Analytica Chimica Acta 859 (2015) 20–28

Contents lists available at ScienceDirect

Analytica Chimica Acta

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

Modeling nonbilinear total synchronous fluorescence data matrices with a novel adapted partial least squares method

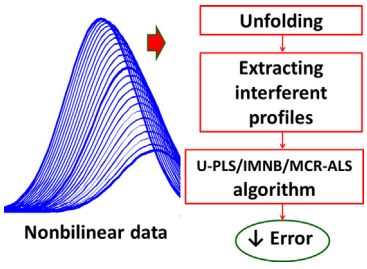
Agustina V. Schenone [a], Adriano de Araújo Gomes [b], María J. Culzoni [a], Andrés D. Campiglia [c], Mário Cesar Ugulino de Araújo [b], Héctor C. Goicoechea [a,*]

[a] Laboratorio de Desarrollo Analítico y Quimiometría (LADAQ), Cátedra de Química Analítica I, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral – CONICET, Ciudad Universitaria, Santa Fe S3000ZAA, Argentina
[b] Laboratório de Automação e Instrumentação em Química Analítica e Quimiometria (LAQA) Universidade Federal da Paraíba, CCEN, Departamento de Química, Caixa Postal 5093, CEP 58051-970, João Pessoa, PB, Brazil
[c] Department of Chemistry, University of Central Florida, Orlando, FL 32816, USA

HIGHLIGHTS

- A new residual modeling for nonbilinear data is presented.
- The problem is addressed by extracting the interference profile with MCR-ALS.
- Total synchronous fluorescence spectroscopy are conveniently modeled.
- The novel algorithm perform better than known second-order algorithms.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 4 October 2014
Received in revised form 3 December 2014
Accepted 5 December 2014
Available online 10 December 2014

Keywords:
Second-order advantage
Synchronous fluorescence
Residual modeling
Ciprofloxacin

ABSTRACT

A new residual modeling algorithm for nonbilinear data is presented, namely unfolded partial least squares with interference modeling of non bilinear data by multivariate curve resolution by alternating least squares (U-PLS/IMNB/MCR-ALS). Nonbilinearity represents a challenging data structure problem to achieve analyte quantitation from second-order data in the presence of uncalibrated components. Total synchronous fluorescence spectroscopy (TSFS) generates matrices which constitute a typical example of this kind of data. Although the nonbilinear profile of the interferent can be achieved by modeling TSFS data with unfolded partial least squares with residual bilinearization (U-PLS/RBL), an extremely large number of RBL factors has to be considered. Simulated data show that the new model can conveniently handle the studied analytical problem with better performance than PARAFAC, U-PLS/RBL and MCR-ALS, the latter modeling the unfolded data. Besides, one example involving TSFS real matrices illustrates the ability of the new method to handle experimental data, which consists in the determination of ciprofloxacin in the presence of norfloxacin as interferent in water samples.

* Corresponding author. Tel.: +54 3424575206x190.
E-mail address: hgoico@fbcb.unl.edu.ar (H.C. Goicoechea).

http://dx.doi.org/10.1016/j.aca.2014.12.014
0003-2670/

1. Introduction

Nowadays, there is a widespread acceptance in the analytical community of the growing need for the study of increasingly complex samples by improving the available analytical methods. In the field of chemometrics, this fact is linked to the intensive research of new multi-way data generation and its consequent modeling with suitable algorithms [1–3]. Multi-way instrumental data carry the intrinsic potential to achieve the second-order advantage, which in principle permits analyte quantitation in samples containing unexpected constituents (i.e., sample concomitants not included in the calibration set). This property is especially interesting in the field of complex sample analysis owing to the fact that it allows building calibration models with a limited number of standards, maintaining quality of analyte prediction ability regardless of the presence of interferences, and without the need of previous chromatographic or electrophoretic separation.

A literature search on the available second-order algorithms capable of coping with the second-order advantage reveals an interesting fact: most of them call for the property of trilinearity. According to Olivieri et al. [1], the following requirements are necessary to fulfill the multilinearity property: (I) constancy of profiles across the different samples for each component; (II) signal linearly related to the analyte concentration; and, (III) bilinear signal for a given sample. The property of bilinearity implies that a single component data matrix can be decomposed into the product of two vectors, each containing the component profile in one of the two data dimensions (a well-known example of this kind of data is an excitation–emission fluorescence matrix, EEM). This can be condensed in the concept of rank of a matrix, i.e., the number of bilinear terms needed to reproduce the data matrix. If the rank is close to the number of responsive constituents, the matrix is bilinear. When the three-way array built with standards and test data matrices follow the trilinear structure, parallel factor analysis (PARAFAC) [4] seems to be the appropriate algorithm, because of its inner trilinear structure. A three-way array can then be built with a set of second-order **X** data matrices. Each data element in the array can be modeled by the following equation:

$$x_{ijk} = \sum_{n=1}^{N} a_{in} b_{jn} c_{kn} + e_{ijk} \quad (1)$$

where N is the total number of chemical constituents generating the measured signal, a_{in} is the relative concentration or score of component n in the i-th sample, and b_{jn} and c_{kn} are the intensities in the instrumental channels (or data dimensions) J and K, respectively. The values of e_{ijk} are the elements of the three-dimensional array **E**, representing the residual error, and having the same dimensions as **X**. Summation in Eq. (1) implies that the individual signals of the matrix constituents are additive. Usually the loadings are normalized to unit length, and collected into the loading matrices **B** and **C**, of size $J \times N$ and $K \times N$ respectively. The scores are collected into the score matrix **A** (size $I \times N$), which reflects the relative concentration values of the various constituents in all samples. This information is used in the context of calibration to build a univariate plot. This plot allows the prediction of analyte concentrations in unknown samples by projecting its score onto the fitted line.

Very recently, Olivieri and Escandar discussed the different kind of second-order data that can be gathered in the lab; their characteristics and the most convenient algorithm to be employed in each case [5]. Especial attention is paid on what they called “the Cinderella type 3 data”, nontrilinear data which involve nonbilinear data matrices, i.e., different spectral profiles for a component in a single sample. This is one of the most difficult problems to be solved. A reduced number of articles have been published dealing with nonbilinear data from single sample components [6–11] generated by (1) tandem mass-spectrometry (MS^2) [6], (2) total synchronous fluorescence spectroscopy matrix (TSFSM) [7–9], (3) ultrafast time-resolved spectroscopy [10], and (4) two-dimensional nuclear magnetic resonance spectroscopy [11]. TSFSM data sets are not bilinear due to the fact that synchronous spectral profiles vary with the offset between excitation and emission wavelengths [7–15]. MS^2 data sets are not bilinear because each fragment of a single compound has a specific MS pattern in the second MS dimension, making it impossible to describe the MS^2 data in terms of one MS profile in each dimension [6].

To the best of our knowledge, there are only two publications, which have been reported by our group, exploiting the second-order advantage on TSFSM data for quantitation purposes [8,9]. Essentially, synchronous spectroscopy consists in simultaneously scanning both the excitation and the emission monochromators while maintaining a constant wavelength interval ($\Delta\lambda$) between them. The judicious choice of Δl values along with the reduction of excitation and emission bandwidths often provides the spectral simplification needed to overcome severe overlapping in the analysis of multi-component samples.

Spectral simplification is always possible under proper instrumental parameters [16]. Unfortunately, in some cases synchronous fluorescence spectroscopy needs to be combined with a multivariate calibration method. This is particularly true for the direct analysis of complex samples of unknown composition having multiple fluorescence species [14]. In addition to resolving spectral overlapping, the chosen algorithm should be able to handle nonbilinear signals. Although U-PLS/RBL has been proposed for this task [5], the RBL procedure only works properly when data from unexpected components are bilinear. This is due to the fact that RBL models the signals of unexpected components via singular value decomposition (SVD) [17]. Therefore, the number of components needed to model the data may be extremely large, precluding the possibility of modeling the full variability of the data by only using a few of the first components. Thus, RBL will probably fail to model the matrix, not because of inability to model a nonbilinear matrix, but because of the fact that the large series of bilinear components that would be required to do so is truncated in some arbitrary way, i.e., if fifteen RBL components are necessary to model the variability of one interferent, this number should be truncated because it is not practical, or even feasible, especially when more than one interferent should be considered.

In an earlier work [8], the second-order advantage was gained from TSFSM data modeling with both U-PLS/RBL and MCR-ALS, but in the latter case the second-order data were previously unfolded assuming that the unfolded-TSFSM data set have a bilinear structure [12], i.e., the strategy consisted in modeling first-order responses with MCR-ALS, and exploiting the potentiality of the correlation constraint [18].

Herein, we present a novel method that comprises nonbilinear partial least squares followed by the modeling based on MCR-ALS with the correlation constraint. During the interferent modeling step, profiles corresponding to interfering species are extracted by MCR-ALS processing of unfolded data and incorporated into a step which updates the PLS scores in order to minimize the residual of the model. The new algorithm has been named unfolded partial least squares with interference modeling of non bilinear data by multivariate curve resolution by alternating least squares (U-PLS/IMNB/MCR-ALS). Both simulated and experimental data sets are used to compare the prediction ability of the new method to conventional U-PLS/RBL, PARAFAC and the previously mentioned unfolded MCR-ALS.

2. Experimental

2.1. Simulations

Simulations of nonbilinear matrices were carried out in order to illustrate the behavior of the new second-order calibration algorithm U-PLS/IMNB/MCR-ALS and to compare its performance with other algorithms: PARAFAC (which requires the bilinearity fulfillment), U-PLS/RBL (which needs to add a large number of factors in order to succeed) and unfolded MCR-ALS (which can fail if the system is highly complex). The simulated system considered the presence of a single analyte in the calibration samples, and the analyte and one concomitant in the test samples. As such, its proper resolution requires adherence to the second-order advantage. Nonbilinear matrices were generated from the noiseless excitation and emission profiles shown in Fig. 1.

The same spectral profiles were used to generate an excitation–emission matrix for each sample component. Fig. 2 shows the contour plot of an EEM corresponding to a mixture of both components. Data matrices built in this way had a dimension of 100×100. Every row in the box inside the EEM contour plot corresponds to a synchronous fluorescence spectrum (diagonal elements of the matrix starting at sensor 1 and ending at sensor 50) hypothetically recorded at a given $\Delta\lambda$ value.

A calibration set of five matrices only containing the analyte at nominal concentrations of 1.0, 2.0, 3.0, 4.0 and 5.0 (in arbitrary units, a.u.) was built. The size of every matrix was 25×50 (50 data points corresponding to the spectral dimension, and 25 to the offset dimension). One set of 9 test samples was created containing both analyte and interferent. The concentrations were set according to Table 1. Significant amounts of the interferent were established using a central composite design.

Fig. 3 shows the simulated nonbilinear data obtained for the two compounds at each one of the 25 wavelength offsets. It should be noted that although the way of obtaining the simulated data cannot be exactly the same as when real spectra are recorded; the simulated spectra trend is the same as that observed for experimental data.

A further approximation to a real experimental data set was made by adding noise to all simulated spectra from a Gaussian distribution having a standard deviation equal to 1% of the maximum calibration signal. The final values were employed to build the calibration and validation data sets. The second-order nonbilinear data for each of the 9 test samples were then combined to those from the calibration set, and each sixteen-sample data set (five standards in triplicate plus the test sample) was submitted to second-order calibration with PARAFAC, U-PLS/RBL, unfolded MCR-ALS and the novel U-PLS/IMNB/MCR-ALS. Analyte predictions were stored for statistical analysis and future comparison.

2.2. Experimental system

The experimental system consisted in the determination of ciprofloxacin (CPF) in water samples in the presence of norfloxacin (NOR), which was used as the potential interferent due to its common occurrence and the strong spectral overlapping with CPF.

2.2.1. Chemicals and reagents

Analytical reagent-grade chemicals and ultrapure water, obtained from a Milli-Q water purification system from Millipore (USA), were used throughout the experiments. Ciprofloxacin (CPF) and norfloxacin (NOR) were obtained from Fluka (Switzerland). Lichrosolv® HPLC grade methanol and glacial acetic acid were purchased from Merk (Germany). Sodium dodecyl sulfate (SDS), tris-[hydroxymethyl] aminomethan (Tris) and yttrium nitrate were obtained from Cicarelli (Argentina), Applichem (Germany) and Sigma (USA), respectively.

Stock solutions of CPF ($300.0\,mg\,L^{-1}$) and NOR ($250.0\,mg\,L^{-1}$) were prepared by dissolving the appropriate amounts of each fluoroquinolone in methanol. Solutions were stored in the dark at 4 °C. Working standards were freshly prepared by diluting stock solutions to the desired concentrations with Milli-Q water. Tris buffer $0.2\,mol\,L^{-1}$ was prepared in Milli-Q water and adjusted to pH 4 with glacial acetic acid and a pH meter Hanna Instrument (USA). SDS $0.3\,mol\,L^{-1}$ and yttrium $0.02\,mol\,L^{-1}$ solutions were prepared by appropriated dilution of the solid drugs in Milli-Q water.

2.2.2. Calibration and validation samples

Five standard calibration solutions of the analyte (two replicates) were prepared in the concentration range from 100 to $500\,ng\,mL^{-1}$. In addition, five standard solutions of NOR

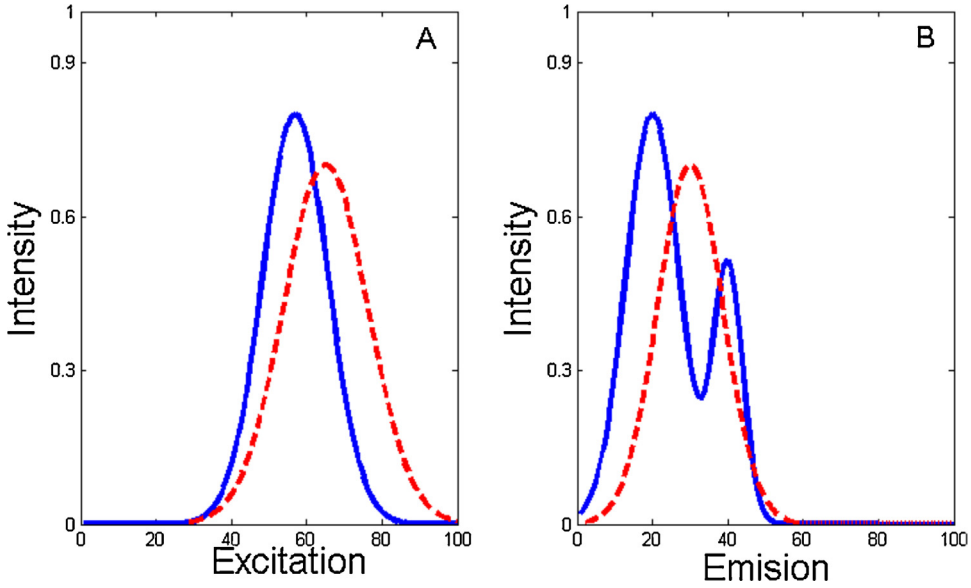


Fig. 1. Simulated excitation (A) and emission (B) spectra of analyte (blue solid line) and interferent (dashed red lines). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

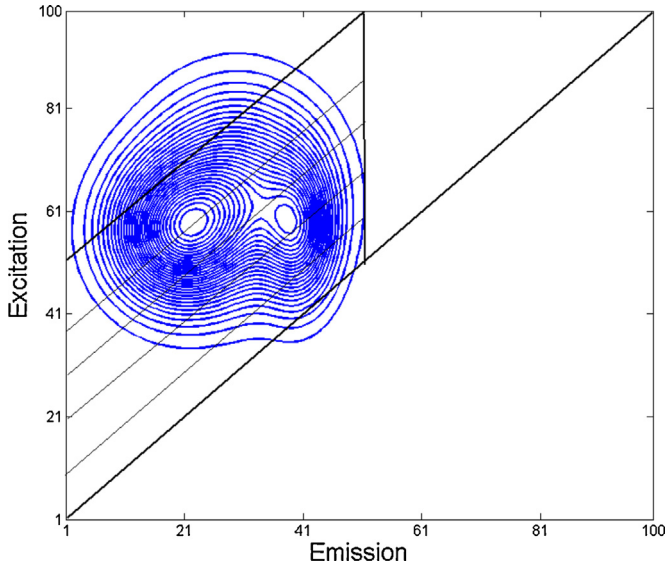


Fig. 2. Simulated excitation–emission matrix for a mixture containing the analyte and a single interferent. The rows inside the box contain the simulated synchronous spectra at different wavelength offsets.

in the concentration range from 200 to 600 ng mL^{-1} were prepared. The validation set consisted of seven samples prepared in concentrations ranging from 150 to 450 ng mL^{-1} and 250 to 550 ng mL^{-1}, for CPF and NOR, respectively.

Calibration and validation samples were prepared by measuring appropriate aliquots of the standard solutions of each quinolone, placing them into 2.00 mL volumetric flasks to obtain the desired concentrations, adding 20.0 μL of yttrium 0.02 mol L^{-1}, 200.0 μL of Tris buffer 0.2 mol L^{-1} and 100.0 μL of SDS solution 0.3 mol L^{-1} and completing to the mark with Milli-Q water. All of the samples were prepared in duplicate. The selected concentrations for the studied drugs are within those found in biological fluids and environmental samples [19].

2.2.3. Instrumentation

All measurements were performed with a Varian Cary Eclipse fluorescence spectrophotometer (Agilent Technologies, Germany) equipped with a xenon flash lamp. Excitation and emission bandpass were kept at 5 nm and the detector's voltage was 600 V. The fluorescence characteristics of CPF and NOR were firstly investigated by recording excitation and emission spectra from 262 to 358 nm and 391 to 509 nm, respectively. Data points were collected at 4 nm steps for excitation and 0.5 nm steps for emission. Hence, the size of each excitation–emission data matrix (EEM) was 237 × 25. Synchronous fluorescence spectra were recorded upon sample excitation between 250.9 and 358 nm, each 4 nm. Synchronous offsets varied from 80 to 200 nm at 5 nm wavelength intervals. The resulting size of each TSFS data matrix was 217 × 25.

2.3. Software for chemometric analysis

All calculations were done using MATLAB 7.10 [20]. PARAFAC and U-PLS/RBL were applied with the MVC2 graphical user interface written in MATLAB by Olivieri et al. [21] and available on the Internet [22]. MCR-ALS was implemented using the concentration correlation constraint developed by R. Tauler [18]. All performed simulations and the new U-PLS/IMNB/MCR-ALS method were implemented using in-house routines.

3. Theory

3.1. The U-PLS/RBL model

The first calibration step of this algorithm employs concentration information solely from known standards and excludes information from the unknown sample [23]. The I calibration data matrices $\mathbf{X}_{c,i}$ (size $J \times K$, where J and K are the number of channels in each dimension) are vectorized (unfolded) and used to calibrate–along with the vector of calibration concentrations $\mathbf{y}$ ($I \times 1$)–a usual U-PLS model. This calibration provides a set of loadings $\mathbf{P}$ and weight loadings $\mathbf{W}$ (both of size $JK \times A$, where A is the number of latent factors), as well as regression coefficients $\mathbf{v}$ (size $A \times 1$). Techniques such as leave-one-out cross-validation allow for the acquisition of the A parameter [24]. In the absence of unexpected agents in the test sample, $\mathbf{v}$ can be employed to estimate the analyte concentration via the following equation:

$$y_u = \mathbf{t}_u^T \mathbf{v} \tag{2}$$

where $\mathbf{t}_u$ is the test sample score obtained by projection of the (unfolded) data for the test sample ($\mathbf{x}_u$) onto the space of the A latent factors:

$$\mathbf{t}_u = (\mathbf{W}^T\mathbf{P})^{-1}\mathbf{W}^T\mathbf{X}_u \tag{3}$$

When unexpected constituents occur in $\mathbf{X}_u$, the sample scores given by Eq. (3) are not suitable for analyte prediction with Eq. (2).

Table 1
Results obtained when modeling simulated data with PARAFAC, U-PLS/RBL, MCR-ALS and U-PLS/IMNB/MCR-ALS

Sample	Analyte nominal concentration (arbitrary unity)	Interferent nominal concentration	Predicted concentration (arbitrary unity)			
			PARAFAC[a]	U-PLS/RBL[b]	MCR-ALS[c]	U-PLS/IMNB/MCR-ALS[d]
V1–1	1.59	1.59	3.31	1.74	1.84	1.60
V1–2	4.41	1.59	6.21	4.56	4.69	4.45
V1–3	3.00	3.00	6.45	3.30	3.47	3.27
V1–4	1.00	3.00	4.38	1.28	1.46	0.89
V1–5	4.41	4.41	9.61	4.84	5.01	4.71
V1–6	5.00	3.00	8.45	5.30	5.00	4.98
V1–7	1.59	4.41	6.76	1.76	2.25	1.63
V1–8	3.00	1.00	4.11	3.10	3.18	3.10
V1–9	3.00	5.00	8.93	3.26	3.75	3.06
REP%	–	–	**152.9**	**10.3**	**18.7**	**5.9**

[a] The number of responsive components was set by analyzing the residuals of the least-squares fit of the three-way data array to the trilinear model ($A = 2$).
[b] The number of interferents was set on 3 for all the samples as a compromise value.
[c] Two components were considered.
[d] One factor extracted by MCR-ALS (see Eq. (9) text).

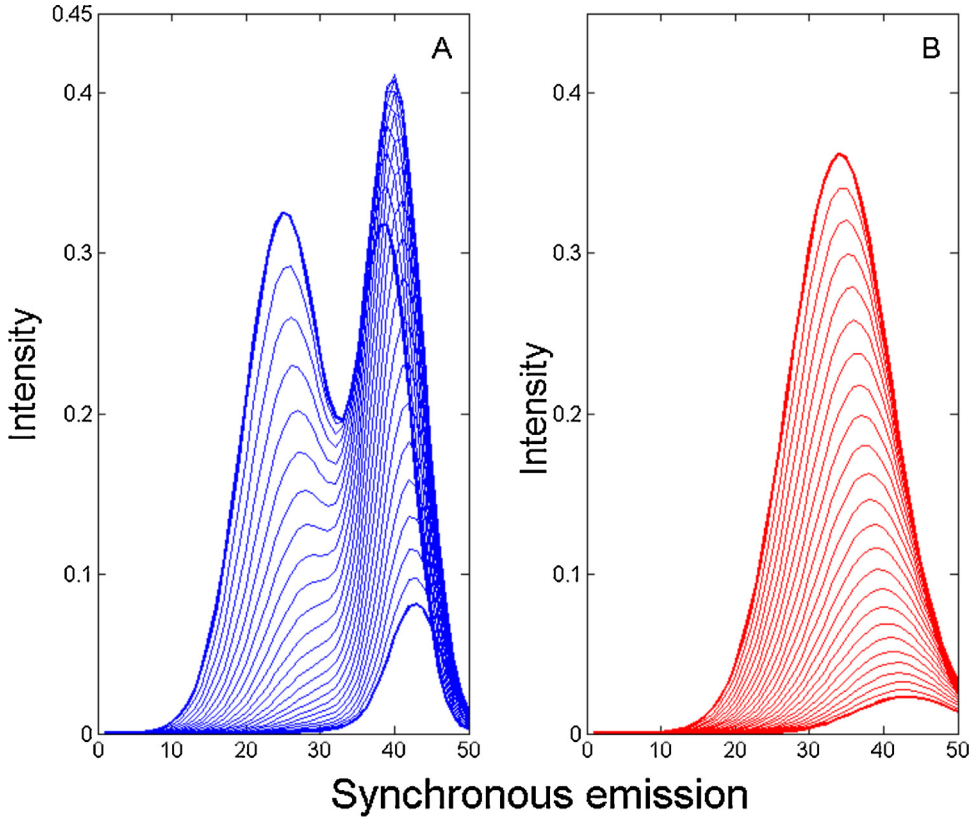


Fig. 3. Simulated synchronous spectra of analyte (A) and inteferent (B) generated at 25 wavelength offsets.

In the presence of unexpected constituents, the residuals of the U-PLS prediction step [s_p, see Eq. (4)] become abnormally large in comparison to the instrumental noise:

$$S_p = \frac{\| \mathbf{e}_p \|}{\sqrt{(JK - A)}} = \frac{\| \mathbf{x}_u - \mathbf{P}(\mathbf{W}^T\mathbf{P})^{-1})\mathbf{W}^T\mathbf{x}_u \|}{\sqrt{(JK - A)}} = \frac{\| \mathbf{x}_u - \mathbf{P}\mathbf{t}_u \|}{\sqrt{(JK - A)}} \quad (4)$$

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

The case of unexpected sample constituents can be handled by RBL, an algorithm that models interference effects via SVD. This is achieved by minimizing the norm of the residual vector ($\mathbf{e}_u$) while fitting the sample data to the sum of the relevant contributions to the sample signal. For single component interference, the following equation applies:

$$\mathbf{x}_u = \mathbf{P}\mathbf{t}_{uRBL} + \text{vec}\left[g_{int}\mathbf{b}_{int}(\mathbf{c}_{int})^T\right] + \mathbf{e}_u \quad (5)$$

where $\mathbf{b}_{int}$ and $\mathbf{c}_{int}$ are the left and right eigenvectors of $\mathbf{E}_p$ and g_{int} is a scaling factor obtained during the following SVD procedure:

$$(g_{int}, \mathbf{b}_{int}, \mathbf{c}_{int}) = \text{SVD}_1(\mathbf{E}_p) \quad (6)$$

where $\mathbf{E}_p$ is the $J \times K$ matrix obtained after reshaping the $JK \times 1$ $\mathbf{e}_p$ vector of Eq. (4), and SVD_1 indicates taking the first principal component.

The RBL procedure consist in maintaining the matrix of loadings $\mathbf{P}$ in Eq. (5) constant at the calibration values and $\mathbf{t}_u$ is varied until $\| \mathbf{e}_u \|$ is minimized. During this minimization step, profiles for the unexpected constituents are continually updated through Eq. (5). The minimization can be carried out using a Gauss–Newton (GN) procedure starting with $\mathbf{t}_u$ from Eq. (4). Once $\| \mathbf{e}_u \|$ is minimized in Eq. (5), the analyte concentrations are provided by Eq. (2), by introducing the final $\mathbf{t}_{uRBL}$ vector found by the RBL procedure.

The number of unexpected constituents N_{unx} can be assessed by comparing the final residuals s_u with the instrumental noise level:

$$S_u = \frac{\| \mathbf{e}_u \|}{\sqrt{JK - N_c - N_{unx}}} \quad (7)$$

where $\mathbf{e}_u$ is from Eq. (5). A plot of s_u versus the trial number of components will show a decreasing trend. The highest s_p value in the plot corresponds to a number of components equal to A, i.e. the number of latent variables used to describe the calibration data. The correct number of components is then graphically obtained from a stabilized s_u value that is compatible with the experimental noise.

3.2. The U-PLS/IMNB/MCR-ALS model

The new procedure described herein (IMNB/MCR-ALS) overcomes the RBL difficulty to handle the nonbilinear structure of data having a reduced number of components. The basic idea of IMNB/MCR-ALS is similar to that of the classical RBL method, which is to minimize the norm of the residual vector $\mathbf{e}_u$ computed by fitting the relevant contributions of the calibration data set and the contribution of interfering concomitants (see Eq. (5)). The main difference is the use of MCR-ALS. This algorithm is implemented to extract a vector profile for each interferent from an augmented matrix $\mathbf{D}$ [of size $(I+1) \times JK$] built with both the I unfolded calibration matrices $\mathbf{x}_{c,i}{}^T$ (of size $1 \times JK$) and the unfolded unknown test sample $\mathbf{x}_u{}^T$ ($1 \times JK$):

$$\mathbf{D} = \left[\mathbf{x}_u{}^T | \mathbf{x}_{c,1}{}^T | | \mathbf{x}_{c,2}{}^T | \ldots | \mathbf{x}_{c,I}{}^T\right] \quad (8)$$

The augmented matrix $\mathbf{D}$ is decomposed by MCR (see below restrictions that need to be applied):

$$D = \mathbf{C}\mathbf{S}^T + \mathbf{E}_{MCR} \quad (9)$$

where $\mathbf{C}$ [of size $(I+1) \times F$, being F the number of components estimated by PCA that should be equal to the number of calibrated

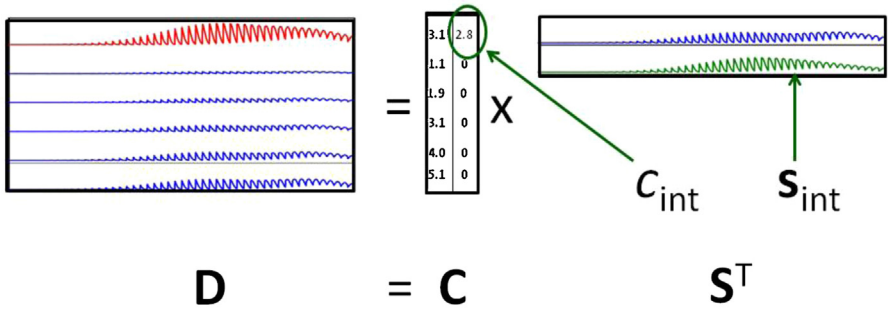


Fig. 4. Schematic representation of the application of MCR-ALS with correlation constraint to the decomposition of a simulated data matrix. The matrix results from five standard unfolded TSFS matrices and one validation sample containing both analyte and interferent. The procedure was used to generate both spectral profile and concentration of the interferent necessary to minimization in Eq. (11).

components plus the number of interfering agents] contains the concentration scores, and $\mathbf{S}^T$ (of size $F \times JK$) contains the profiles of calibrated and interfering components. In the case of a single calibrated analyte and a single interferent:

$$\mathbf{S} = [\mathbf{s}_{int} | \mathbf{s}_c] \quad (10)$$

where $\mathbf{s}_{int}$ ($JK \times 1$) and $\mathbf{s}_c$ ($JK \times 1$) are the vectors containing the profiles of the interferent and the calibrated analyte, respectively (see Fig. 4).

The possibility to apply constraints to better model the shapes of the $\mathbf{C}$ and $\mathbf{S}^T$ profiles occurs during ALS optimization. Non-negativity can be applied to both concentration profiles and their related resolved spectra. Another possibility – which is explored in the present work – is to apply a correlation constraint to obtain concentration values in the real concentration scale [18]. In each iterative cycle of the resolution analysis, this constraint performs a regression model between the calculated MCR-ALS concentration values (in arbitrary units) and the reference concentration values that are known from the calibration samples. The established regression model is then used to update the values of the MCR-ALS concentration profiles to the real concentration scale. The outcome of this step provides a straightforward correlation among the values of the final resolution concentrations and the real concentrations in the sample.

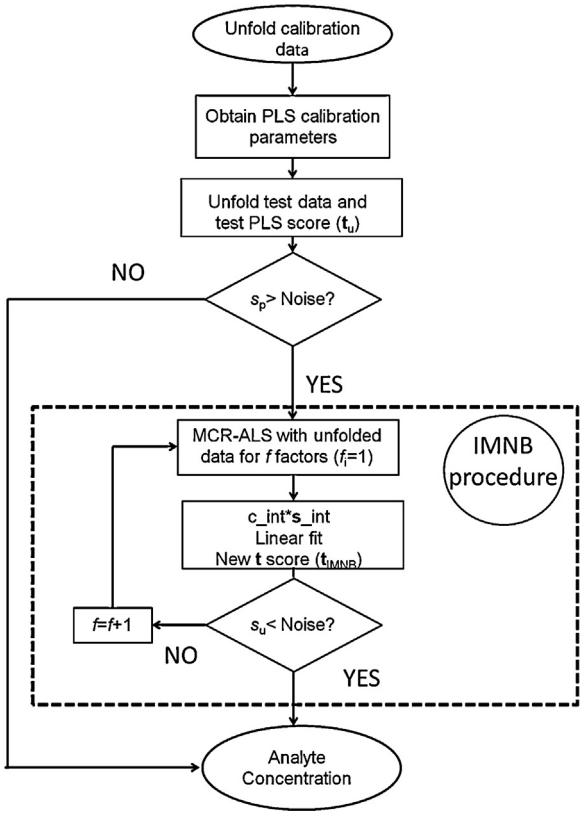


Fig. 5. Scheme illustrating how the new U-PLS/IMNB/MCR-ALS method works. See text for explanation of symbols.

In the final step, the contribution of the interferent, extracted from the $\mathbf{S}^T$ matrix and its corresponding concentration value, is then modeled in the IMNB/MCR-ALS procedure according to the following expression:

$$\mathbf{x}_u = \mathbf{P}\mathbf{t}_{uIMNB} + c_{int}\mathbf{s}_{int} + \mathbf{e}_u \quad (11)$$

where c_{int} is the concentration score for the interferent in the test sample.

Fig. 4 illustrates the decomposition of a matrix built with data extracted from the simulated example and composed of five unfolded calibration matrices and one unknown sample. The unfolded TSFS vector corresponding to the interference spectral profile ($\mathbf{s}_{int}$) and its score (c_{int}) are obtained from the MCR-ALS decomposition.

The minimization of $\| \mathbf{e}_u \|$ with the IMNB/MCR-ALS procedure is carried out by keeping $\mathbf{P}$ constant at the calibration values and obtaining new PLS scores ($\mathbf{t}_{uIMNB}$) with a linear fit by projecting ($\mathbf{x}_u - c_{int}\,\mathbf{s}_{int}$) onto the generalized inverse of $\mathbf{P}$ in Eq. (11). Then, the substitution of $\mathbf{t}_u$ for the final value of the $\mathbf{t}_{uIMNB}$ vector in Eq. (2) provides the analyte concentration in the sample. The complete process is outlined in Fig. 5.

Finally, it should be taken into account that this kind of vectorized structure will behave in a bilinear way only if the nonbilinear deviation pattern of the response is very reproducible among samples. This happens in synchronous fluorescence spectroscopy, but may not happen in other examples of nonbilinear systems.

3.3. PARAFAC

As previously mentioned, PARAFAC [4] appears to be the appropriate algorithm to handle data that follow the trilinear structure, i.e., the three-way array built with a set of bilinear second-order data (see Eq. (1)). In the present work, we use PARAFAC to show how the bilinearity loss affects the modeling and prediction ability of this type of algorithm.

4. Results and discussion

4.1. Simulated data

The strong overlapping that exists among the spectral features of the analyte and the interferent at all $\Delta\lambda$ values (see Fig. 3) calls for an algorithm with the second-order advantage. Table 1

summarizes the results obtained with PARAFAC, U-PLS/RBL, MCR-ALS and U-PLS/IMNB/MCR-ALS. The extremely poor recoveries (relative error of prediction, REP% > 100) obtained for the nine solutions of the test sample data set are due to the inability of PARAFAC to handle the nonbilinear structure of the analyzed data. The number of responsive components ($A = 2$) used to build the calibration model was obtained by analyzing the residuals of the least-squares fit of the three-way data array.

In the application of U-PLS/RBL, the number of latent variables A is assessed with the leave-one-out cross-validation criterion [24]. Since in our case the only responsive component in the calibration samples was the analyte, A was equal to 1. The contribution of unexpected components (N_{unx}) was evaluated by applying the RBL method to the test sample data. The analysis of the residuals s_u obtained with the RBL procedure was done according to Eq. (7). Table 2 shows the variation of the residuals and predictions for validation sample V1–1. The tabulated figures indicate that the residual (0.010) stabilizes at $N_{unx} = 3$. Considering the presence of a single interferent in the test sample, the necessity of introducing a larger number of components to better modeling the contribution of the interference becomes evident. In this case, in which one unexpected compound is considered, three components allow to obtain a recovery of 109.4%. On the other hand, if a large number of components is considered ($N_{unx} = 15$), the prediction notably improves, reaching to a recovery of 100.6%. Unfortunately, this excessive number of components makes the strategy impracticable in the case of two or more unexpected components. The U-PLS/RBL predictions (using $N_{unx} = 3$ as a compromise value, i.e. the large series of bilinear components that would be required to do so is truncated in some arbitrary way) for the nine samples of the validation set are summarized in Table 1. Although the results are better than those obtained with PARAFAC, the predicted concentrations are still far from satisfactory. Once again, it should be stated that the results could be improved using $N_{unx} = 15$, but we consider this as an unfeasible approach.

Similar consideration can be made with the MCR-ALS modeling with unfolded data. In this case, it is expected that the second-order advantage can be exploited with first-order data, as achieved with data presenting lower overlapping [18]. However, predictions are still poor owing to the high complexity of the simulated system (see Fig. 3), although they are better than those gathered with U-PLS/RBL with $N_{unx} = 3$.

The prediction results obtained via the U-PLS/IMNB/MCR-ALS algorithm for the nine validation samples are summarized in the last column of Table 1. The better values obtained with this algorithm in all cases account for its superior ability to modeling the nonbilinear interference. In this case, it can be postulated that using the profile retrieved by MCR-ALS in the minimization step (Eq. (11)) has a similar effect as when using ca. 15 factors in the SVD decomposition of the residual matrix in the RBL procedure (Eq. (6)).

4.2. Experimental system

CPF and NOR are fluoroquinolones, antibacterial agents widely used in human and veterinary medicine. Consequently, these drugs end up in wastewaters coming from hospital and municipal emissions, whereas veterinary drugs are excreted by the animals and are released in the manure. Residues of these antibiotics have been reported in the natural environment in many countries [19]. Therefore, monitoring of low quantities of these compounds from different environmental matrices is essential for human health protection and environmental control.

Recently, Tong et al. reported the enhancement of enrofloxacin (another fluoroquinolone) synchronous fluorescence signal based on the yttrium(III)-perturbed luminescence [25]. Although the strategy is advantageous in the sense of improving the figures of merit, it precludes the discrimination of the signals originated by the analyte and CPF or NOR, used as interferences. Thus, the object of the present work was to record second-order TSFSM and modeling them with a suitable second-order algorithm which allows the quantitation of CPF in the presence of NOR as interferent.

It should be remarked that the excitation and emission spectra are highly overlapped, especially in the excitation spectra (see Fig. 6), and that the modeling of EEMs with U-PLS/RBL furnishes very poor results (data not shown). On the contrary, synchronous spectra present some slight differences that can be appreciated by visual inspection of Fig. 7. Therefore, some improvement in the prediction results could be expected if the data are conveniently modeled.

Table 3 summarizes the results obtained with PARAFAC, U-PLS/RBL, MCR-ALS and U-PLS/IMNB/MCR-ALS. Although PARAFAC provided considerably better results than those reported in Table 1, the worst results were still obtained with this algorithm (relative error of prediction, REP% 55.1). The number of responsive components ($A = 2$) used to build the calibration model was

Table 2
Number of unexpected components, residuals and predicted concentration of analyte in validation sample number 1 (nominal concentration equal to 1.59) when applying U-PLS/RBL. Calibration fit residue = 0.0010.

Number of unexpected components (Nunx)	Sample residual after RBL procedure	Predicted concentration[a]	Recovery (%)[b]
0	0.0670	2.59	162.9
1	0.0100	2.04	128.3
2	0.0026	1.86	117.0
3	0.0010	1.74	109.4
4	0.0010	1.74	109.4
5	0.0010	1.74	109.4
6	0.0009	1.74	109.4
7	0.0009	1.74	109.4
8	0.0009	1.72	108.2
9	0.0009	1.71	107.5
10	0.0009	1.71	107.5
11	0.0009	1.71	107.5
12	0.0008	1.71	107.5
13	0.0008	1.70	106.9
14	0.0008	1.62	101.9
15	**0.0008**	**1.60**	**100.6**

[a] Predicted concentration in arbitrary unities.
[b] Recovery (%) = (predicted/nominal) × 100.

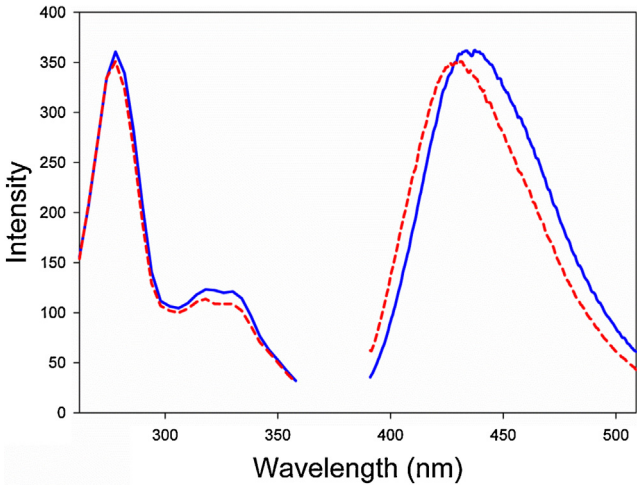


Fig. 6. Excitation and emission spectra of CPF 300 ng mL^{-1} (blue solid lines) and NOR 300 ng mL^{-1} (red dashed lines). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

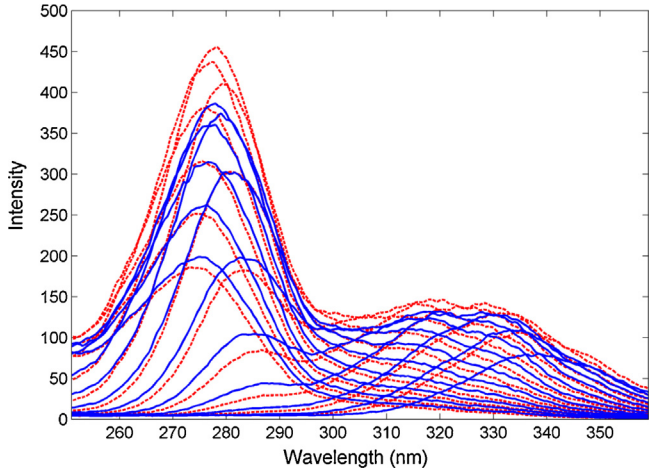


Fig. 7. Synchronous spectra of CPF 300 ng mL^{-1} (blue solid line) and NOR 300 ng mL^{-1} (red dotted line) generated at different wavelength offsets (between 80 and 200 nm, each 10 nm). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

obtained as described previously for the simulated system. The same is true for all calculations involving U-PLS/RBL modeling. Since the calibration samples were pure CPF standard solutions, the number of latent variables (A) was equal to 1. The behaviors of the number of unexpected components (N_{unx}) and the residuals (s_u) were similar to those observed for the simulated system. By using one component for modeling the calibration data and a variable number of unexpected components (ranged from 4 to 8), U-PLS/RBL provides better REP% than PARAFAC (see Table 3). Similar considerations to those for the simulated system can be made when applying MCR-ALS.

The best results are attained with U-PLS/IMNB/MCR-ALS, which provided a REP% value of 11.8. The latter supports the fact that the correct modeling of the interferent signal allows for a similar performance to that obtained when an extremely large number of factors is considered in the RBL procedure, as was commented for the simulated system. Finally, it is important to note that a pre-concentration step is necessary to reach the sensitivity required for water monitoring of these emerging contaminants. In this sense, we achieved a 1000-fold pre-concentration by using solid phase extraction with C18 membranes. Thus, concentration levels of 200 ppt of CPF and NOR could be measured in real samples by the presented method.

5. Conclusions

The second-order advantage for modeling strong spectral overlapping from data with nonbilinear structure has been achieved with a new algorithm that couples unfolded partial least squares with MCR-ALS. The enhanced prediction ability of the new method – U-PLS/IMNB/MCR-ALS – was demonstrated with both simulated and experimental data. Its prediction ability was compared to the performances of three well-known second order calibration methods, namely U-PLS/RBL, MCR-ALS and PARAFAC. In all the investigated systems, the new method showed better prediction ability than the established methodologies.

Further work will be needed to deeply understand the behavior of specific algorithms in relation to the presently analyzed data. In addition, issues such as figures of merit computation should be focused, in order to make better comparisons among different algorithms. In this report, two examples having a single analyte and a single interferent have been studied. Although the results are promising, additional work is being carried out in our lab to confirm these trends in more complex samples in which additional analytes and potential interfering components occur.

Acknowledgments

The authors are grateful to Univers'idad Nacional del Litoral (Projects CAI + D 2012 No. 11-11 and No. 11-7), to CONICET (Consejo

Table 3
Results of the experimental data set corresponding to ciprofloxacin quantitation in the presence of norfloxacin in water samples.

Sample	CFP nominal concentration (ng mL^{-1})	NOR.	Predicted ciprofloxacin concentration (ng mL^{-1}) PARAFAC[a]	U-PLS/RBL[b]	MCR-ALS[c]	U-PLS/IMNB /MCR-ALS [d]
1	410	290	550	390 (4)	360	400
2	450	400	640	460 (5)	460	460
3	300	550	520	320 (8)	430	290
4	190	510	400	290 (7)	350	220
5	300	250	320	300 (4)	310	290
6	190	290	260	230 (7)	260	230
7	150	400	280	230 (7)	280	220
REP%	–	–	**55.1**	**18.3**	**34.5**	**11.8**

[a] The number of responsive components was set by analyzing the residuals of the least-squares fit of the three-way data array to the trilinear model (A = 2).
[b] Value between parenthesis corresponds to number of interferents being considered to stabilization of error s_u (see Eq. (7) text).
[c] Two components were considered.
[d] One factor extracted by MCR-ALS (see Eq. (9) text).

Nacional de Investigaciones Científicas y Técnicas, Project PIP 455) and to ANPCyT (Agencia Nacional de Promoción Científica y Tecnológica, Project PICT 2011-0005) for financial support. A.V.S. thanks CONICET for her fellowship.

References

[1] A.C. Olivieri, G.M. Escandar, A. Muñoz de la Peña, Second-order and higher-order multivariate calibration methods applied to non-multilinear data using different algorithms, Trends Anal. Chem. 30 (2011) 607–617.
[2] G.M. Escandar, N.M. Faber, H.C. Goicoechea, A. Muñoz de la Peña, A.C. Olivieri, R.J. Poppi, Second and third-order multivariate calibration: data, algorithms and applications, Trends Anal. Chem. 26 (2007) 752–765.
[3] G.M. Escandar, H.C. Goicoechea, A. Muñoz de la Peña, A.C. Olivieri, Second- and higher-order data generation and calibration: a tutorial, Anal. Chim. Acta 806 (2014) 8–26.
[4] R. Bro, Second- and higher-order data generation and calibration: a tutorial, Chemom. Intell. Lab. Syst. 38 (1997) 149–171.
[5] A.C. Olivieri, G.M. Escandar, Practical Three-way Calibration, Elsevier, Waltham (MA), USA, 2014.
[6] C.G. Zampronio, S.P. Gurden, L.A. Moraes, M.N. Eberlin, A.K. Smilde, R.J. Poppi, Direct sampling tandem mass spectrometry (MS/MS) and multiway calibration for isomer quantitation, Analyst (Cambridge UK) 127 (2002) 1054–1060.
[7] K. Kumar, A.K. Mishra, Simultaneous quantification of dilute aqueous solutions of certain polycyclic aromatic hydrocarbons (PAHs) with significant fluorescent spectral overlap using total synchronous fluorescence spectroscopy (TSFS) and N-PLS, unfolded-PLS and MCR-ALS analysis, Anal. Methods 3 (2011) 2616–2624.
[8] A.V. Schenone, J.M. Culzoni, A.D. Campiglia, H.C. Goicoechea, Total synchronous fluorescence spectroscopic data modeled with first- and second-order algorithms for the determination of doxorubicin in human plasma, Anal. Bioanal. Chem. 405 (2013) 8515–8523.
[9] K. Calimag-Williams, G. Knobel, H.C. Goicoechea, A.D. Campiglia, Achieving second order advantage with multi-way partial least squares and residual bi-linearization with total synchronous fluorescence data of monohydroxy-polycyclic aromatic hydrocarbons in urine samples, Anal. Chim. Acta 811 (2014) 60–69.
[10] B. Debus, M. Sliwa, H. Miyasaka, J. Abe, C. Ruckebusch, Multivariate curve resolution – alternating least squares to cope with deviations from data bilinearity in ultrafast time-resolved spectroscopy, Chemom. Intell. Lab. Syst. 128 (2013) 101–110.
[11] B.E. Wilson, W. Lindberg, B.R. Kowalsky, Multicomponent quantitative analysis using second-order nonbilinear data: theory and simulations, J. Am. Chem. Soc. 111 (1989) 3797–3804.
[12] K. Kumar, A.K. Mishra, Application of 'multivariate curve resolution alternating least square (MCR–ALS)' analysis to extract pure component synchronous fluorescence spectra at various wavelength offsets from total synchronous fluorescence spectroscopy (TSFS) data set of dilute aqueous solutions of fluorophores, Chemom. Intell. Lab. Syst. 116 (2012) 78–86.
[13] D. Patra, A.K. Mishra, Total synchronous fluorescence scan spectra of petroleum products, Anal. Bioanal. Chem. 373 (2002) 304–309.
[14] D. Patra, A.K. Mishra, Recent developments in multi-component synchronous fluorescence scan analysis, Trends Anal. Chem. 21 (2002) 787–798.
[15] A.K. Sarma, A.G. Ryder, Comparison of the fluorescence behavior of a biocrude oil and crude petroleum oils, Energy Fuels 20 (2006) 783–785.
[16] F. Garcia Sanchez, A.L. Ramos Rubio, C. Cruces Blanco, R. Suau Suarez, Three-dimensional synchronous fluorescence spectrometry for the analysis of three-component alkaloid mixtures, Talanta 37 (1990) 579–584.
[17] J. Öhman, P. Geladi, S. Wold, Residual bilinearization. Part 1: theory and algorithms, J. Chemom. 4 (1990) 79–90.
[18] H.C. Goicoechea, A.C. Olivieri, R. Tauler, Application of the correlation constrained multivariate curve resolution alternating least-squares method for analyte quantitation in the presence of unexpected interferences using first-order instrumental data, Analyst 135 (2010) 636–642.
[19] F. Tamtam, F. Mercier, J. Eurin, M. Chevreuil, B. Le Bot, Ultra performance liquid chromatography tandem mass spectrometry performance evaluation for analysis of antibiotics in natural waters, Anal. Bioanal. Chem. 393 (2009) 1709–1718.
[20] MATLAB 7.10 (2010) The MathWorks Inc. Natick, Massachussets, USA.
[21] A.C. Olivieri, H.-L. Wu, R.-Q. Yu, MVC2: A MATLAB graphical interface toolbox for second-order multivariate calibration, Chemom. Intell. Lab. Syst. 96 (2009) 246–251.
[22] www.iquir-conicet.gov.ar/descargas/mvc2.rar.
[23] S. Wold, P. Geladi, K. Esbensen, J. Øhman, Multi-way principal components-and PLS-analysis, J. Chemom. 1 (1987) 41–45.
[24] D.M. Haaland, E.V. Thomas, Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information, Anal. Chem. 60 (1988) 1193–1202.
[25] C. Tong, X. Zhuo, W. Liu, J. Wu, Synchronous fluorescence measurement of enrofloxacin in the pharmaceutical formulation and its residue in milks based on the yttrium(III)-perturbed luminescence, Talanta 82 (2010) 1858–1863.