

DETERMINATION AND CLASSIFICATION OF POLLUTANTS IN WASTE WATER

FILIP KRAIC¹, ZUZANA KIRÁLYOVÁ², JÁN MOCÁK^{1,2}

¹*Department of Chemistry, University of SS. Cyril & Methodius in Trnava, J. Herdu 2, Trnava, SK-917 01, Slovak Republic (filip.kraic@ucm.sk)*

²*Institute of Analytical Chemistry, Slovak University of Technology, Bratislava, SK-81237, Slovak Republic*

Abstract: This paper deals with waste waters produced by industrial producers during recent three years. Its main purpose is to evaluate the data monitored from discharges of three leather plants where eight traditional variables (COD, BOD, insoluble matters, pH, and the concentrations of ammonia, total nitrogen, chromium and sulphides) were regularly analyzed and quantified.

Chemometrical and statistical methods were approved as very useful tools for characterization and classification of various kinds of water samples considering the environmental and metrological aspects. For this purpose, multivariate (multidimensional) techniques of data analysis and correlation analysis were especially very useful. The mentioned techniques are used in this work to (1) reveal the concealed interrelations of the monitored characteristics of waste waters, (2) compare the results of individual waste water producers and find the most important polluting source during a long time period, and (3) derive some generally valid conclusions regarding the observed results.

Key words: waste water, pollutants, multivariate data analysis

1. Introduction

Pollution of the rivers and seas jeopardizes production of oxygen in water. A severe pollution source is the water waste pollution. Waste water is coming into environment as a consequence of the industrial and agricultural production and the town activities. It has an unhealthy influence on human and environment. Contaminations present in waste water contain various inorganic and organic matters. The most important variables, which characterize the water quality and indicate the extent of pollution are the concentrations of chromium (coded Cr in further text), total nitrogen (coded Ntot), ammonium (NH₄), and sulphide (S₂). Together with chemical oxygen demand, COD, (coded CHSK according to the Slovak designation), biological oxygen demand, BOD₅, (coded BSK₅ according to the Slovak designation), water acidity (pH), and amount of insoluble matters (IM) they were regularly monitored at the output from three industrial sites in the same Slovak town during the years 2006, 2007 and (partly) 2008, which has created the measurement basis for the performed research.

Due to their potential environmental danger, discharges from the leather plants have to be regularly monitored. The most dangerous environmental factor is here the content of chromium, which is very toxic in the oxidation state +VI, therefore most part of Cr(VI) is trapped in the technological process. Also further above mentioned factors must be regularly monitored not only by the plant laboratory but also by an independent accredited analytical laboratory. When the limiting values are exceeded, the level of the respective factor must be immediately properly adjusted.

Characterization and classification of various kinds of water samples from the environmental and metrological aspects was effectively performed using various chemometrical and statistical methods (VONČINA *et al.*, 2007; ŠNUDERL *et al.*, 2007; KANNEL *et al.*, 2007; KRAIC *et al.*, 2008). Therefore in this work several multidimensional (multivariate) techniques of data analysis were used.

2. Material and methods

2.1 Sampling

The waste water samples originated from three relatively similar industrial production sites. The sampling was done once a week by the analysts from a nearby accredited laboratory where the required measurements were made. The following 8 variables were monitored and analyzed: CHSK, BSK, pH, IM, N_{tot}, NH₄, Cr and S₂ (their codes are explained above). The sampling procedures and laboratory analyses were described in detail in our previous papers (KIRÁLYOVÁ *et al.*, 2008a; 2008b). To prevent changes in waste water samples, the analyses started two hours after sampling, which was facilitated by closeness of the analytical laboratory to the monitoring sites.

2.2 Analysis

COD (CHSK) was measured by the spectrophotometrical dichromate method; BOD₅ (BSK₅) was determined after treatment of analyzed water with manganese dioxide by iodometric thiosulphate determination of the present Mn(IV). The pH value was measured potentiometrically using a glass electrode. Insoluble matters (IM), separated on an appropriate filter and filtered off, were dried at 105 °C until the constant weight and determined gravimetrically. Ammonia was distilled off from the analyzed water and determined spectrophotometrically using the Nessler reagent. For the total nitrogen determination, inorganic and organic nitrogen was oxidized by persulphate and mineralized in sulphuric acid. The formed nitrate reacted with 2,6-dimethylphenol (in sulphuric and phosphoric acid medium) and the resulted 4-nitro-2,6-dimethylphenol was determined spectrophotometrically. Chromium was selectively determined by inductively coupled plasma optical emission spectrometry. Sulphide anion was determined by back iodometric titration using thiosulphate standard solution.

2.3 Multidimensional data analysis

Statistical calculations were performed using the following techniques: correlation analysis, principal component analysis (PCA), cluster analysis (CA) and linear discriminant analysis (LDA). In the performed calculations three contemporary software commercial packages were used: STAGRAPHICS Plus 5.1, SPSS 15.0 and SAS JMP 7.0.

3. Results and discussion

3.1 Correlation analysis

The output of correlation analysis is the correlation table, which contains the pair (or Pearson) correlation coefficients expressing the strength of correlation between all possible pairs of variables. Tab. 1 shows the calculated correlation table. The entries of this table are symmetrical according to diagonal. The following conclusions can be derived from the correlation table: (a) The highest correlation is between CHSK and BSK5. (b) Very high correlations exist between CHSK and Ntot, BSK5 and Ntot, as well as CHSK and IM. (c) Considerably significant correlations ($r_{crit} \geq 0.258$ at $p \leq 0.01$) exist between IM and Ntot, BSK5 and IM, IM and pH, as well as NH4 and S2. (d) Significant correlations (at the 95 % or higher probability level, $p \leq 0.05$) exist between BSK5 and pH, as well as IM and NH4 (where both dependences are inverse). All correlation coefficients in Table 1, which are larger or equal than $r_{crit} = 0.184$ are marked by bold faces. (e) No significant correlation was proved in all other pairs.

Table 1. Pearson correlation coefficients exhibiting the strength of correlation between individual pairs of variables

Variable	CHSK	BSK5	IM	pH	NH4	Ntot	Cr	S2
CHSK	1							
BSK5	0.7606	1						
IM	0.5255	0.3195	1					
pH	0.0025	-0.2352	0.2890	1				
NH4	0.0058	0.0389	-0.1923	0.0786	1			
Ntot	0.6386	0.5766	0.4338	0.1337	0.0284	1		
Cr	-0.0473	-0.0326	0.1056	-0.0942	-0.0982	-0.0052	1	
S2	0.0986	0.0329	0.1808	-0.1102	-0.2694	0.0155	0.0973	1

Note: 81 studied objects, i.e. 27 averaged month values were measured at the discharge of three leather plants. Critical values of the correlation coefficient (absolute values): $r_{crit} = 0.144$ ($p = 0.1$), $r_{crit} = 0.184$ ($p = 0.05$), and $r_{crit} = 0.258$ ($p = 0.01$) for $n = 81$. Significant correlations are marked bold.

3.2 Principal component analysis

In principal component analysis, PCA, some natural grouping of the objects (the waste water samples - in this work) and the studied variables might be seen. The principal components, PCs, are calculated as the linear combinations of original variables (SHARMA 1996; KHATTREE and NAIK 2000). According to the computed eigenvalues only three principal components (PCs) were found important; their value was larger than 1, which is usually considered as the criterion. Three kinds of graphical outputs are used in the PCA, namely the scatterplot showing the objects, the loadings plot showing the variables, and the biplot where all objects and variables are depicted together. The advantage of first two outputs is the possibility to obtain not only the 2D graph but also the 3D one, where usually the first two and three most important PCs are used as the axes, respectively. On the other hand, the biplot, even though plotted in two dimensions, provides more complex information about the studied problem.

Fig. 1 exhibits the biplot, which simultaneously represents the samples, depicted here by the numbers, and eight originally utilized chemical descriptors, depicted by the rays starting from the origin and ending at the point determining the variable position. The samples are here categorized according to the sampling site located at the discharge of the plants 1, 2 and 3, respectively.

From the position of variables in the plane PC2 – PC1 the following outcomes can be deduced: (a) Variables CHSK (COD), BSK5 (BOD5) and Ntot provide similar information about the sampled waste water. This is well understandable especially for the first two variables since they are highly correlated; the possibility to convert mutually their values has already been discovered. (b) Similarity (interdependence) of NH4 and pH proves the role of ammonia in changing the pH value of waste water. (c) The opposite position of Cr with respect to pH testifies that they are antagonists; a high level of chromium is reached at lower pH, and vice versa. The achieved PCA results are in accord with the described outputs of cluster analysis introduced in further text.

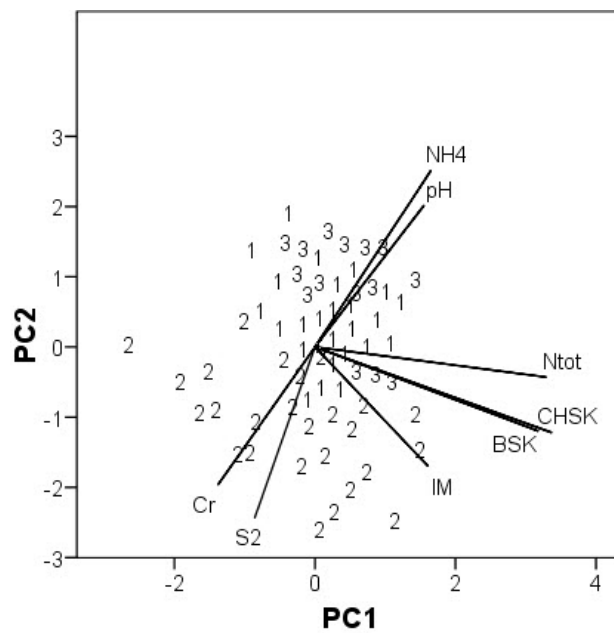


Fig. 1. Biplot PC2 vs. PC1 for 8 measured chemical descriptors and 81 waste water samples coming out from the discharges of three leather plants (coded by numbers 1, 2 and 3). Software SPSS 15.0.

When looking at the sample positions in Fig. 1, it can be concluded that high sulphide and Cr concentration values are characteristic for the waste coming from the second plant. On the other side, high pH and NH4 values are characteristic mainly for the third and partly for the first plant. Since the samples belonging to the plant 2 are localized at low (mostly negative) values of CHSK, BSK and Ntot, it can be concluded that samples from plant 2 exhibit generally low values of these variables. The

occurrence of negative variable values in the PCA is caused by the performed variable standardization, in which the corresponding mean is subtracted from the original variable values and the result is divided by the corresponding standard deviation (e.g. the zero value is achieved for the original mean value of the variable).

3.3 Cluster analysis

Generally, the clustering process in cluster analysis may be performed either with objects or variables (KHATTREE and NAIK 2000). Clustering in this work was made concerns the studied eight variables. The result of the performed cluster analysis is a dendrogram depicted in Fig. 2. The basis for the performed calculations were data on 81 objects representing 27 average month data measured at 3 sampling sites (discharges of three leather production plants) for each of eight investigated variables. In these calculations Ward's method of clustering and squared Euclidean distance were used.

In Fig. 2 three clusters of closely related variables can be seen. The first cluster is formed by CHSK, BSK, connected further to Ntot and IM. The second cluster is formed by pH and NH₄, and the third cluster contains Cr and S₂. The variables forming the same cluster are most similar; the measure of their mutual similarity is given by Distance, which represents the vertical axis of the dendrogram. The results of cluster analysis are in a good agreement with the outputs of correlation analysis and principal component analysis.

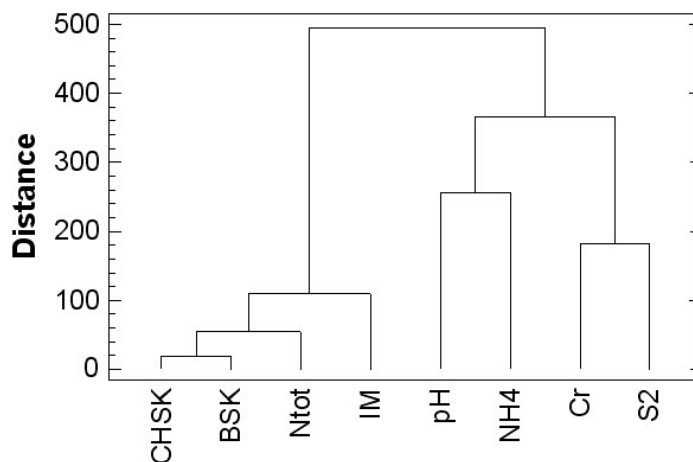


Fig. 2. Cluster analysis of 8 variables measured in waste water analysis of 81 samples obtained during 27 months (average values at 3 sampling sites). Software Statgraphics Plus 5.1. CHSK and BSK denote COD and BOD₅, respectively.

3.4 Linear discriminant analysis

Linear discriminant analysis, LDA, is a supervised learning method, in which the classification model is calculated using the data belonging to the training set where the

categorization of all samples is known before the calculation starts. In the first step (training), the classification model is calculated, by which the studied objects (waste water samples in this case) are re-categorized so that the object (sample) category is either confirmed or the object is assigned to another category (SHARMA, 1996; VANDEGINSTE *et al.*, 1998; KHATTREE and NAIK, 2000). The success in classification is calculated by the ratio of the correctly classified objects to their total number. Then, in the second step, the developed model is used for classification of the objects, which are not included into the training set. They belong either to the test data set, which is utilized for validation of the discriminant model, or the category of the investigated data is completely unknown so that the main goal of LDA is prediction of the object category.

Table 2. Evaluation of linear discriminant analysis classification of the waste water samples divided into categories given by the sampling location (plants 1, 2 and 3).

Evaluated data	Success by	Plant No.	Predicted group membership			Total
			1	2	3	
Training set	Count	1	17	0	10	27
		2	0	27	0	27
		3	4	0	23	27
	%	1	63.0^b	0	37.0	100.0
		2	0	100.0^b	0	100.0
		3	14.8	0	85.2^b	100.0
Cross-validated ^a	Count	1	14	0	13	27
		2	0	26	1	27
		3	6	0	21	27
	%	1	51.9^c	0	48.1	100.0
		2	0	96.3^c	3.7	100.0
		3	22.2	0	77.8^c	100.0

^a In leave-one-out cross validation, each object was classified by the model derived from all objects other than that one.

^b In total **82.7 %** of original grouped cases were correctly classified.

^c In total **75.3%** of cross-validated grouped cases were correctly classified.

In this work, the samples were separated into three categories, assigned by the plant releasing waste water. The classification results are summarized in Table 2. The LDA results testify that the waste water from plants 1 and 3 have a similar level of pollutants therefore the categorization of the samples from these two sources is often interchanged, e.g. 10 samples belonging to plant 3 were classified into group 1 (plant 1), as shown in the first line of the table. On the other hand, 100 % of the samples from plant 2 were classified correctly when the training set is considered and a 96.3 % success was obtained when performing leave-one-out validation. According to the position of the samples belonging to three plant discharges with respect to the first discriminant function expressing the overall pollutant concentration in the LDA

diagram (not shown here) it was proved that the worst pollutant is plant 1, the best cleaning procedures were applied in plant 2.

An attempt was also made to see the seasonal effect on the waste water pollution in the LDA output. Therefore all data were divided into four categories according to the year seasons: Spring (March, April, May), Summer (June, July, August), Autumn (September, October, November) and Winter (December, January, February). The nominal categorical variable *Season* was created with 4 classes Spring, Summer, Autumn and Winter and used as the discrimination criterion. The separation of the data by four seasons was not well demonstrated. However, the Winter and Autumn data were displayed at higher values of the first discriminant function, DF1, which reflects the extent of pollution since DF1 may be explained as the total concentration of the polluting substances. Much more successful was seasonal categorization of the data into two parts: the warmer part of the year (Spring and Summer) and the colder part (Autumn and Winter). In this case the success of discrimination was 76.5 % for the training set and 66.7 % for the leave-one-out validation. In both cases the discrimination results are unambiguously significant. At the same time it means that in cold months the total level of pollution of the waste water from all three investigated plants is larger than in the year season when the temperature is higher.

4. Conclusions

The waste water characteristic parameters, namely COD (CHSK), BOD (BSK5), insoluble matters, pH, ammonium, total nitrogen, chromium, and sulphides were monitored at the discharge of three industrial plants of the same town during almost three years. The obtained analytical values were statistically evaluated and mutual relations and the trends of the individual waste water descriptors were discovered. In total, the most important source of pollution was the industrial plant number 1. With respect to the year seasons, more polluted waste waters are expected in cold months.

Acknowledgement: The support of this work by Slovak Grant Agencies by means of grants VEGA 1/1005/09, VVCE-0004-07 and APVV-0057-06 is highly acknowledged.

References

- KANNEL, P.R., LEE, S., KANEL, S.R., KHAN, S.P.: Chemometric application in classification and assessment of monitoring locations of an urban river system. *Anal. Chim. Acta*, 582, 2007, 390-399.
- KHATTREE, R., NAIK, D.N.: *Multivariate data reduction and discrimination*. SAS Institute, Cary, North Carolina, USA, 2000.
- KIRÁLYOVÁ, Z., KRAIC, F., MOCÁK, J.: Study of Pollutants in Waste Water. In: 28. Proceedings of Industrial Toxicology 2008. Tatranská Štrba, Slovakia, June 18-20, 284-289.
- KIRÁLYOVÁ, Z., ŠNUDERL, K., KRAIC, F., MOCÁK, J.: Determination and classification of pollutants in waste water. *Acta Chim. Slovaca*, 1, 2008, 143-152.

- KRAIC, F., MOCÁK, J., FIKET, Ž., KNIEWALD, G.: ICP MS analysis and classification of potable, spring, and mineral waters. *Chem. Pap.*, 62, 2008, 445–450.
- SHARMA, S.: *Applied Multivariate Techniques*. Wiley, New York, 1996.
- ŠNUDERL, K., SIMONIČ, M., MOCÁK, J., BRODNJAK-VONČINA, D.: Multivariate data analysis of natural mineral waters. *Acta Chim. Slov.*, 54, 2007, 33-39.
- VANDEGINSTE, B.G.M., MASSART, D.L., BUYDENS, L.M.C., DE JONG, S., LEWI, P.J., SMEYERS-VERBEKE, J.: *Handbook of Chemometrics and Qualimetrics: Part B*, Elsevier, Amsterdam, 1998.
- VONČINA, E., BRODNJAK-VONČINA, D., MIRKOVIČ, N., NOVIČ, M.: Chemometric Characterisation of the Quality of Ground Waters from Different Wells in Slovenia. *Acta Chim. Slov.*, 54, 2007, 119-125.