Chemometrics modelling of organic contaminants in fish and sediment river samples

Emma Peré-Trepata a, Lourdes Olivella b, Antoni Ginebreda b, Josep Caixach c, Romà Tauler c,*

a Department of Analytical Chemistry, Universitat de Barcelona, Diagonal 647, 08028 Barcelona, Spain
b Agència Catalana de l’Aigua, Provença 204-208, 08036 Barcelona, Spain
c Institute of Chemical and Environment Research, CID-CSIC, Jordi Girona 18-26, 08034 Barcelona, Spain

Received 7 September 2005; received in revised form 12 April 2006; accepted 18 April 2006
Available online 9 June 2006

Abstract

Chemometric methods are applied in the analysis and interpretation of large multivariate data sets obtained in environmental monitoring studies. Multiple fish and sediment river samples from different sampling sites at different geographical locations and during different campaigns and/or sampling time periods were collected and analyzed as a part of an extensive multi annual monitoring program. Concentrations of organic microcontaminants (like hexachlorobenzene, hexachloro-butadiene, hexachlorocyclohexane, dichloro-diphenyldichloroethane and metabolites, and polychlorinated biphenyls PCBs, among others) were determined in fish and sediment river samples from Catalonia (northeast Spain). Chemometrics modeling and interpretation of these data sets allowed the investigation of the main contamination patterns in the rivers under study and the investigation of their distribution. The results achieved in this study are intended to be a contribution to quality assessment and evaluation of the global situation of the contamination of surface waters in Catalonia, and to support public policies of environmental control and protection in the region under study.

© 2006 Elsevier B.V. All rights reserved.

Keywords: Chemometrics modelling; Principal component analysis; Organic contaminants; Fish; Sediments; Catalonia

1. Introduction

Chemometric methods (Massart et al., 1998) provide powerful tools for the modelling and interpretation of large, environmental, multivariate data sets generated within environmental monitoring programs (Einax et al., 1997). The goal of these studies is the computation, screening and graphical display of patterns in large data sets, looking for possible contamination sources and their distribution. Principal component analysis (PCA) is one of these multivariate data analysis methods frequently used in environmental data exploratory studies (Jolliffe, 2003; Wold et al., 1987). PCA allows the transformation and visualization of complex data sets into a new simpler perspective in which the more relevant environmental information can be more easily perceived. Using PCA, contamination patterns may be identified and their geographical and temporal distributions may be investigated. In order to improve interpretation of contamination patterns derived from
PCA, varimax orthogonal rotation of principal components has been proposed (Jolliffe, 2003; Cattel, 1978). Principal component analysis has been applied in previous studies by different authors to different types of environmental data sets: waters, biota and sediments (Manz et al., 2001; Backe et al., 2004; Mujunen et al., 1996; Dietze et al., 2001; Peré-Trepat et al., 2004). Other chemometrics methods have been applied to the investigation of environmental data like PLS (Mujunen et al., 1996; Dietze et al., 2001), MCR–ALS (Tauler et al., 2000), PARAFAC and Tucker3 (Tauler et al., 2004). The use of multivariate factor analysis chemometric methods like those proposed in the present work has also been revised in different books (Massart et al., 1998; Einax et al., 1997; Malinowski, 2002).

In the present work, the case of study chosen was a large environmental data set obtained in a study of fish and sediment samples from natural surface waters in rivers of Catalonia River Basin Area (north east of the Iberian Peninsula) and including multiple organic contaminants. In the frame of this extensive multi annual environmental monitoring program from the Water Catalan Agency, 80 sediment samples and 102 fish samples from the entire geographical area of Catalonia were analyzed during the years 1998–2001. The presence of organic compounds in natural surface waters is attributed to the presence of different industrial, agricultural and urban wastewater point and diffuse contamination sources. Most of the compounds studied in this work (see in Experimental data section a detailed list) are under EU priority lists of compounds (like Directives 76/464/EEC, 2000/60/EC and 2455/2001/CE) and should be eliminated or substantially reduced because of their toxicity, environmental persistence or bioaccumulation. Catalonia geographical area is one of the most industrialized areas of Spain, and it is of interest to evaluate the environmental situation of this area. Although this investigation gives results which could be considered only of concern for the particular area under study, the obtained results and conclusions are of general interest from an environmental point of view to other river basin areas, specially those which are close to Mediterranean coasts, which have the same type of climate, hydrology, vegetation and human activities (industrial, agricultural, urbanization) operating over the river water systems. This work is also of interest from a chemometrics point of view, specially for the comparison of results obtained by PCA, and varimax rotation methods, often used in the environmental studies (Cattel, 1978; Einax et al., 1997; Massart et al., 1998). Some other recent publications are concerned about the presence of persistent organic compounds in Catalonia (Eljarrat et al., 2001; Calvo et al., 2002; Larrazabal et al., 2003; Teixidó et al., 2001), which were also analyzed in different types of environmental compartments. Also, some other recent examples exist proposing rather similar approaches for the resolution and interpretation of major contamination sources of surface waters operating over different river basins over the world (Fharnham et al., 2002).

Summarizing, the two main objectives of this work have been: (a) the investigation of main long term diffuse contamination sources of organic microcontaminants in this case, in the Catalonia River Basin Area; and (b) the deduction of their geographical distribution and of their accumulation effects in sediment and fish samples to contribute to the evaluation of the environmental health of the surface waters of the region under study. To achieve these two goals, multivariate data analysis (chemometric) methods based on principal component analysis (PCA) and on varimax orthogonal rotation are applied and compared.

2. Experimental data

Sampling was performed from spots previously decided by the Catalan Water Agency. Sediments were manually collected from the river and divided into two parts. The P destined to organic analysis was kept in an aluminium foil, while that destined to metal analysis was contained on a plastic polyethylene bag. Both parts were kept at 4 °C during the transport, and stored at −10 °C until their final analysis. Fish were caught by electric fishing, identified, kept in cold for transport (4 °C) and stored frozen as above, until the analysis. Specimens were weighed, measured, and subjected to macroscopic examination in order to assess their general sanitary status. Fish samples were ground, homogenized and stored at −10 °C prior to chemical analysis.

102 fish samples were collected at the 45 different geographical sites given in the map of Fig. 1. There were different types of fish such as: barbel, catfish, bleak, carp and trout, which can be separated in two main groups: salmonids (trout) and cyprinids (the other four). 80 sediment samples were collected at 36 geographical sites (see map and sampling site allocation in Fig. 1) close to fish sampling sites. In the first years, more than one fish was taken in every site, every year. Later, fish and sediment samples were taken only one time per year, and not every site was studied every year. Fish and sediment samples were obtained over a period of four years, starting at 1998 until 2001. The geographical area under study where these compounds were analyzed...
covered different small and medium size rivers in the Catalonia region such as (from North to South Catalonia coast): Muga, Fluvià, Ter, Tordera, Besòs, Llobregat, Foix, Gaià, Francolí, Segre, Ebre and Garona rivers (see location of these rivers in Fig. 1). With exception of the last three rivers, the others are typical Mediterranean rivers, characterized by short length and small catchment areas, steep slopes and drastic flow variations between the dry summer season and sudden flow increases after the fall and spring rains that often cause floods and damages.

These data sets have been analyzed as they were provided by the Catalan Water Agency and no attempt was performed to have an optimal design of the best sampling sites for the purpose of environmental source identification. It should be noted therefore, that because of these sampling limitations, information about temporal evolution of the river contamination sources in Catalonia rivers could only be obtained in a limited way. These provisory results should be confirmed with new data obtained using a better designed monitoring sampling plan including more recent years. Some work is pursued at present in this direction.

2.1. Analysis of organic microcontaminants

The analytical methodology used to determine PCBs and organochlorine pesticides is based in the US-EPA 1668A method. Soxhlet extraction was made with n-hexane/dichloromethane for 24 h. Prior to the extraction,
sample was spiked with $^{13}$C-labeled internal standards. Cleanup was made with Florisil® and Alumina 5% deactivate for fish and sediments, respectively. The purified extracts were analyzed by HRGC/HRMS using an Autospec-Q mass spectrometer (Micromass, Manchester UK) coupled to HP gas chromatograph (Palo Alto, CA, USA). The capillary column used was a DB-5 (J and W Scientific, Folsom, CA, USA) fused silica (60 m × 0.25 mm ID, 0.28 μm film thickness). The injection mode was split-less and helium was used as carrier gas. The operating conditions of mass spectrometer were: electron ionization (70 eV ionization energy), acquisition mode SIM and minimum static resolution used were 8000 (10% definition valley). Accuracy and precision of methods were evaluated by analyzing certified reference materials (SRM 1944 and SRM 1939a) and by participation in intercalibration exercises. The Water Agency is accredited by ISO9001 and by the more specific ISO17025 for some of the analytical techniques used in this work.

The total number of variables was 19 including the following compounds and total sums of them: HCBu (hexachlorobutadiene), HCB (hexachlorobenzene), HCH (total hexachlorocyclohexanes, α, β and γ isomers), $o,p$-DDE ($ortho,para$-dichlorodiphenyldichloroethylene), $p,p$-DDE ($para,para$-dichlorodiphenyl-dichloro-ethylene), $o,p$-DDD ($ortho,para$-dichlorodiphenyl-dichloroethane), $p,p$-DDD ($para,para$-dichlorodiphenyl-dichloroethane), $o,p$-DDT ($ortho,para$-dichlorodiphenyl-trichloroethane), $p,p$-DDT ($para,para$-dichlorodiphenyl-trichloroethane), DD X (total dichlorodiphenyl chloroethanes), POCs (total persistent organic compounds) and seven BCR indicator polychlorinated biphenyls (no. IUPAC PCB#28, PCB#52, PCB#101, PCB#118, PCB#138, PCB#153, PCB#180), and total PCBs.

3. Chemometrics methods

3.1. Data pretreatment methods

Data sets were organized in two data tables or data matrices. Matrix FO for organic compounds in fish data samples and matrix SO for organic compounds in sediment samples. The rows of these two data matrices identified the different fish and sediment samples at the different geographical sites and sampling dates. The columns (variables) of these two data matrices have the different analyzed chemical compounds. As it is explained below these two data sets are analyzed individually and simultaneously in a column-wise augmented data matrix formed by concatenation of the fish matrix on top of the sediments matrix and keeping the columns or variables (same analyzed organic compounds) in common. The dimensions of the two individual data matrices were: FO (102×19) and SO (80×19). The dimensions of the augmented data matrix [FO; SO] were 182×19.

In determining the best data pretreatment method, a compromise was sought to find the method which provided the easiest and more optimal interpretation of possible contamination sources. The following data pretreatment methods were performed: (a) Estimation of missing data, using PCA imputation functions in PLS Toolbox 3.5 software (Eigenvector Research, Mason, WA, USA). (b) Values below limit of detection were replaced by half of their limit of detection (Farnham et al., 2002). (c) Units of concentration for organic compounds for fish and sediment samples were selected to be picograms ($10^{-12}$ g, pg) in one gram of sample, i.e. pg/g. This allows application of log transformation without creating negative values. (d) Concentration values were log-transformed, i.e. decimal logarithm of all data matrix values was calculated. Log transformation of experimental data has been recommended for skewed data sets (Grimalt et al., 1993; Zitko, 1994), like those usually found in environmental studies where the majority of the values are low values with a minor contribution of high values. (e) Log concentration values of each compound in the different samples were mean centered, i.e. each log concentration value was subtracted by the mean of the log concentration values of the same compound in the different samples (mean of each column variable of the data matrix). (f) Log concentration values of each compound in the different samples were scaled, i.e. each log concentration value was divided by the standard deviation of the log concentration values of the same compound in the different samples (standard deviation of each column variable of the data matrix). (g) Log concentration values of each compound in the different samples were autoscaled, i.e. previous mean centering and unit variance scaling pretreatment methods were combined.

The two experimental data matrices described in the Experimental data section were first analyzed individually and then simultaneously by principal component analysis (PCA) and by varimax orthogonal rotation.

3.2. Principal component analysis (PCA)

PCA assumes a bilinear model to explain the observed data variance using a reduced number of components. See previous references (Jolliffe, 2003; Wold et al., 1987) for a detailed description of this well-known methodology in chemometrics and other
multivariate statistical data analysis methods. This bilinear decomposition may be written by the element-wise equation:

\[ d_{ij} = \sum_{n=1}^{N} u_{in} v_{jn} + e_{ij} \]  (1)

where \( d_{ij} \) is one of the entries of the experimental data matrix (measure of concentration of one organic compound) from the \( i \)th row (a particular sample) and the \( j \)th column (variable or considered organic compound), \( u_{in} \) is the corresponding \( n \)th score element for the sample \( i \), \( v_{jn} \) is the corresponding \( n \)th loading element for the variable \( j \) and \( e_{ij} \) is the residual not modelled by the sum of \( N \) components or contributions. The same bilinear equation can be written in matrix form as:

\[ \mathbf{D} = \mathbf{U} \mathbf{V}^T + \mathbf{E} \]  (2)

where \( \mathbf{D} \) is the experimental data array arranged in a data matrix. Eq. (2) describes the decomposition (matrix factorization) of matrix \( \mathbf{D} \) on two matrices, the loadings matrix \( \mathbf{V}^T \) and the scores matrix \( \mathbf{U} \). The loadings matrix \( \mathbf{V}^T \), identify the main sources of data variance by means of their chemical composition (composition loadings), which eventually may be related with the main patterns and sources of contamination. The scores matrix \( \mathbf{U} \), provides sample scores for these data variance patterns (i.e. it will indicate the geographical and temporal sample distribution of these patterns). PCA solves Eq. (2) under orthogonal constraints. Each successively extracted principal component explains maximum variance. The determination of the complexity of the model in PCA (i.e. the number of principal components) is performed as a compromise between different goals, model simplicity (few components), maximum variance explained by the model (more components), and model interpretability.

3.3. Varimax orthogonal rotation

Varimax orthogonal rotation of principal components obtained by PCA is proposed as a method to facilitate interpretability of PCA loadings. Once varimax orthogonal rotation is performed, factor loadings keep their orthogonality but they are not directed anymore towards directions of maximum explained variance nor the scores are orthogonal, i.e. they overlap. Varimax rotation was performed using Kaiser’s VARIMAX algorithm for orthogonal rotation of principal components which maximized the total loadings simplicity (Jolliffe, 2003; Cattel, 1978).

Most of data treatments, algorithms, computer programs and graphics were developed and performed under MATLAB computer and visualization environment (Release 12, The Mathworks, Natick MA, USA). For PCA and varimax rotation, PLS Toolbox v3.5 (Eigenvector Research, Mason, WA, USA) was used.

4. Results and discussion

Statistic parameters of the studied organic compounds are summarized in Table 1. Concentration distributions of organic compounds among samples were strongly biased towards low values as it is clearly seen from the large observed differences between mean and median values. Median values were always lower than mean values, in both fish and sediment samples. DDE compounds are usually at higher mean and median concentrations in fish than in sediment samples, whereas DDT compounds show the opposite trend, with their mean and median concentration values higher in sediment samples than in fish samples. DDD compounds show an intermediate behavior. HCB had higher mean and lower median concentration values in sediment than in fish samples, with large variability among samples (large standard deviations). For PCBs, considerably larger mean and median values were observed for the fish samples than for the sediment samples. Generally the accumulation of organic compounds occurs differently in sediment samples than in fish samples depending on the type of considered compound. Some organic compounds (like PCBs and \( p \), \( p\)-DDE) accumulated more in fish samples, while others (DDT, POCs and HCH) seem to be more accumulated in sediment samples. It should be taken into account however, that obtained values for fish samples reflect average fish concentrations, since sampling did not differentiate among fish tissues.

PCB concentrations were highly correlated in both fish and sediment samples; in many circumstances the correlation coefficients were close to one, especially for sediment samples, demonstrating a major common pattern of distribution of all of them. Other interesting high pairwise correlations were also observed for other individual organic compounds. Hexachlorobenzene (HCB) in fish samples was highly correlated with many of the other organochloride compounds (like DDE, DDD and DDT) and POC for example, but it was not correlated at all with the sum of hexachlorocyclohexanes (HCH), nor with \( p \cdot p\) -DDT. This last compound showed a rather low concentration and independent behaviour in fish samples. In general, organic compounds not included in the PCBs family, gave lower
pairwise correlations in sediment samples than in fish samples. Multivariate correlations are better investigated by means of principal component analysis and will be discussed below.

For PCA calculations, the results shown correspond to the selected data pretreatment method finally applied for the two data matrices FO and SO, which was log transformation followed by column autoscaling. These two matrices were analyzed individually and simultaneously (see data pretreatment). In Table 2, PCA explained variances obtained in each case are given. The use of log transformation previous to autoscaling allowed an improved analysis and interpretation of the main sources of data variance, specially in sample scores plots. When data were not log-transformed before autoscaling, results were extremely influenced by the more contaminated fish and sediment samples of Flix (sampling sites 66 in map of Fig. 1). These samples were so influent that the rest of samples appeared strongly clustered in a very small region close to the coordinates centre of the scores plot and without any distinction among them, even after autoscaled. See below for the comparison of results obtained with and without log data transformation when both type of samples (fish and sediments) were simultaneously analyzed.

4.1. Organic compounds in fish river samples (FO data matrix)

When principal component analysis (PCA) was applied to the investigation of concentrations of organic compounds in fish samples (log autoscaled FO data matrix, 102 samples × 19 variables), five principal components explained as much as 85.3% of data variance. In the upper half of Fig. 2, loadings of the two first components are given. PC1 (51.7% of explained data variance in Table 2) describes a general diffuse contamination pattern for most of the organic compounds. PC2 (17.5% of explained data variance) had negative loadings for PCBs and positive loadings for DDEs. In Table 2, the explained variances for each component are given for fish and sediments separately and for fish and sediments combined.

Table 1
Summary of descriptive statistic parameters for organic compounds concentrations (units 10⁻¹² g/g or pg/g) referred to wet sample for fish and to dry sample for sediments

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>S.D.</th>
<th>Max. value</th>
<th>Min. value</th>
<th>N samples analyzed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fish</td>
<td>Sediments</td>
<td>Fish</td>
<td>Sediments</td>
<td>Fish</td>
<td>Sediments</td>
</tr>
<tr>
<td>HCBu</td>
<td>790</td>
<td>1617</td>
<td>75</td>
<td>200</td>
<td>3130</td>
<td>6905</td>
</tr>
<tr>
<td>HCB</td>
<td>5425</td>
<td>22628</td>
<td>693</td>
<td>400</td>
<td>28011</td>
<td>140971</td>
</tr>
<tr>
<td>Total HCH</td>
<td>6303</td>
<td>6801</td>
<td>1279</td>
<td>1000</td>
<td>18541</td>
<td>27972</td>
</tr>
<tr>
<td>o,p-DDE</td>
<td>1788</td>
<td>1981</td>
<td>35</td>
<td>250</td>
<td>11002</td>
<td>13446</td>
</tr>
<tr>
<td>p,p-DDE</td>
<td>46544</td>
<td>11020</td>
<td>14100</td>
<td>950</td>
<td>137115</td>
<td>56232</td>
</tr>
<tr>
<td>o,p-DDD</td>
<td>2509</td>
<td>2176</td>
<td>251</td>
<td>325</td>
<td>12026</td>
<td>7295</td>
</tr>
<tr>
<td>p,p-DDD</td>
<td>6096</td>
<td>4327</td>
<td>1929</td>
<td>500</td>
<td>20012</td>
<td>11409</td>
</tr>
<tr>
<td>o,p-DDT</td>
<td>2166</td>
<td>7658</td>
<td>35</td>
<td>125</td>
<td>12612</td>
<td>45839</td>
</tr>
<tr>
<td>p,p-DDT</td>
<td>7086</td>
<td>64631</td>
<td>99</td>
<td>500</td>
<td>59353</td>
<td>425858</td>
</tr>
<tr>
<td>Total DDX</td>
<td>66190</td>
<td>91771</td>
<td>18103</td>
<td>4725</td>
<td>199733</td>
<td>492404</td>
</tr>
<tr>
<td>Total POCs</td>
<td>78709</td>
<td>122691</td>
<td>23835</td>
<td>7400</td>
<td>229549</td>
<td>635345</td>
</tr>
<tr>
<td>PCB#28</td>
<td>7600</td>
<td>1346</td>
<td>2100</td>
<td>400</td>
<td>23056</td>
<td>5836</td>
</tr>
<tr>
<td>PCB#52</td>
<td>7070</td>
<td>1364</td>
<td>3267</td>
<td>900</td>
<td>12946</td>
<td>3414</td>
</tr>
<tr>
<td>PCB#101</td>
<td>13412</td>
<td>3126</td>
<td>6500</td>
<td>1468</td>
<td>24155</td>
<td>5476</td>
</tr>
<tr>
<td>PCB#118</td>
<td>9339</td>
<td>2461</td>
<td>5700</td>
<td>907</td>
<td>12912</td>
<td>4083</td>
</tr>
<tr>
<td>PCB#138</td>
<td>21669</td>
<td>4702</td>
<td>10400</td>
<td>1720</td>
<td>48821</td>
<td>8621</td>
</tr>
<tr>
<td>PCB#153</td>
<td>27239</td>
<td>4066</td>
<td>12700</td>
<td>1465</td>
<td>67901</td>
<td>8722</td>
</tr>
<tr>
<td>PCB#180</td>
<td>18991</td>
<td>3171</td>
<td>6700</td>
<td>1257</td>
<td>64120</td>
<td>5787</td>
</tr>
<tr>
<td>Total PCBs</td>
<td>104296</td>
<td>21031</td>
<td>51800</td>
<td>9400</td>
<td>233827</td>
<td>35794</td>
</tr>
</tbody>
</table>

Statistic parameters (mean, median, standard deviation (S.D.) and maximum and minimum values) for organic compounds in fish and sediment samples are given. In every column, two values are shown; the first one is for fish samples and the second one is for sediments. The number of samples analyzed is before missing values were estimated (see Chemometrics methods).

4.1. Organic compounds in fish river samples (FO data matrix)

When principal component analysis (PCA) was applied to the investigation of concentrations of organic compounds in fish samples (log autoscaled FO data matrix, 102 samples × 19 variables), five principal components explained as much as 85.3% of data variance. In the upper half of Fig. 2, loadings of the two first components are given. PC1 (51.7% of explained data variance in Table 2) describes a general diffuse contamination pattern for most of the organic compounds. PC2 (17.5% of explained data variance) had negative loadings for PCBs and positive loadings for DDEs. In Table 2, the explained variances for each component are given for fish and sediments separately and for fish and sediments combined.

Table 2
% Explained variances by PCA and varimax

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PCA b Varimax b</td>
<td>PCA b Varimax b</td>
<td>PCA b Varimax b</td>
</tr>
<tr>
<td>1</td>
<td>51.7</td>
<td>37.7</td>
<td>52.2</td>
</tr>
<tr>
<td>2</td>
<td>17.5</td>
<td>31.5</td>
<td>20.4</td>
</tr>
</tbody>
</table>

a [FO; SO] Augmented data matrix (182 × 19).
b Log transformation + autoscaled pretreatment.
for the other organic compounds (HCB, HCH, DDX, total POCs). Thus, PC2 is describing the contrast between the group of PCBs and the group of the remaining organic compounds. When considered together, PC1 and PC2, already explained a large amount of the data variance (close to 70%), describing the two main contamination patterns of organic compounds operating over fish samples in the investigated geographical region. PC3 and PC4 loadings (not shown in Fig. 2) explained smaller independent contamination patterns for $p,p$-DDT and HCHs, correlated either positively or negatively between them, and with other compounds like HCBu in PC3, and with $o,p$-DDT in PC4. PC5 and higher described less than 5% of variance and are not further discussed since they refer mostly to individual specific patterns and experimental data noise.

Varimax PC1 and PC2 rotated loadings obtained in the analysis of the same FO data matrix (also in the lower half of Fig. 2) provided a confirmatory view of the two main contamination patterns of organic compounds operating over fish samples already detected for the straight PCA analysis (without factor axes rotation) giving above. First varimax component explained approximately a 38% of data variance and it had mostly only contributions from the PCBs group of compounds. This component described most of the variation observed for these compounds in fish samples and it is the more important contamination pattern detected on them. Second varimax component explained approximately a 31% of data variance and described the contamination coming from the other organic compounds. Distinction between these two contamination patterns (PCBs and other organic compounds) is enhanced after varimax rotation of principal components. Fig. 3 gives PCA and varimax scores and loadings biplot for the two first components, with fish samples marked with different symbols depending on their acquisition year. Whereas PC1 and PC2 PCA axes are represented by orthogonal broken lines, PC1 and PC2 varimax axes are represented by rotated orthogonal continuous lines in the same plot. This figure shows that the most contaminated fish samples were 66 (Ebre river in Flix), 41 (Llobregat river close to Barcelona) and 36 (Llobregat river close to Berga), all of them on the right side on the PCA axes, indicating large positive scores of PC1 (general contamination). See map of Fig. 1 for
samples identification. It is also possible to detect that many of the 1998 and 1999 fish samples had scores situated on the right half of the biplot because of their higher concentration levels of organics, and specially of PCBs (more in the right down quadrant). And that many of the 2000 and 2001 samples gave most of the scores situated in the upper left quadrant of PCA axes because of their lower concentration levels of organics and PCBs. However, because samples from different years were not always acquired in the same locations, it is difficult to extract definitive conclusions about the tendency over years on decreasing contamination trends by organic compounds in fish samples. This possible decreasing trend would be in agreement with the current regulation of PCBs, whose production and use has been discontinued since 1977 in the USA and since 1986 in Catalonia. This conclusion, however, should be confirmed with a more systematic data acquisition and analysis in the more recent years. Concerning HCB, HCH, DDX and total POCs, practically no changes were observed on average with the years. Similar conclusions can be deduced either using PCA axes or rotated varimax axes, although in the later case it is easier the distinction among those fish samples more contaminated by PCBs from those fish samples more contaminated by the rest of organic compounds. Fish samples 66 from Flix in 1998, appear to be the more highly contaminated fish samples by the two families of compounds since they are situated in the extreme of the first quadrant (upper right diagonal) of varimax axes. On the contrary, fish samples from the Pyrenees region are mostly situated in the third (down left) varimax quadrant, indicating that they are rather unpolluted for the investigated organic compounds. See for instance fish sample 32 which was obtained close to the headwaters of Llobregat river.

In Fig. 4, PC1 vs. PC2 PCA and varimax scores plot is given as a function of the type of fish. Salmonid type of fish (solid triangles), commonly only present in river headwaters, should give relatively lower contamination levels of organic compounds. Cyprinids (solid squares), which are generally located in middle river watercourses and at river mouths, should have significantly higher contamination levels of these...
organic compounds. This separation as a function of the type of fish is confirmed for most of samples along PC1, which is determined approximately by all the organic compounds. Observe that most of the salmonids (solid triangles) are on the left side of the scores plot, and that very few of them are situated in the more contaminated quadrants either for PCA or varimax axes (right side of the scores plot).

4.2. Organic compounds in sediment river samples (SO data matrix)

When principal component analysis (PCA) was applied to the investigation of concentrations of organic compounds in sediment samples (log autoscaled SO data matrix, 80 samples × 19 variables), the first five principal components explained 86.7% of data variance. Like for fish samples, the first two principal components already described a great amount of data variance (72.6%) and their loadings are given in Fig. 5 (upper half). PC1 (52.2% of explained data variance in Table 2) described a general diffuse contamination pattern where all organic compounds had large contributions and all were positively correlated among them. And also like for fish samples, PC2 (20.4% of explained data variance in Table 2) allowed the differentiation between contamination from PCBs from contamination from other organic compounds. Therefore, a first very important conclusion of this work is that for both, fish and sediment samples (FO and SO data matrices), PC1 and PC2 described nearly identical contamination patterns of organic compounds, with also very similar percentages of explained variances (see Table 2). When Figs. 3 and 5 are compared, it is clear that PC1 and PC2 loading plots for FO and SO data sets are very similar.

Varimax PC1 and PC2 rotated loadings obtained in the analysis of the same SO data matrix (Fig. 5, lower half) confirms these results. First varimax component explained approximately a 38% of data variance and it has mostly only contributions from the PCBs group of compounds. It gives a loadings profile very similar to first varimax loadings profile obtained for fish samples (Fig. 3). The second varimax component explained approximately 34% data variance and describes the contamination coming from the other organic compounds. This separation as a function of the type of fish is confirmed for most of samples along PC1, which is determined approximately by all the organic compounds. Observe that most of the salmonids (solid triangles) are on the left side of the scores plot, and that very few of them are situated in the more contaminated quadrants either for PCA or varimax axes (right side of the scores plot).

Fig. 4. Fish (FO matrix) PCA and varimax PC1 vs. PC2 scores plot as a function of the type of fish.
compounds. Like for fish samples, distinction between these two contamination patterns (PCBs and other organic compounds) is enhanced after varimax rotation of the two first principal components.

Fig. 6 gives PCA and varimax scores and loadings biplot for the two first components, with sediment samples marked with different symbols depending on their acquisition year. Like for fish samples in Fig. 3, PC1 and PC2 PCA axes are represented by orthogonal broken lines, and PC1 and PC2 varimax axes are represented by rotated orthogonal continuous lines in the same plot. This figure shows again that the most contaminated sediment samples were acquired at Ebre river in Flix (66), mouth of Besòs and Llobregat rivers close to Barcelona (31, 42) and Ter river in Manlleu (13), Ter Manlleu, in a region where intensive farming and industry is known to have a rather high environmental impact in surface and underground waters, sediments and soils. Few sediment samples were taken in river upper stream locations, but in general it is confirmed that less contaminated sediment samples (on the left of the plot and/or third varimax quadrant) are in the more remote areas away from urban, intensive agriculture and industrial anthropogenic inputs. PCB contamination in sediment samples showed no particular annual trend. That seemed to indicate that the last years tendency of possible decrease of contamination by PCBs observed for fish samples was not observed for sediment samples yet. However, the lack of a more systematic sampling design over the years and over the whole geographical area under study did not allow us to extract more reliable temporal conclusions.

4.3. Simultaneous analysis of organic compounds in fish and sediment river samples ([FO; SO] augmented data matrix)

Results of PCA applied simultaneously to fish and sediment samples ([FO; SO] column-wise augmented data matrix) are given in Table 2 and in loadings and scores biplot of Fig. 7. Fish are plotted in solid circles and sediments in solid triangles. Information about contamination patterns already described by PCA in
previous sections in the individual analysis of each compartment (fish or sediment) was confirmed here in a more synthetic way, when PCA was applied simultaneously to fish and sediment samples (but separately autoscaled after log data transformation). Like for the individual analysis of fish and sediment samples, the two fist principal components explained a large amount of data variance (70.1%), confirming again the high similarity between the identity and nature of the two main contamination patterns operating over fish and sediment samples investigated in this work. The fist component describes a general contamination pattern with contribution of all variables (Fig. 7) and with an explained variance of 51.6%, similar to that explained individually for fish (51.7%) and sediments (52.2%). This first component allows for a ranking of a less to more contaminated sediment and fish samples, and also allows for the comparison among them (see below). Samples from Ebre River at Flix (66) both for fish and sediments were rather similar and the more contaminated ones, followed by those from the mouth of Llobregat and Besòs rivers close to the Barcelona industrial area (41, 31) and from middle Ter river in Manlleu (13). The second component, explains a similar percentage of explained variance (18.5%) than that found in the individual PCA analysis (17.5% for FO, and 20.4% for SO), and it discriminates (Fig. 7) between samples more contaminated by PCBs from samples more contaminated by the other group of variables under study (HCBu, HCB, total HCH and DDXs). First varimax rotated principal component (37.3% of explained variance) is now explaining mostly only PCBs variance and second varimax rotated principal component (32.8%) is explaining variance from the rest of organic compounds.

Also interesting is the comparison between Figs. 7 and 8, both for simultaneous analysis of fish and sediment samples by PCA using different data pretreatments. In Fig. 7, log-transformed data matrices from each compartment, fish (FO) and sediments (SO), were first autoscaled independently and then concatenated in the column-wise augmented matrix [FO;SO] before PCA. In Fig. 8, the same log-transformed data from

---

**Fig. 6.** Sediment (SO matrix) PCA and varimax PC1 vs. PC2 scores and loadings biplot. PCA axes in broken lines and varimax axes are in continuous lines. Sample scores are plotted by different symbols and colours according to sampling year and distinguished from loadings (x).
each compartment, fish and sediments, were now only scaled (not mean-centered) instead of autoscaled (as in Fig. 7) and then concatenated to give the augmented data matrix and before PCA. When the two plots, Figs. 7 and 8, are compared the effect of mean-centering appears clearly. Whereas no differences were observed between fish and sediment samples for autoscaled log data (symbols for fish samples and for sediments are intermixed over the whole plot of Fig. 7), a rather clear distinction was obtained instead for the two type of samples when log-transformed data were only scaled and not mean-centred. In this case, most fish samples (solid circles) appear on the lower right hand of the plot and most of the sediment samples (solid triangles) on the upper left hand of the plot. The reason for that is that mean-centering removes the differences among fish and sediment samples which are related with the mean content of the analyzed compounds. Since PCBs gave larger negative loadings in PC2 also for scaled log data (results not shown), this is interpreted saying that in fish samples, PCB compounds accumulated at higher concentrations than in sediments (Fig. 8) although in terms of variance within each group the same patterns of concentration changes were observed (Fig. 7).

Finally, also interesting is the comparison between Figs. 7 and 9. Whereas in Fig. 7, fish and sediment data were log-transformed before autoscaling and PCA, in Fig. 9, the same data were not log-transformed before autoscaling and PCA. In the latter case, when data were not log-transformed, the much larger scores of samples from Flix (66), obscure the display of other differences among the rest of samples because they appeared nearly all clustered close to coordinates origin, not allowing any other data interpretation nor extraction of further information apart from the extreme behaviour of samples (no. 66 in map of Fig. 1) from Ebre river in Flix, which could be also derived from the direct observation of the data. These results prove that log transformation pretreatment was adequate in this case, since it allowed the identification of similar contamination patterns (loadings) than without log transformation but with the advantage of giving a more clear distribution of samples (scores) according to these.

Fig. 7. Fish and sediment ([FO:SO] augmented data matrix, 182 fish and sediment samples × 19 variables) PCA and varimax PC1 vs. PC2 scores and loadings biplot for autoscaled log data. PCA axes in broken lines and varimax axes are in continuous lines. Sample scores are plotted by different symbols and colours according to type of sample fish or sediment and distinguished from loadings (x).
contamination patterns in a single PCA and interpretation of fewer plots.

5. Conclusions

The most relevant conclusions extracted by application of the chemometrics methods used in this work can be summarized as follows: (1) It has been demonstrated that the same main contamination patterns of organic compounds (of anthropogenic origin) operate on fish and sediment samples collected during this study. (2) It has been possible to identify and differentiate a main contamination pattern of PCBs, with higher accumulations for fish than for sediment samples. This is particularly significant considering the disparity and large number of samples analyzed during this investigation, covering different years, geographical spots and diverse environments (industrial, agricultural, and rural areas). (3) It has been also possible to identify and differentiate a second major pattern of the organic microcontaminants including compounds like hexachlorobutadiene, hexachlorobenzene, hexachlorocyclohexans and the DDX family of organic compounds. (4) Some evidence is obtained about an annual decreasing trend of PCBs contamination in fish samples, specially in the more recent years. (5) It has been possible to corroborate that salmonid types of fish, usually present in headwater samples, were less contaminated by PCBs and other organic compounds than cyprinid types of fish. (6) The more heavily contaminated spots (marked and numbered in map of Fig. 1) were in low stream Ebre river (Flix), in low streams Llobregat and Besòs rivers (Barcelona), and in mid stream Ter river (Les Masies de Roda and Manlleu). (7) Although the results obtained by the different chemometric methods used in this work (PCA and varimax rotation) are slightly different, enhancing different aspects of the data sets under study, their final interpretation is in agreement whatever is the approach used.

Results obtained in this work are intended to be an initial contribution for a more exhaustive investigation about geographical distribution of contamination patterns, and about their correlation with land uses and
point contamination patterns operating over the geographical area under study.

Acknowledgments

Research project No. BQU2003-00191 from MCyT, Spain, is acknowledged. E. Peré-Trepat thanks her PhD grant from Universitat de Barcelona and Catalan Water Agency (Agència Catalana de l’Aigua).

References


Fig. 9. Fish and sediment (FO:SO) PCA PC1 vs. PC2 scores and loadings biplot for autoscaled data (non-log-transformed). Sample scores are plotted by different symbols and colours according to type of sample fish or sediment and distinguished from loadings (×).


