DE GRUYTER
OPEN

Application of multivariate statistical techniques in the water quality assessment of Danube river, Serbia

Danijela Voza[1][*], Milovan Vukovic[1], Ljiljana Takic[2], Djordje Nikolic[1],
Ivana Mladenovic-Ranisavljevic[2]

[1] University of Belgrade, Serbia
Technical Faculty in Bor
[2] University of Niš, Serbia
Faculty of Technology

[*] Corresponding author's e-mail: ddakic@tf.bor.ac.rs

Abstract: The aim of this article is to evaluate the quality of the Danube River in its course through Serbia as well as to demonstrate the possibilities for using three statistical methods: Principal Component Analysis (PCA), Factor Analysis (FA) and Cluster Analysis (CA) in the surface water quality management. Given that the Danube is an important trans-boundary river, thorough water quality monitoring by sampling at different distances during shorter and longer periods of time is not only ecological, but also a political issue. Monitoring was carried out at monthly intervals from January to December 2011, at 17 sampling sites. The obtained data set was treated by multivariate techniques in order, firstly, to identify the similarities and differences between sampling periods and locations, secondly, to recognize variables that affect the temporal and spatial water quality changes and thirdly, to present the anthropogenic impact on water quality parameters.

Introduction

Since the time immemorial, people have always settled their communities along the fertile river alluviums. In addition to being a life source for many plant and animal species, rivers have always provided opportunities for people to be engaged in activities that could ensure their survival and development. Also, rivers have important multi-usage components, such as sources of drinking water, the irrigation, fishery and energy production (Iscen et al. 2008). Almost every sector of economy, including agriculture, industry, transportation and public water supplies, depends on this water body.

Despite providing the conditions for numerous developmental activities of the modern world, the quality of water in rivers is now at a very low level. This is caused by easy river accessibility for the wastewater disposal. Because of the fact that drinking water could become a scarce resource in the future in most countries, the water quality evaluation has become a significant issue recently. Since rivers constitute the main inland water resources for domestic, industrial and irrigation purposes, it is our primary task to prevent and control river pollution as well as to ensure reliable information on water quality for effective management (Singh et al. 2004).

The European Union significantly contributes to the protection and preservation of water resources and environment, considering them as the base of sustainable development in the 21st century. Hence, the EU has decided not to leave such an important issue to independent decision making by individual countries within the Union, but to make a unique and coherent strategy for the environment protection and water management. In order to solve this problem, in 2000 the EU established the Water Framework Directive, which requires mutual action of the Member States regarding the protection, restoration and improvement of water bodies (EU, 2000). However, significant and positive issues in this regard have not been achieved yet. Huge amounts of industrial, domestic sewage and agricultural waste still end up in rivers and lakes preventing thus the balance of these ecosystems.

The surface water quality is influenced by both natural (precipitation rate, weathering processes and soil erosion) and anthropogenic (urban, industrial, agricultural activities and increasing exploitation of water resources) factors (Carpenter et al. 1998; Jarvie et al. 1998, Singh et al. 2004). Seasonal variations in precipitation, surface run-off, ground water flow, interception and abstraction strongly affect the river discharge and the concentrations of river water pollutants (Vega et al. 1998, Pejman et al. 2009). Water quality is determined by its physical, chemical and biological features.

Water monitoring management, during a long-term period and at many sampling sites, offers large and complicated data sets consisting of all kinds of water parameters, which are not only difficult to analyze and interpret but also to extract comprehensive information from (Zhang et al. 2011). The problem of (multi)collinearity between the explanatory variables has not yet been fully resolved in various studies dealing with river water quality. Ignoring this potentially

confounding phenomenon may lead to models that assume misleading inferences about the effects of various parameters on the river water quality (Varanka et al. 2012). Collinearity among explanatory variables can cause the situation where the variable that explains the variation in the response variable better in statistical terms, remains in the model, whereas the variable that is more causal in reality is excluded from the regression model (Mac Nelly et al. 2004). To solve the problems caused by collinearity, researchers have conventionally used two statistical techniques: excluding some of the most inter-correlated variables or summarizing the variation of several explanatory variables into composite variables using, for example, principal components analysis (Quinn et al. 2002).

The application of different multivariate approaches (cluster analysis, principal component analysis, factor analysis and discriminant analysis) helps in the interpretation of complex data matrices, provides a better indication of water quality and ecological status of the studied systems, allows for the identification of possible factors/sources that affect water bodies, provides a valuable tool for reliable management of water resources, and finally offers a rapid solution to the pollution problem (Simeonov et al. 2003, Simeonova et al. 2003, Singh et al. 2004, Iscen et al. 2008, Han et al. 2009, Varol et al. 2012, Gudas et al. 2013, Boyacioglu 2014).

The aim of this study is to observe Danube River water quality in its course through Serbia. Danube, as European Unions' first and continents' the second longest river, plays an important trans-boundary role. Danube River is 2,857 km long and has a basin of 817,000 km^2 (with 10 riparians); it connects western and eastern Europe. Having in mind that Danube River is one of the most important natural ecosystems in Europe, special attention must be paid to its evaluation, pollution, and protection problems in each riparian-country. Given these assumptions, the careful monitoring of Danube water quality in long-term or short-term time periods, at different sampling sites, is not only an ecological but also a political issue. Not surprisingly, Danube River is constantly a subject of various studies. In recent years, the focus is on the water pollution of Danube River Basin (Guien et al. 1998, Pawellek et al. 2002, Woitke et al. 2003, Brankov et al. 2012), water quality temporal and spatial trends (Onderka et al. 2008) as well as predicting values of water quality parameters (Takic et al. 2012).

Although, in many countries, the regular monitoring systems produce a significant amount of analytical data available to various users, there is still a lack of summarizing studies that consider all aspects of the river system and take into account all possible information from the row data sets. According to Simenonova et al. (2003), the application of multivariate statistical methods (chemometrics and environmentrics) seems to be the only approach to gain the proper knowledge regarding water quality. In the case of Danube River flow through Serbia, it is of special importance to detect trends in the concentrations of the main chemical parameters determined in measuring sites. Equally important it is to discover the seasonal behaviour of the components as well as to identify possible sources of river pollution. Namely, long-term changes of parameters, indicators of Danube water quality, differ in time and space.

Therefore, in this study, the extensive data sets were processed by multivariate techniques in order to identify similarities and differences between monitoring periods and locations then, to determine variables that have an effect on the temporal and spatial changes in water quality and to present the impact of natural and anthropogenic factors on water quality of Danube River in its course through Serbia.

Materials and methods

Monitoring area

Danube River flows through Serbia in the length of 588 km, from Bezdan to Prahovo. It stretches south-easterly, from the border between Serbia, Hungary and Croatia to the Timok confluence, on the border between Serbia, Bulgaria and Romania. The largest tributaries of Danube in Serbia are Tisa, Tamiš and Danube-Tisa-Danube Canal on the left, and Drava, Sava, Velika Morava and Timok on the right. Also, there are two hydroelectric power stations on its banks: Iron Gate I and Iron Gate II.

The water quality of Danube River in Serbia could be mostly described as moderately polluted, while some tributaries and its lower course do not reach this pollution level. In the structure of Danube registered polluters, industry is dominant, followed by agriculture, settlements, energetic sector and others (Milanovic et al. 2010). Two large cities are located on the banks of Danube in Serbia: Belgrade (which with 1.7 million inhabitants is the third largest city on this river) and Novi Sad (300,000 inhabitants), and a host of smaller towns (Apatin, Bačka Palanka, Pančevo Smederevo, Donji Milanovac and Kladovo) and villages. In addition, none of these settlements have an adequate system for wastewater treatment.

Stations and sampling period

The starting point for monitoring the quality of River Danube water in Serbia, is the Data Fund of the RHSS for the year 2011. Monitoring was carried out monthly, from January to December 2011, at 17 measuring points. The locations of the sampling sites obtained with the data are shown in Fig. 1. The research of Danube river water quality status is covered by seventeen hydrological monitoring stations at a certain distance from the confluence and then along the river:

1. Bezdan – the input profile (1,425.59 km),
2. Apatin (1,401 km),
3. Bogojevo (1,367.4 km),
4. Bačka Palanka (1,298.6 km),
5. Novi Sad (1,254.98),
6. Slankamen (1,215.5 km),
7. Čenta (1,189 km),
8. Zemun (1,174 km),
9. Pančevo (1,154.6 km),
10. Belgrade-Vinča (1,145.5 km),
11. Smederevo (1,116 km),
12. Banatska Palanka (1,076.6 km),
13. Veliko Gradište (1,059.2 km),
14. Dobra (1,021 km),
15. Tekija (956.2 km),
16. Brza Palanka (883.8 km) and
17. Radujevac – the output profile (852 km).

Obviously, seventeen monitoring stations cover almost completely Danube River flow through Serbia.
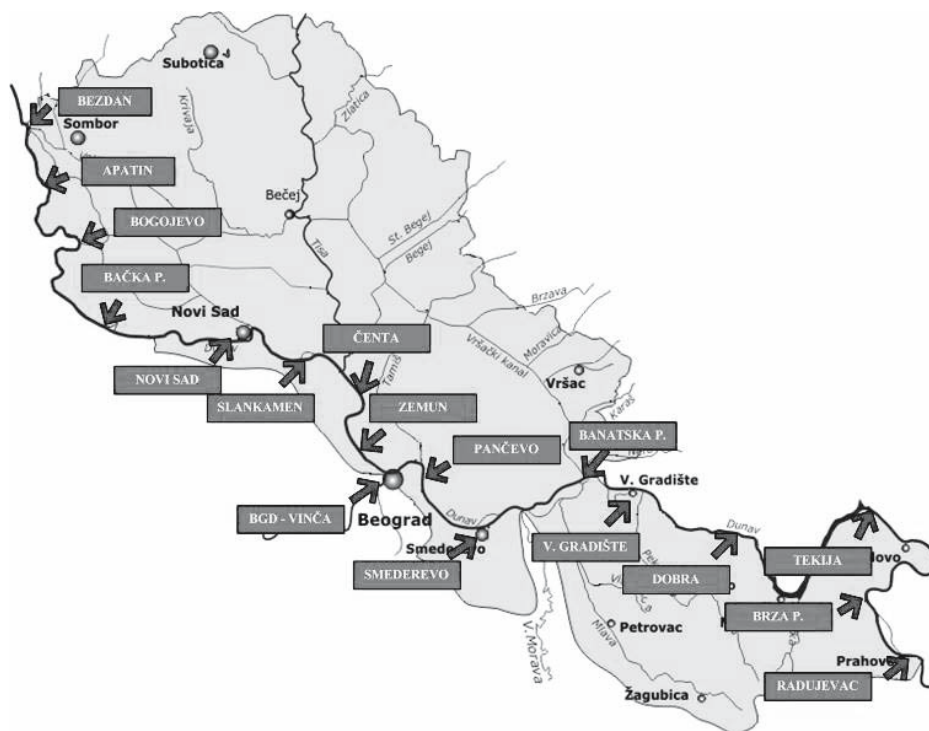
Fig. 1. Sampling stations on the Danube in Serbia

Monitored parameters and analytical methods

The data for 17 water quality monitoring stations consisted of 11 water quality parameters. The selected parameters included: water temperature (T), pH value, suspended solids (SM), dissolved oxygen (RK), ammonium (NH_4), nitrate (NO_3), orthophosphate (PO_4), silicate (SiO_2), sulphur dioxide (SO_4), chloride (Cl^-) and biochemical oxygen demand (BPK_5). As it can be noticed, the chemical indicators were selected for the assessment of water quality. All water quality parameters were expressed in milligram per liter (mg 1^{-1}), except T (°C) and pH.

Among the above indicators, three chemical parameters represent the so-called mobile water-soluble forms of nitrogen and phosphorous – nitrate (NO_3), ammonium (NH_4) and orthophosphate (PO_4). These chemical compounds are of crucial environmental importance because of their role in formation of nutrition loads in rivers – when nutrient loads are too high, eutrophication occurs, causing problems such as water quality deterioration, changes in the biotic population structure, oxygen depletion etc. (Ani et al. 2011).

The methods chosen to conduct chemical analyses included standard analytical techniques as routinely applied in the control laboratories of the monitoring network. Potentiometry, gravimetry, titrimetry, and spectrophotometry are standard analytical techniques widely used in the surface water quality assessment.

At the sampling point, the water temperature was measured and pH value determined according to SRPS H.Z1.111 method; the biochemical oxygen consumption (BPK_5) was determined by EPA 360.2 method, suspended solids according to 13.060.30 SRPS H.Z1.160 method, phosphates according to the standard analytical method APHAS AWWA WEF 4500, while the total nitrogen oxides were determined according to SRPS ISO 5664.

Statistical analysis

The possibility of multivariate statistical techniques application in water quality analysis has been proved repeatedly (Wunderlin et al. 2001, Reghunath et al. 2002, Singh et al. 2004, Milanovic et al. 2010, Varol et al. 2012). In this study, the multivariate data analysis was performed using cluster analysis, principal component analysis and factor analysis. They were applied to experimental data, standardized through z – scale transformation in order to avoid misclassifications arising from different orders of magnitude of both numerical values and variance of the parameters analyzed (Liu et al. 2003, Simeonov et al. 2003, Shrestha et al. 2007). All the statistical computations were made using the software package SPSS V. 17.0.

Cluster Analysis (CA), the unsupervised pattern recognition technique, reveals the intrinsic structure of a data set without making a-priori assumptions about the data to classify the objects of the system into categories or clusters based on their nearness or similarity (Vega et al. 1998). Hierarchical clustering is the most common approach, where clusters are formed sequentially starting by the most similar pair of objects and forming higher clusters in a step-by-step pattern.

Principal Component Analysis (PCA) is a powerful technique used for recognition of patterns which can explain the variance of large data sets of inter-correlated variables and transform them into smaller sets of independentindepended variables (principal components) (Simeonov et al. 2003). Principal components provide information on the most significant parameters, which describe the whole data sets through the data reduction with a minimum loss of original information (Vega et al. 1998, Helena et al. 2000). In short, the PCs are weighted linear combinations of the original variables.

Factor Analysis (FA) follows PCA. FA further reduces the contribution of less significant variables obtained from PCA, and a new group of variables, known as varifactors (VFs), is extracted by rotating the axis defined by PCA. VF can include unobservable, hypothetical, latent variables.

Results and discussion

The basic statistics of the one-year data set on the river water quality are summarized in Table 1.

Spatial similarities

In order to define the similarity groups between sampling stations, the cluster analysis was applied. A dendrogram (Figure 2) was obtained as a result, according to which 17 monitoring stations could be grouped into three statistically significant clusters in a distant connection $(D_{link} / D_{max}) * 100$.

Hierarchical clustering was performed according to Ward's method, with a squared Eucledian distance (Singh et al. 2004, Pejman et al. 2009, Varol et al. 2012). The Eucledian distance usually gives the similarity between two samples

Table 1. Results of water quality parameters from 17 sampling stations
(T – temperature; SM – suspended solids; RK – dissolved oxygen; BPK5 – biological oxygen demand)

Sampling station		T	SM	RK	pH	NH_4	NO_3	PO_4	SiO_2	Cl^-	SO_4	BPK_5
Bezdan	Mean	12.69	15.00	11.37	8.25	0.047	1.557	0.041	6.783	24.742	41.42	2.508
	Std.dev	7.977	4.632	2.278	0.265	0.042	0.463	0.024	2.53	6.327	8.426	1.533
	Max	22.6	22	14.60	8.80	0.130	2.37	0.070	11.00	34.80	55	5.80
	Min	1.5	7	8.40	8.00	0.010	1.01	0.002	0.60	15.70	28	1.00
Apatin	Mean	13.26	15.55	11.19	8.26	0.049	1.702	0.046	6.8273	24.663	41.818	2.491
	Std.dev	8.824	5.973	2.466	0.157	0.039	0.562	0.025	2.261	6.485	11.188	1.617
	Max	23.8	25	15.40	8.60	0.130	2.73	0.078	10.40	38.00	61	5.80
	Min	1.4	5	8.30	8.10	0.010	0.83	0.003	1.30	16.20	24	1.20
Bogojevo	Mean	12.96	21.4	10.81	8.18	0.052	1.606	0.049	7.54	23.57	38.5	2.10
	Std.dev	9.30	13.41	2.42	0.181	0.034	0.589	0.027	1.523	5.119	11.058	1.365
	Max	24.00	52	15.40	8.50	0.120	2.61	0.092	10.70	31.90	61	5.50
	Min	2.4	7	7.70	7.80	0.010	0.80	0.003	5.80	18.30	25	1.0
Bačka P.	Mean	13.33	25.22	10.11	8.133	0.054	1.584	0.050	8.033	21.933	38.000	1.722
	Std.dev	9.300	19.273	1.952	0.122	0.041	0.501	0.015	1.580	5.487	7.566	0.734
	Max	23.9	59	12.60	8.30	0.120	2.40	0.068	11.10	33.00	48	3.50
	Min	2.3	5	7.2	8.00	0.020	1.02	0.020	6.30	16.30	26	1.10
Novi Sad	Mean	13.69	16.75	9.83	8.03	0.097	1.528	0.046	5.883	22.300	39.917	2.527
	Std.dev	8.596	7.629	1.958	0.206	0.041	0.519	0.022	3.11	5.30	8.273	0.674
	Max	23.9	28	12.80	8.30	0.190	2.30	0.078	11.00	32.10	52	3.60
	Min	1.5	5	7.30	7.50	0.050	0.66	0.012	1.00	14.9	24	1.50
Slankamen	Mean	13.44	13.800	10.09	8.14	0.089	1.494	0.042	6.033	21.743	36.667	2.980
	Std.dev	9.188	9.052	2.078	0.117	0.039	0.453	0.018	1.250	6.559	9.647	3.19
	Max	24.2	37	12.90	8.30	0.150	2.31	0.065	6.90	32.20	49.00	11.80
	Min	1.5	5	7.80	8.00	0.050	0.82	0.011	4.60	15.00	27.00	0.90
Čenta	Mean	13.71	19.556	10.28	8.18	0.068	1.485	0.044	6.467	21.957	36.333	2.000
	Std.dev	9.267	8.172	2.157	0.132	0.043	0.430	0.021	2.25	6.175	7.146	0.729
	Max	24.1	34	13.10	8.40	0.150	2.37	0.068	8.70	31.70	48	3.10
	Min	1.6	10	7.70	8.00	0.010	0.85	0.011	4.20	14.90	29	0.90
Zemun	Mean	15.19	24.181	9.255	8.08	0.152	0.782	0.057	5.420	22.373	37.727	1.746
	Std.dev	7.889	35.858	2.530	0.133	0.074	0.402	0.056	1.589	7.421	8.200	0.710
	Max	25.0	130	12.7	8.30	0.260	1.70	0.218	7.90	36.00	26	3.50
	Min	5.9	2	6.30	7.80	0.050	0.40	0.002	3.20	14.00	50	1.00
Pančevo	Mean	13.00	29.80	9.78	8.11	0.080	1.232	0.051	8.344	24.750	41.700	1.778
	St.dev	9.676	25.223	1.914	0.160	0.043	0.294	0.022	2.956	4.855	6.075	0.438
	Max	24.6	76	12.40	8.30	0.150	1.71	0.082	13.50	36.00	53	2.60
	Min	1.3	5	7.20	7.80	0.020	0.76	0.013	4.60	20.30	33	1.20
Bgd.Vinča	Mean	15.81	16.55	9.85	8.19	0.128	0.736	0.037	4.740	20.882	34.546	1.755
	Std.dev	8.216	22.065	2.301	0.178	0.063	0.266	0.014	0.944	5.141	6.023	0.789

Table 1. Results of water quality parameters from 17 sampling stations
(T – temperature; SM – suspended solids; RK – dissolved oxygen; BPK5 – biological oxygen demand) – cont.

| Sampling station		T	SM	RK	pH	NH_4	NO_3	PO_4	SiO_2	Cl^-	SO_4	BPK_5
	Max	25.7	80	12.9	8.5	0.220	1.40	0.052	5.80	33.00	44	3.60
	Min	6.9	2	7.30	7.9	0.040	0.40	0.004	2.80	15.00	26	1.00
Smederevo	Mean	15.72	16.64	9.24	8.07	0.257	0.682	0.041	4.540	21.782	33.636	1.946
	Std.dev.	8.418	25.378	2.426	0.127	0.396	0.352	0.016	1.094	6.475	5.608	0.848
	Max	26.0	89	12.9	8.4	1.440	1.60	0.060	6.20	33.00	41	3.90
	Min	6.2	1	6.8	7.9	0.050	0.30	0.007	2.90	16.00	23	1.00
Banatska P	Mean	14.77	20.90	9.27	7.88	0.141	1.270	0.046	6.882	24.436	38.818	1.782
	Std.dev	8.905	11.318	2.082	0.147	0.043	0.321	0.015	2.341	4.909	10.722	0.586
	Max	26.2	40	12.70	8.20	0.220	1.72	0.062	11.00	34.70	58	2.90
	Min	2.2	5	6.60	7.70	0.080	0.82	0.019	4.00	18.30	23	1.00
V.Gradište	Mean	15.17	6.33	9.26	8.09	0.314	0.617	0.050	5.858	19.833	35.333	2.150
	Std.dev	8.446	5.015	2.323	0.239	0.668	0.374	0.028	1.972	4.783	8.500	0.632
	Max	28.2	18	13.50	8.60	2.400	1.30	0.116	9.40	29.00	45	3.10
	Min	3.8	2	6.40	7.70	0.010	0.10	0.003	2.00	10.00	23	1.30
Dobra	Mean	14.93	6.42	10.13	8.15	0.100	0.567	0.053	5.444	19.833	36.727	1.925
	Std.dev	8.541	5.534	2.188	0.224	0.068	0.303	0.028	2.688	3.786	12.823	0.422
	Max	27.1	22	14.4	8.70	0.250	1.10	0.119	11.90	25.00	69	2.60
	Min	3.6	2	7.3	7.90	0.010	0.20	0.014	2.00	13.00	26	1.10
Tekija	Mean	15.28	4.89	9.78	8.22	0.112	0.600	0.051	5.489	22.333	36.917	1.842
	Std.dev	8.785	5.422	2.591	0.233	0.082	0.316	0.025	3.253	3.985	10.501	0.500
	Max	28.1	21	15.30	8.70	0.300	1.00	0.102	13.50	31.00	53	2.80
	Min	2.6	0.70	5.70	8.00	0.010	0.10	0.005	1.60	15.00	18	1.00
Brza P.	Mean	15.04	4.23	9.40	8.13	0.081	0.625	0.041	5.033	21.000	34.167	1.908
	Std.dev	8.934	5.238	2.208	0.230	0.060	0.644	0.027	1.427	3.075	10.071	0.545
	Max	27.0	20	13.10	8.60	0.200	1.30	0.093	8.30	25.00	48	2.60
	Min	2.1	0.50	6.90	7.80	0.010	0.10	0.003	3.60	15.00	13	1.10
Radujevac	Mean	14.29	8.13	9.05	7.79	0.095	0.625	0.137	5.717	20.500	36.833	1.958
	Std.dev	9.073	7.404	2.413	0.156	0.054	0.387	0.112	1.389	4.275	8.569	0.523
	Max	28.0	28	12.40	8.10	0.200	1.40	0.356	8.40	30.00	49	2.80
	Min	2.3	0.50	6.30	7.50	0.030	0.10	0.034	3.60	15.00	23	1.20
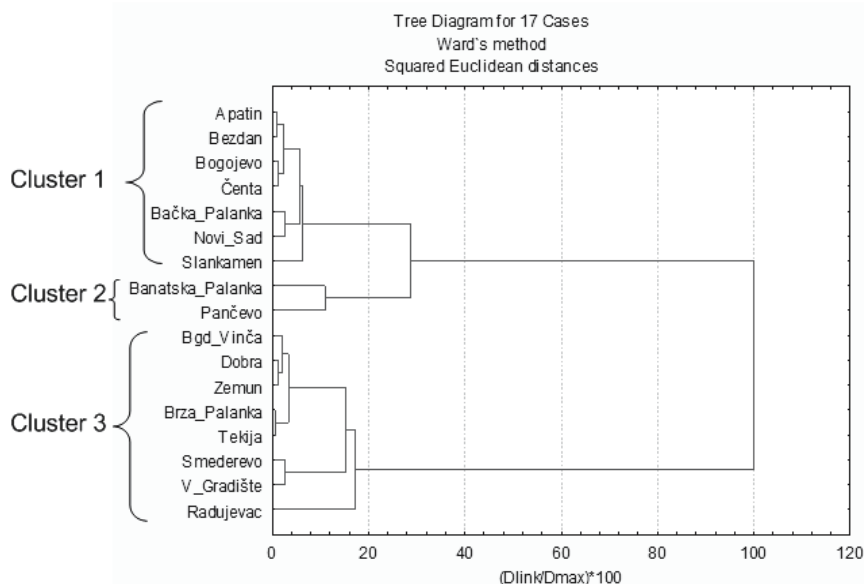
Fig. 2. Dendrogram showing hierarchical clustering of sampling stations on Danube River in Serbia

and the distance can be represented by the difference between analytical values from the samples (Otto, 1998).

Cluster 1 consists of seven sampling stations on Danube in Serbia (Apatin, Bezdan, Bogojevo, Čenta, Bačka Palanka, Novi Sad and Slankamen), cluster 2 includes two sampling stations (Banatska Palanka and Pančevo), while the third cluster consists of eight sampling stations (Belgrade--Vinča, Dobra, Zemun, Brza Palanka, Tekija, Smederevo, V. Gradište and Radujevac). The locations within the same cluster have similar characteristics and pollution source. Therefore, for a rapid water quality assessment, only one site from each cluster is enough, and it can serve as a good indicator of water quality of the whole group. On this basis, it can be concluded that the sampling stations in cluster 1 are located in areas of high pollution.

Furthermore, cluster 2 and all corresponding sampling stations indicate moderate pollution areas. Finally, cluster 3 could be classified as the low polluted region of Danube in Serbia. This result of spatial grouping, partly (cluster 1 and cluster 2) matches a division of Serbian Danube into the upper and lower section. The upper section covers the stretch from the Hungarian border to Belgrade, while the lower section, which is strongly influenced by the Iron Gates I and II dam complex, covers the stretch from Belgrade to the Bulgarian border.

Temporal similarities

As a result of the hierarchical CA applied on the data sorted monthly, the dendrogram (Figure 3), which indicates the existence of two clusters or periods, was obtained. Period 1 includes January, February, March, April, November and December. Other months (May, June, July, August, September and October) belong to the second period. Based on the results, it can be concluded that this temporal pattern of water quality does not match the traditional classifications to four seasons (spring, summer, autumn, winter) or the dry and wet seasons.

Principal component analysis/factor analysis

PCA/FA was applied to the normalized data in order to make a comparison between patterns of the water samples chemical structure as well as to identify the factors that influence each other (Singh et al. 2004). PCA of the whole data set shows

that four principal components account for 68.95% of the total water quality data set variance (Table 2).

The first principal component (PC1), which accounts for 33.1% of the total variance, has strong positive loadings (>0.75) on dissolved oxygen (RK), Cl^- i SO_4, but strong negative loading on water temperature (T). There is evident moderate loading on NO_3 and weak positive loading on SiO_2. This group can be interpreted as a mineral component of the river water (Singh et al. 2004). This clustering points to a common origin for these variables, similar to the dissolution of limestone, marl and gypsum soils (Vega et al. 1998).

In the second PC (PC2), which accounts for 16.04% of the total variance, it can be noticed that there are moderate positive loadings on pH value and BPK_5 but moderate negative loadings on orthophosphates. It can be concluded that PC2 is connected with anthropogenic pollution sources (Vega et al. 1998) caused by the use of fertilizers in agricultural activities.

PC3 (10.43% of the total variance) has strong positive loadings on suspended solids, moderate positive ones on SiO_2 and weak positive loadings on NO_3. This factor, loaded by solids, indicates that their origin lies in run-off from the fields with a high load of solids and waste disposal activities (Singh et al. 2004).

Finally, the principal component PC4, which participates with 9.4% of the total variance, is characterized by a strong negative loading of ammonium and weak positive loading on nitrite, which also leads to the pollution of the source of anthropogenic origin.

The results, obtained by PCA/FA, indicate that the changes in water quality mainly occur under the influence of soluble salts (natural) and organic pollutants (anthropogenic) and nutrients. Given that 11 out of 17 originally selected variables are still needed to account for 69% of the total variance of the data, it can be concluded that the factor analysis did not cause a significant data reduction.

Conclusions

This study proves the utility of multivariate statistical methods, cluster analysis (CA), principal component analysis (PCA) and
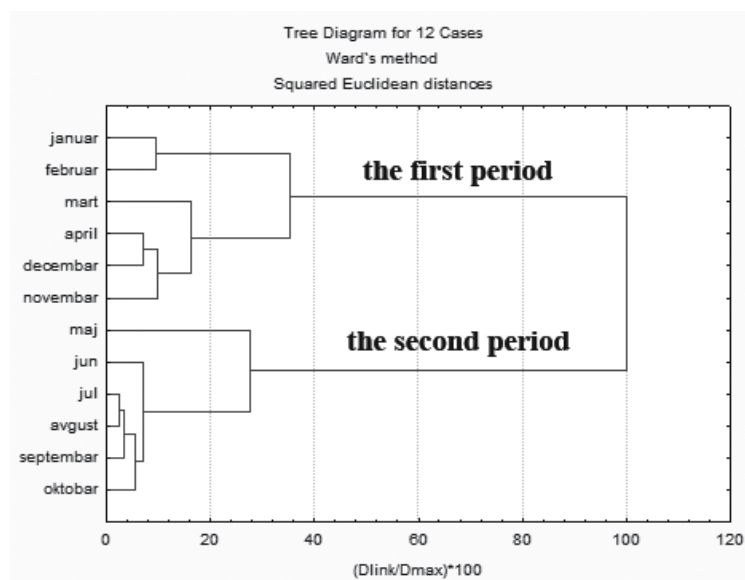

Fig. 3. Dendrogram showing hierarchical clustering of monitoring period

Table 2. Loadings of experimental variables (11) on significant principal components
(with Varimax rotation) for the Danube River data set

	Component			
	1	2	3	4
T	**-.901**	.093	-.139	.161
SM	-.033	.142	**.836**	-.090
RK	**.870**	.291	.100	-.058
pH	.188	**.725**	-.200	.206
NH_4N	.117	-.110	.056	**-.893**
NO_3N	**.527**	-.021	**.432**	**.407**
PO_4P	-.158	**-.736**	-.111	.199
SIO_2	**.304**	**-.366**	**.610**	.049
Cl	**.840**	.083	.026	.029
SO_4	**.851**	-.059	.008	-.011
BPK_5	-.126	**.585**	.021	.126

factor analysis (FA) above all, for the analysis and interpretation of the complex water quality data sets. With these techniques, it is possible not only to identify and distribute pollution sources, but also to gain a better insight into the temporal and spatial changes in surface water quality.

The application of the above mentioned methods on the water quality data of Danube through Serbia, led to the following conclusions:

1. the spatial cluster analysis grouped 17 sampling stations on Danube in three clusters which could be divided into groups with low, moderate and high level of pollution;
2. with the temporal cluster analysis, 12 months of monitoring were divided in two periods;
3. PCA/FA allocated for the most important factors responsible for water quality changes, due to which, the following are recognized as main pollution sources of Danube River in Serbia: mineral salts (natural origin), organic pollution and nutrients (anthropogenic origin).

Defining the spatial and temporal similarities of the water quality of Danube River in Serbia, provides a more efficient organization of sampling station network and data collecting, which is a prerequisite for creating an optimal sampling strategy and reducing associated costs. The application of multivariate statistical methods to the monitoring data provides a collection of new types of information about sampling and the chemical structure of the water body.

Acknowledgments

References

Ani, E.C., Hutchins, M., Kraslawski, A. & Agachi, P.S. (2011). Mathematical model to identify nitrogen variability in large rivers, *River Research and Applications,* 27, pp. 1216–1236.

Brankov, J., Milijašević, D. & Milanović, A. (2012). The assessment of the surface water quality using the Water Pollution Index: a case study of the Timok River (the Danube River Basin), Serbia, *Archives of Environmental Protection*, 38, 1, pp. 49–61.

Boyacioglu, H. (2014) Spatial differentiation of water quality between reservoirs under anthropogenic and natural factors based on statistical approach, *Archives of Environmental Protection*, 40 (1), pp. 41–50.

Carpenter, S.R., Caraco, N.F., Corell, D.L., Howarth, R.W., Sharpley, A.N. & Smith, V.H. (1998). Nonpoint pollution of surface waters with phosphorous and nitrogen, *Ecological Applications,* 8, pp. 559–569.

EU. *Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for community action in the field of water policy: 2000,* (http://ec.europa.eu/environment/water/water-framework/index_en.html (05.03.2013)).

Gudas, M. & Povilaitis, A. (2013). Factors affecting seasonal and spatial patterns of water quality in Lithuanian rivers, *Journal of Environmental Engineering and Landscape Management,* 21, pp. 26–35.

Guien, C., Martin, J. M., Tankéré, S.P.C., Mousty, F., Trincherini, P., Bazot, M. & Dai, M.H. (1998). On trace metal geochemistry in the Danube River and Western Black Sea, *Estaurine Coastal and Shelf Science,* 47, pp. 471–485.

Han, S., Kim, E. & Kim, S. (2009). The water quality management in the Nakdong River watershed using multivariate statistical techniques, *KSCE Journal of Civil Engineering,* 13, pp. 97–105.

Helena, B., Pardo, R., Vega, M., Barrado, E., Fernandez, J.M. & Fernandez, L. (2000). Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis, *Water Research,* 34, pp. 807–816.

Iscen, F.C., Emiroglu, Ö., Ilhan, S., Arslan, N., Yilmaz, V. & Ahiska, S. (2008). Application of multivariate statistical techniques in the assessment of surface water quality in Ulabat Lake, Turkey, *Environment Monitoring and Assessment,* 144, pp. 269–276.

Jarvie, H.P., Whitton, B.A. & Neal, C. (1998). Nitrogen and phosphorous in east-coast British rivers: speciation, sources and biological significance, *Science of the Total Environment,* 210/211, pp. 79–109.

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

Mac Nelly, R. & Walsh, C.J. (2004). Hierarchical partitioning public--domain software, *Biodiversity and Conservation,* 13, pp. 659–660.

Milanović, A., Kovačević-Majkić, J. & Milivojević, M. (2010). Water quality analysis of Danube river in Serbia: Pollution and protection problems, *Glasnik srpskog geografskog društva,* 90, pp. 47–68. (in Serbian)

Onderka, M. & Pekarova, P. (2008). Retrieval of the suspended particulate matter concentrations in the Danube River from Landsat ETM data, *Science of the Total Environment,* 397, pp. 238–243.

Otto, M. Multivariate methods, in: *Analytical Chemistry,* Kellner R., et al. (Eds.).Wiley-VCH, Weinheim 1998.

Pawellek, F., Fraunstein, F. & Veizer, J. (2002). Hydrochemistry and isotope geochemistry of the upper Danube River, *Geochimica et Cosmoshimica Acta,* 66, pp. 3839–3853.

Pejman, A.H., Nabi Bidhendi, G.R., Karbassi, A.R., Mehrdadi, N. & Esmaeili Bidhendi, M. (2009). Evaluation of spatial and seasonal variations in surface water quality using multivariate statistical techniques, *International Journal of Environmental Science and Technology,* 6, 3, pp. 467–476.

Quinn, G.P. & Keough, M.J. (2002). *Experimental Design and Data Analysis for Biologists,* Cambridge: Cambridge University Press 2002.

Reghunath, R., Murthy, T.R. & Raghavan, B.R. (2002). The utility of multivariate statistical techniques in hydrogeochemical studies: an example from Karnataka, India, *Water Research,* 36, pp. 2437–2442.

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 and Software,* 22, pp. 464–475.

Simeonov, V., Stratis, J.A., Samara, C., Zachariadis, G., Voutsa, D., Anthemidis, A., Sofoniou, M. & Kouimtzis, T. (2003). Assessment of the surface water quality in Northern Greece, *Water Research,* 37, pp. 4119–4124.

Simeonova, P., Simeonov, V. & Andreev, G. (2003). Water quality study of the Struma River Basin, Bulgaria, *Central European Journal of Chemistry,* 2, pp. 121–136.

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, pp. 3980–3992.

Takić, Lj., Mladenović–Ranisavljević, I., Vuković, M. & Mladenović, I. (2012). Evaluation of the ecochemical status of the Danube in Serbia in terms of water quality parameters, *The Scientific World Journal,* doi: 10.1100/2012/930737

Varol, M., Gökot, B., Bekleyen, A. & Şen, B. (2012). Water quality assessment and apportionment of pollution sources of Tigris river (Turkey) using multivariate statistical techniques – a case study, *River Research and Applications,* 28, pp. 1428–1438.

Varanka, S. & Luoto, M. (2012). Environmental determinants of water quality in boreal rivers based on partitioning methods, *River Research and Applications,* 28, pp. 1034–1046.

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

Woitke, P., Wellmitz, J., Helm, D., Kube, P., Lepom, P. & Litheraty, P. (2003). Analysis and assessment of heavy metal pollution in suspended solids and sediments of the River Danube, *Chemosphere,* 51, pp. 633–642.

Wunderlin, D.A., Diaz, M.P., Ame, 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: Suquia river basin (Cordoba, Argentina), *Water Research,* 35, pp. 2881–2894.

Zhang, X., Wang, Q., Liu, Y., Wu, J. & Yu, M. (2011). Application of multivariate statistical techniques in the assessment of water quality in the Southwest New Territories and Kowloon, Hong Kong, *Environmental Monitoring and Assessment,* 173, pp. 17–27.