MULTIVARIATE STATISTICAL METHODS FOR CHARACTERIZATION OF WASTE WATER QUALITY

DARJA KAVŠEK[1], DARINKA BRODNJAK VONČINA[2]

[1]*Regional Technological Centre Zasavje, Chemical-Technological Laboratory, Nasipi 48, 1420 Trbovlje, Slovenia (darjakavsek@gmail.com)*
[2]*Faculty of Chemistry and Chemical Engineering, University of Maribor, Smetanova 17, 2000 Maribor, Slovenia (darinka.brodnjak@uni-mb.si)*

Abstract: The aim of this work is focused on water quality classification of the waste waters and evaluation of pollution by the monitoring measurements during period 2006-2008. Environmental monitoring was performed in the region of Trbovlje, Slovenia, with two sampling sites and 15 chemical and physicochemical water quality parameters (pH, temperature, suspended solids, settling matter, chemical oxygen demand, biochemical oxygen demand, AOX (adsorbable organic halogens), total phosphorus, ammonium, nitrite, sulphate, chloride, fluoride, sulphide and mineral oil content) monitored in monthly periods (total of 60 objects x 15 variables). For handling the results different chemometric methods were employed, such as basic statistical methods for the determination of mean and median values, standard deviations, minimal and maximal values of measured parameters and their mutual correlation coefficients, the principal component analysis (PCA), cluster analysis (CA), and linear discriminant analysis (LDA). Monitoring of general pollution of waste waters and following measuring parameters which are above permitted concentration level can be used for searching of pollution source and for planning prevention measures from pollution, as well. The study allows drawing new information from the data sets such as patterns of similarity between sampling locations, sources of pollution in the environment, seasonal behavior of chemical contents and time trends.

Key words: waste waters, water quality, chemometrics, principal component analysis, classification.

1. Introduction

The experimental data set, carried out throughout the years 2006-2008, was composed of analytical parameters from 2 waste water sources, first is the cave water from Savski rov and the second is the industrial wastewater. Through this period the quality of the water was followed and the classification has been made according to sampling sites. From 28 variables 15 variables, carrying the most useful information, were selected and investigated. The aim of this work is to find the correlation between sampling sites and the variables obtained by chemical measurements, which can be used to construct a fast decision model for separating different waste water quality samples.

Chemometric methods have been often used for the classification and comparison of different water samples (SIMEONOVA and SIMEONOV, 2007; MASSART *et al.,* 1997). Some examples are, for instance, the differentiation of water sources in a vast southwest area of Paris by principal component analysis (PCA) (DE LUCA *et al.,* 2008), application of chemometric techniques to the analysis of Suquia River water quality (ALBERTO *et al.,* 2001), identification of sources of bottom waters in the Weddel Sea by PCA and target estimation (LINDEGREN and JOSEFSON, 1998),

determination of correlation of chemical and sensory data in drinking waters by factor analysis (MENG *et al.,* 1997), to name just a few. Chemometric methods have been used also for evaluating environmental data of Lagoon water (CARRER and LEARDI, 2006), San Francisco Bay and Estuary (JARMAN *et al.,* 1997), and Muggia Bay in Northern Adriatic Sea (BARBIERI *et al.,* 1998).

The quality of the waste waters was studied through the years 2006-2008. Altogether 15 preselected characteristic features were measured for 60 samples collected and analysed during this period. Several chemometric methods were applied in order to visualize multivariate data and to enable a quick classification of samples, regarding the source location within the studied time period.

2. Materials and methods

A standard method was used for sampling (ISO 5667-01:1996 (E)). Water was collected in polyethylene bottles 0.5 m below the surface. All glass and plastic ware used for sampling and analyses were rinsed with milli-Q water.

Standard analytical methods (WATER QUALITY, 1980, 1996, 1998, 2000, 2002, 2004, 2005) were used for the determination of 15 physico-chemical variables**.** All reagents were analytical grade. The milli-Q system was used for purifying the water.

The 60 samples are characterized by 15 physico-chemical variables: (1) pH, (2) water temperature, (3) suspended solids, (4) settling matter, (5) chemical oxygen demand (6) biochemical oxygen demand, (7) adsorbable organic halogens (AOX), (8) total phosphorus, (9) ammonium, (10) nitrite, (11) sulphate, (12) chloride, (13) fluoride, (14) sulphide, and (15) mineral oil content. The results of all measurements have been investigated by different chemometric methods (MASSART *et al.,* 1997): the basic statistical methods for the determination of mean and median values, standard deviations, minimal and maximal values of measured variables and their mutual correlation coefficients. The PCA was applied for grouping of water samples due to measured variables. All the calculations and plots in the following (PCA) section were done with the Teach/Me software (TEACH/ME DATALAB 2.002, 1999) using Teach/Me Data Analysis option.

3. Results and discussion

3.1 Statistical screening of data

First the mean and median values and standard deviation were determined. After excluding 2 samples, which were discovered as outliers, namely, first sample with very high content of sulphate (more ten times higher than in all other samples, but still under the legislation permitted level) and the second one with high content of suspended solids. The mutual correlation was sought for all measured variables. The maximal correlation coefficient of the data was found between measurements of sulphate content and pH (r = -0.82) and between suspended solids and settling matter (r = 0.77). The correlation between suspended solids and settling matter is expected to be high. The negative correlation between sulphate content and pH shows that samples from two sampling sites are mainly differentiated according to these two parameters.
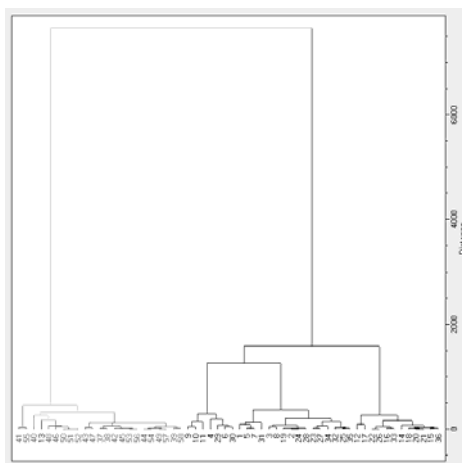
Fig. 1. Dendrogram of 58 samples at 2 different sampling sites (classes).

Cluster analysis resulted in a dendrogram shown in Fig. 1, where all 58 samples are divided into a number of clusters, depending on the level of similarity. Clustering is based on Ward distance. First group of samples, namely sampling site "Savski rov" (the right-most cluster) is well distinguished from second sampling site. Clusters of samples from sampling sites 1 and 2 are well defined. According to the content of pollution parameters from the previous mentioned sampling sites it can be concluded that sample 13 from sampling site 1 is different from all others and is clustered in cluster 2.

3.2 Principal Component Analysis (PCA)

PCA was performed in order to get an overall impression about the correlation of 58 water samples, described with physical and chemical variables, with the quality of water in different sampling sites. PCA was applied on the matrix composed of 58x15 elements. 58 rows represent waste water samples composed of 15 variables. Data were additionally pre-processed in two different ways. First the "Column centering" of the data was used and second, the autoscaling of individual variables was performed, called "Column standardization".

The PCA with column standardized data were further analysed for formed clusters. It was found from the score plots of the first and the second PCs that samples are well separated according to sampling sites. Clusters of samples from sampling sites 1 and 2 are well defined.

The scores and loadings plots of PCA of the waste water samples represented with 15 variables are shown in Figure 2. It is evident from Figure 2 that samples from two sampling sites are well separated. The inspection of particular parameters shows that the content of sulphate is much higher for all samples from sampling site 1, while pH for all samples from sampling site 1 is always lower than for the sampling site 2 (Fig. 4).
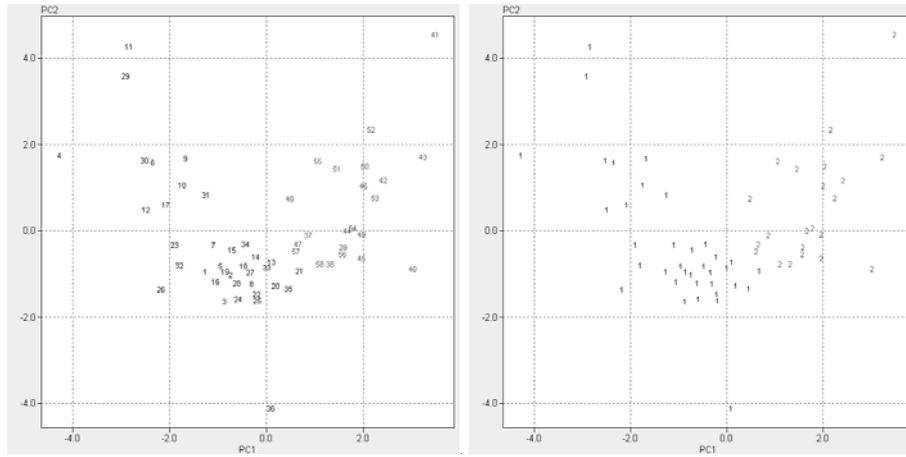
Fig. 2. PCA for all waste water samples from 2 different sampling sites denoted by class numbers 1 and 2.
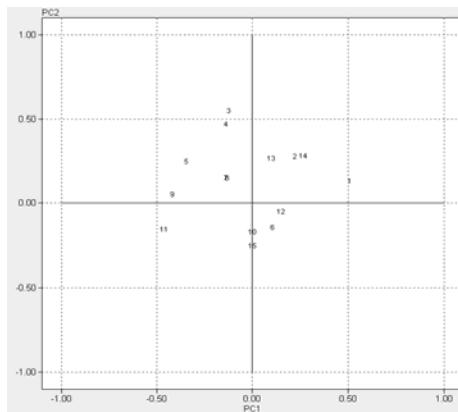

Fig. 3. PCA for 58 waste water samples from 2 different sampling sites; loadings in 15 PC axes are shown.
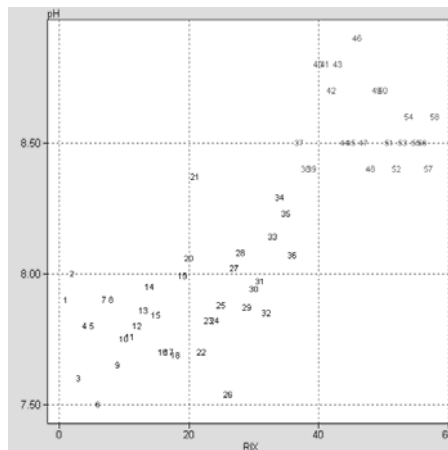

Fig. 4. pH value for all samples from two different sampling sites.

Separation of the two classes is obtained by PC1 (labels 1, 11 and 9, see Fig. 3) which is accounted for the pH value and content of ammonium and sulphate, while the second principal component PC2 is accounted for suspended solids, settling matter and mineral oil content (labels 3, 4 and 15, see Fig. 3).

3.3 Linear discriminant analysis (LDA)

Linear discriminant analysis confirms the predetermined classes. It was found from LDA that samples are well separated according to sampling sites (Fig. 5).
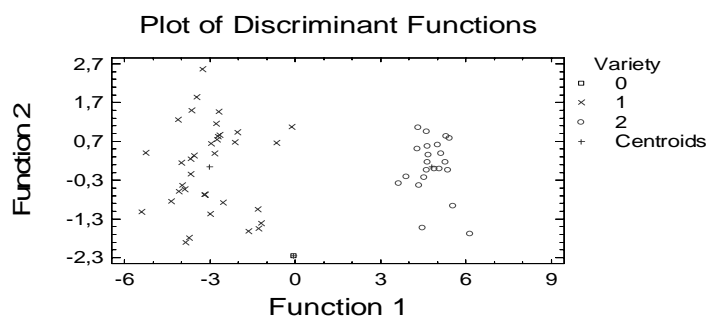

Fig. 5. Discriminant function for waste waters.

4. Conclusions

The study gives the opportunity to follow quality of waste waters at different sampling sites. Monitoring of general pollution of waste waters and following measuring parameters which are above permitted concentration level can be used for searching of pollution source, for planning of prevention action and for the protection from pollution.

References

ALBERTO, W.D., DEL PILAR, D.M., VALERIA, A.M., FABIANA, P.S., CECILIA, H.A., DE LOS ANGELES, B.M.: Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality, A case study: Suquia River basin (Cordoba-Argentina). Water Res., 35, 2001, 2881-2894.

BARBIERI, P., ADAMI, G., FAVRETTO, A., REISENHOFER, E., A chemometric survey of three sites in Muggia Bay (Northern Adriatic Sea): meteorological effects on heavy metal patterns in surface coastal waters. Fresenius J. Anal. Chem., 361, 1998**,** 349-352.

CARRER, S., LEARDI, R.: Characterizing the pollution produced by an industrial area: chemometric methods applied to the Lagoon of Venice. Sci. Total Environ., 370, 2006, 99-116.

DE LUCA, M., OLIVERIO, F., IOELE, D., HUSSON, G. P., RAGNO, G.: Monitoring of water quality in South Paris district by clustering and SIMCA classification. Int. J. Environ. Anal. Chem., 88, 2008, 1087-1105.

Deutsche einheisverfahren zur wasser-, abwasser- und schlammuntersuchung bestimmung des volumenanteils der absetzbaren stoffe im wasser und abwasser (H9), DIN 38409-H9-2:1980.

Deutsche einheitsverfahren zur wasser-, abwasser- und schlammuntersuchung physikalische unf physikalisch-chemische bestimmung der temperatur (C4), DIN 38404:C4:2000.

JARMAN, W.M., JOHNSON, G.W., BACON, C.E., DAVIS, J.A., RISEBROUGH, R.W., RAMER, R.: Levels and patterns of polychlorinated biphenyls in water collected from the San Francisco Bay and Estuary, 1993-1995. Fresenius J. Anal. Chem., 359, 1997**,** 254-260.

MASSART, D.L., VANDEGINSTE, B.G.M., BUYDENS, L.M.C., DE JONG, S. LEWI, P.J., VERBEKE, J.S.: Handbook of Chemometrics and Qualimetrics: Part A, Elsevier, Amsterdam, 1997.

MENG, A.K., SUFFET, I.H.: A procedure for correlation of chemical and sensory data in drinking water samples by principal component factor analysis. Environ, Sci. Technol., 31, 1997*,* 337-345.

SIMEONOVA, P., SIMEONOV, V.: Chemometrics to evaluate the quality of water sources for human consumption. Microchim. Acta, 156, 2007, 315-320.

TEACH/ME, SDL - SOFTWARE DEVELOPMENT LOHNINGER; TEACH/ME DATALAB 2.002, 1999**,** Springer, Berlin, Developed by H. Lohninger and the Teach/Me people.

SOIL QUALITY- Determination of oil content- Method by infrared spectrometry and gas chromatographic method, ISO TR 11046 (E).

WATER QUALITY-SAMPLING- PART 10: Guidance on sampling of waste water ISO 5667-01: 1996 (E).

WATER QUALITY- Determination of pH, ISO 10523:1996 (E).

WATER QUALITY- Determination of suspended solids by filtration through glass-fibre filters, ISO 11923:1998 (E).

WATER QUALITY- Determination of chemical oxygen demand, ISO 6060:1996 (E).

WATER QUALITY- Determination of biochemical oxygen demand after n days (BODn)- Part 2: Method for undiluted samples, ISO 5815-2:2002 (E).

WATER QUALITY- Determination of adsorbable organically bound halogens (AOX), ISO 9562:2005 (E).

WATER QUALITY- Determination of phosphorus- Ammonium molybdate spectrometric method, ISO 6878:2004 (E).

WATER QUALITY- Determination of ammonium- Distillation and titration method, ISO 5664:1996 (E).

WATER QUALITY- Determination of dissolved anions by liquid chromatography of ions- Part 2: Determination of bromide, chloride, nitrate, nitrite, orthophosphate and sulfate in waste water, ISO 10304-2:1998(E).