ISSN 1848-0071
UDC 574.64+519.2=111
Recieved: 2012-06-16
Accepted: 2012-09-07
Original scientific paper

MULTIVARIATE ANALYSIS OF POLLUTION INDICATORS OF MUNICIPAL WASTEWATER IN SPLIT AREA

MLADEN PERIŠIĆ, MIRJANA ČURLIN[1], MARIJA BRALIĆ[2], MARIJO BUZUK[2], LJUBICA ŠIMAC PERAICA[3]

The County of Split and Dalmatia, Department of environmental protection, Split, Croatia
[1]Faculty of Food Technology and Biotechnology, University of Zagreb, Zagreb, Croatia
[2]Faculty of Chemistry and Technology, University of Split, Split, Croatia
[3]Water and Sewer service, Laboratory for Wastewater, Split, Croatia
e-mail: mcurlin@pbf.hr

This paper presents the results of continuous monitoring of 11 pollution indicators of municipal wastewater that were discharged into the natural recipient (the sea) at three locations in the Split area during four years (2006-2009). Experimental data were analyzed by basic statistical methods for determination of mean and median values, standard deviations, minimal and maximal value of the measured parameters and their mutual correlation coefficients, normality test for all parameters and by different chemometric methods, such as cluster analysis (CA), factor analysis (FA) and principal component analysis (PCA).
Multivariate statistical techniques for evaluation and interpretation of large complex data sets used in this study provide better insight into the information about water quality and make a decision concerning the design of monitoring network for effective management of water resources much easier.
Key words: municipal wastewater, pollution indicators, statistical analysis, factor analysis, principal component analysis.

Multivarijantna analiza pokazatelja onečišćenja komunalne otpadne vode šireg područja Splita. Ovaj rad predstavlja rezultate kontinuiranog monitoringa 11 pokazatelja onečišćenja komunalne otpadne vode koja se ispušta u prirodni recipijent (more) na tri lokacije na širem području grada Splita u razdoblju od 4 godine (2006-2009). Na dobivenom setu podataka provedena je osnovna statistička analiza određivanja srednjih vrijednosti i medijana, standardne devijacije, minimalnih i maksimalnih vrijednosti mjerenih parametara i njihovih međusobnih korelacijskih koeficijenata, te su provedene različite kemometrijske metode kao što su analiza klastera (CA), faktorska analiza (FA) i analiza glavnih komponenata (PCA). Ovo istraživanje predstavlja nužnost i korisnost multivarijantne statističke tehnike za vrednovanje i interpretaciju velikih složenih skupova podataka dajući bolje informacije o kakvoći vode i oblikovanju nadzora s ciljem učinkovitog upravljanja vodenim resursima.
Ključne riječi: komunalna otpadna voda, pokazatelji onečišćenja, statistička analiza, faktorska analiza, analiza glavnih komponenata.

INTRODUCTION

Wastewater is discharged into environment recipients (lakes, rivers, the sea) as a consequence of the industrial and agricultural production as well as the town activities, and it has an unhealthy influence on the humans and the environment. Safe disposal of municipal and industrial wastewater is an essential requirement under the Environmental Protection Act, which establishes maximum permissible levels of various pollutants to be discharged into a natural recipient. This places an emphasis on the proper treatment of sewage/wastewater prior to its final disposal. For the selection of an appropriate treatment process and the design of parameters to maintain stable conditions for optimum performance of the wastewater treatment plants (WWTPs) round the year, detailed information about the sources, composition and levels of pollutants in the wastewater is needed [1]. The quality of water resources is undermined to the extent that natural mechanisms of purification cannot compensate for the consequences of pollution, to which they are exposed daily. Systematic environmental monitoring of pollution indicators of municipal wastewater is important for determination of its quality, before discharging it in a natural recipient. Also, statistical analysis of pollution indicators provides valuable information that can be used in on-line monitoring of water quality in an outlet stream. This is also a necessary step in a preliminary study for process design of a wastewater treatment plant based on mathematical models, thus enabling effective process control and optimisation.

All pollution indicators were measured continuously and the result of these measurements is a complex data matrix comprised of a large number of physico-chemical parameters [2] which are often difficult to interpret and draw meaningful conclusions from. A structured approach to monitoring and analysis of pollution indicators is required. Multivariate analysis provides a methodology to extract and interpret information from large amounts of data. It is also a necessary step in preliminary study for process design of biological wastewater treatment plant based on mathematical models, and provides for effective process control and optimization. To monitor water quality and to make qualitative and quantitative decisions based on real data has become a challenge for environmental engineers during all stages of the process, from data collecting, storage and processing up to the analysis and interpretation of the results [3].

Cluster analysis is an exploratory data analysis tool for solving classification problems. Cluster analysis groups the objects (monitoring locations) into clusters on the basis of similarities and dissimilarities between object.

Factor analysis, which includes principal component analysis (PCA), is a very powerful technique applied to reduce the dimensionality of a data set consisting of a large number of inter-related variables, while retaining the variability in the data setas much as possible. This reduction is achieved by transforming the data set into a new set of variables, principal components (PCs), which are orthogonal (non-correlated) and arranged in decreasing order of importance [4]. Principal component analysis provides information on the most meaningful parameters, which describe the whole data set, rendering data reduction with minimum loss of original information [5]. This study presents the results of monitoring program that was performed during 4 years (2006-2009) and it was subjected to CA, PCA and FA. The aim of this study was to emphasize the importance of data reduction techniques of PCA/FA in the analysis of various pollution indicators and to determine by CA the similarities among three drain locations located in wider Split area.

EXPERIMENTAL

Case study area

The city of Split is located in the warmest region of northern Mediterranean coast and it is industrial, university and business center of the region, with the population of about 200 000. Sewer system has two main basins that collect wastewater. The south basin collects water from the parts of town that stretch from east to west, while the north basin collects wastewater from the northern part of Split. This basin is characterized by three major collectors from all over the secondary sewage network. Through these collectors the wastewater is brought to a mechanical device. After mechanical (primary) treatment, the separation of sand, grease and oil, the water is discharged into the Brač Channel.

Figure 1. The area with indicated locations of sampling
Slika 1. Područje s označenim mjestima uzorkovanja

The data set and statistical procedure

The used data were obtained by measuring the parameters that indicate the quality of the water at the three drain locationsin wider Split area.

The value of pH, total suspended solids (TSS), chemical oxygen demand (COD), biochemical oxygen demand (BOD_5), concentration of chloride (Cl^-), sulphate (SO_4^{2-}), orthophosphates (o-PO_4^{3-}), ammonia (N-NH_4^+), nitrites (N-NO_2^-), nitrates (N-NO_3^-) and total phosphorus (P-tot) were experimentally determined according to standard methods for water and wastewater analysis [6] in Water and Sewer service, Laboratory for Wastewater, Split (Table 1).

Table 1. Selected pollution indicators, units of measurement and symbols applied in the analysis
Tablica 1. Odabrani pokazatelji onečišćenja, mjerne jedinice i simboli korišteni za analizu

Parameter*	Symbol
pH	-
Total suspended solids	TSS
Chemical oxygen demand	COD
Biochemical oxygen demand	BOD_5
Chloride	Cl^-
Sulphate	SO_4^{2-}
Phosphates	$P\text{-}PO_4^{3-}$
Amonium	$N\text{-}NH_4^+$
Nitrites	$N\text{-}NO_2^-$
Nitrates	$N\text{-}NO_3^-$
Detergenti	Detergents
Total phosporus	P- tot

* Units of parameters is mg L^{-3} except pH

Before the statistical analysis of the data set of pollution indicators the pollutions indicators data set, inconsistent data and missing values (outliers) were eliminated. After the exclusion of samples that were not analyzed for all the parameters, 741 water samples remained from location 1, 401 from location 2 and 762 from location 3. Thus, the database contained around 1900 groups with 11 parameters each, from all three locations, for the period between January 2006 and December 2009. Basic statistical analysis and all chemometric analyses were carried out using the Statistical software 10.0 [7].

Monitoring locations and selected pollution indicators are presented in Fig.1 and Table 1. The spatial CA and FA/PCA were performed on the data set standardized through z-scale transformations. The z-scaled variables (z_{ij}) corresponding to the measured raw variables (x_{ij}) are expressed as:

$$z_{i,j} = \frac{x_{i,j} - x_m}{\sigma}$$

where x_m and σ denote the mean and standard deviation, respectively. Standardization tends to minimize the influence of the difference of variance of variables, eliminate the influence of different units of measurement and render the data dimensionless [8].

The similarities/dissimilarities were quantified through Euclidian distance measurement; the distance between two objects (monitoring locations), i and j, is given as:

$$d_{i,j}^2 = \sum_{i=1}^{m} (z_{i,k} - z_{j,k})^2$$

where $d_{i,j}$ denotes the Euclidean distance, $z_{i,k}$ and $z_{j,k}$ are the values of variable *k* for objects *i* and *j*, respectively, and m is the number of variables [9]. For two-way PCA, Statistica 10.0 software was used and the number of significant PCs was chosen by criteria of eigenvalues greater than unity, which explains more variance than a single original variable.

DISSCUSION

Based on the parameters for the three drain locations, descriptive statistical analysis was performed, and the results are presented in Table 2. Median values for all pollutions were lower than their mean values, except for biochemical oxygen demand (BOD_5) and pH value at location 1 and chemical oxygen demand (COD) at locations 2 and 3. Coefficients of variation were quite different, with the highest values for nitrite nitrogen ($N\text{-}NO_2$) at locations 1 and 2, and for detergent concentration at location 3. The highest coefficients of variation at location 1were obtained for TSS, Cl^-, SO_4^{2-}, O-P and $N\text{-}NO_3^-$, at location 2 for TSS and Cl^-, at location 3 for TSS, Cl^-, SO_4^{2-}, $N\text{-}NO_2^-$ and $N\text{-}NO_3^-$. This can be explained by the lack of composite daily wastewater sample, because the sample for analysis was taken dailyand the sewer system collects wastewater and stormwater.

Also, the geographical location of Split and the Mediterranean climate have great influence especially on the concentration of suspended solids. The smallest coefficient of variation, about 4 %, for all locations was for the pH value. The pH value is very important for biological balance in natural systems (e.g. the sea). The pH of the wastewater can disturb the balance of natural systems.the measured pH values were between 6.3 and 8.8, which is in accordance with the regulations (pH interval 6-9). Thus, water within this pH interval should not cause any biological problems in the outlet, bearing in mind buffer characteristics of the sea water, so water with the pH variation of 0.5-1 units of that of the sea water (pH=8.2) should not cause significant changes.

Table 2 shows the proximity between the mean and median values for detergent concentrations at location 3. This can be explained by the household activities since the majority of wastewater at that location originates from households.

Table 2. Statistical analysis of pollution indicators at three locations
Tablica 2. Statistička analiza pokazatelja onečišćenja na tri lokacije

Variable	Descriptive statistics - Location 1										
	Mean	Median	Min	Max	Variance	Std.Dev.	Coef.Var.	Skewness	Std.Err. Skewness	Kurtosis	Std.Err. Kurtosis
pH	7,90	7,99	6,37	8,81	0,11	0,32	4,11	-0,90	0,0898	0,64	0,1794
TSS	265,94	241,00	33,00	3713,00	34262,20	185,10	69,60	9,98	0,0898	168,51	0,1794
COD	435,52	419,82	51,61	1530,72	18688,36	136,71	31,39	1,33	0,0898	6,63	0,1794
BOD 5	252,26	255,43	71,21	672,73	4767,51	69,05	27,37	0,44	0,0898	3,63	0,1794
Cl-	70,11	63,03	6,53	920,80	2304,15	48,00	68,46	10,29	0,0898	158,86	0,1794
SO4=	59,76	56,40	1,64	772,04	1610,12	40,13	67,15	9,08	0,0898	144,48	0,1794
O - P	6,88	6,46	0,61	110,13	18,93	4,35	63,20	18,14	0,0898	429,12	0,1794
N-NH4+	43,78	42,27	3,10	109,26	323,12	17,98	41,06	0,65	0,0898	0,72	0,1794
N-NO2-	0,07	0,06	0,00	2,16	0,01	0,11	159,92	9,55	0,0898	157,03	0,1794
N-NO3-	2,18	1,47	0,00	12,27	3,54	1,88	86,08	1,42	0,0898	1,80	0,1794
Deterg	4,46	4,05	0,23	27,18	5,16	2,27	50,88	2,04	0,0898	13,50	0,1794
P-Tot	11,46	10,48	3,33	29,68	17,91	4,23	36,91	1,29	0,0898	2,32	0,1794

Variable	Descriptive Statistics - Location 2										
	Mean	Median	Min	Max	Variance	Std.Dev.	Coef.Var.	Skewness	Std.Err. Skewness	Kurtosis	Std.Err. Kurtosis
pH	7,72	7,71	6,77	9,21	0,12	0,35	4,53	0,29	0,1219	0,28	0,2431
TSS	258,21	225,00	55,00	2767,00	41434,05	203,55	78,83	7,66	0,1219	81,50	0,2431
COD	428,20	406,80	249,04	887,59	9906,27	99,53	23,24	1,16	0,1219	2,30	0,2431
BOD 5	272,91	277,19	108,09	452,49	2663,54	51,61	18,91	-0,31	0,1219	0,03	0,2431
Cl-	288,34	194,76	6,39	1597,04	64168,52	253,32	87,85	1,90	0,1219	4,25	0,2431
SO4=	61,62	57,67	4,25	357,65	1194,22	34,56	56,08	3,65	0,1219	24,95	0,2431
O - P	7,00	6,94	0,84	18,05	5,81	2,41	34,44	0,45	0,1219	1,20	0,2431
N-NH4+	54,11	51,05	4,78	122,71	513,69	22,66	41,88	0,74	0,1219	0,41	0,2431
N-NO2-	0,05	0,00	0,00	0,78	0,01	0,10	199,02	3,03	0,1219	12,31	0,2431
N-NO3-	3,47	3,21	0,06	11,35	3,39	1,84	53,09	0,82	0,1219	0,95	0,2431
Deterg	5,24	4,95	0,23	19,33	4,40	2,10	40,06	1,94	0,1219	8,29	0,2431
P- Tot	11,54	10,64	1,56	40,86	19,68	4,44	38,43	1,91	0,1219	7,29	0,2431

Variable	Descriptive Statistics - Location 3										
	Mean	Median	Minimum	Maximum	Variance	Std.Dev.	Coef.Var.	Skewness	Std.Err. Skewness	Kurtosis	Std.Err. Kurtosis
pH	7,811	7,860	6,430	8,600	0,081	0,285	3,642	-1,117	0,089	2,211	0,177
TSS	206,063	199,000	23,000	920,000	10610,727	103,008	49,989	2,185	0,089	10,057	0,177
COD	367,036	372,355	87,530	1241,590	14830,679	121,781	33,180	0,721	0,089	2,947	0,177
Cl-	685,608	416,830	48,340	5653,290	41931,608	861,355	125,634	3,107	0,089	11,404	0,177
BOD 5	172,584	170,785	47,800	398,960	1634,766	40,432	23,428	0,371	0,089	1,658	0,177
SO4=	78,974	72,000	6,170	888,800	2898,101	53,834	68,166	6,222	0,089	76,639	0,177
O - P	5,926	5,475	1,000	14,140	5,157	2,271	38,322	0,808	0,089	0,771	0,177
N-NH4+	43,162	41,355	2,530	120,600	350,358	18,718	43,367	0,815	0,089	1,014	0,177
N-NO2-	0,080	0,076	0,001	0,728	0,005	0,071	89,627	1,872	0,089	11,447	0,177
N-NO3-	1,896	1,303	0,018	10,740	2,929	1,711	90,262	1,911	0,089	3,812	0,177
Deterg	10,552	3,705	0,100	5023,000	33063,443	181,834	1723,281	27,598	0,089	761,780	0,177
P- tot	9,833	9,320	1,820	24,520	12,418	3,524	35,838	1,021	0,089	1,379	0,177

Many methods such as FA and PCA require variables to conform to a normal distribution. Prior to multidimensional chemometric analysis, the normality of the distribution of all variables was checked by analyzing the histograms and skewness and kurtosis indices, and by applying the Shapiro-o-Wilk statistical test [7]. At all locations the distribution of six variables was far from normal and six variables were within normal distribution. The original data demonstrated that pH, COD, BOD_5, $N\text{-}NH_4^+$, $N\text{-}NO_3^-$ and P-Tot were almost normally distributed, whereas the other parameters, TSS, Cl^-, SO_4^{2-}, O-P, $N\text{-}NO_2^-$ and detergent concentration, were positively skewed, with kurtosis coefficients significantly greater than zero (95 % confidence). After log transformation of these other parameters, all skewness and kurtosis values were significantly reduced (Fig.2).

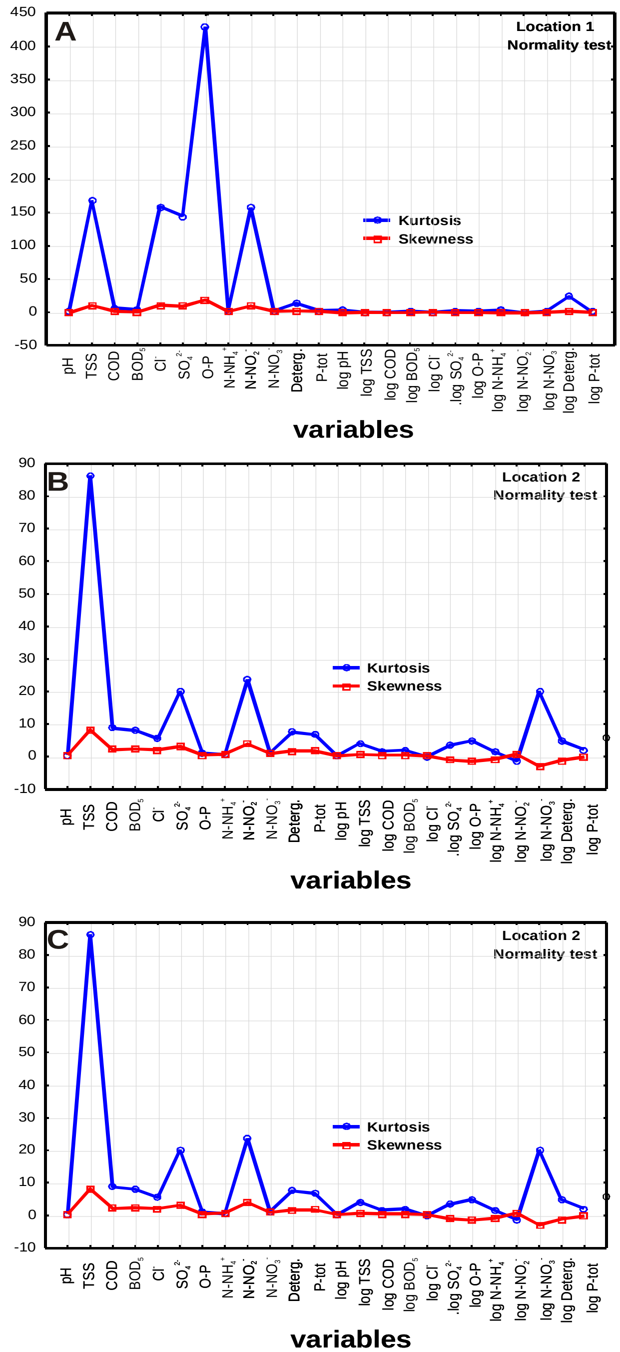

Figure 2. Skewness and Kurtosis coefficients of raw and log transformed data. a) location 1; b) location 2; c) location 3

Slika 2. Koeficijenti zakrivljenosti I spljoštenosti krivulja raspodjele za originalne i log transformirane podatke. a) lokacija 1; b) lokacija 2; c) lokacija 3

The analysis of pollution indicators in municipal wastewater and the applied descriptive statistical analysis gave information on the quality of wastewater. There were some conclusions for the load of the Adriatic Sea, primarily the Brač Channel as the recipients of the wastewater from the largest town on the Croatian side of the Adriatic.

The consequences of that pollution were related to the hydrodynamic conditions which affect the retention time and the degradation of pollutants carried with the wastewater.

Cluster analysis was applied to detect spatial similarity for grouping of sample locations in the monitoring area. It was performed to identify the sewage drains on the basis of similarities/dissimilarities in their wastewater characteristics, in order to group them for proposed wastewater treatment plant. Spatial CA produced a dendrogram with one group from locations 1 and 2 at $(D_{link}/D_{max})\times100<45$ (Fig. 3).

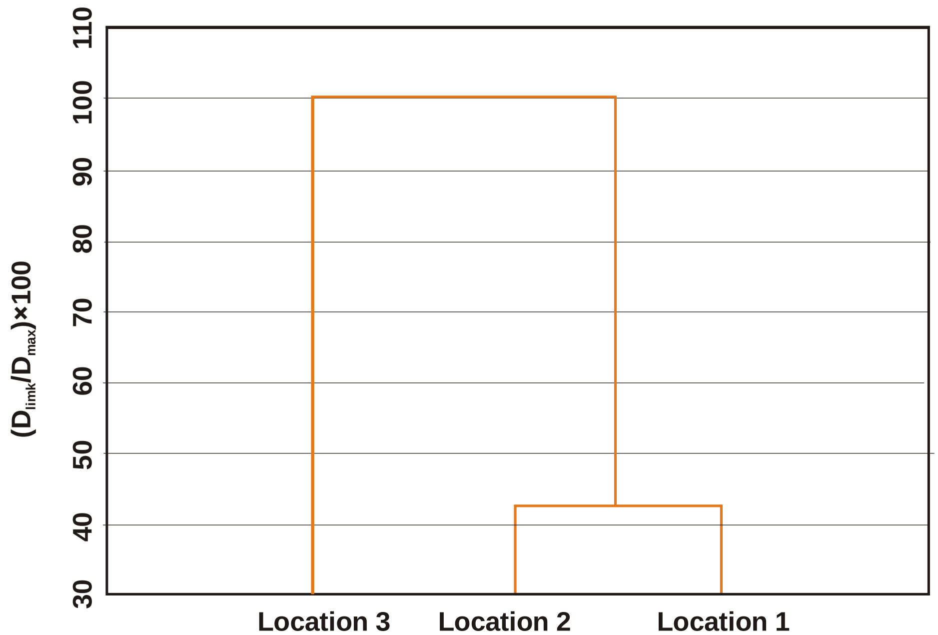

Figure. 3. Dendrogram of the CA according to Ward for spatial clustering of monitoring location
Slika 3. Dendrogram Klaster analize prema Ward-u za prostorno grupiranje prema lokacijama

The dendrogram shows that all monitoring locations may be generally grouped into two main clusters (groups). Cluster I is formed by locations 1 and 2 and cluster II by location 3. The classification between clusters I and II showed significant variation because the sites in these clusters had similar features, natural background and were affected by similar sources. The same technique was used by other workers in the field of environmental monitoring [10, 11]. The drain from the clusters has similar characteristic features and background source type. Locations 1 and 2 collect domestic wastewater, industrial wastewater, and partly storm-water. Location 3 collects mainly domestic wastewater from household activities (about 60%) and storm-water, without industrial wastewater, from the old town with high population density.

FA/PCA was applied to normalized data of six parameters which were normally distributed. It was performed to compare compositional patterns between the analyzed samples and to identify the factors that influence each one. On the basis of screeplot for the PCA, up to 99 % of the normalized data set variability is gathered in the two new variables (components).

All information about six pollution indicators at three drain locations can be presented in reduced space and explained by two calculated variables (components). Based on the fact that the average eignevalue of the autoscaled data is just one, only the PCs with eignevalues greater than one are considered important. The position of variables in the plane 1×2 (Fig. 4) shows that variables COD, BOD_5 and P-tot provide similar information about the pollution of the wastewater. This can be explained by common transport of an organic and mineral polluting load along the collector. Further pollution based on nitrogen components (variables $N\text{-}NO_3^-$ and $N\text{-}NH_4^+$) provides similar information.

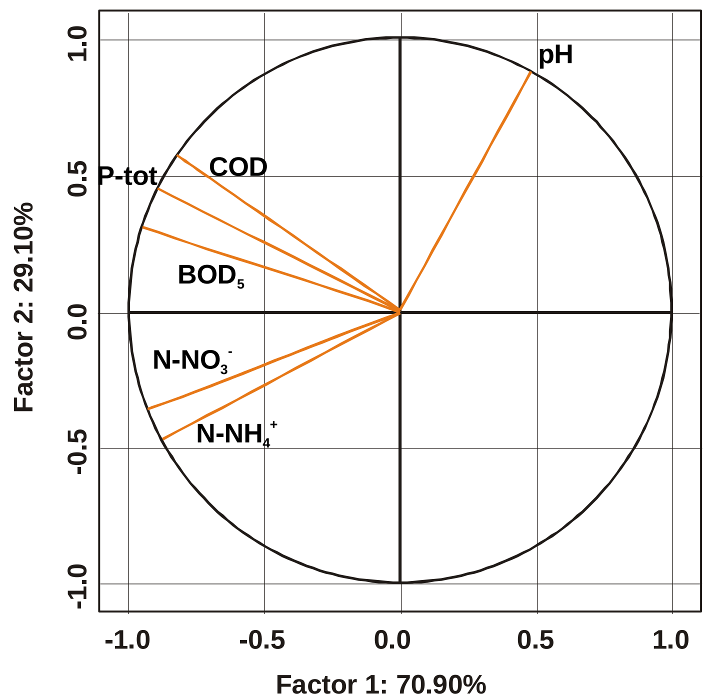

Figure 4. Factorial distribution of variable space on the plan 1×2
Slika 4. Raspodjela varijabli u faktorskom planu 1×2

The cumulative variance for two PCs was 99% and the results of PCA analysis are in good agreement with CA results. The monitoring locations clustered previously in groups I and II can now be discriminated (Fig. 5).

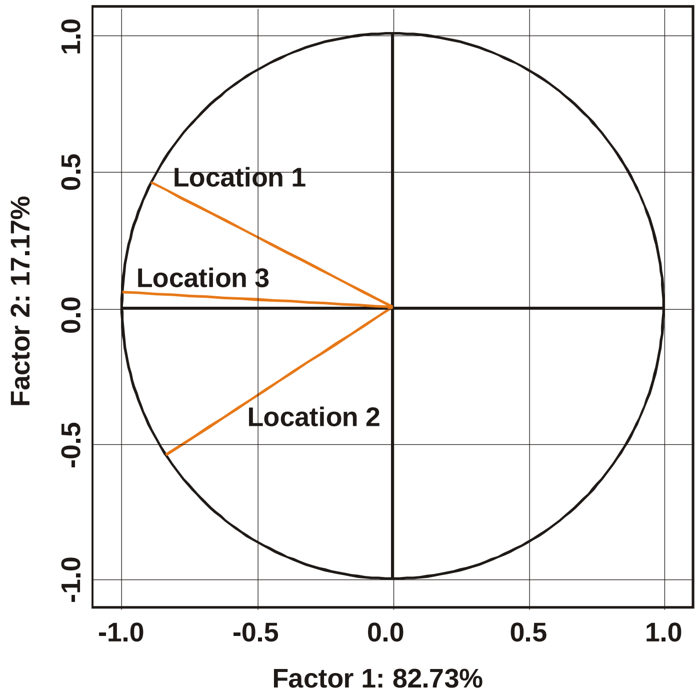

Figure 5. Factorial distribution of locations on the plan 1×2
Slika 5. Raspodjela lokacija u faktoskom planu 1×2

CONCLUSION

Tourism is a major economic factor in the region of Dalmatia, so the protection of the sea as a natural recipient has great importance. Monitoring and supervision of the quality of wastewater is particularly expressed at the drain locations in Split. Routine monitoring programs of the wastewater quality generate complex multidimensional data that need multivariate statistical techniques for their analysis and interpretation of the underlying information.

A multivariate statistical technique for evaluation and interpretation of large complex data set presents a necessity and usefulness for effective management of water resources. Normality test of original data demonstrated that pH, COD, BOD_5, N-NH_4^+, N-NO_3^- and P-tot were almost normally distributed.CA showed that monitoring locations may be generally grouped into two main clusters (groups). The first group consisted of locations 1 and 2, while location 3 is individual.

By applying FA, it was possible to identify the main sourcesof pollution at the three locations.

PCA allowed the reduction of six variables to two PCs to explain 99 % variance of the data set and confirmed dissimilarities between locations.

REFERENCES

[1] Metcalf and Eddy, Wastewater engineering: Treatment, Disposal, Reuse, McGraw-Hill Inc., New York, London, Hamburg, 1991.

[2] Curis W. Bryant, Wat. Sci. and Technol., 31 (1995) (12) 211-218.

[3] L. Mutihac, R. Mutihac Anal. Chim.Acta 612 (2008) 1-18

[4] U. C. Panda, S. K. Sundaray, P. Rath, B. B. Nayak, D. Bhatta, J. Hidrol. 331 (2006) 434-445

[5] B. Helena, R. Pardo, M. Vega, E. Barrado, J. M. Fernandez, L. Fernandez, Water Res. 34 (2000) 807-816

[6] APHA, Standard Methods for the examination of Water and Wastewater, 18the. American Public Health Association, Washington, D. C. 1992.

[7] StatSoft, Inc. (2010) STATISTICA (data analysis software system) ver. 10,0 http://www.statsoft.com

[8] K.P. Singh, A. Malik, S. Simba, Anal. Chim. Acta 538 (2005) 355-374.

[9] J. Zupan, Kemometrija in obdelava eksperimentalni podatkov, Kemijski inštitut Ljubljana, Inštitutno verevije, Ljubljana, 2009.

[10] D. Kavšek, D. Brodnjak Vončina, Nova biotechnologica 9 (2009) 3, 265-270.

[11] F. Kraic, Z. Kiralyova, J. Mocak, Nova biotechnologica 9 (2009) 2, 217-224.