the Alqueva's reservoir, Portugal, *Environmental Monitoring and Assessment*, **165**, 539-552.

Popa P., Murariu G., Timofti M., Georgescu L.P., (2018), Multivariate statistical analyses of Danube River water quality at Galati, Romania, *Environmental Engineering and Management Journal*, **17**, 1249-1266.

Popescu I., Cioaca E., Pan Q., Jonoski A., Hanganu J., (2015), Use of hydrodynamic models for the management of the Danube Delta wetlands: The case study of Sontea-Fortuna ecosystem, *Environmental Science & Policy*, **46**, 48-56.

Shrestha S., Kazama F., (2007), Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji River basin, Japan, *Environmental Modelling & Software*, **22**, 464-475.

Simeonov V., Simeonova P., Tsitouridou R., (2004), Chemometric quality assessment of surface water: Two case studies, *Chemical Engineering Ecology*, **11**, 449–469.

Singh K.P., Malik A., Mohan D., Sinha S., (2004), Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India): A case study, *Water Research*, **38**, 3980–3992.

Szolgayova E., Parajka J., Bloschl G., Bucher C., (2014), Long term variability of the Danube River flow and its relation to precipitation and air temperature, *Journal of Hydrology*, **519**, 871–880.

Teodoru C., Wehrli B., (2005), Retention of sediments and nutrients in the Iron Gate I Reservoir on the Danube River, *Biogeochemistry*, **76**, 539–565.

Vargha M., Takats Z., Marialigeti K., (2005), Degradation of atrazine in a laboratory scale model system with Danube River sediment, *Water Research*, **39**, 1560–1568.

Varol M., Sen B., (2009), Assessment of surface water quality using multivariate statistical techniques: a case study of Behrimaz Stream, Turkey, *Environmental Monitoring and Assessment*, **159**, 543–553.

Wunderlin D.A., Díaz M.P., Amé M.V., Pesce S.F., Hued A.C., Bistoni M.A., (2001), Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquía River Basin (Córdoba-Argentina), *Water Research*, **35**, 2881–2894.

Zhou F., Liu Y., Guo H., (2007), Application of multivariate statistical methods to water quality assessment of the watercourses in Northwestern New Territories, Hong Kong, *Environmental Monitoring and Assessment*, **132**, 1–13.