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†

Héctor C. Goicoechea,*[a] Alejandro C. Olivieri[b] and Romà Tauler[c]

Received 28th October 2009, Accepted 23rd December 2009
First published as an Advance Article on the web 15th January 2010
DOI: 10.1039/b922547a

Correlation constrained multivariate curve resolution-alternating least-squares is shown to be a feasible method for processing first-order instrumental data and achieve analyte quantitation in the presence of unexpected interferences. Both for simulated and experimental data sets, the proposed method could correctly retrieve the analyte and interference spectral profiles and perform accurate estimations of analyte concentrations in test samples. Since no information concerning the interferences was present in calibration samples, the proposed multivariate calibration approach including the correlation constraint facilitates the achievement of the so-called second-order advantage for the analyte of interest, which is known to be present for more complex higher-order richer instrumental data. The proposed method is tested using a simulated data set and two experimental data systems, one for the determination of ascorbic acid in powder juices using UV-visible absorption spectral data, and another for the determination of tetracycline in serum samples using fluorescence emission spectroscopy.

Introduction

In recent years, the subject of analyte determination in the presence of unexpected sample constituents, *i.e.*, those not taken into account in the calibration phase, has been a field of major scientific improvements. Recent reviews and perspective articles on the subject highlight the importance of analyte quantitation regarding the analysis of complex and natural samples, even when potential interferences occur in a given specimen.[1–5] A variety of strategies have been discussed, most of them taking advantage of the intrinsic property of higher-order instrumental data, *i.e.*, data having two or more inter-modulated measurement orders or modes per sample.[5] In these cases, the so-called second-order advantage[6] can be implemented in well-known algorithms for the adequate processing of higher-order (multi-way or multi-mode) data, as has been exploited in a substantial number of recently proposed analytical methods (mainly but not restricted to the spectroscopic field).[1]

The problem of the appearance of measurement interferences in an unknown sample is ubiquitous. Components present in both calibration and validation samples are regularly called 'expected', because analysts expect them to be present in most test samples, and therefore include them into a sufficiently representative training sample set. The expected components can be further divided into 'calibrated' and 'uncalibrated', referring to whether or not calibration concentrations are available for each of them. Truly unknown samples may contain additional components which are not expected, hence the name 'unexpected' components. Although these potential interferences may generate a signal that overlaps with the analyte of interest, they will not always produce a real interference, because they will not always lead to a systematic error in the analyte determination.[7] Whether the interference will be actual or will only remain as potential, depends on the type of instrumental signals and on the calibration methodology. Univariate calibration, for example, cannot detect sample components producing an interfering signal. First-order calibration may compensate for potential interferences, usually by inclusion of the latter in the calibration set. This methodology does also allow one to flag new samples containing unexpected components, a property known as the first-order advantage.[6] On the other hand second- and higher-order calibration methods can compensate for potential interferences which are not included in the calibration set, which is the basis of the second-order advantage.[6]

Since first-order instrumental data can be measured using unsophisticated equipment, different procedures have been devised in order to quantitate analytes from first-order multivariate data in the presence of unexpected substances (interferences). One alternative is to select a spectral window where the contribution from the interferences is relatively low in comparison to the analyte contribution. This is for instance possible by resorting to the so-called net analyte signal regression plots.[8] The net analyte signal is the part of the total signal which can be uniquely ascribed to the analyte of interest,[9] and the regression plot corresponds to the net analyte signal for a test sample as

[a]*Laboratorio de Desarrollo Analítico y Quimiometría (LADAQ), Cátedra de Química Analítica I, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral, Ciudad Universitaria, Santa Fe, S3000ZAA, Argentina. E-mail: hgoico@fbcb.unl.edu.ar*
[b]*Departamento de Química Analítica, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario and Instituto de Química Rosario (IQUIR-CONICET), Suipacha 531, Rosario, S2002LRK, Argentina*
[c]*Institute of Environmental Assessment and Water Research, CSIC, Barcelona, 08036, Spain*

† Electronic supplementary information (ESI) available: details of the U-PLS/RBL and BBLS/RBL models. See DOI: 10.1039/b922547a

a function of that for the pure analyte. This plot should ideally be a straight line.[8] Significant deviations from linearity indicate the presence of unexpected constituents, opening the possibility of selecting only those sensor regions where these plots are linear. This has been accomplished in the framework of classical least-squares[8] and also in hybrid linear regression analysis.[10,11]

Another alternative is the so-called secured-principal component regression analysis.[12] In this variant of the well-known principal component regression (PCR) technique, test sample data are analyzed after conventional calibration in two steps: 1) investigating first whether sample data are consistent with the calibration model or not, and 2) extracting from the measured spectrum those spectroscopic features not adequately modeled by the principal components model. This corrected spectrum is then re-evaluated by conventional PCR.[12]

Recently, however, a different method has been proposed which intends to model rather than to discard the effect of unexpected components: this is the so-called augmented classical least-squares method,[13] in which the incorporation of prior information into the model can improve the accuracy of analyte quantitation in the presence of unknown interferences, especially when the pure component contributions for the interferents are known and taken into account. The latter is, however, the major weakness of this procedure, since it requires a thorough knowledge of the qualitative composition of the analytical system under scrutiny.

A second possibility is to employ the multivariate curve resolution-alternating least-squares (MCR-ALS)[14] method. This method iteratively applies natural constraints during an alternating least-squares optimization, such as: 1) non-negativity in concentrations and spectral profiles, 2) unimodality (as for chromatographic profiles), 3) closure relationships in closed systems. A *correlation constraint* has been previously proposed, where the objective was the quantitative analysis of mixtures of metal ions by voltammetry[15] or the quantitative determination of the components in mixtures of analytes in pharmaceutical using UV spectrophotometric data and of protein and humidity contents in agricultural samples using near infrared spectrophotometric data.[16] In this latter case, the experimental spectra of wheat samples were strongly affected not only by component interferences but also by physical and matrix effects.

The purpose of this work is to further investigate the ability of the proposed correlation constraint in MCR-ALS for analyte quantitation in complex mixtures, in the presence of unexpected interferences, and to check for the achievement of the second-order advantage, considering and processing first-order multivariate data. To assess the achievement of this advantage, in the case of the experimental examples investigated in this work, the results obtained using MCR-ALS with the correlation constraint have been compared with those obtained previously using the unfolded partial-least squares residual bilinearization (U-PLS/RBL)[17,18] and the bilinear least-squares/RBL (BLLS/RBL)[19] methods. These two methods have already proved to achieve the so-called second order advantage from instrumental data.[20,21] As will be shown, the use of the correlation constraint in MCR-ALS to model small subsets of first-order data allows one to achieve a performance comparable to the one obtained when using second-order data, at least in the application to the particular test cases herein investigated.

Experimental section

General

The two experimental systems analyzed in this work were: 1) the determination of ascorbic acid in powder juices using spectral-pH matrices[21] and 2) the determination of tetracycline in serum samples using excitation-emission fluorescence matrices.[20] Apart from these two experimental systems, an additional simulated data system (see below) was analyzed to better ascertain the properties of the proposed algorithm.

Electronic absorption measurements for system 1 were carried out on a Perkin Elmer Lambda 20 spectrophotometer (Perkin Elmer, Waltham, Massachusetts, USA), using 1.00 cm quartz cells and 2 nm slit width. Fluorescence measurements for system 2 were carried out on a Varian Cary-Eclipse spectrofluorimeter equipped with a xenon flash lamp, using 1.00 cm quartz cells. The spectra were saved in ASCII format, and transferred to a PC AMD Athlon 2200 for subsequent manipulation.

Reagents

All samples were prepared using analytical-reagent grade chemicals. For the experimental system 1, a stock solution of ascorbic acid (1.00×10^{-2} mol L^{-1}) was prepared by dissolving the compound in doubly distilled water. Auxiliary solutions of KH_2PO_4 (pH = 4.27), HCl and NaOH, all of them 1.00 mol L^{-1}, were prepared by convenient dissolution or dilution in doubly distilled water. For the experimental system 2, a stock solution of tetracycline, containing 1.000 g L^{-1}, was prepared by dissolving the compound in doubly distilled water. This solution was stable for at least 2 months in a refrigerator (4 °C). Serum samples were prepared by spiking blank sera with appropriate amounts of the stock solution of tetracycline.

Calibration and validation sets

For experimental system 1, three standard solutions of ascorbic acid were prepared by placing appropriate volumes of the stock solution in 5.00 mL volumetric flasks and completing to the mark with a buffer solution (see below). The concentrations of these standard solutions were 0.500, 0.750 and 1.000×10^{-4} mol L^{-1}. Linearity in the absorbance-concentration relation was verified in this concentration range. Three test samples were prepared: a) sample 'Ma', a blank specimen containing only background juice and no analyte [a commercial (Dink) powder orange juice dissolved in 1000 mL of distilled water followed by 1 : 5 dilution], b) sample 'Mb', a blank specimen spiked with ascorbic acid 0.30×10^{-4} mol L^{-1}, and c) sample 'Mc', a blank specimen spiked with ascorbic acid 0.50×10^{-4} mol L^{-1}.

For experimental system 2, and owing to the presence of analyte-background interactions, the calibration set was built by adding pure standards of tetracycline to sera obtained from healthy patients. Five standards were prepared for tetracycline calibration, with the following concentrations: 0.5, 1.0, 2.0, 3.0 and 4.0 μg mL^{-1}. The selected concentrations cover the therapeutic range of tetracycline. After suitable dilution (1 : 500), the latter concentrations rendered values which were previously verified to lie in the known linear fluorescence-concentration range, *i.e.*, 0.167–1.333 μg mL^{-1}. Appropriate amounts of the

stock solution of tetracycline were added to 1.00 ml of serum in order to obtain the desired concentrations. The samples were then homogenized, and 2.00 mL of magnesium acetate 5×10^{-3} mol L^{-1} in ammonium chloride 3 mol L^{-1} were added.[22] After homogenization, and within a period of 30 min from the sample preparation (in order to avoid signal intensity changes), the emission spectra were recorded at $\lambda_{exc} = 340$ nm. Eight test samples were prepared, all of them composed of sera from healthy patients (different than those used for the calibration set) and tetracycline, with the addition, in certain cases, of four drugs which are occasionally co-administered with tetracycline, *i.e.*, salicylate, paracetamol, ibuprofen and doxorubicin. The final concentration of tetracycline was 1.3μg mL^{-1} in the three samples and 1.8 μg mL^{-1} in the remaining five samples.

Spectra for the experimental systems 1 and 2 were actually collected in Ref. 20 and 21.

Theory

Bilinear least-squares with residual bilinearization (BLLS/RBL) and unfolded partial least-squares with residual bilinearization (U-PLS/RBL)

These two methods were described in detail in previous works.[20,21] A brief description of them is provided in the Supplementary Information. In Ref. 20 and 21 they were employed to analyze both of the experimental systems described in this work.

Multivariate curve resolution alternating least squares (MCR-ALS)

The spectral data matrix **D** analyzed by the MCR-ALS method contains in its rows the individual spectra measured for the different analyzed samples, and in its columns the sample signals measured at each spectral wavelength. In the first step, a rough estimation of the number of components is obtained, which can be simply performed by visual inspection of singular values or principal component analysis (PCA) plots,[23,24] or by carrying our leave-one-out cross-validation, as suggested in Ref. 25. This initial number of components can be afterwards refined considering larger or lower number of components, and checking for their fit and reliability (see below). The bilinear model assumed by MCR methods is analogous to the generalized Lambert-Beer's law, where the individual responses of each component are additive. In matrix form, this bilinear model is expressed as:

$$\mathbf{D} = \mathbf{C}\mathbf{S}^{T} + \mathbf{E} \quad (1)$$

where **D** (size $I \times J$) is the matrix of experimental data, (I is the number of samples or spectra and J the number of wavelengths), **C** (size $I \times K$) is the matrix of concentration profiles of the K components present in the samples, $\mathbf{S}^{T}$ (size $K \times J$) is the matrix of pure spectra, whose K rows contain the K pure component spectra and **E** (size $I \times J$) is a matrix collecting the experimental error and the variance not explained by the bilinear model of eqn (1).

The resolution is accomplished using an optimization as follows. The iterative Alternating Least Squares, ALS, procedure is initialized using an initial estimation of the spectral or concentration profiles for each intervening species. Different methods are used for this purpose such as EFA[15] or the determination of the purest variables.[16–18] In this work, initial estimations were based on the detection of purest variables which were more easily obtained and preferred. If the initial estimations are the spectral profiles, the unconstrained least-squares solution for the concentration profiles can be calculated from the expression:

$$\mathbf{C} = \mathbf{D}\,(\mathbf{S}^{T})^{+} \quad (2)$$

where $(\mathbf{S}^{T})^{+}$ is the pseudoinverse of the spectral matrix $\mathbf{S}^{T}$, which is equal to $[\mathbf{S}(\mathbf{S}^{T}\mathbf{S})^{-1}]$ when $\mathbf{S}^{T}$ is full rank.[26] If the initial estimations were the concentration profiles, the unconstrained least-squares solution for the spectra can be calculated from the expression:

$$\mathbf{S}^{T} = \mathbf{C}^{+}\mathbf{D} \quad (3)$$

where $\mathbf{C}^{+}$ is the pseudoinverse of **C** [$\mathbf{C}^{+} = (\mathbf{C}^{T}\mathbf{C})^{-1}\mathbf{C}^{T}$], when **C** is full rank.[26] Both steps can be implemented in an alternating least-squares cycle, so that at each iteration new **C** and $\mathbf{S}^{T}$ matrices are obtained. During these iterative recalculations of **C** and $\mathbf{S}^{T}$, a series of constraints are applied to give physical meaning to the obtained solutions, and to limit their possible number for the same data fitting.[27] Iterations continue until an optimal solution is obtained that fulfils the postulated constraints and the established convergence criteria. For example, non-negativity constraints are applied to the concentration profiles, due to the fact that the concentrations of the chemical species are always positive values or zero. Non-negativity constraints are also applied for UV-Vis, fluorescence or near-infrared spectra. Unimodality is a constraint which can be applied to profiles having a single maximum, as in the case of chromatographic profiles. Finally, closure constraints may be applied for the fulfilment of chemical mass balance equations among different chemical species in equilibrium or in kinetics.

In this paper a correlation constraint is also applied during the MCR-ALS analysis to establish calibration models for the quantitative determination of analytes in the presence of unknown interferences.[15,16] This correlation constraint consists of a series of steps performed during each iteration step of the ALS optimization. The concentrations of a particular analyte in calibration samples at each ALS iteration, are forced to be correlated to previously known reference concentration values of the analyte in these samples. A local linear model between the ALS estimated values and the nominal concentrations is built by least-squares linear regression. Concentration values are then updated according to the predicted values using the estimated parameters of the local model. More details about the implementation of this constraint in previous works can be found elsewhere.[15,16]

Data simulations

A simulated data set was especially prepared to validate the use of the MCR-ALS method with the correlation constraint in the analysis of multi-component synthetic mixtures having two calibrated analytes, and a single potential interference appearing in the test samples along with the analytes. Noiseless profiles for the analytes and for the potential interference are shown in Fig. 1. Using the analyte profiles shown in Fig. 1, a calibration set of 10 samples was built having random concentrations (both

nominal analyte concentrations were taken as random numbers, distributed with equal probability in the range 0–1). To produce the calibration data set, the simulated spectrum for a typical sample was obtained as the sum of the contributions of the two analytes. For the 100 test samples, both analytes were considered to be present at nominal concentrations which were also taken at random from the range 0–1. These test samples did also contain the potential interference, at concentrations taken at random from the range 0–1. Once the calibration and test data set matrices were built, noise was added to all signals (standard deviation = 0.005 units). The added noise represents *ca.* 1% with respect to the maximum calibration concentration and signal respectively. The simulated system consists of two calibrated analytes, present in the training sample set, and a series of test samples which do also contain a spectral interference or unexpected component. The calibration and test spectra were joined into a single data matrix of size 110 × 50.

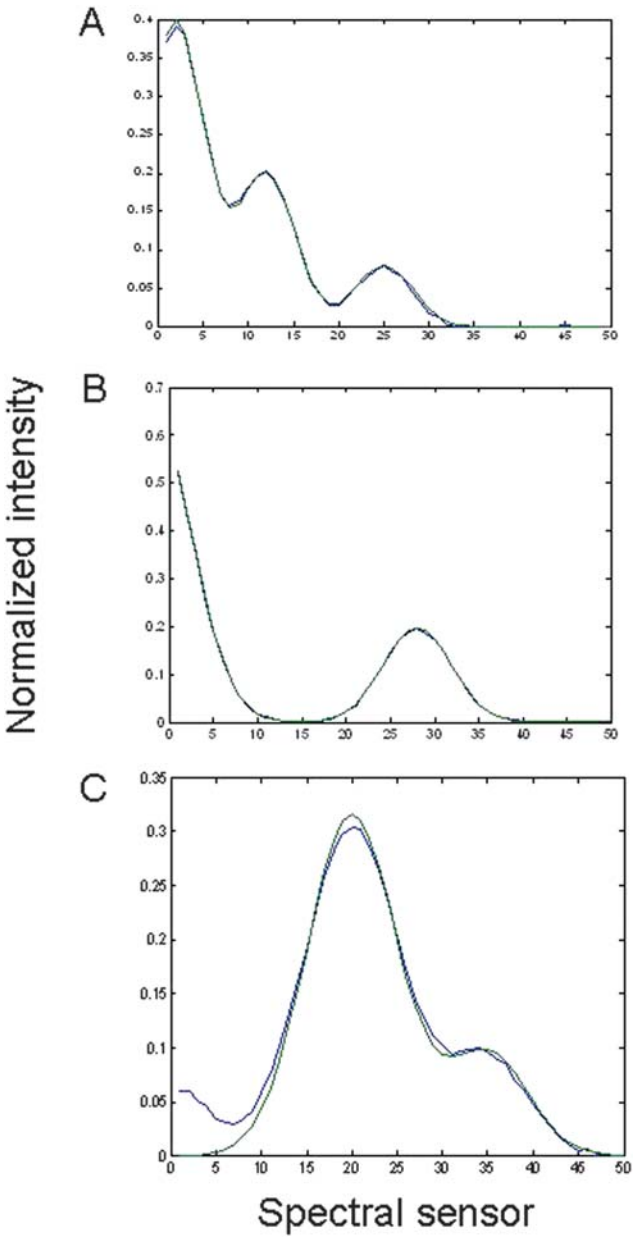

Fig. 1 Simulated profiles normalized to both unit length and unit concentration, employed for simulating the data and MCR-ALS retrieved spectral profiles. A) analyte 1, B) analyte 2, C) interferent.

With respect to the analyte calibration and test concentrations, the nominal values employed for building the calibration and test spectra were not directly employed for calibrating the model for analyte quantitation and for comparing the model predictions with the nominal values respectively. Instead, and in order to mimic a real analytical experiment, in which the sample preparation always carries some degree of uncertainty in the final analyte concentrations, gaussian-distributed noise with a standard deviation of 0.01 units was added to all nominal concentrations. The final values were employed for calibration and for comparison of predicted with nominal analyte concentrations in the test samples. This intends to resemble the existence of an average error of 0.01 concentration units in the analyte concentration during the preparation of all samples. Since the maximum concentration is 1 unit, then the concentration uncertainty level introduced in these simulations is also 1%, analogous to that employed in the case of the signal noise (see above).

Software

All simulations were done using MATLAB computer and visualization environment.[28] Correlation constrained MCR-ALS was implemented using a MATLAB code which is available on request from R. Tauler (e-mail: Roma.Tauler@idaea.csic.es).

Results and discussion

Simulated data

The simulated data matrix **D** was built with the spectra for the calibration and all test samples (see Fig. 1), and it has 10 calibration and 100 test samples measured at 50 different spectroscopic channels. Before submitting the **D** matrix to MCR-ALS decomposition, initial estimates of the three component spectra were obtained using SIMPLISMA.[16–18] In the present case, the number of components was known to be three because of the previous knowledge of three absorbing analytes in the system. Nevertheless, in a general case, the first step is an estimation of the number of components by visual inspection of a PCA plot or by cross-validation analysis. Non-negativity constraints were applied to both concentration and spectral profiles. The two calibration vectors (size 10 × 1) containing the nominal concentrations for each analyte in the calibration samples were provided as initial inputs, as well as the information that the interferent was not present in the calibration samples. This information was provided during the ALS optimization as a masking matrix of concentration values of the same size as the **C** matrix, in which the known calibration values are used for the corresponding correlation and equality constraints.[20,21]

After the MCR-ALS optimization, the retrieved spectral profiles were compared with those used in the data simulation (Fig. 1). As can be seen, the agreement is totally satisfactory, with excellent spectral recoveries for both analytes. Although the recovery of the interferent profile is not perfect, the analytical results were nevertheless satisfactory. With regards to the predicted concentrations of both analytes in the test samples, the root mean square error (RMSE) for analyte 1 was 0.060

concentration units, corresponding to a 12% relative error of prediction (REP) with respect to the mean calibration concentration (0.5 units). For analyte 2, in turn, the corresponding values were RMSE = 0.030, REP = 6%.

It should be noticed that the simulations implied a minimum average concentration error of 0.01 units (see above), introduced in order to resemble a real analytical experiment. In comparison, the RMSE values furnished by the multivariate model are slightly larger, as expected from the fact that error propagation to the final concentrations arises from several sources: 1) uncertainties in analyte concentrations during sample preparation, 2) uncertainties in the experimentally measured signals for each sample, and 3) model errors attributed to the least-squares fitting employed to calibrate the model.

Experimental system 1

In a previous work, this system was studied by processing second-order spectral-pH data.[21] pH was varied between 1.83 and 6.80 for each sample (including the calibration standards and the test samples). This pH variation was obtained by adding appropriate volumes of either acid or basic solutions to thirteen volumetric 5.00 mL flasks. In this way, both species (ascorbic acid/acorbate) were assured to be present in equilibrium at significant concentrations for each of these thirteen solutions. Their absorption spectra were then recorded between 200 and 300 nm, at 1 nm intervals, giving a data matrix of size 13 × 101 data points for each sample. The application of the bilinear least-squares/residual bilinearization (BLLS/RBL) method gave acceptable predictions for test samples. It is interesting to remark that in the test samples not only was spectral overlapping severe, but also the absorbance due to the background was significantly larger than that of the analyte, adding a considerable challenge to the determination of its concentration.

In the present work, only the spectra recorded at the single pH = 1.83 were used. At this pH the only species present was the ascorbic acid (the interference was also present in test samples). The data matrix **D** was built with three calibration spectra and three test samples (size 6 × 101). The number of components was estimated by visual inspection of a PCA plot and by cross-validation, and it was two, as previously commented. Initial estimates were also obtained using SIMPLISMA. Fig. 2 clearly shows the presence of a non-calibrated interference (the background juice), which is present in the test samples.

Fig. 3 shows the retrieved spectral profiles of the analyte and of the interference after MCR-ALS application. A visual inspection of the acid ascorbic pure normalized spectrum and of its MCR-ALS resolved spectrum indicates an excellent agreement between them. The MCR-ALS resolved spectrum corresponding to the background juice sample is also shown in Fig. 3. As can be observed, spectra overlapping occurs in the whole wavelength range. Concerning the predicted concentrations of ascorbic acid in the test samples, the root mean square error (RMSE) for them was 0.03×10^{-4} mol L^{-1}, which corresponds to a relative error of prediction of 14.7% with respect to the mean calibration concentration (0.75×10^{-4} mol L^{-1}). As it can be seen in Table 1, these values are close to those obtained using the BLLS/RBL method when they were applied to pH modulated second order spectral data matrices of the same system.[21] The main advantage in this case however is the fact that MCR-ALS was applied to simpler single pH first order data and the correct recovery of the qualitative spectral information not only of the analyte, but also of the interference (background juice).

Table 1 Predictions obtained by BLLS/RBL and MCR-ALS on real powder juice samples spiked with known concentrations of ascorbic acid

	Nominal	BLLS/RBL predictions		MCR-ALS predictions	
Sample	(mol L^{-1} × 10^4)	(mol L^{-1} × 10^4)	Recovery (%)	(mol L^{-1} × 10^4)	Recovery (%)
Ma	0.000	0.100	—	0.020	—
Mb	0.300	0.300	100.0	0.296	98.7
Mc	0.500	0.470	94.0	0.555	111.0
RMSE (mol L^{-1} × 10^4)		0.02		0.03	

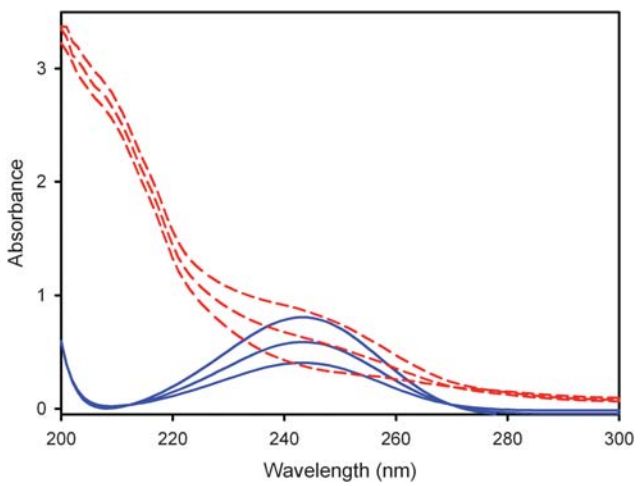

Fig. 2 UV spectra corresponding to ascorbic acid standard solutions (0.500, 0.750 and 1.000 × 10^{-4} mol L^{-1}) in solid lines, and three blank samples spiked with ascorbic acid (0.000, 0.300 and 0.500 × 10^{-4} mol L^{-1}) in dashed lines.

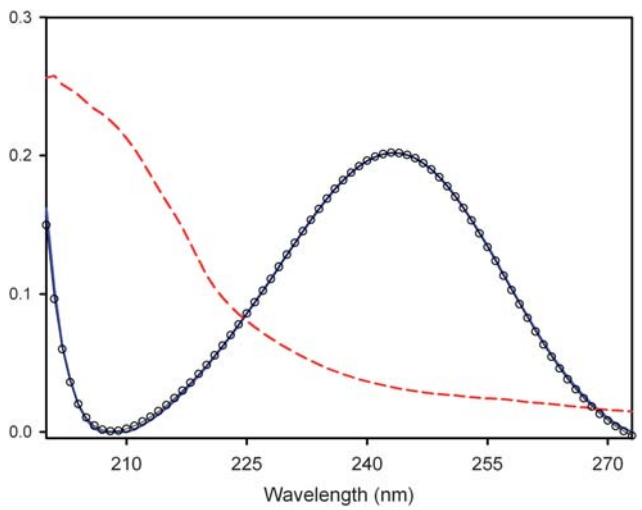

Fig. 3 Comparison between the MCR-ALS retrieved spectral profile (solid line) with the acid ascorbic pure normalized spectrum (circles). The dashed line is the profile for the interference as extracted by MCR-ALS.

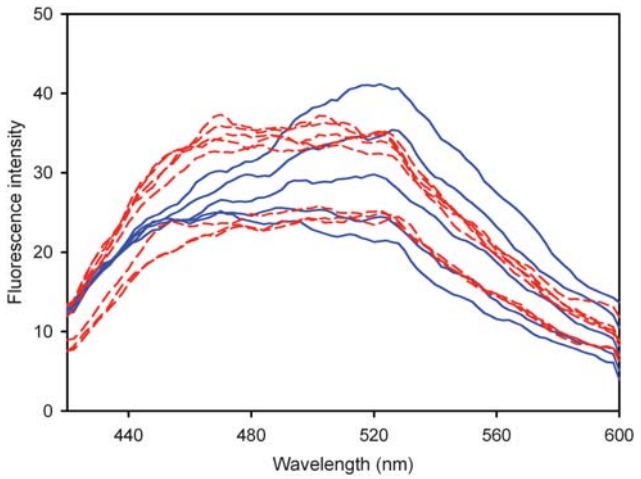

Fig. 4 Emission spectra (400–600 nm) recorded when exciting at λ_{exc} = 340 nm five tetracycline standard solutions (0.5, 1.0, 2.0, 3.0 and 4.0 μg mL^{-1}, solid lines), and eight serum samples spiked with tetracycline (see Table 2, dashed lines).

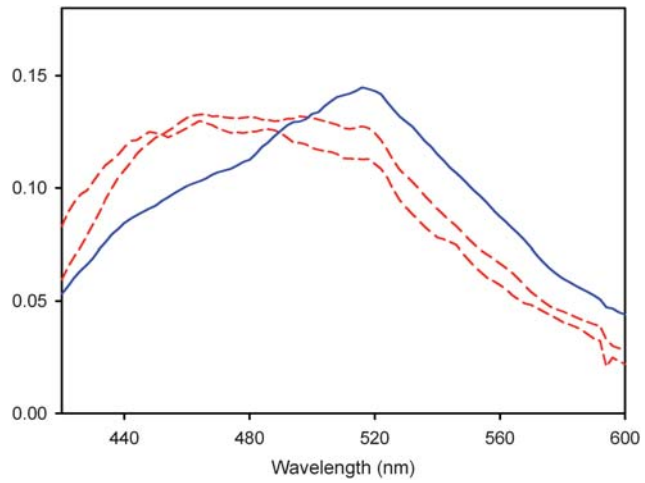

Fig. 5 Retrieved MCR-ALS profiles when analyzing sample number 6 (see Table 2) which contains the more complex serum sample.

Experimental system 2

Concentrations of tetracycline in serum samples were previously estimated using excitation-emission fluorescence matrices (EEMs) and the unfolded partial least-squares followed by residual bilinearization (U-PLS/RBL) method, which is considered to be a powerful method achieving the so-called second-order advantage. By contrast, in the present work emission spectra were recorded at only a single excitation wavelength (λ_{exc} = 340 nm). Fig. 4 shows the data matrix **D** built using five calibration spectra and eight test sample spectra (size of 13 × 101). As it can be seen, significant differences are observed for several samples, especially in the spectral region 440–485 nm. These spectral differences can be ascribed to both endogenous serum components (not present in the serum sample employed to build the calibration set) and to the possible interferences added to the test samples. For each kind of sample serum being analyzed, the number of components was estimated by visual inspection of a PCA plot and also by cross-validation analysis. As an example, Fig. 5 shows the three MCR-ALS profiles obtained in this experiment when analyzing one of the most complex samples (number 6 in Table 2).

For the predicted concentrations of tetracycline in the test samples (Table 2), the RMSE values was 0.3 μg mL^{-1}, corresponding to a 14.7% of relative error of prediction (REP) with respect to the mean calibration concentration (2.1 μg mL^{-1}). Table 2 also shows that these values obtained are even better than those previously obtained by implementation of U-PLS/RBL on second-order data for the study of the same system.[20] These results therefore confirm the power of the proposed approach and are in agreement also with the preliminary results reported previously about the use of the correlation constraint in the MCR-ALS method.

Table 2 Predictions obtained by U-PLS/RBL and MCR-ALS on serum samples spiked with know concentrations of tetracycline

	Nominal	U-PLS/RBL predictions		MCR-ALS predictions	
Sample	(μg mL^{-1})	(μg mL^{-1})	Recovery (%)	(μg mL^{-1})	Recovery (%)
1	1.30	1.44	110.8	1.02	78.5
2	1.30	1.52	116.9	1.26	96.9
3	1.30	1.52	116.9	1.27	97.7
4	2.80	3.23	115.4	3.20	114.3
5	2.80	3.24	115.7	3.07	109.6
6	2.80	3.23	115.4	3.04	108.6
7	2.80	3.28	117.1	3.21	114.6
8	2.80	3.40	121.4	3.37	120.4
RMSE (μg mL^{-1})		0.4		0.3	
REP (%)		17.0		14.7	

Conclusion

From the results obtained in the analysis of the three present data examples, it can be concluded that the extension of the MCR-ALS method, including the proposed correlation constraint, allowed for the accurate quantitative determination of analyte concentrations using spectrophotometric or fluorescence first-order data, even in the presence of unexpected interferences. This allows for the design of calibration procedures where the samples used at the calibration stage do not contain the unknown interferences, *i.e.*, when only external standard mixture samples of the analytes are available, not containing the interferents. An additional advantage of the MCR-ALS method (either when the correlation is applied or not) is the possible recovery of the spectral information of analytes and of possible unknown interferences in the analyzed samples. This extra information is not available in most of the other multivariate first order calibration methods, except for the causal modelling using a classical least squares (CLS) multi-component calibration method where all the spectra of the components in the calibration and test mixtures are known, or where all the concentrations of components in the mixtures are known in the calibration step. The presence of unknown interferences in the samples disturb this type of data analysis in practice and precludes its use in the analysis of natural samples where interferents are ubiquitous,

such as in different examples shown in this work. The generalization of the conditions where the second-order advantage can be extended to uncalibrated interferences using the proposed approach is difficult, because it will depend on the complexity of the system under study. For instance, it will be difficult to achieve for the interferences in a many-component system with significant unresolved rotational freedom, and with strong overlap of their profiles in the two modes. On the contrary, it will be much easier to be achieved for simpler systems, with a few number of components and with interferences having additional selectivity in their profiles. Many intermediate situations are then possible and more work is needed to clarify this aspect completely.

Acknowledgements

Universidad Nacional del Litoral, Universidad Nacional de Rosario, CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas), ANPCyT (Agencia Nacional de Promoción Científica y Tecnológica), and MCyI (CTQ2006-15052-C02-01 grant) for financial support.

References

1 G. M. Escandar, N. M. Faber, H. C. Goicoechea, A. Muñoz de la Peña, A. C. Olivieri and R. J. Poppi, *TrAC, Trends Anal. Chem.*, 2007, **26**, 752–765.
2 V. Gómez and M. Pilar Callao, *Anal. Chim. Acta*, 2008, **627**, 169–183.
3 M. Martínez Galera, M. D. Gil García and H. C. Goicoechea, *TrAC, Trends Anal. Chem.*, 2007, **26**, 1032–1042.
4 R. Bro, *Crit. Rev. Anal. Chem.*, 2006, **36**, 279–293.
5 A. C. Olivieri, *Anal. Chem.*, 2008, **80**, 5713–5720.
6 K. S. Booksh and B. R. Kowalski, *Anal. Chem.*, 1994, **66**, 782A–791A.
7 W. E. Van der Linden, *Pure Appl. Chem.*, 1989, **61**, 91–95.
8 J. Ferré and F. X. Rius, *Anal. Chem.*, 1998, **70**, 1999–2007.
9 A. Lorber, *Anal. Chem.*, 1986, **58**, 1167–1172.
10 H. C. Goicoechea and A. C. Olivieri, *Analyst*, 1999, **124**, 725–731.
11 H. C. Goicoechea and A. C. Olivieri, *Talanta*, 1999, **49**, 793–800.
12 F. Vogt and B. Mizaikoff, *Anal. Chem.*, 2003, **75**, 3050–3058.
13 W. Saeys, K. Beullens, J. Lammertyn, H. Ramon and T. Naes, *Anal. Chem.*, 2008, **80**, 4951–4959.
14 A. DeJuan, E. Casassas, R. Tauler, in R. A. Myers (Ed.), *Encyclopedia of Analytical Chemistry*, Vol. 11, Wiley, Chichester, 2002, pp. 9800–9837.
15 M. C. Antunes, J. E. J. Simao, A. C. Duarte and R. Tauler, *Analyst*, 2002, **127**, 809–817.
16 T. Azzouz and R. Tauler, *Talanta*, 2008, **74**, 1201–1210.
17 J. Öhman, P. Geladi and S. Wold, *J. Chemom.*, 1990, **4**, 79–90.
18 A. C. Olivieri, *J. Chemom.*, 2005, **19**, 253–265.
19 M. Linder and R. Sundberg, *Chemom. Intell. Lab. Syst.*, 1998, **42**, 159–178.
20 M. J. Culzoni, H. C. Goicoechea, A. P. Pagani, M. A. Cabezón and A. C. Olivieri, *Analyst*, 2006, **131**, 718–732.
21 H. C. Goicoechea and A. C. Olivieri, *Appl. Spectrosc.*, 2005, **59**, 926–933.
22 H. Poiger and C. Schlatter, *Analyst*, 1976, **101**, 808–814.
23 M. Maeder and A. Zilian, *Chemom. Intell. Lab. Syst.*, 1988, **3**, 205–213.
24 J. Jaumot, R. Gargallo, A. de Juan and R. Tauler, *Chemom. Intell. Lab. Syst.*, 2005, **76**, 101–110.
25 R. G. Brereton, *Chemometrics. Data Analysis for the laboratory and chemical plant*; Wiley, Chichester, UK, 2003; p 199.
26 G. H. Golub, C. F. Van Loan, *Matrix computation*, Second ed., The Johns Hopkins University Press, Baltimore, 1989.
27 R. Tauler, A. Smilde and B. R. Kowalski, *J. Chemom.*, 1995, **9**, 31–58.
28 *MATLAB 7.0*, The Mathworks, Natick, Massachussets, USA, 2007.

