JOURNAL OF CHEMOMETRICS
J. Chemometrics 2000; **14**: 171–185

Application of PARAFAC for calibration with excitation–emission matrix fluorescence spectra of three classes of environmental pollutants

Renée D. JiJi[1], Greger G. Andersson[2] and Karl S. Booksh[1]*

[1]*Department of Chemistry and Biochemistry, Arizona State University, Tempe, AZ 85287, USA*
[2]*Scotia LipidTeknik AB, PO Box 6686, S-113 84 Stockholm, Sweden*

SUMMARY

Parallel factor analysis (PARAFAC) is applied to three calibrations of a field-portable, cuvette-based, single-measurement, excitation–emission matrix fluorometer. In the first example the fluorometer is calibrated based on interactions between a non-fluorescent DDT-type pesticide and a fluorescent dye. PARAFAC is employed to deconvolve the fluorescence profiles of dissociated and complexed dye states. Calibration is performed based on the intensity of dye–pesticide fluorescence. In the second example, weighted PARAFAC (W-PARAFAC) is applied to determination of three polynuclear aromatic hydrocarbons (PAHs). The weighted algorithm is required to incorporate saturated channels of the CCD detector into the calibration model. In the third example, W-PARAFAC is applied to calibration of two carbamate pesticides. The weighted algorithm is required to account for Rayleigh and Raman scattering overlapping with the fluorescence spectra. For theses three applications, parts-per-trillion to parts-per-billion detection limits are observed in aqueous solutions. Copyright © 2000 John Wiley & Sons, Ltd.

KEY WORDS: parallel factor analysis (PARAFAC); calibration; polynuclear aromatic hydrocarbon (PAH); pesticide; fluorescence

INTRODUCTION

The combination of hyphenated instrumentation and multiway spectral deconvolution methods provides a promising paradigm for quantitative environmental monitoring. Many hyphenated instruments provide bilinear data well suited for multiway methods that assume the parallel factor analysis (PARAFAC) model. The multiway spectral deconvolution methods enable accurate and reliable discrimination of the analyte signal in the presence of unknown and uncalibrated spectral interferents. This capability is often referred to as the 'second-order advantage.' As a practical consequence, the second-order advantage allows for construction of field-portable sensors that rely on statistical discrimination, not complete instrumental separation, of the target analytes.

Two decades ago, Hirschfeld listed 66 hyphenated instrumental methods that currently, or could in

* Correspondence to: K. N. Booksh, Department of Chemistry and Biochemistry, Arizona State University, Tempe, AZ 85287, USA.
Contract/grant sponsor: Camille and Henry Dreyfus Foundation
Contract/grant sponsor: University of Hawaii Sea Grant Program

the future, exist [1]. With hyphenated methods the output of one instrument, say a chromatograph, modulates the input of a second instrument, say an array detector. Consequently, a matrix of data (time × wavelength) is collected per sample analyzed. Examples of hyphenated instrumentation include excitation–emission matrix (EEM) fluorescence spectrometers, gas chromatograph–mass spectrometers (GC–MS) and tandem mass spectrometers (MS). Comparatively, UV–IR spectroscopy would not be a hyphenated method since the UV spectra do not modulate the IR spectra.

Multiway modeling has been applied extensively for quantitative analysis of data from hyphenated instrumentation. Modifications of the basic PARAFAC model to account for expected non-linearities in the instrumental signal have been cataloged [2]. Applications of these multiway calibration techniques include flow injection analysis [3], GC–GC [4], HPLC–spectroscopy [5,6], tandem MS [7] and EEM fluorescence [8,9]. The resolution power of multiway methods is essential in the performance of two environmental sensors utilizing kinetic–spectroscopic analysis [10–12] and chromatographic–spectroscopic analysis [13].

Of particular interest for *in situ* environmental analyses is fluorescence spectroscopy. Fluorescence spectroscopy provides excellent sensitivity for a range of environmentally relevant analytes, including polynuclear aromatic hydrocarbons (PAHs) and many pesticides. PAHs, such as benzanthracenes and chrysenes, are listed in the Clean Water Act (CWA) as section 307 Toxic Pollutants. Nineteen specific PAHs are further listed in the CWA as section 307 Priority Pollutants and recognized as carcinogenic under California Proposition 65. Other than naphthalene, PAHs have little to no commercial value and are not commercially produced. However, PAHs enter the environment from numerous sources: leaching from creosote used to preserve marine lumber; fuel spills; and combustion of fossil fuels, wood and garbage. These PAHs, and specifically the carcinogenic PAHs, are largely fat-soluble and tend to bioconcentrate as they move up the food chain. Consequently, it is prudent to monitor the levels of PAHs in the environment. Pesticides are an integral part of crop management strategies. However, environmental complications of pesticide applications include injury to non-target species, downstream impact of pesticides, and decrease in pesticide efficacy due to overexposure. Often pesticides applied to the soil are not readily degraded; this can result in repeated exposure of the surrounding area to the pesticide long after application. Therefore a method for real-time, on-site analysis of pesticides would be advantageous for environmental management.

One problem hindering wider application of fluorescence spectroscopy for environmental monitoring is the intrinsic lack of selectivity in excitation and emission fluorescence measurements. The broad character of fluorescence peaks usually prohibits finding an explicit excitation/emission wavelength pair that is unique to a given analyte. Employment of multivariate calibration methods, such as partial least squares regression, is unsatisfactory owing to the inability of multivariate calibration methods to provide reliable estimates in the presence of uncalibrated interferents [14]. In theory, multidimensional calibration methods avoid the shortcomings of univariate and multivariate calibration of EEM instruments.

To simultaneously exploit the resolution abilities of multiway calibration and the sensitivity of fluorescence analysis, a single-measurement EEM fluorometer was constructed for field analyses of pesticides and PAHs [15]. The applicability and utility of multiway calibration methods for calibration of and quantification with the EEM fluorometer are demonstrated in three analyses. In the first analysis, methoxychlor, a non-fluorescent DDT-type pesticide, is quantified based on interactions with a fluorescent dye. In the second example, PARAFAC calibration is applied to determination of PAHs. Here W-PARAFAC is employed to extend the dynamic range of calibration beyond the threshold of the detector. In the third example, W-PARAFAC is employed to mitigate the effects of overlapping Rayleigh and Raman scattering for determination of natively fluorescent pesticides.

THEORY

Trilinear model

Parallel factor analysis employs the trilinear model [16]

$$R_{i,j,k} = \sum_{n=1}^{N} \hat{X}_{i,n}\hat{Y}_{j,n}\hat{Z}_{k,n} + E_{i,j,k} \tag{1}$$

where the kth slice of the trilinear cube $\underline{\mathbf{R}}$ is the $I \times J$ matrix of data collected from the instrumental analysis of the kth sample. $\underline{\mathbf{E}}$ is the collection of model and random residual errors from fitting this trilinear model. In this paper, underlined boldface capitals represent three-way arrays, boldface capitals represent two-way arrays or matrices, and scalars are italicized. In the case of EEM fluorometry each matrix contains the excitation and emission profiles of the fluorescent components in each sample. Thus the N columns of the $\hat{\mathbf{X}}$, $\hat{\mathbf{Y}}$ and $\hat{\mathbf{Z}}$ matrices correspond to the estimates of the true excitation profiles, $\mathbf{X}$, emission profiles, $\mathbf{Y}$, and relative concentrations, $\mathbf{Z}$, of the N unique components in the samples. This results in $I \times N$, $J \times N$ and $K \times N$ size matrices for $\hat{\mathbf{X}}$, $\hat{\mathbf{Y}}$ and $\hat{\mathbf{Z}}$ respectively. The analyst supplies the number of factors, N, employed by the model. For these applications the ideal number of factors was chosen by visually inspecting the excitation and emission profiles to determine their reasonability and comparing root mean square errors of calibration for successive models. In addition, knowledge of the system and the number of analytes present in the samples were also considered.

Alternating least squares (ALS) procedure

The N columns of $\hat{\mathbf{X}}$, $\hat{\mathbf{Y}}$ and $\hat{\mathbf{Z}}$ are estimated using an ALS procedure. The following is a three-step summary of the ALS procedure.

Step 1. The PARAFAC algorithm begins with an initial guess of the $\mathbf{X}$ and $\mathbf{Y}$ starting profiles, while the initial $\mathbf{Z}$ profiles are determined by solving the equation

$$\mathbf{R}_{\mathrm{C}} = \mathbf{C}\mathbf{Z}^{\mathrm{T}} \tag{2a}$$

such that

$$\hat{\mathbf{Z}}^{\mathrm{T}} = \mathbf{C}^{+}\mathbf{R}_{\mathrm{C}} \tag{3a}$$

with $\mathbf{C}^{+}$ being the pseudoinverse of $\mathbf{C}$, which can be calculated from the normal equations or singular value decomposition of $\mathbf{C}$. In Equation (2a), $\mathbf{R}_{\mathrm{C}}$ is an $IJ \times K$ matrix constructed by unfolding the K slices of $\underline{\mathbf{R}}$ in the XY plane, where $R_{C_{(j-1)I+i,k}} = R_{i,j,k}$. Similarly, $\mathbf{C}$ is an $IJ \times N$ matrix formed from N columns of $\hat{\mathbf{X}}$ and $\hat{\mathbf{Y}}$, where $C_{(j-1)I+i,n} = \hat{X}_{i,n}\hat{Y}_{j,n}$. Random numbers were used to generate the initial $\mathbf{X}$ and $\mathbf{Y}$ starting profiles.

Step 2. Updated estimates of the $\mathbf{X}$ profiles are found by solving the equation

$$\mathbf{R}_{\mathrm{A}} = \mathbf{A}\mathbf{X}^{\mathrm{T}} \tag{2b}$$

such that

$$\hat{\mathbf{X}}^{\mathrm{T}} = \mathbf{A}^{+}\mathbf{R}_{\mathrm{A}} \tag{3b}$$

$\mathbf{R}_A$ is constructed analogously to $\mathbf{R}_C$ by unfolding the I slices of $\underline{\mathbf{R}}$ in the YZ plane. This forms a $JK \times I$ matrix for $\mathbf{R}_A$, and similarly to **C, A** is a $JK \times N$ matrix, where $A_{(k-1)J+k,n} = \hat{Y}_{j,n}\hat{Z}_{k,n}$.

Step 3. Updated estimates of the **Y** profiles are found by solving the equation

$$\mathbf{R}_B = \mathbf{B}\mathbf{Y}^T \tag{2c}$$

such that

$$\hat{\mathbf{Y}}^T = \mathbf{B}^+\mathbf{R}_B \tag{3c}$$

Like $\mathbf{R}_A$ and $\mathbf{R}_C$, $\mathbf{R}_B$ is constructed by unfolding the J slices of $\underline{\mathbf{R}}$ in the XZ plane. This forms an $IK \times J$ matrix for $\mathbf{R}_B$; **B** is an $IK \times N$ matrix, where $B_{(k-1)I+k,n} = \hat{X}_{i,n}\hat{Z}_{k,n}$.

The algorithm proceeds iteratively, cycling through Steps 1–3 until it converges. At each iteration the most recent estimates of **X** and **Y** are used to determine $\hat{\mathbf{Z}}$ (or **Y** and **Z** to determine $\hat{\mathbf{X}}$, or **X** and **Z** to determine $\hat{\mathbf{Y}}$, depending on the equation currently being solved). Thus the squared residual penalty function

$$\sum_{i,j,k}\left[\left(R_{i,j,k} - \sum_{n=1}^{N}\hat{X}_{i,n}\hat{Y}_{j,n}\hat{Z}_{k,n}\right)^2\right] \tag{4}$$

is minimized by minimizing the following expressions at each iterative cycle:

$$\|\mathbf{R}_A - \mathbf{A}\hat{\mathbf{X}}^T\|_F^2 \tag{5a}$$

$$\|\mathbf{R}_B - \mathbf{B}\hat{\mathbf{Y}}^T\|_F^2 \tag{5b}$$

$$\|\mathbf{R}_C - \mathbf{C}\hat{\mathbf{Z}}^T\|_F^2 \tag{5c}$$

Note that for any matrix **Q** the following equation gives the squared Frobenius or Euclidean norm of **Q**:

$$\|\mathbf{Q}\|_F^2 = \sum_{i=1}^{I}\sum_{j=1}^{J} Q_{i,j}^2 \tag{6}$$

Convergence is achieved when the correlation of the most recent estimates of **X**, **Y** and **Z** and the values of $\hat{\mathbf{X}}$, $\hat{\mathbf{Y}}$ and $\hat{\mathbf{Z}}$ from the previous step is greater than $1-10^{-9}$ [17,18].

W-PARAFAC

In spectral data, some of the measured values deviate significantly from the underlying trilinear model. For example, the non-linear Rayleigh and Raman scattering are inefficiently modeled owing to their lack of inherent factors assumed by the trilinear model; for the PAH analyses some spectral channels have become saturated and effectively become missing data. A weight matrix $W_{i,j,k}$ may be constructed to mitigate the effect of corrupt data points from the model. By assigning questionable data an arbitrarily small weight in the least squares optimization, the degree of fit by the model to these data will have no effect on the final determination of model parameters. This yields an optimization penalty function

$$\sum_{i,j,k}\left[W_{i,j,k}\left(R_{i,j,k}-\sum_{n=1}^{N}\hat{X}_{i,n}\hat{Y}_{j,n}\hat{Z}_{k,n}\right)^2\right] \tag{7}$$

which is minimized by minimizing the following expressions at each appropriate step in the iterative cycle [19,20]:

$$\|\mathbf{W}_A \circ (\mathbf{R}_A - \mathbf{A}\hat{\mathbf{X}}^T)\|_F^2 \tag{8a}$$

$$\|\mathbf{W}_B \circ (\mathbf{R}_B - \mathbf{B}\hat{\mathbf{Y}}^T)\|_F^2 \tag{8b}$$

$$\|\mathbf{W}_C \circ (\mathbf{R}_C - \mathbf{C}\hat{\mathbf{Z}}^T)\|_F^2 \tag{8c}$$

Here $\mathbf{W}_A$, $\mathbf{W}_B$ and $\mathbf{W}_C$ are constructed by unfolding $\underline{\mathbf{W}}$ equivalently to unfolding $\underline{\mathbf{R}}$ to construct $\mathbf{R}_A$, $\mathbf{R}_B$ and $\mathbf{R}_C$, and ° is the Hadamard product indicating element-wise multiplication.

EXPERIMENTAL

EEM fluorometer

The EEM fluorometer was constructed in the manner of Booksh *et al.* [21] and is described in Reference [15]. Light from a 75 W xenon arc lamp (Osram Sylvania Inc., Danvers, MA) was focused on to the 2·75 mm entrance slit of a 150 mm imaging spectrograph (Acton Research Corp., Acton, MA). This excitation spectrograph was fitted with a 600 groove/mm grating blazed at 300 nm. The 2·75 mm exit slit was rotated 90 ° to provide a 60 nm band of excitation light. The excitation spectrograph was fitted to a sample chamber (Acton Research Corp., Acton, MA) containing a cuvette holder. The excitation light was then focused on the near edge of the cuvette, and the emitted fluorescence was collected through the 1·00 mm entrance slit of a second 150 mm imaging spectrograph (Acton Research Corp., Acton, MA) at a 90 ° angle. This emission spectrograph could be fitted with one of two gratings: a 600 groove/mm grating blazed at 300 nm or a 300 groove/mm grating blazed at 500 nm. An astronomy-grade, thermoelectrically cooled SBIG ST6 CCD camera (Santa Barbara Instrument Group, Santa Barbara, CA) was employed for collection of the resulting EEM fluorescence spectra. The camera was computer-controlled through KestrelSpec 3·2 (Catalina Scientific, Tucson, AZ). A 60 s integration time was employed for all samples, and the CCD camera was maintained at $-20{\cdot}1$ °C.

Methoxychlor and Nile Red

Nile Red is a solvatochromic dye [22] which has been shown to form a fluorescent complex with methoxychlor [23]. Nile Red and methoxychlor (95% purity) were obtained from Aldrich and used without further purification. Stock solutions of 140 ppm Nile Red and 400 ppm methoxychlor were prepared in reagent-grade acetone. Three sets of solutions were prepared with identical methoxychlor concentrations and varied Nile Red concentrations of 0, 14 and 28 ppb. In-house distilled water was employed for preparation of the methoxychlor standards, which ranged from 0 to 999 ppb. Solutions to be used as unknowns were prepared using pond water, with concentrations ranging from 0 to 747 ppb. All solutions were allowed to equilibrate for approximately 3 h to ensure consistent results.

The emission spectrograph was fitted with the 300 groove/mm grating blazed at 500 nm, resulting in a 60 nm × 160 nm EEM spectrum. The excitation spectrograph was centered at 515 nm and the emission spectrograph was centered at 650 nm. Four replicate spectra of each solution were acquired. Two-factor PARAFAC and W-PARAFAC models were found to work best for this application.

Polynuclear aromatic hydrocarbons

Aqueous solutions of polynuclear aromatic hydrocarbons with overlapping EEM spectra were prepared for investigating the dynamic range of the detector. Fluorene (98% purity), naphthalene (99% purity) and phenanthrene (98% purity) were obtained from Aldrich and used without further purification. Twenty-four one- to three-component aqueous solutions were prepared from stock 11, 18 and 12 ppm ethanol solutions of fluorene, naphthalene and phenanthrene respectively. The final fluorene, naphthalene and phenanthrene concentration ranges were 0·5–25, 3·5–100 and 2·0–50 ppb respectively.

The emission spectrograph was fitted with the 300 groove/mm grating blazed at 500 nm, resulting in a 60 nm $\times$ 160 nm EEM spectrum. In aqueous solution, fluorene has two excitation maxima at 270 and 290 nm and an emission maximum at 305 nm. Naphthalene has a single excitation maximum at 280 nm and an emission maximum at 345 nm. Phenanthrene has an excitation maximum at 270 nm and an emission maximum at 355 nm. Therefore the excitation spectrograph was centered at 280 nm and the emission spectrograph was centered at 330 nm. Four to six replicate spectra were acquired for single-component solutions, four replicate spectra were acquired for multiple-component solutions, and 24 spectra of pure water were acquired. Three-factor PARAFAC and W-PARAFAC models were found to work best for this application.

Natively fluorescent pesticides

Technical-grade carbaryl (98·4% purity) was obtained from Rhône-Poulenc and carbofuran (98% purity) was obtained from Aldrich. One hundred per cent ethanol and in-house distilled water were used for all dilutions, and all reagents were used without further purification. Stock solutions of 42 ppm carbaryl and carbofuran were prepared in 10% (v/v) ethanol. Six single-component solutions of carbaryl and carbofuran were prepared with concentration ranges of 2·5–80 and 62·5–2000 ppb respectively.

The emission spectrograph was fitted with the 600 groove/mm grating blazed at 300 nm, resulting in a 60 nm $\times$ 80 nm EEM spectrum. The excitation spectrograph was centered at 280 nm and the emission spectrograph was centered at 340 nm. Four replicate spectra of each solution and 16 replicate spectra of pure water were acquired. Three-factor PARAFAC models were found to work best for this application. When a high cut-off value for determining weights was employed, a three-factor W-PARAFAC model was found to work the best. With lower cut-offs, a two-factor W-PARAFAC model provided equivalent performance to the three-factor model. Consequently, the two-factor model was employed.

Data analysis

Analysis and deconvolution of the EEM spectra were performed in the Matlab 5·2 (MathWorks, Natick, MA) working environment. The PARAFAC decomposition and linear regression programs were written in-house. The programs were executed on IBM-compatible computers with Intel Pentium 233 MHz processors. Prior to importing the data into Matlab, spectra were converted to text files by internal converters in KestrelSpec.

For all applications, linear regression was carried out on the Z values of known standards in order to calculate the regression line. Subsequently, the concentrations of the analytes in each sample were estimated. The root mean square error (RMSE) was then calculated using the equation

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{k=1}^{K}(c - \hat{c})^2}{K}} \tag{9}$$

where c represents the true value, $\hat{c}$ represents the predicted value and K represents the total number of samples. Two RMSEs are calculated. The RMSE of calibration (RMSEC) is calculated based on the fit of the model to just the standard samples. This statistic is consequently employed to aid in determining the optimal number of factors in the PARAFAC model. The RMSE of estimation (RMSEE) is calculated from the ability of the PARAFAC model to predict the concentration of non-standard, 'unknown', samples. This statistic is employed to judge the overall efficacy of the calibration procedure. The standard deviation of the blank was calculated using the equation

$$\sigma_{\mathrm{blank}} = \sqrt{\frac{\sum_{k=1}^{K_\mathrm{B}}(\hat{c}_k - \bar{\hat{c}})^2}{K_\mathrm{B} - 1}} \tag{10}$$

where $\hat{c}_k$ is the predicted analyte concentration in the kth blank and $\bar{\hat{c}}$ is the mean of the predicted concentrations in the K_B blanks.

RESULTS AND DISCUSSION

Application to novel sensors (PARAFAC on DDT-type pesticides)

Nile Red is a solvatochromatic laser dye [22] which has proven useful as a hydrophobic probe. Nile Red has been employed to study the hydrophobic surfaces [24] and binding sites of proteins [25] and as a selective intracellular dye for lipids and lipoproteins [26]. In addition, Nile Red has recently been shown to interact with the DDT-type pesticide methoxychlor in aqueous solution [23]. The fluorescence of Nile Red in aqueous solution is severely quenched and red-shifted [24,25] and centered at approximately 665 nm [23,24]. The blue shift of the excitation and emission spectra of Nile Red with increasing solvent polarity [24–26] is also observed in the presence of methoxychlor [23]. However, in the presence of selected proteins and methoxychlor, two distinct fluorescence maxima are observed that may be attributed to the solvated Nile Red and the hydrophobically bound Nile Red [23,25].

The use of an EEM fluorometer to study the interaction of Nile Red with methoxychlor elucidates the presence of two distinct EEM spectra (Figures 1(a) and 1(b)). The EEM spectrum of Nile Red in aqueous solution is shown in Figure 1(a); addition of methoxychlor results in an additional peak, which is shown in Figure 1(b). Three sets of standards and unknowns were prepared with 0, 14 and 28 ppb Nile Red to explore the effect of Nile Red concentration on the fluorescence of the Nile Red–methoxychlor complex. The standards and unknowns were prepared with methoxychlor concentrations of 0, 664, 830 and 996 ppb and 0, 581 and 747 ppb respectively. The unknowns were prepared in pond water to examine the effect of organic material on the fluorescence of Nile Red. No additional fluorescence peaks were observed, and fluorescence of the solvated Nile Red was unaffected.

PARAFAC was employed to resolve the excitation and emission profiles of the solvated Nile Red and the Nile Red–methoxychlor complex (Figures 2(a) and 2(b)). The solvated Nile Red has excitation and emission maxima at approximately 535 and 665 nm respectively. The estimated fluorescence profile of Nile Red is negative at 600 nm. This is an artifact of the model and is probably a result of the Nile Red reabsorbing some of the emitted fluorescence from the Nile Red–methoxychlor complex.

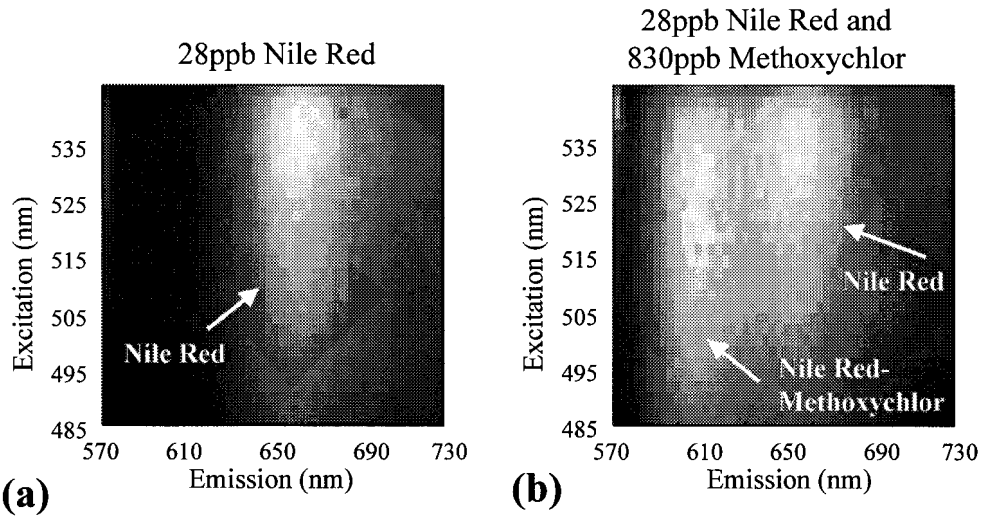

Figure 1. (a) EEM spectrum of 28 ppb Nile Red in aqueous solution. (b) EEM spectrum of 28 ppb Nile Red and 830 ppb methoxychlor in aqueous solution.

Both the excitation and emission profiles of the Nile Red–methoxychlor complex are blue-shifted. The excitation profile of the complex is much broader with respect to the excitation profile of Nile Red, with a maximum at approximately 530 nm. The emission profile of the complex is blue-shifted about 55 nm to 610 nm. As expected, there were no discrete EEM spectra present in the set without Nile Red, confirming that the additional peak may be attributed to the interaction of Nile Red and methoxychlor.

A previous study by Hassoon and Schechter reported a linear increase in the fluorescence intensity of the Nile Red–methoxychlor complex with increasing methoxychlor concentration [23]. The results of this study suggest that the fluorescence of the Nile Red–methoxychlor compound increases linearly within a range of methoxychlor concentrations. However, outside of this range the increase in

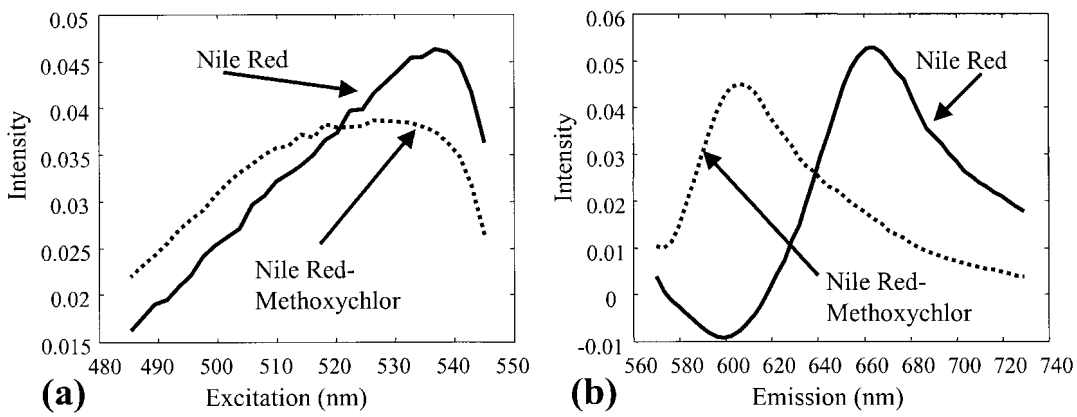

Figure 2. (a) Estimated excitation profiles of Nile Red and Nile Red–methoxychlor complex in aqueous solution. (b) Estimated emission profiles of Nile Red and Nile Red–methoxychlor complex in aqueous solution.

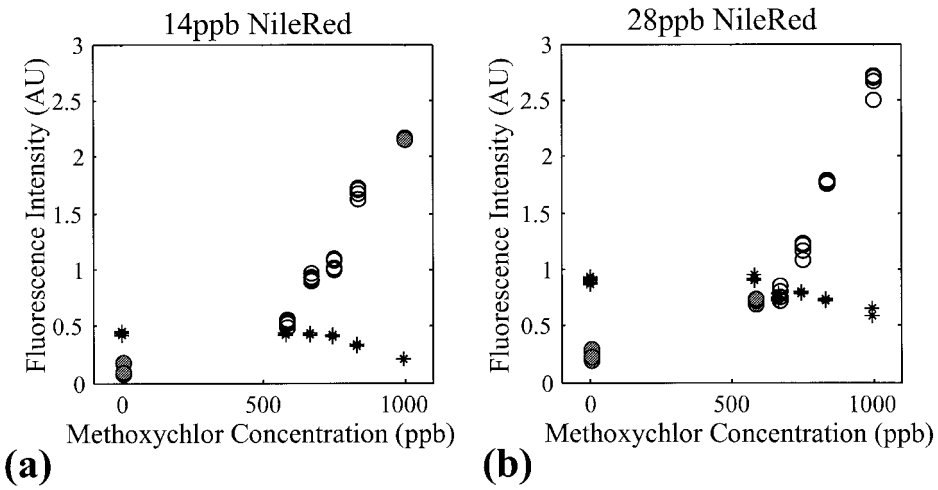

Figure 3. Plots of estimated fluorescence intensity of Nile Red and Nile Red–methoxychlor complex in (a) 14 ppb Nile Red and (b) 28 ppb Nile Red versus methoxychlor concentration. Open circles represent methoxychlor concentrations for which the fluorescence intensity of the Nile Red–methoxychlor complex increases linearly with methoxychlor concentration. Gray-filled circles represent the methoxychlor concentrations for which the fluorescence intensity of the Nile Red–methoxychlor complex increases non-linearly with methoxychlor concentration. Stars represent the fluorescence intensity of the solvated Nile Red.

fluorescence intensity is non-linear. Application of PARAFAC for deconvolution of the solvated and complexed fluorescent profiles does show a decrease in fluorescence intensity for solvated Nile Red and an increase in fluorescence intensity for complexed Nile red with increasing methoxychlor. The open circles in Figures 3(a) and 3(b) represent the range within which the increase in fluorescence intensity of the Nile Red–methoxychlor complex is linear; the gray-filled circles represent the values that do not fall within this range. In addition, this range appears to be dependent on the concentration of Nile Red. The range of methoxychlor concentrations for which there is a linear increase in fluorescence of the formed complex is 581–747 and 664–996 ppb with 14 and 28 ppb Nile Red respectively.

The use of 14 ppb Nile Red resulted in root mean square errors of estimation (RMSEE) of 42 and 65 ppb for the pond water samples with methoxychlor concentrations of 581 and 747 ppb respectively. Increasing the Nile Red concentration to 28 ppb resulted in RMSEE of 71 and 18 ppb for the corresponding samples. The limit of detection for methoxychlor employing this method is dependent on the concentration of Nile Red. At high concentrations of Nile Red and low concentrations of methoxychlor the fluorescence of the Nile Red–methoxychlor complex is either too weak to be quantified or undetectable. Future studies will involve lowering the Nile Red concentrations to sub-ppb concentrations in order to lower the detectable concentration of methoxychlor.

Extending dynamic range (PARAFAC on PAHs)

For natively fluorescent analytes with high quantum efficiency, such as PAHs, detector saturation often becomes a concern. In instances when the relative concentration of a target analyte is low compared to overlapping interferents, simply reducing the integration time to avoid saturation is unsatisfactory; when the interferent signal is within the range of the detector, the analyte signal-to-

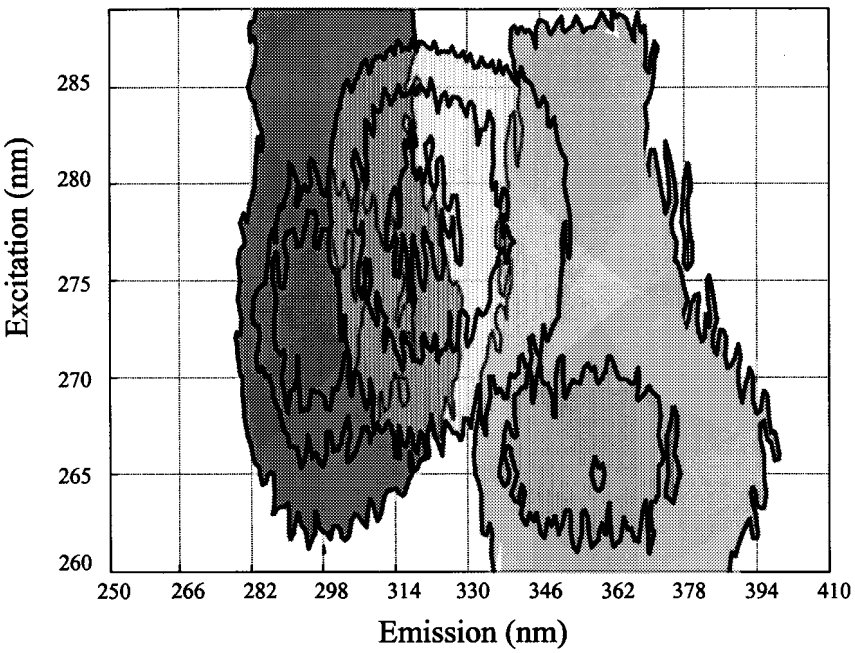

Figure 4. Contour plot of EEM spectra for fluorene (left), naphthalene (center) and phenanthrene (right).

noise ratio may be insufficient for accurate quantification. W-PARAFAC can alleviate this problem by treating the saturated channels as missing values. W-PARAFAC has the advantage over eigenvalue-based deconvolution methods, such as direct trilinear decomposition, in that missing values do not necessitate discarding all variables that share one of the indices. (That is, if the *ijk* measurement is missing from the data set **R**, either the *i*th row, *j*th column or *k*th slice must be removed prior to analysis.) There are two strategies for incorporating missing values in PARAFAC. The first is to assign the missing value a weight of zero and apply a weighted least squares optimization. This is the strategy implemented here. The second strategy is to assign the missing measurement a random value and iteratively refine the assignment with parameter estimates from the model. Both strategies perform equivalently [27] and the relative efficiency of the strategies depends on the size and nature of the data set [28].

In this example the ability of W-PARAFAC to extend the dynamic range of analysis is demonstrated on mixtures of three PAHs. A contour plot presenting the degree of overlap among the three analytes is shown in Figure 4. Naphthalene (center) is much less fluorescent than fluorene (left) and phenanthrene (right); consequently, naphthalene generally demonstrates higher RMSEE. A typical surface plot of a sample's EEM spectra is shown in Figure 5. The spectrum on the left is from a mixture of 70 ppb naphthalene, 5 ppb fluorene and 17 ppb phenanthrene. Note that some channels in the center of the spectrum are saturated and must be treated as missing values. The spectrum on the right is of the same sample after a twofold dilution. In this application the choice of weights is easily determined. Saturation is accentuated by the digitization of the CCD camera voltage feed. At the chosen camera settings the camera provides 2^{16} discrete output levels ranging from 1 to 2^{16}. Consequently, the cut-off level for deciding whether to assign a weight of zero or one is defined as 95% of 2^{16}. This 5% buffer protects against any non-linearities that may occur when the detector is near saturation.

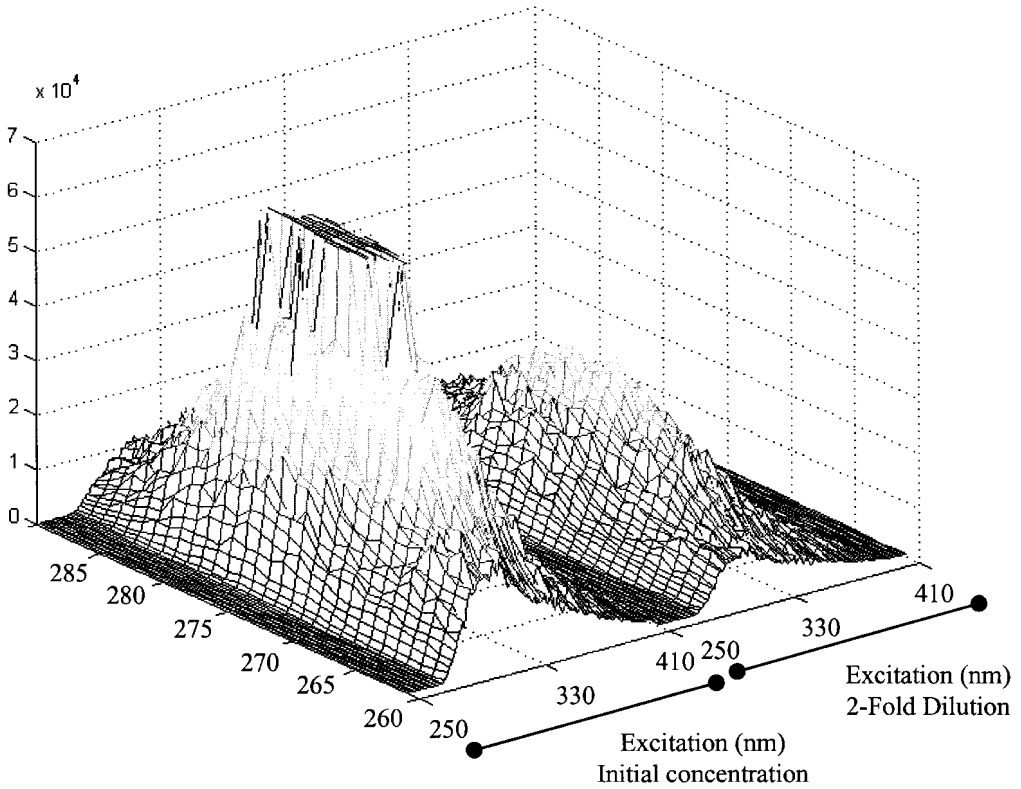

Figure 5. Mesh plot of fluorene/naphthalene/phenanthrene mixture at full strength (left) and half concentration. The plateau on the left peak shows the effects of detector saturation.

There is significant improvement in the RMSEE upon applying W-PARAFAC such that the saturated channels are treated as missing values compared to applying PARAFAC with the unaltered, recorded values. Here only the twelve blanks and replicates of two pure standards (below the saturation level) are employed for calibration (Table I, columns 2 and 3). The RMSEE for naphthalene improves by 35% upon applying W-PARAFAC. The RMSEE for fluorene improves by 15% and the RMSEE of phenanthrene improves by 50%. Note that this improvement in accuracy of

Table I. Comparison of straight and weighted PARAFAC for estimation of three PAHs

PAH	RMSEE[a] (straight)	RMSEE[a] (weighted)	σ_{blank} (straight)	σ_{blank} (weighted)	Improve[b]	Degrade[b]	Z (paired difference)[c]
Naphthalene	7·74	5·00	0·17	0·14	59	13	4·2 (>99·99%)
Fluorene	0·33	0·28	0·021	0·023	27	6	3·0 (99·87%)
Phenanthrene	0·56	0·28	0·013	0·016	52	15	5·0 (>99·99%)

[a]Root mean square error of estimation. Blank (twelve replicates) and two pure standards (four to six replicates) as calibration set.
[b]Improvement or degradation defined as change in absolute error greater than three standard deviations of the blank.
[c]Large-sample (> 30 degrees of freedom) paired difference test comparing square error of estimation with and without weighting. The confidence level is given parenthetically.

estimation does not come at the expense of precision. There is no significant difference in the standard deviations of estimating the blank samples (Table I, columns 4 and 5) for the three hydrocarbons.

The improvement derived from W-PARAFAC is further seen in head-to-head comparisons of the estimated concentration in each sample for each analyte (Table I, columns 6–8). For naphthalene, 59 of the samples experienced a significant improvement in the error of estimation, while only 13 experienced a significant increase in prediction error. Here 'significant' is defined as a magnitude greater than three standard deviations of the blank. Whether two through ten standard deviations are employed, there is no real change in the interpretation of these data. Applying a paired difference test [29] on the squared prediction errors shows that the assumption that straight PARAFAC and W-PARAFAC perform equivalently can be rejected with greater than 99·9% confidence. Comparable conclusions can be drawn from equivalent analysis of the fluorene and phenanthrene prediction errors.

Improving limits of detection (PARAFAC on carbamate pesticides)

The calibration of the two slightly fluorescent carbamate pesticides presented here, carbaryl and carbofuran, is plagued by two problems that necessitate application of W-PARAFAC. First, the fluorescence Stokes shift for each pesticide is not completely resolved from the Raman Stokes and Rayleigh scattering (Figure 6). Second, the Rayleigh scattering is sufficiently intense to saturate the detector in roughly half the samples. Consequently, W-PARAFAC is needed to mask the most intense Rayleigh and Raman scattering to achieve acceptable detection limits. The Rayleigh and Raman profiles are not efficiently modeled by PARAFAC; there is no intrinsic profile in the 'excitation' or 'emission' order that can be extracted. Consequently, when the PARAFAC model is applied to the spectra without accounting for the scattering, the model fails to converge.

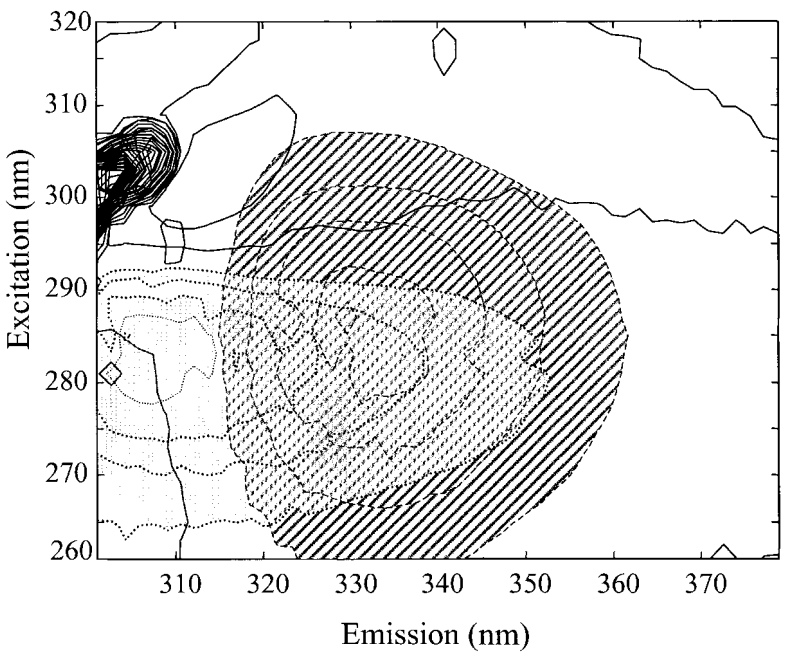

Figure 6. Contour plot of carbaryl (striped) and carbofuran (shaded) EEM spectra. The Rayleigh and Raman background (dense black lines, upper left corner) have been minimized by subtracting the mean of twelve blanks.

Table II. Comparison of PARAFAC preprocessing and weighting strategies for mitigation of Rayleigh and Raman scattering

		Carbaryl		Carbofuran	
Substract mean blank	Weight cut-off	Factors	RMSEC (σ_{blank}) (ppb)	Factors	RMSEC (σ_{blank}) (ppb)
Yes	None	3	6·6 (5·9)	3	168 (149)
Yes	100000	3	6·3 (6·1)	3	157 (152)
Yes	50000	3	2·5 (2·2)	3	51 (41)
Yes	25000	2	2·0 (0·8)	3	42 (24)
Yes	15000	2	1·2 (0·6)	2	29 (15)
Yes	10000	2	3·1 (2·5)	2	31 (14)
Yes	7500	2	1·3 (0·6)	2	31 (14)
No	100000	3	20 (16)	3	292 (239)
No	50000	3	4·5 (4·0)	3	113 (110)
No	25000	2	2·1 (1·5)	2	51 (37)
No	15000	2	3·4 (3·5)	2	65 (68)
No	10000	2	4·4 (3·6)	2	96 (78)
No	7500	2	4·2 (3·3)	2	73 (54)

One possibility for mitigating the effects that the Rayleigh and Raman scattering have on PARAFAC modeling is to subtract a scattering blank from each of the samples prior to analysis. While this may be successful for idealized laboratory analyses, for real-world samples with indeterminate particulates and Raman-active species this approach would be impractical. However, blank-subtracted spectra are analyzed here as a target comparison for W-PARAFAC.

When the mean blank is subtracted from the spectra to minimize the contribution of scattering in the EEM spectra, and a W-PARAFAC model is employed to assign any saturated channels a weight of zero, a moderately acceptable RMSEC and standard deviation of the blank are achieved. Here W-PARAFAC is only considering missing measurements; no effort is made to mitigate the effect of non-saturated scattering. For carbaryl, this is 6·6 ppb and 5·9 ppb respectively. The less fluorescent carbofuran achieves only an RMSEC of 168 ppb and a σ_{blank} of 149 ppb. The relatively poorer figures of merit are also a function of increased overlap with the scattering of carbofuran compared to carbaryl.

The RMSEC and σ_{blank} can be improved by more aggressively down-weighting the entire Rayleigh and Raman scattering prior to PARAFAC analysis. A weight matrix is constructed by analyzing the mean blank to determine the channels that contained the most intense Rayleigh and Raman signals. This weight matrix is constructed such that these channels are given a weight of zero in every sample. It can be seen in Table II, rows 3–9 that the fit and precision of the calibration model initially improve with decreasing the cut-off for the weight matrix. Eventually there is little improvement associated with rejecting measurements, as an increasing amount of fluorescent signal will be excluded along with the scattering background. Also note that as more scattering is rejected, a simpler model (fewer factors) may be employed. Consequently, a cut-off of 15 000 units is near to optimal for blank-subtracted data.

The RMSEC and σ_{blank} are roughly two to three times larger without background subtraction than when background subtraction is successfully applied prior to W-PARAFAC. When only the channels with the most intense background are eliminated (cut-off of 100 000), the RMSEC and σ_{blank} are exceedingly large. However, both figures of merit improve rapidly with small decreases in the cut-off level. As with background subtraction, simpler models are appropriate with more aggressive cut-off levels.

Although the RMSEC and σ_{blank} are higher without than with background subtraction, these results should be put into context. First, in real applications a background will not be available; therefore background subtraction is not an option. Second, the σ_{blank} and RMSEC are comparable to the Resource Conservation and Recovery Act (Superfund) Universal Treatment Standard for waste water (6 ppb) and non-waste water (140 ppb). Third, although the detection limit for carbofuran might be too high for practical utility in aqueous solutions, the detection limit is sufficient following extraction and concentration in a less polar solvent. This permits optical detection as an alternative to chromatographic analysis. Consequently, the use of organic solvents, and associated secondary waste, is minimized.

CONCLUSIONS

The utility of multiway calibration in application of field-portable sensors has been demonstrated. Combining multiway data deconvolution with hyphenated instrumentation can provide a 'green' alternative complete physical separation of analyte and background prior to analysis. With judicious choice of calibration methodology, many potential problems, such as saturation and non-bilinear backgrounds, can be mitigated. However, there are still numerous research areas that may be addressed to improve application of the multiway calibration methods. Examples include automated model selection and validation, on-the-fly determination of statistical confidence limits for the calibration models, and alternative models for incorporation of the non-bilinear sample background.

ACKNOWLEDGEMENTS

The authors gratefully thank the Camille and Henry Dreyfus Foundation and the University of Hawaii Sea Grant Program for financial support. The authors also thank Rhône-Poulenc for the donation of the bulk carbaryl.

REFERENCES

1. Hirschfeld T. The hyphenated methods. *Anal. Chem.* 1980; **52**: 297A.
2. Booksh KS, Kowalski BR. Calibration method choice by comparison of model basis functions to the theoretical instrumental response function. *Anal. Chim. Acta* 1997; **348**: 1–9.
3. Saurina J, Hernandez-Cassou S, Tauler R. Continuous flow titration system for the generation of multivariate spectrophotometric data in the study of acid–base equilibria. *Anal. Chim. Acta* 1995; **312**: 189–198.
4. Bruckner CA, Prazen BJ, Synovec RE. Comprehensive two-dimensional high-speed gas chromatography with chemometric analysis. *Anal. Chem.* 1998; **70**: 2796–2804.
5. Prazen BJ, Synovec RE, Kowalski BR. Standardization of second order chromatographic spectroscopic data for optimal chemical analysis. *Anal. Chem.* 1998; **70**: 218–225.
6. Wu H-L, Shibukawa M, Oguma K. An alternating trilinear decomposition algorithm with application to calibration of HPLC–DAD for simultaneous determination of overlapped chlorinated aromatic hydrocarbons. *J. Chemometrics* 1998; **12**: 1–26.
7. Wang YD, Borgen O, Kowalski BR, Gu M, Turecek F. Advances in second-order calibration. *J. Chemometrics* 1993; **7**: 117–130.
8. Bro R. Exploratory study of sugar production using fluorescence spectroscopy and multiway analysis. *Chemometrics Intell. Lab. Syst.* 1999; **46**: 133–147.
9. Booksh KS, Muroski AR, Myrick ML. A high pressure arc lamp based imaging spectrofluorometer for determination of dissolved hydrocarbons in ocean water: Part 2. Calibration and quantitation of naphthalene and styrene. *Anal. Chem.* 1996; **68**: 3539–3544.
10. Tauler R, Smilde AK, Henshaw JM, Burgess LW, Kowalski BR. Multicomponent determination of chlorinated hydrocarbons using a reaction-based chemical sensor. 2. Chemical speciation using multivariate curve resolution. *Anal. Chem.* 1994; **66**: 3337–3344.
11. Smilde AK, Tauler R, Henshaw JM, Burgess LW, Kowalski BR. Multicomponent determination of

chlorinated hydrocarbons using a reaction-based chemical sensor. 3. Medium-rank second-order calibration with restricted Tucker models. *Anal. Chem.* 1994; **66**: 3345–3351.
12. Booksh KS, Henshaw JM, Burgess LW, Kowalski BR. The second order standard addition method with an application to *in situ* environmental monitors. *J. Chemometrics* 1995; **9**: 263–282.
13. Booksh KS, Lin Z, Wang Z, Kowalski BR. Extension of the trilinear decomposition method with an application to the flow probe sensor. *Anal. Chem.* 1994; **66**: 2561–2569.
14. Booksh KS, Kowalski BR. Theory of analytical chemistry. *Anal. Chem.* 1994; **66**: 782A–791A.
15. JiJi RD, Cooper GC, Booksh KS. Excitation–emission matrix fluorescence based determination of carbamate pesticides and polycyclic aromatic hydrocarbons. *Anal. Chim. Acta* 1999; **397**: 61–72.
16. Bro R. PARAFAC. Tutorial and applications. *Chemometrics Intell. Lab. Syst.* 1997; **38**: 149–172.
17. Mitchell BC, Burdick DS. An empirical comparison of resolution methods for three-way arrays. *Chemometrics Intell. Lab. Syst.* 1993; **20**: 149–161.
18. Mitchell BC, Burdick DS. Slowly converging PARAFAC sequences: swamps and two-factor degeneracies. *J. Chemometrics* 1994; **8**: 155–168.
19. Andersson GG, Dable BK, Booksh KS. Weighted parallel factor analysis of multi-way data for quantitation in the presence of nonlinear instrumental responses. *Chemometrics Intell. Lab. Syst.* 1999; **49**: 195–214.
20. JiJi RD, Booksh KS. Mitigation of Rayleigh and Raman spectral interferents in multi-way calibration of excitation–emission matrix fluorescence spectra. *Anal. Chem.* 2000; **72**: 718–725.
21. Muroski AR, Booksh KS, Myrick ML. A high pressure arc lamp based imaging spectrofluorometer for determination of dissolved hydrocarbons in ocean water: Part 1. Instrumentation and background correction. *Anal. Chem.* 1996; **68**: 3534–3538.
22. Basting D, Ouw D, Schafer FP. The penoxazones: a new class of laser dyes. *Opt. Commun.* 1976; **18**: 260–262.
23. Hassoon S, Schechter I. A sensitive fluorescence probe for DDT-type pesticides. *Anal. Chim. Acta* 1998; **368**: 77–82.
24. Sackett DL, Wolff J. Nile Red as a polarity-sensitive fluorescent probe of hydrophobic protein surfaces. *Anal. Biochem.* 1987; **167**: 228–234.
25. Brown MB, Edmonds TE, Miller JN, Riley DP, Seare NJ. Novel instrumentation and biomedical applications of very near infrared fluorescence. *Analyst* 1993; **118**: 407–410.
26. Greenspin P, Mayer EP, Fowler SD. Nile Red: a selective fluorescent stain for intracellular lipid droplets. *J. Cell Biol.* 1985; **100**: 965–973.
27. Kiers HAL. Weighted least squares fitting using ordinary least squares algorithms. *Psychometrica* 1997; **62**: 2625–2661.
28. Bro R. Multiway analysis in the food industry. *PhD Thesis*, Royal Veterinary and Agricultural University, Denmark, 1988.
29. Vandeginste BGM, Massart DL, Buydens LMC, de Jong S, Lewi PJ, Smeyers-Verbeke J. *Data Handling in Science and Technology 20A. Handbook of Chemometrics and Qualimetrics: Part B.* Elsevier: Amsterdam, 1997.

