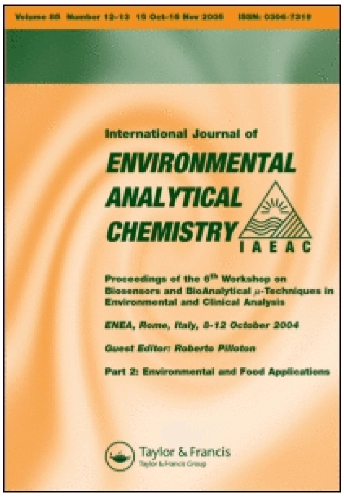

International Journal of Environmental Analytical Chemistry
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713640455

Multivariate Exploratory Data Analysis of the Organic Micropollutants Found in the Llobregat River (Catalonia, Spain)

Elisabeth Teixidó[a]; Lourdes Olivella[a]; M. Figueras[a]; Antoni Ginebreda[a]; Romà Tauler[b]
[a] Department de Laboratori, Agència Catalana de l'Aigua, Barcelona, Spain [b] Department of Analytical Chemistry, University of Barcelona, Barcelona, Spain

To cite this Article Teixidó, Elisabeth , Olivella, Lourdes , Figueras, M. , Ginebreda, Antoni and Tauler, Romà(2001) 'Multivariate Exploratory Data Analysis of the Organic Micropollutants Found in the Llobregat River (Catalonia, Spain)', International Journal of Environmental Analytical Chemistry, 81: 4, 295 — 313

To link to this Article: DOI: 10.1080/03067310108044250

URL: http://dx.doi.org/10.1080/03067310108044250

Intern. J. Environ. Anal. Chem., Vol. 81, pp. 295–313
Reprints available directly from the publisher
Photocopying permitted by license only

MULTIVARIATE EXPLORATORY DATA ANALYSIS OF THE ORGANIC MICROPOLLUTANTS FOUND IN THE LLOBREGAT RIVER (CATALONIA, SPAIN)

ELISABETH TEIXIDÓ[a], LOURDES OLIVELLA[a],
M. FIGUERAS[a], ANTONI GINEBREDA[a]
and ROMÀ TAULER[b,*]

[a]*Departament de Laboratori, Agència Catalana de l'Aigua, Provença 204-208, 08036 Barcelona, Spain;* [b]*Department of Analytical Chemistry, University of Barcelona, Martí Franqués 1, Barcelona 08028, Spain*

(Received 23 March 2001; In final form 26 June 2001)

The presence of organic micropollutants in the Llobregat River (Catalonia, Spain) has been investigated and the results interpreted by Principal Component Analysis (PCA). A deeper insight into the interpretation of the results was achieved by the extension of this analysis of the same samples including general water quality analytical parameters such as: total organic carbon, oxidability, biochemical oxygen demand after 5 days, dissolved oxygen, pH, suspended solids, anionic surfactants, total hydrocarbons, ammonia, nitrates, phenol index and total coliforms. PCA afforded the exploration of the main pollution sources of the Llobregat River as well as the characterization of the geographical and temporal distribution of these pollution sources.

Keywords: Principal component analysis; River pollution; Organic micropollutants

INTRODUCTION

It is widely accepted that the number of organic chemicals currently used, either in the domestic or in the industrial domain, may easily exceed

---

[*]Corresponding author. Fax: +34-93-4021233. E-mail: roma@apolo.qui.ub.es

300,000 compounds[1–3]. Therefore, many of these compounds of anthropogenic origin, together with their metabolites or conversion by-products can potentially act as environmental contaminants. Of course, the above figure is quite impressive, and obviously escapes any attempt of exhaustive legislative enumeration. Thus, for instance, focusing the attention on water, the Directive 76/464/EEC [Annex I Directive 76/464/CEE: Directive of Council, 4th May, 1976] concerning toxic harmful substances for the aquatic environment, only considers 132 specific organic compounds. From the analytical point of view, the extraordinary development of chromatographic techniques coupled to mass spectrometry based detectors, with powerful qualitative identification capabilities (GC-MS, HPLC-MS and many "hyphenated" related sub-families) have opened wide possibilities in the field of the analysis of organic micropollutants.

Due to such analytical huge progress, environmental experimental data regarding organic micropollutants are relatively easily available for a routine control laboratory. But a second question still remains, as it is their further handling and interpretation, in order to throw some insight in the origin and classification of the pollution sources. In this context, the use of multivariate statistical techniques, such as Principal Component Analysis (PCA) has revealed very promising[4–6]. Improving water quality management, would be greatly allowed by the development of such a kind of data analytical tools.

In the present paper, a chemometrical interpretation of the analytical data regarding organic micropollutants in the Llobregat river basin (Catalonia, Spain) (Fig. 1) is performed, and the correlation between these analytical data and other general analytical physico-chemical parameters commonly used in water characterization (Table I) is attempted.

The Llobregat River (and its main tributaries Cardener and Anoia), covering a basin area of 5,110 km^2, plays a significative role for Catalonia, since it provides the main supply of water for domestic and industrial use of Barcelona and its surroundings (3.2 million inhabitants). On the other hand, and due to the same reason, it also receives the impact of wastewater (treated and untreated) and diffuse pollution caused by such a heavy human activity.

Figure 1 shows the geographical situation of the selected sampling points (15) along the river course. Water samples analyzed (34) were gathered throughout our routine monitoring campaigns, during 1997–1998, (see Table I) every point being analyzed at least two times within that period.

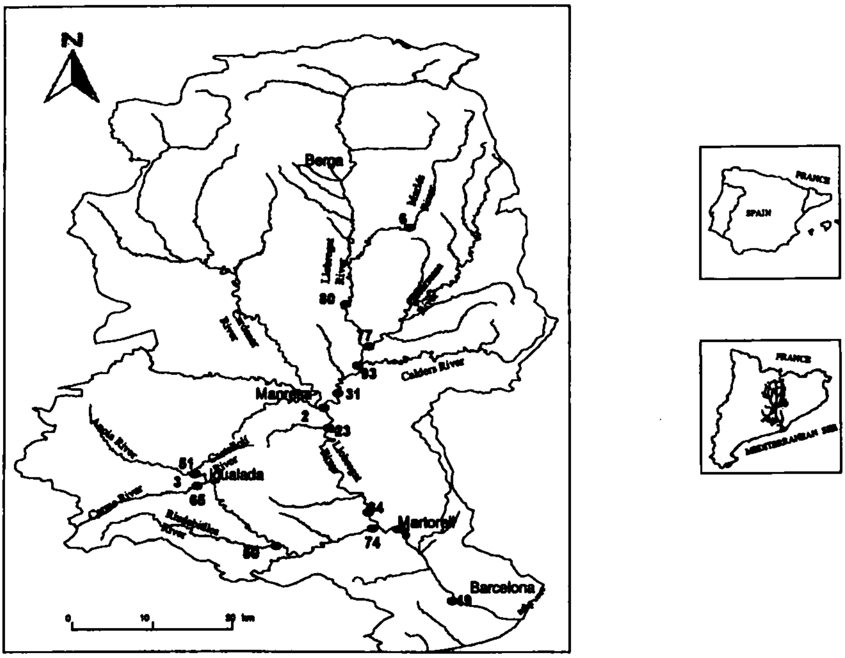

FIGURE 1 Location of study area and the selected sampling points (15) along the Llobregat River course (Catalonia, Spain). Identification numbers of different sample sites and main activities of the surrounding areas are given in Table I.

METHOD

Data pre-treatment

In this work, 34 samples and concentration values of 94 different organic compounds (Table II) composed the first selected data set. This data set was correspondingly arranged in a matrix of 34 rows (samples) and 94 columns (organic compounds). Since all the analyte concentrations were measured in the same scale units (μg/l), data were only variable mean-centered but not scaled (i.e., each element was subtracted by its column mean). Limit of detection was roughly established at 0.01 μg/l since lower values were not significant. To avoid the undesirable presence of "zeros", all concentration values below the limit of detection, as well as, compounds not detected were also considered equal to 0.01 μg/l.

A second data set and data matrix was built with the same samples (34) and 14 general quality parameters (see Table III). Since these parameters are very different (physico-chemical, microbiological and quality indices)

TABLE I Sampling sites description

Code	*Date*	*Location*	*Discharge*
2J	01/97	Cardener River	Urban and food industry
5J	01/97	Llobregat River	Textile industry and others industries
6J	01/97	Marlès River	Urban discharge und farms
23J	01/97	Llobregat River	Textile industry
31J	01/97	Llobregat River	Textile industry and food industry
49J	01/97	Llobregat River	Industrial in general
77J	01/97	Gabarressa River	Food and textile industry
80J	01/97	Llobregat River	Urban
84J	01/97	Llobregat River	Textile industry
93J	01/97	Calders River	Food industry
2A	04/97	Cardener River	Urban and food industry
5A	04/97	Llobregat River	Textile industry and others industries
23A	04/97	Llobregat River	Textile industry
31A	04/97	Llobregat River	Textile industry and food industry
49A	04/97	Llobregat River	Industrial in general
77A	04/97	Gabarressa River	Food and textile industry
80A	04/97	Llobregat River	Urban
84A	04/97	Llobregat River	Textile industry
3U	06/97	Anoia River	Urban, industrial and leather goods
50U	06/97	Riudebitlles River	Paper mill and wine producing
74U	06/97	Anoia River	Food industry and leather goods
2D	12/97	Cardener River	Urban and food industry
5D	12/97	Llobregat River	Textile industry and others industries
6D	12/97	Marlès River	Urban discharge and farms
23D	12/97	Llobregat River	Textile industry
31D	12/97	Llobregat River	Textile industry and food industry
49D	12/97	Llobregat River	Industrial in general
80D	12/97	Llobregat River	Urban
84D	12/97	Llobregat River	Textile industry
93D	12/97	Calders River	Food industry
3M	03/98	Anoia River	Urban, industrial and leather goods
51M	03/98	Castellolí River	Paper mill and wine producing
65M	03/98	Carme River	Paper mill
74M	03/98	Anoia River	Food industry and leather goods

and thus not measured in the same scale units, the data were auto-scaled prior to PCA analysis (each element was subtracted by its column mean and divided by the standard deviation of its column).

Principal Component Analysis

In this work, Principal Component Analysis (PCA) is used to investigate the correlations between 94 organic compounds and 14 water general quality parameters. PCA is used to investigate the eventually existing relationships between samples and between variables (concentration of organic compounds and quality parameters) and to deduce how many independent

TABLE II Analyzed organic compounds

Code #[a]		*Organic Compounds*
Chlorinated aromatic hydrocarbons		
1	1	Dichlorotoluene
2	2	Trichlorobenzene
3		Chlorodimetoxybenzene
4		Trichloronaphthalene
5		Tetrachloronapathalene
6		Pentachloronaphtalene
Aromatic hydrocarbons		
7	3	Trimethylbenzene
8	4	Methylpropylbenzene
9		2,3-dihydro-1H-indene
10	5	Methylnaphthalene
11	6	1,2,3,4-tetrahydronaphthalene
12	7	Naphthalene
13	8	1,2,3,4-tetrahydro-5-methylnapthalene
14	9	1,2,3,4-tetrahydrodimethylnapthalene
15	10	1,1'-biphenyl
16		Dimethylnaphthalene
17	11	1,2-dihydroacenaphtilene
18	12	Methyl,1,1'-biphenyl
19	43	Phenanthrene
20	14	Fluorene
21	15	2,3-dihydro-1,1,3-trimethyl-3-phenyl-1H-indene
22	16	Anthracene
23	17	Methyldibenzothiophene
24	18	Methylphenantrene
25	19	Dimethylphenantrene
26	20	Pyrene
27	21	Flurorantene
28	22	Benzofluorene o methylpyrene
Phenols		
29		Methyl phenol
30		Ethyl phenol
31	23	1,1'-biphenyl-ol
Acids and esters		
32	24	Propanoic acid ester
33	25	Benzoic acid ester
34	26	Butyl hexadecanoate
35	27	Butyl octadecanoate
Alcohols and ethers		
36	28	2-ethylhexanol
37		1-phenoxy-propanol
38	29	2,6-dimethyloct-7-en-2-ol
39	30	α,α,α-dimethylbenzenemethanol
40	31	Dimethylbenzeneethanol
41	32	2-phenoxy ethanol
42	33	2-propanol derivatives
43	34	2-ethylhexyl 2-ethylhexanoate
44	35	α-phenylbenzenemetanol
45	36	Ethylphenoxybenzene

TABLE II (Continued)

Code #[a]		*Organic Compounds*
Ketones and aldehydes		
46	37	2-heptanone 4,6-dimethyl
47		Decanal
48		1-(3-aminophenyl)ethanone
49	38	Isophorone
50	39	p-isopropylphenylacetaldehyde
51	40	2,2,6-trimethyl-2-cyclohexene-1,4-dione
52	41	4-(1,1-dimethylethyl)cyclohexanone
53	42	2-H-benzopyran-2-one
54	43	Methanone diphenyl
55	44	((4-dimethylamino)phenyl)phenyl metanone
Plasticizers		
56	45	Dimethyl phthalate
57	46	Triisobutyl phosphate
58	47	Diethyl phthalate
59	48	Tri-*n*-butyl phosphate
60	49	Bis(2-methylpropyl) phthalate
61		1,3-dichloro-2-propanol phosphate
62	50	di-*n*-butyl phthalate
63	51	Acetyl tributyl citrate
64	52	Bis(2-ethylhexyl)phthalate
Pesticides and Herbicides		
65		Methyl-dimethylcarbamoditioat
66	53	1,3-benzodioxoles,4,7-dimetoxy-5-(2-propanyl)
67	54	Simazine
68	55	Diazinon
69	56	Tertbutylazine
70	57	Fenitrothion
71	58	Ethion
Surfactants and degradation products		
72	59	NPE($N = 0$)
73	60	NPE($n = 1$)
74		Linear alkyl benzene C11
75	61	Linear alkyl benzene C12
76	62	Linear alkyl benzene C13
Nitrogen containing compounds		
77	63	Benzothiazole
78		3-methyl, 1H-indole
79	64	2H-indole-2-one,1,3-dihydro-,1,3,3-trimethyl
Terpenes		
80		Limonene
81	65	Linalool
82	66	Camphor
83	67	5-methyl-2-(1-methylethyl)cyclohexanol
84	68	α-terpineol
Sterols		
85		Ethylcoprostanol
86	69	Coprostanol
87	70	Cholesterol
88	71	Sitosterol

TABLE II (Continued)

Code #[a]		*Organic Compounds*
Miscellaneous		
89		Dimethyl tetrasulphide
90	72	Tetrabutylurea
91	73	Geosmin
92	74	2-(methylthio)benzothiazole
93		Phenantrenequinone
94	75	Squalene

[a] First and second column codes give the variable identification number in data set A and in reduced data set A, respectively.

sources of them are needed to explain the observed data variance. PCA quantifies the variable interactions by computing the eigenvectors of the covariance data matrix[7,8]. PCA provides a projection of the objects (*scores*) from the high-dimensional variable space on to a reduced dimensional space defined by the new set of eigenvectors of principal components; i.e. using a linear combination (*loadings*) of the original variables which explains maximum variance and are orthogonal. Since the observed variance in the variables is usually highly correlated, only a reduced number of principal components is needed to explain most of it. Explained data variance, scores and loadings plots[7,8] give a powerful and graphical tool for the representation of multidimensional data allowing the identification and interpretation of the more important underlying sources of data variance. Varimax rotation of PCA scores and loadings[9] was also attempted to ascertain an optimal interpretation of the deduced factors and environmental contamination sources. In this work PCA calculations have been performed with the S-Plus 4.5 software package (Mathsoft, Inc., Cambridge) and MATLAB PLS Toolbox 2.0 (Eigenvector. Research Inc., Manson, Washington, 1998).

EXPERIMENTAL

Sampling

Water samples were collected from the already numbered points of the quality network established by the Catalan institution of Water (*'Agència Catalana de l'Aigua'*), indicated in Fig. 1 during the period comprised between January 1997 till March 1998. Sampling points of this network were chosen to achieve the quality objectives of the Spanish law. Depending on the use and the human activities developed along the river, Catalan institution of Water established where the sampling points must be, how often they have to be collected and which analysis must be done. Therefore, the samples 2, 5, 6, 23, 31, 49, 77, 80, 84 and 93 were collected three times a year (except for 5, 77 and 93 that were collected only two times for technical reasons). The samples 3, 50, 51, 65 and 74 were collected one or two times a year.

Samples were kept in 1 liter glass bottles fitted with Teflon-lined caps, leaving no headspace. After sampling, they were preserved in cold (not exceeding 5°C) until the moment of performing the analysis.

Chemicals and reagents

Suprasolve-grade dichloromethane, ethyl acetate, n-hexane, isooctane, 2-propanol and acetone were purchased from Merck (Darmstadt, Germany). Anhydrous Na_2SO_4 was purchased form Panreac (Barcelona, Spain) and was heated overnight at 400°C and kept dry until use. Surrogate standards nitrobenzene-d_5, 2-fluorobiphenyl and 4-terphyenyl-d_{14} and internal standards anthracene-d_{10} and decachlorobiphenyl were from Supelco (Bellefonte, PA, and U.S.A.).

Sample liquid–liquid extraction procedure

Liquid–liquid extraction was carried out according to the method 625, from the U.S. Environmental Protection Agency[10]. One liter of sample was added with surrogate standards (nitrobenzene-d_5, 2-fluorobiphenyl and 4-terphenyl-d_{14}) and extracted twice with dichloromethane (150 ml and 100 ml) by stirring for 10 min. The organic extracts were combined and dried with anhydrous Na_2SO_4. Dichloromethane was removed under reduced pressure, first in a round bottom flask and further in a conic flask, until a volume of 0.5 ml. The concentrate was transferred to a 1-ml conic vial, washing the flask with isooctane, and dried under N2 stream until a final volume of 100 µl for HRGC/MS analysis. Internal standard mixture (Anthracene-d_{10} and decachlorobiphenyl) was added to the extract.

HRGC-MS analysis

Final extracts were analyzed by HRGC-MS. The analyses were performed on an integrated quadrupol GC/MS MD-800 from Fisons (Manchester, U.K.). Helium was used as carrier gas (at a constant flow of 1.2 ml/min) in a DB-5MS column (30 m × 0.25 mm i.d., 0.25 μm film thickness). The program was from 90°C (held 5 min) to 240°C at 6°C/min and to 310°C (maintained for 10 min) at 10°C/min. Injector temperature was 280°C, and the injection mode was splitless for 90 s. The scanning was 40–500 m/z at 1 scan/s. MS spectra were compared with NIST spectra library (62,235 spectra) and with Wiley spectra library 5th ed. (138,111 spectra), and each compound was identified according to its best mass spectrum fitting.

Quantification

After its positive identification, a characteristic ion of the compound mass spectrum was selected (Table II), and the peak area was integrated. For the particular case of nonylphenol ethoxylate (NPEO), linear alkyl benzene (LAB), tri-, tetra- and pentachloronaphthalenes, the molecular ions were chosen and all the peak areas, appearing in a retention time interval, were added. In each sample, the peak of mass ion 188 corresponding to anthracene (internal standard) was integrated. A semiquantification of the organic compounds was calculated using the following internal normalization formula:

$$C = (A_{\text{ion}}/A_{IS})/C_{IS}$$

where C is the concentration of the compound (μg/l), A_{ion} = area of compound's characteristic ion, A_{IS} = area corresponding to ion mass 188 (internal standard) and C_{IS} = concentration of internal standard (μg/l). This formula assumes equal sensitivities for all the measured ions, which is not obviously the case in MS analysis and it has been considered adequate only for qualitative/semi-quantitative purposes. However, this is not a crucial objection, since the aim of the present study is more the elucidation of the relative pattern distribution and variability of the organic micropollutants found in the river (temporal and spatial), than providing accurate concentration values.

Organic compound description and analysis

As mentioned above, a list of 94 different organic compounds have been identified on the course of the sampling and analysis campaign. They are reported in Table II together with the characteristic ion used in quantification. Most of them have been unequivocally identified and classified in a rather "pragmatic" way, using different criteria related with their chemical generic families (i.e., aromatic hydrocarbons, alcohols etc.) or their use (pesticides, plasticizers etc.). Many of them (35%) have been previously reported to be present in wastewater effluents.

General water quality analytical parameters

It was considered of interest to compare the quality status of the samples, obtained from organic micropollutants, with that resulting from other quality measurements. These parameters were performed according to well established methods, described elsewhere[10,11]. The quality parameters used belong to several classes (Table III): physico-chemical (suspended solids, total organic carbon, oxidability, biochemical oxygen demand, total hydrocarbons, anionic surfactants, ammonia, phenol index, pH, dissolved oxygen and nitrates), microbiological (total coliforms), and two water quality indices, such as ISQA[12] and a biological index, the so called BMWP' (or BMWPC adapted to southern Mediterranean rivers)[13]. Biological index BMWP was measured following the method originally described in[14], and modified in[13]. It is based on an identification of family aquatic macro-invertebrates and assignation of a puntuation, this gets five groups of water quality. ISQA is a physico-chemical index of water quality, used by the local water authorities and computed from some simple physico-chemical parameters (temperature, oxidability, conductivity, suspended solids). It is scaled between 0 and 100[13].

TABLE III Quality parameters analysed

Code[a]	*Water Quality Parameters*	*Code*
95	Oxidability	OXID
96	Biochemical organic demand	BOD_5
97	Total organic carbon	TOC
98	phenols	PHEN
99	ammonium	NH_4
100	pH	pH
101	nitrates	NO_3
102	anionic surfactants	SURFA
103	dissolved oxygen	$DisO_2$
104	total hydrocarbons	HYDRC
105	suspended solids	SS
106	total coliforms	COLIF
107	ISQA	ISQA
108	BMWPC	BMWPC

[a] Variable identification number in data set B.

RESULTS AND DISCUSSION

PCA of organic microcontaminants (data set A)

In a first analysis, the mean-centered data set A (34 samples × 94 organic compounds) was examined. With four principal components (PC) 79.4% of the total variance was explained (see Table IV), the remaining PC's accounting for less than 10% of the data variance each one. Components explaining little data variance have not been taken into consideration in the analysis and they were assumed to be not relevant for the purposes of this work, i.e., the detection and distribution of the main Llobregat river pollution sources by organic compounds.

Loadings plots for the first four principal components (PC) are given in Fig. 2. The percentage of explained variance for the first PC was around 31.2%. Two butyl esters (variables 34, 35) and tributyl phosphate (variable 59) had high positive loadings in the first PC, and methyl and ethyl phenol (variables 29, 30) plasticizers (variables 56, 58), 3-methyl-indole (variable 78), sitosterol (variable 88) and squalene (variable 94) had high negative loadings. The second PC explains 21.4% of the total mean centered data variance and showed positive contribution of all variables contributing also to the first PC (variables 29, 30, 34, 35, 56, 58, 59, 78, 88 and 94). These organic compounds should be considered the more significant contaminants of the region under study. The third PC accounts for 13.9% of the total variance, which is loaded by negative contribution of plasticizers (variables 56, 57, 59), associated to an industrial origin. In the fourth PC (12.9% of explained variance) tributyl phosphate (variable 59) had a high

TABLE IV PCA results of the different data sets

Data set	*PC1*	*PC2*	*PC3*	*PC4*
34 samples and 94 organic compounds (data set A)	31.2[a] (31.2)[b]	21.4 (52.6)	13.9 (66.5)	12.9 (79.4)
23 samples and 75 organic compounds (reduced data set A)	57.1 (57.1)	20.5 (77.6)	7.3 (84.8)	4.3 (89.1)
32 samples and 94 organic compounds and 14 quality parameters (data set B)	21.4 (21.4)	16.0 (37.4)	10.6 (48)	10.3 (58.3)

[a] % percent of variance.
[b] (cum), percent of cumulative variance.

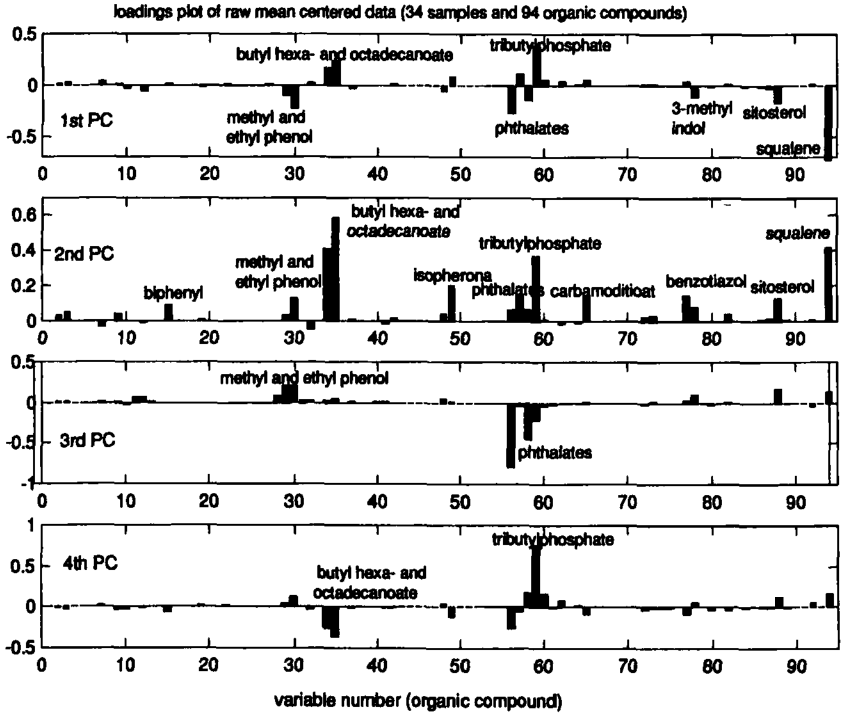

FIGURE 2 Loadings plot of the four principal components for the matrix of 34 samples and 94 variables (organic compound concentrations, data set A). See Table II for variable identification. More significant variable contributions are explicitly identified in the Figure.

positive loading, whereas butyl hexa- and octadecanoate (variables 34, 35) had high negative loadings.

Figure 3 gives the PCA scores for PC1 and PC2 (Up) and for PC1 and PC3 (Down). Samples 3M and 77J had high scores for PC1 and PC2, appearing well differentiated from the other remaining samples and also between them. These two samples are representative of two independent contamination sources, which are geographically separated. Sample 3M is representative of high concentrations of the two fatty acids butyl esters and tributyl phosphate and of low concentrations of squalene, whereas sample 77J is mostly related with high concentrations of the two butyl esters, tributyl phosphate and also with high concentrations of squalene. This shows the presence of two different independent sources for squalene, a first one, which is highly positively correlated with the two butyl esters and tributyl phosphate, and another source, which is independent of them. A cluster of samples collected in January appeared grouped at the left side

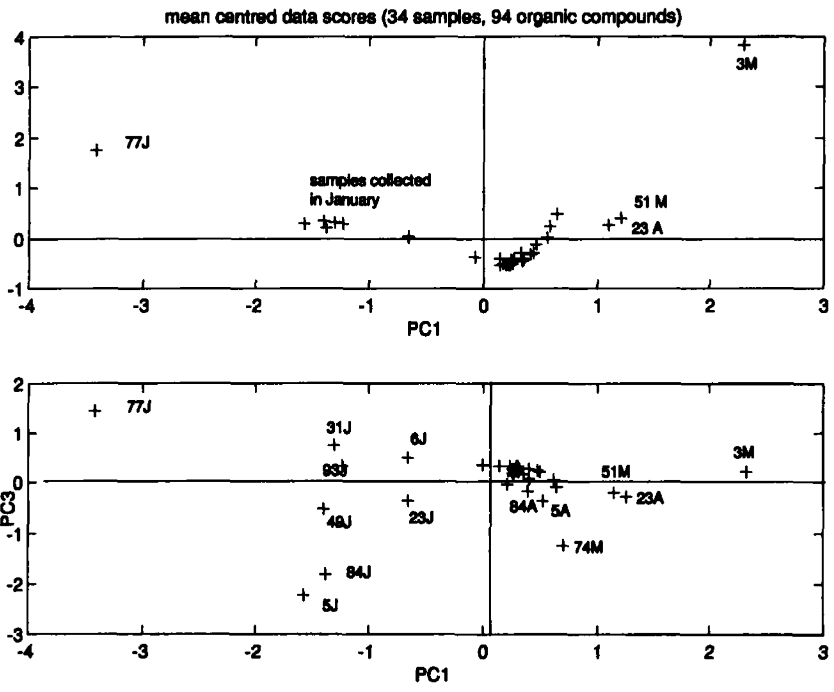

FIGURE 3 Scores plots of the principal components for the matrix of 34 samples and 94 variables (organic compound concentrations, data set A). Up PC1 vs. PC2; down PC1 vs. PC3. Only samples with higher score values are identified (see Table I).

of the PC1 axes. The reason for this was that in January phthalate derivative compounds had higher concentrations than during the rest of the months studied. A clear distribution north-to-south (i.e., up to downstream) of the January samples is observed in PC3 scores (Fig. 3, down). Positive PC3 scores are observed for 77J, 31J, 6J and 93J samples which are located in the upstream of Llobregat River, whereas samples 23J, 49J, 84J and 5J, which are located in the downstream of the River, gave negative PC3 scores. Therefore, this geographical separation between samples observed in the third PC, could be associated to gradual river quality degradation from up to downstream. Thus, samples at the downstream of the river, exhibiting poor water quality, reflect a significative urban and industrial pollution that agrees with the presence of plasticizers (phthalates and phosphates). On the contrary, samples at the upstream of the river, characterized by a better general quality, showed a pollution profile mostly composed by products of faecal origin (i.e., methyl-indole), natural products and some industrial products (i.e., phenols).

Since this foregoing analysis was strongly biased due to the singular character of some of the samples, a second analysis was performed after the elimination of the most conspicuous ones, i.e., those samples corresponding to January (including sample 77J) and the sample 3M (see Table I and Fig. 3). At the same time, the organic compounds, which were only present in these samples, were also eliminated (see Table II). This gave rise to a new data set with only 23 samples and 75 variables (organic compounds), which was again subjected to PCA. With three components, 84.8% of the total variance was now explained (see Table IV). First PC accounts for 57.1% of the total variance, which is highly negatively loaded by tributyl phosphate (variable 48, Fig. 4). Second PC (20.5% of explained variance) was positively loaded by plasticizers (variables, 46,47), propanoic acid ester (variable 24) and 2-(methylthio) benzothiazole (variable 74) whereas it was negatively loaded by tributyl phosphate (variable 48). Finally, in the third PC (7.2% of explained variance), trimethylbenzene (variable 3) and diethyl phthalate (variable 47) had positive loadings and propanoic acid ester (variable 24)

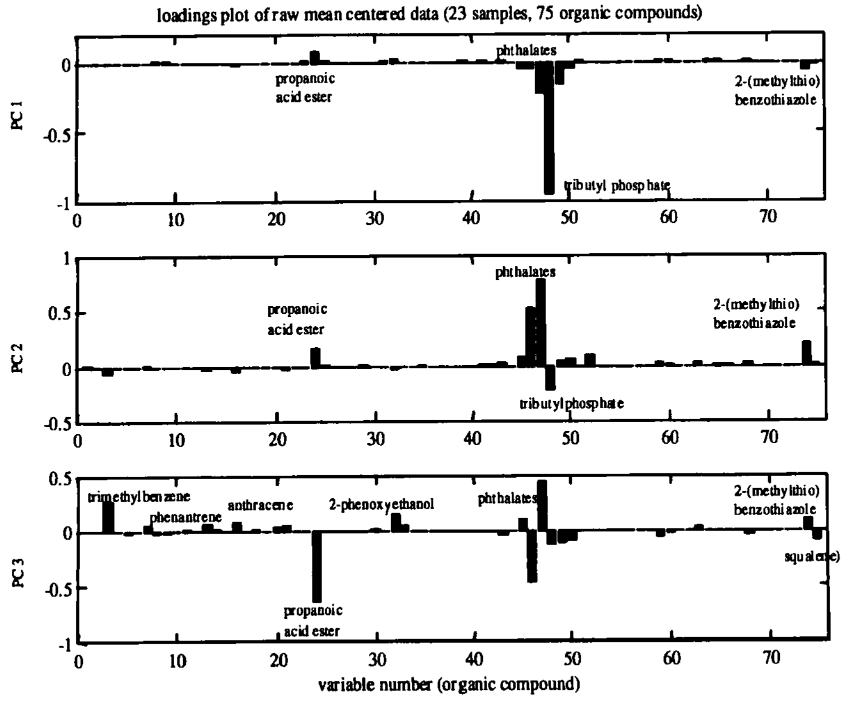

FIGURE 4 Loadings plot of the three principal components for the reduced matrix of 23 samples and 75 variables (organic compounds concentrations, reduced data set A). See Table II for variable identification. More significant variable contributions are explicitly identified.

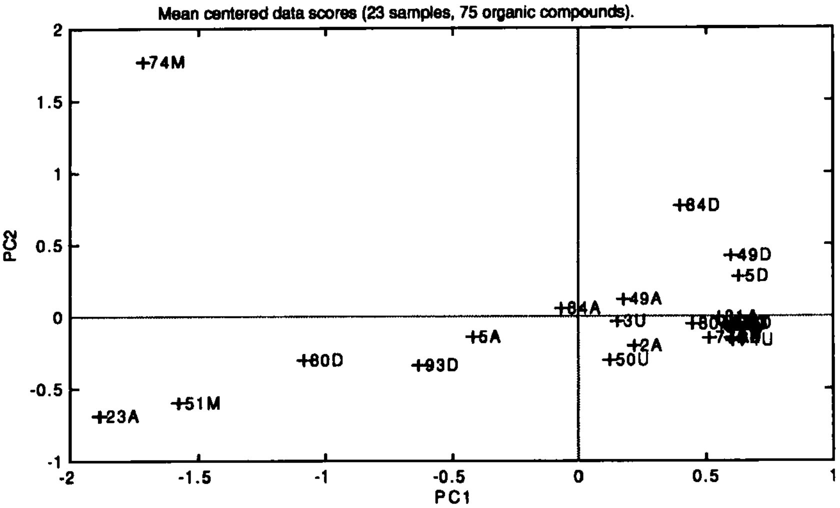

FIGURE 5 Scores plot of the two principal components for the reduced matrix of 23 samples and 75 variables (organic compound concentrations, reduced data set A). Only samples with higher score values are identified (see Table I).

and triisobutyl phosphate (variable 46) had negative loadings. Although diethyl phthalate was present in almost all samples, a very high concentration of this compound was found in sample 74M. High concentration values of the other two compounds (triisobutyl phosphate and propanoic acid ester) were found in samples 49D and 5D.

Figure 5 gives the PC1 and PC2 scores for the reduced sample data set. A large number of the samples had low scores values. However, some samples had high scores values for the first PC: samples 74M, 23A, 51M and 80D. All these samples had high concentration values of tributyl phosphate. The sample 74M also had high PC2 scores, since it had high concentration values of plasticizers (diethyl phthalate, triisobutyl phosphate and tributyl phosphate). In view of these results, plasticizers were found again to be the main source of pollution in Llobregat River. This is not strange, since it is well known, that these compounds are used in many industrial applications and domestic uses in the area, are ubiquitous and constitute a well known water contamination background in many other populated and industrial areas.

In order to investigate alternative interpretations of the detected environmental contamination sources varimax rotation[9] of previous PCA scores and loadings was attempted. The obtained angle of rotation in all cases was very small, proving that the PCA resolved scores and loadings had already achieved high simplicity values[9] and that the varimax rotated factors were very similar to straight PCA factors. Therefore, they gave very close interpretations to those already obtained by PCA without varimax rotation.

Simultaneous PCA of organic microcontaminants and of general quality parameters (data set B)

A new PCA study was performed including the concentration of the 94 organic microcontaminants and the 14 general quality parameters. The interest was to investigate the possible correlation between microcontaminants and general quality parameters not deduced from previous independent analysis. Before PCA was performed a data pretreatment was applied. The 94 variables corresponding to the concentration of the organic microcontaminants were mean-centered, while the 14 variables corresponding to the general quality parameters were first standardized (divided by their standard deviation) and then halved to have scale values similar to the organic microcontaminant concentrations.

As for the PCA of data set A, samples 3M and 77J had a very singular character. Thus, after elimination of these two samples, a new data set (data set B) was built up (second column in Table I). The resultant new matrix had 32 samples and 108 variables (94 organic compounds and 14 general quality parameters). A large number of principal components were required for the explanation of data variance (ten components to explain the 86% of the data variance). In Table VI (last row) PCA results are shown. The first PC explained 21.4% of the total variance and was loaded positively by dissolved oxygen, ISQA, BMWPC and squalene (94) and it was negative loaded by all the rest of contamination parameters of organic matter and plasticizers (triisobutyl phosphate and diethyl phthalate). This component described well the water quality of the River. A better quality in water is related to the presence of low concentrations of organic matter, total coliforms and anionic surfactants; and is also related to larger values of the quality indexes and higher concentrations of dissolved oxygen. It is worth mentioning that, the two quality indexes of water, ISQA and BMWPC, which were calculated independently, were positively correlated between them and also with dissolved oxygen, and negatively correlated with contamination organic matter parameters. As a conclusion, the two proposed indexes, ISQA and BMWPC (these indexes score from 0, poor quality, to 100, highest quality), were confirmed to be good indicators of the river water quality. Second PC (with 16% of the total variance) gave high positive loadings of tri-*n*-butyl phosphate, BOD, TOC, total hydrocarbons, ISQA and BMWPC and high negative loadings of phenol, phthalates, pH, nitrates and suspended solids. Third PC accounted for 10.6% of the total variance, and was attributed mostly to tri-*n*-butyl phosphate and other plasticizers, all of industrial origin. Trimethylbenzene, tri-*n*-butyl phosphate, dissolved oxygen, suspended solids and total coliforms gave high negative loadings in the fourth PC (10.3% of explained variance), whereas propanoic acid ester, triisobutyl phosphate, diethyl phthalate, phenols, anionic surfactants and BMWPC gave high positive loadings for the same PC. Thus, this is interpreted as if the presence of organic contaminants like phenols and surfactants did not affect the water biological quality (BMWPC). Fig. 6 gives the PCA scores for PC1 and PC3 (up) and for PC2 and PC4 (down). In PC1, samples

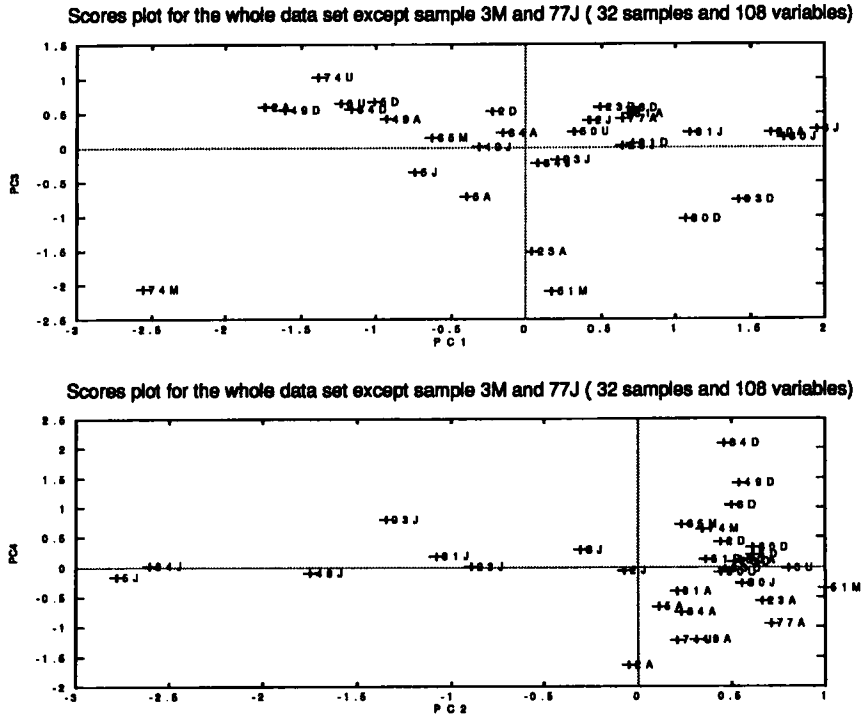

FIGURE 6 Scores plot of the principal components for the matrix of 32 samples and 108 variables (organic compounds plus quality parameters, data set **B**). Up PC1 vs PC3; down PC2 vs PC4.

with higher positive scores are located upstream and samples with higher negative scores are located downstream. Therefore PC1 described the general quality of the Llobregat River. On one side, the more polluted water samples were located in Anoia River and at the downstream of Llobregat River, in agreement with the areas which have the highest industrial and urban load. On the other side the samples with better water quality were located at the upper and middle parts of the Llobregat River. Samples 74M, 51M and 23A had high negative scores for PC3 (see Fig. 6 up). These samples had high amounts of plasticizers, especially of tributyl phosphate, which is the variable with larger loading in PC3. By plotting the scores of PC2 and PC4 (see Fig. 6 down), the samples may be grouped in 3 clusters: (i) a first cluster containing samples collected in January (on the left side of the plot) with high input values of phenol, phthalates, pH, nitrates and suspended solids, (ii) a second cluster containing samples collected in December and March (on the upper right side of the plot) with high input values of propanoic acid ester, triisobutyl phosphate, diethyl phthalate, phenols, anionic surfactants and BMWPC; and (iii) a third cluster formed by samples collected in April and June (in the lower right side of the plot) with high input values of trimethylbenzene, tri-*n*-butyl phosphate, dissolved oxygen, suspended solids and total coliforms. In summary, this plot allowed the discrimation of samples according a temporal distribution which was related to changing seasonal fluxes of major contaminants.

CONCLUSIONS

Principal Component Analysis of 94 organic microcontaminants measured in 34 samples at the Llobregat River (Catalonia, Spain) allowed the identification of the main sources of river industrial and domestic contamination in the studied region. Main contaminants were found to be butyl esters, phenol, 3-methyl indol, propanoic acid ester, sitosterol, squalene and specially, plasticizer compounds such as phthalates and phosphates. Simultaneous Principal Component Analysis of the 14 quality parameters and of the 94 organic microcontaminats showed that pollution from ammonium, nitrates and anionic surfactants was correlated with organic matter. Biological (BMWPC) and physico-chemical (ISQA) water quality indexes were confirmed to be good indicators of the general quality of the river water. Temporal and geographical distribution of samples was deduced. Samples collected in January when the river flow was lower than in spring showed a clear contamination pattern following a distribution up to downstream in agreement with the increasing contamination levels in the same direction. The more polluted areas were confirmed to be located at the downstream of the Llobregat River, especially in the Anoia River affluent.

References

[1] M. Valls, J.M. Bayona and J. Albaigés. *Intern. J. Environ. Anal. Chem.*, **39**, 329–348 (1990).
[2] N. Paxéus, P. Robinson and P. Balmér. *Water Sci. Technol.*, **25**, 11, 249–256 (1992).
[3] N. Paxéus. *Water Res.*, **30**, 5, 1115–1122 (1996).
[4] J.O. Grimalt and J. Olivé. *Anal. Chim. Acta.*, **278**, 159–176 (1993).
[5] J. Salau, R. Tauler, J.M. Bayona and I. Tolosa. *Environ. Sci. Tech.*, **31**, 12, 3482–3490 (1997).
[6] R. Tauler, D. Barceló and E.M. Thurman. *Environ. Sci. Tech.*, **34**, 3307-3314 (2000).
[7] B.G.M. Vandeginste, D.L. Massart, L.M.C. Buydens, S. De Jong, P.J. Lewi and J. Smeyers. *Handbook of chemometrics and qualimetrics*. Elsevier Science, Amsterdam, Holland (1998).
[8] J.W. Einax, H.W. Zwanziger S. Geiß. *Chemometrics in Environmental Analysis*. VCH Verlagsgesellschaft, Weinheim, Germany (1997).
[9] H.F. Kaiser. *Psychometrika*, **23**, 187–200 (1958).
[10] US Environmental Protection Agency Method 625. *Guidelines Establishing Test Procedures for the Analysis of Pollutants under the Clean Water Act: Final Rule and Interim Final Rule and Proposed Rule* (1984).
[11] *Standard methods for the examination of water and wastewater*, 20th Edn. American Public Health Association, Washington, D.C. (1998).
[12] R. Queralt, Ll. X. Godé, C. Serena, Ll. Matia, Ll. and R. Ariño. *Technologia del agua*, **143**, 17–24 (1995).
[13] G. Benito, G.M.A. Puig. *Technología del agua*, **191**, 43–56 (1999).
[14] P.D. Armitage, D. Moss, J.F. Wright and M.T. Furse. *Water Res.*, **17**, 333–347 (1993).