Acta Chimica Slovaca, Vol.1, No. 1, 2008, 143 – 152

Determination and classification of pollutants in waste water

Zuzana Királyová[a], Katja Šnuderl[b], Filip Kraic[c], Jan Mocak[a,c,]*

[a] *Institute of Analytical Chemistry, Slovak University of Technology, SK-81237 Bratislava, Slovakia*
[b] Faculty *of Chemistry and Chemical Engineering, University of Maribor, SI-2000 Slovenia*
[c] *Department of Chemistry, University of Ss. Cyril & Methodius, SK-91701 Trnava, Slovakia*

**jan.mocak@stuba.sk*

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 content of ammonia, total nitrogen, chromium and sulphides) were regularly analyzed and quantified. Mutual relations of these variables in waste waters were discovered using statistical techniques, mainly multivariate data analysis, and some general conclusions were found regarding the trends of pollution with respect to its source as well as the year season.

Keywords: waste water, pollutants, multivariate data analysis

Introduction

Water Pollution

Pollution of the rivers and seas jeopardizes production of oxygen in water. A severe pollution source is 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, (code CHSK according to the Slovak designation), biological oxygen demand, BOD5, (code 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.

Chemometrical and statistical methods were approved as very useful for characterization and classification of various kinds of water samples from the environmental and metrological aspects (Vončina et al. 2007, Šnuderl et al. 2007, Kannel et al. 2007, Kraic et al. 2008). For this purpose especially multivariate (multidimensional) techniques of data analysis, correlation analysis and/or ANOVA were very useful. Therefore in this work the mentioned techniques are used to (1) reveal the 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.

Experimental

Sampling

The samples of waste waters from three similar industrial production sites were sampled once a week by the analysts from an accredited laboratory where the required measurement conditions were created for obtaining correct results. Eight variables were monitored and analyzed: CHSK, BSK, pH, IM, Ntot, NH3, Cr and S2 (the codes are explained above). The sampling as well as the laboratory analyses was performed by relevant methods (Collective of authors 1965, Horáková et al. 1986). Since the waste water samples have usually a changing character, the analyses were made shortly after sampling so that the time between the end of sampling and the start of analyses was up to two hours (STN EN ISO 1994). The wastewater monitoring sites and the analytical laboratory were located close together.

Analysis

COD (CHSK) was measured by dichromate method spectrophotometrically at 585 nm (Horáková et al. 1986, STN ISO 6060 2001). BOD5 (BSK5) was determined using a titration method using the reaction with manganese dioxide and iodometric determination of the formed Mn(IV); the excess of iodine was titrated by thiosulphate standard solution (STN EN 1899-1 2001, 1899-2 2001, 258 13 1996). The pH value was measured by a glass electrode using the temperature compensation of the slope pH/electrode potential (STN ISO 10523 2005). Insoluble matters (IM) were separated on an appropriate filter, then filtered off, dried at 105 °C until the constant weight and determined gravimetrically (Horáková et al. 1986, STN EN 872 1999).

In analysis of ammonia it was first separated from the sample by distillation, followed by the reaction with the Nessler reagent and spectrophotometric measurement of absorbance at 425 nm. In the total nitrogen determination inorganic and organic nitrogen is oxidized by persulphate and mineralized in sulphuric acid. Then the formed nitrate reacts with 2,6-dimethylphenol in sulphuric and phosphoric acid medium under formation of 4-nitro-2,6-dimethylphenol, which is determined spectrophotometrically at 345 nm.

Chromium was selectively determined by inductively coupled plasma optical emission spectrometry (ICP-OES) at the wavelength 267.7 nm (STN EN 1233 1999). Sulphide anion was determined by iodometric titration; the excess of iodine was titrated by thiosulphate standard solution (Tkáčová 1979, STN 83 0520-16 1976, 83 0530-31a 1979). For quality assurance, the control samples were analysed together with the series of the waste water samples. In this process, the quality management system and internal process control established in the laboratory were applied.

Chemometrics methods and software

Mostly used multidimensional data analysis techniques were principal component analysis (PCA), cluster analysis (CA), and linear discriminant analysis (LDA). A majority of calculations was done by commercial software packages Statgraphics Plus, ver. 5.1 and SPSS, ver. 15.0.

Results and Discussion

Correlation analysis

The output of correlation analysis is the correlation table, which contains pair (or Pearson) correlation coefficients expressing the strength of correlation between all possible pairs of variables. The entries in this table are symmetrical according to diagonal. Table 1 shows the correlation table achieved under the same conditions, which were described in the part of cluster analysis. The following conclusions from the correlation table: (a) The highest correlation is between CHSK and BSK5. (b) A very high correlation is between CHSK and Ntot, BSK5 and Ntot, and CHSK and IM. (c) A very significant correlation ($r_{crit} \geq 0.258$ at $p \leq 0.01$) is between IM and Ntot, BSK5 and IM, IM and pH, NH4 and S2. (d) A significant correlation (at the 95 % or higher probability level, $p \leq 0.05$) is between BSK5 and pH, IM and NH4 (both dependences are inverse!). All correlation coefficients in Table 1 larger or equal than $r_{crit} = 0.184$ are marked by bold faces. (e) No correlation was proved in all other pairs.

Cluster Analysis

Generally, the clustering process in cluster analysis may be performed either with objects or variables (Khattree and Naik 2000). In this work clustering was made for the studied eight variables. The result of cluster analysis is a dendrogram depicted in Fig. 1. 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

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

81 studied objects, i.e. 27 averaged month values measured at the discharge of 3 leather plants. Critical values of the correlation coefficient (absolute values): r_{crit} = 0.144 (p = 0.1), r_{crit} = 0.184 (p = 0.05), r_{crit} = 0.258 (p = 0.01) for n = 81. Significant correlations are marked bold.

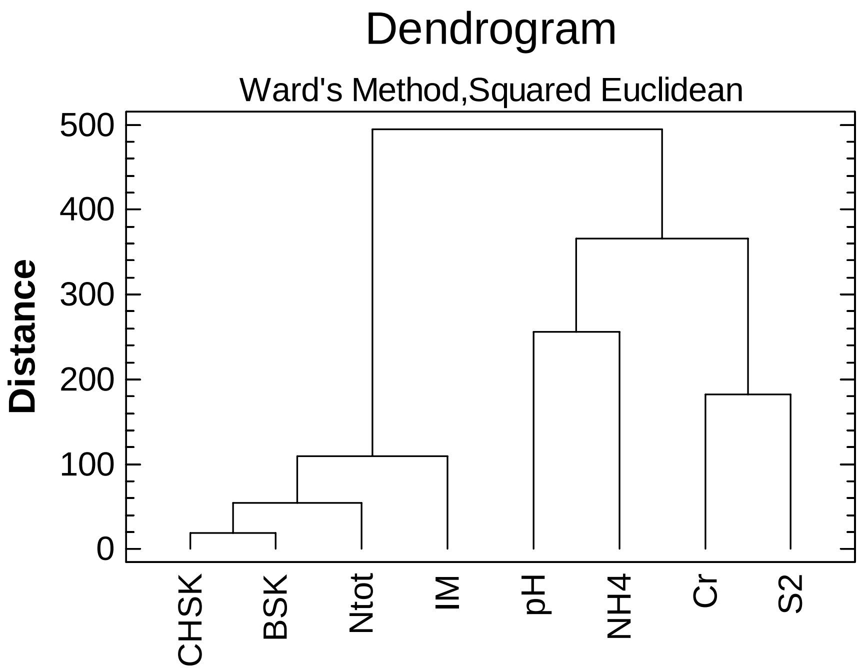

Fig. 1. Cluster analysis of 8 variables measured in waste water analysis of 81 samples obtained during 27 months (average values at 3 sampling sites). Software Statgraphics Plus 5.1. CHSK and BSK5 denote COD and BOD5, respectively.

eight investigated variables. Ward's method of clustering and squared Euclidean distance was used in these calculations.

In Fig. 1 three clusters of closely related variables can be seen. The first cluster is formed by CHSK, BSK, further connected 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 mutual similarity is given by Distance, which represents the vertical axis of dendrogram. The results of cluster analysis were confirmed also by correlation analysis and principal component analysis.

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 eigen values 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 scatterplot showing the objects, loadings plot showing the variables, and biplot where the objects and variables are depicted together. The advantage of first two is the possibility to obtain besides the 2D graph also the 3D one where usually the first three most important PCs are used as the axes. On the other hand, the biplot, even though plotted in two dimensions, provides more information about the studied problem.

Fig. 2 exhibits 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 at the discharge of the plants 1, 2 and 3.

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.

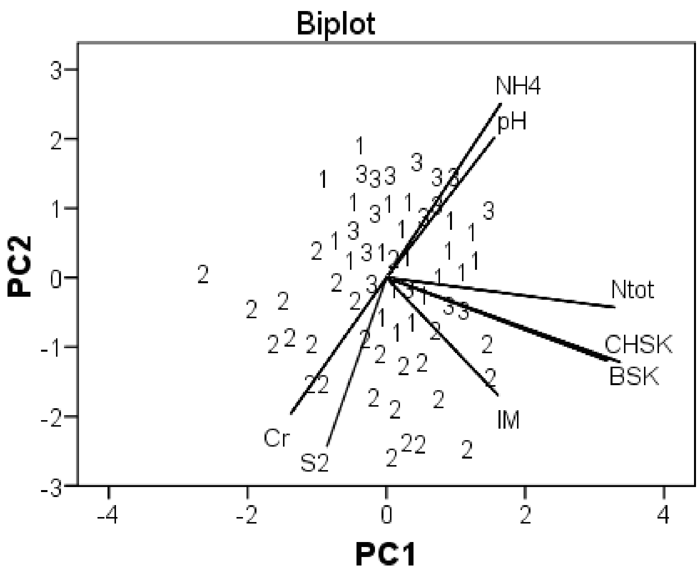

Fig. 2. Biplot PC2 vs. PC1 for 8 measured chemical descriptors and 81 samples from the discharges of three leather plants (coded 1, 2 and 3). Software SPSS 15.0.

When looking at the samples position in Fig. 2 it can be concluded that high sulphide and Cr values are characteristic for the waste 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 possible due to performed 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).

Linear Discriminant Analysis

In linear discriminant analysis, LDA, which is a supervised learning method, the classification model is calculated on the basis of the training data set where the categorization of each datum is known in advance. In this process, the calculation algorithm is trained to distribute the studied objects into given categories (Sharma 1996, Vandeginste et al. 1998, Khattree and Naik 2000); the category of the object is confirmed or it is assigned to another category. Classification success is given by the ratio of the correctly classified objects to their total number. In the second step, the developed model is used for classifications of the objects (waste water samples in this work) not included into the training set and belonging either to the test data set (used for validation of the discriminant model) or the object category is completely unknown (used for prediction of the object category). In the investigated problem, the categories were made according to the plant releasing the waste water. The classification results are summarized in Table 2.

The LDA results have shown that the waste water from the plants 1 and 3 have similar level of pollutants and the samples from these two sources are often interchanged (e.g. 4 samples belonging to plant 1 are classified into group 3, i.e. plant 3). The samples from plant 2 were classified with full success (100 %) when the training set is concerned and with 96.3 % when the sample category was made by leave-one-out technique. According to the position of the samples (belonging to three plant discharges) with respect to the first discriminant function (expressing the overall pollutant concentration) it was proved that the worst pollutant is plant 1, the best cleaning procedures were applied in plant 2.

Table 2. Success in classification of waste water samples by the sampling location (plants 1, 2 and 3) using linear discriminant analysis.

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 case is classified by the functions derived from all cases other than that case; [b]in total 82.7 % of original grouped cases were correctly classified (after calculating discriminant model); [c]in total 75.3% of cross-validated grouped cases were correctly classified.

In LDA an attempt was also made to see the seasonal effect on the waste water pollution. 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 four classes Spring, Summer, Autumn and Winter) and used as the LDA criterion. The separation of the data by four seasons was not very convincing, however, the Winter and Autumn data were displayed at higher values of the first discriminant function, which reflects the extent of pollution (the total concentration of the polluting substances). Much more successful was seasonal division of the data into two parts: the first half year (Spring and Summer) and the second half year (Autumn and Winter). In this case the success for the training set (used for calculating the discriminant model) was 62/81 = 76.5 % and that for the leave-one-out validation was 66.7 %, which are clearly significant results. At the same time it means that in cold months the total level of pollution for the waste water of three investigated plants is larger than in the part of the year when the temperature is higher.

Conclusion

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

This work was supported by courtesy of the Slovak Grant Agencies (VEGA 1/3584/06 and APVV-0057-06).

References

Collective of authors (1965) Unified methods of chemical analysis of waters (in Slovak), SNTL, Prague.

Horáková M, Lische P, Grünwald A (1986) Chemical and physical methods of water analysis (in Slovak), Alfa, Bratislava.

Kannel PR, Lee S, Kanel SR, Khan SP (2007) Analytica Chimica Acta 582: 390–399.

Khattree R, Naik DN (2000) Multivariate data reduction and discrimination, SAS Institute, Cary, North Carolina.

Kraic F, Mocak j, Fiket Ž, Kniewald G (2008) Chemical Papers 62: 445–450.

Sharma S (1996) Applied Multivariate Techniques, John Wiley, New York.

STN EN 1899-1 (75 7369) (2001) Water Quality. Determination of BOD (BSK) after n days (BSKn). Part 1. Dilution water and vaccine water with allylthiocarbamide addition, Slovak Technical Standard.

STN EN 1899-2 (75 7369) (2001) Water Quality. Determination of BSK after n days (BSKn). Part 2.: Method for non diluted sample, Slovak Technical Standard.

STN EN 25813 (75 7462) (1996) Water Quality. Determination of soluble oxygen. Iodometric method, Slovak Technical Standard.

STN EN 872 (75 7365) (2005) Water Quality. Determination of insoluble matters. Filtration method through the filter from glass fibres.

STN EN 1233 (75 7452) (1999) Water Quality. Chromium determination. Atomic absorption spectrophotometric method, Slovak Technical Standard.

STN EN ISO 5667-3 (75 7051) (1994) Water Quality - Sampling. Part 3: Guide for conservation of samples and manipulation with them, Slovak Technical Standard.

STN ISO 10523 (75 7371) (2005) Water Quality. pH measurement, Slovak Technical Standard.

STN ISO 6060 (2001) Water Quality. Determination of COD (CHSK), Slovak Technical Standard.

STN 83 0520-16 (1976) Physico-chemical analysis of drinking water. Determination of sulphide and sulphane, Slovak Technical Standard.

STN 83 0530-31a (1979) Chemical analysis of surface water. Determination of sulphide and sulphane, Slovak Technical Standard.

Šnuderl K, Simonič M, Mocak J, Brodnjak-Vončina D (2007) Acta Chimica Slovenica 54: 33–39.

Tkáčová J (1979) Titration method of determination of sulphides and sulphane in waste water by Slovak Standard STN 83 0530-31a (in Slovak), VUVH (Research Institute of Water Economy), Bratislava.

Vandeginste BGM, Massart DL, Buydens LMC., de Jong S, Lewi PJ, Smeyers-Verbeke J (1998) Handbook of Chemometrics and Qualimetrics: Part B, Elsevier, Amsterdam.

Vončina E, Brodnjak-Vončina D, Mirkovič N, Novič M (2007) Acta Chimica. Slovenica 54: 119–125.