ELSEVIER

Analytica Chimica Acta 361 (1998) 233–240

ANALYTICA CHIMICA ACTA

Quantitative multicomponent analysis of polycyclic aromatic hydrocarbons in water samples

J. Guiteras*, J.L. Beltrán, R. Ferrer

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

Received 29 July 1997; received in revised form 9 December 1997; accepted 15 December 1997

Abstract

A comparative study between three different multivariate calibration methods, classical least squares (CLS), principal component regression (PCR) and partial least squares regression (PLSR) was carried out. The calibration models for all three methods were obtained from a combination of two synchronous fluorescence spectra (recorded at 50 and 100 nm wavelength increments) for each standard of a calibration set of 70 standards, each containing ten polycyclic aromatic hydrocarbons (anthracene, benz[a]anthracene, benzo[a]pyrene, chrysene, fluoranthene, fluorene, naphthalene, perylene, phenanthrene and pyrene). The predictions of the model were compared with the relative root mean squared difference (RRMSD) obtained from the results of an external validation set, formed by 15 independent mixtures. Finally, the PLSR and PCR models were used for the determination of the above mentioned PAHs in spiked natural water samples at concentration levels between 4 and 20 ng ml^{-1}. Recoveries ranged from 80 to 120% in most cases, although fluorene gave significantly lower results. © 1998 Elsevier Science B.V.

Keywords: Multivariate calibration; Partial least squares; Principal component regression; Classical least squares; Polycyclic aromatic hydrocarbons

1. Introduction

Spectrofluorimetry, because of its high sensitivity and selectivity, is a powerful analytical technique for the analysis of chemical pollutants in environmental samples. The selectivity of spectrofluorimetry is based on the fact that relatively few compounds show intrinsic fluorescence and emission intensity depends on two variables, excitation and emission wavelengths. Therefore, fluorescent compounds can be determined by means of either their excitation or their emission spectra, or even by their synchronous spectra.

As a consequence of the high selectivity of spectrofluorimetry, data obtained in this way are very useful to establish multivariate calibration models in order to determine several compounds without any previous physical separation.

Full spectrum methods, which means calibration methods where the whole spectrum is used, such as partial least squares regression (PLSR), principal component regression (PCR) and classical least squares (CLS), have been found useful for quantitative analysis in spectroscopic techniques.

*Corresponding author. Fax: +34 3 4021233; e-mail: jacinto@zeus.qui.ub.es

0003-2670/98/$19.00 © 1998 Elsevier Science B.V. All rights reserved.
PII S0003-2670(98)00014-2

The first two methods are adequate if there are unknown interfering compounds present in the samples that affect the spectra. PLSR differs from other calibration procedures, such as PCR, in that it uses the concentration data from the training and the spectral data in modeling, whereas PCR only uses the spectral data [1]. Therefore, PLS can reduce the influence of dominant, but irrelevant, factors and, in some cases, yields models of lower dimensionality, in order to achieve better correlations with concentrations during predictions. The third calibration method, CLS, is used when all the compounds in the sample are known. CLS supposes that the fluorescence intensity (I) is proportional to the concentrations of the components:

$$I = k\phi_F P_o \epsilon bc$$

where k is a proportionality constant, ϕ_F the quantum yield, P_o the intensity of the excitation radiation, ϵ the molar absorptivity, b the cell path length and c the molar concentration.

The success of the CLS method is often limited by the correct selection of the spectral wavelength bands used in the calibration [2].

These calibration methods have been applied to UV–Visible absorbance spectra [3–5], fluorescence spectra [6,7], and infrared spectra [8,9]. The great advantage of multicomponent analysis is the speed of the determination, because a separation step can often be avoided.

Polycyclic aromatic hydrocarbons (PAHs) are a group of compounds whose mutagenic and/or carcinogenic effects are well known. These substances can be produced in natural and anthropogenic processes and they can be found in many different kinds of samples, both biological (e.g. meat, fish) [10–12] and environmental (e.g. soils, sediments, airborne particulates, water) [13–18]. For this reason, their detection and monitoring has become an important problem and this has led to the development of new analytical methods with improved selectivity and sensitivity.

In the present paper, PLS, PCR and CLS multivariate calibration methods have been applied to the detection and quantification of ten polycyclic aromatic hydrocarbons in a micellar medium (to enhance fluorescence and reduce quenching effects), without any previous separation, using the synchronous fluorescence spectra recorded at three different wavelength increments. For this purpose, the synchronous spectra of 70 mixtures of these compounds were used as calibration set, to establish the model, and 15 mixtures were used as external validation set to determine the predictive capacity of the method.

2. Experimental

2.1. Reagents

Stock standard solutions (about 200 μg ml^{-1}) of anthracene, benz[a]anthracene, benzo[a]pyrene, chrysene, phenanthrene, fluoranthene, fluorene, naphthalene, perylene and pyrene were prepared by dissolving the pure solid (Supelco) in either methanol or acetonitrile, depending on its solubility.

Acetonitrile and methanol were of analytical reagent quality (Merck). Doubly distilled water (Milli-Q+, Millipore) and Brij-35 (polyoxyethylenlaurylether, Merck) were used in the surfactant solutions.

2.2. Apparatus

An Aminco Bowman Series 2 spectrofluorimeter, equipped with a quartz cell (1 cm pathlength) was used for fluorescence measurements.

Preprocessing of the raw data and all calculations were performed with an IBM RS/6000 computer, using the algorithms from the PLS_ToolBox [17], written in MATLAB language (MathWorks). CLS and PCR algorithms were written by the authors in MATLAB language.

2.3. Procedure

Seventy standards, with concentrations ranging between 0 and 20 ng/ml, were prepared by dilution of known amounts of the stock solution in 25 ml of a 40 cmc ($3.6 \cdot 10^{-3}$ mol/l) solution of Brij-35. This concentration range was selected because it allowed well-defined synchronous fluorescence spectra to be obtained and it was low enough to avoid saturation of the photomultiplier in the experimental conditions used (16 nm for excitation and emission slit widths). The synchronous spectra of each solution were recorded at 10, 50 and 100 nm wavelength increments. Emission ranges were from 210 to 560 nm for the

10 nm increment, from 250 to 600 nm for the 50 nm increment and from 300 to 650 nm for the 100 nm increment, and readings were taken for each nanometer. However, the synchronous spectra obtained at 10 nm wavelength increments provided an information that was insufficient to allow discrimination of the different compounds and, therefore, these spectra were not used in further calculations.

3. Results and discussion

Both PLS and PCR are factor analysis methods, where the first step is the determination of the number of factors that allow the system to be modelled without overfitting the concentration data. For this purpose, a cross-validation method, leaving out one sample at a time, was used [19]. This means that for a calibration set with N standards, the PLS or PCR calibration was carried out with $N-1$ standards, and data thus obtained were used to calculate the concentration of the sample left out. This process was repeated for the N standards and the cumulative prediction error sum of squares (CUMPRESS), was calculated by the equation:

$$\text{CUMPRESS} = \sum_{i=1}^{N}\sum_{j=1}^{c}(\hat{y}_{i,j} - y_{i,j})^2$$

where N is the number of standards, c is the number of compounds, $\hat{y}_{i,j}$ and $y_{i,j}$ are the predicted and real concentration, respectively, of the j compound in the i standard.

The optimum model was selected taking the minimum number of factors that yielded a CUMPRESS that did not have any significant differences with the minimum CUMPRESS. For this purpose, the statistical F was used [1]. The value of F was calculated as

$$F_k = \frac{\text{CUMPRESS}(k)}{\text{CUMPRESS(min)}}$$

where F_k is the calculated value, CUMPRESS(k) is the CUMPRESS value obtained in a model with k factors, and CUMPRESS(min) is the minimum value of CUMPRESS obtained. The minimum number of factors obtained was 15 for the PLS model and 13 for the PCR model (Fig. 1).

After the number of factors of each calibration method was chosen, the models obtained for PLS and PCR were applied to an external validation set of fifteen standards. The prediction ability of the models can be expressed either in terms of the root mean square difference (RMSD):

$$\text{RMSD} = \left[\frac{1}{V}\sum_{i=1}^{V}(\hat{z}_i - z_i)^2\right]^{0.5}$$

where V is the number of samples in the validation set, and $\hat{z}_i$ and z_i are the predicted and real concentrations of a given PAH in the i sample of the validation set, or in terms of relative root mean square difference

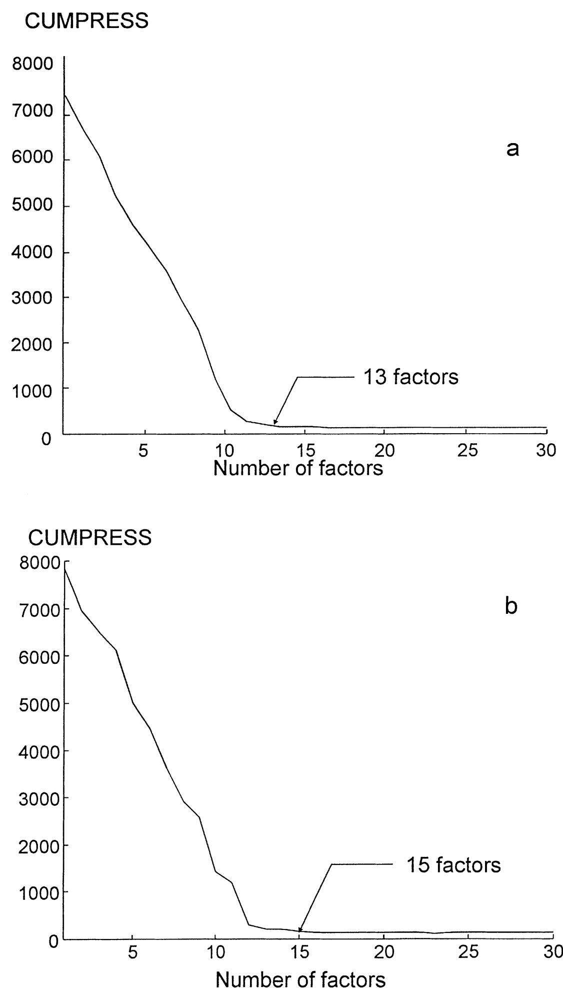

Fig. 1. Value of CUMPRESS vs. the number of factors: a) PCR model and, b) PLSR model.

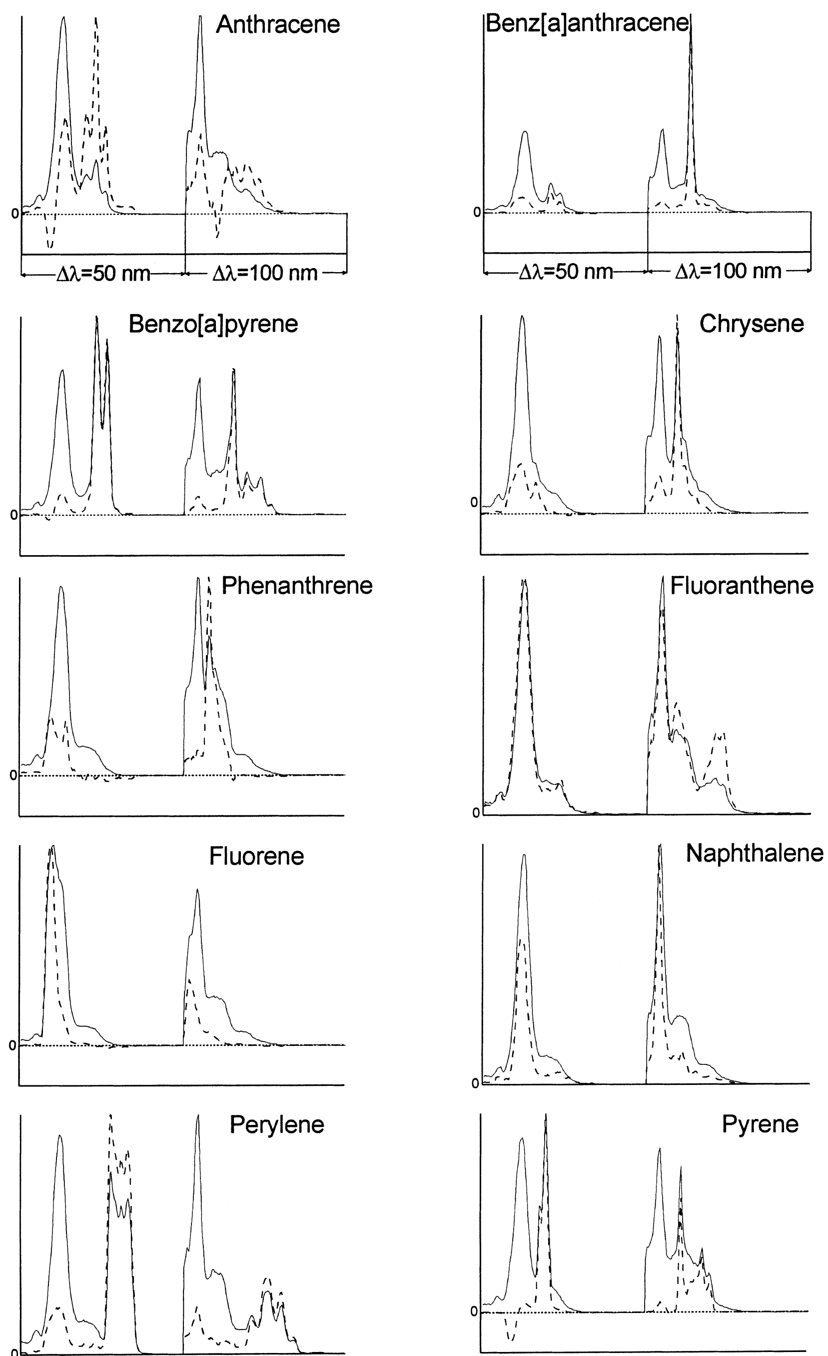

Fig. 2. Synchronous spectra at $\Delta\lambda$ of 50 and 100 nm for the ten PAHs. Solid line: real spectra, dashed line: spectra obtained from CLS.

Table 1
RRMSD (%) values for the PLSR, PCR and CLS calibration methods, using the full spectra

PAH	PLS	PCR	CLS [a]	CLS [b]	CLS [c]
Anthracene	7.34	10.23	48.14	13.40	27.02
Benz[a]anthracene	8.25	7.04	6.80	6.85	7.88
Benzo[a]pyrene	6.16	6.45	11.19	6.61	7.13
Chrysene	13.89	13.44	23.04	20.27	24.61
Fluoranthene	10.80	16.87	176.75	28.14	96.29
Fluorene	14.42	14.42	26.95	25.45	28.00
Naphthalene	9.29	15.95	82.14	16.62	66.18
Perylene	4.07	4.25	3.86	3.73	3.57
Phenanthrene	7.29	8.54	11.15	10.95	10.91
Pyrene	12.24	11.23	6.65	6.31	6.28

[a] Calculations carried out for ten compounds.
[b] Calculations carried out by adding the Brij-35 as the eleventh compound.
[c] Calculations carried out by subtracting the spectrum of the Brij-35 and considering ten compounds.

(RRMSD(%)):

$$\mathrm{RRMSD}(\%) = 100\left(\frac{\mathrm{RMSD}}{\bar{z}}\right)$$

where $\bar{z}$ indicates the mean value of z_i.

The values of RRMSD(%) obtained for each PAH, using PCR, PLS and CLS methods, are shown in Table 1. Results indicate that the prediction models obtained by PLS are slightly better than those obtained by PCR, because the values of RRMSD for anthracene, fluoranthene, naphthalene and phenanthrene are lower by PLS than by PCR, while results for the remaining compounds are similar.

For the classical least squares (CLS) calibration method, which has not a previous step of factor analysis, three models were tested: a) the full synchronous spectra were used, and ten components were modelled; b) the full synchronous spectra were used, but the surfactant (Brij-35) was added to the concentration data as the eleventh compound, using a concentration (ng ml^{-1}) equal to the unity for each sample; c) the full synchronous spectra were used, but the spectrum of the Brij-35 was subtracted from the spectra of the calibration and the validation sets.

Results shown in Table 1 indicate that CLS gave poor predictions in most cases, whatever the model, except for benz[a]anthracene, benzo[a]pyrene, perylene and pyrene, where predictions were similar to those obtained by PLS. The only advantage of CLS lies in the fact that, as shown in Fig. 2, it can calculate the pure spectra of all compounds modelled.

The three multivariate calibration methods studied (PLS, PCR and CLS) are full-spectrum methods. This means that a full-spectrum is used in the spectral calibration matrix, unlike other multivariate calibration methods, such as ILS (inverse least squares), where the number of wavelengths used cannot exceed the number of calibration mixtures. Even so, the use of a full spectrum may sometimes introduce more noise

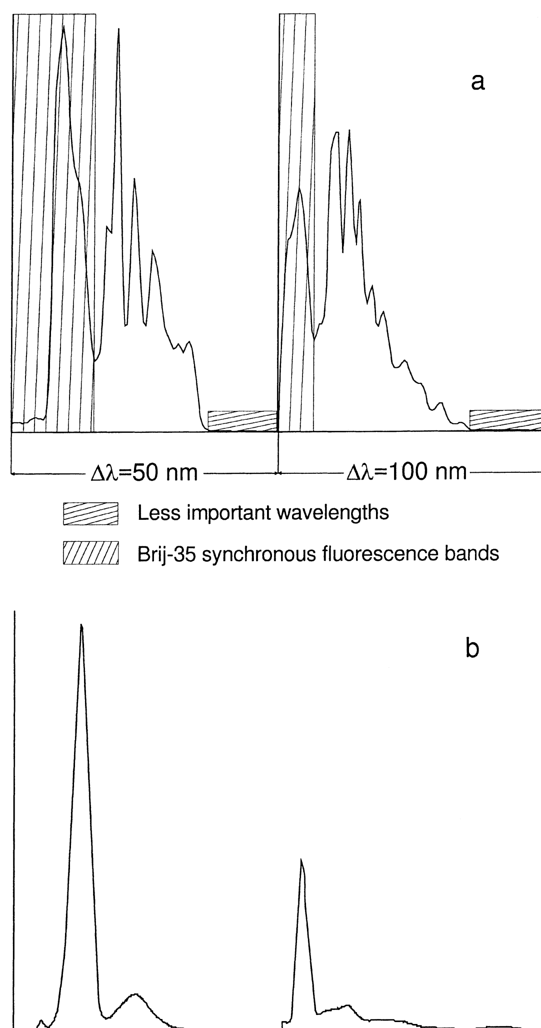

Fig. 3. a) Irrelevant zones and surfactant bands of the synchronous spectra, b) synchronous spectra of Brij-35.

than relevant information of the compounds in the samples, especially in the spectrum zones where no significative bands appear.

For this reason, experimental data were pretreated before the PLSR, PCR and CLS were calculated. Two different pretreatrements were tested: a) wavelengths where no significant fluorescence appeared were suppressed, b) wavelengths where the main fluorescence emission bands from the surfactant appeared were also suppressed (Fig. 3).

Results are shown in Tables 2 and 3. By comparing them with those in Table 1, where full spectra were used in calculations, it is evident that the suppression of the zones where no fluorescence appears has not modified the results, which clearly indicates that these zones do not have any influence. When the wavelengths where the emission bands of the surfactant appear were also suppressed, some relevant information was also lost and, therefore, predictions were generally worse, except for compounds which, like benz[a]anthracene, benzo[a]pyrene, perylene and pyrene, have intense emission bands that are well differentiated from those of the surfactant.

PCR and PLSR calibration models were used to determine PAHs in natural water samples. For this purpose, 25 ml samples of natural water were spiked with PAHs at four concentration levels (4, 9, 12 and 20 ng ml^{-1}), with all PAHS in a sample having similar concentrations, and enough Brij-35 solution was added to reach a concentration equivalent to

Table 2
RRMSD (%) values for the PLSR, PCR and CLS calibration methods, suppressing the less relevant zones from the spectra

PAH	PLS	PCR	CLS [a]	CLS [b]	CLS [c]
Anthracene	7.34	10.23	48.14	13.40	27.01
Benz[a]anthracene	8.25	7.03	6.80	6.86	7.88
Benzo[a]pyrene	6.16	6.45	11.19	6.61	7.11
Chrysene	13.88	13.44	23.04	20.27	24.61
Fluoranthene	10.80	16.87	176.75	28.13	96.29
Fluorene	14.42	14.42	26.94	25.45	28.01
Naphthalene	9.29	15.94	82.15	16.62	66.18
Perylene	4.06	4.25	3.86	3.73	3.57
Phenanthrene	7.29	8.54	11.15	10.95	10.92
Pyrene	12.25	11.25	6.65	6.31	6.28

[a] Calculations carried out for ten compounds.
[b] Calculations carried out by adding the Brij-35 as the eleventh compound.
[c] Calculations carried out by subtracting the spectrum of the Brij-35 and considering ten compounds.

Table 3
RMSD values for the PLSR, PCR and CLS calibration methods, supressing the less relevant zones and the Brij-35 bands from the spectra

PAH	PLS	PCR	CLS [a]	CLS [b]	CLS [c]
Anthracene	10.46	53.39	10.11	6.96	8.02
Benz[a]anthracene	6.33	6.44	8.23	8.64	7.67
Benzo[a]pyrene	6.14	4.07	7.78	7.19	7.72
Chrysene	26.00	22.53	24.80	27.60	23.7
Fluoranthene	17.11	66.25	25.09	17.80	21.80
Fluorene	76.50	76.01	157.87	87.09	106.86
Naphthalene	57.11	61.64	158.93	187.74	115.07
Perylene	3.98	1.23	5.46	4.15	4.42
Phenanthrene	15.91	21.77	19.01	18.24	17.96
Pyrene	10.91	10.82	11.97	11.25	11.13

[a] Calculations carried out for ten compounds.[b] Calculations carried out adding the Brij-35 as the eleventh compound.[c] Calculations carried out subtracting the spectrum of the Brij-35 and considerating ten compounds.

40 cmc. The synchronous fluorescence spectra, recorded at wavelength increments of 50 and 100 nm, were used to calculate the concentrations. The results, shown in Table 4, indicated a good agreement, in the case of PLSR, between the amounts added and found, and only fluorene had recoveries (ratio amount found/amount added) lower than 30%, which can probably be attributed to interferences from the sample matrix, concretely to fluorescence from Brij-35. As could be expected, results obtained by PCR, while still good, showed a worse agreement with added values.

Table 4
Results obtained in natural spiked water samples for PCR and PLS models

Compound	PCR			PLS	
	Added (ng ml^{-1})	Found ± std (ng ml^{-1}) [a]	Recovery ± std(%)	Found ± std (ng ml^{-1}) [a]	Recovery ± std(%)
Anthracene	4.2	6.1±0.3	145.2±7.1	4.5±0.4	106.1±9.9
	8.9	9.3±0.2	104.5±2.2	6.6±1.0	74.3±11.5
	12.9	10.4±0.5	80.6±3.9	8.2±0.9	63.7±7.2
	17.9	15.2±0.9	84.9±5.0	11.0±1.6	61.6±9.2
Benz[a]anthracene	3.9	4.3±0.3	110.3±7.7	4.2±0.1	108.5±1.7
	8.2	8.7±0.5	106.1±6.1	8.4±0.1	102.2±0.2
	11.9	11.1±0.6	93.3±5.0	11.3±0.5	93.8±0.9
	16.4	16.9±0.9	103.0±5.5	16.5±0.3	100.4±1.8
Benzo[a]pyrene	4.7	5.5±0.4	117.0±8.5	4.9±0.2	103.8±3.7
	10.0	11.0±0.8	110.0±8.0	10.1±0.6	101.7±5.6
	14.3	13.9±0.8	97.2±5.6	13.2±0.5	92.1±3.6
	19.3	21.2±1.2	109.8±6.2	20.1±0.9	103.7±4.9
Chrysene	3.8	2.3±0.2	60.5±5.3	4.1±0.4	106.4±11.1
	8.1	6.5±0.2	80.2±2.5	7.1±0.4	87.9±5.2
	11.7	9.0±0.3	76.9±2.6	9.2±0.6	78.8±5.1
	16.2	14.9±0.2	92.0±1.2	13.4±0.7	82.9.6
Fluoranthene	4.7	9.5±0.6	202.1±12.8	6.7±0.1	143.2±2.3
	10.0	15.3±0.4	153.0±4.0	11.1±0.5	110.9±5.4
	13.1	17.5±0.7	133.6±5.3	13.5±0.6	93.4±4.1
	20.0	26.±0.7	130.0±3.5	20.2±0.7	101.3±3.4
Fluorene	4.5	−0.2±0.3	—	0.3±0.2	7.5±3.5
	9.7	2.8±0.2	28.9±2.1	1.6±0.1	16.8±0.9
	13.9	5.2±0.2	37.4±1.4	3.5±0.1	25.4±0.7
	19.3	8.8±0.3	45.6±1.6	5.4±0.1	28.0±0.6
Nephthalene	5.2	1.5±0.4	28.8±7.7	3.8±0.3	74.6±5.8
	11.0	5.5±0.2	50.0±1.8	8.1±0.6	73.4±5.1
	13.9	9.3±0.4	66.9±2.9	10.9±1.1	68.7±6.8
	21.9	14.6±0.4	66.7±1.8	17.8±1.0	81.2±4.7
Perylene	3.8	5.1±0.3	134.2±7.9	4.7±0.1	123.8±1.8
	8.0	9.9±0.6	123.7±7.5	8.7±0.1	108.2±0.6
	11.5	12.5±0.5	108.7±4.3	11.2±0.1	96.8±0.4
	16.0	19.2±0.8	120.0±5.0	16.7±0.1	104.5±0.9
Phenanthrene	4.3	2.1±0.4	48.8±9.3	4.8±0.2	113.0±5.7
	9.1	5.3±0.3	58.2±3.3	8.6±0.2	95.1±2.5
	13.1	7.5±0.2	57.3±1.5	10.5±0.4	80.6±3.0
	18.1	10.9±0.2	60.2±1.1	15.5±0.5	85.4±2.5
Pyrene	4.1	5.7±0.6	139.0±14.6	4.2±0.2	103.4±3.7
	8.7	11.2±1.3	128.7±14.9	8.4±0.1	96.6±1.2
	12.6	13.5±1.0	107.1±7.9	11.0±0.1	87.7±0.8
	17.4	20.6±1.9	118.4±0.9	16.2±0.4	93.1±2.6

[a] Mean of three independent samples.

4. Conclusions

A previous study of the synchronous spectra of the analytes and their mixtures is very useful to determine the less relevant spectral zones, which can be eliminated without any adverse effect on the predictive performance of the models. This avoids the use of large amounts of irrelevant data, which will only increase the time of analysis and the requirements of computer memory. Calculation times required to model the seventy standards of the calibration set when using full spectra were 9 min 35 s for PLS, 3 min for PCR and about 5 s for CLS. These times were reduced by 10% when the non-relevant areas of the spectra were suppressed. The suppression of the spectrum of the surfactant is not recommended, because some information is also lost and predictions are generally worse, except for compounds having bands that are well differentiated from those of the surfactant.

The best predictive performances were achieved when using the partial least squares regression modelling, although PCR gave also good results. This was confirmed when these two procedures were used for the determination of PAHs in spiked water samples. The concentrations predicted by the PLS procedure were, in most cases, within 80–120% of the real values.

References

[1] D.M. Haaland, E.V. Thomas, Anal. Chem. 60 (1988) 1193.
[2] P.C. Schmidt, B.W. Glombitza, Trends Anal. Chem. 14 (1995) 45–49.
[3] P.X. Zhang, D. Littlejohn, Chem. Intel. Lab. Syst. 34 (1996) 203.
[4] A. Espinosa Mansilla, F. Salinas, M. Del Olmo, I.P. Paya, App. Spec. 50 (1996) 449.
[5] F. Navarro-Villoslada, L.V. Pérez-Arribas, M.E. León-González, L.M. Polo-Diez., Anal. Chim. Acta (313) (1995) 93–101.
[6] R.D.B. Jimenez, A.I.J. Abizanda, F.J. Moreno, J.J.A. Leon, Clin. Chim. Acta 249 (1996) 21.
[7] R. Bro, J. Chemometr. 10 (1996) 47.
[8] F.R. Vandevoort, K.P. Memon, J. Sedman, A.A. Ismail, J. Am. Oil Chem. Soc. 73 (1996) 411.
[9] J. Dubois, F.R. Vandevoort, J. Sedman, A.A. Ismail, H.R. Ramaswamy, J. Am. Oil Chem. Soc. 73 (1996) 787.
[10] A. Bjorseth (Ed.), Handbook of Polycyclic Aromatic Hydrocarbons, Marcel Dekker, New York, 1983.
[11] C.A. Menzie, B.B. Potocki, J. Santodonato. Environ. Sci. Technol. 26 (1992) 1278.
[12] 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.
[13] A.I. Krilov, I.O. Kostyuk, N.F. Volynets, J. Anal. Chem. 50 (1995) 494.
[14] H.G. Kicinski, S. Adamek, A Kettrup, Chromatographia 28 (1989) 203.
[15] E.R. Brouwer, A.N.J. Hermans, H. Lingeman, U.A.Th. Brinkman, J. Chromatogr. A 669 (1994) 45.
[16] G. Codina, M.T. Vaquero, L. Comellas, F.B. Puig, J. Chromatogr. A 673 (1994) 21.
[17] S.A. Wise, M.M. Schantz, B.A. Benner, Jr., M.J. Hays, S.B. Schiller, Anal. Chem. 67 (1995) 1171.
[18] C. Escrivá, E. Viana, J.C. Moltó, Y. Picó, J. Mañes, J. Chromatogr. 676 (1994) 375.
[19] A. Garrido-Frenich, M. Martínez-Galera, J.L. Martínez-Vidal, M.D. Gil-García, J. Chromatogr. A 727 (1996) 27.