Principal component analysis of surface water quality data of the River Drava in eastern Croatia (24 year survey)

Vlatka Gvozdić, Josip Brana, Nela Malatesti and Danijela Roland

ABSTRACT

The River Drava is one of the major, inexhaustible water sources not only for Croatia, but also for the other European countries it flows through. This study is based on the observations of 15 water variables at three sampling stations in the lower River Drava over a 24 year period. Although the obtained results revealed an improvement of most of the parameters, the values of some of them (i.e. NH_4-N, NO_3-N, BOD_5, total coliform and heterotrophic bacteria) are still above the approved limits for water Class II. The results of principal component analysis (PCA) confirmed an existence of three clearly separated zones. The first zone corresponds to a rural upstream part of River Drava, which is characterised with low level pollution. The influences of untreated domestic waters become more noticeable in the second more densely populated suburban zone (II) located upstream of the city of Osijek. According to the results of the PCA, untreated wastewaters from Osijek are becoming contributing factors to the high pollution level of the river in the third (III) suburban zone. This study shows the usefulness of the PCA method for analysis and interpretation of complex data sets as well as for determination of pollution sources.

Key words | Drava river, principal component analysis, water quality

Vlatka Gvozdić (corresponding author)
Nela Malatesti
Danijela Roland
Department of Chemistry,
University of J. J. Strossmayer,
Franje Kuhača 20,
HR-31000 Osijek,
Croatia
E-mail: *Vlatka.Gvozdic@kemija.unios.hr*

Josip Brana
Department of Physics,
University of J. J. Strossmayer,
Trg Ljudevita Gaja 6,
HR-31000 Osijek,
Croatia

INTRODUCTION

The source of the River Drava lies in Tyrol (North Italy) and runs through five different European countries: Italy, Austria, Slovenia, Hungary and Croatia. It ends its course through Croatia near the city of Osijek, where it flows into the River Danube 21 km downstream. As an inter-state river the water quality of the Drava is important for the countries through which it flows but also for the quality of the River Danube, of which the Drava is a tributary.

The section of the River Drava from Legrad to the Drava's flow into the Danube is called the Lower Drava (Figure 1) and it represents an important factor in the social development of the area. The River Drava provides a vital resource for water supply, irrigation, fishing, navigation and recreation, but is also a natural wastewater recipient from towns and settlements located in its river basin. From an economic point of view, the most important use of the River Drava is hydro-energy (Tušar & Mijušković-Svetinović 2000).

However, the environmental value of Drava is recently getting increased attention. The Drava is extremely important for conservation of biodiversity in both Croatia and at the European level (Maričić 2005). This is because its water course and surrounding wetland area are among the best preserved in Central Europe, while the Danube at its confluence with the Drava determines the flood area of the Nature Park Kopački Rit. The banks of the Drava are higher than the ground in the Kopački Rit itself. Therefore, under the increased water levels the changes in the Drava's water quality may influence the water quality of the bigger canals and subsequently that of the numerous, overgrown, almost invisible canals, old small backwater canals and depressions (Bogut *et al.* 2010). It would be extremely valuable to investigate the full potential of surface water of the River Drava for drinking in this area, as very high arsenic concentrations have been recorded in the groundwater that is being used for drinking in eastern Croatia

doi: 10.2166/hydro.2012.063

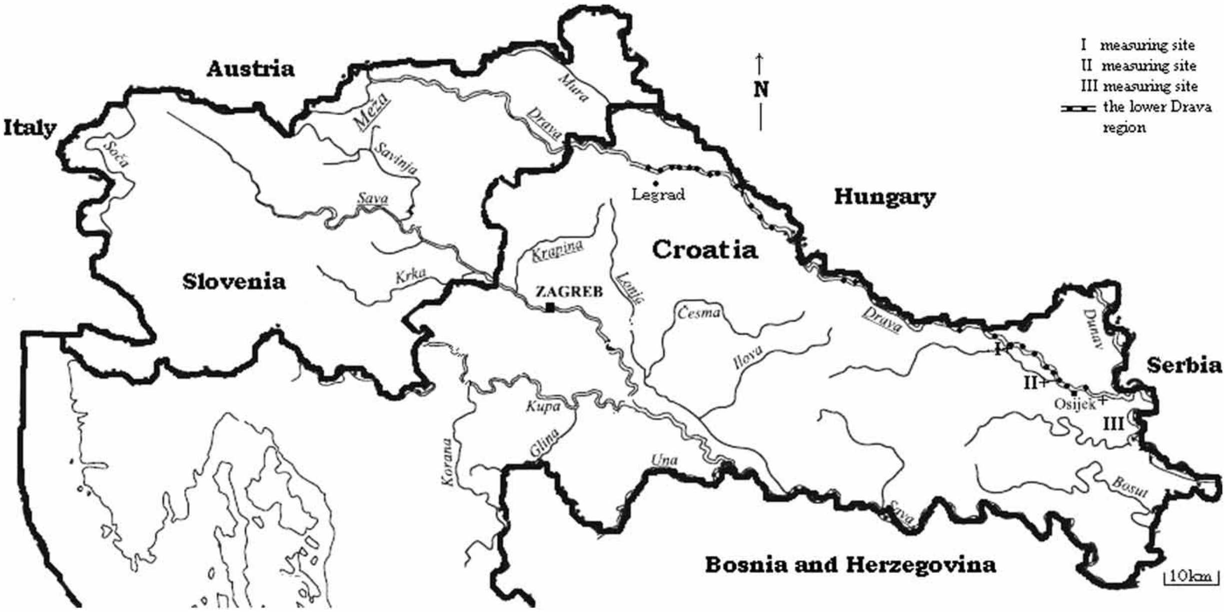

Figure 1 | Map of locations of measuring sites.

(Romić *et al.* 2011). Moreover, it is believed that almost 200,000 people daily are drinking water with arsenic concentrations ranging from 10 to 610 $\mu g\ L^{-1}$ (Habuda-Stanić *et al.* 2007). Therefore, the objective of the present research is to provide information on the physicochemical and microbiological status of the River Drava's water. This is in order to appreciate the impacts of anthropogenic activities on the quality of the river's water as well as discuss its suitability for human consumptions in compliance with the Water Regulation of Waters Classification (Croatian Water Classification Act 1998). According to the Croatian Water Classification Act, waters are divided into five classes according to their purpose and their purity grade.

The human activities are a major factor determining the surface water quality through things such as: atmospheric pollution, effluent discharges, use of agricultural chemicals and livestock wastes. All these are ultimately making the water unsuitable for human use (Brodnjak-Vončina *et al.* 2002; Vitale *et al.* 2002; Milovanović 2007; Enache *et al.* 2009; Živadinović *et al.* 2010). In the urban environment, river runoffs from the highways and streets introduce numerous contaminants into the surface water, including nitrogen, suspended solids, bacteria and organic contamination. Long-term measurements produce a large number of variables that are often difficult to interpret.

As a result, the particular problem in the case of surface quality monitoring is a complexity deriving from analysis of large sets of data. The methods based on descriptive (Helena *et al.* 2000; Bengraïne & Marhaba 2003; Simeonov *et al.* 2003; Çamdevýren *et al.* 2005; Ouyang *et al.* 2006; Chau & Muttil 2007) and predictive data mining (DM) techniques have proven to be effective for investigating water quality data (Cheng *et al.* 2002; Chau 2006; Lin *et al.* 2006; Muttil & Chau 2006; Wang *et al.* 2009; Wu *et al.* 2009). The application of principal component analysis (PCA) for the interpretation of a large and complex volume of data offers a better understanding of water quality and ecological status of the study systems, while at the same time it allows the identification of possible factors/sources that influence the surface water systems, thus providing a valuable tool for rapid solution of pollution problems (Simeonov *et al.* 2003).

The present study is rather specific as it deals with the largest volume of data covering the broadest range of water quality parameters for the River Drava that has ever been accumulated and analysed by means of PCA. In this study, three data matrices, which were obtained during the 24-year monitoring program (1985–2008), were subjected to PCA in order to extract the main parameters that are the most important in assessing variations in River Drava water quality, and to obtain information regarding spatial variations.

MATERIALS AND METHODS

Study area

The selected sampling stations are situated in the vicinity of the main pollutant sources such as the agriculture, residential land, and a city zone.

The first zone (I) is the agricultural area upstream of the industrial town Belišće, the second zone (II) is the suburban area upstream of the city of Osijek and the third zone (III) includes the suburban area located downstream of Osijek.

Belišće is situated near the border with Hungary at 45,41° N, 18,25° E (with a population of approximately 11,900). Main occupations in this area include intensive vegetable production, farming, livestock, chemical and mechanical wood processing, sawmill and paper production industry. The other important settlement is a suburban part of Osijek, located upstream of Osijek (measurement site III) with approximately 7,100 inhabitants.

The most significant centre in terms of the number of inhabitants, with a population of approximately 120,000 at 45,32° N, 18,44° E is Osijek. It has an elongated shape and its area is parallel with the River Drava's course in an east–west direction.

Data

The data were obtained from Croatian Waters, the State Water Management Agency responsible for surface and ground water quality monitoring. The weekly measurements included: water level (WL, cm), temperature (T, °C), pH, electrical conductivity (EC, μScm^{-1}), total suspended solids (TSS, mg L^{-1}), dissolved oxygen (DO, mgO$_2$ L^{-1}), oxygen saturation (OS, %sat), chemical oxygen demand (COD, mgO$_2$ L^{-1}), 5-day biochemical oxygen demand (BOD$_5$, mgO$_2$ L^{-1}), ammonia-nitrogen (NH$_4$-N, mg L^{-1}), nitrite nitrogen (NO$_2$-N, mg L^{-1}), nitrate nitrogen (NO$_3$-N, mg L^{-1}), total inorganic nitrogen (TIN, mg L^{-1}). Monthly measurements included: total coliform bacteria (TCol., MPNC/100 ml) and heterotrophic bacteria (Het., cfu ml^{-1}).

The surface water quality is monitored by the Public Health Institute of Osječko-baranjska County. Current legislation in the Republic of Croatia regulates standards and designates five classes for quality of the surface waters (Croatian Water Classification Act). The first-class quality includes surface waters used for drinking or in the food industry either in their natural state or after disinfection, as well as the surface water used for rearing high-quality species of fish. The second class quality implies water used in its natural state for swimming and recreation and water which after appropriate treatment can be used for drinking and other industrial purposes. The third class quality includes water used in industry and agriculture with no specific requirements on its quality. The fourth class quality includes water which is used only after having been treated and in the areas with water shortage. The fifth class quality refers to the waters that generally cannot be used.

Data treatment

In the present study, the statistical analysis of the water quality parameters was performed using the statistical software Statistica, version 7.0. Beside basic descriptive statistics (median, minimum, maximum, 10th and 90th percentile values) of measured variables, the obtained physicochemical and microbiological data were subjected to PCA in order to evaluate the most significant parameters in the river water quality assessment.

The result of PCA can be influenced by uneven sampling, missing values or observations below detection limits of analytical methods which can be changed during the data collection period. These influences in our case are not strong enough to produce discard of important information, influences on mutual correlations or possible occurrences of outliers in PCA, because such cases were below 2% of total data and were quite random. Before data processing, a few missing values as well as observations below detection limits were replaced with a mean value of the nearest neighbours and with values equal to one half of detection limits, respectively. For the purposes of the PCA method, we used the average monthly values of physicochemical variables since the samples for microbiological analysis were collected on a monthly basis. The original data were also standardised in order to avoid misclassifications arising from different orders of magnitude of measured parameters. For this purpose, the data were mean (average) centered and divided by the relevant standard deviations.

Data analysis (PCA)

Although basic statistics have the ability of extracting precise characteristics from observed data, the information so obtained are limited, thus, for more comprehensive insight into river water network, we need the power of the multivariate statistical analysis. Multivariate statistical techniques, such as PCA, attempt to explain the correlations between the observations in terms of not directly observable underlying factors as well as to identify the possible sources that influence water quality.

The basic idea of PCA is to reduce the dimensionality of a data set containing a large number of correlated variables, while retaining as much as possible of the variation present in the data. This reduction is achieved by transforming original variables into a new set of variables, i.e. the principal components (PCs), which are uncorrelated, and are ordered so that first few retain most of the variation present in all of the original variables. The PCs can be rotated to a more interpretable structure by varimax or other rotation algorithms. In general:

$$\mathrm{PC}_i = l_{1i}X_1 + l_{2i}X_2 + \cdots + l_{ni}X_n \tag{1}$$

where PC_i is ith PC and l_{ji} is the loading of the observed variable $X_{j.}$ The absolute value of the loadings in relation to the considered PC is a measure of the variable importance for the PC model.

In this study the PCs were estimated by the PCA method and then rotated by means of the Varimax rotation which ensured that each variable is maximally correlated with only one component, and is in a near zero association with the other components. The number of PCs was defined according to the eigenvalue >1 criterion (Jolliffe 1986; Vandenginste *et al.* 1998). The component loadings were used to determine the relative importance of the variables measured in this study. The original data were standardised in order to avoid misclassifications arising from different orders of magnitude of measured parameters. For this purpose the data were mean (average) centered and divided by the relevant standard deviations.

RESULTS AND DISCUSSION

In this section we present the spatial analysis of the 15 physicochemical and microbiological parameters that have been monitored in three sampling stations during a 24-year period. Current legislation in the Republic of Croatia regulates standards and assigns five classes for the quality of surface waters. Therefore, the statistical parameters 10th percentile (for DO and OS) and 90th percentile (for pH, EC, COD, BOD_5, NH_4-N, NO_2-N, NO_3-N, TIN, total coliform and heterotrophic bacteria) must be considered. Tables 1a–1c show descriptive statistics for all the monitored parameters in the three sampling stations over the 24-year period. The parameters that are not included in the classifications are marked with an asterisk.

Through the analysis of the data it becomes obvious that the median values of BOD_5, COD, TIN, the total coliform and the heterotrophic bacteria increase gradually during the river passage downstream. According to the current regulations and the results from Tables 1a–1c, it is evident that during the 24-year period pH, EC and DO meet the standards for the Class I while OS, COD and NO_2-N fall into Class II. However, the concentrations of NH_4-N, NO_3-N and TIN were slightly higher, with the 90th percentile values being just above the limit for Class II quality.

However, a somewhat more pronounced problem for the Drava's water quality is the BOD_5 values, the total coliform and the heterotrophic bacteria. This occurs especially at measuring station III, where total coliform bacteria counts only meet the criteria for Class V. The values of the BOD_5 in samples from the sites I and III indicate that organic compounds that fall into the river are not completely destroyed by the self-purification processes, while the oxygen regime (DO and OS) was still good at all three observed sites.

The water samples of the Drava contained lower median concentrations of NH_4-N than median concentrations of NH_4-N in Eastern Europe (1.0–0.25 mg L^{-1}) and slightly less than median concentrations in the Danube River basin (0.23–0.1 mg L^{-1}) (Serbian Environmental Protection

Table 1a | Median, minimum, maximum, 10th and 90th percentile, class of the water and standard deviation (SD) at sampling site I

Variable	Median	Minimum	Maximum	Percentile	Class	S.D.
WL/cm	174	−134	502	311	*	85
T/°C	12.8	0.0	26.0	22.0	*	7.1
pH	7.8	6.9	9.5	8.2	I	0.3
EC/μS cm^{-1}	320	190	580	399	I	55
TSS/mg L^{-1}	40	2	220	70	*	25
DO/mg O_2 L^{-1}	9.7	4.7	24.4	7.6	I	2.5
OS/% sat	92.5	44.9	253.9	71.5	II	21.4
COD/mg O_2 L^{-1}/	3.3	1.0	10.0	5.8	II	1.4
BOD_5/mg O_2 L^{-1}	2.50	0.03	25.13	5.58	III	2.17
NH_4-N/mg L^{-1}	0.150	0.001	0.970	0.290	III	0.120
NO_2-N/mg L^{-1}	0.012	0.001	0.140	0.020	II	0.007
NO_3-N/mg L^{-1}	1.90	0.01	8.20	3.10	III	0.81
TIN/mg L^{-1}	2.10	0.08	8.50	3.25	III	0.88
TCol/MPNc/100 mL	9.2×10^2	2	5.4×10^4	3.0×10^3	III	5.0×10^3
Het/cfu/mL	1.2×10^3	30	6.8×10^4	8.3×10^3	II	7.6×10^3

Agency, SEPA 2008; European Environment Agency, EEA 2009; Babović *et al.* 2011). The median concentrations of NO_3-N in the River Drava's water are relatively low compared to other rivers in western Europe (4.4–3.8 mg L^{-1}) (EEA 2009). These values are similar to those obtained in other rivers in eastern Europe, the Danube and the River Sava (2.1–2.0 1.6–0.95 and 1.3–1.2 mg L^{-1} NO_3-N, respectively) (SEPA 2008; Babović *et al.* 2011). In the period between 1992 and 2006, median BOD_5 values in European rivers decreased from 5 to 2 mg O_2 L^{-1}, and in the Danube

Table 1b | Median, minimum, maximum, 10th and 90th percentile, class of the water and standard deviation (SD) at sampling site II

Variable	Median	Minimum	Maximum	Percentile	Class	S.D.
WL/cm	30	–128	429	202	*	108
T/°C	12.8	0.0	27.0	22.0	*	7.2
pH	7.8	6.9	9.6	8.2	I	0.3
EC/μS cm^{-1}	325	147	681	418	I	62
TSS/mg L^{-1}	48	2	250	80	*	26
DO/mg O_2 L^{-1}	9.6	4.4	23.1	7.5	I	2.5
OS/% sat	90.3	40.8	261.2	70.0	II	21.1
COD/mg O_2 L^{-1}	3.6	0.9	11.0	6.3	II	1.6
BOD_5/mg O_2 L^{-1}	2.70	0.03	17.50	5.77	III	2.27
NH_4-N/mg L^{-1}	0.190	0.001	0.900	0.350	III	0.130
NO_2-N/mg L^{-1}	0.012	0.001	0.300	0.020	II	0.010
NO_3-N/mg L^{-1}	1.90	0.23	4.5	3.10	III	0.77
TIN/mg L^{-1}	2.16	0.08	6.70	3.30	III	0.84
TCol/MPNc/100 mL	2.3×10^3	4	6.7×10^4	6.7×10^3	III	5.8×10^3
Het/cfu/mL	2.7×10^3	80	1.5×10^5	1.5×10^4	III	1.8×10^4

Table 1c | Median, minimum, maximum, 10th and 90th percentile, class of the water and standard deviation (SD) at sampling site III

Variable	Median	Minimum	Maximum	Percentile	Class	S.D.
WL/cm	19	−128	429	198	*	107
T/°C	12.5	0.000	27.0	22.0	*	7.2
pH	7.8	6.9	9.2	8.2	I	0.3
EC/μS cm^{-1}	327	164	577	418	I	60
TSS/mg L^{-1}	50	2	182	80	*	25
DO/mg O_2 L^{-1}	9.6	4.1	24.2	7.4	I	2.5
OS/% sat	90.2	36.8	249.1	70.2	II	21.0
COD/mg O_2 L^{-1}/	3.8	1.3	18.2	6.5	II	1.8
BOD_5/mg O_2 L^{-1}	2.91	0.08	37.20	6.20	III	2.71
NH_4-N/mg L^{-1}	0.190	0.001	3.88	0.38	III	0.20
NO_2-N/mg L^{-1}	0.012	0.001	0.066	0.02	II	0.01
NO_3-N/mg L^{-1}	1.90	0.22	8.30	3.20	III	0.85
TIN/mg L^{-1}	2.20	0.07	8.80	3.4	III	0.93
TCol/MPNc/100 mL	3.3×10^3	13	1.6×10^6	1.6×10^5	V	8.2×10^4
Het/cfu/mL	6.1×10^3	64	2.3×10^5	3.4×10^4	III	2.5×10^4

basin they decreased from 3 to 2.6 mg O_2 L^{-1} (Živadinović *et al.* 2010). These values represent the closest comparison to those observed in the water of the River Drava. Many factors affect the quality of the River Drava, and although the quality of the Drava and Mura rivers have improved over the past years in Slovenia, the Drava comes into Croatia with a lower quality than desired Class II (Vitale *et al.* 2002). The lower River Drava flow with its tributaries is a wastewater recipient for a number of cities and other human settlements situated along the river banks. Since 1973 there have been several purification wastewater systems built in many small towns located upstream (Mijušković-Svetinović *et al.* 2001). However, the city of Osijek with its 120,000 inhabitants and its industry does not have a wastewater treatment plant, therefore the entire wastewater of this city is discharged into the Drava as well as into the nearby waters. Even though the River Drava manages to keep its physicochemical parameters at the acceptable levels I and II (or just above the level II at some points), at the very end of its course (in the close proximity to the National Park Kopački Rit and its estuary into the Danube), the untreated wastewater from the city of Osijek and the surrounding area deteriorates Drava's water quality practically to the point of becoming completely unusable.

Principal component analysis

Tables 2a–2c show the factor loadings obtained by PCs analysis with the varimax rotation. The strong (>0.75) and the moderate (0.75–0.50) (Liu *et al.* 2003) factor loadings are made bold and considered for interpretation.

As can be seen in Table 1a, six PCs with eigenvalues >1 assuming 75.7% of the total variation were obtained at measurement site I. This is enough to give adequate representation of the data. As shown by the factor loadings matrix, the first PC, accounting for 20.7% of the total variation, was correlated heavily with the T and EC and moderately with the WL. The second PC, accounting for 16.2%, was strongly correlated with BOD, DO and OS, and the third PC, accounting for 12.9%, was influenced strongly by the COD and NH_4-N. The remaining variables were represented by the rest of the PCs accounting successively for less of the total variation. PC4 (11.3%) and PC5 (7.7%) loaded heavily on NO_3-N, TIN, the total coliform and the heterotrophic bacteria, respectively. PC6 (7%) exhibited moderate correlation with TSS and pH .

It is evident from the comparison of the results from sites I and II that the loadings of the first and the fourth factor are exchanged.

From Table 2b (station II) it is noticeable that dominant component PC1 was strongly influenced this time with

Table 2a | Rotated principal component loadings at site I

	PC1	PC2	PC3	PC4	PC5	PC6
WL/cm	−**0.664**	−0.128	−0.104	0.142	0.119	−0.217
T/°C	−**0.829**	0.039	−0.145	−0.283	−0.031	0.151
pH	−0.169	0.189	−0.114	0.083	−0.048	0.675
EC/μS cm^{-1}	**0.820**	0.111	−0.164	0.036	0.108	0.028
TSS/mg L^{-1}	−0.233	0.010	0.347	−0.024	−0.058	−**0.661**
DO/mg O$_2$ L^{-1}	0.455	**0.811**	0.034	0.136	0.005	0.017
OS/% sat	−0.167	**0.904**	−0.057	−0.086	−0.034	0.137
COD/mg O$_2$ L^{-1}/	−0.050	0.033	**0.857**	0.027	0.078	0.039
BOD$_5$/mg O$_2$ L^{-1}	0.054	**0.810**	0.019	0.082	0.094	−0.107
NH$_4$-N/mg L^{-1}	0.128	−0.056	**0.804**	0.177	0.052	−0.286
NO$_2$-N/mg L^{-1}	0.208	0.142	−0.295	0.369	0.108	−0.430
NO$_3$-N/mg L^{-1}	0.073	0.013	0.001	**0.959**	−0.002	0.033
TIN/mg L^{-1}	0.071	0.039	0.133	**0.950**	0.013	−0.021
TCol/MPNc/100 mL	0.022	−0.020	0.148	0.033	**0.766**	0.041
Het/cfu/mL	0.057	0.068	0.003	−0.016	**0.763**	−0.060
Expl.Var.%	20.7	16.2	12.9	11.3	7.7	7.0
Prp.Totl.%	20.7	36.9	49.8	61.1	68.6	75.8

Table 2b | Rotated principal component loadings at site II

	PC1	PC2	PC3	PC4	PC5	PC6
WL/cm	0.173	−0.048	0.018	−0.489	−0.189	−0.463
T/°C	−0.233	−0.016	0.181	−0.837	−0.015	0.177
pH	0.151	0.104	0.434	−0.034	−0.280	0.577
EC/μS cm^{-1}	0.076	0.140	0.128	0.809	−0.018	0.068
TSS/mg L^{-1}	0.034	−0.035	−0.698	−0.183	0.003	−0.178
DO/mg O$_2$ L^{-1}	0.094	0.845	−0.040	0.410	−0.023	0.005
OS/% sat	−0.082	0.896	0.094	−0.231	−0.052	0.149
COD/mg O$_2$ L^{-1}/	0.017	0.059	−0.723	0.142	−0.121	0.292
BOD$_5$/mg O$_2$ L^{-1}	0.075	0.790	−0.032	0.125	0.079	−0.163
NH$_4$-N/mg L^{-1}	0.176	−0.072	−0.743	0.104	0.062	−0.124
NO$_2$-N/mg L^{-1}	0.283	0.049	0.152	0.321	−0.128	−0.563
NO$_3$-N/mg L^{-1}	0.961	0.006	0.013	0.092	0.026	−0.001
TIN/mg L^{-1}	0.955	0.030	−0.108	0.085	0.047	−0.042
TCol/MPNc/100 mL	0.065	0.145	−0.256	−0.144	0.453	0.312
Het/cfu/mL	0.080	−0.033	0.045	0.037	0.882	−0.032
Expl.Var.%	19.7	15.9	12.7	10.9	8.7	7.7
Prp.Totl.%	19.7	35.6	48.3	59.2	67.9	75.6

Table 2c | Rotated principal component loadings at site III

	PC1	PC2	PC3	PC4	PC5	PC6
WL/cm	0.230	0.104	−0.107	−0.073	−0.487	0.205
T/°C	−0.305	0.006	−0.101	0.233	−0.809	−0.002
pH	0.091	0.093	−0.081	0.734	−0.094	0.021
EC/μS cm^{-1}	0.077	0.139	0.151	0.122	0.787	0.001
TSS/mg L^{-1}	0.096	−0.164	0.165	−0.476	−0.253	0.020
DO/mg O$_2$ L^{-1}	0.111	0.855	0.067	−0.070	0.425	0.013
OS/% sat	−0.135	0.928	−0.009	0.122	−0.214	0.009
COD/mg O$_2$ L^{-1}/	0.069	−0.024	0.636	−0.126	0.032	0.189
BOD$_5$/mg O$_2$ L^{-1}	0.051	0.391	0.736	0.065	0.173	−0.078
NH$_4$-N/mg L^{-1}	0.039	−0.121	0.843	0.019	0.094	−0.067
NO$_2$-N/mg L^{-1}	0.226	0.031	−0.108	0.060	0.377	0.183
NO$_3$-N/mg L^{-1}	0.970	−0.003	−0.079	−0.021	0.081	0.001
TIN/mg L^{-1}	0.956	−0.037	0.162	−0.014	0.108	−0.016
TCol/MPNc/100 mL	−0.096	−0.001	0.100	0.826	−0.048	−0.030
Het/cfu/mL	0.019	−0.010	0.026	0.009	0.013	−0.951
Expl.Var.%	19.8	15.2	12.5	11.6	9.3	7.9
Prp.Totl.%	19.8	35	47.5	59.1	68.4	76.3

NO$_3$-N and TIN. The variations in number of the total coliform bacteria in this case appeared not to be significant for any component.

The component loading patterns obtained for PC1, PC2 and PC3 at monitoring site III (Table 2c) were similar to those for site II. The noticeable difference is in component PC4 (11.6%) which was highly correlated with total colliform bacteria and pH.

As shown in Tables 2b and 2c, only TIN and NO$_3$-N dominated the first PC, whereas NH$_4$-N and NO$_2$-N were distributed on different factors (NH$_4$-N being associated with COD, BOD$_5$ or TSS on factor 3), or they do not have high loadings on any remaining PCs (Table 2a). The reason for this probably lies in the fact that NH$_4$-N and NO$_2$-N values have small standard deviations compared to NO$_3$-N and TIN, thus the effect of nutrient processes are subdivided into several PCs. Such a pattern suggests for NO$_3$-N different sources and behaviour than that of the NH$_4$-N and NO$_2$-N.

The PCA indicates that the most relevant variables defining water quality at sites II and III are related to NO$_3$-N and TIN, and less relevant ones are BOD$_5$, COD or NH$_4$-N which can be assigned to nutrients and organic components of water, in both regions. The suburban part upstream of the city (II) is populous; thus the influence of domestic sewage as well as agricultural activities on the loadings proved to be considerable (Table 2b). The same situation exists at region III located downstream (Table 2c). In contrast, station I (less populated-rural part) was not affected by nutrient variables. Bacterial indicators like total coliforms are widely applied to assess the degree of pollution of river systems from various external sources (Kirschner *et al.* 2009). Although, most of these bacteria are widely distributed in the environment, they indicate anthropogenic impact such as human or animal excreta or the availability of easily degradable organic material. The evaluation of microbiological results (Table 1c) and the results of PCA (Table 2c) showed the highest pollution at location III downstream from Osijek. This clearly indicates that the bacterial contamination is caused by wastewater discharges. This situation represents a potential public health hazard, because the permissible 90th percentile values of total coliform bacteria in the surface water for water Class II is 5×10^1 to 5×10^2/100 ml. In fact, the wastewater coming from the city, together with the industry and sanitary wastewater, is being discharged into the River Drava without being previously treated (Mijušković-Svetinović *et al.* 2001). Consequently, the quality of the Drava water downstream of the city is significantly lower according to the number of the total coliform bacteria, so it only meets the Class V criteria. Although the pollution caused by the industry has decreased over recent years (Vitale *et al.* 2002) as a result of reduced economic activity and technological improvements, pollution from the urbanised area along the river is still severe and the volumes of domestic sewage and the surface runoffs tend to grow with the course of urbanisation. The implementation of wastewater treatment plants and appropriate wastewater treatment measures according to the European Union directive (European Council 1991) will have great potential to reduce microbiological pollution in the Drava River and thus improve water quality. In this study, PCA analysis demonstrated deteriorating conditions of water quality due to city wastewater, as well as relatively better situations in the upper (rural part) of the Drava River. Use of PCA makes it possible to identify the major processes controlling variation in the studied aquifer and can be useful in aquifers where hydrochemical and microbiological data are available.

CONCLUSIONS

This paper demonstrates the usefulness of the PCA method when a voluminous and complex dataset is used in obtaining information about different influences that can affect the surface water quality. PCA has helped to identify the factors responsible for the main surface water quality variations during a 24 year period. The results of PCA showed that six factors accounted for about 76% of the variance in the data sets. Although the PCA did not result in a significant variables reduction, it helped to extract the sources responsible for variations in river water quality at different sampling sites. The results of PCA showed that the level of pollution increases downstream. While differences between first two, rural and suburban sites are visible already on the first component, the differences between the two suburban stations located up and downstream of the city exist only on one of the remaining components (PC4). The most worrying results come from the microbiological parameters at site III. At this suburban region located downstream from the city, the rotated component loadings identified that the variables responsible for water quality are mainly related to the nutrients (high loads of NO_3-N, TIN), the organic matter, and the total coliform bacteria thus reflecting anthropogenic activities. This study has shown that the effluents from Osijek have a big impact on the water quality of the River Drava in that it becomes inappropriate for any use once downstream from the city. This study is important for understanding the River Drava's water quality and may be useful for future activities regarding water resources pollution control and implementation of the Water Framework Directive 2000/60/EC 9 (EC 2000). The finding that the pollution level of the River Drava's water downstream of Osijek is quite serious implies an urgent need for constructing a wastewater treatment plant as well as introducing more comprehensive water quality measurements, especially in its downstream section.

ACKNOWLEDGEMENT

The authors thank Croatian Waters State Water Management Agency for making relevant data available.

REFERENCES

Babović, N., Marković, D., Dimitrijević, V. & Marković, D. 2011 Some indicators of water quality of the Tamiš River. *Chem. Ind. Chem. Eng. Q.* **17**, 107–115.

Bengraïne, K. & Marhaba, T. F. 2003 Using principal component analysis to monitor spatial and temporal changes in water quality. *J. Hazard. Mater.* **B100**, 179–195.

Bogut, I., Čerba, D., Vidaković, J. & Gvozdić, V. 2010 Interactions of weed bed invertebrates and Ceratophylumsemersum stands in a floodplain lake. *Biologia.* **65**, 113–121.

Brodnjak-Vončina, D., Dobčnik, D., Novič, M. & Zupan, J. 2002 Chemometrics characterisation of the quality of river water. *Analytica Chim. Acta*. **462**, 87–100.

Çamdevýren, H., Demýr, N., Kanik, A. & Keskýn, S. 2005 Use of principal component scores in multiple linear regression models for prediction of *Chlorophyll-a* in reservoirs. *Ecol. Modell.* **18**, 581–589.

Chau, K. W. 2006 Particle swarm optimization training algorithm for ANNs in stage prediction of Shing Mun River. *J. Hydrol.* **329**, 363–367.

Chau, K. W. & Muttil, N. 2007 Data mining and multivariate statistical analysis for ecological system in coastal waters. *J. Hydroinf.* **9**, 305–317.

Cheng, C. T., Ou, C. P. & Chau, K. W. 2002 Combining a fuzzy optimal model with a genetic algorithm to solve multiobjective rainfall-runoff model calibration. *J. Hydrol.* **268**, 72–86.

Croatian Water Classification Act 1998 Official bulletin of the Republic of Croatia, Narodne novine. No.77, 1998, Zagreb, Croatia.

Enache, I., Birghila, S. & Dumbrava, A. 2009 The Danube river water quality characteristics in the Braila Town. *Ovidius Univ. Ann. Chem.* **20**, 146–152.

European Council 1991 Council directive 91/271/EEC. 21 May 1991 concerning urban-waste treatment. *Off. J.* **L135**, 40–52.

European Council 2000 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. *Off. J.* **L327**, 22/12/2000 0001–0073.

European Environment Agency (EEA) 2009 Report: Oxygen Consuming Substances in Rivers–Assessment published January 2009. EEA, Copenhagen, Denmark.

Habuda-Stanić, M., Kuleš, M., Kalajdžić, B. & Romić, Ž. 2007 Quality of groundwater in eastern Croatia. The problem of arsenic pollution. *Desalination* **210**, 157–162.

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 Res.* **34**, 807–816.

Jolliffe, I. T. 1986 *Principal Component Analysis*. Springer, Berlin, pp. 63–147.

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

Lin, J. Y., Cheng, C. T. & Chau, K. W. 2006 Using support vector machines for long-term discharge prediction. *Hydrolog. Sci. J.* **51**, 599–612.

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. *Sci. Total. Environ.* **313**, 77–89.

Maričić, S. 2005 Analysis of the one of the rare natural retentin in the middle Danube. In *Proceedings of the Ninth International Symposium on Water Management and Hydraulic Engineering*. 4–7 September, 2005, Ottenstein, Austria, pp. 383–393.

Mijušković-Svetinović, T., Maričić, S. & Tušar, B. 2001 The wastewater of the city of Osijek. VII International Symposium on Water Management and Hydraulic Engineering, Poland.

Milovanović, M. 2007 Water quality assessment and determination of pollution sources along the Axios/Vardar River, southeastern Europe. *Desalination* **213**, 159–173.

Muttil, N. & Chau, K.W. 2006 Neural network and genetic programming for modelling coatal algal blooms. *J. Environ. Pollut.* **28**, 223–238.

Ouyang, Y., Nkedi-Kizza, P., Wu, Q. T., Shinde, D. & Huang, C. H. 2006 Assessment of seasonal variations in surface water quality. *Water Res.* **40**, 3800–3810.

Romić, Ž., Habuda-Stanić, M., Kalajdžić, B. & Kuleš, M. 2011 Arsenic distribution, concentration and speciation in groundwater of the Osijek area, eastern Croatia. *Appl. Geochem.* **26**, 37–44.

Serbian Environmental Protection Agency (SEPA) 2008 *National Report on the State of the Environment*. Beograd, Serbia.

Simeonov, V., Stratis, J. A., Samara, C., Zachariadis, G., Voutsa, D., Anthemidis, A., Sofouniou, M. & Kouimtzis, Th. 2003 Assessment of the surface water quality in northern Greece. *Water Res.* **37**, 4119–4124.

Tušar, B. & Mijušković-Svetinović, T. 2000 The water quality of multipurpose accumulations. In *20th Conference of the Danubian countries on Hydrological Forecasting and Hydrological Basses of Water Management*, 4–8 September, Bratislava, Slovakia.

Vandeginste, B. G. M., Massart, D. L., Buydens, L. M. C., De Jong, S., Lewi, P. J. & Smeyers-Verbeke, J. 1998 *Handbook of Chemometrics and Qualimetrics, Part A*. Elsevier, Amsterdam, pp. 519–556.

Vitale, K., Marijanović Rajčić, M. & Senta, A. 2002 Waters in Croatia between practice and needs: Public health challenge. *Croat. Med. J.* **43**, 485–492.

Wang, W. C., Chau, K. W., Cheng, C. T. & Qiu, L. 2009 A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *J. Hydrol.* **374**, 294–306.

Wu, C. L., Chau, K. W. & Li, Y. S. 2009 Predicting monthly streamflow using data-driven models coupled with data-preprocessing techniques. *Water. Resour. Res.* **45**, W08432.

Živadinović, I., Ilijević, K., Gržetić, I. & Popović, A. 2010 Long-term changes in the eco-chemical ststus of the Danube River in the region of Serbia. *J. Serb. Chem. Soc.* **75**, 1125–1148.

First received 13 May 2011; accepted in revised form 7 December 2011. Available online 2 March 2012