ELSEVIER

Available online at www.sciencedirect.com
ScienceDirect

Desalination 213 (2007) 297–310

DESALINATION

www.elsevier.com/locate/desal

Pattern recognition techniques for the classification of wastewater samples based on their UV-absorption spectra and their fractions after applying MW-fractionation techniques

Melina E. Kotti, Nikolaos A. Parisis, Athanasios G. Vlessidis*,
Nicholaos P. Evmiridis

Laboratory of Analytical Chemistry, Department of Chemistry, University of Ioannina, 451 10 Ioannina, Greece
Tel. +30 2651098401; Fax +30 2651044831; email: avlessid@cc.uoi.gr

Received 4 December 2005; revised accepted 2 April 2006

Abstract

Different category-wastewater samples were collected from the inlets of biological treatment plants (installed in hospitals, industries and municipality) and from the body of polluted natural surface water (PNSW) systems (lakes, rivers). UV absorption spectral data of each wastewater system or their fractions as obtained by fractionation with membranes of various pore openings or with gel permeation chromatography were treated by supervised (neural network) and unsupervised (cluster analysis) pattern recognition methods with the target to classify them in clusters that include exclusively samples of the same category. The results based on neural network method applied to $\log_{10}$(UV-absorption spectra) of 80 wastewater samples gave a prediction score of around 77% for all category-samples. The cluster analysis method applied to the 1st derivative of $\log_{10}$(UV-absorption spectra) of 79 wastewater samples gave a promising classification for one of the four category wastewater samples and the others were grouped in sub-clusters of the same cluster without clear separation. Fractionation through membrane dialysis of two extremely non-similar UV-spectra samples from each category showed that the cluster analysis was more successful when UV-absorption spectra of the high molecular weight fractions were used in cluster analysis. Fractionation with GPC-chromatography gave chromatographic peaks and peak-spectra that are different for each category of wastewater samples; this method revealed that the MW of absorbing species are different and the absorption intensities are significantly different between the inlet feeds of the three types of wastewater treatment plants.

Keywords: Wastewater; Pattern recognition techniques; MW-fractionation techniques; Wastewater treatment plants

---

*Corresponding author.

Presented at the International Conference on New Water Culture of South East European Countries-AQUA 2005, 21–23 October 2005, Athens, Greece.

doi:10.1016/j.desal.2006.04.089

1. Introduction

Quality parameters of wastewater systems are determined by formal global analytical methods of BOD, COD, TOC, TN-Kjeldahl, TPhosphates, TSS; however, the procedures are time consuming and with poor reproducibility. Sensors have, also, been proposed for in vitro and in situ determinations, but they are limited by interference and fouling problems; therefore there is a demand for a simple and reliable technique.

Spectrophotometry is a good choice because of its reasonable sensitivity, simplicity, rapidity and accuracy. UV-absorptiometry was used for the estimation of organic pollutants. Relationships between absorption and wastewater quality parameters were obtained [1]; derivative methods of molecular spectrophotometry were used for determinations of nitrates and phenolics [2], multi-wavelength absorptiometry for nitrates, Cr(VI), organic matter in grinding sludges and wastewater characterization [3–7].

Spectrophotometers apart from obtaining spectra of absorbing species in batch systems it can be used as detectors in FIA-, and chromatography flow systems of analysis for in vitro or in situ determinations [8]. The technique can be used with absorption on a single, dual-wavelength procedure, or by obtaining the whole spectra of wastewater which gives a wealth of useful information to be treated with statistical resolution techniques such as polynomial modeling [9–11,14–16] for the determination of UV-light absorbing species in wastewater-samples or pattern recognition methods for classifying them in categories according to the source of pollution [12,13,17–21].

With the development of CCD-detectors it became possible to obtain spectra instantaneously and therefore, the UV–Vis detectors of the chromatography and FIA flow systems provide the chromatogram of the relevant wastewater sample together with the absorption spectra of each eluted peak in the chromatogram at the same time. The rapid scan of wastewater spectra followed by statistical data-treatment with proper pattern recognition computer software provides a quick method for classifying wastewater systems into categories of pollution sources. A more accurate classification can be obtained by combining wastewater spectral data and/or spectral data obtained from the fractionation of wastewater and treat them statistically for classification of wastewater system.

In this work we report an elegant, precise and sensitive method based on the treatment of UV-absorption spectra of samples from wastewater treatment plant inlets of municipalities (MBTP-samples), hospitals (HBTP-samples), dairy product industries (DPIBTP-samples) and polluted natural surface waters (lakes, rivers) (L- and R-samples) with pattern recognition techniques of supervised and unsupervised learning. Furthermore, the chemometric treatments are applied to fractions of wastewater samples fractionated by dialysis through cellulose membranes of different MWCO and with gel permeation chromatography (GPC).

2. Experimental part

2.1. Methodology

The samples were collected from the inlets of different biological treatment plants and/or the aquatic volume of surface waters (lakes and/or rivers) in brand new polypropylene bottles of 1 L volume, and were filtered initially through a glass Millipore 0.8 μm filter and afterwards though an acetic cellulose filter of a 0.2 μm pore size. The UV spectra (200–400 nm) of the filtered samples were, then, obtained and the quality parameters were determined. The fractionation of the wastewater samples were achieved using either cellulose membranes, or gel permeation chromatography.

2.1.1. Technique of dialysis membranes

Two membranes by SERVA were selected; one of regenerated cellulose with MWCO 12,000–19,000 Da and the other of acetic ester of cellulose with MWCO 500 Da. The membranes were submitted in the advised preprocessing as reported in the special membranes' protocol.

2.2. Methodology

A sample of 10 mL, appropriately diluted, was suitably placed inside the MWCO 12,000–19,000 Da membrane and the whole was immersed in a container with 10 mL of distilled water and left alone for 24 h to reach equilibrium. The absorption spectra of solutions inside and outside the membrane were obtained, respectively. Then the outside solution from dialysis through MWCO 12,000–19,000 Da membrane was introduced inside the MWCO 500 Da membrane and left alone for 24 h with an equal volume of distilled water to achieve equilibrium. The UV-absorption spectra of the solutions inside and outside the membrane were, then, obtained.

2.2.1. Gel chromatography

A gel column with dimensions of 1.6 cm diameter and 100 cm length, was filled with Sephadex G-25 Medium (50–150 mm dry spherule size, having an exclusion limit of 5000 Da, and a complete penetrability limit of 1000 Da), was used. Buffer solution of phosphates at pH = 6.4 was used as the mobile phase.

2.3. Methodology

A sample of 5–7 mL is injected on the top of column using a syringe. A peristaltic pump was used to achieve a constant flow rate of the mobile phase at 0.76 mL/min. Fractions (~3.5 mL) were collected in the fraction collector tubes at the exit of the column. The spectra of the fractions were obtained using a UV–Vis spectrophotometer in the range of 200–400 nm. Duplicate-chromatographs were obtained by UV-absorption at 220 and 260 nm.

2.3.1. Pattern recognition techniques

A number of eighty wastewater samples from inlets of Biological Treatment Plants of Municipalities (MBTP), Public Hospitals (PHBTP) University Hospitals (UHBTP), Dairy Product industries (DPIBTP), and polluted surface waters (L and R) were collected and prepared for running the UV-absorption spectra, and further treatment with neural network and cluster analysis pattern recognition techniques, using the STATISTICA software for Windows v7.0 program by StatSoft, Inc.

Neural networks: According to this methodology, the $\log_{10}$ of the absorption values in the region 210–329 nm (120 wavelengths) were coded in four categories of wastewater. The neural network RBF (Radial Basis Function) was used for the classification of samples; it was developed in three planes of neurons. From the 80 samples (spectra), 60 were used for the training of the neural network and the remaining 20 were used for finding the effectiveness of the training model.

Cluster analysis: A hierarchical cluster analysis technique was applied on the sample data using Ward's method and Euclidean distances. The spectra were pre-processed in one of the following methods (i) by calculating the first derivative of the $\log_{10}$ of the raw absorption data in the region 200–400 nm, or (ii) by normalising the absorption values of each wastewater sample in the region 210–329 nm (120 wavelengths) with the formula

$$\%A_{\lambda i} = 100 \times \left(A_{\lambda i} / \sum_{i=1}^{319} A_{\lambda i} \right)$$

3. Results and discussion

Seventy nine UV-absorption spectra of wastewater samples from MBTP-, HBTP-, DPIBTP-inlets and PNSW are preprocessed with both methods (i) and (ii) described in the experimental part and the resulted spectra were treated with cluster analysis technique; the dendrograms obtained after taking the first derivative are shown in Figs. 1 and 2, respectively. According to the dendrogram in Fig. 1 the samples are grouped in four main sub-clusters (IB1, IB2, IIA, and IIB) while according the dendrogram in Fig. 2, the samples are grouped in the four main sub-clusters (I, IIA1, IIA2, and IIB). In Tables 1 and 2 are tabulated the number of samples, from each of the four categories of wastewater samples involved in this work, that are included in each of the four clusters found by each of the two preprocess-methods, respectively.

From results in Table 1 it is found that 85% of category MBTP-samples are included in cluster II (IIA, IIB), 71% of category HBTP-samples in cluster IB, 86% of category DPIBTP-samples in cluster II and 93% of category L-samples in cluster II. From results in Table 2 it is found that 83% of category MBTP-samples are found in cluster IIB; 75% of category HBTP-samples are included in cluster IIA; 57% of category DPIBTP-samples are included in cluster IIB; 53% of category PNSW(L)-samples are included in cluster IIB. In both cases the prediction score for the MBTP- and HBTP-samples is quite satisfactory if we take into account that the quality

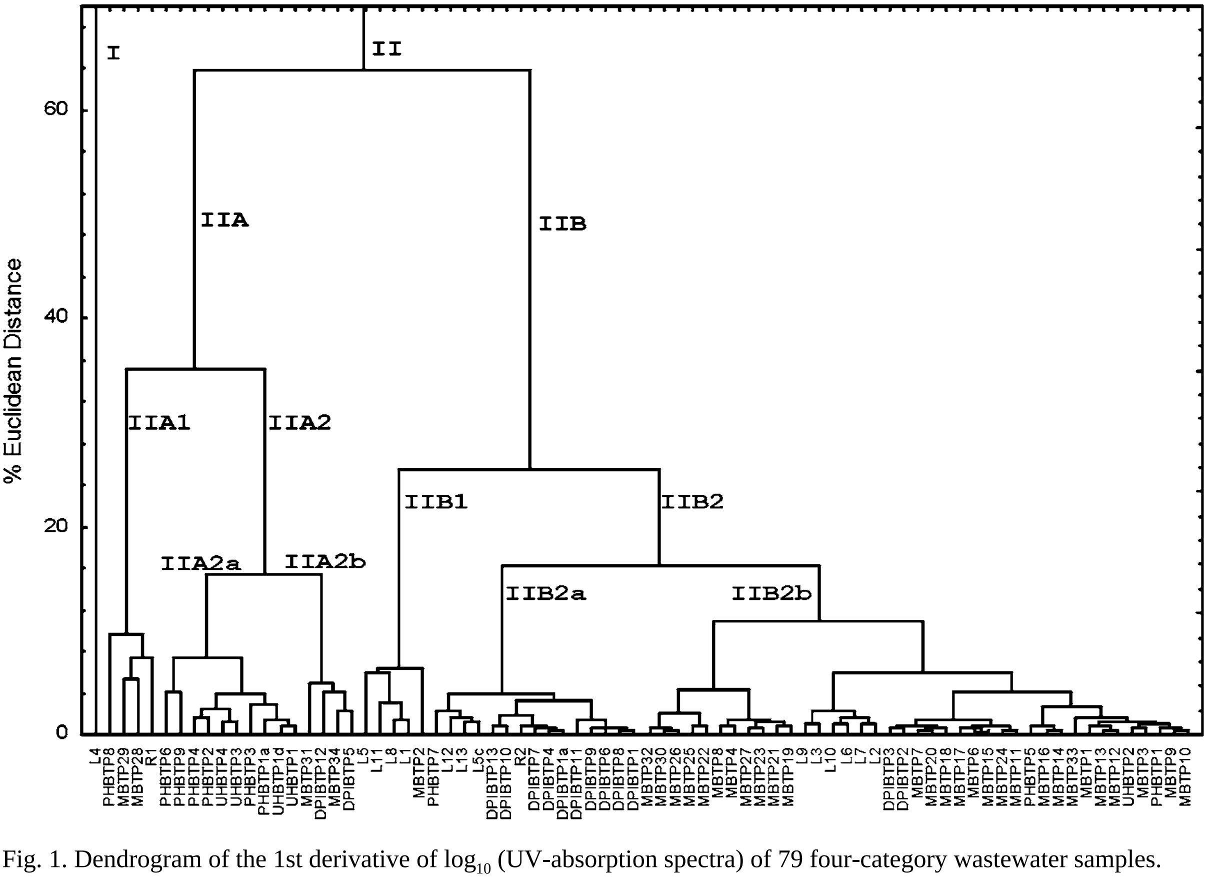

Fig. 1. Dendrogram of the 1st derivative of $\log_{10}$ (UV-absorption spectra) of 79 four-category wastewater samples.

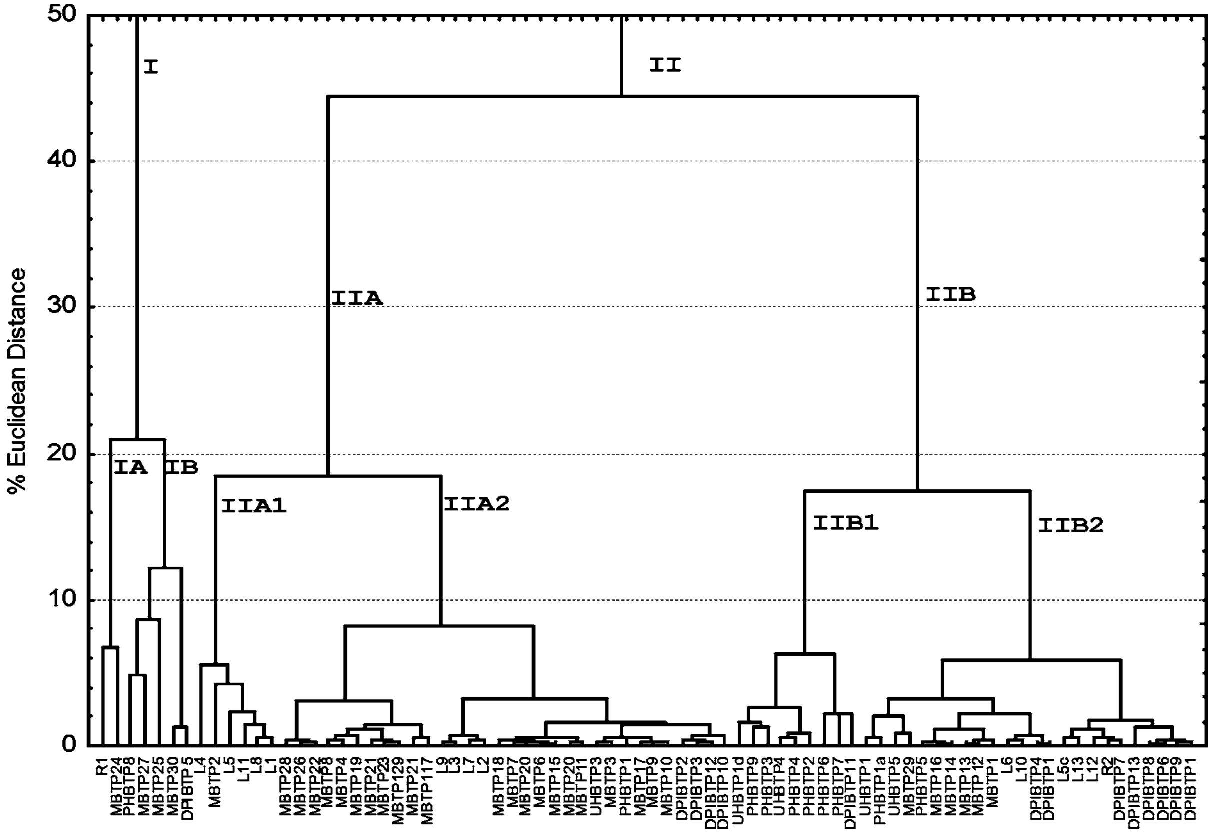

Fig. 2. Dendrogram of the 1st derivative of the normalized UV-absorption spectra of 79 four-category wastewater samples.

parameters of wastewater feeds and the operational regimes of the plants may differ widely. However, the MBTP samples cannot be distinguished from DPIBTP and L category samples. On the other hand both methods give same number of main (% Euclidean distance > 25) sub-clusters, similar classification order and, approximately, the same sample-population in each cluster from each category. This is strong evidence that the clusters formed are not restricted from the actual

Table 1
Classification of 79 normalized and their first derivative UV-spectra of wastewater samples from four different categories

Category	Normalized (UV-absorption spectra)				1st derivative normalized (UV-absorption spectra)				
	I	IIA1	IIA2	IIB	I	IIA1	IIA2	IIB1	IIB2
MBTP	4	5	22	2	4	1	22	0	6
HBTP	1	0	2	11	1	0	2	8	4
DPIBTP	1	0	7	6	1	0	4	1	8
R, L	1	5	7	3	1	5	4	0	6

Table 2
Classification of 79 $\log_{10}$ and first derivative of $\log_{10}$ UV-spectra of wastewater samples from four different categories

Category	Log_{10} (UV-absorption spectra)				1st derivative $\log_{10}$ (UV-absorption spectra)				
	IA	IB	IIA	IIB	I	IIA1	IIA2	IIB1	IIB2
MBTP	3	8	3	19	0	2	2	1	27
HBTP	3	3	3	5	0	1	10	0	3
DPIBTP	0	5	3	6	0	0	2	0	12
R, L	13	2	0	0	1	1	0	4	10

statistical treatment. On the other hand wastewater-spectra from samples of specific category that are scattered away from the centroid of their sub-cluster may be assumed to be outliers since their frequency is pretty small.

Therefore, cluster analysis treatment of wastewater sample spectra may be used:

- To discriminate MBTP-samples from HBTP-samples, and
- to recognize the outliers.

Based on the data of neural network technique a prediction score of 77% for all categories of samples tested was obtained which is close to that obtained with cluster analysis technique of the four category-samples. Furthermore, the proximity of % prediction between the two pattern recognition techniques suggests that the classification of the samples is more or less the same and any differences between the methods depends on the outliers; therefore, cluster analysis technique may become useful, also, as a preprocess-method for removing outlier samples before applying the neural network technique for a better prediction score with the latter method.

Another way to improve the prediction score of wastewater category sample is to use molecular weight based fractionation techniques, take the spectral data of the fractions and group them using cluster analysis technique based on the UV-spectra of the obtained fractions.

3.1. Fractionation based on dialysis membranes having different pore openings

For this investigation seven BTP-inlet samples were chosen; two municipal, three hospital, and two dairy product industry. Furthermore, two PNSW samples were included in the lot. The choice of these samples from the lot of 79 was made on the basis that their normalized UV-spectra are grouped in different sub-clusters of the main cluster found from cluster analysis from the 79 samples. Normalized UV-spectra and quality parameters of the chosen samples for this investigation are given in Fig. 3 and Table 3. The results from the cluster analysis are tabulated in Tables 4 and 5 for the normalized and for the

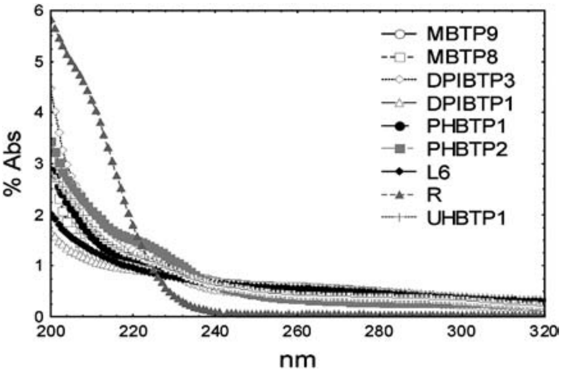

Fig. 3. Normalized UV-absorption spectra of nine samples from MBTP-, HBTP-, DPIBTP-, and PNSW-samples used for fractionation by dialysis through cellulose membranes of different pore openings.

Table 3
Wastewater quality parameters of the four category samples used for fractionation by dialysis through cellulose membranes of different pore openings

Samples	COD	N-NH$_4^+$	P-PO$_4^{3-}$	TSS	pH
MBTP8	334	44.5	5.25	255.5	7.3
MBTP9	462	151	9.75	898.5	7.7
DPIBTP3	–	–	–	–	–
DPIBTP1	–	–	–	–	–
PHBTP1	800	34	3.5	–	–
PHBTP2	794	27.2	2.3	168.5	7.7
UHBTP1	–	–	–	–	–
L6	57	6.4	–	23	8.4
R	80	0.1	0	–	8.1

logarithmically treated UV-absorption spectra, respectively.

These samples were submitted to a dialysis procedure as described in the experimental part and the normalized UV-spectra of each fraction for all involved samples are given in Fig. 4. The results from the cluster analysis are tabulated in Tables 4 and 5 for the normalized and for the logarithmically treated UV-absorption spectra, respectively.

3.2. Investigation of UV-absorption spectra from wastewater and its fractions

3.2.1. Optical inspection of UV-absorption curves

From inspecting the UV-absorption curves in Figs. 3 and 6 it is found that

- Absorption curves of all samples in all fractions converge in the range of 225–230 nm to an isosbestic point thus separating the UV-absorption spectra into two regions.
- The integral of the absorption within region of 200–225 nm varies between the different fractions and within the fractions between the different samples of each fraction.
- The absorption curve features in region of 200–225 nm vary in between the different fractions and within fractions between individual samples of each fraction.
- The absorption curve integral and features within the range of 240–400 nm is similar for all fractions and for all samples within the fractions with a few exceptions.
- The variation of the integral absorption and features over the range of 200–240 nm in the fractions that include compounds of low molecular weight becomes larger than those

Table 4
Classification of different categories wastewater samples and their fractions according to cluster analysis treatment of their normalized UV-spectra and their first and second derivatives

Sample	MW > 12,000			500 < MW < 12,000			MW < 500			Initial sample		
	UV	1st	2nd	UV	1st	2nd	UV	1st	2nd	UV	1st	2nd
R	IA1	IA1	IA1				IIA	IIB1	IIB2	I	I	I
L6	IBI	IIBI	IIA	IA	IA	IIA	IIB2	IIB2	IIB1	IIB1	IIA2	IB2b
MBTP8	II	IIA	IB2	IIA2	IB	I				IIB2	IIB2	IIB1
MBTP9	II	IIB2	IIB2	IIB1	IIB	IIB2	IIB2	IIB2	IIB2	IIB2	IIB2	IIB1
UHBTP	II	IIA	IBI	IIB	IIB	IIB2	IIB2	IIB2	IIB2	IIB2	IIA1	IB2
PHBTP1	II	IIA	IIB2	IB	IA	IIB1				IIB2	IIA1	IB1
PHBTP2	II	IIB2	IIB1	IB	IA	IIA2	IIB2	IIB2	IIB2	IIA1	IIB2	IIB1
DPIBTP1	II	IIA	IB2	IIB	IIA	IIB2	IIB1	IIA	IIA	IIB1	IIA2	IB2b
DPIBTP3	II	IIB2	IIB2	IIA1	IIA	IIB2	I	I	I	IIA2	IIB1	IIA

Table 5
Classification of different categories wastewater samples and their fractions according to cluster analysis treatment of their $\log_{10}$ UV-spectra and their first and second derivatives

Sample	MW > 12,000			500 < MW < 12,000			MW < 500			Initial sample		
	UV	1st	2nd	UV	1st	2nd	UV	1st	2nd	UV	1st	2nd
R	I	I	I				IIA	IIB1	IIB1	I	I	I
L6	IIA	IIA	IIA	I	IIA1	IIA	IIA	IIB2a	IIB2a	IIB1	IIA2b	IIB1
MBTP8	IIB	IIB1	IIB	I	I	I				IIB1	IIB2a	IIB2
MBTP9	IIB	IIB2b	IIB	IIB	IIB	IIB2b	IIB	IIB2b	IIB2b	IIB2	IIB2b	IIB2
UHBTP	IIB	IIB1	IIB	IIA	IIB2a	IIB2a	IIA	IIB2b	IIB2b	IIA	IIA2a	IIB1
PHBTP1	IIB	IIB2a	IIB	IIA	IIA2b	IIB2a				IIA	IIA1	IIB1
PHBTP2	IIA	IIB2b	IIB	IIA	IIA2a	IIB1	IIA	IIB2b	IIB2b	IIB1	IIA2a	IIB2
DPIBTP1	IIA	IIB2a	IIB	IIB	IIB2b	IIB2b	IIA	IIA	IIA	IIA	IIA2b	IIB1
DPIBTP3	IIB	IIB2b	IIB	IIA	IIB1	IIB2b	I	I	I	IIB2	IIB1	IIA

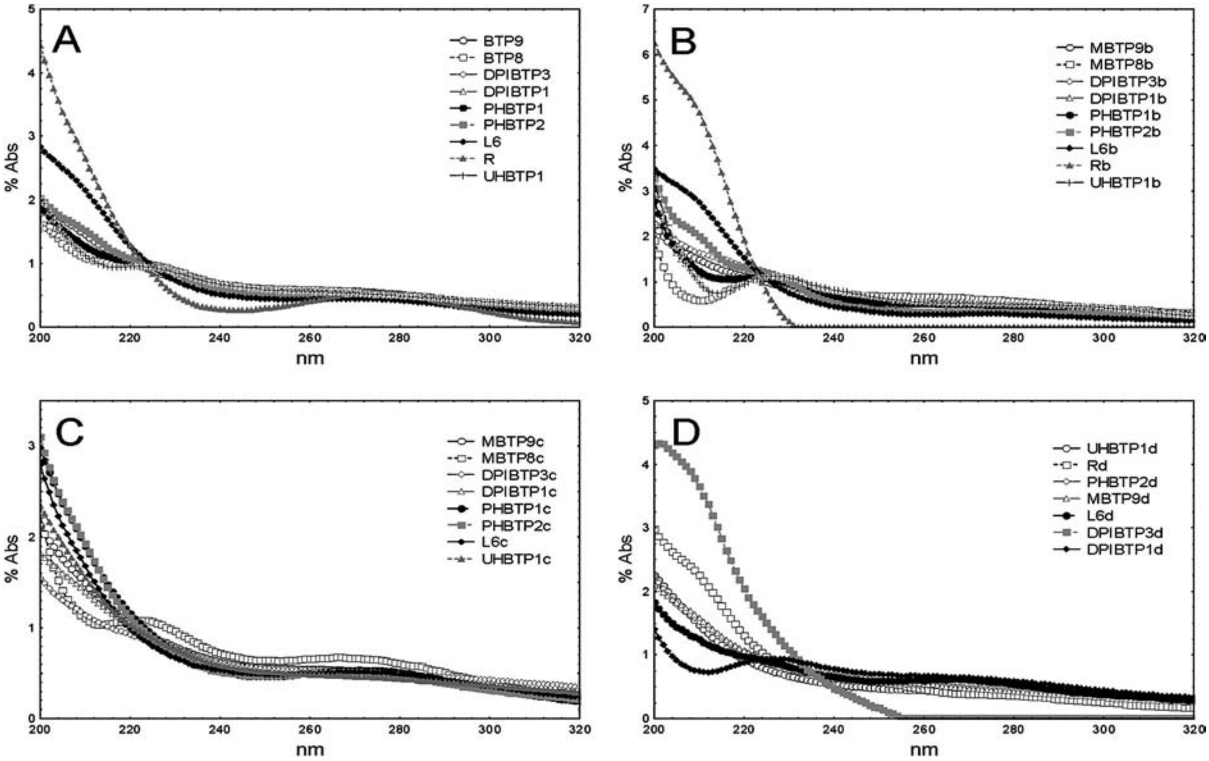

Fig. 4. Normalized UV-spectra for the spectra of the fractions of the nine wastewater samples submitted to the dialysis process: (A) MW > 12 kDa, (B) MW < 12 kDa, (C) 0.5 < MW < 12 kDa, (D) MW < 0.5 kDa.

observed in fractions that contain high MW-ingredients; this suggests that low molecular weight (UV light-absorbing compounds) in their fractions differ in species and quantities in each wastewater category.

3.2.2. Cluster analysis of UV-absorption spectra from wastewater and its fractions

Tables 4 and 5 give the designation of the specific cluster in which each sample is included after cluster analysis treatment of the normal (non-derivatized), first, and second derivative of normalized UV-spectra and $\log_{10}$(UV-spectra) of the initial wastewater and its fractions obtained by dialysis. Specific classification i.e. cluster including exclusively one category of samples was scarce; however selective classification was the rule in cases where samples of the same category were included in the same cluster but other category samples, were, also, included.

Looking over the data in Table 4 under the heading of:

- *Initial sample in UV-column*, it is realized that the R-sample is included in cluster I and all the rest are included in various subdivisions of cluster II.
- *MW > 12 kDa in UV-column*, it is found that the surface water samples are included in cluster I (R in IA and L in IB). All other samples are classified in cluster II. This is important since it is possible though this fraction to distinguish the BTP-inlet samples from the surface water samples; furthermore, cluster analysis of these fraction spectra, distributes the different kinds of surface water samples in different sub-clusters.
- *0.5 < MW < 12 kDa in 1st der. column*, both the DPIBTP-samples are included in sub-cluster IIA and no other category sample is found in there.

Looking over the data in Table 5 under the heading of:

- *Initial sample in UV-column*, it is realized that the R-sample is included in cluster I and all the rest are included in various subdivisions of cluster II.
- *Initial sample in 1st der. column*, it is realized that in sub-cluster IIB2 are included both the MBTP-samples alone. Therefore in this case the MBTP-samples are grouped in a sub-cluster of their own.
- *0.5 < MW < 12 kDa in UV-column*, it is found that all the HBTP-samples are included in sub-cluster IIA.

Table 6 summarizes all cases of specific and selective classification of the samples in their categories obtained by cluster analysis of the normalized and $\log_{10}$ UV-spectra (normal, 1st and 2nd derivative) together with other category-samples found in the same cluster (called interference).

From Table 6 it is realized that no individual preprocess of UV-absorption spectra is capable of classifying all the category-samples into sub-clusters of the correct category that they belong. However, it is found that by fractionating samples on the basis of MW it is possible to change the absorption integral and features especially in the range of 200–225 nm. Such changes suggest the change of the distribution of UV-absorbing ingredients between the obtained fractions. The failure, however, to find a fraction which will classify the samples in their categories is probably due to the fact that the fractionation degree by the method applied in this work is poor and perhaps that the wastewater samples of the different categories were very different; therefore, more powerful fractionation or statistical treatment techniques are needed for a clear distinction and safer prediction of such complicated water-systems.

3.3. Fractionation using gel permeation chromatography (GPC)

One representative wastewater sample from each of the MBTP-, HBTP- and DPIBTP-inlets was chosen and their spectra are shown in Fig. 5.

Table 6
Selectivity and interference from the other categories in the classification of BTP-wastewater samples

Category	Selectivity						Interference
	Normalized UV-spectra			$\log_{10}$ (UV-spectra)			
Samples	Raw	MW > 12 kDa	0.5 < MW < 12 kDa	Raw	MW > 12 kDa	0.5 < MW < 12 kDa	
MBTP	Normal						HBTP-samples (2)
	1st der.						HBTP-samples (1)
	2nd der.						HBTP-samples (1)
		Normal					HBTP-samples (3), DPIBTP-samples (2)
				1st der.			None
					Normal		HBTP-sample (1), DPIBTP-sample (1)
HBTP						Normal	DPIBTP-sample (1)
DPIBTP			1st der.				None
			2nd der.				HBTP-samples (1), MBTP-samples (1)
					1st der.		HBTP-samples (1), MBTP-samples (1)

The quality parameters of the samples chosen are given in Table 7. These samples were submitted to fractionation process employing the GPC-technique. The GPC-chromatograms obtained for the three different category samples are shown in Fig. 6 and the associated chromatographic peak-spectra of each sample in Fig. 7.

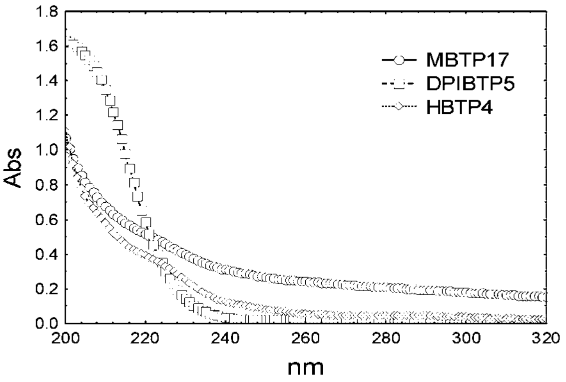

Fig. 5. UV-spectra of the MBTP-, HBTP-, and DPIBTP-sample of the three wastewater samples submitted to GPC-fractionation process.

From the spectra in Fig. 5 it is realized that the spectrum of MBTP-inlet sample is slightly different from the HBTP-inlet sample spectrum. Furthermore, the DPIBTP-inlet sample spectrum is quite different from MBTP-inlet, and HBTP-inlet sample spectra.

From GPC-chromatograms in Fig. 6 it becomes obvious that:

- Fractions with ingredients of lower than 5 kDa molecular weight are eluted at different retention volumes at the chromatograms of the three tested samples, and the fractionation pattern for each category of wastewater sample as detected by the UV-absorption at 220 and 260 nm is different, suggesting that the UV-light absorbing species differ between the categories of wastewater in quantity and molecular weight;
- MBTP-samples show high UV-absorption from intermediate molecular weight components,

Table 7
Quality parameters of the three category biological treatment wastewater samples fractionated with gel permeation chromatography

Sample	Date	COD	N-NH$_4^+$	N-NO$_3^-$	P-PO$_4^{3-}$	TSS	pH
MBTP	28/11/01	446	78	–	18.5	515	7.64
HBTP	05/12/01	655	18.5	9.5	5.5	132.2	9.05
DPIBTP	29/11/01	8500	80	–	40	–	5.62

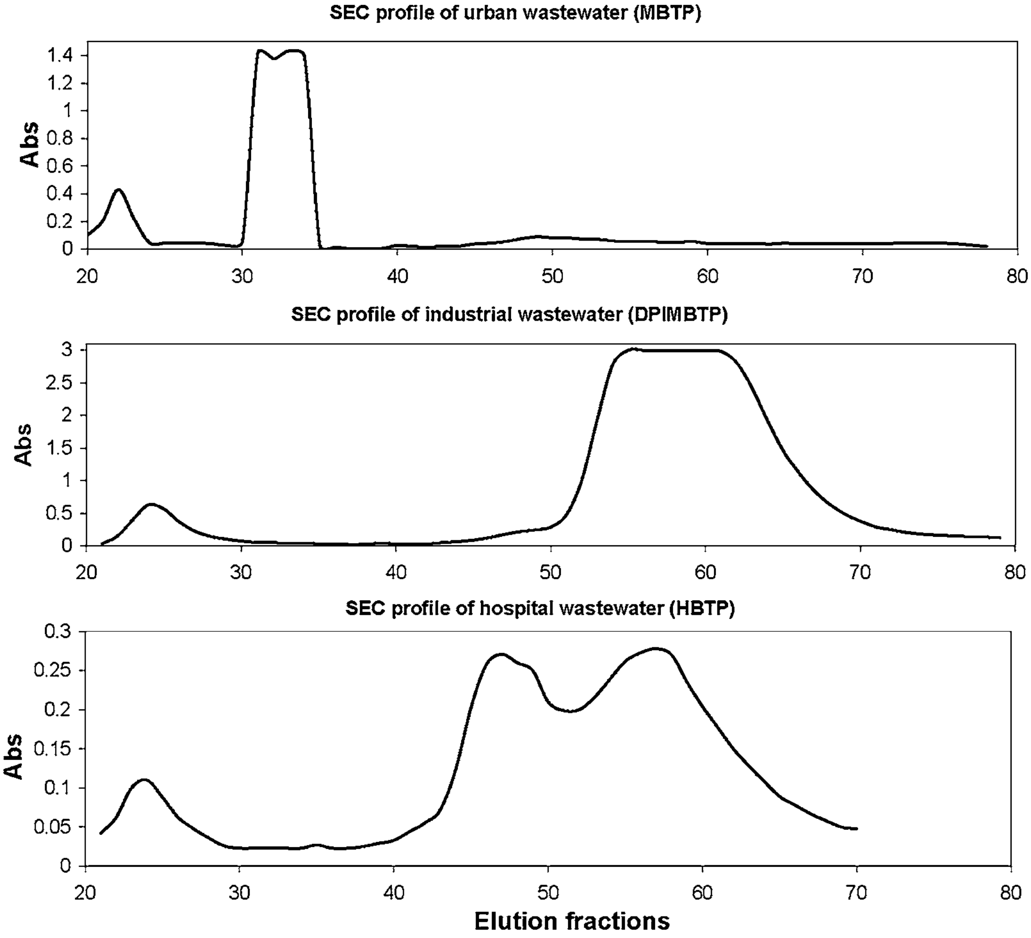

Fig. 6. GPC-chromatograms of the MBTP-, HBTP-, and DPIBTP-samples of the three category wastewater samples submitted to GPC-fractionation process.

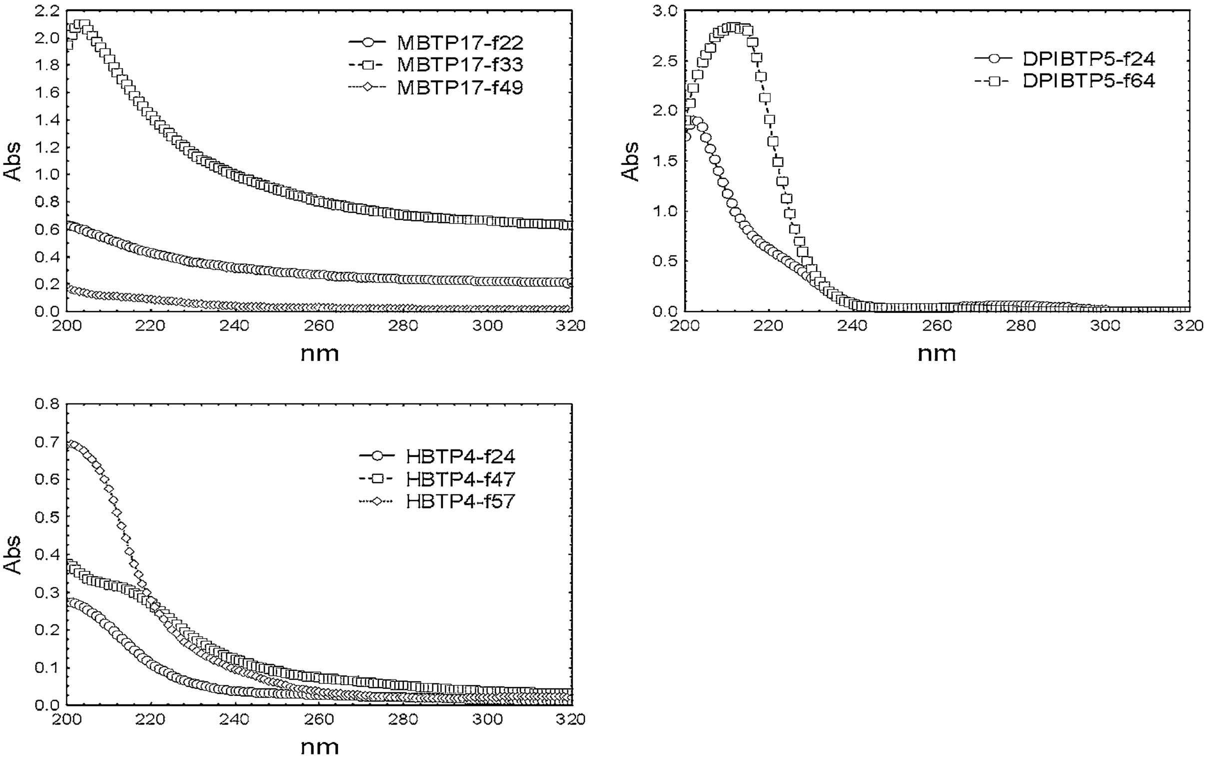

Fig. 7. UV-spectra of the fractions of MBTP-, HBTP-, and DPIBTP-wastewater samples submitted to GPC-fractionation process.

lower from high molecular weight materials and very low from small molecular weight compounds;

- HBTP-samples show relatively high absorption of UV-light from the small and intermediate molecular weight molecules and low absorption from the relatively high molecular weight substances, finally
- DPIBTP-samples show the highest absorption from the relatively low molecular weight compounds and much less from the high molecular weight substances.

From optical inspection of the obtained spectra of chromatographic peaks of each sample it is realized that:

- UV-absorption spectra profile curves within fractions and between samples are quite different.
- Feature-differences of UV-absorption profile curves of the intermediate MW fractions between the three category samples are great.
- UV-absorption intensity is much lower in the HBTP-inlet sample in all fractions compared to the corresponding fractions of the other two category samples.

Because of the safe predictability by GPC-fractionation technique there was no need to proceed with cluster analysis based on UV-spectra of adequate number of samples. From the results of the GPC-fractionation it is evident that wastewater samples from the different biological treatment plant-inlets are different in composition and MW of UV-absorbing ingredients, and the fraction UV-spectra are different, also. Furthermore, the overall intensity of UV-absorption is large for the DPIBTP-inlet sample, smaller for the MBTP-inlet sample and much smaller for the HBTP-inlet sample.

3.4. Discussion on the generalization of the method

The GPC fractionation method seems to be successful with the samples we have investigated. However, we are doing further work with more samples to generalize the method. On the other hand the pattern recognition methods of the initial spectrum are much more rapid and our effort is to make it more accurate.

4. Conclusions

From this investigation it is found that:

(1) Pattern recognition techniques (cluster analysis and neural network) can give a rather good prediction score of the category of wastewater rapidly, just by obtaining and normalizing the UV-spectrum of the wastewater sample in question.

(2) Cluster analysis of the normalized UV-spectra of wastewater category samples can also be used to remove outlier samples before the use of neural network technique for better prediction score.

(3) Widely different samples of the same category treated by cluster analysis can be submitted to a fractionation technique combined with cluster analysis.

(4) Cluster analysis of 1st derivative $\log_{10}$ or normalized (UV-absorption spectra) from fractions of higher molecular weight give better classification of the different category samples.

(5) Fractionation with membrane dialysis was time consuming process and insufficient for an efficient fractionation.

(6) GPC-fractionation is able to predict the category just from the GPC-chromatogram as well as from UV spectra of the fractions obtained.

Appendix A

General equations for multi-component samples according to Beers Law are

$$\sum_{i}^{n} F_{i,k} = \sum_{i}^{n} \varepsilon_{i,k} \cdot c_i \cong \overline{c} \sum_{i}^{n} \varepsilon_{i,k} \tag{1}$$

$$\sum_{j}^{l} \sum_{i}^{n} F_{i,j} = \sum_{j}^{l} \sum_{i}^{n} \varepsilon_{i,j} \cdot c_i \cong \overline{c} \sum_{j}^{l} \sum_{i}^{n} \varepsilon_{i,j} \tag{2}$$

where F stands for absorbance, k for the specific wavelength, j for any wavelength and i for the dissolved absorbing species in the sample. Also n stands for all absorbing species dissolved in the sample, c is the mean concentration of the n dissolved absorbing species involved in the sample, and l for the total number of wavelengths in the spectra measured.

This approximation is reasonable since most of the dissolved species involved in wastewater samples are in the ppm concentration level.

The base 10 logarithm of Eq. (1) and its derivative are shown in Eqs. (3) and (4) respectively:

$$\log \sum_{i}^{n} F_{i,k} \cong \log \sum_{i}^{n} \varepsilon_{i,k} + \log \overline{c} \tag{3}$$

$$\frac{\mathrm{d} \log \sum_{i}^{n} F_{i,k}}{\mathrm{d}\lambda} = \frac{\mathrm{d} \log \sum_{i}^{n} \varepsilon_{i,k}}{\mathrm{d}\lambda} \tag{4}$$

The normalization of the absorption data at each specific wavelength in the spectra is approximated by Eqs. (5) and (6) respectively:

$$\frac{\sum_{i}^{n} F_{i,k}}{\sum_{j}^{l} \sum_{i}^{n} F_{i,j}} \cong \frac{\sum_{i}^{n} \varepsilon_{i,k}}{\sum_{j}^{l} \sum_{i}^{n} \varepsilon_{i,j}} \tag{5}$$

$$\frac{\mathrm{d}}{\mathrm{d}\lambda} \left[\frac{\sum_{i}^{n} F_{i,k}}{\sum_{j}^{l} \sum_{i}^{n} F_{i,j}} \right] \cong \frac{\mathrm{d}}{\mathrm{d}\lambda} \left[\frac{\sum_{i}^{n} \varepsilon_{i,k}}{\sum_{j}^{l} \sum_{i}^{n} \varepsilon_{i,j}} \right] \tag{6}$$

References

[1] R. Briggs and K.V. Melbourne, Water Treatm. Exam., 17 (1968) 107–120.

[2] A.R. Hawthorne, S.A. Morris, R.L. Moody and R.B. Gammage, J. Environ. Sci. Health, Part A, 19 (1984) 253–266.

[3] O. Thomas and S. Gallot, Fresenius J. Anal. Chem., 338 (1990) 234–237.

[4] O. Thomas, S. Gallot and N. Mazas, Fresenius J. Anal. Chem., 338 (1990) 238–240.

[5] O. Thomas, S. Gallot and E. Naffrechoux, Fresenius J. Anal. Chem., 338 (1990) 241–244.

[6] B. Jourdan, J. Ollivier and O. Thomas, Environ. Technol., 24 (2003) 597–603.

[7] S. Vaillant, M.F. Pouet and O. Thomas, Talanta, 50 (1999) 729–736.

[8] A. Höskuldsson, J. Chemom., 2 (1988) 211–228.

[9] D.A. Skoog, F.J. Holler and T.A. Nieman (Eds.), Principles of Instrumental Analysis, Saunders Golden Sunburst Series, 5th edn., Saunders College Publishing, Holt Rinehart and Winston, 1998.

[10] W. James (Ed.), Advances in Modeling the Management of Stormwater Impact, Taylor & Francis–CRS Press, 1996.

[11] S.E. Jorgensen (Ed.), A Systems Approach to the Environmental Analysis of Pollution Minimization, Taylor & Francis–CRC Press, 1999.

[12] J. Fan and I. Gijbels (Eds.), Local Polynomial Modeling and its Applications, Monographs on Statistics and Applied Probability, Vol. 86, Taylor & Francis–CRC Press, 1996.

[13] J.P. Marques de Sá (Ed.), Pattern Recognition: Concepts, Methods, and Applications, Springer, Berlin, New York, 2001.

[14] B.R. Kowalski, J. Chem. Inf. Comput. Sci., 15 (1975) 201–203.

[15] B. Clement and O. Thomas, Environ. Technol., 16 (1995) 367–377.

[16] E.H. Smith and W.J.R. Weber, Water Air Soil Pollut., 53 (1990) 279–295.

[17] R. Gourdon, C. Comel, P. Vermande and J. Veron, Water Res., 23 (1989) 167–173.

[18] M. Vidal and M.T. Melgar, Water Air Soil Pollut., 121 (2000) 367–377.

[19] M.E. Kotti, A.G. Vlessidis, N.C. Thanasoulias and N.P. Evmiridis, Water Res. Manage., 19 (2005) 77–94.

[20] D.L. Massart and L. Kaufman (Eds.), Interpretation of Analytical Chemical Data by the Use of Cluster Analysis, Wiley, New York, 1983.

[21] M.A. Sharaf, D.L. Illman and B.R. Kowalski, Chemometrics, in: P.J. Elving, J.D. Winefordner and I.M. Kolthoff (Eds.), Chemical Analyses: A Series of Monographs on Analytical Chemistry and its Applications, Vol. 82, Wiley–Interscience, New York, 1986.