JOURNAL OF CHEMOMETRICS
J. Chemometrics 2003; **17**: 608–617
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/cem.829

Optimizing the implementation of the PARAFAC method for near-real time calibration of excitation–emission fluorescence analysis

Michelle L. Nahorniak and Karl S. Booksh*

Department of Chemistry and Biochemistry, Arizona State University, Tempe, AZ 85287-1604, USA

Received 14 May 2002; Revised 12 September 2003; Accepted 17 October 2003

A field-portable, single-exposure excitation–emission matrix (EEM) fluorometer is used in conjunction with parallel factor analysis (PARAFAC) for sub-ppb polycyclic aromatic hydrocarbon (PAH) determinations in the presence of spectral interferents. Several strategies for bringing multiway calibration methods such as PARAFAC into the field were explored. It was shown that automated methods of PARAFAC model selection can be as effective as manual selection. In addition, it was found that there is not always a single best model to employ for prediction. Second, the effect that reducing data density by systematically decreasing calibration set size and spectral resolution has on PARAFAC speed and prediction accuracy was investigated. By decreasing data density, the computational intensity of the PARAFAC algorithm can be reduced to increase the plausibility of on-the-fly data analysis. It was found that reducing eight sample PAH calibration sets to two or three calibration standards significantly decreased computation intensity yet generated adequate predictions. It was also found that spectral resolution can be decreased to reach an optimal compromise between calibration accuracy and analysis speed while minimizing instrumental requirements.

KEYWORDS: PARAFAC; excitation–emission matrix; polycyclic aromatic hydrocarbons; experimental design; *N*-way calibration

1. INTRODUCTION

Section 307 Toxic Pollutants of the United States' Clean Water Act (CWA) lists polycyclic aromatic hydrocarbons (PAHs) as toxic pollutants. The toxic effects of PAHs are well documented [1,2]. Nineteen PAHs are further listed as Priority Pollutants under the CWA. Although the determination of PAHs in aquatic systems is of great concern owing to their suspected carcinogenic properties (see the Environmental Response Conservation and Recovery Act (CERCLA)), there are several problems associated with PAH detection in aqueous systems such as oceans, lakes, run-off and ground water. First, owing to the extremely low water solubility of PAHs, their concentrations in water are very low. For example, total PAH concentrations may range from sub-parts-per-billion (sub-ppb) in highly contaminated sewage to sub-parts-per trillion (sub-pptr) in ground water supplies [3]. Preconcentration techniques such as solid phase extraction (SPE) are traditionally required prior to analysis [4,5]. This increases analysis time, cost, reagent use and the possibility of sample loss.

A second problem is that both natural and waste waters usually contain a number of organic species, often in relatively high concentrations. In addition, because several different PAHs can be generated by a single process, samples are usually presented as complex mixtures. Gas and/or liquid chromatography, which are relatively slow compared with the actual PAH detection, are often employed to separate individual components of a sample prior to analysis. Because these chromatographic methods are time-consuming and not readily field-portable, samples need to be collected and brought back to the laboratory for analysis at a later time. When samples are transported prior to analysis, problems such as PAH degradation [6] and absorption by the sampling vessels [7] can occur.

Because of increasing environmental concerns, there is a strong and growing demand for sensitive, real time, on-site analytical methods for detecting PAHs in both natural and waste waters. For such a method to be practical in the field, traditional sample pretreatments must be minimized or eliminated and analytes probed in their native environment. This would simplify procedures, save considerable time and

*Correspondence to: K. S. Booksh, Department of Chemistry and Biochemistry, Arizona State University, Tempe, AZ 85287-1604, USA.
E-mail: kbooksh@asu.ed

money and eliminate several sources of error. Such a method could be used to make an initial 'map' of a large area to determine spatial variations and to identify sites where further, more in-depth testing may be called for.

2. BACKGROUND

Polycyclic aromatic hydrocarbons are natively fluorescent and the fluorescence spectrum can be easily observed without chemical modification of the compounds. Although fluorescence spectroscopy is an extremely sensitive analytical technique [8], traditional fluorescence spectra are broad and featureless. Owing to this lack of selectivity, fluorescence detection is often coupled with separation techniques such as high-performance liquid chromatography (HPLC) [9]. Traditional fluorescence detection methods measure total fluorescence at a single excitation and emission wavelength pair where the analyte of interest is known to fluoresce [10]. This limits detection to a single, known analyte and does not account for interfering fluorophores at or near the same wavelength pair.

Collecting a two-dimensional total fluorescence spectrum, termed an excitation–emission matrix (EEM) spectrum, increases selectivity. In EEM spectroscopy a total fluorescence spectrum is obtained by systematically varying the excitation and emission wavelengths and collecting the resulting $I \times J$ data matrix. The ensemble spectrum may be collected by a photomultiplier tube or silicon diode detector by rastering across all excitation and emission wavelengths. However, an alternative instrumental design is to collect the entire EEM spectrum in one measurement with a two-dimensional charge-coupled device (CCD) camera. In either case, each of the I rows in the EEM spectrum is the emission spectrum at the ith excitation wavelength. Each of the J columns in the EEM spectrum is the excitation spectrum at the jth emission wavelength. If viewed from above as a contour plot, it is much like a topological map (Figure 1).

Although scanning each excitation wavelength at each emission wavelength and combining the individual spectra produces an EEM spectrum, the data collection time for such a practice is impractical for *in situ* analysis. In addition, since each scan is taken at a different time, in a dynamic environment, the individual spectra do not truly represent the same sample. Although never practical for routine use, single-measurement EEM fluorescence was first introduced in 1977 by Christian and co-workers [11,12]. More recently, Muroski *et al.* [13] revisited single-measurement EEM spectroscopy, employing the relatively recently developed planar array CCD.

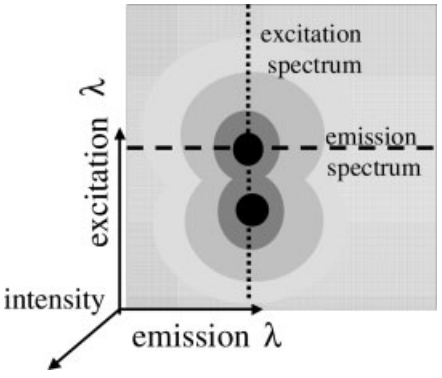

Figure 1. Depiction of an excitation–emission matrix viewed as a contour plot. Each row represents an emission spectrum at a particular excitation wavelength, and each column an excitation spectrum at a particular emission wavelength.

The real advantage of multiway data such as EEM spectra is the ability to employ powerful, multiway deconvolution methods on all the available data to maximize the information content extracted from a data set. Algorithms such as PARAFAC utilize the interrelated relationships between multiple attributes ('ways') to simultaneously extract multiple species, even in the presence of unknown, uncalibrated interferents. In the case of EEM spectra, PARAFAC can be used to deconvolute overlapping spectra, even if no unique spectral portion exists, thus eliminating or minimizing the need for sample pretreatments and separations.

Parallel factor analysis (PARAFAC) is one of several deconvolution methods for multiway data. Two-way data can consist of multiple samples each represented by a set of variables, such as NIR spectra of a single analyte at different concentrations, or of one sample represented by two sets of interacting variables, such as in GC–MS. In EEM spectroscopy, modulating individual data points of an excitation spectrum according to emission wavelength produces second-order data. A third 'way' can be added by combining correlated second-order data. In the case of EEM, several EEM spectra (e.g. a concentration or time series) can be stacked to form a three-way 'data cube' (Figure 2). This data cube can be mathematically broken down into a set of trilinear components,

$$R_{i,j,k} = \sum X_{i,n} Y_{j,n} Z_{k,n} + E_{i,j,k} \qquad (1)$$

where $R_{i,j,k}$ is the fluorescence intensity of sample k at excitation wavelength i and emission wavelength j, N is the number of unique spectral profiles found in the data cube, the columns of **X**, **Y** and **Z** are the estimations of the pure excitation, emission and concentration profiles respectively and $E_{i,j,k}$ is the error matrix. PARAFAC uses an iterative, least squares-type algorithm to simultaneously

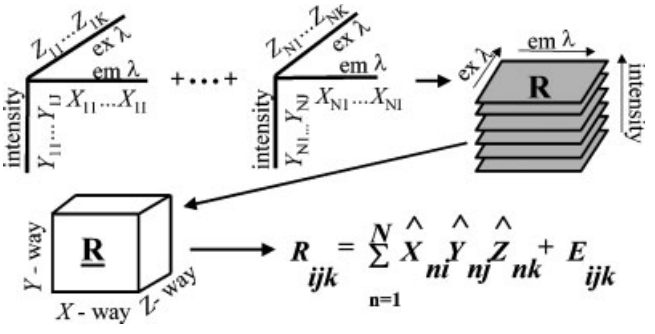

Figure 2. Stacking correlated EEM spectra forms a three-way data cube. The cube can be mathematically modeled by a set of trilinear components, where $R_{i,j,k}$ is the fluorescence intensity of sample k at excitation wavelength i and emission wavelength j, N is the number of unique spectral profiles found in the data cube, the columns of **X**, **Y** and **Z** are the estimations of the pure excitation, emission and concentration profiles respectively and $E_{i,j,k}$ is the error matrix.

extract the N pure spectral profiles, even in the presence of unknown, uncalibrated interferents [14,15]. This decomposition of $\underline{\mathbf{R}}$ is unique in $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ except for a scaling factor [16].

Because of the inherent complexity of environmental samples, the utility of multiway calibration methods in environmental analyses is becoming increasingly explored. For example, da Silva and Oliveira [17] use PARAFAC decomposition of UV–vis spectra to examine ligand substitution kinetic methods in mixtures of heavy metal ions. Johnson *et al.* [18] apply a direct trilinear decomposition (DTD) algorithm and a three-way multivariate curve resolution–alternating least squares (MCR–ALS) technique to data from liquid chromatography coupled with diode array detection (LC–DAD) to study herbicide hydrolysis. Crouch *et al.* [19] evaluate the use of the three-way calibration procedures N-way PLS (nPLS) and PARAFAC for kinetic–spectrophotometric determinations of metal ions.

The utility of multiway methods method for deconvoluting collections of EEM spectra in particular has also been demonstrated. Beltran *et al.* [20] employ the three-way multivariate calibration methods Tri-PLS2 and PARAFAC to resolve individual EEM spectra from mixtures of 10 polycyclic aromatic hydrocarbons. Wentzell *et al.* [21] describe the application of trilinear decomposition (TLD) to EEMs of PAH mixtures quenched by nitromethane. Muroski *et al.* [22] demonstrate the use of EEM in conjunction with PARAFAC for hydrocarbon determination in ocean water. JiJi *et al.* [23] apply PARAFAC to EEM data for the calibration of DDT-type pesticides, carbamate pesticides and PAHs.

3. THEORY

Despite its efficacy, there are several problems associated with employing a multiway analysis method such as PARAFAC in the field. Real time analysis is limited by the required user input and the speed of the algorithm. To effectively realize the utility of three-way analyses, especially for real time, *in situ* analysis, both the instrumentation and the data analysis methods must be optimized to reach a compromise between the reliability of the results and the 'effort' (simplicity, speed, cost) exerted to obtain that information. Optimization of spectral resolution and calibration density and simultaneous automation of model selection will minimize both user expertise and computational intensity requirements. When the data density is minimized, not only the computation time but also the computer power required for analysis can be minimized.

Although a wealth of analytical information can be extracted from multiway data, a problem arises as the data become increasingly dense. For example, an excitation–emission matrix (EEM) that consists of 50 emission spectra at 30 different excitation wavelengths, or alternatively 30 excitation spectra at 50 different emission wavelengths, forms a 1500-element matrix. If an EEM is collected at 10 different analyte concentrations, the data become a 15 000-element cube with the three ways being excitation wavelength, emission wavelength and concentration. The larger the data cube becomes, the more computationally intensive data analysis becomes. Decreasing spectral resolution and including the minimum number of standards possible to achieve target levels of accuracy and precision can minimize the speed of generating a PARAFAC model. The density of the excitation $\times$ emission $\times$ sample cube can be reduced in any or all of the three dimensions or 'ways'. The excitation and emission ways can be reduced by either collecting data over a narrower wavelength range, by 'compressing' the data by decreasing the instrumental resolution of the grating or the CCD, or by spectral averaging. The sample way can be reduced by decreasing the number of calibration standards.

Several examples of efforts to reduce computation time can be found in the literature. Alsberg and Kvalheim [24] present B-spline methods to compress multiway data for more efficient handling of very large data sets, and Alsberg *et al.* [25] apply the method to second-order FT-IR spectra. Bro and de Jong [26] propose a modification of the standard non-negativity-constrained linear least squares regression algorithm to speed up PARAFAC-type iterative algorithms when applying non-negativity constraints. Bro and Andersson [27] develop a compression approach that can speed up estimation in constrained factor models. Anderrson and Bro [28] attempt to improve the speed of the Tucker-3 multiway algorithm to make it fast enough to be suitable for large data arrays. While any of these methods for algorithmically increasing PARAFAC speed are beneficial, this paper focuses on simple instrumental and experimental approaches to handle the same problem. It is probable that the algorithmic and experimental design approaches could be combined for greater improvements.

A second limitation to PARAFAC is that, although the true number of latent factors, N, is unknown in most situations, it must be specified by the analyst. If the correct number of factors is chosen, the model predicts the true spectral profiles of each component. If the chosen N is too small, a portion of the variance in unaccounted for, which results in the incomplete deconvolution of overlapped spectra. If N is chosen too high, one of several situations may occur, including the appearance of degenerate, extraneous or inverse factors. An automated method is needed to efficiently and consistently select an appropriate calibration model.

For an analytical calibration with a specific set of data the 'best' PARAFAC model is the one that results in the lowest prediction error, which is commonly represented by the root mean square error of prediction according to the equation

$$\mathrm{RMSEP} = \left(\sum (c_i - \hat{c}_i)^2 / N\right)^{1/2} \qquad (2)$$

where RMESP is the square root of the sum of squared errors between predicted and actual concentrations of the analyte of interest in an unknown and N is the total number of 'unknown' mixtures used in the model. However, when transferred to a different set of data, that particular choice of N may or may not give the best prediction. Therefore robustness should also be considered when choosing the appropriate model for a given situation. More importantly, the actual prediction error generally cannot be determined, since typically the actual analyte concentration in the 'unknown' is unknown. Therefore other strategies must be used to determine the correct model (i.e. the number of factors) to use for a specific circumstance. Three general strategies were explored in this investigation, involving the figures of merit

(1) calibration error, (2) residual spectra and (3) number of PARAFAC iterations.

Calibration error is a common analytical figure of merit, defined in this paper as the root mean square error of calibration according to the equation

$$\mathrm{RMSEC} = \left(\sum (c_i - \hat{c}_i)^2 / M\right)^{1/2} \tag{3}$$

where RMESC is the square root of the sum of squared errors between predicted and actual concentrations of the standard solutions and M is the total number of standards (calibration points) used to create the model. Although this is actually a measure of how well the model predicts the concentration of the standards, not the unknowns, the relationship between the RMSEC and the predictive ability of a model is explored.

Residual matrices, defined here as the differences between actual and predicted EEM matrices, were used to determine how well a PARAFAC model fits the data. A predicted spectrum $\mathbf{A}$ was formed by reconstructing the EEM according to the equation

$$\mathbf{A}_K = \mathbf{X}\mathrm{diag}(\mathbf{Z}_K)\mathbf{Y}^{\mathrm{T}} \tag{4}$$

where $\mathbf{X}$ is the predicted excitation spectrum, diag($\mathbf{Z}_K$) is a matrix of zeros with the predicted intensity profiles for the Kth sample along the diagonal and $\mathbf{Y}^{\mathrm{T}}$ is the transpose of the predicted emission spectrum. Two ways of calculating the 'size' of a residual matrix are to determine its 2-norm or Frobenius norm. The 2-norm (or p-norm) of a matrix $\mathbf{A}$ is defined as the square root of the largest eigenvalue of $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ [29]. The Frobenius norm of a matrix $\mathbf{A}$ is the square root of the sum of the squares of all the entries of the matrix, which can be represented as

$$\|\mathbf{A}\|_{\mathrm{F}} = \left(\sum_{i=1}^{I}\sum_{j=1}^{J} a_{ij}^2\right)^{1/2} \tag{5}$$

where $\|\mathbf{A}\|_{\mathrm{F}}$ is the Frobenius norm of the $I \times J$ matrix $\mathbf{A}$, and a_i and a_j are respectively the rows and columns of $\mathbf{A}$. Therefore the 2-norm has a magnitude and directional character because it is associated with the largest eigenvalue–eigenvector pair; on the contrary, the Frobenius norm is purely a measure of magnitude because it is not constrained to measure variance in any particular direction. In a series of PARAFAC models ranging from 1 to N factors, as n increases and the model fit improves, both the 2-norm and the Frobenius norm of the residuals will decrease, but at different rates. When the residual matrix is purely random (i.e. is only noise with no directional variation), the 2-norm becomes a measure of magnitude only and the 2-norm and Frobenius norm should decrease at the same rate as the residuals decrease. At this point the model is modeling noise or 'overfitting' the data. This can be used as an indication that too many factors have been used in the PARAFAC model.

PARAFAC is an iterative algorithm. In this paper, 'iterations' is defined as the number of times PARAFAC cycles through the least squares algorithm before it converges. The authors have often observed that, when too few factors are used in a PARAFAC model, the algorithm converges after relatively few iterations. Similarly, when too many factors are used, the PARAFAC model often fails to converge before a preset number of iterations (2000) signals its termination. Therefore it was anticipated that the relative numbers of iterations of different models could be used to help determine the appropriateness of a particular model.

4. EXPERIMENTAL

The EEM fluorescence spectra of 10 different PAHs were examined. The three PAHs with the most spectral overlap were used in this study. Stock solutions of 6.2 mg l^{-1} 1,2-benzanthracene, 7.4 mg l^{-1} benzo(*a*)pyrene and 5.6 mg l^{-1} benz(*e*)acephenanthylene (Aldrich, St. Louis, Mo.) in absolute ethanol were prepared by heating and stirring in 1 l pyrex volumetric flasks. Serial dilutions with deionized water were performed to acquire eight aqueous calibration standards containing between 0 and 1 ppb 1,2-benzanthracene, benzo(*a*)pyrene and benz(*e*)acephenanthylene. Twelve aqueous mixtures containing two or three of the PAHs, each < 1 ppb, were prepared to represent analytes of unknown concentrations in the presence of unknown, uncalibrated interferents. These mixtures will be referred to as 'unknowns'. All solutions were prepared in 100 ml pyrex volumetric flasks using 1, 3, 4 and 5 ml pyrex transfer pipettes on the morning of analysis.

The single-measurement EEM fluorometer used was modeled after that of Muroski *et al.* [13] and has been described previously [30,31]. Small, inexpensive equipment has been chosen to facilitate field use and battery operation. It consists of a 75 W Xe lamp (14 V, 5 A), two 0.4 nm resolution SpectraPro imaging spectrographs with focal lengths of 150 mm and f/4 optics (Acton Research Corporation, Acton, Mass.) separated by a sample chamber (Acton Research model SC-447), and a low-resolution imaging charge-coupled device (CCD) (ST-6B Santa Barbara Instrument Group, Santa Barbara, Calif.) with a 750×242 pixel array, an 11 μm × 27 μm pixel size and a total area of 8.6 mm × 6.5 mm. The exit slit of the first spectrograph is axially rotated to produce an incident spectrum such that the excitation photons are spatially dispersed across the sample. This resolution of excitation wavelengths is maintained when light is collected at 45° from incident and focused onto the slit of the second spectrograph. From the second spectrograph the light is dispersed onto the CCD. EEM data are collected on a standard laptop or desktop using KestrelSpec software from Catalina Scientific (Tucson, AZ, USA).

The 150 grooves/mm excitation grating (blazed at 300 nm) used in the excitation spectrograph and the 300 grooves/mm emission grating (blazed at 500 nm) used in the emission spectrograph generated 60 nm × 160 nm excitation–emission matrices centered at and 285 and 470 nm respectively. The CCD pixels were binned 3×8, resulting in 250×30 pixels of 33 μm × 216 μm each. EEMs of the PAH standards and mixtures were randomly collect using 1–2 min exposure times. To correct for bias and thermal noise, a dark exposure was acquired immediately prior to each spectrum and subtracted from the raw image.

The KestrelSpec image files were converted into text files and imported into the MATLAB work environment on Pentium 200, 600 and 933 MHz computers, depending upon availability. The size of each 250×30 pixel EEM was reduced to 200×30 pixels by removing the spectral region

containing no germane information. The entire data set consisted of 98 200×30 EEM spectra: duplicate spectra of eight 1,2-benzanthracene and eight benzo(*a*)pyrene standards, single spectra of seven benz(*e*)acephenanthylene standards, triplicate spectra of 12 PAH mixtures, and 23 deionized water spectra acquired periodically as blanks.

The PARAFAC model used has been described earlier [32]. Because the issue of model selection is under investigation, PARAFAC models were constructed in 'sets'. Each model set included a one-, two-, three-, four- and five-factor PARAFAC calibration model of the same data cube. Sets of PARAFAC models were constructed from data cubes consisting of the entire data set, various subsets of the data, and different densities of selected data subsets. For example, model sets were constructed from data cubes containing from one to eight standard EEMs and from zero to 12 mixture EEMs. To calculate best case scenario detection limits, sets of PARAFAC models were constructed from a data cube of five < 0.5 ppb standard solutions and 10 blanks. Sets of PARAFAC models were constructed to investigate the effects of spectral resolution. To represent different spectral resolutions, data cubes were compressed by averaging excitation and/or emission spectra in the $I \times J$ excitation–emission matrices. This resulted in data cubes representing spectral resolutions of 2–60 and 1.28–64 nm/pixel in the excitation and emission ways respectively. To investigate PARAFAC model stability, several model sets were repeatedly constructed (20 times) from the same data cubes.

For each multifactor model the factor representing the analyte of interest was chosen based on the largest correlation between the predicted and actual relative concentrations of standards of the standard solutions. For each model in a set the following were determined: the root mean square error or calibration (RMSEC), the root mean square error of prediction (RMSEP), the 2-norms of the residual matrices, the Frobenius norms of the residual matrices, the 2-norm/Frobenius norm ratio of the residuals corresponding to the standards and the mixtures, the number of iterations undergone before model convergence, and the number of floating point operations executed during model construction. A residual matrix for a model was obtained by subtracting the average of the difference matrices between each predicted and actual EEM spectrum included in the model construction.

5. RESULTS

5.1. Data density

The four-factor PARAFAC model was deemed to be the best model for the $55 \times 200 \times 30$ point data cube (nine stock solutions, 12 mixtures, two replicated of each, 13 blanks) employed for 1,2-benzanthracene calibration on prediction based on visual inspection of the resolved excitation and emission profiles (Figure 3), relationship between resolved concentration profile and true concentrations for standards and unknowns (Figure 4), and prediction error (Table I). Three of the four factors can be attributed to (A) 1,2-benzanthracene, (B) benzo(*a*)pyrene and (C) benz(*e*)acephenanthylene, which can be determined from the X block (resolved excitation spectra) and Y block (resolved emission spectra) loadings (Figure 3). Although there is no physical interpretation of the fourth factor D in terms of a fluorescence spectrum, the four-factor model 'correctly' deconvolutes the individual fluorescence spectra of the three known PAHs more adequately than the three-factor model. This fourth 'background' factor in the excitation direction seems to be an artifact of the lack of vertical flat field of the CCD detector. Whatever its source, the 'background factor' is not linearly additive and thus theoretically cannot be modeled as a true PARAFAC factor. However, in this and other cases, its inclusion improves both model fit and prediction (Table I).

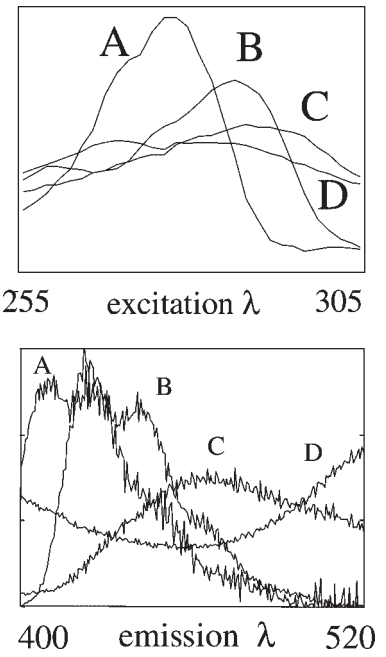

Figure 3. Resolved excitation and emission profiles of (A) 1,2-benzanthracene, (B) benzo(*a*)pyrene, (C) benze(*e*)acephenanthylene and (D) the 'background' using a four-factor PARAFAC model.

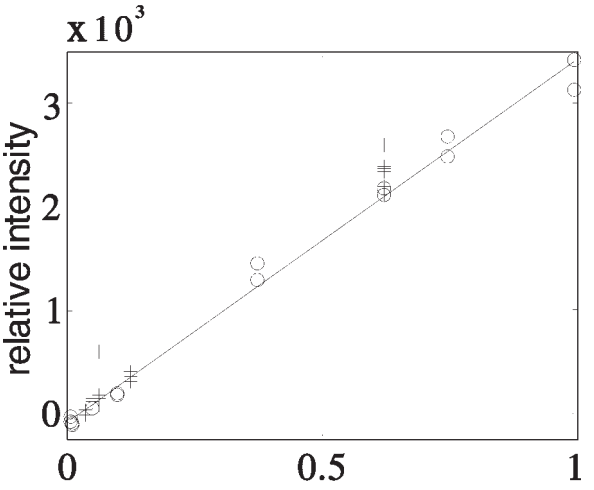

Figure 4. Resolved relative concentrations of 1,2-benzanthracene standards (circles) and mixtures (crosses).

The Z block loadings (fluorescence intensity), shown as circles in Figure 4, corresponding to the relative concentrations of the 1,2-benzanthracene standards were used to predict the 1,2-benzanthracene concentrations in the 12

Table I. Figures of merit for the simultaneous analysis of three PAHs

Sensitivity[a]	Model[b]	RMSEP[c]	RMSEC[d]	LOD[e]	LOD[f]
1,2-Benzanthracene					
1.8×10^3	1	802	50	10	141
	2	$*551^{95\%}$	43	10	74
	3	$*101^{95\%}$	51	86	102
	4	$***61^{95\%}$	**56**	**10**	**68**
	5	374	56	—	68
Benzo(*a*)pyrene					
7.4×10^3	1	696	44	25	68
	2	227	49	5	109
	3	$***160^{90\%}$	**23**	**5**	**31**
	4	$***103^{80\%}$	**21**	**5**	**10**
	5	103	21	5	10
Benz(*e*)acephenanthylene					
2.6×10^2	1	838	83	46	68
	2	$***133^{95\%}$	**47**	**14**	**109**
	3	140	51	10	31
	4	125	27	36	10
	5	125	27	36	10

[a]Change in integrated spectral intensity with concentration (counts/pptr).
[b]Number of factors used in PARAFAC model.
[c]Root mean square error of prediction (pptr): see Equation (1).
[d]Root mean square error of calibration (pptr): see Equation (2).
[e]Limit of detection: 3 × standard deviation of predicted concentration of (five) blanks (pptr).
[f]Limit of detection: 3 × standard deviation of predicted concentration of (10) blanks (pptr).
*Signifies a statistically significant decrease in prediction error from the previous model as determined by a paired t test: $t = x/(s/\sqrt{n})$, where x is the mean of $c_i - \hat{c}_i$, s is the standard deviation of $c_i - \hat{c}_i$, and n is the sample size.
%Confidence level of t test.

mixtures, which are shown as crosses. The root mean square error of prediction and the limit of detection using a four-factor model are 60 and 55 pptr respectively. Average 'best case' scenario figures of merit for simultaneous analysis of all five standards of each PAH are shown in Table I. Note that the 'best model' depends upon the figure of merit of interest. However, two generalizations can be made from models generated with this data set: when too few factors are used, (1) detection limits are often misleadingly low compared with the RMSEP owing to unresolved interfering species in the unknown samples and (2) calibration error is often minimized. Because these two affects are not true physical representations of the system, it was found that residuals were more effective than either detection limits (Table I) or calibration errors for predicting the number of factors to use in a model (see Figure 8).

5.2. Calibration set size

Decreasing the number of standards included in the PARAFAC model is one way to decrease the data density. Figure 5 illustrates the relationship found between the number of standards and the number of floating point operations undergone during the construction of the PARAFAC model, as well as the relationship between the number of inherent factors in the data and the number of floating point operations undergone to generate the PARAFAC model. (The number of floating point operations carried out by the computer during PARAFAC model construction was assumed to be directly related to computation time used as a measure of computational toil.) The line with data points indicated by circles in Figure 5 summarizes PARAFAC models constructed using from two to seven 1,2-benzanthracene standards and one unknown containing 0.620, 0.740 and 0.560 ppb 1,2-benzanthracene, benzo(*a*)pyrene and benz(*e*)acephenanthylene respectively. This three-component data set was best modeled using a three-factor PARAFAC model. By examination of the deconvoluted spectra it was seen that the three factors correspond to the three PAHs. The line with data points indicated by squares summarizes PARAFAC models constructed using from two to seven 1,2-benzanthracene standards and one unknown containing 0.6200 and 0.5600 ppb 1,2-benzanthracene and benz(*e*)acephenanthylene respectively. As expected, this two-component data set was best modeled using a two-factor PARAFAC model. By examination of the deconvoluted spectra it was seen that the two factors correspond to 1,2-benzanthracene and benz(*e*)acephenanthylene.

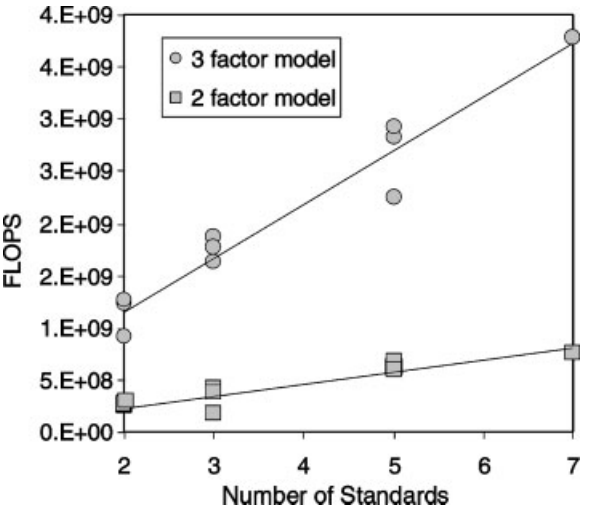

Figure 5. Number of MATLAB floating point operations (FLOPs) undergone before PARAFAC model convergence versus the number of standards included in the data cube.

It can be seen from Figure 5 that doubling the total number of spectra (standards and unknowns) approximately doubles the number of floating point operations (FLOPs) (on average, each additional standard spectrum included in the PARAFAC model added 1×10^8 floating point operations to the model construction). Doubling the number of FLOPs doubles the time it takes to complete the model, assuming that a reasonable model is chosen, i.e. the model is not greatly over- or undersimplified by choosing excessively too few or too many factors. Conversely, cutting the data in half decreases the number of FLOPs. However, decreasing the number of standards included in a model only slightly degrades its predictive capability, as shown in Figure 6, where the average RMSEP is plotted versus the number of calibration points.

A typical relationship between the number of calibration points used and the resulting prediction error of the PARAFAC model can be seen in Figure 6. Each value in Figure 6 represent the mean RMSEP of three PARAFAC models, each containing a different set of standards and one of the 12 unknowns, for each of the three PAHs, at seven different

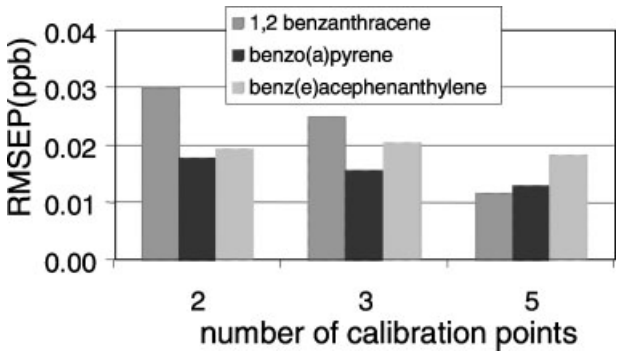

Figure 6. Average root mean square error of prediction versus the number of calibration points used in the data cube for three PAHs.

resolutions, for a total of 63 models constructed from 63 different sets of data. In the case of 1,2-benzathracene, reducing the number of calibration points from five to two increases the prediction error from about 0.01 to 0.03 ppb. The same reduction in calibration points increases the benzo(*a*)pyrene prediction error from about 0.01 to 0.02 ppb but has relatively little effect on the benz(*e*)acephenanthylene prediction error. Although reducing the number of calibration standards in the PARAFAC model may increase the prediction error (10–20 pptr in this case), for many field or on-line applications this may be a fair trade-off for the gain of the corresponding increase in analysis speed. However, this decrease in calibration points significantly decreases the number of floating point operations needed to finish the model, which significantly decreases the run time of the PARAFAC algorithm. Note that the change in prediction error in going from five to two calibration standards is about equal to the limit of detection with seven calibration standards. In addition, the use of fewer calibration points reduces the time and money spent preparing the standards.

5.3. Spectral resolution

One of the biggest obstructions to the PARAFAC algorithm's speed is the size of the individual EEM data matrices. For a given range of spectral coverage the matrix size can be reduced by decreasing the resolution of the spectrofluorometer gratings or increasing the binning on the CCD detector during data acquisition, or by mathematical spectral averaging of the EEM after collection. Because the former involves actually collecting fewer data, it decreases data acquisition time as well as data analysis time. The latter only decreases data analysis time. In this study, only the latter was carried out in an effort to determine the optimal parameters for the former.

Decreasing data density by decreasing spectral resolution in either the excitation or emission direction was found to significantly decrease the number of PARAFAC model FLOPs yet have little degradative effect on predictive ability. Root mean square errors of prediction were calculated for models using the same spectra of eight benzo(*a*)anthracene standards and 12 unknowns at 10 different excitation resolutions between 2 and 60 nm/pixel and a constant emission resolution of 0.64 nm/pixel and at 17 different emission resolutions between 1.28 and 64 nm/pixel and a constant excitation resolution of 2 nm/pixel. A linear relationship was found between the number of FLOPs and the total number of pixels, with a slope of 2×10^7 FLOPs per pixel for a four-factor model and 1×10^6 FLOPs per pixel for a three-factor model, with R^2 values of 0.849 and 0.6792 respectively. However, for typical three- and four-factor models, only a slight increase in prediction error was found as the total number of pixels decreased from 3000 to 60. For example, average prediction error rose by only approximately 0.08 ppb over a range of 3000–60 total pixels/EEM for four-factor PARAFAC models. This represents approximately a 98% increase in analysis speed yet only about an 8% rise in prediction error. The increase in prediction error is trivial compared with the practical quantitation limits (PQLs) of EPA methods 8100 and 8270, which are the advisory test methods for benzo(*a*)pyrene, 1,2-benzanthracene and benz(*e*)acephenanthylene set by the Resource Conservation and Recovery Act (RCRA). The PQL of method 8100, which employs gas chromatography and flame ionization detection for PAH detection, is 200 ppb [30]. The PQL of method 8270 for the analysis of semivolatile organic compounds by GC–MS is 10 ppb [30].

As well as reducing the number of floating point operations needed to construct a PARAFAC model, reducing spectral resolution by using a lesser groove density grating or a larger CCD pixel size decreases the time needed to acquire a spectrum (exposure time). Higher resolution means there are more pixel intensity values for the computer to read. In this case, these values are read as one emission vector for each excitation value/pixel/unit, causing a significant reduction in the read-out time when the excitation resolution is decreased. The relationship between CCD read-out time and total number of pixels was found to be linear, with a slope of 6×10^{-4} s/pixel and an R^2 value of 0.9981. Increasing the surface area of each pixel reduces resolution but results in more photons hitting each pixel. Therefore, in addition to increasing read-out speed, decreasing spectral resolution correlated with either a decrease in the actual exposure time or an improvement in the signal-to-noise ratio of the spectra.

It was found that, when optimized, reducing spectral resolution by means of hardware (increasing CCD pixel size or decreasing the groove density of the grating) can lead to better prediction, primarily owing to boxcar signal averaging of the spectra. Figure 7 demonstrates a typical result obtained by decreasing the density of data used to construct a PARAFAC model by decreasing spectral resolution. Each point is the average of RMSEPs resulting from nine PARAFAC models constructed with different combinations of spectra from a set of eight benzo(*a*)anthracene standards and three unknowns containing benzo(*a*)anthracene and one or two interfering PAHs (1,2-benzanthracene and/or benz(*e*)acephenanthylene). Three models were derived from two standards and one unknown, three others from three standards and one unknown, and three from five standards and one unknown.

It can be seen from Figure 7 that decreasing the excitation spectral resolution by binning the data in the excitation direction and thereby decreasing the number of rows in each EEM matrix actually improves the predictive ability of the PARAFAC model as a result of spectral averaging.

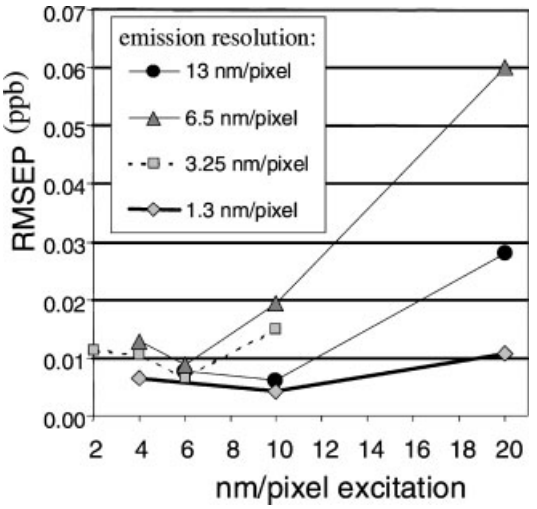

Figure 7. Average root mean square error of prediction versus excitation resolution at four different emission resolutions.

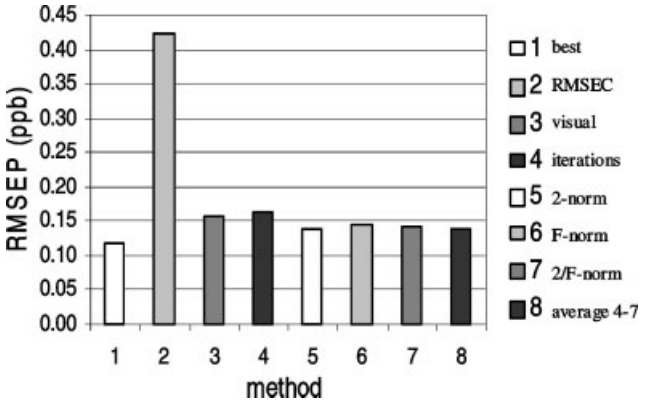

Figure 8. Average root mean square error of predictions using various methods for PARAFAC model selection: 1, the best model; 2, root mean scale error of calibration; 3, visual inspection; 4, iterations until model convergence; 5, 2-norm; 6, Frobenius norm; 7, ratio of 2-norm to Frobenius norm; 8, using the average number of factors predicted by methods 4–7.

However, at some point the benefit of spectral averaging is outweighed by the loss of spectral resolution and, as a result, the predictive ability worsens. In Figure 7 this occurs within a spectral resolution range of 6–10 nm/pixel in the excitation wavelength dimension while keeping the emission resolution constant. For example, at a constant emission resolution of 6.5 nm/pixel, as the excitation resolution decreases from 4 to 6 nm/pixel, the RMSEP improves from 0.013 to 0.009 ppb. As the resolution is further reduced to 10 and 20 nm/pixel, the RMSEP increases to 0.019 and 0.060 ppb respectively. Similarly, as the excitation spectral resolution increases from 2 to 6 nm/pixel at a constant emission resolution of 3.25 nm/pixel, the prediction error decreases from 0.011 to 0.007 ppb. However, when the excitation resolution further decreases to a resolution of 10 nm/pixel, the RMSEP increases to 0.015 ppb, which is higher than at any of the other three resolutions. Similar trends exist when the excitation resolution is held constant and the emission resolution is varied. The same trend was observed for analyses of other PAHs with different calibration ranges.

5.4. Automation of model selection

Traditionally, the number of factors to use in a PARAFAC model must be chosen by the analyst. In the case of EEM, these 'factors' refer to resolved excitation and emission profiles. The number of factors to include in a PARAFAC model can be chosen by visually inspecting the resolved profiles from models created using different numbers of factors. Often, in simple systems, if one too many factors are used, two factors are almost identical, one factor appears to model random noise, or one factor is the 'opposite' or negative of another factor. These effects can be seen upon visual inspection by the expert analyst and often can be used to select the 'correct' model in a simple system. For our purposes the 'correct' model is stable and provides low prediction errors. By stable it is meant that, when the same number of factors is used on the same data set, the same model results, regardless of the initialization vector. In the simple system used, this usually was the one found to be the 'best' by the expert (the authors). In this investigation it was found that several simple methods can be implemented for rapid, automated estimation of the number of factors inherent in a data set.

Figure 8 shows the average root mean square errors of 1,2-benzanthracene prediction (from 52 data sets at full resolution, consisting of models built with two to seven standards and one unknown mixture) resulting from different methods of model selection. The x axis represents the methods used for model selection in the following order: root mean square error of calibration (RMSEC); visual inspection of the resolved factors by the analyst (visual); number of iterations until model convergence (iterations); the 2-norm of the residual matrix (2-norm); the Frobenius norm of the residual matrix (F-norm), the ratio of the 2-norm to the Frobenius norm (2/F-norm); and using the average number of factors predicted by the iteration method, the 2-norm and the Frobenius norm. The category 'best' in Figure 8 refers to the ensemble of the lowest RMSEP out of all the models, even if it was a complete fluke that a model resulted in the lowest prediction error. Since, in true unknowns, actual concentrations are not known, the model with the lowest prediction error cannot actually be determined, so RMSEP obviously cannot be used to determine the best model in a real scenario.

These results suggest that one or several of these methods (methods 4–7 in Figure 8) could be implemented into the PARAFAC algorithm in order to automate the model selection process and thus require no user input or expertise in the field of chemometrics. However, applications of the automated model selection protocols are needed to verify the generalizability beyond EEM fluorescence spectroscopy. For the application presented here, it can be seen that using calibration error is not a good indicator of prediction ability. This is understandable, because the goal here is to obtain the best prediction of unknown samples, not the standards. The average RMSEP found for the model with the least error in calibration is 0.41 ppb, which is approximately three times as large as the prediction error associated with choosing the model using any of the other four criteria. Choosing a model using visual inspection, the number of iterations, the 2-norm, the Frobenius norm and the ratio of the 2-norm to the

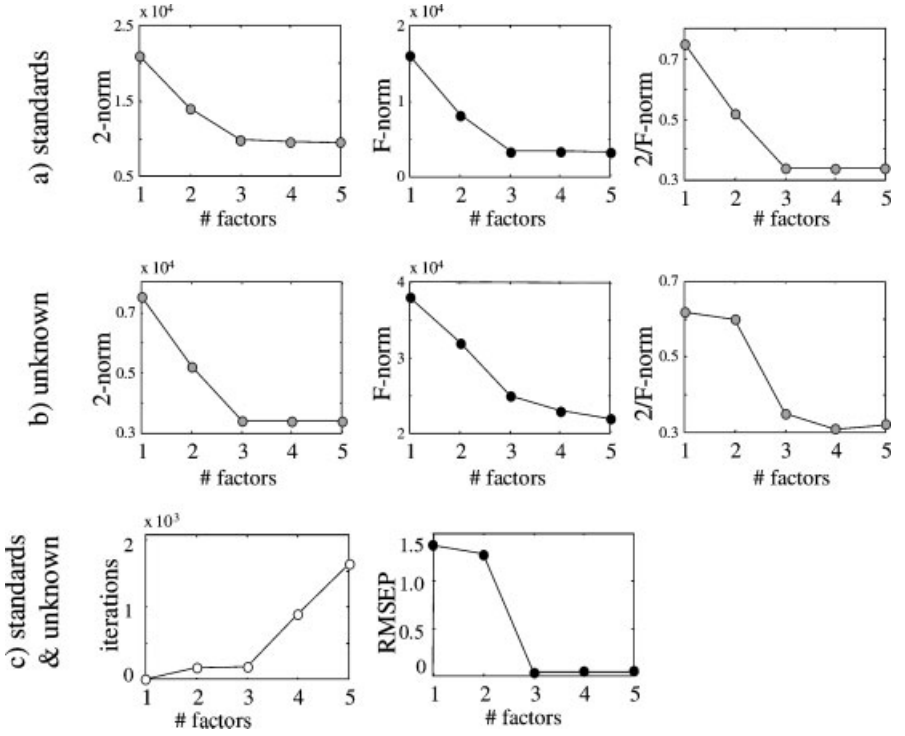

Figure 9. Average 2-norm, Frobenius norm and 2-norm/Frobenius norm ratio of the residuals for one-, two-, three-, four- and five-factor PARAFAC models for 1,2-benzanthracene calibration where a three-factor model is the 'best' choice.

Frobenius norm of the residuals resulted in approximately the same prediction error of about 0.15 ppb. Similarly, choosing a model by talking the average number of factors predicted by the iteration method, the 2-norm, the Frobenius norm and the ratio of norms resulted in a prediction error of 0.15 ppb.

5.5. Stability

Figure 9 illustrates the successful model prediction by the various selection methods. The average 2-norm, Frobenius norm and 2-norm/Frobenius norm ratio of the residuals calculated for the 1,2-benzanthracene calibration standards and the mixture for the one-, two-, three-, four- and five-factor models are shown in Figure 9. From one to two factors and from two to three factors there is a considerable decrease in the residuals. However, adding a fourth or fifth factor to the PARAFAC model has little effect on the residuals. Although these 'extra' factors do not hinder prediction, including them in the data set can significantly increase analysis time.

PARAFAC models were reconstructed several times to determine the robustness of the models and if the resulting predictions made using the various automation methods were adequately robust. For example, using different random starting vectors, 20 sets of PARAFAC models were constructed from the same data cube to predict 1,2-benzanthracene in a PAH mixture containing 0.62 ppb 1,2-benzanthracene, 0.74 ppb benzo(*a*)pyrene and 0.56 ppb benz(*e*)acephenanthylene. The average root mean square error of prediction for the 20 models was 0.037 ppb, with a standard deviation of 0.0002 ppb or 0.03%. This reproducibility demonstrates that it is unlikely that the model will get trapped in a practically different local minimum, regardless of the random starting vector choice.

6. CONCLUSIONS

Decreasing the density of EEM data in any one of the three 'ways' can significantly speed up PARAFAC model construction without significant loss of prediction accuracy. When density is optimized, prediction accuracy can remain adequate or even improve. In addition, several simple methods can be implemented into the algorithm to achieve rapid, automated estimation of the inherent number of factors in a data set, thus simplifying PARAFAC use. The combination of data density optimization and model selection automation can make simple, rapid, multiway data analysis possible. Using these methods in conjunction with field-portable, battery-operated, multiway instrumentation makes feasible near-real time, on-site multiway analyses.

The EEM fluorometer and corresponding data analysis methods are currently being optimized for field deployment. The EEM has been adapted to run off DC power and is thus field-portable for on-site analysis. The 75 W Xe lamp requires 14 V and draws 5 A h^{-1}; it can be powered in the field for 15 h using two 12 V, 75 A h^{-1} deep cycle batteries. Concurrent with optimization for field deployment is the development of a fiber optic interface to the fluorometer that will permit true *in situ* analysis of many target analytes [31].

REFERENCES

1. El-Alawi YS, McConkey BJ, Dixon DG, Greenberg BM. Measurement of short- and long-term toxicity of

polycyclic aromatic hydrocarbons using luminescent bacteria. *Ecotoxicol. Environ. Safety* 2002; **51**: 12–21.
2. Saunders CR, Shockley DC, Knuckles ME. Fluoranthene-induced neurobehavioral toxicity in F-344 rats. *Int. J. Toxicol.* 2003; **22**: 263–276.
3. Futoma DJ, *et al. Poly Aromatic Hydrocarbons in Water Systems*. CRC Press: Boca Raton, FL, 1981.
4. Wittkamp BL, Hawthorne SB, Tilotta DC. Determination of aromatic compounds in water by solid phase microextraction and ultraviolet absorption spectroscopy. 1. Methodology. *Anal. Chem.* 1997; **69**: 1197–1203.
5. Algarra M, *et al.* Direct fluorometric analysis of PAHs in water and in urine following liquid solid extraction. *J. Fluoresc.* 2000; **10**: 355–359.
6. Botello AV, *et al.* Presence of PAHs in coastal environments of the south-east Gulf of Mexico. In *Proceedings of the Thirteenth International Symposium on Polynuclear Aromatic Hydrocarbons*. Gordon and Breach: Bordeaux, 1991.
7. Hertz HS, *et al. Anal. Chem.* 1978; **50**: 429A.
8. Schulman SG. Fluorescence spectroscopy: where we are and where we're going. In *Fluorescence Spectroscopy: New Methods and Applications*, Wolfbies OS (ed.). Springer: New York, 1993; 3–12.
9. Reeve Analytical Website. 8200 OEM filter fluorimeter. [Online]. Available: http://www.reeveltd.demon.co.uk/ [7 September 2001].
10. Turner Designs Website. TD-4100 oil in water monitor. [Online]. Available: http://www.oilinwatermonitors.com/ [13 February 2002].
11. Johnson DW, Callis JB, Christian GD. Rapid scanning fluorescence spectroscopy. *Anal. Chem.* 1977; **49**: 747A–757A.
12. Warner IM, *et al.* Analysis of multicomponent fluorescence data. *Anal. Chem.* 1977; **49**: 2155–2159.
13. Muroski AR, Booksh KS, Myrick ML. Single-measurement excitation/emission matrix spectrofluorometer for determination of hydrocarbons in ocean water. 1. Instrumentation and background correction. *Anal. Chem.* 1996; **68**: 3534–3538.
14. Bro R. PARAFAC. Tutorial and applications. *Chemometrics Intell. Lab. Syst.* 1997; **38**: 149–171.
15. Smilde AK. Three-way analyses—problems and prospects. *Chemometrics Intell. Lab. Syst.* 1992; **15**: 143–157.
16. Kruskal JB. Rank, decomposition, and uniqueness for 3-way and *N*-way arrays. In *Multiway Data Analysis*, Coppi R, Bolasso S (eds). Elsevier Science: Amsterdam, 1989.
17. Da Silva J, Oliveira CJS. PARAFAC decomposition of three-way kinetic–spectrophotometric spectral matrices corresponding to mixtures of heavy metal ions. *Talanta* 1999; **49**: 889–897.
18. Johnson K, De Juan A, Rutan SC. Three-way data analysis of pollutant degradation profiles monitored using liquid chromatography–diode array detection. *J. Chemometrics* 1999; **13**: 331–341.
19. Crouch SR, Coello S, Maspoch S, Porcel M. Evaluation of classical and three-way multivariate calibration procedures in kinetic–spectrophotometric analysis. *Anal. Chim. Acta* 2000; **424**: 115–126.
20. Beltran J, Ferrer R, Guiteras J. Multivariate calibration of polycyclic aromatic hydrocarbon mixtures from excitation–emission fluorescence spectra. *Anal. Chim. Acta* 1998; **373**: 311–319.
21. Wentzell PD, Nair SS, Guy RD. Three-way analysis of fluorescence spectra of polycyclic aromatic hydrocarbons with quenching by nitromethane. *Anal. Chem.* 2001; **73**: 1408–1415.
22. Muroski AR, Booksh KS, Myrick ML. Single measurement excitation/emission matrix spectrofluorometer for determination of hydrocarbons in ocean water. 2. Calibration and quantitation of naphthalene and styrene. *Anal. Chem.* 1996; **68**: 3539–3544.
23. JiJi RD, Andersson GG, Booksh KS. Application of PARAFAC for calibration with excitation–emission matrix fluorescence spectra of three classes of environmental pollutants. *J. Chemometrics* 2000; **14**: 171–185.
24. Alsberg BK, Kvalheim OM. Compression of *N*th-order data arrays by B-splines. 1. Theory. *J. Chemometrics* 1993; **7**: 61–73.
25. Alsberg BK, Nodland E, Kvalheim OM. Compression of *N*th-order data arrays by B-splines. 2. Application to 2nd-order FT-IR spectra. *J. Chemometrics* 1994; **8**: 127–145.
26. Bro R, de Jong S. A fast non-negativity-constrained least squares algorithm. *J. Chemometrics* 1997; **11**: 393–401.
27. Bro R, Andersson CA. Improving the speed of multiway algorithms: Part II. Compression. *Chemometrics Intell. Lab. Syst.* 1998; **42**: 105–113.
28. Andersson CA, Bro R. Improving the speed of multiway algorithms: Part I. Tucker3. *Chemometrics Intell. Lab. Syst.* 1998; **42**: 93–103.
29. Golub GH, Van Loan CF. *Matrix computations* (3rd edn). Johns Hopkins University Press: Baltimore, MD, 1996; 57.
30. JiJi RD, Nahorniak ML, Fruitman E, Booksh KS. Application and calibration of a field portable, excitation–emission matrix fluorometer for analysis of environmental contaminants. *Proc SPIE* 1999, 3856, 73–82.
31. JiJi RD, Cooper GA, Booksh KS. Excitation–emission matrix fluorescence based determination of carbamate pesticides and polycyclic aromatic hydrocarbons. *Anal. Chim. Acta* 1999; **397**: 61–72.
32. JiJi RD, Booksh KS. Mitigation of Rayleigh and Raman spectral interferences in multiway calibration of excitation–emission matrix fluorescence spectra. *Anal. Chem.* 2000; **72**: 718–725.