View Article Online / Journal Homepage / Table of Contents for this issue
Dynamic Article Links

Analytical Methods

Cite this: *Anal. Methods*, 2011, **3**, 2616

www.rsc.org/methods

PAPER

Published on 06 October 2011. Downloaded by State University of New York at Stony Brook on 29/10/2014 14:21:05.

Simultaneous quantification of dilute aqueous solutions of certain polycyclic aromatic hydrocarbons (PAHs) with significant fluorescent spectral overlap using total synchronous fluorescence spectroscopy (TSFS) and *N*-PLS, unfolded-PLS and MCR-ALS analysis

Keshav Kumar and A. K. Mishra*

Received 17th August 2011, Accepted 25th August 2011
DOI: 10.1039/c1ay05513e

The possibility of simultaneous quantification of dilute aqueous solutions of five polycyclic aromatic hydrocarbons (PAHs) with substantial overlap of fluorescence spectra without preseparation was examined by using total synchronous fluorescence spectroscopy (TSFS) with multivariate methods like *N*-way partial least square (*N*-PLS), unfolded-PLS and multivariate curve resolution-alternating least square (MCR-ALS) analysis. The PAHs chosen were anthracene, benzo[a]pyrene, chrysene, perylene and pyrene. Two calibration sets were made using different approaches. Even with significant spectral overlap, the *N*-PLS, unfolded-PLS and MCR-ALS models were found to be robust for the quantification of all five PAHs in the calibration set where the amounts of PAHs were changed with the fixed relative ratio. The second calibration set, which is relatively difficult to analyse, consisted of samples where amounts of PAHs were changed randomly. It was found that even for this set, predictions were satisfactory. The three calibration models based on *N*-PLS, unfolded-PLS and MCR-ALS in the second set were relatively more robust for chrysene, benzo[a]pyrene and perylene compared to other PAHs. Root mean square error of calibration (RMSEC) and root mean square error of prediction (RMSEP) and regression parameters were calculated for all the three calibration models. The various parameters show that the combination of TSFS and *N*-PLS, unfolded-PLS and MCR-ALS analysis can be used for the simultaneous quantification of PAHs in water without any preseparation even when there is substantial overlap of fluorescence.

Introduction

Polycyclic aromatic hydrocarbons (PAHs) are a well known class of environmental pollutants.[1–4] The increase in the number of industries and hydrocarbon driven vehicles are the two major reasons why the level of PAHs in the environment has increased in recent years. Oil spills, industrial wastes, run-off from the roads and sewage are a few common reasons why these PAHs enter into the water.[2,5] PAHs are found to be carcinogenic and mutagenic,[3,4] as a result their presence in water above a certain level is a serious issue. According to the Bureau of Indian standards the amount of total PAHs in drinking water should not exceed 0.2 ng ml^{-1} and in particular the amount of benzo[a] pyrene in water should be less than 0.01 ng ml^{-1}. Gas chromatography with mass spectrometry or liquid chromatography with UV-visible or fluorescence spectroscopy is generally used for the analysis of PAHs in water.[6–9] These techniques are expensive, time consuming and require laborious sample preparation steps. PAHs are fluorescent in nature which allows the use of fluorescence spectroscopy, a simple, sensitive and cost-effective technique for their analysis in water. Excitation-emission matrix fluorescence (EEMF)[10] and synchronous fluorescence spectroscopy (SFS)[11] are the two fluorescence techniques which are commonly used for the analysis of samples containing a mixture of fluorophores with extensive overlap of absorption and emission spectra. EEMF and SFS have been used successfully for the analysis of dissolved organic matter in water.[12–17] However, the complex fluorescence spectral behaviour of PAH mixtures does not allow the use of a fluorescence intensity based univariate calibration model for their reliable quantitative interpretation. In recent years there have been a number of reports in the literature which show that mixtures of PAHs in water can be successfully analysed using a combination of EEMF with multivariate methods such as partial least square (PLS), parallel factor analysis (PARAFAC), alternating trilinear decomposition (ATLD), multivariate curve resolution alternating least square (MCR-ALS) and *N*-way partial least square (*N*-PLS).[5,18–20] Combinations of SFS and multivariate methods have also been used for the analysis of PAHs in water.[21–23]

Department of Chemistry, Indian Institute of Technology-Madras, Chennai, 600036, India. E-mail: mishra@iitm.ac.in; Fax: (+91) 44-22574202; Tel: (+91) 44-22574207

This journal is © The Royal Society of Chemistry 2011

View Article Online

Published on 06 October 2011. Downloaded by State University of New York at Stony Brook on 29/10/2014 14:21:05.

Total synchronous fluorescence spectroscopy (TSFS), introduced by Patra and Mishra,[24,25] is similar to EEMF and allows simultaneous visualization of the fluorescence response of all the fluorophores of a mixture in a single three dimensional (3D) plot. TSFS plots are constructed by plotting the excitation or emission wavelength along the *x*-axis, different wavelength offset values along the *y*-axis and synchronous fluorescence intensity values along the *z*-axis. One of the major advantages of TSFS plots is that the Rayleigh scattering of the excitation light, which appears as prominent ridges in the EEMF plots, is easily avoided by TSFS. Scattered light does not contain any fluorescence information and its removal from the EEMF data set is necessary before data analysis, whereas no such correction is required in the TSFS data set. TSFS alone and along with multivariate methods has been used for the qualitative analysis of petroleum products, natural oils and food samples;[24,26–30] it has also been used for the diagnosis of cancer[31] and PAT application in bioprocesses.[32]

The objective of the present work is to see whether TSFS along with multivariate methods such as *N*-PLS, MCR-ALS and unfolded-PLS can be used for the simultaneous quantification of PAHs with overlapping fluorescence in dilute aqueous solutions without any preseparation. The choice of dilute aqueous solutions and samples with significant fluorescence overlap, which makes the system difficult to analyze, is deliberate. To the best of our knowledge TSFS along with multivariate methods have not been used so far for the quantitative analysis of multi-fluorophores in aqueous samples. The analyses were carried out using five polycyclic aromatic hydrocarbons, anthracene, chrysene, benzo[a]pyrene, perylene and pyrene, having significant overlap of florescence spectra. Two different calibration sets were chosen for the analyses. In the first set, a fixed ratio of concentration was maintained among the samples and their overall concentration in water was varied within the ng ml^{-1} levels. Since in more realistic cases, the amount as well as the relative ratios of these PAHs in water samples are expected to vary randomly, the second calibration set was made in which the concentration of the PAHs in a particular sample was permitted to vary randomly.

Materials and methods

Chemicals and sample preparation

Anthracene (AN), benzo[a]pyrene (BP), chrysene (CY), perylene (PE) and pyrene (PY) were obtained from Aldrich and used as such without any further purification. A stock solution (S1) was prepared by dissolving 11.25 mg of AN, 12.25 mg of BP, 7.5 mg of CY, 12.00 mg of PE and 9.25 mg of PY in 500 ml of spectroscopic grade methanol.

In the set of experiments involving the first calibration set C1 of 19 samples and validation set V1 of 6 samples, solutions were prepared by diluting the stock solution (S1) with triple distilled water. The amounts of these five PAHs were changed from sample to sample, however, the relative ratio of the PAHs was kept the same in each sample of the C1 and V1 set.

For the second calibration set C2 and its corresponding validation set V2, separate stock solutions were prepared for each of the five PAHs by dissolving 23.5 mg of AN, 25.5 mg of BP, 52.5 mg of CY, 22.5 mg of PE and 20.5 mg of PY in 500 ml of methanol, and they were labelled as S2, S3, S4, S5, and S6 respectively. C2 set of 17 samples and V2 set of 6 samples were made by taking the different amounts of these PAHs from their respective stock solutions followed by dilution with triple distilled water. To see whether the fluorescence intensity which is directly proportional to concentration, has any effect on the prediction abilities of these multivariate models, CY in most of the samples of the C2 set was taken in excess compare to other PAHs. The amounts of all the five PAHs in the standard solutions of the calibration sets C1 and C2 are given in Table 1 and Table 2, respectively.

Instrumentation and data acquisition

Fluorescence measurements were performed using a Fluoromax 4 (Horiba Jobin Yvon) spectrofluorometer, with a xenon lamp of 150 W as the excitation source. The band pass for the excitation and emission monochromator was kept at 5 nm. Synchronous fluorescence spectra were recorded in the excitation wavelength

Table 1 Amounts of AN, BP, CY, PE, and PY in calibration set C1

Sample	AN (ng ml^{-1}) × 10^{-1}	BP (ng ml^{-1}) × 10^{-1}	CY (ng ml^{-1}) × 10^{-1}	PE (ng ml^{-1}) × 10^{-1}	PY (ng ml^{-1}) × 10^{-1}
1	0.90	0.98	0.60	0.96	0.74
2	2.70	2.94	1.80	2.88	2.22
3	3.60	3.92	2.40	3.84	2.96
4	4.50	4.90	3.00	4.80	3.70
5	5.40	5.88	3.60	5.76	4.44
6	7.20	7.84	4.80	7.68	5.92
7	8.10	8.82	5.40	8.64	6.66
8	9.00	9.80	6.00	9.60	7.40
9	9.90	10.78	6.60	10.56	8.14
10	11.70	12.74	7.80	12.48	9.62
11	12.60	13.72	8.40	13.44	10.36
12	13.50	14.70	9.00	14.40	11.10
13	14.40	15.68	9.60	15.36	11.84
14	16.20	17.64	10.80	17.28	13.32
15	17.10	18.62	11.40	18.24	14.06
16	18.00	19.60	12.00	19.20	14.80
17	19.80	21.56	13.20	21.12	16.28
18	21.60	23.52	14.40	23.04	17.76
19	22.50	24.50	15.00	24.00	18.50

View Article Online

Published on 06 October 2011. Downloaded by State University of New York at Stony Brook on 29/10/2014 14:21:05.

range of 250–650 nm with 14 different wavelengths offset from 20 to 150 nm with an interval of 10 nm. The concentrations of the five PAHs were low in all the samples therefore synchronous spectra were collected in the corrected mode which takes care for the possible instrument biases in the lamp intensity.

N-way partial least square (*N*-PLS) analysis

N-PLS[33–35] is an extension of PLS to handle three or more dimensional data sets. The *N*-PLS model is fitted in such a way that it maximizes the covariance between dependent and independent variables. This is achieved by simultaneously fitting a multilinear model of the dependent variables ($\underline{\mathbf{Y}}$), a multilinear model of the independent variables ($\underline{\mathbf{X}}$) and a regression model relating the two decomposition models. Here, $\underline{\mathbf{X}}$ ($\mathbf{I} \times \mathbf{J} \times \mathbf{K}$) and $\underline{\mathbf{Y}}$ ($\mathbf{I} \times \mathbf{M} \times \mathbf{N}$) are decomposed as,

$$\mathbf{X} = \mathbf{T}\,(\mathbf{W}^{\mathbf{K}}|\otimes|\mathbf{W}^{\mathbf{J}})^{\mathbf{T}} + \mathbf{E} \quad (1)$$

$$\mathbf{Y} = \mathbf{U}\,(\mathbf{Q}^{\mathbf{N}}|\otimes|\mathbf{Q}^{\mathbf{M}})^{\mathbf{T}} + \mathbf{F} \quad (2)$$

where **X** and **Y** are unfolded matrices of size ($\mathbf{I} \times \mathbf{JK}$) and ($\mathbf{I} \times \mathbf{MN}$), respectively. **I**, **J**, and **K** are the number of samples, excitation wavelength, and wavelength offsets, respectively. **T** and **U** are the first mode score matrices, $\mathbf{W}^{\mathbf{J}}$ and $\mathbf{Q}^{\mathbf{M}}$ are the second mode weight matrices, $\mathbf{W}^{\mathbf{K}}$ and $\mathbf{Q}^{\mathbf{N}}$ are the third mode weight matrices, **E** and **F** are the residual matrices of $\underline{\mathbf{X}}$ and $\underline{\mathbf{Y}}$, respectively. The symbol $|\otimes|$ represents the Khatri-Rao product. The weight matrices have the maximal covariance with score matrices **T** and **U**. A regression model relating the two decomposition model is given as,

$$\mathbf{U} = \mathbf{T}\,\mathbf{B} + \mathbf{G} \quad (3)$$

where B is the regression matrix and G is the residual matrix.

Unfolded-PLS analysis

In the unfolded-PLS method a three-way array is unfolded along the first mode and PLS analysis is done to make the calibration model.[33]

Multivariate curve resolution-alternating least square (MCR-ALS) analysis

MCR-ALS is a well explained method in literature,[36–38] briefly it is used to extract the concentration and spectral profiles of each component from the spectral data matrix of a mixture. MCR-ALS decomposes the spectral data matrix, **A** ($\mathbf{I} \times \mathbf{V}$), into two matrices, **C** ($\mathbf{I} \times \mathbf{D}$) which contains a relative concentration profile and **S** ($\mathbf{V} \times \mathbf{D}$) matrix of pure component spectra, where **I** is the number of samples, **V** is the wavelength and **D** is the number of components in the mixture.

$$\mathbf{A} = \mathbf{C}\,\mathbf{S}^{\mathbf{T}} + \mathbf{H} \quad (4)$$

H is the residual matrix in eqn (4). The MCR-ALS method starts with an initial estimate of the concentration or spectral profiles, if the initial estimates are concentration profiles then the estimated spectral profile is given by eqn (5). In the next step the concentration profile is estimated using eqn (6), a number of iteration steps are performed between eqn (5) and (6) to get the physically meaningful concentration and spectral profile.

$$\mathbf{S}^{\mathbf{T}} = \mathbf{C}\,(\mathbf{C}^{\mathbf{T}}\,\mathbf{C})^{-1}\,\mathbf{A} \quad (5)$$

$$\mathbf{C} = \mathbf{A}\,\mathbf{S}\,(\mathbf{S}^{\mathbf{T}}\,\mathbf{S})^{-1} \quad (6)$$

Since concentration and spectral values can not be negative, the above two equations were solved under the non-negative constraints. The obtained concentration profile is regressed with the actual concentration to make calibration plots which can be used to predict the PAHs in new samples.

Root mean square error of calibration (RMSEC) and root mean square error of prediction (RMSEP)

RMSEC measures the difference between the predicted and reference value for samples of calibration. It is defined as,

$$\mathrm{RMSEC} = \sqrt{\frac{\sum_{i=1}^{n}(y_{\mathrm{pred}} - y_{\mathrm{ref}})^2}{n}} \quad (7)$$

Table 2 Amounts of AN, BP, CY, PE, and PY in calibration set C2

Sample	AN (ng ml^{-1}) $\times 10^{-1}$	BP (ng ml^{-1}) $\times 10^{-1}$	CY (ng ml^{-1}) $\times 10^{-1}$	PE (ng ml^{-1}) $\times 10^{-1}$	PY (ng ml^{-1}) $\times 10^{-1}$
1	5.64	8.16	4.20	9.00	3.28
2	7.52	10.20	8.40	1.80	4.92
3	9.40	2.04	12.60	3.60	6.56
4	3.76	6.12	21.00	7.20	1.64
5	5.35	7.70	19.80	8.50	1.54
6	7.14	9.62	3.96	10.20	3.09
7	8.92	11.55	7.92	3.40	4.63
8	3.57	5.77	15.84	6.80	7.72
9	3.57	7.70	19.80	8.50	3.09
10	8.92	3.85	11.88	5.10	7.72
11	3.57	3.85	19.80	5.10	9.27
12	5.35	5.77	23.76	6.80	3.09
13	8.92	9.62	11.88	1.70	6.18
14	8.92	11.55	7.92	1.70	6.18
15	3.57	5.77	15.84	5.10	9.27
16	5.35	7.70	19.80	6.80	3.09
17	7.14	9.62	3.96	8.50	4.63

 This journal is © The Royal Society of Chemistry 2011

View Article Online

Published on 06 October 2011. Downloaded by State University of New York at Stony Brook on 29/10/2014 14:21:05.

where y_{pred} and y_{ref} are the predicted and the reference value of the i^{th} sample respectively and n is the number of samples used.[35,39] Eqn (7) can also be used for the calculation of RMSEP of the validation or testing set, where n is the number of samples used for the validation.[35,39]

Software used

N-PLS, unfolded-PLS and MCR-ALS analysis were carried out using PLS_Toolbox 5.0.3 written in MATLAB language.

Results and discussion

The synchronous fluorescence spectra (SFS) of all the samples of the calibration and validation sets were collected with different wavelength offsets and the TSFS data set of these samples was constructed. Fluorescence of the various mixtures of five hydrocarbons was mainly seen in the excitation range of 290–470 nm with the wavelength offsets starting from 20 to 150 nm. The TSFS contour plots of the pure AN, BP, CY, PR and PY in water are shown in Fig. 1. From the figure it can be seen that there is a substantial overlap of the spectra of all the five PAHs, as a result the simultaneous quantification of these five PAHs in water using a univariate calibration method is difficult. Multivariate methods which involve the analysis of more than one variable at a time can be used for the analysis of such complex mixtures without any separation.

Before carrying out the analysis the TSFS and concentrations data needed to be arranged. For the *N*-PLS analysis the TSFS data of C1 and C2 were arranged in a three way array of dimensions (sample × excitation wavelength × wavelength offset), 19 × 401 × 14 and 17 × 401 × 14, respectively. Three way arrays of TSFS data were unfolded along the first mode, thereby two dimensional data set of sizes 19 × 5614 and 17 × 5614 were obtained for the C1 and C2 calibrations sets, respectively for the unfolded-PLS and MCR-ALS analyses. Concentration data of C1 and C2 sets were arranged in matrices of dimensions (sample × components), 19 × 5 and 17 × 5, respectively. Each column of these matrices contains the concentration information of a particular hydrocarbon in all the samples. Since there are five PAHs in all the samples, 5 factors were used for *N*-PLS, unfolded-PLS and MCR-ALS analyses. In order to verify the choice of the optimum number of factors, principal component analysis (PCA)[39] was used. PCA is a technique which is generally used for the estimation of a possible number of factors or to reduce the dimension of the data set.[38–40] TSFS data of both the calibration sets C1 and C2 were unfolded along the first mode and mean-centred before the PCA analysis. Leave one out cross-validation (LOOCV)[35,39] was used for the development of reliable PCA models without over fitting the data. PCA models with different numbers of factors (*i.e.* principal component) for the C1 and C2 set were made using the data of all the samples except one. The obtained models were then used to estimate the left out sample in the respective calibration sets. This procedure was repeated so that every sample of both the calibration sets was left and estimated once. Root mean square error of cross-validation (RMSECV)[35] values were calculated and plotted against the number of factor, shown in Fig. 2. From the plots it can be seen that the PCA model developed using five factors for the C1 and C2 set has a minimum RMSECV value. Hence, the 5 factors were found to be valid for the multivariate analyses.

N-PLS, unfolded-PLS and MCR-ALS analyses of calibration set C1

The *N*-PLS models of 5 factors for the calibration sets C1 were made using the pre-processed (*i.e.* mean centred or autoscaled) and unprocessed (*i.e.* actual) three way array of TSFS and concentration data. It was found that the *N*-PLS model developed without pre-processing the data gives a better result; therefore *N*-PLS analysis was performed without any pre-processing. The *N*-PLS model was found to explain the 89.44% and 99.97% variances of the spectral and concentration data sets, respectively. The unfolded-PLS analysis of the calibration set C1 was carried out using the unfolded three way array of TSFS and the concentration data of the C1 set. The spectral and concentration data were mean centred before the analysis, the unfolded-PLS model of 5 factors explained the more than 99% variances of the spectral and concentration data sets. MCR-ALS analysis of the C1 set was done with the unprocessed unfolded TSFS data set. The analysis was initiated with an estimate of the concentration profile, since concentrations and spectral profile cannot be negative; the analysis was subjected to the non-negative constraints. The MCR-ALS model of 5 factors captured 97.27% variance of the spectral data set. The concentration profiles of each of the PAHs obtained from the MCR-ALS model were regressed with their actual concentrations to make the calibration model. Linear regression equations, relating the actual amounts of each of the PAHs and the corresponding *N*-PLS,

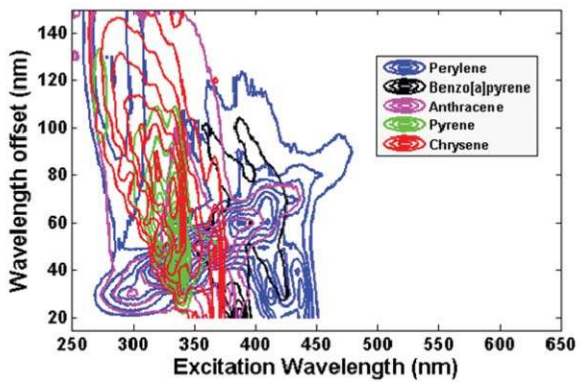

Fig. 1 TSFS contour plots of anthracene, benzo[a]pyrene, chrysene, perylene, and pyrene.

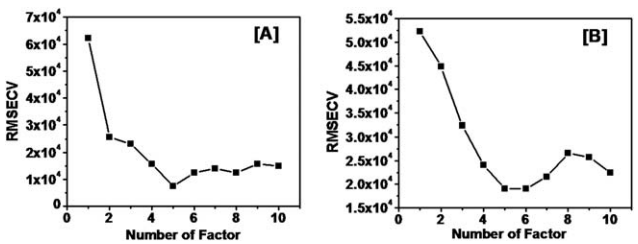

Fig. 2 RMSECV *versus* number of factor plot of the PCA models of [A] C1 and [B] C2 calibration set.

View Article Online

Published on 06 October 2011. Downloaded by State University of New York at Stony Brook on 29/10/2014 14:21:05.

Table 3 Linear regression equations of *N*-PLS, unfolded-PLS and MCR-ALS models for the calibration set C1

	N-PLS		Unfolded-PLS		MCR-ALS	
PAH	Reg. eq.	R^2	Reg. eq.	R^2	Reg. eq.	R^2
AN	$y = 1.00*x + 0.0012$	0.999	$y = 1.00*x + 0.0002$	0.999	$y = 0.99*x + 0.0091$	0.999
BP	$y = 1.00*x + 0.0014$	0.999	$y = 1.00*x + 0.0003$	0.999	$y = 0.99*x + 0.0099$	0.999
CY	$y = 1.00*x + 0.0008$	0.999	$y = 1.00*x + 0.0002$	0.999	$y = 0.99*x + 0.0061$	0.999
PE	$y = 1.00*x + 0.0013$	0.999	$y = 1.00*x + 0.0003$	0.999	$y = 0.99*x + 0.0097$	0.999
PY	$y = 1.00*x + 0.0010$	0.999	$y = 1.00*x + 0.0002$	0.999	$y = 0.99*x + 0.0075$	0.999

unfolded-PLS and MCR-ALS predicted amounts, are given in Table 3. The RMSEC values for these five PAHs in all the three models were also calculated and are reported in Table 4. The low RMSEC values and linear regression equation with slope close to one and intercept close to zero shows that the actual amounts of the PAHs in the samples of the C1 set are correlated with the model predicted amounts.

In order to test the prediction abilities of these three models, the validation set (V1) of six samples were used. The actual amount of PAHs in the samples of the V1 set and predicted

Table 4 RMSEC and RMSEP values of the *N*-PLS, unfolded-PLS and MCR-ALS models for the C1 set

	N-PLS		Unfolded-PLS		MCR-ALS	
PAH	RMSEC (ng ml^{-1}) × 10^{-1}	RMSEP (ng ml^{-1}) × 10^{-1}	RMSEC (ng ml^{-1}) × 10^{-1}	RMSEP (ng ml^{-1}) × 10^{-1}	RMSEC (ng ml^{-1}) × 10^{-1}	RMSEP (ng ml^{-1}) × 10^{-1}
AN	0.24	0.75	0.23	0.68	0.55	0.90
BP	0.26	0.82	0.25	0.74	0.59	0.98
CY	0.16	0.50	0.15	0.45	0.36	0.60
PE	0.25	0.80	0.24	0.72	0.58	0.96
PY	0.20	0.62	0.19	0.56	0.45	0.74

Table 5 Actual and *N*-PLS, unfolded-PLS and MCR-ALS predicted amounts of PAHs in the samples of validation set V1

			Predicted amount (ng ml^{-1}) × 10^{-1}		
Sample	PAH	Actual amount (ng ml^{-1}) × 10^{-1}	*N*-PLS	Unfolded-PLS	MCR-ALS
1	AN	1.80	1.66	1.63	2.09
	BP	1.96	1.8	1.78	2.27
	CY	1.20	1.1	1.09	1.39
	PE	1.92	1.77	1.74	2.23
	PY	1.48	1.36	1.34	1.72
2	AN	6.30	6.22	6.12	6.07
	BP	6.86	6.77	6.66	6.61
	CY	4.20	4.15	4.08	4.05
	PE	6.72	6.63	6.52	6.48
	PY	5.18	5.11	5.03	4.99
3	AN	10.80	10.19	10.4	11.97
	BP	11.76	11.09	11.32	13.03
	CY	7.20	6.79	6.93	7.98
	PE	11.52	10.87	11.09	12.76
	PY	8.88	8.38	8.55	9.84
4	AN	15.30	16.01	16.02	16.19
	BP	16.66	17.44	17.44	17.62
	CY	10.20	10.68	10.67	10.79
	PE	16.32	17.08	17.08	17.26
	PY	12.58	13.17	13.17	13.31
5	AN	18.90	17.36	17.53	17.71
	BP	20.58	18.91	19.08	19.28
	CY	12.60	11.58	11.68	11.8
	PE	20.16	18.52	18.69	18.89
	PY	15.54	14.28	14.41	14.56
6	AN	0.70	20.33	20.31	19.61
	BP	22.54	22.13	22.11	21.35
	CY	13.80	13.55	13.54	13.08
	PE	22.08	21.68	21.66	20.92
	PY	17.02	16.71	16.70	16.13

 This journal is © The Royal Society of Chemistry 2011

View Article Online

Published on 06 October 2011. Downloaded by State University of New York at Stony Brook on 29/10/2014 14:21:05.

results of all the three models are summarized in Table 5. All the three models predicted amounts of PAHs which were close to the actual amounts. The RMSEP value, a measure of the accuracy with which a model predicts the unknown concentration, of all the three models were calculated for each of the five PAHs, and are given in Table 4. The small RMSEP, RMSEC values and parameters of linear regression show that the obtained *N*-PLS, unfolded-PLS and MCR-ALS models are robust.

The above results demonstrate that the combination of multivariate methods and TSFS is capable of doing simultaneous quantification of five PAHs in water even at ng ml^{-1} level. As mentioned earlier the amount of PAHs in the C1 set changes from sample to sample with a fixed relative ratio. Therefore, even though there are significant spectral overlaps of these PAHs, the fluorescence intensities at all wavelengths change in a definite manner from one sample to another, as a result calibration modelling with such a spectral data set becomes relatively easy. Hence, in order to verify the strength of the combination of TSFS and multivariate methods, *N*-PLS, unfolded-PLS and MCR-ALS analyses were further carried out with the calibration set C2 which is a more realistic set where the amount as well as the relative ratios of the PAHs vary in a random fashion.

N-PLS, unfolded-PLS and MCR-ALS analysis of calibration set C2

TSFS and concentration data of the C2 calibration set were mean-centred along the first mode, and a *N*-PLS model of 5 factors was made. These 5 factors were found to capture 80.77% and 91.31% variances of the spectral and concentration data sets, respectively. The unfolded-PLS model of 5 factors was made using the mean centred unfolded TSFS and concentration data, the model explained the 97.05% and 91. 54% variances of the spectral and concentration data set. MCR-ALS analysis was carried out with unfolded TSFS data and an initial estimate of the concentration data, the MCR-ALS model captured 83.37% variance of the spectral data. As in the previous set C1, here also the MCR-ALS analysis was done with the non-negative constraints. The estimated amounts obtained from the MCR-ALS model were regressed with the actual amounts of PAHs. The linear regressions were carried out between the actual amounts of the PAHs in the samples of the calibration sets and the *N*-PLS, unfolded-PLS and MCR-ALS model predicted amounts. The regression equations and the RMSEC values of these three calibration models for the five PAHs are given in Table 6 and 7, respectively. From the regression parameters it can be inferred that the *N*-PLS, unfolded-PLS and MCR-ALS models are more accurate in predicting the amounts of CY in most of the calibration set samples, which can be attributed to the fact that its fluorescence intensity is higher compare to other hydrocarbons and it is probably so because the majority of the samples of the calibration set C2 contain relatively large amounts of CY compared to other PAHs. The amounts of BP and PE were also predicted accurately by the *N*-PLS and unfolded-PLS model. The MCR-ALS model predictions were accurate for BP in most of the samples but for unknown reasons its estimations for the PE amounts were not impressive. All the three models were found to make relatively poor estimations of AN and PY amounts which may be due to less fluorescence intensity and significant spectral overlap with other hydrocarbons. The prediction abilities of the models were tested using the six samples of validation set V2; the actual and predicted amounts of the PAHs are given in Table 8. The RMSEP values of the calibration models for each of the PAHs were also calculated and are shown in Table 7. The small RMSEC and RMSEP values give an impression that the obtained calibration models are robust for all the five PAHs, however regression parameters clearly shows this is not the case, calibration models are relatively more reliable for CY, BP and PE compare to AN and PY. We can also say that the accurate estimation of a component becomes easy if it has higher fluorescence intensity compare to other components, for example in the present case the fluorescence due to CY in the samples of C2 set is higher compare to other PAHs and its estimation in most of the

Table 6 Linear regression equations of *N*-PLS, unfolded-PLS and MCR-ALS models for the calibration set C2

	N-PLS		Unfolded-PLS		MCR-ALS	
PAH	Reg. eq.	R^2	Reg. eq.	R^2	Reg. eq.	R^2
AN	$y = 0.62*x + 0.240$	0.619	$y = 0.60*x + 0.250$	0.600	$y = 0.54*x + 0.290$	0.540
BP	$y = 0.85*x + 0.110$	0.847	$y = 0.86*x + 0.100$	0.860	$y = 0.74*x + 0.190$	0.743
CY	$y = 0.99*x + 0.018$	0.989	$y = 0.99*x + 0.018$	0.987	$y = 0.93*x + 0.092$	0.932
PE	$y = 0.87*x + 0.077$	0.868	$y = 0.88*x + 0.072$	0.877	$y = 0.55*x + 0.270$	0.547
PY	$y = 0.76*x + 0.120$	0.765	$y = 0.78*x + 0.110$	0.780	$y = 0.65*x + 0.180$	0.649

Table 7 RMSEC and RMSEP values of the *N*-PLS, unfolded-PLS and MCR-ALS models for the C2 set

	N-PLS		Unfolded-PLS		MCR-ALS	
PAH	RMSEC (ng ml^{-1}) × 10^{-1}	RMSEP (ng ml^{-1}) × 10^{-1}	RMSEC (ng ml^{-1}) × 10^{-1}	RMSEP(ng ml^{-1}) × 10^{-1}	RMSEC (ng ml^{-1}) × 10^{-1}	RMSEP (ng ml^{-1}) × 10^{-1}
AN	1.33	0.80	1.37	0.89	1.47	0.88
BP	1.05	0.96	1.00	0.95	1.35	1.21
CY	0.73	0.63	0.73	0.64	1.67	1.30
PE	0.95	0.71	0.92	0.81	1.77	1.41
PY	1.16	0.67	1.12	0.68	1.41	0.90

View Article Online

Published on 06 October 2011. Downloaded by State University of New York at Stony Brook on 29/10/2014 14:21:05.

Table 8 Actual and *N*-PLS, unfolded-PLS and MCR-ALS predicted amount of PAHs in the samples of validation set V2

			Predicted amount (ng ml^{-1}) × 10^{-1}		
Sample	PAH	Actual amount (ng ml^{-1}) × 10^{-1}	*N*-PLS	Unfolded-PLS	MCR-ALS
1	AN	8.46	7.99	7.93	8.08
	BP	6.12	6.90	7.05	6.93
	CY	10.50	10.78	10.88	12.15
	PE	2.70	3.59	3.60	3.45
	PY	5.74	6.57	6.41	5.30
2	AN	6.58	7.91	8.16	7.39
	BP	9.18	8.25	8.5	10.00
	CY	6.30	6.06	5.98	6.36
	PE	5.40	5.70	5.20	4.89
	PY	4.10	5.34	5.43	4.46
3	AN	6.25	5.62	5.70	6.04
	BP	5.77	5.30	5.34	5.78
	CY	15.84	16.99	16.83	15.52
	PE	6.80	5.96	5.76	5.32
	PY	5.41	5.86	6.04	7.28
4	AN	6.58	5.45	5.38	6.08
	BP	4.08	5.63	5.63	4.88
	CY	16.8	17.62	17.75	11.49
	PE	5.40	6.59	6.77	4.45
	PY	4.10	4.48	4.30	5.21
5	AN	8.92	8.57	8.37	7.38
	BP	10.59	10.06	9.94	8.28
	CY	9.90	9.86	9.97	11.49
	PE	1.70	1.74	2.00	4.45
	PY	6.18	6.20	6.06	5.21
6	AN	6.25	6.28	6.25	7.32
	BP	8.66	7.60	7.62	7.45
	CY	11.88	11.42	11.30	10.46
	PE	7.65	7.75	7.89	6.82
	PY	3.86	3.57	3.55	4.07

samples are more accurate. On comparing the various measures of the calibration models it can be concluded that the *N*-PLS and unfolded-PLS are more robust compared to the MCR-ALS model.

The calibration models of the C1 and C2 set were also validated by selecting some samples from the C2 and C1 set, respectively the predictions made by the models are given in Table 9 and 10.

Table 9 Validation of *N*-PLS, unfolded-PLS and MCR-ALS models of C1 set by the samples of C2 set

			Predicted amount (ng ml^{-1}) × 10^{-1}		
Sample	PAH	Actual amount (ng ml^{-1}) × 10^{-1}	*N*-PLS	Unfolded-PLS	MCR-ALS
1	AN	7.52	8.65	9.10	8.43
	BP	10.20	9.34	9.45	10.87
	CY	8.40	8.28	7.95	9.06
	PE	1.80	2.96	2.10	2.67
	PY	4.92	6.20	6.27	4.42
2	AN	8.92	8.61	9.08	9.63
	BP	11.55	11.99	12.3	12.01
	CY	7.92	7.48	8.79	8.53
	PE	3.40	5.15	4.74	4.26
	PY	4.63	4.75	4.96	4.34
3	AN	8.92	6.60	6.87	7.27
	BP	3.85	3.69	4.20	4.23
	CY	11.88	11.21	12.54	11.75
	PE	5.10	5.34	4.86	4.41
	PY	7.72	7.79	8.32	9.17
4	AN	7.14	6.89	7.77	9.76
	BP	9.62	8.89	8.47	9.29
	CY	3.96	3.20	2.98	2.26
	PE	8.50	7.81	8.00	6.64
	PY	4.63	4.02	4.35	4.11

 This journal is © The Royal Society of Chemistry 2011

View Article Online

Published on 06 October 2011. Downloaded by State University of New York at Stony Brook on 29/10/2014 14:21:05.

Table 10 Validation of *N*-PLS, unfolded-PLS and MCR-ALS models of C2 set by the samples of C1 set

Sample	PAH	Actual amount (ng ml^{-1}) × 10^{-1}	Predicted amount (ng ml^{-1}) × 10^{-1}		
			N-PLS	Unfolded-PLS	MCR-ALS
1	AN	2.70	3.34	2.63	1.70
	BP	2.94	2.43	2.56	2.59
	CY	1.80	1.36	1.97	1.76
	PE	2.88	3.71	4.34	2.41
	PY	2.22	1.87	1.82	1.60
2	AN	5.40	5.87	4.67	4.89
	BP	5.88	5.23	4.95	5.14
	CY	3.60	3.64	5.17	3.23
	PE	5.76	5.21	5.31	5.01
	PY	4.44	4.23	4.82	4.98
3	AN	7.20	7.67	6.69	5.20
	BP	7.84	7.21	8.51	6.87
	CY	4.80	4.07	4.12	4.87
	PE	7.86	7.34	7.25	7.61
	PY	5.92	5.56	4.84	5.23
4	AN	8.10	8.74	9.01	10.02
	BP	8.82	8.06	7.43	8.23
	CY	5.40	5.87	3.23	5.21
	PE	8.64	9.20	8.23	8.79
	PY	6.62	7.36	6.16	6.23

The obtained results show that the combination of TSFS with multivariate methods can be used for the quantitative analysis of the five PAHs in water without any preseparation.

Conclusions

In the present work it was shown that the TSFS technique along with multivariate methods can be used for the simultaneous quantification of PAHs in water. The strength of the technique was tested by selecting those PAHs which have significant spectral overlap with each other. Two calibration sets C1 and C2 were made with different approaches. *N*-PLS, unfolded-PLS and MCR-ALS models were found to be robust in predicting the amounts of all the five hydrocarbons present in the samples of the C1 set. The obtained calibration models of the C2 set, which is more realistic and comparatively difficult to analyse because of random variation in the amounts of PAHs and significant spectral overlap, were found to be relatively more robust for the quantification of the CY, BP and PE compared to AN and PY. RMSEC and RMSEP values of the calibration models for C1 and C2 were found to be small which shows that the developed models are capable of estimating the PAHs amount simultaneously without any preseparation. In summary, all the three models *N*-PLS, unfolded-PLS and MCR-ALS developed using TSFS data were found to be reliable in the concentration ranges for which these models were developed.

Acknowledgements

Keshav Kumar is thankful to the Council of Scientific and Industrial Research (CSIR) New Delhi for providing the fellowship.

References

1 A. Biorseth, *Handbook of polycyclic aromatic hydrocarbons*, Marcel Decker, New York, 1983.
2 C. R. Spier, G. G. Vadas, S. L. Kaattari and M. A. Unger, *Environ. Toxicol. Chem.*, 2011, **30**, 1557–1563.
3 T. Vo-Dinh, *Chemical analysis of polycyclic aromatic compounds*, Wiley-Interscience, New York, 1989.
4 J. M. Neff, *Polycyclic aromatic hydrocarbons in the aquatic environment: sources, fate and biological effects*, Applied Sciences, London, 1979.
5 H. Wang, Y. Zhang and X. Xiao, *Anal. Sci.*, 2010, **26**, 1271–1276.
6 E. Manoli and C. Samara, *TrAC, Trends Anal. Chem.*, 1999, **18**, 417–428.
7 Z. Kiss, J. V. Puchonv and J. Hlavay, *J. Chromatogr., A*, 1996, **725**, 261–272.
8 M. B. Deschamps, J. J. Daudin and E. Barriuso, *J. Chromatogr., A*, 2007, **1167**, 143–153.
9 F. P. Koltalo, K. Oukebdane, L. Robin and P. L. Desbene, *Talanta*, 2007, **71**, 1825–1833.
10 J. H. Rho and J. L. Stuart, *Anal. Chem.*, 1978, **50**, 620–625.
11 J. B. F. Lloyd, *Nat. Phys. Sci.*, 1971, **231**, 64–65.
12 M. Z. Bieroza, J. Bridgeman and A. Baker, *Drinking Water Eng. Sci.*, 2010, **3**, 63–70.
13 A. Baker, E. Tipping, S. A. Thacker and D. Gondar, *Chemosphere*, 2008, **73**, 1765–1772.
14 C. L. Muller, A. Baker, R. Hutchinson, I. J. Fairchild and C. Kidd, *Atmos. Environ.*, 2008, **42**, 8036–8045.
15 S. E. Cabaniss and M. S. Shuman, *Mar. Chem.*, 1987, **21**, 37–50.
16 G. M. Ferrari and M. Mingazzini, *Mar. Ecol.: Prog. Ser.*, 1995, **125**, 305–315.
17 A. Vodacek, *Remote Sens. Environ.*, 1989, **30**, 239–247.
18 J. L. Beltrán, R. Ferrer and J. Guiteras, *Anal. Chim. Acta*, 1998, **373**, 311–319.
19 H. B. Wang, Y. J. Zhang, X. Xiao, S. H. Yu and W. Q. Liu, *Anal. Methods*, 2011, **3**, 688–695.
20 M. V. Bosco and M. S. Larrechi, *Talanta*, 2007, **71**, 1703–1709.
21 D. Patra and A. K. Mishra, *Talanta*, 2001, **55**, 143–153.
22 J. L. Beltrán, R. Ferrer and J. Guiteras, *Talanta*, 1998, **45**, 1073–1080.
23 M. Algarra, V. Jiménez, P. F. de Violet and M. Lamotte, *Anal. Bioanal. Chem.*, 2005, **382**, 1103–1110.
24 D. Patra and A. K. Mishra, *Anal. Bioanal. Chem.*, 2002, **373**, 304–309.
25 D. Patra and A. K. Mishra, *TrAC, Trends Anal. Chem.*, 2002, **21**, 787–798.
26 A. G. Ryder, *J. Fluoresc.*, 2004, **14**, 99–104.
27 A. K. sarma and A. G. Ryder, *Energy Fuels*, 2006, **20**, 783–785.
28 A. Nevin, D. Comelli, G. Valentini and R. Cubeddu, *Anal. Chem.*, 2009, **81**, 1784–1791.
29 A. Sahar, T. Boubellouta, S. Portanguen, A. Kondjoyan and ÉDufour, *J. Food Sci.*, 2009, **74**, 534–539.
30 K. I. Poulli, G. A. Mousdis and C. A. Georgiu, *Food Chem.*, 2007, **105**, 369–375.
31 T. Dramićanin, B. Dimitrijević and M. D. Dramićanin, *Appl. Spectrosc.*, 2011, **65**, 293–297.

View Article Online

32 A. P. Teixeira, T. M. Duarte, M. J. T. Carrondo and P. M. Alves, *Biotechnol. Bioeng.*, 2011, **108**, 1852–1861.
33 R. Bro, *J. Chemom.*, 1996, **10**, 47–61.
34 R. Bro, Ph.D. Thesis, *Multi-way analysis in the food industry*, University of Amsterdam, 1998.
35 B. M. Wise, N. B. Gallaghar, R. Bro and J. M. Shaver, *PLS_Toolbox 4.0. Eigen-vector research*, 2006.
36 R. Tauler and B. Kowalski, *Anal. Chem.*, 1993, **65**, 2040–2047.
37 R. Tauler, *Chemom. Intell. Lab. Syst.*, 1995, **30**, 133–146.
38 T. Azzouz and R. Tauler, *Talanta*, 2008, **74**, 1201–1210.
39 R. Kramer, *Chemometric techniques for quantitative analysis*, Marcel Decker, New York, 1998.
40 M. Maeder and A. Zilian, *Chemom. Intell. Lab. Syst.*, 1988, **3**, 205–213.

Published on 06 October 2011. Downloaded by State University of New York at Stony Brook on 29/10/2014 14:21:05.

 This journal is © The Royal Society of Chemistry 2011