ELSEVIER

Analytica Chimica Acta 373 (1998) 311–319

ANALYTICA CHIMICA ACTA

# Multivariate calibration of polycyclic aromatic hydrocarbon mixtures from excitation–emission fluorescence spectra

J.L. Beltrán*, R. Ferrer, J. Guiteras

*Departament de Química Analítica, Universitat de Barcelona, Avgda, Diagonal 647, E-08028 Barcelona, Spain*

Received 27 November 1997; received in revised form 16 June 1998; accepted 20 June 1998

**Abstract**

The excitation–emission fluorescence spectra (EEM) of mixtures of 10 polycyclic aromatic hydrocarbons (PAHs) have been analyzed using different multivariate calibration procedures (partial least squares regression, PLSR; and parallel factor analysis, PARAFAC). The compounds studied were anthracene, benz[a]anthracene, benzo[a]pyrene, chrysene, phenanthrene, fluoranthene, fluorene, naphthalene, perylene and pyrene.

Several algorithms (PARAFAC, PLS1, PLS2, Tri-PLS1 and Tri-PLS2) were tested for the quantification of these compounds from the EEM spectra of a set of standards, at ng $ml^{-1}$ levels. The best results were obtained by the application of a three-way multivariate calibration method (Tri-PLS2). This method was selected for the determination of samples of tap and mineral waters spiked with all these PAHs. Excitation and emission wavelength ranges were between 240 and 300 nm and 310 and 478 nm, respectively. In the working conditions, an EEM spectrum covering this range could be recorded in about 2 min.


*Keywords:* Multivariate calibration; Partial least squares; PARAFAC; Polycyclic aromatic hydrocarbons; EEM fluorescence

## 1. Introduction

Polycyclic aromatic hydrocarbons (PAHs) are a wide group of polluting compounds that are of anthropogenic and natural origin, and therefore they can be found in all kinds of samples [1–3]. Moreover, they are extremely hazardous, which has led the environmental protection agency (EPA) to include 16 of them in the list of priority pollutants [4].

The determination of PAHs in complex mixtures is usually done by gas chromatography coupled to mass spectrometers or by liquid chromatography (LC) with either UV–visible or fluorescence detectors [4,5]. For routine analysis, LC combined with fluorescence detectors is very suitable, because it is relatively simple, and provides good sensitivity and selectivity. There is, however, a constant need to improve existing methods, not only in order to obtain better selectivity and limit of detection but also to shorten the time required for an analysis. This reduction in time is usually achieved in the steps of sample pretreatment, preconcentration or clean-up, but can also be obtained in the measuring step. In fact, several authors have described the use of the full excitation–emission fluorescence spectra (also known as EEM: excitation–emission matrix) for the identification and quantification of PAHs [6–10] in relatively simple

*Corresponding author. Tel.: +34-93-402-1283; fax: +34-93-402-1233; e-mail: joseluis@zeus.qui.ub.es

0003-2670/98/$19.00 
*PII* S0003-2670(98)00420-6

mixtures. However, the circumstance that these compounds have different fluorescence excitation and/or emission characteristics, combined with the additional wealth of information offered by EEMs and the power of such chemometric tools as *n*-parallel factor analysis (PARAFAC) or partial least squares regression (PLSR) makes it possible, theoretically at least, to perform a complete analysis of complex mixtures of PAHs – including quantification – without the need of a previous separation by chromatographic means. This assumption has been tested in this paper, where samples containing 10 different PAHs have been analyzed without any previous separation.

The EEM spectrum of a single substance (i) is considered as a bilinear response, as it is actually a matrix obtained by the product of two vectors, related with their corresponding excitation and emission spectra ($\phi_{ex,i}$ and $\phi_{em,i}$, respectively). It is also proportional to the concentration of the substance ($C_i$). For a mixture of *nc* fluorescent compounds, the obtained EEM spectrum is equal to the addition of the *nc* corresponding spectra.

$$\mathrm{EEM} = \sum_{i=1}^{nc} C_i \cdot \phi_{ex,i} \cdot \phi_{em,i}.$$

EEM can be generated in several ways, the most common being by successive recording of the fluorescent emission spectra at different excitation wavelengths using a conventional spectrofluorimeter. Other systems, such as video fluorometers [6–8] or CCD cameras [9,10] can also be used to detect the emission fluorescence; they are much faster in recording the emission spectra than the scanning spectrofluorimeters based on photomultiplier tubes (PMT), but these latter ones have the advantage of greater sensitivity.

In this work, a conventional spectrofluorimeter has been used to record the EEM spectra at a medium-fast scanning speed (25 nm/s), which allows to obtain a complete EEM spectrum in about 7 min. Initially, the excitation and emission wavelength ranges were from 200 to 350 nm and from 250 to 550 nm, respectively, but, later, shorter ranges – excitation wavelengths from 240 to 300 nm and emission wavelengths from 310 to 478 nm – were selected. In this way, the size of the data matrix was substantially reduced and the effect of the Rayleigh scattering was minimized. An additional advantage was that the time required to record an EEM spectrum was shortened to about 2 min.

There are several multivariate calibration procedures that can be used for the treatment of EEM fluorescence data, in order to quantify the compounds present in a mixture. However, parallel factor analysis (PARAFAC) [10–12] and partial least-squares regression (PLSR) [13–16] appear to be the best suited for the analysis of complex mixtures (that is, those containing more than three or four components) and, in consequence, only these have been tested in this paper.

PARAFAC is based on the decomposition of the data matrix by iterative procedures. Given a number of factors (the number of fluorescent species), two vectors related with the excitation and emission spectra and one scalar (related with the concentration) are found for each factor. This method has been applied for the resolution of mixtures of several compounds from EEM fluorescence data [10,11]. The main drawback of PARAFAC lies in the application to complex mixtures, where the resolution of the system by iterative procedures is a lengthy task. Moreover, depending on the range of excitation and emission wavelengths, the Rayleigh scattering is a strong interference, as it is essentially a nonlinear signal, and it cannot be modeled properly by a single two-vector product. For this particular reason, PARAFAC has been applied only to the reduced data set of this work.

PLSR is a factor analysis method that has been widely used in multicomponent quantitative analysis from several spectral data, such as IR, UV–visible or fluorescence [15,17]. This procedure is based on the regression of chemical concentrations on latent variables (or factors) obtained from the spectral data. The main advantage of PLSR calibration procedures is that they can model a system even in the presence of interfering signals, provided that they are included in the calibration step. There are two main PLSR algorithms [13], named PLS1 and PLS2. In PLS1, a calibration model is built for each component; this means that for a 10-component sample (as in this work) 10 different models have to be built. PLS2 is better suited for multicomponent analysis, as it can model several components simultaneously.

Both PLS1 and PLS2 are designed to work with a single spectrum (first-order data) for each sample and they cannot be directly applied to EEM spectra. To

overcome this problem, the emission spectra contained in a single EEM spectrum are concatenated in order to obtain an 'unfolded' spectrum [18].

Recent developments in PLSR algorithms have led to their extension to data of higher orders, mainly to second-order data. This new procedure, named multilinear PLS [19], has been successfully applied to the resolution of multicomponent mixtures from EEM fluorescence spectra. As in normal PLSR, there is also the possibility to determine only one analyte or several analytes simultaneously. The corresponding algorithms are named Tri-PLS1 and Tri-PLS2, because they are applied to Tri-linear data.

In the present work, different algorithms (PARAFAC, PLS1, PLS2, Tri-PLS1 and Tri-PLS2) have been used for the resolution of mixtures containing 10 PAHs, making use of their EEM spectra, obtained in a Brij-35 micellar medium to enhance their native fluorescence and to protect them from quenching effects. The PAHs studied are: anthracene (Ant), benz[a]anthracene (Baa), benzo[a]pyrene (Bap), chrysene (Cry), phenanthrene (Phen), fluoranthene (Fla), fluorene (Flr), naphthalene (Naph), perylene (Per) and pyrene (Pyr). The procedure has been applied to the determination of these PAHs in spiked water samples (mineral water and network supply) at concentration levels between 4 and 20 ng $ml^{-1}$ with good results, except for naphthalene.

## 2. Experimental

### 2.1. Reagents

Stock standard solutions (about 200 μg $ml^{-1}$) of each PAH were prepared by dissolving the pure solid (Supelco) in either methanol or acetonitrile, depending on its solubility. Brij-35 (Polyoxyethylenlaurylether, Merck) was used without further purification. Acetonitrile and methanol were of analytical reagent grade (Merck) and were used without further purification. Doubly distilled water (Mili-Q+, Millipore) was used to prepare the standard solutions.

### 2.2. Apparatus

An Aminco Bowman Series 2 spectrofluorimeter, equipped with a 1.00 cm quartz cell, was used to obtain the EEM spectra. Slit widths were set to 16 nm, both for the excitation and emission monochromators. All measurements were made in the ratio mode.

Data treatment was done in an IBM/RS 6000 workstation, using several MATLAB™ programs, as PARAFAC [11], the PLS-Toolbox [20] for PLS1 and PLS2 procedures, and the multilinear PLS [19] for Tri-PLS1 and Tri-PLS2.

### 2.3. Procedures

Standard mixtures of PAHs were prepared by dilution of known amounts of the stock solutions to 25 ml and addition of the amount of Brij-35 required to obtain a surfactant concentration of about 0.0036 mol $l^{-1}$ (which corresponds to 40 times its critical micellar concentration). A calibration set of 70 standards and an independent validation set of 14 standards were prepared in this way. The concentration of PAHs in these standards ranged between 0 and 20 ng $ml^{-1}$. The water samples were spiked with all PAHs at four different concentration levels (about 4, 9, 12 and 20 ng $ml^{-1}$), so that all compounds in a sample had similar concentrations, and Brij-35 was added to obtain the same concentration as in the standards. EEM spectra of standards and samples were collected along a period of a month to include the possible day-to-day instrumental uncertainty.

The EEM fluorescence spectra were obtained by recording the emission spectra (from 250 to 550 nm, at 3 nm intervals) corresponding to excitation wavelengths ranging between 200 and 350 nm, set at 5 nm steps. This means that a single EEM spectrum consisted of 31 emission spectra, measured at 101 wavelengths, which makes a total of 3131 data points. The reduced data set selected for latter calculations (240–300 nm for excitation and 310–478 nm for emission) had only 741 data points.

In Fig. 1, the excitation and emission spectra of the PAHs studied are shown. A contour map of a mixture of the 10 compounds is displayed in Fig. 2, where the scattering bands are clearly visible. The inner zone indicated in this figure corresponds to the reduced range in excitation and emission wavelengths. The difference between a full EEM spectrum and the reduced data set is also appreciated in the three-dimensional spectrum displayed in Fig. 3.

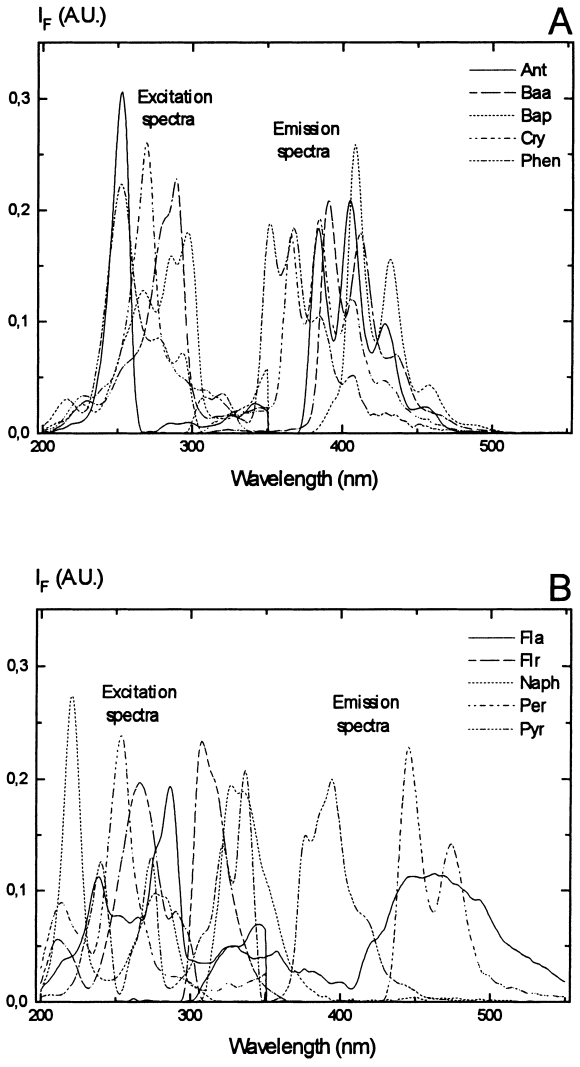


Fig. 1. Excitation and emission spectra of (A) anthracene, benz[a]anthracene, benzo[a]pyrene, chrysene and phenanthrene; (B) fluoranthene, fluorene, naphthalene, perylene and pyrene.

## 3. Data treatment

The calibration data sets (both the complete spectra and the reduced data) were analyzed using the different algorithms in order to ascertain the optimum number of factors required to model the system. For PLS models, this task was done by binary cross-validation [21]. In this procedure, the calibration set was split in two groups (containing 35 samples each); samples of group A were used as standards to calculate the concentration of the analytes in the samples of group B, and then the B-group was used as calibration set to quantify the samples in the A-group. The squared errors in the predicted concentrations for both groups were added, thus obtaining the prediction error sum of squares (PRESS):

$$\text{PRESS} = \sum_{i=1}^{ns}\sum_{j=1}^{nc}(\hat{y}_{i,j} - y_{i,j})^2,$$

where *ns* is the number of samples, and *nc* the number of components to be determined (e.g., for PLS1, $nc=1$; and for PLS2, $nc=10$), $\hat{y}_{i,j}$ the predicted concentration for the *j* component in the *i* solution, and $y_{ij}$ is the corresponding true concentration. PRESS was calculated as a function of the number of factors considered in the model. In each case, the number of factors yielding a minimum in PRESS were retained to build the final model.

Table 1 shows the results obtained for the different algorithms and spectral ranges. They have been calculated for individual compounds and for all the PAHs simultaneously (last column). These values indicate that, in general, the number of factors needed to build the model is lower when Tri-PLS procedures are used. However, there is not a clear trend when the full spectrum and the reduced data are compared. Moreover, the number of factors is very different for each PAH (between 6 and 23), depending on the particular compound and the algorithm applied. This dispersion can be partially attributed to their different relative fluorescence intensities: for example, fluorene, that shows the highest relative fluorescence, requires the lowest number of factors, only 6, while fluoranthene and naphthalene, that show the lowest relative fluorescence, need the highest number of factors (21 and 23, respectively). A trend in this sense can be observed in Fig. 4, where the number of factors needed for each compound is plotted versus their relative fluorescent intensities (taking the fluorescence intensity of the fluoranthene as the unity). However, the number of factors may depend on many other variables, such as the spectral overlapping between the different compounds, which can be very strong, as indicated by their individual excitation and emission spectra represented in Fig. 1.

As noted previously, PARAFAC only can be applied to the reduced data set. The number of factors selected

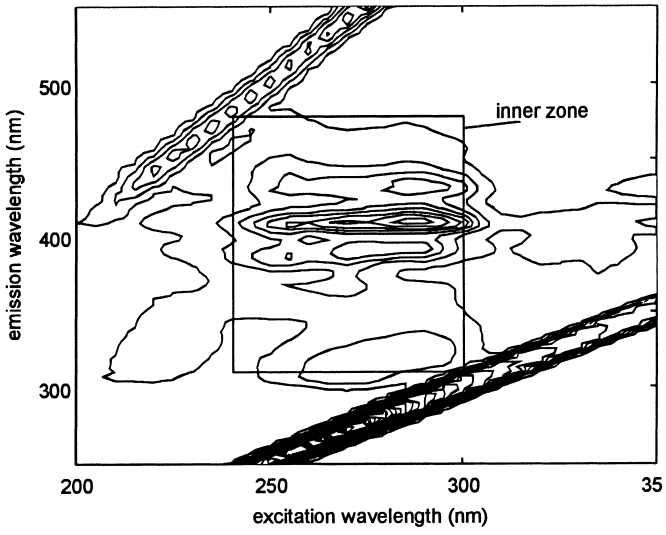


Fig. 2. Contour map of the EEM fluorescence spectrum of a mixture of the 10 PAHs.

Table 1
Number of factors obtained by cross-validation for the different PLS models

| | Ant | Baa | Bap | Cry | Phen | Fla | Flr | Naph | Per | Pyr | All |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Full spectrum* | | | | | | | | | | | |
| PLS1 | 12 | 16 | 8 | 18 | 17 | 21 | 8 | 23 | 15 | 14 | 23[a] |
| *Tri*-PLS1 | 14 | 7 | 12 | 11 | 17 | 19 | 6 | 10 | 11 | 16 | 17[b] |
| *Reduced data* | | | | | | | | | | | |
| PLS1 | 19 | 9 | 13 | 20 | 13 | 18 | 17 | 11 | 18 | 12 | 17[a] |
| *Tri*-PLS1 | 15 | 13 | 13 | 20 | 13 | 15 | 6 | 14 | 10 | 11 | 16[b] |

[a]Using PLS2.
[b]Using Tri-PLS2.

in this case (12) was obtained as giving the better spectral fit to all standards simultaneously.

## 4. Results and discussion

After each model was established, the whole set of standards (70 spectra) was used as calibration set to determine the concentration of the external validation set and of the mineral and tap water spiked with PAHs.

The external validation samples were prepared in the same way as the calibration standards, but they were not included in the cross-validation step. Ten of these 14 samples contained mixtures of all PAHs at different concentration levels, but the remaining four had only from one to five compounds, in order to check the errors at zero concentration. To evaluate the degree of fit for each procedure, the percentage of relative error was determined as the relative root mean squared errors (RRMSE):

$$\mathrm{RRMSE}(\%) = 100\sqrt{\frac{\sum_{i=1}^{ns}\sum_{j=1}^{nc}(\hat{y}_{i,j}-y_{i,j})^2}{\sum_{i=1}^{ns}\sum_{j=1}^{nc}(y_{i,j})^2}},$$

where the notation is the same as in the definition of the PRESS function.

The RRMSE values obtained from the different procedures, both for determination of single compounds and for the quantification of all PAHs simultaneously (PARAFAC, PLS2 and Tri-PLS2 methods)

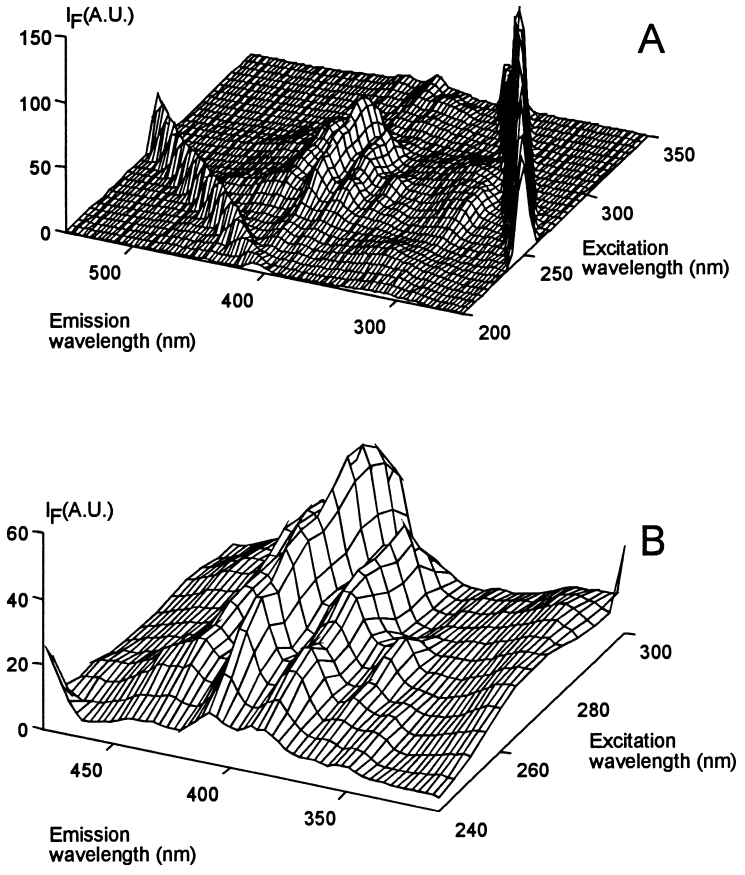


Fig. 3. EEM fluorescence spectrum of a mixture of the 10 PAHs: (A) full excitation and emission wavelength ranges; (B) reduced data set.

are shown in Table 2. As it might be expected, in all cases the best results were obtained from the reduced data set, as a consequence of the reduction of the interfering scattering bands. Although for each individual compound the minimum error corresponded to a different procedure, it is evident that Tri-PLS2 gave significantly better results than standard PLS2 (mean error of 9.29% against 12.48%) when all compounds are determined simultaneously. Data plotted in Fig. 5 show the good agreement between the true concentrations and the predicted values when the Tri-PLS2 procedure was applied to the reduced data set, as the dispersion of results obtained for a concentration of PAHs equal to zero is only of about 1 ng $ml^{-1}$.

It should be noted that PARAFAC give the worse results in most cases. This fact is related with the different approaches applied in PARAFAC and PLS algorithms: in the PARAFAC calibration, only the spectra are adjusted, whereas in PLS calibration both the spectra and the concentration of the standard are taken into account to build the model. Otherwise, the time needed for the calculations by using PARAFAC is much higher than by using the other procedures.

The procedure offering best results (reduced data in conjunction with Tri-PLS2) was selected to be used for the simultaneous quantification of all PAHs in the spiked mineral and tap water samples (five and four samples, respectively). Initially the model selected from the standards was applied, to test the prediction ability in samples which are fairly different from the standards. Results, given in the first row of Table 3, indicated that only benzo[a]anthracene, benzo[a]pyrene and pyrene gave an acceptable fit (RRMSE between 7% and 17%). The application of PLS2

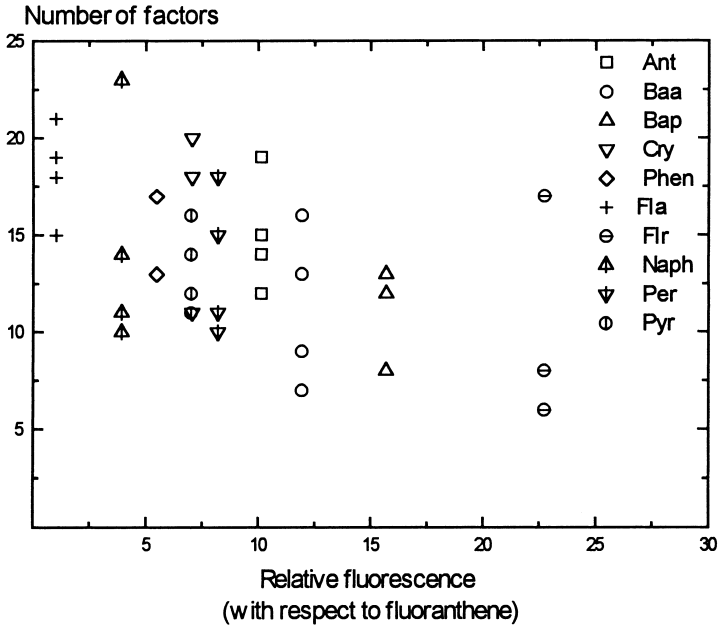


Fig. 4. Number of factors determined by cross-validation versus the relative fluorescence intensities of PAHs.

Table 2
RRMSE values for the external validation set

| | Ant | Baa | Bap | Cry | Phen | Fla | Flr | Naph | Per | Pyr | All |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Full spectrum* | | | | | | | | | | | |
| PLS1 | 6.69 | 8.83 | 8.94 | 24.59 | 10.52 | 9.42 | 11.84 | 13.64 | 8.60 | 10.41 | |
| PLS2 | 6.75 | 6.98 | 8.11 | 24.85 | 9.20 | 11.35 | 11.27 | 15.21 | 8.32 | 9.52 | 11.04 |
| Tri-PLS1 | 7.23 | 8.16 | 7.05 | 19.90 | 8.86 | 10.25 | 14.47 | 20.38 | 8.28 | 10.22 | |
| Tri-PLS2 | 8.66 | 6.85 | 8.72 | 20.99 | 8.89 | 13.39 | 14.78 | 17.33 | 7.60 | 9.45 | 11.42 |
| *Reduced data* | | | | | | | | | | | |
| PARAFAC | 11.61 | 8.06 | 7.80 | 26.14 | 6.98 | 14.62 | 16.85 | 44.48 | 11.62 | 11.20 | 17.26 |
| PLS1 | 6.20 | 6.97 | 4.96 | 18.77 | 6.47 | 7.77 | 8.85 | 14.09 | 4.91 | 10.84 | |
| PLS2 | 7.37 | 4.78 | 7.55 | 16.52 | 6.67 | 10.47 | 9.17 | 32.47 | 7.82 | 11.92 | 12.48 |
| Tri-PLS1 | 6.36 | 6.34 | 6.03 | 18.50 | 5.99 | 7.42 | 13.50 | 12.59 | 7.33 | 7.07 | |
| Tri-PLS2 | 6.03 | 4.99 | 7.70 | 14.25 | 8.38 | 11.27 | 13.61 | 14.58 | 6.36 | 7.94 | 9.29 |

calibration model, or the use of the full spectral range, did not improve these results, so that new, different models were tested. These models consisted in the determination of groups of several PAHs (instead of testing all of them simultaneously or individual compounds), applying the different procedures (PLS1, PLS2, Tri-PLS1 and Tri-PLS2) to the full EEM spectrum and to the reduced data set. The best model consisted in the simultaneous determination of seven PAHs (benz[a]anthracene, benzo[a]pyrene, chrysene, phenanthrene, fluoranthene, fluorene, and pyrene) by applying the Tri-PLS2 algorithm to the reduced wavelength ranges. The other compounds (anthracene, naphthalene and perylene) were determined individually by using the Tri-PLS1 algorithm with the same spectral range. Results, shown in the last two rows of Table 3 (under final model), indicate that five compounds (benz[a]anthracene, benzo[a]pyrene,

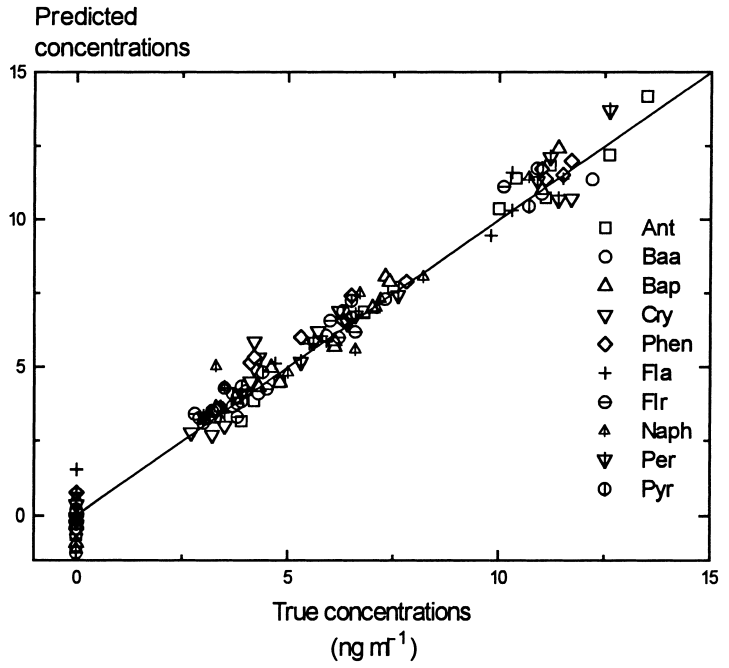


Fig. 5. Predicted and true concentrations of PAHs in the validation data set, as obtained by the *Tri*-PLS2 procedure (data from the reduced wavelength range).

Table 3
RRMSE values for spiked mineral and tap water samples (reduced data set)

| | Ant | Baa | Bap | Cry | Phen | Fla | Flr | Naph | Per | Pyr | All |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Tri-PLS2 | 42.82 | 11.13 | 17.10 | 27.34 | 30.97 | 49.27 | 27.48 | 97.70 | 39.36 | 7.03 | 55.18 |
| *Final model* | | | | | | | | | | | |
| Tri-PLS1 | 16.49 | | | | | | | 45.63 | 14.35 | | |
| Tri-PLS2 | | 7.01 | 10.31 | 13.50 | 27.51 | 10.03 | 8.82 | | | 10.58 | 13.85[a] |

[a]Only for the seven compounds quantified after *Tri*-PLS2.

fluoranthene, fluorene and pyrene) could be determined with errors of about 10%, or less; naphthalene had the worse results (errors of about 46%).

These results are also shown in Fig. 6, where the predicted concentrations are plotted versus the true concentrations of PAHs. The poor results for naphthalene are evident, with predicted values systematically lower than the target values. Results obtained for phenanthrene were generally lower than the real ones, and the rest of the compounds gave acceptable values.

To conclude, it has been shown that a standard spectrofluorimeter can be used to obtain EEM fluorescence spectra in a fast way, provided that it has the sufficient scanning speed (which is common in today's instruments) and the excitation and emission wavelength ranges to be scanned are not very large. The optimization of these ranges, therefore, plays an important role in the time needed to record the EEM spectra. In many cases, spectra containing all the required information can be recorded in only 2 or 3 min. This is also useful for the data treatment, as the reduced data set contains about a quarter of the data points in a full EEM spectrum. Combined with a multidimensional calibration procedure such as Tri-

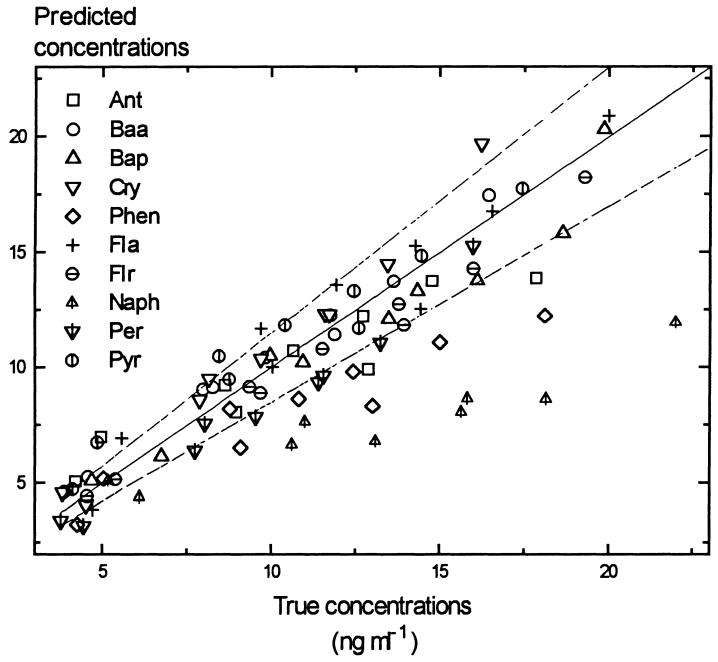


Fig. 6. Predicted and true concentrations of PAHs in the spiked water samples, after the final model. Dashed lines indicate an error interval of ±15%.

PLS, the EEM fluorescence data provide an adequate resolution of multicomponent mixtures without the need of a chromatographic separation. In this sense, it is important to remember that seven of the 10 different compounds tested in this work can be determined simultaneously in one single prediction step. Taking into account the different sensitivities and the strong spectral overlapping, these results indicate that the method described could be used in the screening of PAHs in water samples without the need of a chromatographic separation.

## References

[1] A. Bjorseth (Ed.), Handbook of Polycyclic Aromatic Hydrocarbons, Marcel Dekker, New York, 1983.
[2] C.A. Menzie, B.B. Potocki, J. Santodonato, Environ. Sci. Technol. 26 (1992) 1278.
[3] S.A. Wise, B.A. Benner Jr., R.G. Christensen, B.J. Koster, J. Kurz, M.M. Schantz, R. Zeizier, Environ. Sci. Technol. 25 (1991) 1695.
[4] T.L. Cecil, S.C. Rutan, J. Chromatogr. 556 (1991) 495.
[5] Z. Wang, M. Fingas, LC–GC 13 (1995) 950.
[6] I.M. Warner, G.D. Christian, E.R. Davidson, Anal. Chem. 49 (1977) 564.
[7] C.-N. Ho, G.D. Christian, E.R. Davidson, Anal. Chem. 50 (1978) 1108.
[8] C.-N. Ho, G.D. Christian, E.R. Davidson, Anal. Chem. 52 (1980) 1071.
[9] A.R. Muroski, K.S. Booksh, M.L. Myrick, Anal. Chem. 68 (1996) 3534.
[10] K.S. Booksh, A.R. Muroski, M.L. Myrick, Anal. Chem. 68 (1996) 3539.
[11] R. Bro, Chemom. Intell. Lab. Syst. 38 (1997) 149.
[12] P. Kroonenberg, J. de Leeuw, Psychometrika 45 (1980) 69.
[13] H. Martens, Multivariate Calibration, Wiley, New York, 1983.
[14] K.R. Beebe, B.R. Kowalski, Anal. Chem. 59 (1987) 1007A.
[15] D.M. Haaland, E.V. Thomas, Anal. Chem. 60 (1988) 1193.
[16] I.E. Frank, J.H. Friedman, Technometrics 35 (1993) 109.
[17] L. Norgaard, Talanta 42 (1995) 1305.
[18] S. Wold, P. Geladi, K. Esbensen, J. Öhman, J. Chemom. 1 (1987) 41.
[19] R. Bro, J. Chemom. 10 (1996) 47.
[20] B.M. Wise, PLS-Toolbox for use with MATLAB™, CPAC, Washington, 1992.
[21] S. Wold, Technometrics 20 (1978) 397.