Analytica Chimica Acta 677 (2010) 97–107

Contents lists available at ScienceDirect

Analytica Chimica Acta

journal homepage: www.elsevier.com/locate/aca

Four-way kinetic-excitation-emission fluorescence data processed by multi-way algorithms. Determination of carbaryl and 1-naphthol in water samples in the presence of fluorescent interferents

Rubén M. Maggio, Patricia C. Damiani, Alejandro C. Olivieri*

Departamento de Química Analítica, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Instituto de Química de Rosario (IQUIR-CONICET), Suipacha 531, Rosario, S2002LRK, Argentina

ARTICLE INFO

Article history:
Received 11 May 2010
Received in revised form 28 July 2010
Accepted 28 July 2010
Available online 5 August 2010

Keywords:
Four-way kinetic fluorescence data
Second-order advantage
Unfolded partial least-squares
Residual trilinearization
Parallel factor analysis

ABSTRACT

Four-way data were obtained by recording the kinetic evolution of excitation-emission fluorescence matrices for samples containing the analytes carbaryl and 1-naphthol, two widely employed pesticides, in the concentration ranges 0–363 $\mu g\,L^{-1}$ and 0–512 $\mu g\,L^{-1}$, respectively. The reaction followed was the alkaline hydrolysis of carbaryl to produce 1-naphthol, a fact which introduced strong linear dependencies and multi-linearity losses in the analyzed system. Data processing was performed with unfolded partial least-squares combined with residual trilinearization (U-PLS/RTL) and also with a suitably initialized and restricted parallel factor model (PARAFAC), combined with calibration based on multi-linear regression. U-PLS/RTL is shown to be significantly simpler in its implementation and to provide similar figures of merit. The applied chemometric strategy is able to successfully determine the analytes in water samples containing uncalibrated interferences, such as other commonly employed agrochemicals and also a naturally occurring background signal.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Analytical applications of multi-way calibration models based on luminescence excitation-emission matrix (EEM) data are continuously growing and applied to diverse research fields [1–3]. The inclusion of an additional data dimension increases the selectivity and sensitivity of the analysis [4], as has been accomplished by registering time-dependent EEM information. In these cases, a reaction may produce a fluorescent species from a non-fluorescent analyte, or a natively fluorescent analyte may generate a non-fluorescent species. The time dependence may also arise from naturally time-decaying EEM data [5]. In any case, only a few reports exist on the processing of kinetic-EEM data aimed at the determination of one or several analytes embedded in a complex background. Pertinent examples include the oxidation of catecholamines [6], mixtures of leucovorin or folic acid with methotrexate [7–10], the alkaline hydrolysis of procaine [11] and carbaryl [12], the derivatization of malondialdehyde [13], and the photochemical degradation of several pesticides [14], polycyclic aromatic hydrocarbons [15], and folic acid and its two main metabolites [16].

Several algorithms are available for the convenient processing of four-way data. Quadrilinear models such as parallel factor analysis (PARAFAC) [5,6] and alternating quadrilinear decomposition [12] are appealing because they provide useful physicochemical information, in terms of kinetic and spectral profiles for the different sample components. However, kinetic systems pose challenges to these latter algorithms due to: (1) strong linear dependencies among component profiles, typical of kinetic systems, and (2) multi-linearity losses due to the fact that the reaction may progress during the time required for registering a complete EEM. In these cases, therefore, the analyst may prefer more flexible, latent-structured multivariate methods, such as the multi-dimensional version of partial least-squares (N-PLS) [6]. More importantly, the recently introduced residual trilinearization (RTL) procedure [7,8], when combined with trilinear least-squares (TLLS) [7,8], with both unfolded-PLS (U-PLS) and N-PLS [8,10,11,16], and with an artificial neural network model [13], enables the combined model to achieve the second-order advantage (i.e., the quantitation of analytes in the presence of uncalibrated interferences) [17]. It is important to notice that quadrilinear models such as PARAFAC naturally achieve the latter advantage, whereas the remaining algorithms require the RTL procedure in order to obtain it [4]. As new kinetic-luminescent analytical systems are studied, it is likely that the latter flexible techniques will be more and more appreciated in order to develop automatic, easily implementable methods based on four-way analysis of complex samples.

* Corresponding author. Tel.: +54 341 4372704; fax: +54 341 4372704.
E-mail addresses: olivieri@iquir-conicet.gov.ar, aolivier@fbioyf.unr.edu.ar (A.C. Olivieri).

0003-2670/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.aca.2010.07.045

In this report, we discuss the four-way kinetic-EEM analysis of binary mixtures of carbaryl (1-naphthyl methylcarbamate) and its degradation product 1-naphthol in the presence of interferences. Mixtures of carbaryl and 1-naphthol were previously determined using second-order EEM data [18]. However, carbaryl is significantly less fluorescent than 1-naphthol, and thus its conversion to 1-naphthol leads to a considerable sensitivity improvement. On the other hand, carbaryl alone has been quantified by kinetic-EEM four-way data using a quadrilinear decomposition algorithm [12]. However, this single-analyte system does not present the complex challenges of the one herein reported, where one analyte (1-naphthol) is also the reaction product of the second analyte (carbaryl), introducing strong dependencies in the analyte profiles. The situation becomes even more complex in the presence of interferents. We show a new way to apply PARAFAC in the presence of linear dependencies and correlations, and the advantages of employing the U-PLS/RTL model for studying the present mixtures.

2. Experimental

2.1. Apparatus

All fluorescence measurements were performed on a fast-scanning Varian Cary Eclipse spectrofluorometer equipped with two Czerny-Turner monochromators, a xenon flash lamp, a quartz cell, and connected to a PC microcomputer via an IEEE 488 (GPIB) serial interface.

Excitation-emission fluorescence matrices were recorded in the following ranges: excitation, 244–310 nm each 3 nm, emission, 321–481 nm each 2 nm, time, 16 min (from 2 to 18 min from the beginning of the reaction) each 0.8 min. The third-order array was thus of size $23 \times 81 \times 21$, making a total of 39,123 data points. The slit band widths for the excitation and emission monochromators were fixed at 5 nm, and the detector voltage at 600 V. The cell was thermostated at 35 C. A wavelength scanning speed of 9600 nm/min was employed, so that a complete excitation-emission fluorescence matrix was obtained in ca. 30 s. Notice that this latter time is close to the time within successive EEM recording (48 s), and this is a potential source of multi-linearity loss.

2.2. Reagents

Analytical reagents grade chemicals, pure solvents and doubly distilled water were used in all experiments. Carbaryl, 1-naphthol, fuberidazole, thiabendazole, imacloprid, imazalil, 2,4-dichlorophenoxyacetic acid, carbendazim, neburon, linuron, diuron, chlorsulfuron, isopropturon, mercaptodimethopur, metomil and propoxur were purchased from Sigma–Aldrich Co. (St. Louis, MO).

2.3. Stock standard and working standard solutions

Stock standard solutions of carbaryl (1.23 g L^{-1}) and 1-naphthol (1.12 g L^{-1}) were prepared in 10.00 mL volumetric flasks by dissolving accurately weighed amounts of the drugs in acetonitrile and completing to the mark with the same solvent.

Working solutions of carbaryl (1.23 mg L^{-1}) and 1-naphthol (1.12 mg L^{-1}) were prepared by appropriate dilution of the corresponding stock standard solutions in water, employing 10.00 mL volumetric flasks.

2.4. Calibration samples

A set of 13 calibration solutions (in the form of a central composite design with five replicates of the central point) containing the analytes in the ranges 0.0–363 μg L^{-1} for carbaryl and 0.0–512 mg L^{-1} for 1-naphthol were prepared adding 1 mL of phosphate buffer (0.1 M, pH 10.2) to the corresponding mixtures of working standard solutions, and completing to the mark with water in 10.00 mL volumetric flasks. Each sample was measured 2 min after its preparation. The concentrations are collected in Table 1.

Table 1
Calibration concentrations for carbaryl and 1-naphthol.[a].

Sample	Carbaryl	1-Naphthol
1	53	75
2	310	75
3	53	437
4	310	437
5	0	256
6	364	256
7	182	0
8	182	512
9	182	256
10	182	256
11	182	256
12	182	256
13	182	256

[a] All concentrations in μg L^{-1}.

2.5. Validation and test samples

A validation sample set V was prepared containing the analytes in concentrations within the corresponding calibration ranges but different than the specific calibration values. This was done by appropriate dilution of mixtures of the corresponding working standard solutions with 1 mL of phosphate buffer (0.1 M, pH 10.2) and water in 10.00 mL volumetric flasks. Each sample was measured 2 min after its preparation.

An additional spiked sample set S was prepared, containing the analytes in concentration within the calibration ranges, and also fuberidazole and/or thiabendazole as potential interferents. The preparation details were analogous to those described above for the validation set. These interfering test samples are intended to mimic truly unknown samples composed of uncalibrated substances, where an unknown background may occur. The inclusion of known chemical components in these test samples had the purpose of checking whether the multivariate algorithm was able to successfully retrieve their corresponding spectral and time profiles.

Finally, a four-sample set T was prepared from a natural water sample containing unknown fluorescent components. Two of them were spiked one of the analytes, and two with both analytes. The concentrations were different than those employed for the remaining sets. This final set T served to explore the achievement of the second-order advantage by the second-order multivariate procedures under a real situation where unknown sample components may be present in the unknown specimens.

3. Theory

3.1. PARAFAC

The theory of PARAFAC is well known and thus only a brief account is presented here. Four-way data are created by joining the third-order data arrays for the calibration samples and for each of the analyzed test samples. Application of the PARAFAC model to the latter four-way data arrays requires fitting the following expression:

$$X_{ijkl} = \sum_{n=1}^{N} a_{in}\, b_{jn}\, c_{kn}\, d_{\ln} + E_{ijkl} \quad (1)$$

where X_{ijkl} is an element of the four-way array of kinetic-excitation-emission fluorescence signals, a_{in} is the score of component n in sample i, N is the total number of responsive components, b_{jn}, c_{kn} and d_{ln} are the loading elements in the excitation, emission and time dimensions, respectively, and E_{ijkl} is an element of the array of errors not fitted by the model. The scores are collected in matrix **A**, of size $(I_{cal}+1)\times N$, where I_{cal} is the number of calibration samples. The loadings are normalized to unit length, and collected into the profile matrices **B**, **C** and **D**, of size $J\times N$, $K\times N$ and $L\times N$, respectively, where J, K and L are the number of excitation, emission and time channels, respectively. The structure of the model Eq. (1) is called quadrilinear, and the decomposition is usually accomplished through alternating least-squares [19,20].

There are several relevant issues regarding the application of the PARAFAC model for the calibration of four-way data: (1) initializing the algorithm, (2) applying restrictions to the least-squares fit, (3) establishing the number of responsive components, (4) identifying specific components from the information provided by the model and (5) calibrating the model in order to obtain absolute concentrations for a particular component in an unknown sample.

Initializing PARAFAC for the study of four-way arrays can be done using several options implemented in the PARAFAC package [19]: (1) singular value decomposition (SVD) vectors, (2) random orthogonalized values and (3) the best-fitting model of several models fitted using a few iterations. The best initialization option for the present system will be described below in the appropriate section. Notice that during the fit non-negativity restrictions were applied in order to recover physically recognizable profiles, as required by linearly dependent systems such as the kinetic ones.

The number of responsive components (N) can be estimated by several methods. A useful technique is CORCONDIA, a diagnostic tool considering the PARAFAC internal parameter known as core consistency [21]. However, in linearly dependent systems requiring restrictions, a more useful technique is the consideration of the PARAFAC sum of squared errors (SSE), i.e., the sum of squared elements of the array ***E*** in Eq. (1) [22]:

$$\mathrm{SSE}=\sum_{i=1}^{I_{cal}+1}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{l=1}^{L}(E_{ijkl})^2 \tag{2}$$

Usually this parameter decreases with increasing N, until it stabilizes at a value corresponding to the optimum number of components.

Identification of the chemical constituents under investigation is done with the aid of the estimated profiles, mainly the emission and excitation spectra, and comparing them with those for a standard solution of the analyte of interest. This is required since the components obtained by decomposition of ***X*** are sorted according to their contribution to the overall spectral variance, and this order is not necessarily maintained when the unknown sample is changed.

Absolute analyte concentrations are obtained after calibration, because the four-way array decomposition only provides relative values (the scores contained in matrix **A**). Calibration is usually done by means of the set of standards with known analyte concentrations, a procedure which is repeated for each new test sample analyzed. Notice, however, that in the presence of linear dependencies, this simple pseudo-univariate calibration methodology may not be employed and should be replaced by a multiple linear regression (MLR) approach correlating elements of several columns of the **A** matrix with the analyte concentrations. These elements are contained in a sub-matrix $\mathbf{A}^*$ corresponding only to the calibration scores, and to the columns related to the analytes. Fig. 1 explains in detail the applied procedure.

Once built the sub-matrix $\mathbf{A}^*$, the following MLR expression is considered for calibration of analyte n:

$$\mathbf{y}_n=\mathbf{b}_{\mathrm{MLR},n}\mathbf{A}^*+\mathbf{e} \tag{3}$$

where $\mathbf{y}_n$ is a column vector containing the calibration concentrations for the analyte, $\mathbf{b}_{\mathrm{MLR},n}$ is a vector of parameters to be fitted by the MLR model and **e** collects the model errors. Eq. (3) provides the coefficients by least-squares:

$$\mathbf{b}_{\mathrm{MLR},n}=(\mathbf{A}^*)^{+}\mathbf{y}_n \tag{4}$$

where '+' stands for the generalized inverse. Finally, the analyte concentration in the test sample is estimated by applying the model to the test sample scores contained in the vector $\mathbf{a}_u^*$, extracted from the $\mathbf{A}^*$ matrix corresponding to the same columns employed for calibration (see Fig. 1):

$$\mathbf{y}_n=(\mathbf{b}_{\mathrm{MLR},n})^T\mathbf{a}_u^* \tag{5}$$

In the specific case of the analyte carbaryl, the MLR approach correlates a sub-matrix $\mathbf{A}^*$ containing scores which are related to the reaction product 1-naphthol. In this way, the sensitivity of the carbaryl determination considerably increases, because scores for a strongly fluorescent reaction product (1-naphthol), which are also related to carbaryl through the hydrolysis reaction, are included in the quantitative analysis.

3.2. U-PLS/RTL

In the U-PLS method, the original third-order data are unfolded into vectors before PLS is applied [8,23]. In this algorithm, concentration information is employed in the calibration step, without including data for the unknown sample. The I_{cal} calibration data matrices are first vectorized into $JKL\times 1$ vectors, and then a usual PLS model is built using these data together with the vector of calibration concentrations **y** (size $I_{cal}\times 1$). This provides a set of loadings **P** and weight loadings **W** (both of size $JKL\times A$, where A is the number of latent factors), as well as regression coefficients **v** (size $A\times 1$). The parameter A can be selected by techniques such as leave-one-out cross-validation, as described in Ref. [24].

If no unexpected components occurred in the test sample, **v** could be employed to estimate the analyte concentration according to:

$$\mathbf{y}_u=\mathbf{t}_u^T\mathbf{v} \tag{6}$$

where $\mathbf{t}_u$ is the test sample score, obtained by projecting the vectorized data for the test sample $\mathrm{vec}(\mathbf{X}_u)$ onto the space of the A latent factors:

$$\mathbf{t}_u=(\mathbf{W}^T\mathbf{P})^{-1}\mathbf{W}^T\mathrm{vec}(\mathbf{X}_u) \tag{7}$$

where $\mathrm{vec}(\cdot)$ implies the vectorization operator.

When unexpected constituents occur in $\boldsymbol{X}_u$, then the sample scores given by Eq. (7) are unsuitable for analyte prediction through Eq. (6). In this case, the residuals of the U-PLS prediction step [s_p, see Eq. (8) below] will be abnormally large in comparison with the typical instrumental noise level:

$$s_p=\frac{||\mathbf{e}_p||}{(JKL-A)^{1/2}}=\frac{||\mathrm{vec}(\mathbf{X}_u)-\mathbf{P}\mathbf{t}_u||}{(JKL-A)^{1/2}} \tag{8}$$

where $||\cdot||$ indicates the Euclidean norm.

This situation can be handled by a separate procedure called residual trilinearization, based on the Tucker3 model of the unexpected effects [8]. RTL aims at minimizing the residuals computed while fitting the sample data to the sum of the relevant contributions:

$$\mathrm{vec}(\mathbf{X}_u)=\mathbf{P}\mathbf{t}_u+\mathrm{vec}[\mathrm{Tucker3}(\mathbf{E}_p)]+\mathbf{e}_u \tag{9}$$

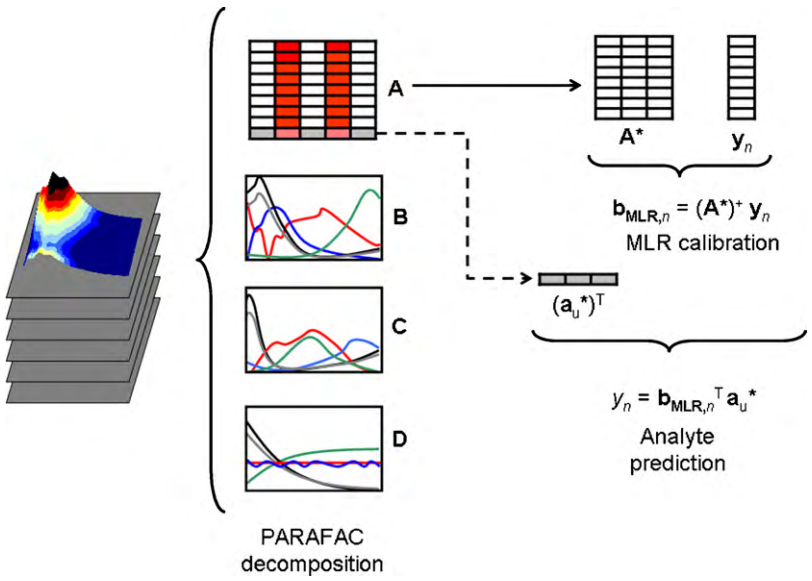

Fig. 1. Scheme illustrating the obtainment of data from the PARAFAC **A** matrix in order to calibrate an MLR model for the quantitation of a specific analyte. From the original **A** matrix provided by PARAFAC decomposition, the calibration scores corresponding to the columns related to the correlated analytes are extracted (white elements) and employed to build a calibration sub-matrix $\mathbf{A}^*$. Those unrelated to the analytes (red elements) are discarded. An MLR model is built with $\mathbf{A}^*$ and the vector **y** of calibration concentrations, providing the regression vector $\mathbf{b}_{MLR,n}$ associated to analyte *n*. Likewise, the test scores corresponding to the unknown sample (gray elements) form the score vector $\mathbf{a}_u^*$, which furnishes the predicted concentration y_u upon multiplication by $\mathbf{b}_{MLR,n}$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

$$\mathbf{E}_p = \text{reshape}(\mathbf{e}_p) \tag{10}$$

where reshape(·) indicates transforming a $JKL \times 1$ vector into a $J \times K \times L$ three-way array and $\mathbf{e}_p$ is from Eq. (8). During this RTL procedure, the loadings **P** are kept constant at the calibration values, and $\mathbf{t}_u$ is varied until the final RTL residual error s_u is minimized using a Gauss–Newton procedure, with s_u given by:

$$s_u = ||\mathbf{e}_u||/(JKL)^{1/2} \tag{11}$$

where $\mathbf{e}_u$ is from Eq. (9). Once this is done, the analyte concentrations are provided by Eq. (6), by introducing the final $\mathbf{t}_u$ vector found by the RTL procedure.

To analyze the presently discussed data, the Tucker3 model in Eq. (9) is constructed by restricting the loadings to be orthogonal, and with no special constraints on the core elements. For a single unexpected component, the Tucker3 model is built with a single component in all dimensions, which is straightforward and provides the three corresponding interferent profiles. For additional unexpected constituents, however, the retrieved profiles no longer resemble true spectra (or time profiles). Moreover, in this latter case, several different Tucker3 models could in principle be constructed, because the number of loadings may be different in each dimension. We notice that the aim which guides the RTL procedure is the minimization of the residual error term s_u of Eq. (11) to a level compatible with the degree of noise present in the measured signals. Therefore, if two unexpected components are considered, for example, one should explore the possible Tucker3 models having one or two loadings in each dimension, and select the simplest model giving a residual value of s_u which is not statistically different than the minimum one. For more unexpected components a similar procedure is recommended. The final Tucker3 model selected to model the unexpected effects is the simplest one which provides a value of s_u which is not statistically different than the noise level.

3.3. Figures of merit

Concerning the figures of merit which can be achieved by the applied algorithms, adequate expressions are known for estimating the sensitivity in the case of U-PLS and classical PARAFAC analysis [25,26]. However, in the latter case they have not been extended to the presently applied MLR calibration approach. For U-PLS regression analysis, the sensitivity for the calibrated analyte $S_{UPLS,n}$ is directly given by the length of the corresponding regression vector, which depends on the calibration parameters (loadings and regression coefficients in score space), i.e.:

$$S_{UPLS,n} = ||\mathbf{b}_{UPLS,n}|| \tag{12}$$

where

$$||\mathbf{b}_{UPLS,n}|| = \mathbf{W}(\mathbf{P}^T\mathbf{W})^{-1}\mathbf{v} \tag{13}$$

In the case of PARAFAC, when pseudo-univariate calibration is possible for analyte *n*, the sensitivity has been shown to be provided by the following expression:

$$S_{PARAFAC,n} = s_n\left\{\left[\left(\mathbf{B}^T\mathbf{B}\right) * \left(\mathbf{C}^T\mathbf{C}\right) * \left(\mathbf{D}^T\mathbf{D}\right)\right]^{-1}\right\}_{nn}^{-1/2} \tag{14}$$

where **B**, **C** and **D** are the loading matrices provided by PARAFAC in the three data dimensions, s_n is the total analyte signal at unit concentration, '*' is the element-wise Hadamard matrix product, and '*nn*' implies selection of the (*n*,*n*) element corresponding to the *n*th. analyte of interest. For MLR calibration, a reasonably adapted expression can be deduced employing error propagation arguments:

$$S_{PARAFAC,n} = \left[\frac{1}{\sum_{m=1}^{M} b_{MLR,m}^2 \left[\left(\mathbf{B}^T\mathbf{B}\right) * \left(\mathbf{C}^T\mathbf{C}\right) * \left(\mathbf{D}^T\mathbf{D}\right)\right]_{mm}^{-1}}\right]^{1/2} \tag{15}$$

where the *m* index runs over the *M* components which are correlated by the MLR approach in order to quantitate the analyte, and $b_{MLR,m}$ is the corresponding element of the regression $\mathbf{b}_{MLR,n}$ vector. When $M = 1$, Eq. (15) reduces to Eq. (14), as expected.

The error propagation Eqs. (12) and (15) can be tested by a complementary methodology, based on Monte Carlo noise addition, as has been discussed previously [26]. In this latter approach, an operational definition of the sensitivity is employed:

$$S_{\text{Monte Carlo},n} = \frac{s_X}{s_y} \tag{16}$$

where s_X is the uncertainty in the signal for the test sample, and s_y is the standard error in the predicted concentration. This approach has been found to be fruitful in uncovering the relation between the error propagation formulas and several different calibration scenarios [26]. Briefly, it involves adding Gaussian distributed noise with a certain standard deviation (in our case 1% of the maximum calibration signal) to a given test sample data, predicting the analyte concentration using a certain algorithm, repeating the calculations a large number of times (usually 1000) and estimating the standard deviation of the distribution of predicted concentrations [26]. Application of Eq. (16) then furnishes an estimate of the component sensitivity, which can be compared with the corresponding expressions (12) or (15).

3.4. Comparison of prediction results

Prediction results concerning analyte concentrations were made using paired *t*-statistics [27]. Specifically, the differences (Δ) between nominal and predicted concentrations (or between concentrations predicted by two methods) for both analytes were computed, followed by comparison of the experimental t_{exp} with the critical t_{crit}, where $t_{exp} = \left|\bar{\Delta}\right| \sqrt{n-1}/s$ ($\bar{\Delta}$ and s are the mean and standard deviation of the different Δ values, n is the number of pairs and $| \, |$ indicates modulus). If $t_{exp} < t_{crit}$ then no significant differences exist between two concentration sets.

3.5. Software

PARAFAC was implemented with the routine provided by Bro at http://www.models.kvl.dk/source/ which is written in MATLAB [28]. The latter routine and the ones for performing U-PLS/RTL calculations were incorporated into a graphical interface, of the type already published for second-order multivariate calibration [29]. The PARAFAC calibration variant applying MLR to the score-concentration data was developed in MATLAB as an in-house routine.

4. Results and discussion

4.1. Carbaryl hydrolysis

Before conducting the analytical protocol, a series of exploratory experiments was performed in order to select an appropriate working pH. This is due to the fact that the kinetics of the carbaryl hydrolysis is strongly pH-dependent. In a previous four-way analysis of this analyte, the selected pH was 9.3, because under the latter conditions the reaction was slow enough to permit the recording of a sufficient number of EMMs [12]. However, using a fast-scanning spectrofluorimeter, the speed of EEM recording is considerably increased, and thus a faster kinetics is allowed, significantly reducing the time required for analysis. In our case, after checking the following pH values: 9.1, 10.2, 11.0 and 12.1, the final working pH was selected as 10.2, which provided a reasonable compromise between hydrolysis time and EEM registering speed.

Fig. 2 shows the excitation-emission fluorescence contour plots which can be recorded in the working wavelength ranges as a function of time, when carbaryl is hydrolyzed to produce 1-naphthol at pH 10.2. As can be seen, the fluorescence intensity developed with time is significantly larger than that for carbaryl alone, prompting to the application of a kinetic-spectroscopic analysis for increasing the sensitivity of its determination. However, spectral overlapping with the hydrolysis product and also to other test sample components (see below) makes it necessary to resort to multivariate calibration techniques.

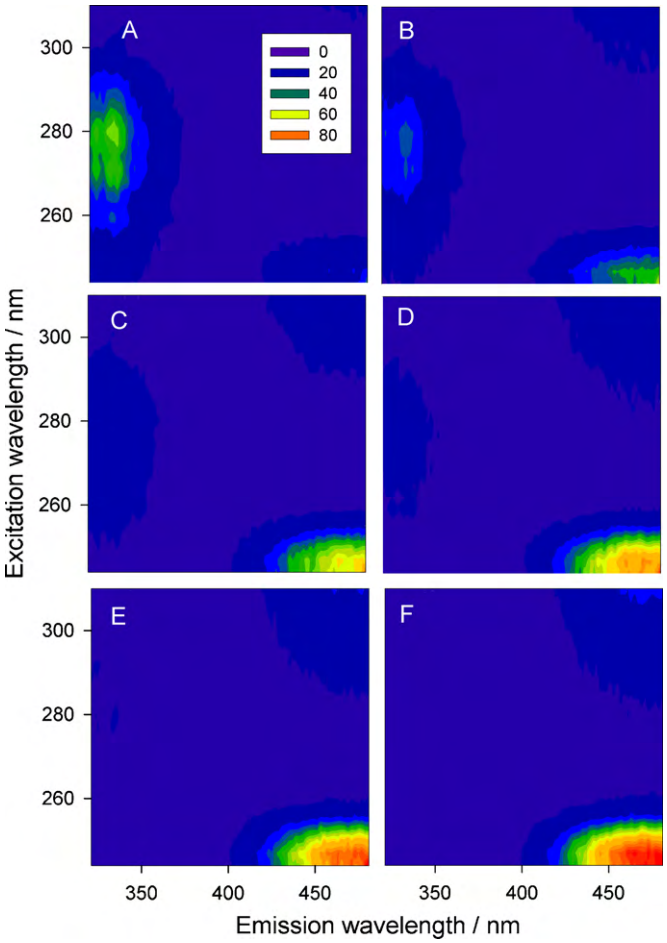

Fig. 2. Excitation-emission fluorescence contour plots (excitation, 244–310 nm each 3 nm, emission, 321–481 nm each 2 nm) recorded during carbaryl hydrolysis to produce 1-naphthol at pH 10.2 and 35 C after the following reaction times: (A) 120, (B) 210, (C) 300, (D) 390, (E) 470 and (F) 930 s. The insert in plot (A) shows the fluorescence intensities in arbitrary units.

In the same wavelength regions and under the same experimental conditions, only two other agrochemicals evidence fluorescence emission: fuberidazole and thiabendazole. A contour plot of a typical mixture of these foreign compounds is shown in Fig. 3A. The EEM surfaces for these potential interferents remain almost constant in time during the carbaryl reaction. Fig. 3B, on the other hand, shows the contour plot for the fluorescence signal of a naturally occurring water samples having a responsive background. This water sample was employed to generate the sample set T.

4.2. Determination of analytes in the validation sample set V: PARAFAC

The validation set of samples V contain both analytes, at concentrations different than those employed for calibrating the multivariate models. In order to simultaneously quantitate the analytes, four-way arrays were created from the third-order data (kinetic-excitation-emission third-order arrays) for each validation sample and all calibration samples. In order to obtain physically reasonable profiles using PARAFAC, two simultaneous strategies were applied: (1) initialization was made by the best results of a series of small runs involving singular value decomposition (SVD) vectors and several sets of orthogonal random loadings, and (2) scores and loadings were restricted to be non-negative during the alternating least-squares fitting phase. The number of fluorescent components in the four-way arrays was estimated by analyzing the

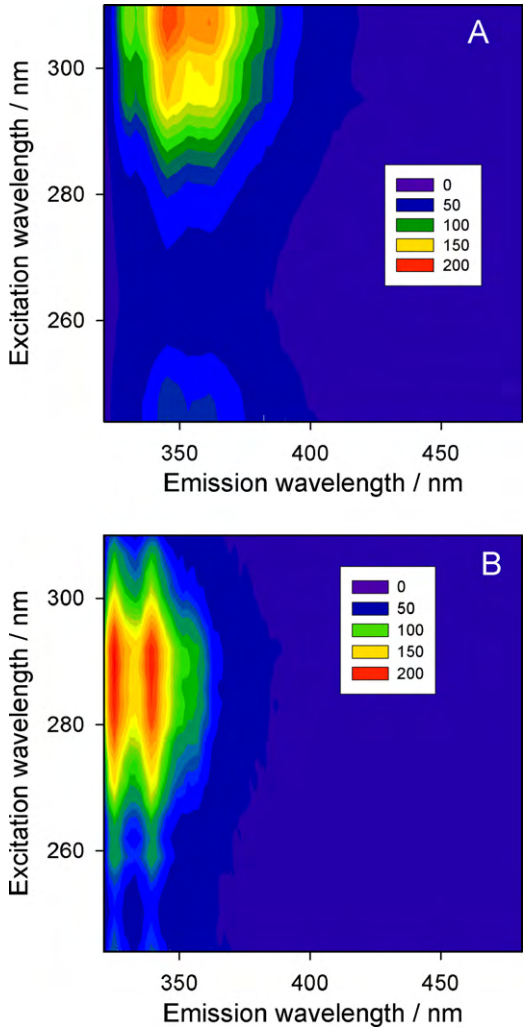

Fig. 3. Excitation-emission fluorescence contour plot (excitation, 244–310 nm each 3 nm, emission, 321–481 nm each 2 nm) for (A) a mixture of thiabendazole (212 $\mu g\,L^{-1}$) and fuberidazole (8 $\mu g\,L^{-1}$) at pH 10.2 and 35 C, and (B) a natural water sample with a fluorescent background under identical conditions. The inserts show the intensities in arbitrary units.

SSE value for increasing number of components. Three PARAFAC components were required to describe the variability in these data arrays, providing reasonable emission and excitation profiles. The kinetic profiles, however, deserve some attention.

The need of three PARAFAC components to describe the contribution of the presently studied analytes can be explained on the following basis. Since both carbaryl and 1-naphthol are fluorescent, and hydrolysis of carbaryl produces 1-naphthol, the excitation-emission matrix for a given sample will vary as a function of time as:

$$\mathbf{X}(t) = \mathbf{S}_{NAP}c_{NAP} + \mathbf{S}_{NAP}c_{CAR}f(t) + \mathbf{S}_{CAR}c_{CAR}[1 - f(t)] \quad (17)$$

where $\mathbf{X}(t)$ is the excitation-emission matrix data matrix at time t for a given sample, $\mathbf{S}_{CAR}$ and $\mathbf{S}_{NAP}$ are the excitation-emission data matrices for pure carbaryl and 1-naphthol respectively at unit concentration, c_{CAR} and c_{NAP} are the corresponding nominal concentrations in this particular sample, and $f(t)$ is a function of time describing the kinetics of the conversion of carbaryl into 1-naphthol.

If c_{NAP} and c_{CAR} were constant in a series of experiments, then only two components will suffice to model Eq. (12), which could be written as the sum of two bilinear components:

$$\mathbf{X}(t) = \mathbf{S}_{NAP}[c_{NAP} + c_{CAR}f(t)] + \mathbf{S}_{CAR}c_{CAR}[1 - f(t)] \quad (18)$$

However, in the present case carbaryl and 1-naphthol are independent analytes, and a designed calibration set has been employed, where c_{NAP} and c_{CAR} are different for each sample. Consequently, there is no single function of time which can describe the time evolution of the first two terms of Eq. (17), and hence three PARAFAC components are required. This implies the need of a **B** loading matrix having three columns, two of them equal to the (normalized) emission spectrum of 1-naphthol. Likewise, the **C** matrix will contain three columns, two of them with the excitation 1-nahthol spectrum. The loading matrix **D**, in turn, would be given by:

$$\mathbf{D} = [N_1 c_{NAP}\mathbf{1} | N_2 c_{CAR} f(\mathbf{t}) | N_3 c_{CAR}[1 - f(\mathbf{t})]] \quad (19)$$

where $\mathbf{1}$ is a $K \times 1$ vector having all elements equal to 1, $\mathbf{t}$ is a vector having the K time values, $f(\mathbf{t})$ is the function of time of Eq. (17), and N_1, N_2 and N_3 are normalization constants. However, a **D** matrix composed of linear combinations of the columns of Eq. (18) will also describe the time variation of the instrumental data. This means that the PARAFAC least-squares fit may give linear combinations of the expected ideal kinetic profiles, which may depend on the starting values for the least-squares PARAFAC minimization.

Analysis of a typical validation sample rendered the profiles in the three dimensions which are shown in Fig. 4. The emission and excitation profiles matched the expected spectral properties of carbaryl and 1-naphthol. However, in the case of the time profiles, it is apparent that they are not correct for 1-naphthol, since they consist of: (1) a time-decreasing profile for one of the spectrally active components and (2) a time-increasing profile for the other component (Fig. 4). In principle, one of them would be expected to be constant, due to the 1-naphthol originally present in the solution (see above). This suggests the presence of linear combinations in the time mode, which are expressed into corresponding correlations in the PARAFAC score matrix **A**.

One alternative to cope with the above problem is to correlate all relevant columns of **A** with the analyte concentrations using the MLR approach described in Section 3.1, as has been previously done in the modelling of four-way fluorescence data for other kinetic systems [6,11]. In this case, the scheme outlined in Fig. 1 is applied, except that all columns of **A** may in principle be useful for quantitation, because no unexpected components occur in these V samples. When PARAFAC results were processed in this latter way, analyte concentrations could be reasonably predicted in the validation set of samples. The statistical summary of this analysis is presented in Table 2. As can be seen, acceptable figures of merit are obtained, in terms of root-mean square error (RMSE) and relative error of

Table 2
Prediction results for validation samples of set V.[a]

Sample	Carbaryl			1-Naphthol		
	Nominal	PARAFAC	U-PLS	Nominal	PARAFAC	U-PLS
1	60	44	55	78	86	71
2	214	203	226	415	384	387
3	96	83	91	275	252	239
4	230	223	245	500	465	481
5	251	229	231	40	57	45
6	270	245	250	225	233	213
7	51	44	47	173	158	155
8	30	26	49	6	27	18
9	83	66	71	390	378	386
	RMSE	15	14	RMSE	21	19
	REP%	8.3	7.6	REP%	8.2	7.3

[a] All concentrations in $\mu g\,L^{-1}$.

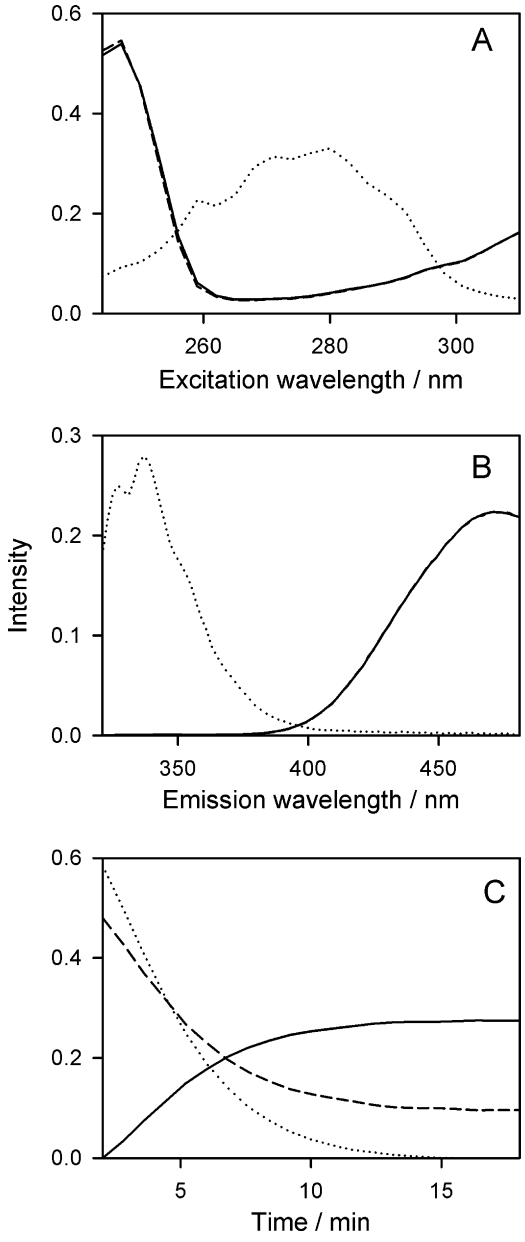

Fig. 4. Emission (A), excitation (B) and kinetic (C) profiles provided by a three-component PARAFAC model (**B**, **C** and **D** matrices, respectively), after processing the array formed by the time evolution of the excitation-emission matrices recorded during carbaryl hydrolysis to produce 1-naphthol at pH 10.2 and 35 C for a sample of set V. The dotted lines in plots (A) and (B) can be ascribed to carbaryl, while the solid and dashed lines correspond to 1-naphthol. The kinetic profiles in plot (C), on the other hand correspond to carbaryl (dotted line), while the solid and dashed lines are linear combinations of the profiles expected for 1-naphthol. The intensities of individual profiles are normalized to unit length.

prediction (REP%), computed with respect to the mean calibration concentration of each analyte.

4.3. Determination of analytes in the validation sample set V: U-PLS

For applying this model to the set of validation samples, it is first required to assess the number of U-PLS latent variables. This can be done by resorting to the well-known leave-one-out cross-validation procedure [24]. In the present case, the number of factors was two for both carbaryl and 1-naphthol, even when the number of PARAFAC components was three, because of the above-mentioned correlations in the 1-naphthol time profiles. Notice that the validation set does not require the RTL procedure for analyte quantitation, because no interferences were added to these samples of set V.

The analytical results are shown in Table 2. As can be seen, the recovery for carbaryl and 1-naphtol were reasonable, and acceptable figures of merit were obtained (RMSE = 14 $\mu g\,L^{-1}$ and REP = 7.6% for carbaryl and RMSE = 19 $\mu g\,L^{-1}$ and REP = 7.3% for 1-naphthol). They are comparable to those achieved using PARAFAC. However, it should be apparent from the previous section that, in comparison with the quantitation based on PARAFAC analysis of the presently studied four-way kinetic-spectral data, the use of U-PLS/RTL is considerably simpler. No special algorithmic modifications, initialization or restrictions are required for successful analyte quantitation, in comparison with other analytical systems not presenting the phenomenon of linear dependency.

Table 3
Comparison of sensitivities (S_n) of U-PLS and PARAFAC for second- and third-order calibrations.[a].

Method/data	Error propagation S_n		Monte Carlo S_n	
	Carbaryl	1-Naphthol	Carbaryl	1-Naphthol
U-PLS/Second-order	2.9	5.9	3.0	6.0
U-PLS/Third-order	6.8	10.1	6.5	9.4
PARAFAC/Second-order	2.9	6.6	3.3	5.9
PARAFAC/Third-order	4.9	5.2	4.9	5.3

[a] Values in AFU L μg^{-1}. AFU = arbitrary fluorescence units.

4.4. Figures of merit

We now compare the relative sensitivities of U-PLS and PARAFAC regarding the present determination using third-order excitation-emission-kinetic data, and also with those which would be obtained by employing second-order excitation-emission data without the time evolution. For U-PLS, Eq. (12) applies to both of these latter cases, except that calibration using second-order data is restricted to using the matrix data for the first reaction time, without resorting to the kinetic evolution of the system. For PARAFAC, an equation analogous to (14) can be employed for second-order data, with the contribution of the **D** matrix removed, and s_n replaced by the corresponding integrated second-order signal at unit concentration. Eq. (15) is used for third-order data.

The obtained results are collected in Table 3. The first issue to be noticed in this table is the satisfactory agreement between error propagation equations and noise addition simulations, which lends support to the present approach, particularly to the novel expression (15). Moreover, U-PLS is seen to provide increased sensitivity for both analytes in going from second- to third-order data, as expected from the measurement of additional matrix data during the kinetic evolution of the system. The relative increase in sensitivity is larger for carbaryl than for naphthol, since the former is converted into the latter by the hydrolysis reaction, and the fluorescence intensity of the reaction product is higher than that of the reagent. Part of the sensitivity for carbaryl is thus gathered from its correlation to the generated naphthol during the reaction.

In the case of PARAFAC, the sensitivities using second-order data are similar to those attained with U-PLS, as expected. An increase in sensitivity is noticed in going to third-order data processing, again as expected. Interestingly enough, however, the relative increase in this case is significantly lower than for U-PLS. It is likely that the MLR calibration required by the strong correlation observed in the component scores after four-way PARAFAC decomposition causes a corresponding decrease in component sensitivity.

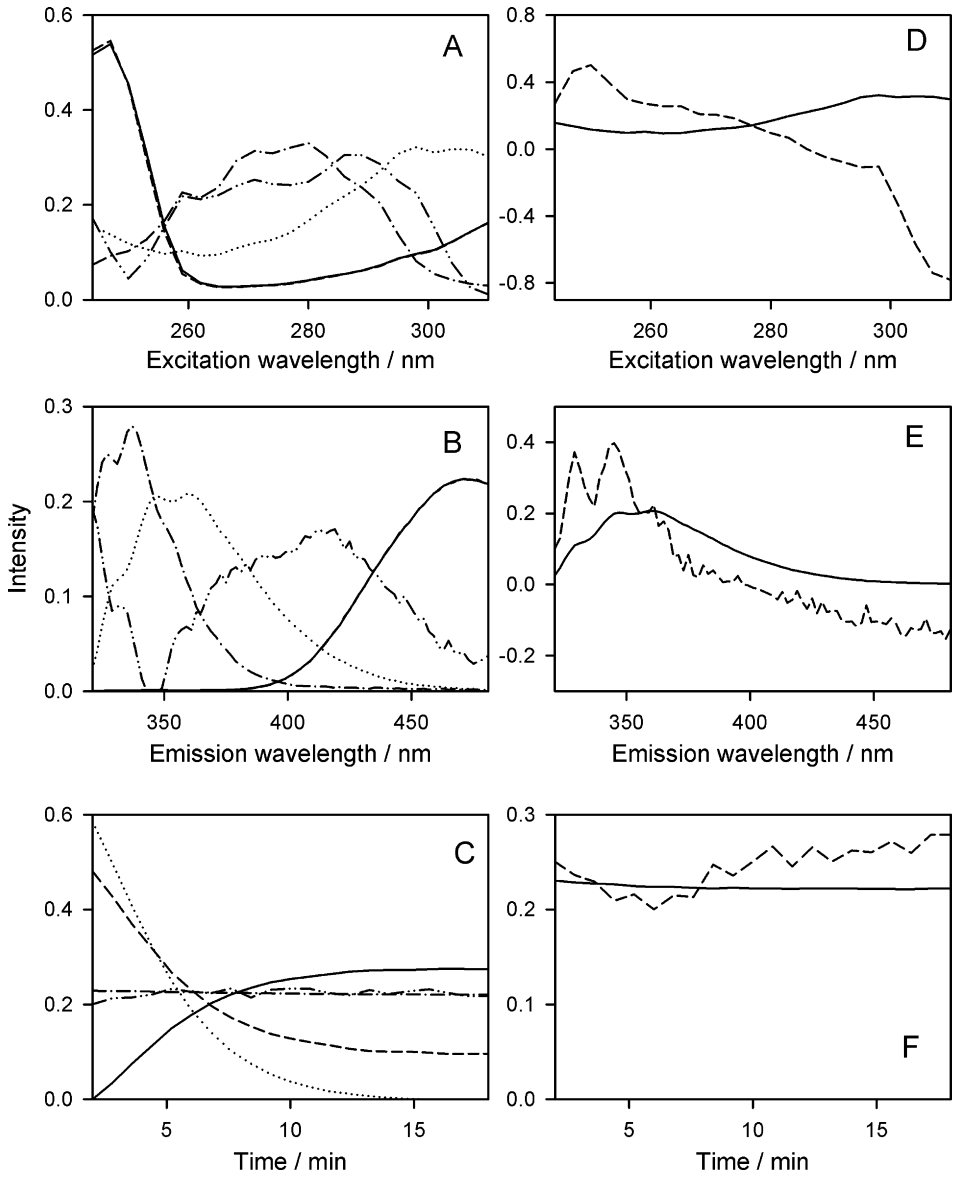

Fig. 5. Emission (A), excitation (B) and kinetic (C) profiles provided by a five-component PARAFAC model (**B**, **C** and **D** matrices, respectively), after processing the array formed by the time evolution of the excitation-emission matrices recorded during carbaryl hydrolysis at pH 10.2 and 35 C for a sample of test S containing both thiabendazole and fuberidazole as potential interferents. The solid, dashed and dotted lines have the same meaning as in Fig. 4, while the dash-dotted and dash-dot-dotted lines correspond to the interferents. Plots (D), (E) and (F) show the corresponding profiles provided for the two interfering components by the U-PLS/RTL model, after processing the same test sample. Solid line, first RTL component, dashed line, second RTL component. The intensities of individual profiles are normalized to unit length.

4.5. Determination of analytes in samples of sets S and T: PARAFAC

All samples of set S contain interfering components which produce fluorescence in the working spectral ranges. These samples intend to mimic a real situation, where unknown responsive components may occur, making it necessary to resort to the second-order advantage in order to quantitate the analytes. The samples of set S were prepared from a natural water sample, by spiking them with the analytes and also with mixtures of other two pesticides: fuberidazole and thiabendazole. These latter compounds are fluorescent in the working spectral ranges, and therefore strongly interfere by overlapping with the analyte signals. Several other pesticides such as imacloprid, imazalil, 2,4-dichlorophenoxyacetic acid, carbendazim, neburon, linuron, diuron, chlorsulfuron, isopropturon, mercaptodimethopur, metomil and propoxur were assayed as potential interferents, but they were not fluorescent under the present experimental conditions. The fact that we know the chemical identity and spectral properties of the interferents in the set of samples S allow us to check the success of the second-order advantage in retrieving correct profiles in the several data dimensions.

PARAFAC decomposition of the four-way arrays formed by each of the samples of set S with the three third-order arrays corresponding to the set of calibration samples proceeded as described above for set V. Similar initialization and restrictions were applied during the fitting phase, and the number of components was established using the same philosophy based on SSE considerations. As expected, five components were required in set S in order to reach stabilization of the SSE parameter for samples containing both interferents, and four components for those containing a single interfering constituent, in agreement with the known composition of these samples (three components were required by the calibrated components, as discussed above for set V). After PARAFAC analysis of a typical sample of set S containing both interferents, the retrieved profiles are shown in Fig. 5A–C in the three dimensions.

Table 4
Prediction results for the spiked test samples of sets S and T.[a]

Sample set S[b]							
Carbaryl			1-Naphthol			Potential interferents	
Nominal	PARAFAC	U-PLS/RTL	Nominal	PARAFAC	U-PLS/RTL	THI	FUB
59	50	31	394	507	394	212	0
246	233	244	39	88	44	0	4
60	55	52	263	330	259	212	4
207	177	185	78	62	88	106	4
96	84	85	423	401	381	0	5
236	201	211	286	253	266	212	8
RMSE	21	19	RMSE	60	19		
REP%	11.3	10.3	REP%	23.4	7.6		
Sample set T							
Carbaryl			1-Naphthol			Potential interferents	
Nominal	PARAFAC	U-PLS/RTL	Nominal	PARAFAC	U-PLS/RTL		
0	–[c]	–[c]	400	383	365	Unknown	
150	152	146	300	289	299	Unknown	
225	210	200	200	191	194	Unknown	
301	332	327	0	–[c]	–[c]	Unknown	

[a] All concentrations in $\mu g\,L^{-1}$.
[b] THI = thiabendazole, FUB = fuberidazole.
[c] Not detected (small negative concentrations were predicted).

Three of them match those shown in Fig. 4, while the remaining ones correspond to the interferences. Notice in Fig. 5C that the time profiles for the interferents are constant, as expected from the fact that they do not react in alkaline media. Finally, the time profiles for 1-naphthol show the same linear correlations discussed above in connection with the sample set V.

Quantitation with PARAFAC is possible in all samples of this set by employing an analogous MLR calibration approach to that discussed above for the validation samples. The scheme of Fig. 1 was employed, in this case by removing from matrix **A** the columns corresponding to the unexpected components. The results for both analytes are shown in Table 4. While acceptable results were obtained for the recovery of carbaryl, the corresponding figures for 1-naphthol appear to be rather disappointing.

The sample set T, on the other hand, was prepared from a natural water carrying an unknown fluorescent background, generated by an unspecified number of uncalibrated sample components. In this set the real limits of the second-order advantage can be tested. The analysis of these samples followed similar guidelines to those detailed above concerning the set S. The total number of components was estimated as four, with retrieved spectral and time profile which are shown in Fig. 6A–C. The additional component which is different than the analytes accounts for the contributing signal from the uncalibrated sample background. Specific prediction results are shown in Table 4, which can be studied by paired *t*-statistics. Since $t_{exp} = 0.8$, lower than $t_{crit} = 2.58$ (95% confidence level and 5 degrees of freedom), the results provided by PARAFAC are statistically comparable to the nominal ones.

4.6. Determination of analytes in samples of sets S and T: U-PLS/RTL

The sample set S containing potential interferents was studied using the U-PLS/RTL model. In the present case it is necessary to exploit the second-order advantage of the four-way data which is provided by the RTL procedure. Hence, for each of these test samples, the number of RTL components should be tuned, in addition to the number of calibration latent variables, which is the same as that employed during analysis of the validation samples of set V. The required number of RTL components was established as two in all samples of set S containing both interferents and one in those containing a single interferent. This conclusion was reached by monitoring the changes in the residual parameter s_u [Eq. (11)] as a function of increasing RTL components. The quantitation of the analytes proceeded as with the V samples, i.e., with no special conditions imposed during the calculations. Fig. 5D–F shows the RTL profiles obtained when studying with U-PLS/RTL the same sample of set S discussed above. As can be seen, the profiles for one of the RTL components resembles those corresponding to one of the interferents (Fig. 5), but the remaining profile is an abstract linear combination of pure spectra with no physical interpretation.

The prediction results for these S samples are shown in Table 4. The recovery of 1-naphthol concentrations shows better predictive ability than the PARAFAC model, probably due to: (1) better analyte resolution by using latent variables, and (2) handling multilinearity losses due to the finite time during which EEM acquisition takes place.

As regards the sample set T, analysis proceeded as described above for set S. A single interfering RBL component was estimated for these samples, allowing to achieve the second-order advantage. This led to retrieved RTL profiles which are shown in Fig. 6D–F, and are seen to be in agreement with the PARAFAC findings for these samples. Quantitative results are collected in Table 4. Paired *t*-statistics was also applied: the experimental t_{exp} was 0.4, lower than the critical $t_{crit} = 2.58$ (95% confidence level and 5 degrees of freedom), confirming that the results provided by U-PLS/RTL are statistically comparable to the nominal values. As can be seen, the recoveries are of a similar quality to those obtained for the set S, confirming the predictive ability of U-PLS/RTL towards these complex samples having a responsive background signal of unknown origin.

4.7. Comparison of PARAFAC and U-PLS/RTL

A comparison of the predictive abilities of PARAFAC and U-PLS/RTL regarding the most challenging set of samples (set T) was finally made, using similar arguments to those employed above in connection with the comparison of predicted vs. nominal concentration values. The results suggest that no significant difference exists between both sets of predictions: paired *t*-statistics gives $t_{exp} = 1.0$, lower than the critical $t_{crit} = 2.58$ (95% confidence level and 5 degrees of freedom).

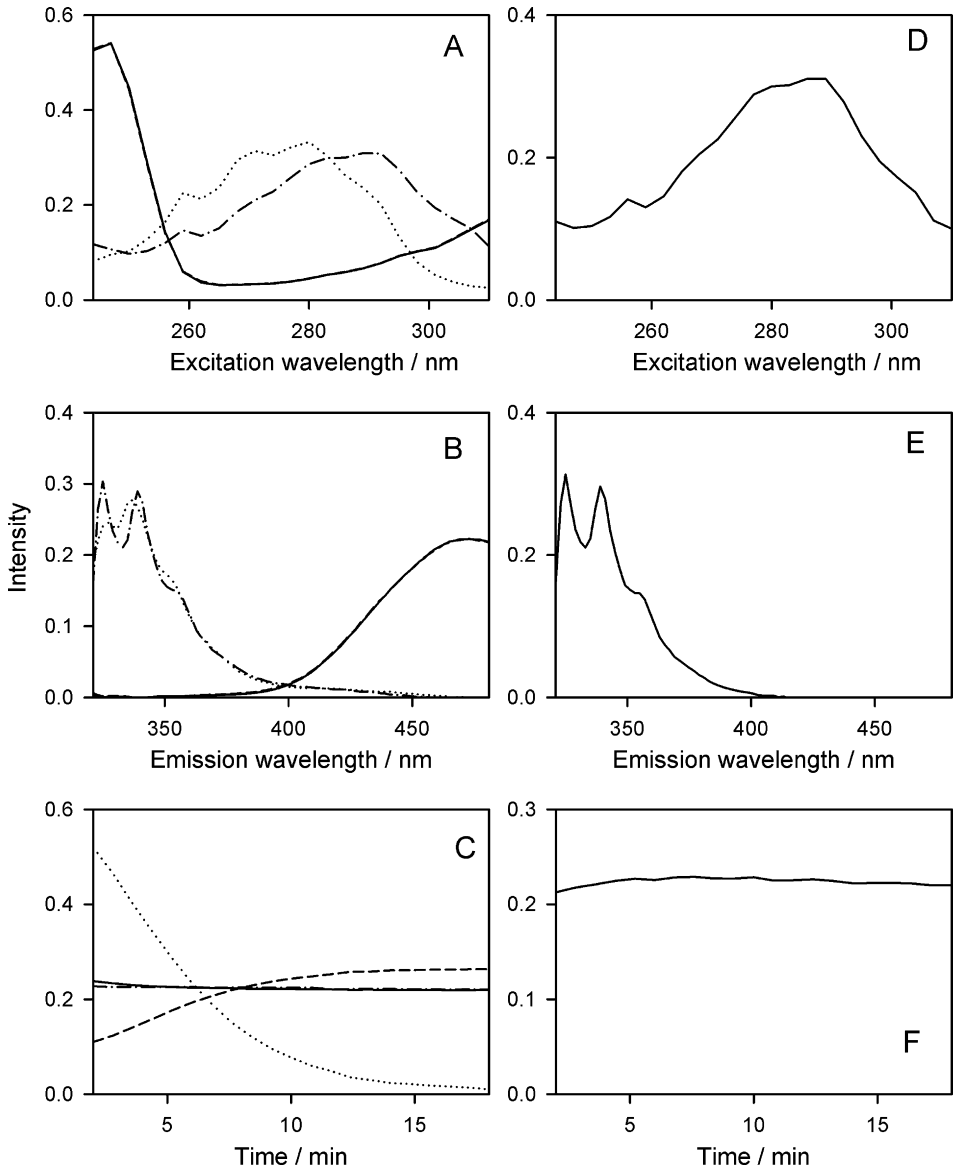

Fig. 6. Emission (A), excitation (B) and kinetic (C) profiles provided by a four-component PARAFAC model (**B**, **C** and **D** matrices, respectively), after processing the array formed by the time evolution of the excitation-emission matrices recorded during carbaryl hydrolysis at pH 10.2 and 35 C for a sample of test T containing an unknown sample background as potential interferent. The solid, dashed and dotted lines have the same meaning as in Fig. 4, while the dash-dotted lines correspond to the interferents. Plots (D), (E) and (F) show the corresponding profiles provided for the interfering component by the U-PLS/RTL model, after processing the same test sample. The intensities of individual profiles are normalized to unit length.

5. Conclusions

The present study reports on a new application of four-way kinetic-excitation-emission fluorescence data, aimed at the quantitative determination of two interconverting analytes embedded in a the complex sample background. The approach based on unfolded partial least-squares combined with residual trilinearization provides simple, accurate and reproducible quantitative analysis for the determination of carbaryl and 1-naphthol as binary mixtures without interference from other assayed pesticides. The proposed method is simple as there is no need for solvent extraction, and has the advantages of being low cost, rapid and environmentally friendly. In addition, the applied algorithm presents similar figures of merit in comparison to classical parallel factor analysis, but is considerably simpler in its computer implementation.

Acknowledgements

The following institutions are gratefully acknowledged for financial support: Universidad Nacional de Rosario, CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas, Project No. PIP 1950) and ANPCyT (Agencia Nacional de Promoción Científica y Tecnológica, Project No. PICT 25825). R.M.M. thanks CONICET for his post-doctoral fellowship.

References

[1] R. Bro, Crit. Rev. Anal. Chem. 36 (2006) 279–293.
[2] G.M. Escandar, N.M. Faber, H.C. Goicoechea, A. Muñoz de la Peña, A.C. Olivieri, R.J. Poppi, Trends Anal. Chem. 26 (2007) 752–765.
[3] R. Bro, Anal. Chim. Acta 500 (2003) 185–194.
[4] A.C. Olivieri, Anal. Chem. 80 (2008) 5713–5720.

[5] H.C. Goicoechea, S. Yu, A.C. Olivieri, A.D. Campiglia, Anal. Chem. 77 (2005) 2608–2616.
[6] R.P.H. Nikolajsen, K.S. Booksh, A.M. Hansen, R. Bro, Anal. Chim. Acta 475 (2003) 137–150.
[7] A.C. Olivieri, J.A. Arancibia, A. Muñoz de la Peña, I. Durán Merás, A. Espinosa Mansilla, Anal. Chem. 76 (2004) 5657–5666.
[8] J.A. Arancibia, A.C. Olivieri, D. Bohoyo Gil, A. Espinosa Mansilla, I. Durán-Merás, A. Muñoz de la Peña, Chem. Intell. Lab. Syst. 80 (2006) 77–86.
[9] A. Muñoz de la Peña, I. Durán Merás, A. Jiménez Girón, Anal. Bioanal. Chem. 385 (2006) 1289–1297.
[10] A. Muñoz de la Peña, I. Durán Merás, A. Jiménez Girón, H.C. Goicoechea, Talanta 72 (2007) 1261–1268.
[11] P.C. Damiani, I. Durán-Merás, A. García Reiriz, A. Jiménez Girón, A. Muñoz de la Peña, A.C. Olivieri, Anal. Chem. 79 (2007) 6949–6958.
[12] S.-H. Zhu, H.-L. Wu, A.-L. Xia, J.-F. Nie, Y.-C. Bian, C.-B. Cai, R.-Q. Yu, Talanta 77 (2009) 1640–1646.
[13] A. García-Reiriz, P.C. Damiani, A.C. Olivieri, F. Cañada-Cañada, A. Muñoz de la Peña, Anal. Chem. 80 (2008) 7248–7256.
[14] M.L. Nahorniak, G.A. Cooper, Y.C. Kim, K.S. Booksh, Analyst 130 (2005) 85–93.
[15] Y.C. Kim, J.A. Jordan, M.L. Nahorniak, K.S. Booksh, Anal. Chem. 77 (2005) 7679–7686.
[16] A. Jiménez Girón, I. Durán-Merás, A. Espinosa-Mansilla, A. Muñoz de la Peña, F. Cañada Cañada, A.C. Olivieri, Anal. Chim. Acta 622 (2008) 94–103.
[17] K.S. Booksh, B.R. Kowalski, Anal. Chem. 66 (1994) 782A–791A.
[18] S.-H. Zhu, H.-L. Wu, A.-L. Xia, Q.-J. Han, Y. Zhang, R.-Q. Yu, Talanta 74 (2008) 1579–1585.
[19] R. Bro, Chemometr. Intell. Lab. Syst. 38 (1997) 149–171.
[20] P. Paatero, Chemometr. Intell. Lab. Syst. 38 (1997) 223–242.
[21] R. Bro, H.A.L. Kiers, J. Chemometr. 17 (2003) 274–286.
[22] M.D. Borraccetti, P.C. Damiani, A.C. Olivieri, Analyst 134 (2009) 1682–1691.
[23] S. Wold, P. Geladi, K. Esbensen, J. Öhman, J. Chemometr. 1 (1987) 41–56.
[24] D.M. Haaland, E.V. Thomas, Anal. Chem. 60 (1988) 1193–1202.
[25] A.C. Olivieri, N.M. Faber, J. Chemometr. 19 (2005) 583–592.
[26] A.C. Olivieri, Anal. Chem. 77 (2005) 4936–4946.
[27] J.N. Miller, J.C. Miller, Statistics and Chemometrics for Analytical Chemistry, Pearson-Prentice Hall, New York, 2005 (Chapter 2).
[28] MATLAB 7.0, The MathWorks Inc., Natick, MA, 2000.
[29] A.C. Olivieri, R.-Q. Yu, H.-L. Wu, Chemometr. Intell. Lab. Syst. 96 (2009) 246–251.