Anal. Chem. **2000,** *72,* 718−725

Mitigation of Rayleigh and Raman Spectral Interferences in Multiway Calibration of Excitation−Emission Matrix Fluorescence Spectra

Renée D. JiJi and Karl S. Booksh*

Department of Chemistry and Biochemistry, Arizona State University,Tempe, Arizona 85287

A weighted parallel factor analysis (W-PARAFAC) model is applied to excitation−emission matrix (EEM) fluorescence spectra of carbamate pesticides to aid with calibration in the presence of Raman scattering. Traditional PARAFAC inefficiently models the Raman scattering, resulting in prediction and calibration errors when a significant background is present. Four different weighting strategies were investigated and compared with subtraction of the appropriate sample background. Using a binary weighting strategy produced superior results, compared with a continuous distribution of weights. Further choice of weighting strategies, which are optimized to include either maximum analyte signal or to exclude a maximum amount of background scattering, is dependent on the degree of overlap and relative signal intensity.

Carbamate pesticides are widely used in agriculture to control insects resistant to the less costly organo-phosphorus pesticides.[1] Two commonly used carbamates are carbaryl and carbofuran. Carbaryl, also known as Sevin, is moderately toxic and often used in household gardens. It can be purchased in almost any nursery. Carbofuran, however, is highly toxic and more closely regulated. A ban on all granular formulations became effective in 1994.[2] Both carbaryl and carbofuran are unstable in water and should pose little threat to the environment, once they have degraded to their less-toxic metabolites.[3] However, although carbaryl and carbofuran hydrolyze readily in aqueous solutions, their effect on the environment and surrounding wildlife can be long-lasting.

The long-term effects of carbamate exposure on nontarget species have been investigated. For example, beekeepers have reported pesticide contamination in their bee colonies after application to neighboring areas. Winterlin et al. reported carbaryl residues found in the bees, honey, and bee bread following exposure.[4] Another study showed that concentrations within the freshwater fish Motsugo were as high as 7.5 part per million (ppm) after 1 day of exposure to water with a 1 ppm concentration of carbaryl.[5] Often, carbamates are applied to the soil where they are not as readily degraded. This can result in repeated exposure of the surrounding environment to pesticides long after application. The results of a study on soil-incorporated carbofuran presented the repeated emergence of carbofuran in the runoff water after a heavy rain for several months after application. Therefore, real-time, on-site analysis of pesticides would be advantageous for environmental management.

EPA methods for carbamate detection employ liquid chromatography (LC) (EPA methods 531.1, 632, and 8318) with detection limits in the part per million (ppm) to part per billion (ppb) range when solid-phase extraction is used. Gas chromatography (GC) based analysis can be used to achieve part per billion (ppb) to part per trillion (ppt) detection limits if solid-phase micro-extraction (SPME) is used. However, carbamates are thermally labile and degrade in the column, so a precolumn derivatization must be performed.[6,7] The consumption of carrier solvents for separation and extraction, which can produce large volumes of waste, is a disadvantage of chromatographic methods. Another drawback is that chromatographic instruments are cumbersome and not readily portable, making on-site analysis a laborious task. Finally, analysis is often lengthy, requiring several hours for multiple samples.

Fluorescence spectroscopy is an attractive alternative to chromatographic methods. Many pesticides are either naturally fluorescent[8] or photodegrade to byproducts with a high fluorescence quantum yield.[9,10] Postcolumn reaction or UV photolysis induced fluorescence is often employed in liquid chromatography to increase the sensitivity of the method.[11,12] The same postcolumn fluorophore-inducing reactions employed in liquid chromatography can be employed to generate analytically useful fluorophores in sample extracts for fluorescence analysis.

Despite these benefits, the wide application of fluorescence techniques for environmental monitoring has been limited by the lack of selectivity of fluorescence spectroscopy. The broad character of both the excitation and emission fluorescence bands curtails the possibility of finding a unique excitation and emission

(1) Hassall, K. A. *The Chemistry of Pesticides: Their Metabolism, Mode of Action and Uses in Crop Protection*; Verlag Cheimie: Weinheim, 1982.
(2) Kamrin, M. A. *Pesticide Profiles: Toxicity, Environmental Impact and Fate*; CRC Press: Boca Raton, 1997.
(3) Francis, B. M. *Toxic Substances in the Environment*; John Wiley & Sons: New York, 1994.
(4) Winterlin, W.; Walker, G. *Arch. Environ. Contam. Toxicol.* **1973**, *1*, 362−374.
(5) Kanawanza, J. *Bull. Environ. Contam. Toxicol.* **1975**, *14*, 346−352.
(6) Liska, I.; Slobodnik, J. *J. Chromatogr., A* **1996**, *733*, 235−258.
(7) Cochrane, W. P. *J. Chromatogr. Sci.* **1979**, *17*, 124−137.
(8) Files, L. A.; Winefordner, J. D. *J. Agric. Food Chem.* **1987**, *35*, 471−474.
(9) Coly, A.; Aaron, J. J. *Analyst* (*Cambridge, U.K.*) **1994**, *119*, 1205−1209.
(10) Coly, A.; Aaron, J. J. *Anal. Chim. Acta* **1998**, *360*, 129−141.
(11) Miles, C. J.; Moye, H. A. *Anal. Chem.* **1988**, *60*, 220−226.
(12) Miles, C. J. *J. Chromatogr.* **1992**, *592*, 283−290.

10.1021/ac990418j CCC: $19.00
Published on Web 01/14/2000

wavelength for each potential analyte. Collection of entire excitation−emission matrix (EEM) fluorescence spectra followed by application of advanced multiway spectral deconvolution and calibration algorithms overcomes the limitations of fluorescence spectroscopy. The use of parallel factor analysis (PARAFAC) to resolve overlapping spectra of target polycyclic aromatic hydrocarbons and pesticides has been demonstrated.[13−15] However, resolution becomes difficult when an analyte's fluorescence signal overlaps the background scattering.

A single-measurement EEM fluorometer lacks the baffles and filters typically present in scanning fluorometers; therefore, the intensity of the Rayleigh and Raman scattering is unmitigated. These diagonal patterns across the spectra are inefficiently modeled by trilinear calibration methods.[16] This problem is easily solved if blank spectra are available; unfortunately, this is not normally the case for environmental samples. An alternate solution is to employ a weighted PARAFAC (W-PARAFAC) model. Results using a W-PARAFAC model are compared with results from nonweighted models that use more conventional methods for background reduction.

THEORY

Three-Way Calibration. Parallel factor analysis employs the trilinear model

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

where the kth slice of the trilinear cube $\mathbf{R}$ is the $I \times J$ matrix of data collected from the instrumental analysis of the kth sample. In the case of excitation−emission matrix 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 excitation profiles, emission profiles, and relative concentrations in the samples, respectively. The analyst supplies the number of factors, N, employed by the model. $\mathbf{E}$ is the collection of model and random errors residual from fitting this trilinear model.

The N columns of $\hat{\mathbf{X}}$, $\hat{\mathbf{Y}}$, and $\hat{\mathbf{Z}}$ are estimated using an alternating least squares (ALS) procedure. The PARAFAC algorithm begins with an initial guess of the X-way and Y-way starting profiles, while the initial Z-way profiles are determined by solving

$$\mathbf{R}_C = \mathbf{C}\mathbf{Z}^T \qquad (2a)$$

such that $\hat{\mathbf{Z}}^T = \mathbf{C}^+\mathbf{R}_C$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 eq 2a, $\mathbf{R}_C$ is a $IJ \times K$ matrix constructed by unfolding the K slices of $\mathbf{R}$ in the IJ plane where $R_{C_{(j-1)I+i,k}} = R_{i,j,k}$. Similarly, $\mathbf{C}$ is a $IJ \times N$ matrix formed from N columns of $\hat{\mathbf{X}}$ and $\hat{\mathbf{Y}}$ where $C_{(j-1)I+i,n} = X_{i,n}Y_{j,n}$.

Updated estimates of the X-way and Y-way profiles are found by solving

$$\mathbf{R}_A = \mathbf{A}\mathbf{X}^T \qquad (2b)$$

such that $\hat{\mathbf{X}}^T = \mathbf{A}^+\mathbf{R}_A$, and

$$\mathbf{R}_B = \mathbf{B}\mathbf{Y}^T \qquad (2c)$$

such that $\hat{\mathbf{Y}}^T = \mathbf{B}^+\mathbf{R}_B$. $\mathbf{R}_A$ and $\mathbf{R}_B$ are constructed analogously to $\mathbf{R}_C$ by unfolding $\mathbf{R}$ in the YZ and XZ planes, respectively. This forms a $JK \times I$ matrix for $\mathbf{R}_A$ and a $IK \times J$ matrix for $\mathbf{R}_B$. Similarly to $\mathbf{C}$, $A_{(k-1)J+k,n} = Y_{j,n}Z_{k,n}$ and $B_{(k-1)I+k,n} = X_{i,n}Z_{k,n}$. The algorithm proceeds iteratively, cycling through eqs 2a, 2b, and 2c until the algorithm converges. At each iteration, the most recent estimates of $\mathbf{X}$ and $\mathbf{Y}$ are used to determine $\hat{\mathbf{Z}}$ (or $\mathbf{Y}$ and $\mathbf{Z}$ to determine $\hat{\mathbf{X}}$, or $\mathbf{X}$ and $\mathbf{Z}$ to determine $\hat{\mathbf{Y}}$, depending on the equation currently being solved). Thus, the squared residual penalty function, $\sum_{i,j,k}((R_{i,j,k} - \sum_{n=1}^{N} X_{i,n}Y_{j,n}Z_{k,n})^2)$ is minimized by minimizing $\|\mathbf{R}_A - \mathbf{A}\hat{\mathbf{X}}^T\|_F^2$, $\|\mathbf{R}_B - \mathbf{B}\hat{\mathbf{Y}}^T\|_F^2$, and $\|\mathbf{R}_C - \mathbf{C}\hat{\mathbf{Z}}^T\|_F^2$ at each iterative cycle. Note, that for any matrix $\mathbf{Q}$, $\|\mathbf{Q}\|_F^2 = \sum_{i=1}^{I}\sum_{j=1}^{J} Q_{i,j}^2$ and is the squared Frobenius, or Euclidian, norm of $\mathbf{Q}$. Convergence is achieved when correlation of the most recent estimates of $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$, 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]

In spectral data, there are often errors present that deviate from the underlying trilinear model. For example, the nonlinear Rayleigh and Raman scattering are inefficiently modeled, due to their lack of true excitation and emission profiles. A weight cube, $W_{i,j,k}$, may be constructed to reduce or eliminate this scattering from the model. For this application, each kth layer is identical and independent of sample. However, having each of the k slices of $\mathbf{W}$ being identical is not necessary. The optimization penalty function, $\sum_{i,j,k}(W_{i,j,k}(R_{i,j,k} - \sum_{n=1}^{N} X_{i,n}Y_{j,n}Z_{k,n})^2)$, is minimized by minimizing $\|\mathbf{W}_A \otimes (\mathbf{R}_A - \mathbf{A}\hat{\mathbf{X}}^T)\|_F^2$, $\|\mathbf{W}_B \otimes (\mathbf{R}_B - \mathbf{B}\hat{\mathbf{Y}}^T)\|_F^2$ and $\|\mathbf{W}_C \otimes (\mathbf{R}_C - \mathbf{C}\hat{\mathbf{Z}}^T)\|_F$ at each appropriate step in the iterative cycle, where $\otimes$ is element-wise multiplication. Here, $\mathbf{W}_A$, $\mathbf{W}_B$, and $\mathbf{W}_C$ are constructed by unfolding $\mathbf{W}$ equivalently to unfolding $\mathbf{R}$ to construct $\mathbf{R}_A$, $\mathbf{R}_B$, and $\mathbf{R}_C$.

Weighting Strategies. There are two classes of weighting strategies—positive and negative (Figure 1). Positive weighting concentrates on enhancing each component's signal, while negative weighting concentrates on eliminating any nonlinearities present in the background. These subtle differences may be carried further to two subgroups—hard and soft weighting. With hard weighting, cutoff values are applied to the weight matrix; intensities above and below the cutoff value are assigned a weight of either one or zero, respectively (Figure 1a,b). Soft weighting applies a range of weights, between one and zero, proportional to the intensities of either the analyte or blank signal (Figure 1c,d). Four weighting strategies are thus produced: hard positive, soft positive, hard negative, and soft negative weighting.

Positive weighting applies an increasing weight matrix from zero to one, dependent on each compound's fluorescent spectrum.

(13) Muroski, A. R.; Booksh, K. S.; Myrick, M. L. *Anal. Chem.* **1996**, *68*, 3534−3538.

(14) Booksh, K. S.; Muroski, A. R.; Myrick, M. L. *Anal. Chem.* **1996**, *68*, 3539−3544.

(15) JiJi, R. D.; Cooper, G. A.; Booksh, K. S. *Anal. Chim. Acta* **1999**, *397*, 61−72.

(16) Bro, R. *Chemom. Intell. Lab. Syst.* **1997**, *38*, 149−171.

(17) Mitchell, B. C.; Burdick, D. S. *Chemom. Intell. Lab. Syst.* **1993**, *20*, 149−161.

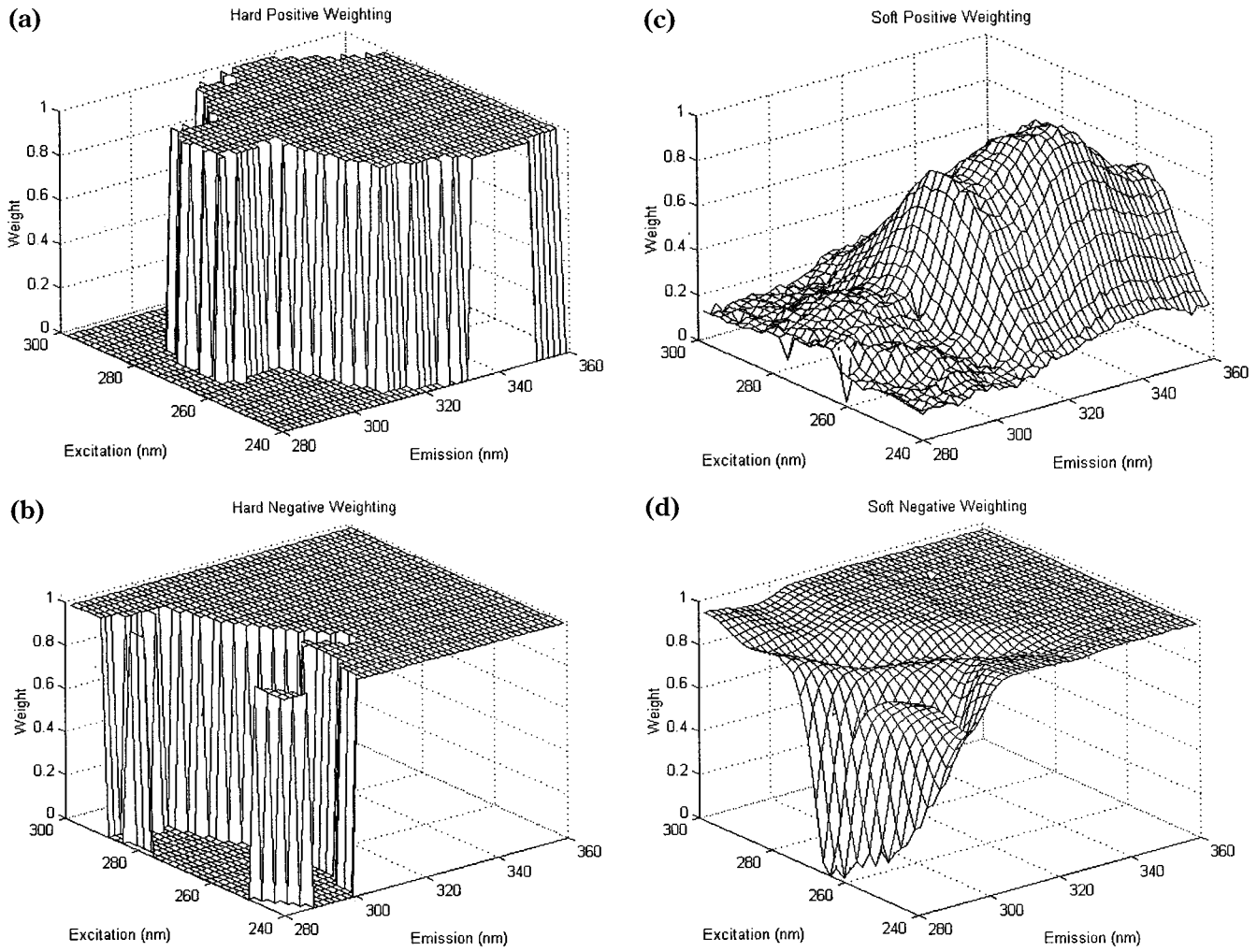

Figure 1. Three-dimensional representation of (a) hard positive weighting, (b) hard negative weighting, (c) soft positive weighting, and (d) soft negative weighting for a mixture of 60 ppb 1-naphthol, carbaryl, and carbofuran.

Positive-weight matrixes were constructed by first subtracting the mean background spectrum from each single-component standard spectrum. An average spectrum for each analyte was created by summing the individual standard spectra and dividing by the number of samples. The average EEM spectrum of carbofuran was then scaled, with the most intense signal areas having a weight of one. This resulted in a single-component weight matrix. A multiple-component weight matrix was created by adding the average EEM spectrum for each analyte together and then scaling to one. The relative fluorescence intensities of each analyte were thus maintained.

Positive weighting is based on each component's fluorescence signal such that analysis may be tailored to compounds of interest while removing or reducing the signal of interferents. For hard-positive weighting, a weight of one is applied to areas with an intensity greater than a baseline noise level; all other areas are given a weight of zero. Several cutoffs were chosen between 5 and 20% of the maximum intensity of the mean standards. This range of cutoffs eliminated any Rayleigh and Raman scattering not directly overlapped with an analyte's signal. On the basis of the ability of the calibration to accurately model the standards, the optimal cutoff was determined. In the case of soft weighting, the fluorescent spectrum is given a weight between zero and one on the basis of relative signal intensity of the pure standard. Each compound's spectrum changes with solvent polarity; consequently, it is necessary to know the type of solvent being used ahead of time.

Contrary to positive weighting, negative weighting is based on removing the sample background spectrum, with the most intense regions of the background assigned a weight of zero. Hard-negative weighting assigns weights of zero to areas with intensity of the blank above a predetermined cutoff value, and all other areas are assigned weights of one. Soft weighting assigns decreasing values from one to zero in order of increasing intensity. Negative weighting is independent of the fluorescent components in the samples, so information concerning the contents of a solution is conserved. The Rayleigh and Raman scattering intensities will vary with solvent type and the quantity of dissolved particles in solution. If a blank is not available, pure solvent may be used, correcting for any intensity differences.

Common problems and benefits exist for both hard and soft weighting. The primary advantage of hard weighting is that assigning certain measurements a weight of zero may eliminate background interference. However, any information in these areas may be lost. Instead of applying strict values of one or zero, soft weighting could be used to avoid any loss of data, although resolution of a compound's signal from the background scattering may be impossible if they are directly overlapped.

EXPERIMENTAL SECTION

EEM Fluorometer. The EEM fluorometer was constructed in the manner of Murowski et al.[13,14] and previously described in JiJi et al.[15] Light from a 75 W Xe arc lamp (Acton Research Corp., Acton, MA) was focused through 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 (gr)/mm ruled grating blazed at 300 nm. The spectrograph was modified such that the 1-mm exit slit is rotated 90°. The excitation spectrograph was attached to a sample chamber (Acton Research Corp., Acton, MA) where the excitation spectrum was focused near the front edge of a cuvette. The emitted fluorescence was collected perpendicular to the excitation radiation and focused onto the 1-mm slit of a second 150-mm imaging spectrograph (Acton Research Corp., Acton, MA). For the methanol samples, the emission imaging spectrograph contained a 600 gr/mm ruled grating blazed at 500 nm, resulting in a 60-nm by 80-nm EEM spectrum. For the aqueous samples, the emission imaging spectrograph contained a 300 gr/mm ruled grating blazed at 500 nm, resulting in a 60-nm by 160-nm EEM spectrum. The resultant spectra were recorded on a thermoelectrically cooled SBIG ST6 CCD camera (Santa Barbara Instrument Group, Santa Barbara, CA). The camera was computer controlled through KestrelSpec 3.2 (Catalina Scientific, Tucson, AZ).

Sample Preparation and Analysis. Technical-grade carbaryl (98.4% purity) was donated by Rhône-Poulenc and 1-naphthol (99% purity) and carbofuran (98% purity) were obtained from Aldrich. For the first experiment, anhydrous reagent-grade methanol (Aldrich) was used for all dilutions and all reagents were used without further purification. Stock solutions of carbaryl, 1-naphthol, and carbofuran were prepared by diluting 1.4, 1.3, and 1.3 mg, respectively, in methanol to 100 mL. Seven standard solutions, from 5 to 100 ppb, were individually prepared for 1-naphthol, carbaryl, and carbofuran. In addition, three mixtures containing two of the analytes at 60 ppb each and one mixture containing all three analytes at 60 ppb each were also prepared.

Four replicate EEM spectra were collected for the 1-naphthol and carbaryl standard solutions. Five replicate spectra were collected for each of the carbofuran standard solutions. Four EEM spectra were collected for each of the four mixtures, and eight spectra of pure methanol, to be used as blanks, were collected. Each replicate spectrum was included as a layer of the three-way array. The excitation and emission spectrographs were centered at 270 and 300 nm, respectively, with a 60-s integration time chosen for the detector. The detector temperature was maintained at −20 °C.

EEM spectra of the standards were collected on the first day of the experiment, and the EEM spectra of the mixtures were collected on the second day of the experiment. On the second day, the instrument sensitivity varied slightly from that on day one, resulting in absolute errors. However, relative compositions could be calculated. Therefore, the root-mean-squared error (RMSE) of percent prediction is presented in Table 2, for the mixtures.

For the second experiment, stock solutions were prepared in methanol to avoid hydrolysis of the analytes. A stock solution of 1-naphthol was prepared by diluting 11.9 mg in methanol to 50 mL. Stock solutions of carbaryl and carbofuran were prepared by diluting 107.9 and 102.6 mg in methanol to 10 mL, then 0.1 mL of each primary stock solution was further diluted in methanol to 100 mL. In-house-distilled water was used for preparation of standard and mixture solutions. Two single-component standards were prepared for each analyte. The standard concentrations were 12 and 119 ppb for 1-naphthol, 270 and 1079 ppb for carbaryl, and 10 and 103 ppm for carbofuran. Six mixtures containing two analytes and three mixtures containing all three analytes were prepared (Table 1). The concentration range of 1-naphthol in the mixtures was from 12 to 95 ppb, 108 to 809 ppb for carbaryl, and 21 to 72 ppm for carbofuran. Four replicate spectra of each sample and eight replicate background spectra were acquired. The excitation and emission spectrographs were centered at 290 and 355 nm, respectively, with a 60-s integration time, and the detector temperature was maintained at −20 °C.

Table 1. Individual Analyte Concentrations for Multicomponent Solutions in Water

mixture	1-naphthol (ppb)	carbaryl (ppb)	carbofuran (ppm)
1	71	540	
2	24	108	
3	36		62
4	83		41
5		809	21
6		216	72
7	12	108	10
8	95	647	31
9	48	384	51

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 analyses the data cube was composed of a 30 row by 50 column EEM spectrum for each sample. All replicate spectra were included in the model. This resulted in 30 × 50 × 115 and 30 × 50 × 68 data cubes for samples in methanol and water, respectively. The correct number of factors was determined by simultaneously minimizing the RMSE of prediction, minimizing the limits of detection, and monitoring the predicted spectral profiles for reasonable peak shapes. The limit of detection was determined as three times the standard deviation of the predicted analyte concentration in the blanks and was calculated by

$$\mathrm{LOD} = 3\sqrt{\left(\sum_{k=1}^{K_\mathrm{B}} (c_\mathrm{k} - \bar{c})^2/(K_\mathrm{B} - 1)\right)} \qquad (3)$$

where c_k is the predicted analyte concentration in the kth blank and $\bar{c}$ is the mean of the predicted concentrations in the K_B blanks. The RMSE was calculated by

$$\mathrm{RMSE} = \sqrt{\left(\sum_{k=1}^{K} (c - \hat{c})^2/K\right)} \qquad (4)$$

where c represents the true value, $\hat{c}$ represents the predicted value, and K represents the total number of samples. For calculation of the RMSE of prediction, c represents the concentration of the analyte. For calculation of the RMSE of percent prediction, c

Table 2. Figures of Merit for Carbofuran Solutions in Methanol

background correction method	subtract mean blank	truncate Rayleigh scattering	soft positive weighting	soft negative weighting
sensitivity[a] (1×10^4) (total counts/ppb)	1.9	1.6	1.9	1.8
LOD[b] (ppb)	3.1	5.0	5.3	4.0
carbofuran[c] (ppb)	3.8	5.6	5.6	4.7

[a] Sensitivity represents the integrated area under the resolved 1 ppb spectra of the analyte. [b] The LOD is determined by three standard deviations from the blank. [c] Root mean-squared error of prediction in the single-component standards.

represents the percentage of each analyte present in a mixture. For example, a solution made up of 60 ppb each of 1-naphthol and carbaryl has fifty percent 1-naphthol and carbaryl.

Spectra of the mixtures in methanol were collected 1 day following collection of the standard spectra. Because of daily variations in environmental conditions, such as temperature, and the instrumental response, the exact concentration of each component could not be predicted from the standard spectra. Therefore, the RMSE of percent prediction was calculated. This is sufficient for the purposes of this study because it allows for a qualitative evaluation of the performance of this instrument and W-PARAFAC. In addition, this problem can be curtailed by collecting all spectra on the same day, minimizing the variations in conditions throughout the experiment.

The sensitivity was determined by taking the cross product of the estimated excitation and emission spectra for each analyte and normalizing the total area under the curve to unit area to determine a scalar, Z. The change in Z with concentration ($\Delta Z/\Delta c$) is thus defined as the sensitivity for each analyte.

RESULTS AND DISCUSSION

Environmental monitoring of aqueous pesticides may be impaired by overlap of the pesticides' fluorescence signals with the Raman scattering. Figure 2a shows the Rayleigh scattering in the top left corner of the image and the Raman scattering just below, overlapping the carbofuran signal. Both types of scattering have distinct diagonal shapes. By applying weights to the EEM spectra during analysis, effects of the Rayleigh and Raman scattering on the model are reduced. Figure 2 parts a and b highlight the effective weight of each channel of the EEM after applying weights to an EEM spectrum of pure carbofuran in methanol. Distinguishing the carbofuran signal from the Raman scattering is difficult due to the intensity of the background scattering. When hard-negative weighting is applied to the same EEM spectra, the carbofuran signal becomes more discernible due to the effective removal of the Rayleigh and Raman scattering.

Multilinear decomposition of both the Rayleigh and Raman scattering is inefficient because they have no intrinsic excitation and emission profiles.[16] Therefore, the scattering must be removed prior to data analysis. Accepted methods for accomplishing this include truncating the Rayleigh scattering, setting the scattering to a nominal baseline value, measuring only the emission wavelengths above the excitation wavelengths, and subtracting a blank spectrum.

However, these methods for dealing with Rayleigh and Raman scattering fall short when a signal overlaps the Raman scattering. Truncating the Rayleigh scattering resulted in nonsensical estimations of the excitation and emission spectra. In addition, any area

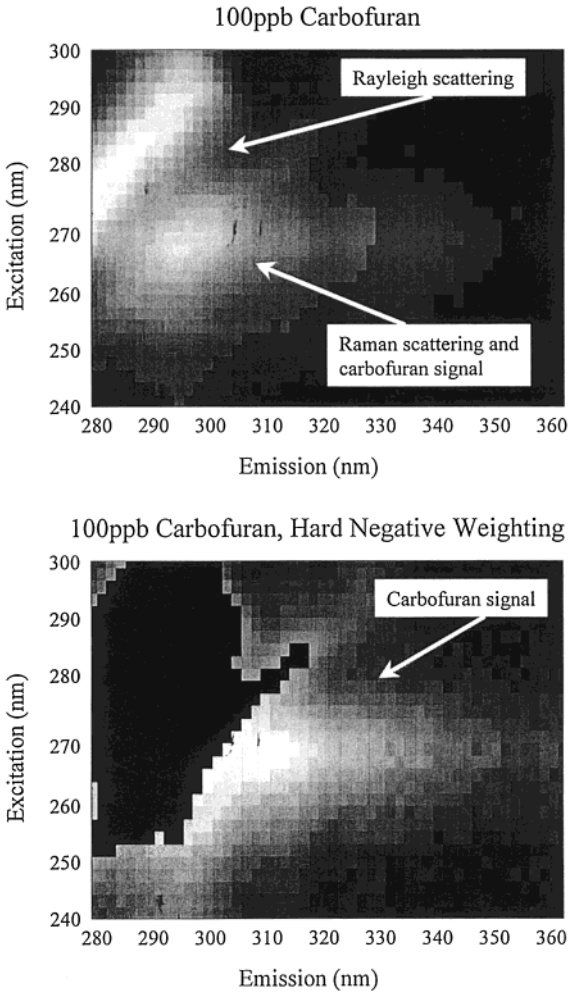

Figure 2. (a) Recorded EEM spectrum of 100 ppb carbofuran in methanol with no background correction and (b) a hard negative weight matrix applied to the same EEM spectrum.

set to zero will be modeled as having no response or signal. Therefore, for further evaluations, the truncated areas of the spectra were set to a nominal value based on the 1,1 pixel of a mean background EEM spectrum. However, this leaves the Raman scattering uncorrected. This is sufficient in the case of 1-naphthol, which is highly resolved from the scattering and has a large fluorescence cross section. However, truncation of the background scattering is unreasonable when the analyte's signal is overlapped with the background scattering. This is the case with the pesticides carbofuran and carbaryl. Truncation of both the Rayleigh and Raman scattering would remove an excessive amount of the analyte's signal, resulting in higher limits of

detection (LOD) and root-mean-squared errors (RMSE) of prediction for carbofuran and carbaryl.

Measuring only the emission wavelengths above the excitation wavelengths is also impractical. For example, the emission maximum of carbofuran in methanol is 300 nm, which lies at the top of the excitation wavelength range, 240−300 nm. A large portion of the carbofuran signal would be lost, possibly raising detection limits further. Another solution is to subtract a blank spectrum from each EEM spectrum. However, a blank is not normally available for in-situ analysis of complex environmental samples. An added disadvantage of background subtraction is the inability of the model to account for the variability of the Rayleigh and Raman scattering due to changes in sample makeup and detector saturation. Therefore, when an analyte signal is overlapped with the background scattering, an alternate method of reducing or removing the effects of the scattering is necessary.

W-PARAFAC overcomes many of the limitations of current methods for removing Rayleigh and Raman scattering. Contrary to truncating the Rayleigh scattering, the Raman scattering is also effectively removed in a W-PARAFAC model. Areas that are given a weight of zero are estimated and not modeled as zeros. This is a significant contrast with truncation, which assigns an actual value to an area. Enough of the anlaytes' signals are left in the spectrum for the model to produce reliable and accurate predictions. Another advantage is the elimination of the need to subtract a blank spectrum. When compared with conventional methods, W-PARAFAC lowers the LOD when an analyte's signal is overlapped with the scattering and, in addition, often reduces the RMSE of prediction.

The four weighting strategies, hard positive, hard negative, soft positive, and soft negative, are evaluated using a preliminary set of samples in methanol and selected figures of merit. Figures of merit for the four weighting strategies are compared with subtraction of the appropriate blank. Note that the figures of merit for the blank subtraction are the best case scenario, although the proper blank is usually not available. The most appropriate weighting strategies are then applied to a second set of samples in aqueous solution. The performance of each of the strategies will be addressed in the following sections.

Soft Weighting. Soft weighting applies a weighting proportional to the intensity of either the analyte signal or sample background. Soft-positive weighting enhances the signal of one or more analytes, while decreasing background interference. Because positive weighting is based on each analyte's unique signal, a priori knowledge of the sample makeup and solvent type is necessary. In contrast, negative weighting decreases the weight given to the background scattering, with minimal effect on the analytes' signals. Negative weighting may be performed with no prior knowledge of the sample makeup.

To determine the efficacy of soft weighting for carbofuran prediction, initial models excluded the carbaryl and 1-naphthol standards. The limits of detection for carbofuran were 3.1, 5.0, 5.3, and 4.0 ppb when subtracting the mean blank spectrum, truncating the Rayleigh scattering, using soft-positive weighting, and using soft-negative weighting, respectively (Table 2). Initial results for soft-positive weighting were encouraging, so it was applied to a data set containing all standards and mixtures. Inclusion of 1-naphthol and carbaryl provides a means to evaluate soft-positive weighting on multiple analytes with varying degrees of resolution. 1-Naphthol is totally resolved from the Raman scattering with no overlap; carbaryl is moderately resolved from the Raman scattering, with only part of its signal being overlapped.

With only the carbofuran signal enhanced, the model failed to converge. Enhancement of a single analyte was determined to be impractical for field studies, because environmental samples will normally contain fluorescent background interferents that should be included in the PARAFAC model. Subsequently, a weight matrix in which the signals of all three components were enhanced was applied to the previous data set. The LOD for carbofuran was 66 ppb—six times that for subtracting a blank spectrum (Table 3). The RMSE of carbofuran prediction in the 1-naphthol, carbaryl, and carbofuran standards and mixtures were 78, 40, 33, and 29 ppb, respectively. These are significantly higher than other types of weightings. The RMSE is expected to be higher for the samples than for the blanks, due to the proportional nature of measurement error with fluorescent intensity.

An additional problem encountered with soft-positive weighting is a result of the direct overlap of the carbofuran signal with the Raman scattering. The Raman scattering is enhanced along with the carbofuran signal, resulting in convolution of the carbofuran signal and the Raman scattering. The Raman scattering is interpreted falsely as additional fluorescence and, consequently, an abnormally high estimated sensitivity.

Soft-positive weighting results in adequate prediction errors of 2.6 and 5.1 ppb (Table 3) for 1-naphthol and carbaryl, respectively, but it is not effective for carbofuran prediction when compared with other methods. This is probably due to the weaker fluorescence signal of carbofuran and the greater overlap with the Raman scattering. In addition, there is no benefit to using soft-positive weighting for 1-naphthol and carbaryl prediction because comparable results may be obtained by simply truncating the Rayleigh scattering (data not shown) when this degree of inherent resolution exists.

Soft-negative weighting diminishes the effect of the Rayleigh and Raman scattering without pretreatment. However, the intensity of the Rayleigh scattering was extremely high, so assignment of any positive weight still prevented the model from converging. Therefore, it was necessary to first truncate the Rayleigh scattering, then apply soft-negative weighting in order to reduce the influence of Raman scattering. Soft-negative weighting was consequently applied to the limited data set mentioned previously, containing only the carbofuran standards. The LOD for carbofuran was 4.0 ppb, only 0.9 ppb higher than subtracting a mean blank and 1.0 ppb lower than truncating the Rayleigh scattering. However, when soft-negative weighting was applied to the complete data set, the limits of detection and RMSE values for carbofuran were 33 and 8.1 ppb, respectively. This is as much as three times higher than subtracting out a background spectrum (Table 3). The poor prediction for carbofuran and low sensitivity may be attributed to convolution of the Raman scattering and carbofuran signal, as with soft-positive weighting.

The overlap of the carbofuran signal and the Raman scattering is the main limitation of both soft-positive and soft-negative weighting. The carbofuran signal becomes indistinguishable from the background, causing both the Raman scattering and the carbofuran signal to simultaneously be either enhanced or reduced. The combination of the carbofuran signal and the Raman scattering may be avoided by using hard weighting rather than soft weighting.

Hard Weighting. Contrary to soft weighting, hard weighting applies a weight of either one or zero, on the basis of a predetermined cutoff value. For hard-negative weighting, a value

Table 3. Figures of Merit for Single- and Multiple-Component Solutions in Methanol

background-correction method	subtract mean blank	hard positive weighting	soft positive weighting	hard negative weighting	soft negative weighting
Prediction of 1-Naphthol					
sensitivity[a] (1×10^4) (total counts/ppb)	8.4	8.8	8.4	8.4	9.9
LOD[b] (ppb)	3.1	4.3	3.4	1.1	3.0
1-naphthol[c] (ppb)	**2.5**	**2.4**	**2.6**	**2.5**	**2.9**
carbaryl[d] (ppb)	2.2	**1.0**	**1.2**	**1.1**	**1.3**
carbofuran[d] (ppb)	1.3	**0.57**	2.2	**0.51**	2.6
mixtures[e] (ppb)	7.2	**4.9**	19	6.3	**3.9**
Prediction of Carbaryl					
sensitivity[a] (1×10^4) (total counts/ppb)	3.5	4.2	3.8	4.4	4.2
LOD[b] (ppb)	**5.0**	7.8	6.9	**5.0**	9.6
1-naphthol[d] (ppb)	6.1	8.2	**4.2**	7.8	7.7
carbaryl[c] (ppb)	**2.6**	3.3	5.1	**2.5**	5.1
carbofuran[d] (ppb)	**2.6**	6.1	12	5.5	18
mixtures[e] (ppb)	10	**2.7**	14	5.8	8.4
Prediction of Carbofuran					
sensitivity[a] (1×10^4) (total counts/ppb)	1.7	2.1	4.6	2.4	0.50
LOD[b] (ppb)	**11**	18	66	**10**	33
1-naphthol[d] (ppb)	**5.5**	12	78	**5.0**	14
carbaryl[d] (ppb)	7.1	11	40	**3.8**	16
carbofuran[c] (ppb)	5.3	7.6	33	**3.4**	8.1
mixtures[e] (ppb)	10	**6.3**	29	9.0	10

[a] Sensitivity represents the integrated area under the resolved 1 ppb spectra of the analyte. [b] The LOD is determined by three standard deviations from the blank. [c] Root mean-squared error of prediction of the analyte in the single-component standards that were used to construct the model. [d] Root mean-squared error of prediction of the analyte in the alternate single-component standards. The actual value of analyte in alternate standards is assumed to be zero. [e] Root mean-squared error of percent prediction in the multiple-component solutions. The methods with the lowest equivalent detection limits and prediction errors are in bold type as determined by an *F*-test ($F_{.95}$ $\nu_1 = 30$ and $\nu_2 = 30$).

of zero is assigned to areas of the solvent spectrum in which the intensity is above the predetermined cutoff value. For hard-positive weighting, a weight of zero is applied to areas of the pure analyte spectrum in which the intensity is below the predetermined cutoff value. With hard weighting, there is some loss of signal that is not encountered with soft weighting.

Analyses on the multiple-component solutions in methanol demonstrated that hard weighting performed superior to soft weighting overall. For carbofuran prediction, detection limits and prediction errors were consistently lower for hard-negative weighting when compared with hard-positive weighting as shown in Table 3. Although employment of hard-negative weighting resulted in lower detection limits for 1-naphthol, carbaryl, and carbofuran when multiple analytes were present in solution, hard-positive weighting resulted in more accurate prediction of each analyte. When positive weighting was employed, the RMSE of percent carbofuran prediction was 6.3 ppb, 2.7 ppb lower than that obtained when negative weighting was employed (Table 3). Although this value is comparatively small, the moderate overlap of the carbofuran signal with the carbaryl signal may favor positive weighting. A more significant difference was found for carbaryl prediction in the mixtures. The RMSE of percent carbaryl prediction is 2.7 ppb (Table 3); this is two times lower than that obtained when hard-negative weighting is used and four times lower than that obtained when a background spectrum is subtracted out. This indicates that focusing the analysis on the analytes and removing exterior information may facilitate carbaryl prediction.

Prediction errors for single-component standards and 1-naphthol in multicomponent solutions were lowest when hard weighting was employed. Hard-negative weighting resulted in detection limits of 1.1, 5.0, and 10 ppb, respectively, for 1-naphthol, carbaryl, and carbofuran (Table 3).

It should be noted that in the model in which subtraction of a background spectrum was the applied background correction, only seven of the eight blank spectra were included. One background spectrum was visibly more intense and had a significant effect on the RMSE of prediction. When this blank was included in the model, the limits of detection for carbofuran were raised to 33 ppb from 10 ppb. However, the limit of detection of 1-naphthol lowered from 3.1 to 1.4, implying that naphthol prediction becomes more accurate with increasing background variance while the inaccuracies in subtracting the background Raman scattering degrade prediction of the overlapped carbofuran signal. Incorporating this blank in the complete data set had no effect on the figures of merit when positive or negative weighting was employed.

Water Samples. Figure 3 presents an EEM spectrum of 1-naphthol, carbaryl, and carbofuran in aqueous solution. The diagonal patterns of the Rayleigh and Raman scattering are apparent in the top left corner of the EEM spectrum. In aqueous solution, the fluorescence signal of 1-naphthol is red shifted compared with that in solutions in methanol, with an excitation maximum at 295 nm and an emission maximum above 435 nm. The carbaryl signal is shifted toward the Raman scattering, overlapping both the Raman scattering and the carbofuran signal. The excitation and emission maxima for carbaryl are 285 and 310 nm, respectively. Carbofuran is also highly overlapped with the Raman scattering, with excitation and emission maxima of 280 and 300 nm, respectively.

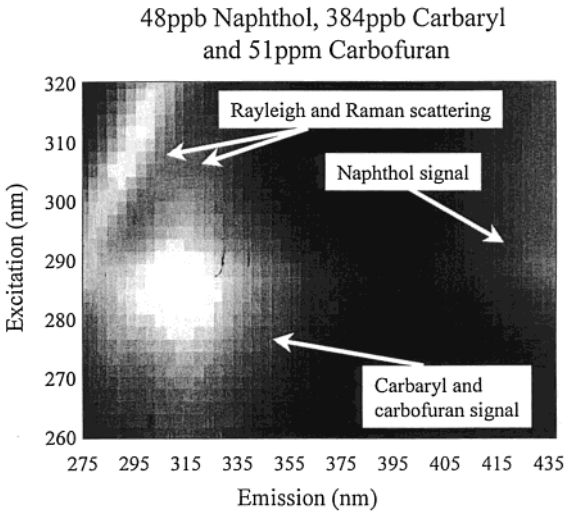

Figure 3. Recorded EEM spectrum of 48 ppb 1-naphthol, 384 ppb carbaryl, and 51 ppm carbofuran in aqueous solution.

Because of the limitations of soft weighting and the increased overlap of the carbaryl signal with the Raman scattering, only subtraction of a background spectrum and hard weighting are presented for background correction. Soft-weighted models often converged to nonsensical solutions. Contrary to the previous experiment, hard-positive weighting resulted in the lowest detection limits for all three analytes. Prediction errors for 1-naphthol and carbaryl in pure standards and multiple-component solutions were also lowest when a hard-positive weighting was employed compared with the hard-negative weighting (Table 4).

In the case of carbofuran, the LOD was decreased more than 3 ppm with both types of weighting, compared with subtracting a background spectrum out (Table 4). Also, the difference of 0.2 ppm in detection limits for hard-positive and -negative weighting strategies is statistically insignificant for carbofuran prediction. In addition, hard-negative weighting performed superior for prediction of carbofuran in the presence of other analytes. The RMSE of carbofuran prediction in the 1-naphthol, carbaryl, carbofuran standards and in multiple-component solutions were 2.2, 2.9, 1.5, and 3.2 ppm, respectively (Table 4).

Table 4. Figures of Merit for Single- and Multiple-Component Solutions in Water

background-correction method	subtract mean blank	hard positive weighting	hard negative weighting
	Prediction of 1-Naphthol		
sensitivity[a] (1×10^8) (total counts/ppb)	1.3	1.1	1.4
LOD[b] (ppb)	**1.8**	**2.0**	2.9
1-naphthol[c] (ppb)	1.8	**0.9**	2.0
carbaryl[d] (ppb)	**1.4**	3.1	**1.2**
carbofuran[d] (ppb)	**0.4**	2.3	0.9
mixtures[e] (ppb)	**6.2**	**5.2**	**5.8**
	Prediction of Carbaryl		
sensitivity[a] (1×10^7) (total counts/ppb)	2.3	1.0	2.2
LOD[b] (ppb)	**22**	**20**	36
1-naphthol[d] (ppb)	**8.7**	24	22
carbaryl[c] (ppb)	5.3	**3.2**	17
carbofuran[d] (ppb)	**13**	62	51
mixtures[e] (ppb)	**16**	29	38
	Prediction of Carbofuran		
sensitivity[a] (1×10^5) (total counts/ppb)	1.8	1.7	2.1
LOD[b] (ppm)	5.7	**1.9**	**2.1**
1-naphthol[d] (ppm)	**2.1**	**1.6**	**2.2**
carbaryl[d] (ppm)	**2.0**	5.6	2.9
carbofuran[c] (ppm)	**1.4**	**1.6**	**1.5**
mixtures[e] (ppm)	**4.0**	5.3	**3.2**

[a] Sensitivity represents the integrated area under the resolved 1 ppb spectra of the analyte. [b] The LOD is determined by three standard deviations from the blank. [c] Root mean-squared error of prediction of the analyte in the single-component standards that were used to construct the model. [d] Root mean-squared error of prediction of the analyte in the alternate single-component standards. The actual value of analyte in alternate standards is assumed to be zero. [e] Root mean-squared error of prediction in multiple-component solutions. The methods with the lowest equivalent detection limits and prediction errors are in bold type as determined by an *F*-test ($F_{.95}$ $\nu_1 = 30$ and $\nu_2 = 30$).

CONCLUSIONS

The effects of Rayleigh and Raman scattering cannot always be efficiently mitigated by background subtraction, so alternatives must be investigated. W-PARAFAC does successfully mitigate problems associated with the presence of a Rayleigh and Raman background and often performs superior to subtraction of the appropriate blank. Hard weighting methods achieved better results compared with soft weighting strategies. Soft weighting led to convolution of the Raman scattering and the carbofuran signal, resulting in imprecise estimations of sensitivity and distortions to the estimated signal. Although this paper did not discuss in depth the use of soft weighting on signals from other analytes, it can be inferred that soft weighting would result in convolution of the analytes' signals with one another. Future studies of soft weighting may provide insight on how to minimize this effect. Hard positive and negative weighting strategies seemed to perform equivalently well in all cases except carbofuran prediction in methanol. This can be attributed to the greater amount of Raman scattering present in the EEMs of solutions in methanol. For the solutions in methanol, weighting strategies achieved better or equivalent results to subtraction of a background in 14 out of 15 instances. In only one instance did subtraction of a background spectrum result in better prediction. In conclusion, hard weighting should be employed rather than soft weighting in cases where there is no blank available. Positive and negative weightings are suitable for most situations and should be applied according to the goals of the researcher. In the case of the analyte's signal being highly overlapped with especially intense background scattering, hard negative weighting should be employed.

ACKNOWLEDGMENT

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

Received for review April 20, 1999. Accepted November 21, 1999.

AC990418J