**Application of chemometric methods to the investigation of main
microcontaminant sources of endocrine disruptors in coastal and harbour
waters and sediments**

Emma Peré-Trepat¹, Mira Petrovic², Damià Barceló² and Romà Tauler^{1,*}

¹Department of Analytical Chemistry, University of Barcelona, Diagonal 647, 08028 Barcelona, Spain

² Department of Environmental Chemistry, IIQAB-CSIC, Jordi Girona 18-26, 08034 Barcelona, Spain

[*roma@apolo.ubi.es](mailto:roma@apolo.ubi.es)

**Application of chemometric methods to the investigation of main
microcontaminant sources of endocrine disruptors in coastal and harbour
waters and sediments**

Emma Peré-Trepat¹, Mira Petrovic², Damià Barceló² and Romà Tauler^{1,*}

¹Department of Analytical Chemistry, University of Barcelona, Diagonal 647, 08028 Barcelona, Spain

² Department of Environmental Chemistry, IIQAB-CSIC, Jordi Girona 18-26, 08034 Barcelona, Spain

*roma@apolo.qui.ub.es

ABSTRACT

Identification, resolution and distribution of main microcontaminant sources of endocrine disruptors in harbour and coastal waters and sediments of Spain are investigated using chemometric methods. Eighteen different endocrine disruptor chemical compounds including nonionic surfactants, their degradation products, and linear alkylbenzene sulfonates, found in a total number of 74 samples, (35 water samples and 39 sediment samples), corresponding to a period of 16 months, from March 1999 to July 2000, and in 32 different geographical sites along the Mediterranean Spanish coast such as the Barcelona, Tarragona, Almeria harbour, Malaga and the Bay of Cadiz were investigated. Main contamination sources of these endocrine disruptor compounds were investigated and interpreted according to their chemical composition and according to their resolved geographical distribution profiles.

KEYWORDS: Principal Component Analysis (PCA), Alternating Least Squares Resolution (ALS), Surfactants, Degradation products, Coastal waters, Marine sediments.

INTRODUCTION

Chemometric methods [1] provide powerful tools for the analysis and interpretation of large, environmental, multivariate data sets generated within environmental monitoring programs [2]. The goal of Chemometrics when applied to this type of data sets is to derive environmental knowledge. Principal Component Analysis [PCA, 3, 4] is one of these multivariate statistical methods frequently used in exploratory data analysis. PCA allows the transformation and visualization of complex data sets into a new perspective in which the more relevant information is made more obvious. Using PCA, contamination sources may be identified and their geographical and temporal distributions investigated. Multivariate Curve Resolution using non-negative Alternating Least Squares (MCR-ALS) is a multivariate method [5, 6] designed for similar tasks, and may be considered a complementary tool to PCA for the improved identification and resolution of the main contamination sources including their composition profiles and their geographical and temporal distributions [7, 8, 9].

In the present work, main sources of endocrine disruptors like non-ionic surfactants, their degradation products and linear alkylbenzene sulfonates in harbour and coastal waters of Spain have been investigated. Surface-active compounds used in industrial processes and in households have one of the highest production rates among all organic synthetic compounds. The total quantity of surfactants produced all over the world in 1996 was more than 10 million metric tons with a predicted increase of 3.6% for the year 2005 [10]. Approximately one half of this production is emitted via wastewater treatment plants (WWTPs) into surface waters. Directive 91/271/European Economic Community (EEC) [11] indicates that all wastewater in the territories of the European Union have to be properly collected and subjected to secondary (biological treatment with secondary settlement) or equivalent treatment before being discharged into estuaries or coastal waters. However, in spite of this European Directive, a significant portion of WWTP sludge is still discharged directly into surface waters. In particular, Spain is one of the European countries that still discharges untreated wastewaters and sewage sludge to the sea. For instance, in 1998, Spain discharged 57.000 tons of sludge in surface waters. There have been several studies on the occurrence of nonionic surfactants, overall APEO and their degradation products, in rivers and lakes [12-17], estuaries and coastal areas [18-25]. In Spain the number of studies generated about

nonionic surfactants in the marine environment is very low [26]. The number of studies about linear alkylbenzene sulfonates (LAS) and their carboxylic degradation products in Spanish coastal waters is somewhat higher [27-31]. The endocrine disruptor compounds studied (shown in Fig. 1) in this work were: alcohol ethoxylates (AEO), nonylphenol ethoxylates (NPEO), coconut diethanolamides (CDEA), polyethyleneglycol (PEG), nonylphenol (NP), octylphenol (OP) and linear alkylbenzene sulfonates (LAS). Whereas LAS contamination is usually originated in clothes washing up liquid detergents, alcohol ethoxylates contamination usually comes from dishes washing up liquid detergents and coconuts. Diethanolamine ion products are related with personal care products. NPEO is used as pesticide adjuvant, paint additives or antifoamers. Aerobic and anaerobic degradation of NPEO yields short ethoxy chain like nonylphenol (NP). Alkylphenol ethoxylates yield octylphenol ethoxylates (OP). Polyethyleneglycol is formed in manufacturing process of ethoxylated surfactants, and it is present in large amounts because it is too expensive to be separated from the final products. PEG is a degradation product of AEO too. Some of these surfactants biodegrade to non-toxic compounds before reaching the environment, but recent concern has focussed on alkylphenol ethoxylates (APEO) and their degradation products as endocrine disruptors. The main environmental concern is not the toxicity of these compounds but rather their estrogenic potential, confirmed by numerous *in vitro* and *in vivo* studies [26]

In this work, all these compounds were investigated both in coastal and harbour sediment and water samples. One of the priority tasks was to determine the environmental distribution of these synthetic organic compounds in coastal areas and to assess the environmental hazards of these compounds. Chemometrical assessment of the multivariate data sets generated in this monitoring study was considered of importance to determine the potential impact of endocrine disruptor compounds in the environmental regions under study. Summarizing, the main objective of this work were: 1) the investigation and identification of the main sources of endocrine disruptor compounds in Spanish coastal and harbour waters and sediments; 2) the description of the nature and composition of these contamination sources; and 3) the estimation of their geographical distribution in the area under study.

EXPERIMENTAL DATA

Experimental details concerning sampling strategy and quality assurance have been reported in a previous work [26]. 35 water samples were collected at 14 different geographical sites, and 39 sediment samples were collected at 26 different geographical sites (see Map and sampling site allocation in Figure 2 and their description in Table 1). The studied area includes hot spots on the Spanish coast, such as the mouths of the Besos and Llobregat rivers, sites near outflows of municipal and industrial wastewaters in harbour areas of Barcelona, Tarragona and Almeria, and various harbours with intense yacht transit. Only in 8 sites (B3, B3-N, B3-S, B3-E, B3-W, T3, T4 and A1) water and sediment samples were taken from the same site. In the other sites, either a water or a sediment sample was only taken. Water samples were obtained over a period of 12 months, starting at July 1999 until July 2000, but not every month in every site. Sediment samples were collected on two occasions from Barcelona and Tarragona at March 1999 and 2000, and from Almeria at June and December 1999 (in Almeria and Almerimar sport harbours and in Aguadulce Harbour). In February 2000, sediment samples were also collected along Cadiz in the Atlantic coast (Cadiz, San Fernando and Sancti Petri) and along Cadiz and Malaga in the Mediterranean coast (Sotogrande, Duquesa, Estepona, Banus and Marbella).

Attention is focussed on the analysis of nonionic polyethoxylene surfactants (alcohol ethoxylates (AEOs and APEOs), their degradation products alkylphenoxycarboxylates and alkylphenols, nonionic surfactants containing an amide group (coconut diethanol amides [CDEAs]) and linear alkyl benzene sulfonates [LAS]. The studied compounds are shown in Figure 1. In a previous work [26], the analysis of these compounds using solid-phase extraction liquid chromatography atmospheric pressure chemical ionization/electrospray ionization mass spectrometry (SPE-LC-APCI-ESI-MS) method was optimized for seawater and marine sediment samples. In this work, the attention is focussed on the chemometrical analysis of these previously obtained results and on how from this analysis environmental knowledge can be derived.

CHEMOMETRIC METHODS

Data sets were organised in two data matrices, one for water samples and another for sediment samples. The rows of these two data matrices identify the different water and sediment samples at the different geographical sampling sites and dates. The columns (variables) of these two data matrices identify the different analysed chemical compounds. When a particular compound was not detected, its concentration value was set equal to half its detection limit.

Data pre-treatment methods included column (variables) mean centering, column norm scaling, autoscaling (column mean centering plus column norm scaling) and log transformation. Mean centering is usually applied to remove constant background contributions, which are considered of little interest for data variance interpretation. However, data mean centering results in a loss of quantitative information about the origin of the scale of the variables, which may be important in environmental source apportionment studies [32-35].

When all the variables are in the same scale units and have similar magnitudes, the usually considered positive effects of using column norm scaling should be considered with caution. Column norm scaling to unit variance may have a notorious and undesirable effect in some cases since it may overweigh variables with poor signal to noise ratios and/or with values close to detection limit. This problem is frequently encountered with water samples and not so much for sediment samples. In fact, column scaling is equivalent to consider that uncertainties are constant for all the elements of the same column, i.e. that in all the samples, the analysis of a particular compound has the same measurement error. This is highly dependent on the analytical method used in these measurements. Unfortunately, in this work, uncertainties of individual measurements for each data entry were not available and more rigorous ways of individual scaling [32] of the variables were not possible. Nevertheless, errors in the measurement of one particular variable (matrix column, concentration of one chemical compound) were similar for all the samples, since the same analytical method was applied for all of them (SPE-LC-APCI-ESI-MS, see ref 26) and these errors were assumed to be independent of analysed samples. When the same errors are expected for

all the measurements of one variable, column norm scaling previous to PCA is an adequate way to give similar weight to different measured variables.

Log transformation of experimental data has been recommended for skewed data sets, like those usually found in environmental studies where the majority of the values are low values with only a minor contribution of high values at widely different magnitudes. With the log transformation data pre-treatment, a more symmetrical distribution of experimental data is obtained. However log transformation pretreatment may also have undesired effects, since it may destroy the internal data linear structure and produce spurious factors. Therefore, it is recommended to investigate and compare the number of components needed to explain the same amount of variance using and not using log transformation. If the number of components is practically the same, log transformation may be a useful alternative to column norm scaling when no error estimates are available (as in this case) and produce results easier to interpret (see below in the results section). In order to remove negative values from input data after log calculation and allow application of non-negativity constraints in ALS treatment, a constant value, usually equal to 1, has been recommended to be added to all the entries [37]. In this way, log values were always non-negative.

Principal Component Analysis, PCA [3, 4] assumes a bilinear model to explain the observed data variance using a reduced number of components. See previous references [3, 4] for a detailed description of this well-known methodology in chemometrics. This bilinear data decomposition may be written by equation:

$$d_{ij} = \sum_{n=1}^N u_{in} v_{jn} + e_{ij} \quad \text{equation 1}$$

where d_{ij} is the i-row sample j-column variable (compound identification) element of the experimental data matrix, u_{in} is the corresponding n- score element for the sample i, v_{jn} is the corresponding n loading element for the variable j and e_{ij} is the residual not modeled by the sum of N components or contributions sources. The same equation can be written in matrix form as:

$$\mathbf{D} = \mathbf{U} \mathbf{V}^T + \mathbf{E} \quad \text{equation 2}$$

where $\mathbf{D}$ is the experimental data array arranged in a data matrix. Equation 2 describes a matrix factorization on two matrices, the loadings matrix $\mathbf{V}^T$ and the scores matrix $\mathbf{U}$. PCA solves equation 2 under orthogonal constraints for the columns of the $\mathbf{U}$ scores matrix and orthonormal constraints for rows of the $\mathbf{V}^T$ loadings matrix. PCA solutions are unique in the sense that every principal component is extracted sequentially in the direction of the maximum explained data variance. Only one direction exists with these properties for each PC. Rotation (orthogonal or not) of these principal components give new solutions fitting equally well the data, but pointing towards different directions and therefore they should not be considered principal component analysis solutions, but rotated solutions from those found by PCA. The determination of the complexity of a PCA model (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. Loadings matrix $\mathbf{V}^T$, identifies the nature of the main contamination sources by means of their chemical composition (chemical composition loadings). Scores matrix $\mathbf{U}$, gives the sample scores for these contamination sources, i.e. it gives the geographical and temporal distribution of contamination sources.

Whereas PCA provides an optimal least squares solution of the bilinear model described by equations 1 and 2 under orthogonal constraints and maximum explained variance, multivariate curve resolution using alternating least squares (MCR-ALS) provides a non-negative least squares solution of the same equation without using orthogonal constraints [5, 6] nor maximum explained variance for each individual component. In MCR-ALS, equation 2 is not solved sequentially as it is for PCA, but for all the components simultaneously and the resolved non-orthogonal loadings do overlap (as also probably do the true contamination sources). Whereas PCA orthogonal solutions of a two-way data matrix are unique, non-negative ALS solutions of the same equation for a two-way data matrix may be not unique, i.e. they may be rotationally ambiguous [5]. Loadings scaling to unit norm (normalization like in PCA) were also applied to scale MCR-ALS solutions of Equation 1. Hopefully, MCR-ALS solutions will be more similar to the true sources of data variance and will be physically and chemically more meaningful than those found by PCA. MCR-ALS results provide an alternative way to PCA to resolve and interpret real environmental sources of data variance. In this

particular case, this refers to the resolution of the main sources of endocrine compounds in the particular area under study, as well as of their geographical distribution. Other authors have also proposed similar approaches in source apportionment and receptor modeling of atmospheric data sets. [32-36]. The same approach can be easily extended to environmental data sets obtained in monitoring studies of other environmental compartments (like surface waters and sediments). The key aspect to recognize here is that exploratory analysis of environmental data sets is a mixture analysis problem and that similar methodologies to those used in other mixture analysis fields can be used for environmental analysis studies too. A similar multivariate curve resolution approach has been proposed to solve other spectrometric mixture resolution problems in the study of complex chemical systems [see for instance references 38 and 39], and also in the resolution and apportionment of contamination sources in the environmental studies [7, 8].

Most of data treatments, algorithms, computer programs, graphics algorithm and software were used under MATLAB computer and visualization environment (Release 12, The Mathworks, Natick MA, USA). For PCA, PLS Toolbox v2.1 (Eigenvector Research, Mason, WA, USA) was used. For MCR-ALS, home made (<http://www.ub.es/gesq/mcr/mcr.htm>) software was used.

RESULTS AND DISCUSSION

In a first analysis, the water data set (35 water samples and 21 chemical compounds) was examined without any data pre-treatment. Some of the analysed compounds had very small contributions (variables C14LAS, OP, C10EO, C14EO, C7DEA) and they were eliminated from this first data analysis. The first variable (LAS total) was also eliminated because it had no influence in the results (it is the sum of the concentration of LAS compounds: C10LAS, C11LAS, C12LAS, C13LAS and C14LAS). In these water samples, most of the concentrations of the analysed endocrine disruptor compounds were low.

The reduced data set (35 samples and 15 variables) was mean-centred and analyzed with and without (column norm) variance scaling. PCA results without variance scaling were preferred for several reasons. On one hand the concentration of all the compounds were at the same scale units and of similar size. On the other hand, when the water data set was variance scaled, the variables with lower signal to noise ratios were dominating too much the variance and they were more difficult to interpret. When data were only mean centered, an easy identification of the main sources of endocrine disruptor compounds was possible as well as a good geographical distribution of them. In Figure 3A, box plot of the finally analysed water data samples is given to illustrate their distribution among different variables.

Three principal components already explained as much as 93.1% of the data variance of this water data set. Mean centering the variables had little effect in this case (water samples) since mean values were always close to zero for all the different variables. The amount of variance explained by these three principal components when data were not mean centered was similar (95.6%), and the obtained loadings and scores were also similar to those given next for mean-centered data. Loading plots of the three principal components are given in Figure 4A. First PC explaining 57.3% of the data variance showed high negative contributions for C10LAS, C11LAS, C12LAS and C13LAS compounds. This first principal component was clearly identified with a major input contamination source from linear alkylbenzene sulfonates (LAS). Percentage of explained variance for the second PC was 25.8%. CDEA total and C11CDEA had high positive loadings in this second PC and it was therefore identified as the main source of

these diethanolamide compounds. The third PC accounts for 10% of the total variance, and was loaded mostly by high negative contributions of C13LAS and especially by PEG. Forth and higher PCs explained lower data variance amounts ($< 3\%$) and they were mostly related with small individual contributions, noise and experimental errors and they were not considered relevant for this study. PCA allowed the identification of three main contamination sources of the studied endocrine disruptor compounds. A first major contamination source due to linear alkylbenzene sulfonates (LAS) compounds, a second contamination source mostly due to coconut diethanolamides (CDEA) compounds and a third contamination source mostly due to polyethylenglycol (PEG) contamination.

Geographical distribution of these three main contamination sources was investigated in scores plots. Figure 4B gives the PCA scores for PC1 and PC2. This plot shows that the samples with the largest negative scores in the PC1 (high concentration of LAS group compounds) axis were from Barcelona (BCN, see Map Figure 2: B1, B2 and B3) area (samples 1-9), obtained in January, March and May 2000, and samples 22 (site T4 Nov.1999), 23 and 24 (sites T3 and T4 January 2000) were from Tarragona (TAR) area. On the other hand, when PC2 is considered, Barcelona samples appeared well differentiated from Tarragona samples. This second principal component mostly distinguishes samples where total CDEA and C11DEA had different concentration levels, i.e. whereas samples of the Barcelona area had relatively low concentrations of total CDEA and C11DEA, samples from Tarragona area had higher concentrations of these two compounds. Negative PC3 scores (not shown) were observed for samples 1, 2, 3 and 7 from Barcelona and for samples 21 and 22 from Tarragona, having all of them high concentrations of PEG and much lower concentrations of C13LAS. Almeria water samples were clustered around the center of the coordinate axes in the scores plots indicating that they were the less contaminated samples.

Loadings and scores obtained by Multivariate Curve Resolution were very similar to those obtained by PCA and they are not shown here for brevity. In this case, experimental data were not mean centered and non-negativity constraints were applied during the Alternating Least Squares optimisation. The degree of overlap between the ALS components was not severe (approximately a 10%) and this is the reason why the components obtained by the two methods, PCA and non-negative ALS were rather

similar. The conclusion is that the three deduced contamination sources of endocrine disruptors in water samples deduced either by PCA or MCR-ALS were very similar and that they were rather independent. Geographical distribution of these sources was also similar using these two approaches and this coincidence reinforces the reliability and interpretation of them.

PCA and MCR-ALS were then applied to the 39 sediment samples including the concentration of all 21 endocrine disruptor compounds (Figure 1). Apart from Barcelona, Tarragona and Almeria areas (see Map, Figure 2), sediment samples from Cadiz Bay and from Malaga were included in the analysis in this case. Before PCA and ALS were applied to sediment samples, different data pre-treatment methods were investigated. Measured values of concentrations of endocrine disruptors in sediment samples were larger and varied more among sediment samples than among water samples; larger accumulation and persistent effects of endocrine compounds were occurring in sediment samples. In Figure 3B, box plot of the finally analysed sediment samples is also given to illustrate their distribution among different variables.

For sediment samples, before PCA was applied, data were log transformed and autoscaled (in the columns/variables direction). The linear structure of the data set was mostly kept after log transformation, since no additional principal components were needed to explain the same amount of data variance. Differently to water samples, mean centering and norm column scaling (autoscaling) after log transformation facilitated in this case (sediment samples) the PCA interpretation of the main contamination sources. Other attempted pre-treatment methods did not improve the identification and distribution of the different contamination sources of endocrine disruptors in sediment samples.

First principal component explained 42.7% of the total data variance, second PC explained 16%, third component explained 10.8% and fourth component 7.1% of the data variance. Figure 5A gives PCA loadings for these fourth main principal components. First PC had high negative loadings for all variables and identified a main highly diffuse contamination source of all endocrine disruptors associated with their steady accumulation in sediment samples. Second PC had positive loadings for NPEO, total AEO, C10EO, C12EO and PEG (mostly related with alcohol ethoxylate and

nonylphenol ethoxylate discharges to the environment) whereas it had negatives loadings for the LAS group of compounds. Therefore, second PC distinguished a second source of these compounds discharged to the environment in an independent way and not simultaneously as shown in first PC. Third PC had high negative loadings for total AEO, C12EO and C13EO and positive loadings for total CDEA, C9DEA and C11DEA. Fourth PC gave high loadings of OP and C7DEA. Again, third and fourth components identified minor sources of the endocrine disruptors different to those expressed by PC1 and PC2.

Samples were well differentiated between them according to their sampling sites in the scores plot PC1 versus PC2 (see different identification and colour symbols in Figure 5B). Samples on the left side of the plot were highly contaminated (high negative PC1 loadings and scores) and those on the right of the plot (positive PC1 loadings and scores) were low contaminated. Barcelona samples (circles) were clustered in the negative PC1-PC2 scores region. Some Barcelona samples had relatively high inputs of LAS group of compounds coming from a secondary source of them as it is shown by the relative high negative PC2 score values. For instance, sample 8 in the mouth of the Besos river close to Barcelona city (site B3-S March 00) was a high contaminated sample, giving high negative PC1 and PC2 scores. Tarragona samples (labelled '+' in Figure 5B) instead, are situated at the right of the PC1-PC2 scores plot, showing that they had a lower contribution of the general accumulation contamination source expressed in PC1. However, some of the Tarragona samples had relatively high positive scores for PC2, especially samples 11, 14 (sites T1 and T4 March 99) and 15 (site T1 March 00). Almeria samples (labelled '*' in Figure 5B) were distributed in two groups: December samples (sample numbers 24-27) were located approximately in the same place as low contaminated Barcelona samples, whereas June samples (sample numbers 19-23) had relatively high positive PC2 scores and loadings, showing that they were more contaminated by this second independent source of alcohol ethoxylate and nonylphenol ethoxylate compounds with high loadings in PC2 (see above). Most of Cadiz and Malaga samples (labelled 'x' in Figure 5B) except sample 30 in San Fernando, were situated at the right of the plot with positive PC1 loadings and scores, i.e. samples with low inputs of the first contamination source of endocrine disruptors defined by PC1. They were distinguished between them according to the different inputs of the second contamination source defined by PC2. Samples 31, 33 and 34,

having positive PC2 scores in Cadiz and Malaga had relatively higher inputs of alcohol ethoxylate and nonylphenol ethoxylate compounds and samples 32, 36, 37, 38 and 39 had relatively high inputs of LAS group of compounds. Sample 30 (San Fernando -site SF-, Cadiz) was an exceptionally contaminated sample and it was clearly distinguished from all the other samples in the extreme left of the PC1-PC2 scores plot, with high negative PC1 loadings and scores. In fact this sample was one of the samples showing a larger general accumulation of all endocrine disruptors studied in this work.

MCR-ALS method was also applied to the 39 sediment samples including the concentration of the same 21 variables as for PCA. Data pre-treatment was in this case, column data scaling without mean centering to allow source apportionment using non-negative constraints in the ALS optimisation procedure. MCR-ALS gave also an easy interpretation of main contamination sources and a good separation of their geographical distribution in the fourth areas (Barcelona, Tarragona, Almeria and Cadiz-Malaga). In contrast to PCA, log preliminary transformation of the whole data set was not needed for ALS, since a good separation between samples was already achieved without using this data pretreatment.

Although in the analyses of sediment samples, relative magnitudes of loadings and scores resolved by ALS differ somewhat from those obtained by PCA (data were not mean centred in this case), the qualitative interpretation of the main contamination sources of endocrine disruptors resolved in both cases was again, like for water samples, rather similar. In the case of ALS however, semiquantitative source apportionment in the different samples is easier. This fast apportionment of the different contamination sources in samples is cumbersome in PCA due to the loss of quantitative information occurred when data are mean centered and due to the application of orthogonality constraints. Since loadings and scores may be positive and negative, they cannot be simply added to get the whole source apportionment in samples and variables (32-36).

Finally, an additional simultaneous study of the 35 water samples and of the 39 sediment samples was performed. Data pretreatment selected to be applied before PCA was in this case log transformation and autoscaling. Autoscaling could be applied in two different ways: i) considering water and sediment samples together as a single data matrix (*j-scaling*); and ii) considering water and sediment samples as two different data matrices and setting them together after autoscaling. A better separation between

contamination sources for both type of samples was obtained in case (i). The first principal component explained 76.9% of the data variance and it described the main contamination source of endocrine disruptor compounds with high positive loadings for all variables. The percentage of explained variance for the second PC was instead, of only around 5.22%. This second PC, described a secondary source differentiating the contamination of LAS group of compounds (large negative PC1 loadings) from the contamination of most of the other compounds (positive PC2 loadings). Scores plot are given in Figure 6A. With *j-scaling*, water (labeled 'x') and sediment samples (labeled 'o') were clustered in two well differentiated groups. Whereas sediment samples were at the right of the plot giving positive scores indicating larger contributions of the first general contamination source, water samples were clustered at the left of the plot indicating lower contributions of this first contamination source. First PC describes graphically the expected accumulation of endocrine disruptors in sediment samples compared to water samples. Comparison of the different sediment and water samples allows a distinction between the different samplings sites. For instance, sediment sample 8 from Barcelona in the mouth of the Besos river (site B3-S March 00) and sediment sample 30 from Cadiz Bay (site SF Feb. 00) were the most contaminated samples according to this first resolved general contamination source. Second principal component explained what samples in each group (water and sediment samples) were more contaminated by the LAS group of compounds, or more contaminated by NPEO, C10EO, PEG, C9DEA and C11DEA compounds. In fact conclusions derived from a more detailed analysis of this plot would be in agreement with those obtained in previous analysis of individual analysis of water and sediment samples. An obvious conclusion is therefore, the expected accumulation of endocrine disruptors in sediment samples compared to water samples, thing that it is also in agreement with a detailed observation of the concentration values of these compounds in sediment samples compared to water samples (see for instance reference 26). However, PCA scores plots provides an easy, fast and very illustrative way to represent graphically this pattern in two dimensions, showing in relative terms how the different sediment and water samples were affected by this general tendency and showing also its geographical distribution.

In case of (ii) pretreatment, the two type of samples were not clustered as in Figure 6A, but in contrast, a better separation between the four geographical areas was obtained (see in Figure 6B, Barcelona labelled 'o', Tarragona labelled '+', Almeria labelled '*',

and Cadiz-Malaga samples labelled 'x'), although they gave similar PC1 and PC2 loading profiles than those obtained using *j-scaling* in Figure 6B, i.e. the same contamination sources were identified. Results applying non-negative MCR-ALS simultaneous analysis of water and sediment samples after column data scaling gave similar results to the already described in previous analysis of water or sediment samples using PCA and ALS and the results are not given here for brevity.

CONCLUSIONS

Chemometric methods were used for the investigation and identification of the main environmental microcontaminant sources of endocrine disruptors like non-ionic surfactants, their degradation products and linear alkylbenzene sulfonates in coastal seawaters and sediments of East and South of Spanish coast and harbours. Although hot spots with point sources of municipal and industrial discharges were identified (Besos mouth river close to Barcelona and San Fernando in Cadiz Bay), it can be concluded that the investigated compounds were wide-spread contaminants in the marine environment of the Mediterranean Spanish coasts and harbours. In comparison, LAS contamination (clothes washing liquid detergents) was predominantly allocated around the Barcelona harbour, alcohol ethoxylates contamination (dishes washing up liquid detergents) was more important in Tarragona and Almeria harbours and coconuts diethanolamide contamination (personal care products) gave high peak values around the Bay of Cadiz and Malaga coast. Large accumulation effects of most of these contaminants were observed for sediment samples collected near discharges of domestic or industrial wastewaters in the studied harbour areas, compared to less contaminated sea water samples.

ACKNOWLEDGEMENTS

This research has been funded by the CICYT AMB98-0913-C03-01 and PPQ 2001-1805-C03 MCYT Projects.

E. Peré-Trepat acknowledges a PhD grant from the University of Barcelona.

REFERENCES

- [1] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. de Jong, P.J. Lewi, and J. Smeyers-Verbeke (1998) Handbook of Chemometrics and Qualimetrics, Elsevier Science, Amsterdam
- [2] J.W. Einax, H.W. Zwaninger and S. Geiss (1997) Chemometrics in Environmental Chemistry, VCH: Weinham, Germany
- [3] I.T. Joliffe (1986) Principal Component Analysis, Springer, New York
- [4] S. Wold, K. Esbensen and P. Geladi (1987) Chemom. Intell. Lab. Syst. 2:37
- [5] R. Tauler, A.K. Smilde and B.R. (1995) J. Chemom. 9:3
- [6] R. Tauler (1995) Chemom. Intell. Lab. Syst. 30:133
- [7] R. Tauler, D. Barceló and E.M. Thurman (2000) Environ. Sci. Technol. 34:3307
- [8] J. Salau, R. Tauler, J.M. Bayona and I. Tolosa (1997) Environ. Sci. Technol. 31: 3482
- [9] R. Tauler, D. De Almeida Azevedo, S. Lacorte, P. Viana and D. Barceló (2001) Environ. Technol. 22:1043
- [10] D.R. Karsa (1998) Chem. Ind., 9:685
- [11] European Economic Community, Urban waste water treatment. (1991) Directive 91/271/EEC Official J. Eur. Communities L135:40
- [12] C. Crescenzi, A. Di Corcia, R. Samperi and A. Marcomini (1995) Anal. Chem. 67: 1797
- [13] C.G. Naylor, J.P. Mieure, W.J. Adams, J.A. Weeks, F.J. Castaldi, L.D. Ogle and R.R. Romano (1992) J. Am. Oil Chem. Soc. 69:695
- [14] D.T. Bennie, C.A. Sullivan, H.B. Lee, T.E. Peart and R. Maguire (1997) J. Sci. Total Environ. 193:263
- [15] S.A. Snyder, T.L. Keith, D.A. Verbrugge, E.M. Snyder, T.S. Gross, K. Kannan and J.P. Giesy (1999) Environ. Sci. Technol. 33:2814
- [16] E.R Bennet and C.D. Metcalfe (1998) Environ. Toxicicol. Chem. 17:1230
- [17] M. Ahel, W. Giger and C. Schaffner (1994) Water Res. 28:1143
- [18] A. Marcomini, G. Pojana, A. Sfriso and J.M. Quiroga Alonso (2000) Environ. Toxicol. Chem. 19:2000
- [19] P. Lee Ferguson, C.R. Iden and B.J. Brownawell (2000) Anal. Chem. 72:4322
- [20] M.A. Blackburn, S.J. Kirby and M.J. Waldock (1999) Mar Pollut. Bull. 38:109
- [21] D.Y. Shang, R.W. Macdonald and M.G. Ikononou (1999) Environ. Sci. Technol. 33:1366
- [22] J.S. Khim, K. Kannan, D.I. Villeneuve, C.K. Koh and J.P. Giesy (1999) Environ. Sci. Technol. 33:4199

- [23] J. Riu, P. Eichhorn, J.A. Guerrero, Th.P. Knepper and D. Barceló (2000) *J. Chromatogr. A*, 889:221
- [24] C.M. Lye, C.L.J. Frid, M.E. Gill, D.W. Cooper and D.M. Jones (1999) *Environ. Sci. Technol.*, 33:1009
- [25] A. Marcomini, B. Pavón, A. Sfriso and A.A. Orio (1990) *Mar Chem.*, 29:307
- [26] M. Petrovic, A. Rodríguez Fernández-Alba, F. Borrull, R.M. Marce, E. González Mazo and D. Barceló (2002) *Environ. Toxicol. Chem.* 21:37
- [27] J. Riu, E. Gonzalez-Mazo, A. Gomez-Parra and D. Barceló (1999) *Chromatographia* 50:275
- [28] V.M. Leon, E. Gonzalez-Mazo and A. Gomez-Parra (2000) *J. Chromatogr. A* 889: 211
- [29] E. Gonzalez-Mazo, J.M. Quiroga, D. Sales and A. Gomez-Parra (1997) *Toxicol. Environ. Chem.* 59:77
- [30] E. Gonzalez-Mazo, J.M. Forja and A. Gomez-Parra (1998) *Environ. Sci. Technol.* 32:1636
- [31] E. Gonzalez-Mazo, J.M. Quiroga and A. Gomez-Parra (1999) *Cienc. Mar.* 25:367
- [32] P. Paatero (1993) *Chemom. Intell. Lab. Syst.* 18:183
- [33] P. Paatero (1997) *Chemom. Intell. Lab. Syst.* 37:23
- [34] R. C. Henry (2002) *Chemom. Intell. Lab. Syst.* 60:43
- [35] Ph. K. Hopke (1995) *The mixture resolution problem applied to airborne particle source apportionment. Chemometrics in Environmental Chemistry. Volume 2, Part H; Springer:Berlin*
- [36] Ph. K. Hopke (1991) *Chemom. Intell. Lab. Syst.* 10:21
- [37] V. Zitko (1994) *Mar. Pollut. Bull.* 28:718
- [38] M. Vives, R. Gargallo and R. Tauler (1999) *Anal. Chem.* 71:4328
- [39] J.M. Díaz-Cruz, M.S. Diaz-Cruz, C. Ariño, E. Esteban and R. Tauler (2001) *J. Electroanal. Chem.* 516:110
- [40] Montgomery, D.C.; Runger, G.C. *Applied Statistics and Probability for Engineers*; Wiley: New York, 1999

FIGURE CAPTIONS

Figure 1. Chemical structures and identification acronyms of studied compounds.

Figure 2. Map of the studied areas showing the locations of the sampling sites (see their description in Table 1).

Figure 3 Box plot statistical distribution (see reference 40) of the endocrine compounds analysed in this study. A) water samples mean centered data set; B) sediment samples autoscaled log data (plus one) set.

Figure 4A. PCA loadings for water samples (35 samples and 15 variables). From top to bottom PC1 to PC3. In CxLAS, CxEo and CxDEA, x = number of C atoms.

Figure 4B. PC1 vs PC2 scores plot for water samples (35 samples and 15 variables). Sample identification is: 'o' Barcelona samples, '+' Tarragona samples and '*' Almeria samples. Symbol sample map (Figure 2 and Table 1) identification is: (1)B1 January00, (2)B2 January00, (3)B3 January00, (4)B1 March00, (5)B2 March00, (6)B3 March00, (7)B1 May00, (8)B2 May00, (9)B3 May00, (10)B1 July00, (11)B2 July00, (12)B3 July00, (13)B3-N July00, (14)B3-S July00, (15)B3-E July00, (16)B3-W July00, (17)T3 July99, (18)T4 July99, (19)T3 Sept.99, (20)T4 Sept.99, (21)T3 Nov.99, (22)T4 Nov.99, (23)T3 January00, (24)T4 January00, (25)T3 March00, (26)T4 March00, (27)T3 May00, (28)T4 May00, (29)T3 July00, (30)T4 July00, (31)A1 January00, (32)A1-N January00, (33)A1-S January00, (34)A1-E January00, (35)A1-W January00.

Figure 5A. PCA loadings for sediment samples (39 samples and 21 variables). From top to bottom PC1 to PC4. In CxLAS, CxEo and CxDEA, x = number of C atoms.

Figure 5B. PC1 vs PC2 scores plot for sediment samples (39 samples and 21 variables). Samples identification is: 'o' Barcelona samples, '+' Tarragona samples, '*' Almeria samples and 'x' Cadiz samples. Symbol sample map (Figure 2 and Table 1) identification is: (1)B3 March99, (2)B3-N March99, (3)B3-S March99, (4)B3-E March99, (5)B3-W March99, (6)B3 March00, (7)B3-N March00, (8)B3-S March00,

(9)B3-E March00, (10)B3-W March00, (11)T1 March99, (12)T2 March99, (13)T3 March99, (14)T4 March99, (15) T1 March00, (16)T2 March00, (17)T3 March00, (18)T4 March00, (19)A1 June99, (20)A2 June99, (21)A4 June99, (22)AM June99, (23)AD June99, (24)A1 Dec.99, (25)A3 Dec.99, (26)A4 Dec.99, (27)AM Dec.99, (28)AD Dec.99, (29)C Feb.00, (30)SF Feb.00, (31)SP Feb.00, (32)SG Feb.00, (33)D Feb.00, (34)E1 Feb.00, (35)E2 Feb.00, (36)BN1 Feb.00, (37)BN2 Feb.00, (38)M1 Feb.00, (39)M2 Feb.00.

Figure 6A. PC1 vs PC2 scores plot for water and sediment samples analysed simultaneously. The whole data set was log transformed and autoscaled as a single data set. Samples identification is: 'x' water samples and 'o' Sediment samples. Identification of water samples is given in Figure caption 3B. Identification of sediment samples is given in Figure caption 5B.

Figure 6B. PC1 vs. PC2 scores plot for water and sediment samples analysed. Water and sediment samples were log transformed and autoscaled independently. Samples identification is: 'o' Barcelona samples, '+' Tarragona samples, '*' Almeria samples and 'x' Cadiz samples. Identification of water samples: (1)B1 January00, (2)B2 January00, (3)B3 January00, (4)B1 March00, (5)B2 March00, (6)B3 March00, (7)B1 May00, (8)B2 May00, (9)B3 May00, (10)B1 July00, (11)B2 July00, (12)B3 July00, (13)B3-N July00, (14)B3-S July00, (15)B3-E July00, (16)B3-W July00, (17)T3 July99, (18)T4 July99, (19)T3 Sept.99, (20)T4 Sept.99, (21)T3 Nov.99, (22)T4 Nov.99, (23)T3 January00, (24)T4 January00, (25)T3 March00, (26)T4 March00, (27)T3 May00, (28)T4 May00, (29)T3 July00, (30)T4 July00, (31)A1 January00, (32)A1-N January00, (33)A1-S January00, (34)A1-E January00, (35)A1-W January00. Identification of sediment samples: (36)B3 March99, (37)B3-N March99, (38)B3-S March99, (39)B3-E March99, (40)B3-W March99, (41)B3 March00, (42)B3-N March00, (43)B3-S March00, (44)B3-E March00, (45)B3-W March00, (46)T1 March99, (47)T2 March99, (48)T3 March99, (49)T4 March99, (50) T1 March00, (51)T2 March00, (52)T3 March00, (53)T4 March00, (54)A1 June99, (55)A2 June99, (56)A4 June99, (57)AM June99, (58)AD June99, (59)A1 Dec.99, (60)A3 Dec.99, (61)A4 Dec.99, (62)AM

Dec.99, (63)AD Dec.99, (64)C Feb.00, (65)SF Feb.00, (66)SP Feb.00, (67)SG Feb.00,
(68)D Feb.00, (69)E1 Feb.00, (70)E2 Feb.00, (71)BN1 Feb.00, (72)BN2 Feb.00,
(73)M1 Feb.00, (74)M2 Feb.00.

Table 1. Description of sampling sites (see their location in Figure 2)

Sample code	Location (Spain)	Description of location	Samples analyzed
B1	Barcelona	River Besos mouth	Water
B2	Barcelona	River Llobregat mouth	Water
B3	Barcelona	Outflow of sludge and municipal wastewater	Water/Sediment
B3-N	Barcelona	1 Km North of B3	Water/Sediment
B3-S	Barcelona	1 Km South of B3	Water/Sediment
B3-E	Barcelona	1 Km East of B3	Water/Sediment
B3-W	Barcelona	1 Km West of B3	Water/Sediment
T1	Tarragona	Outflow of industrial plant	Sediment
T2	Tarragona	Outflow of industrial plant	Sediment
T3	Tarragona	Outflow of industrial plant	Water/Sediment
T4	Tarragona	Outflow of industrial plant	Water/Sediment
A1	Almería	Outflow of municipal wastewater	Water/Sediment
A1-N	Almería	North of A1	Water
A1-S	Almería	South of A1	Water
A1-E	Almería	East of A1	Water
A1-W	Almería	West of A1	Water
A2	Almería	5 m from A1	Sediment
A3	Almería	Harbour	Sediment
A4	Almería	Sport harbour	Sediment
AM	Almerimar	Sport harbour	Sediment
AD	Aguadulce	Harbour	Sediment
C	Cadiz	Nautic club, sport harbour	Sediment
SF	San Fernando	Close to an untreated urban wastewater effluent	Sediment
SP	Sancti Petri	Nautic club, sport harbour	Sediment
SG	Sotogrande	Nautic club, sport harbour	Sediment
D	Duquesa	Nautic club, sport harbour	Sediment
E1	Estepona 1	Nautic club, sport harbour	Sediment
E2	Estepona 2	Nautic club, sport harbour	Sediment
BN1	Banus 1	Nautic club, sport harbour	Sediment
BN2	Banus 2	Nautic club, sport harbour	Sediment
M1	Marbella 1	Nautic club, sport harbour	Sediment
M2	Marbella 2	Nautic club, sport harbour	Sediment