Evaluation of complex spectral-pH three-way arrays by modified bilinear least-squares: determination of four different dyes in interfering systems

Nilda R. Marsili,[a] Adriana Lista,[b] Beatriz S. Fernandez Band,*[b] Héctor C. Goicoechea*[a] and Alejandro C. Olivieri*[c]

This article reports on the first application of a modified version of the bilinear least-squares model to absorbance-pH second-order data recorded for complex samples. The latter are composed of fruit drink powders, where four different analytes and additional background components occur. The analytes are the common juice colorants tartrazine, yellow sunset, allura red and indigo carmine. The data have been measured after generating a double pH gradient within a flow injection system. The selected chemometric methodology adequately exploits the second-order advantage, needed to take into account the background interferents present in real samples. Due to severe spectral overlapping between the acid and basic forms of each of the colorants in the working pH range, other second-order multivariate calibration methods such as parallel factor analysis and multivariate curve resolution-alternating least-squares could not be successfully applied to the presently studied samples. Recoveries of 94.8, 104.7, 109.3 and 105.3% were obtained for yellow sunset, indigo carmine, allura red and tartrazine respectively in the real test samples.

Introduction

Multi-way analysis is becoming popular due to the availability of high-order instrumental data and the proliferation of chemometric algorithms for data processing.[1] A particularly appealing property of multi-way data is the second-order advantage,[2] a term coined to describe the analysis in the presence of unsuspected interferents. Second-order, i.e., matrix data for a given sample can be produced in a variety of ways, among which an attractive possibility is to take advantage of the analyte acid–base properties. For example, a sample can be subjected to a pH gradient within a flow injection analysis (FIA) system, with the FIA peak subsequently registered at a number of wavelengths by diode array detection. Spectra can be obtained at different times along the FIA peak, equivalent to recording spectra at different pH values. This FIA/diode array detection mode represents an important experimental advantage.[3–5] The absorbance changes due to variations in the proton-transfer species composition with the pH give rise to absorbance-pH matrix data, and grouping these type of data for a set of samples will produce a three-way array.

A variety of chemometric methods is available to analyze three-way data, namely generalized rank annihilation (GRAM),[6] parallel factor modeling (PARAFAC),[7,8] multivariate curve resolution coupled to alternating least-squares (MCR-ALS)[9] and the prize-winning bilinear least-squares (BLLS).[10,11] However, absorbance-pH matrix data pose special challenges to these multivariate algorithms, due to the fact that the species' concentrations are correlated along the pH dimension, i.e., the sum of individual species corresponding to a given analyte is constant. This introduces linear dependencies in the pH data dimension, requiring the application of constraints during data processing for successful calibration and prediction. Specific software developed for linear dependent data are some versions of PARAFAC (PARAFAC2,[12] PARALIND[13]), MCR-ALS, residual bilinearization RBL[14] and a recent modification of BLLS.[15] The most popular alternatives seem to be MCR-ALS and PARAFAC, whose software is easily available through the internet.[16] However, as shown in the present report, these two algorithms are not able to handle extreme collinearities which may also be present in the spectral dimension, as those shown by the analytes herein studied. Therefore, we have explored the alternative use of a recently modified BLLS methodology, which takes into account the presence of pH-equilibrating species.[15] This latter method is shown to be highly efficient in analyzing data with severe correlations in both the spectral and pH dimensions.

In this report, we focus on the analysis of dyes in foodstuff. Color is an important characteristic of foods because it allows one to capture the desired aesthetic and organoleptic qualities of a particular foodstuff, and this is the main reason for the importance allotted to colorants in food industry. Some synthetic colors may be pathogenic, especially when they are consumed in excess, and thereby they should be evaluated by both manufacturers and health organizations.[17,18]

[a] Cátedra de Química Analítica, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral, Ciudad Universitaria, Santa Fe (S3000ZAA) CC. 242, Argentina.
E-mail: hgoico@fbcb.unl.edu.ar
[b] Laboratorio FIA-Química Analítica, Departamento de Química, Universidad Nacional del Sur, Av. Alem 1253, Bahía Blanca (B8000CPB), Argentina. E-mail: usband@criba.edu.ar
[c] Departamento de Química Analítica, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Suipacha 531, Rosario S2002LRK. E-mail: aolivier@fbioyf.unr.edu.ar

Tartrazine (T, trisodium 5-hydroxy-1-(4-sulfophenyl)-4-(4-sulfophenylazo) pyrazole-3-carboxylate), yellow sunset (Y, 6-hydroxy-5-[(4-sulfophenyl)azo]-2-naphtalenesulfonic acid disodium salt), allura® red (R, 6-hydroxy-5-[(2-methoxy-5-methyl-4-sulfophenyl)azo]-2-naphtalenesulfonic acid disodium salt) and indigo carmine (I, 5,5′-(2-(1,3-dihydro-3-oxo-2H-indazol-2-ylidene)-1,2-dihydro-3H-indol-3-one) disulfonate) are four synthetic dyes available as yellow (T and Y), red (R) and blue (I) powders, and are present in powder drinks either alone or in suitable combinations. Several methods have been proposed for the determination of these colorants, liquid chromatographic techniques being the recommended ones for analyzing mixtures of several colorants.[19–22] Owing to the complexity of these type of samples, other separative techniques such as capillary electrophoresis have been applied.[23,24] Spectrophotometric methods, especially when assisted by chemometric tools, have also been discussed for solving mixtures of two or three colorants.[25,26] Additional techniques have been employed, including differential pulse polarography,[27] square wave adsorptive voltammetry[28] and mass spectrometry.[29]

The present work attempts to demonstrate that the combination of pH gradient/diode array detection in a flow system is a convenient way of simultaneously measuring the concentration of four colorants with strongly overlapping spectra, in the presence of uncalibrated interferents. Critical to the success of the analytical determination is the employment of an adequate chemometric methodology for data analysis in the presence of strong correlations.

Experimental

Apparatus

Spectrophotometric measurements were performed on a Hewlett Packard 8452A spectrophotometer with a diode array detector and a Hellma 178-010-QS flow cell with an inner volume of 18 µL. A Gilson Minipuls 3 peristaltic pump was used as the propulsion device. A Rheodyne 5041 injection valve was used. Reactor and tubing of PTFE (id 0.5 mm) were used.

Reagents

All solutions were prepared from analytical grade chemicals and purified water (18 mΩ) by using a B-pure system. The buffer solution was prepared by mixing 96.1 mL of potassium dihydrogen phosphate (Mallinckrodt) (9.073 g L^{-1}) and 3.9 mL of di-sodium phosphate (Mallinckrodt) (9.464 g L^{-1}). The final pH of this solution was 5.5. Three commercial samples were analyzed: Tang orange flavor (J1), Tang tangerine flavor (J2) and Clight apple flavor (J3) (Kraft Foods Argentina S.A.), all purchased in the local market. The available pK_a values for the analytes are as follows:[30] tartrazine, 9.4, yellow sunset, 10.4, allura red, 11.4 and indigo carmine, 12.2.

Flow injection methodology

A simple FIA assembly (shown in Fig. 1) was used to generate a double on-line pH gradient. Standard solutions and samples,
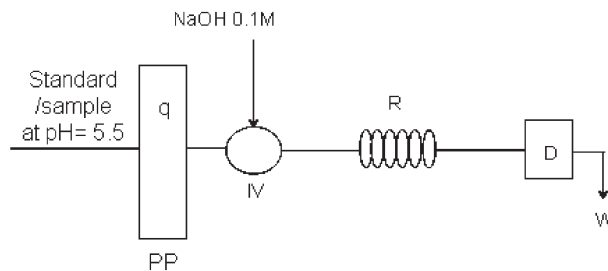
Fig. 1 Flow injection system used for the generation of a double pH gradient: IV, injection valve; R, reactor; D, diode array detector; W, waste; $q = 0.54$ mL min^{-1}.

prepared as described below, were used as a carrier stream, and 50 µL of NaOH 0.1 M were injected to create the pH gradient. For each FIA peak, spectra were collected in the range of 340–750 nm each 2 nm. The number of different points along the pH dimension was 76, taken at intervals of 2 s. In this way, matrices of size 206 × 76 were registered for each injected sample.

Preparation of calibration and validation samples

Pure standard solutions of the four colorants were used for calibration. The standards were prepared from stock solutions of 1.000 g L^{-1}. Six standard solutions of each analyte were prepared with concentrations of 2.0, 4.0, 8.0, 12.0, 16.0 and 20.0 mg L^{-1} to be used in the calibration step, and thus the calibration set consisted of 24 different solutions. Linearity was verified in this concentration range.

A ten sample set was built for validating the employed multivariate model. The analyte concentrations arose from a Plackett–Burman design (Table 1). The extreme concentrations for the design were 4.0 and 8.0 mg L^{-1} for all colorants. They were obtained starting from the stock solutions (see above). Their total spectral-pH evolutions were measured in random order, after injection into the above described FIA system.

Real powder drinks samples

Three real powder drink samples were studied: J1 (containing two colorants: T and A), J2 (three colorants: R, T, and A), and J3 (containing four colorants: I, R, T and A). The samples were prepared by placing 24.0 g of commercial powder in a 250.0 mL flask and completing to the mark with the phosphate buffer solution described above. Each sample was subjected to different dilutions in order to ensure that the final analyte concentrations were within the calibration ranges (this was made by checking that the sample spectra had absorbances of comparable magnitude to those of the calibration samples). The dilutions performed were the following: J1, 1 : 10; J2, 2 : 10 and J3, 3 : 10. Standard additions of the colorants were made in order to perform a recovery study. In this way, in separated samples, three levels of each declared colorant were added. The levels were 2.0, 4.0 and 8.0 mg L^{-1}. This procedure originated six samples for sample J1, nine for sample J2 and twelve for sample J3 (Table 2). Their total spectral evolutions were registered in random order, and in different days than the calibration/test samples.

Table 1 Prediction on model validation samples

Sample	Yellow sunset/mg L^{-1}		Indigo carmine/mg L^{-1}		Allura red/mg L^{-1}		Tartrazine/mg L^{-1}	
	Actual	Pred.	Actual	Pred.	Actual	Pred.	Actual	Pred.
1	4.0	3.8	8.0	8.4	8.0	8.7	8.0	8.4
2	8.0	7.5	8.0	8.5	8.0	8.7	4.0	4.7
3	8.0	7.5	4.0	4.8	8.0	8.7	4.0	4.4
4	8.0	7.5	4.0	4.3	4.0	4.9	8.0	8.2
5	4.0	3.7	8.0	8.1	4.0	4.3	4.0	4.1
6	4.0	3.7	4.0	4.0	8.0	8.3	8.0	8.1
7	8.0	7.5	8.0	8.6	4.0	4.7	8.0	8.5
8	8.0	7.8	4.0	3.9	4.0	4.3	4.0	4.2
9	6.0	5.9	6.0	6.0	6.0	6.3	6.0	6.3
10	6.0	5.8	6.0	6.1	6.0	6.2	6.0	6.2
Mean recovery (%)		94.8		104.7		109.3		105.8

Table 2 Prediction on real samples and real added samples

Sample	Yellow sunset/mg L^{-1}		Indigo carmine/mg L^{-1}		Allura red/mg L^{-1}		Tartrazine/mg L^{-1}	
	Actual[a]	Pred.	Actual[a]	Pred.	Actual[a]	Pred.	Actual[a]	Pred.
J1	—	3.8	—	0.2	—	0.3	—	6.6
J1–T2	—	4.2	—	0.2	—	0.1	8.4	8.2
J1–T4	—	3.4	—	0.2	—	0.2	10.4	10.2
J1–T8	—	3.4	—	−0.3	—	−0.3	14.4	15.5
J1–A2	5.7	5.8	—	0.3	—	0.3	—	5.8
J1–A4	7.7	7.3	—	0.4	—	0.4	—	5.9
J1–A8	11.7	12.7	—	0.8	—	1.1	—	7.3
J2	—	2.3	—	3.1	—	0.4	—	1.6
J2–T2	—	2.2	—	3.4	—	0.4	3.3	3.6
J2–T4	—	2.3	—	3.6	—	0.5	5.3	5.6
J2–T8	—	2.4	—	4.1	—	0.7	9.3	9.7
J2–A2	4.3	4.4	—	2.9	—	0.4	—	1.3
J2–A4	6.3	6.4	—	3.2	—	0.5	—	1.2
J2–A8	10.3	10.7	—	3.0	—	0.7	—	0.7
J2–R2	—	2.2	—	3.2	2.4	2.5	—	1.3
J2–R4	—	2.0	—	3.4	4.4	4.7	—	1.3
J2-R8	—	2.0	—	3.0	8.8	9.4	—	1.1
J3	—	3.1	—	0.1	—	0.7	—	3.8
J3–T2	—	3.1	—	0.3	—	0.9	6.0	5.9
J3–T4	—	3.2	—	0.4	—	1.0	8.0	8.6
J3–T8	—	3.1	—	0.4	—	1.0	12.0	11.5
J3–A2	5.1	5.4	—	0.1	—	0.5	—	3.3
J3–A4	7.1	7.6	—	0.4	—	0.8	—	4.8
J3–A8	11.1	11.7	—	0.0	—	0.8	—	3.7
J3–R2	—	3.0	—	0.0	2.8	2.8	—	3.2
J3–R4	—	3.0	—	0.0	4.8	5.1	—	3.4
J3–R8	—	2.9	—	0.0	8.8	9.6	—	3.4
J3–I2	—	3.3	2.2	2.1	—	0.5	—	3.5
J3–I4	—	3.6	4.2	4.4	—	0.6	—	3.9
J3–I8	—	3.5	8.2	9.2	—	0.9	—	3.9

[a] The actual concentration was computed as the sum of the average value predicted by BLLS on four samples and the amount added in the standard addition procedure (see text).

Data processing

For application of the BLLS algorithm to each test sample, the total absorbance-pH matrices for the calibration set, all of size 206 × 76, were first vectorized in order to apply the mathematical expressions described below. Residual bilinearization was then applied to all samples requiring the second-order advantage. The number of interferents was assessed by increasing a trial number of components until the BLLS prediction residuals stabilize at a value compatible with the instrumental noise level. The latter was estimated by suitable blank replication measurements.

Theory

The original BLLS formulation is discussed in detail in the relevant references, and thus only a brief description is presented here. BLLS employs concentration information into the calibration step, without including data for the unknown sample. This is done in order to estimate pure-analyte matrices at unit concentration (S_n) for specific calibrated analytes (n).

For this purpose, the I calibration data matrices $X_{c,i}$ (each of size $J \times K$) are first vectorized and grouped into a $JK \times I$ matrix V_X:[31]

$$V_X = [\text{vec}(X_{c,1})|\text{vec}(X_{c,2})|\ldots|\text{vec}(X_{c,I})] \qquad (1)$$

where 'vec' implies the unfolding operation. Then a direct least-squares procedure is employed to obtain the pure-analyte information, analogous to first-order classical least-squares multivariate calibration:[31]

$$V_S = V_X Y^{T+} \qquad (2)$$

where Y is an $I \times N_c$ matrix collecting the nominal concentrations of the calibrated analytes, N_c is the number of calibrated analytes, and V_S (size $JK \times N_c$) contains the required S_n matrices in vectorized form:

$$V_S = [\text{vec}(S_1)|\text{vec}(S_2)|...|\text{vec}(S_{N_c})] \qquad (3)$$

To obtain the profiles in both dimensions which are present in the $J \times KS_n$ matrices, two procedures have been discussed in the literature:[10,11] the least-squares (LS) profile estimator and the singular value decomposition (SVD) profile estimator;[10,11] the most reliable and simple seems to be the latter one. In this case, the component profiles are obtained by single-component singular value decomposition (SVD_1) of each of the S_n matrices, obtained after suitable reshaping of the unfolded $\text{vec}(S_n)$:[10,11]

$$(g_n, b_n, c_n) = SVD_1(S_n) \qquad (4)$$

where g_n is the first singular value, and b_n and c_n are the $J \times 1$ and $K \times 1$ left and right eigenvectors of S_n, respectively. This completes the calibration process.

After calibration, concentrations are estimated in test samples by direct least-squares,[10,11] provided no interferents occur in the unknown sample:[31]

$$y_u = S_{cal}^+ \text{vec}(X_u) \qquad (5)$$

where X_u is the $J \times K$ matrix data for the test sample, y_u is an $N_c \times 1$ vector holding the estimated concentrations of the N_c analytes in the sample, and S_{cal} is a calibration $JK \times N_c$ matrix given by:

$$S_{cal} = [g_1(c_1 \otimes b_1)|g_2(c_2 \otimes b_2)|...|g_{N_c}(c_N \otimes b_{N_c})] \qquad (6)$$

where $\otimes$ implies the well-known Kronecker product.

When a single analyte occurs, S_{cal} in eqn (6) contains a single column. However, if the analyte have pH dependent forms, its corresponding S_n will contain information from several equilibrating species, and therefore it may be necessary to consider more than one component for the singular value decomposition in eqn (4). This will render, for a given analyte, different values of g_n, e.g., g_{n1}, g_{n2}, etc., where the first subscript identifies the analyte and the second one the proton equilibrating species. Similarly, different profiles for each dimension (b_{n1}, b_{n2}, c_{n1}, c_{n2}, etc.) will be obtained, which will consist of linear combinations of the true spectral profiles for the equilibrating species of the analyte.

For a single analyte ($n = 1$), an expression similar to eqn (5) is then employed for concentration estimation, where S_{cal} is now given by:

$$S_{cal} = [g_{11}(c_{11} \otimes b_{11})|g_{12}(c_{12} \otimes b_{12})|...|g_{1N_s}(c_{1N_s} \otimes b_{1N_s})] \qquad (7)$$

where N_s is the total number of equilibrating species for the analyte. Hence, even when a single analyte occurs, y_u is in this case an $N_s \times 1$ vector containing the predicted concentration of the calibrated analyte (there are N_s values because each of them is obtained from each of the analyte species).

For more analytes, each of which have equilibrating species, the general equation for S_{cal} will be:

$$S_{cal} = [g_{11}(c_{11} \otimes b_{11})|g_{12}(c_{12} \otimes b_{12})|...|g_{1N_{s1}}(c_{1N_{s1}} \otimes b_{1N_{s1}})|...$$
$$|g_{21}(c_{21} \otimes b_{21})|g_{22}(c_{22} \otimes b_{22})|...|g_{2N_{s2}}(c_{2N_{s2}} \otimes b_{2N_{s2}})|...] \qquad (8)$$

where N_{s1}, N_{s2}, etc. are the number of species associated with each analyte.

The occurrence of uncalibrated compounds in a test sample is investigated after the calibration step is completed. If an interferent occurs, the situation is handled by a separate residual bilinearization (RBL),[10,11] in which profiles for the interferent are found, and incorporated into an expanded version of S_{cal}:

$$S_{exp} = [S_{cal}|g_{int}(c_{int} \otimes b_{int})] \qquad (9)$$

where the profiles b_{int} and c_{int} are obtained by a procedure involving the minimization of the residual matrix E_u, computed while fitting the sample data to the sum of the various component contributions:

$$X_u = \sum_{n'=1}^{N_c} g_{n'} b_{n'} (c_{n'})^T y_{n',u} + g_{int} b_{int} (c_{int})^T + E_u \qquad (10)$$

During the minimization, the interferent profiles are estimated by SVD of a residual matrix E_c, obtained by subtracting the analyte contributions to the total signal X_u:

$$E_c = X_u - \sum_{n'=1}^{N_c} S_{n'} y_{n',u} \qquad (11)$$

$$(g_{int} b_{int} c_{int}) = SVD_1(E_c) \qquad (12)$$

Eqn (10) is suitable for analytes having a single species each. In the event that different analytes have a different number of species, then the latter expression should be modified to accommodate for this possibility:

$$X_u = \sum_{n'=1}^{N_c} \sum_{s'=1}^{N_s} g_{n's'} b_{n's'} (c_{n's'})^T y_{n's',u} + g_{int} b_{int} (c_{int})^T + E_u \qquad (13)$$

An iterative method has been described for applying RBL.[10,11] However, in certain cases the latter procedure has been found to diverge instead of converging to the desired analyte concentrations. We used instead a Gauss–Newton minimization procedure of $\|E_u\|$ in eqn (13), implemented under MATLAB, during which g_{int}, b_{int}, and c_{int} are found by singular value decomposition of E_u. The procedure renders the value of $y_{ns,u}$ required to minimize E_u in eqn (13).

The prevailing idea within BLLS is a two-step calibration–prediction mode, where concentration prediction is guided by a least-squares minimization. The second-order advantage is left for a subsequent stage, in which the matrix residuals are analyzed in order to estimate the interferent profiles. The latter
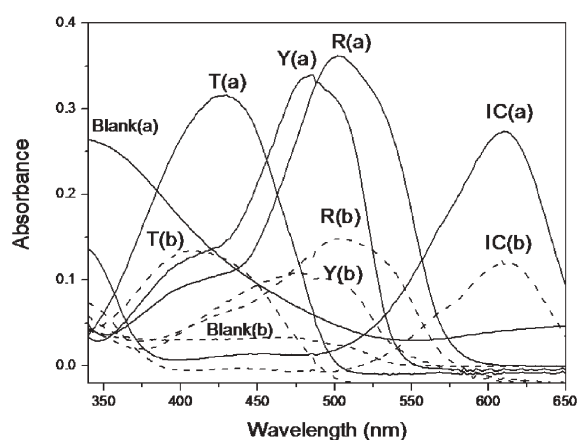
serve to expand the loadings, and to correctly estimate the analyte concentrations, even in the presence of unexpected constituents. Notice that no initialization or constraining procedures are required.

Results and discussion

Spectral behavior of the analytes

Fig. 2 shows the absorption spectra of tartrazine, yellow sunset, indigo carmine and allura red at pH 5.0 and 12.0, where their acid and basic forms can be observed. One of the main difficulties shown by the currently studied system is visually illustrated in Fig. 2: extreme collinearities occur in the spectral dimension, because the acid and basic forms of each analyte display similar spectral shapes. This spectral overlap between analyte species is the main reason complicating the analysis by methods such as PARAFAC or MCR-ALS. The spectrum for a blank powder drink sample (*i.e.*, without colorants) is also shown in Fig. 2 at the pH values selected for the analytes. As can be seen, an intense overlap occurs among the spectra for the analytes and background, a fact which complicates the analysis by first-order multivariate techniques, and highlights the usefulness of the second-order advantage discussed below.

In order to circumvent these difficulties, a suitable methodology for producing second-order data was designed, combining a flow injection system where a double pH gradient is created, and diode array spectrophotometric detection (see below). This type of data permits quantitation even in the presence of unexpected interferents, thanks to the implementation of the powerful second-order advantage. Fig. 3 shows the three-dimensional plot of the total spectral-pH evolution for the validation sample 1 (which contains all four analytes), after injection into the FIA system. The variation in signal along the FIA peak can be readily appreciated, as a decrease in

absorbance near the center of the FIA time dimension, due to the injection of an alkaline solution in the acid carrier stream.
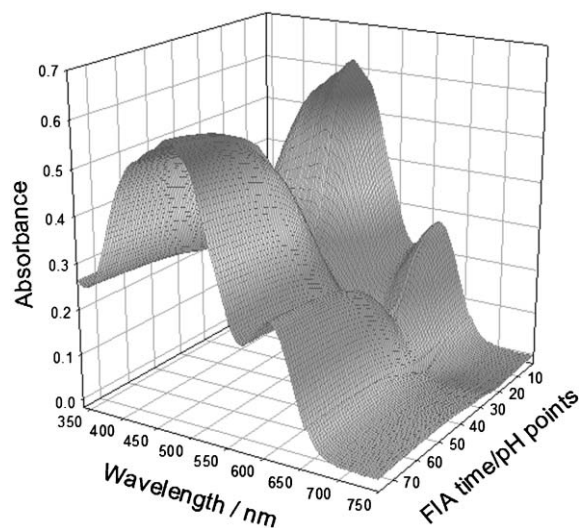
Optimization of the FIA system

In order to check the suitability of the double pH gradient generated in the FIA system, tartrazine solutions prepared with different pH buffers were injected, and the FIA signals were registered at 430 nm. Then, a tartrazine solution of pH 5.5 (the KH_2PO_4/K_2HPO_4 described above) was used as a carrier and a volume of 0.1 M NaOH was injected. In this way, it could be corroborated that a double pH gradient was induced on the carrier stream, with values ranging from 5.5 (the pH of the carrier) to 12.8 close to the center of the FIA peak, to 5.5. It should be noticed that the exact correspondence between pH values and FIA time is not required for the success of the analytical protocol, provided that the generated pH gradient is reproducible from sample to sample.

The optimization of the FIA system was carried out using tartrazine. All the variables were optimized as a compromise between the width and the height of the FIA peak, because the dispersion of the injected volume was responsible for the generation of the pH gradient. As the FIA assembly was simple, the injected volume (30–100 μL), the length of the reactor (1300–1900 mm) and the flow rate (0.36–0.72 ml min^{-1}) were varied, and the optimum values were found to be 50 μL, 1400 mm, and 0.54 ml min^{-1}. The NaOH concentration was also tested between 0.05–0.15 M, and the best was 0.1 M.

Analysis of the synthetic validation set

The use of either PARAFAC (with suitable constraints) or MCR-ALS did not furnish acceptable results for the presently studied samples. The most probable cause is the extreme spectral overlapping between the acid and basic forms of the analytes, which introduces an additional source of linear dependency besides the already existing one in the pH dimension. Therefore, analysis of the set of synthetic samples was carried out by using BLLS. For the successful application

Fig. 2 Solid lines, the four absorption spectra measured for solutions of the pure analytes tartrazine, yellow sunset, allura red and indigo carmine (all at 8.0 ppm), at acid pH (5.0), and indicated as T(a), Y(a), R(a) and IC(a) respectively. Dashed lines, the corresponding spectra at basic pH (12.0), indicated as T(b), Y(b), R(b) and IC(b). Spectra of a solution containing the background of the powder drink are also shown at pH = 5.0 (solid) and pH = 12.0 (dashed), indicated as Blank(a) and Blank(b).

Fig. 3 Three-dimensional plot of the spectral-pH evolution for the validation sample 1, after injection into the FIA system.

of BLLS to the present problem, two parameters should be set: the number of components (*i.e.*, analytes and interferents), and the number of species for each analyte. The number of species per analyte was optimized by trial tests, while the number of interferents was tuned by comparison of the BLLS prediction residuals for each sample with the instrumental noise level (assessed by suitable blank replication). It was found that satisfactory results can be obtained by setting the number of components to four, and the number of species to one for all analytes (see below). The number of components agrees with the known sample composition of the validation set, while the number of species is understandable on the basis of the high spectral collinearity between the acid and basic form of each analyte. In this latter case, the SVD analysis of each of the analyte matrices S_n indicates that the first principal component explains most of the variance, and that no additional components are therefore needed.

The application of BLLS to the validation samples yielded analyte profiles and estimated concentrations. All spectral profiles agree well with those corresponding to standards. Four components and single species for each analyte were included, as suggested by the consideration of BLLS prediction residuals, as shown in Fig. 4 top for a typical validation sample. As can be seen, the model with four components and single analyte species shows a prediction residual comparable to the instrumental noise, while more complex systems only produce a slight decrease in prediction residual which is not justified. Prediction on the synthetic set with the four-component BLLS model yielded the results quoted in Table 2. The latter (recoveries ranged from 94.8 to 109.3%) can be considered as acceptable in view of the complexity of the samples being analyzed. The best results are obtained for IC, as expected since this analyte shows a spectrum which is significantly different in shape in comparison with the remaining three analytes.

Analysis of real powder drink samples

Real powder drink samples and additional ones obtained after analyte standard addition were studied. The analysis of these real samples was made in the manner described above for the synthetic set using BLLS, but using six components instead of four. In this case, the number of chemical species for each analyte was known from the previous analysis of the validation set (*i.e.*, one species for each analyte), while the number of interferents had to be increased to two. This latter number was necessary to decrease the BLLS prediction residuals for each sample, until it stabilizes at a value compatible with the instrumental noise level. Fig. 4B shows that, among the different trial models, the one with six components and a single species for each analyte is the simplest one with a prediction residual comparable to the instrumental level.

Prediction with this BLLS model furnishes the spectral and pH profiles for each component and species, which are shown in Fig. 5A and 5B for a typical real sample. In comparison with Fig. 2, the analyte spectra can be readily matched with those of standards. In the case of the interferent spectral profiles, they consist of principal components of the matrix of prediction residuals, and are thus a linear combination of real spectra. In
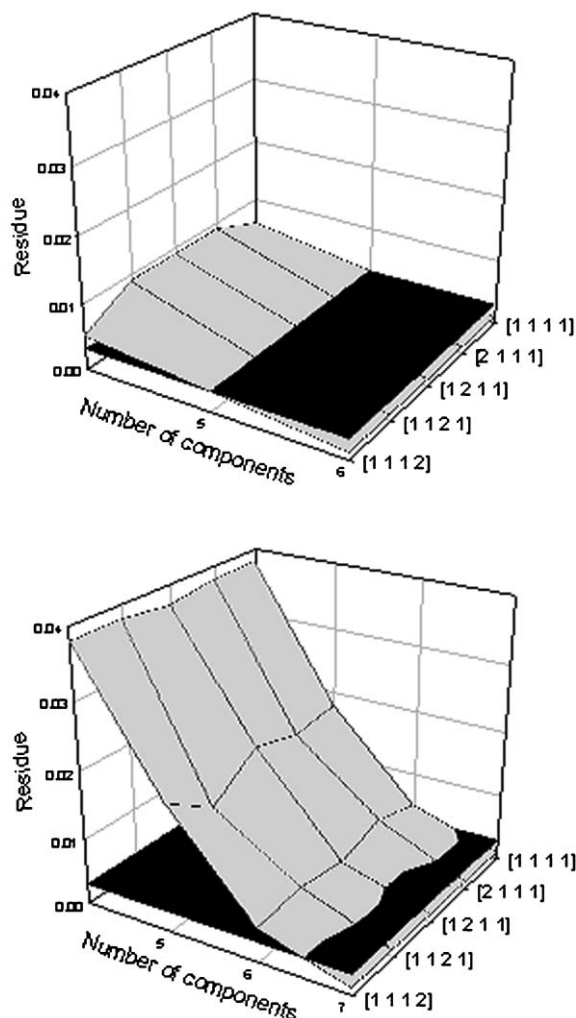
Fig. 4 (Top) Changes in BLLS prediction residuals for a typical validation sample, as a function of the trial number of components, and for five different combinations of equilibrating species for each analyte. The nomenclature is [N_{s1} N_{s2} N_{s3} N_{s4}], where the successive numbers refer to the species for T, Y, R and IC. The horizontal black plane marks the instrumental noise level. (Bottom) Similar to (top) for a typical real sample.

this sense, they are not strictly comparable to those shown for a blank sample in Fig. 2. Nevertheless, BLLS is successful in obtaining the correct analyte profiles to be employed for concentration prediction in complex real samples. The recovered pH profiles shown in Fig. 5B are also indicative of the difficulties found in analyzing the present system: the shapes of the analyte profiles are all similar along the pH dimension, adding to the already strong collinearity in the spectral mode.

The prediction results for the real samples are shown in Table 2. As can be seen, excellent recoveries are obtained in most of the spiked samples. Table 3 shows the obtained probabilities for *t*-paired tests when comparing the amount of added dyes and those predicted by BLLS modeling. As can be seen, only two cases show significant differences (*p* value lower than 0.05). In order to get further insight into the accuracy, linear regression analysis of added *versus* predicted
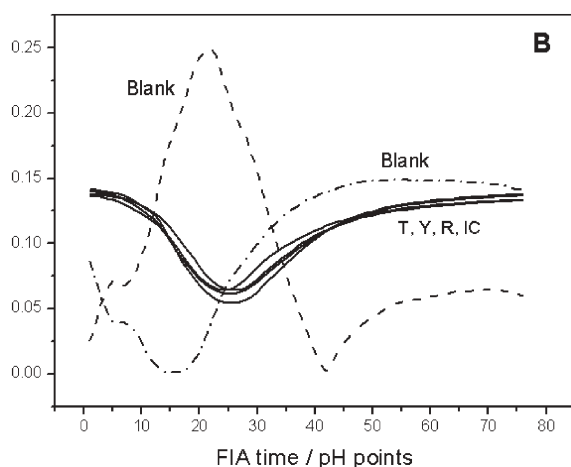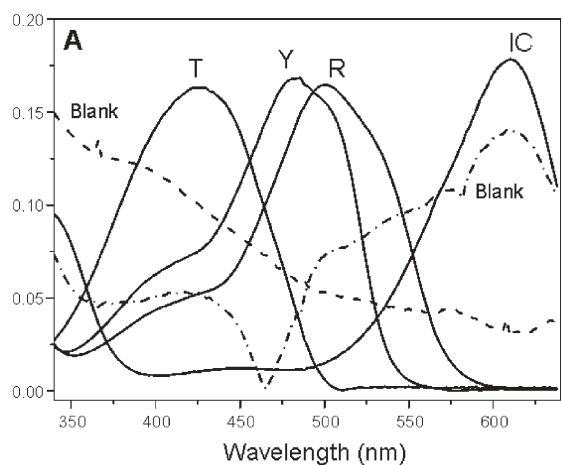
Fig. 5 BLLS extracted spectral (A) and pH (B) profiles for each component in a typical real sample. Analytes can be identified as tartrazine (T), yellow sunset (Y), allura red (R) and indigo carmine (IC). The dotted and dash-dotted lines correspond to the interferents. In the case of the interferents, the absolute values of the profiles have been plotted.

concentration values (Table 2) for yellow sunset and tartrazine were applied on the three real spiked samples. The estimated intercept and slope ($\hat{a}$ and b respectively) were compared with their ideal values of 0 and 1 using the elliptical joint confidence region (EJCR) test, in this case by using an ordinary least squares fitting (OLS).[32] The ellipses are seen to contain the theoretical ($a = 0$, $b = 1$) point for both analytes. This fact is indicative that constant and proportional bias is absent. On the other hand, more precision for the estimation of the yellow sunset concentration can be observed considering the size of the corresponding ellipses. In addition, the analysis of four

Table 3 Probability obtained for *t*-paired tests analyzing predictions on real added samples

Sample	Yellow sunset	Indigo carmine[a]	Allura red[a]	Tartrazine
J1	0.63	—	—	0.64
J2	0.18	—	0.15	0.01
J3	0.03	0.38	0.26	0.90

[a] No addition was performed.

replicates showed that the largest coefficient of variation (CV) was near to 10%. Similar CV values were found for the analyte concentrations in real samples without standard addition.

Conclusions

Second-order data generated in various ways provide, after adequate chemometric analysis, valuable information concerning underlying physical phenomena. More importantly, they permit quantitation of selected analyes in complex multi-component samples, even in the presence of uncalibrated constituents. The presently analyzed example challenges the most advanced second-order multivariate methodologies with data showing strong spectral overlapping coupled to linear dependencies in the pH dimension. Bilinear least-squares appears to be a promising technique for processing these data, yielding analyte profiles and concentrations in samples where a strong background occurs.

Acknowledgements

References

1 N. M. Faber, R. Bro and P. K. Hopke, *Chemom. Intell. Lab. Syst.*, 2003, **65**, 119–137.
2 K. S. Booksh and B. R. Kowalski, *Anal. Chem.*, 1994, **66**, 782A–791A.
3 J. Saurina, S. Hernández-Cassou, R. Tauler and A. Izquierdo-Ridorsa, *Anal. Chem.*, 1999, **71**, 2215–2220.
4 J. Saurina, *Rev. Anal. Chem.*, 2000, **19**, 157–178.
5 J. Diewok, A. De Juan, R. Tauler and B. Lendl, *Appl. Spectrosc.*, 2002, **56**, 40–50.
6 E. Sanchez and B. R. Kowalski, *Anal. Chem.*, 1986, **58**, 496–499.
7 R. Bro, *Chemom. Intell. Lab. Syst.*, 1997, **38**, 149–171.
8 C. M. Andersen and R. Bro, *J. Chemom.*, 2003, **17**, 200–215.
9 R. Tauler, *Chemom. Intell. Lab. Syst.*, 1995, **30**, 133–146.
10 M. Linder and R. Sundberg, *Chemom. Intell. Lab. Syst.*, 1998, **42**, 159.
11 M. Linder and R. Sundberg, *J. Chemom.*, 2002, **16**, 12.
12 R. Bro, *Multi-way analysis in the food industry*, PhD Thesis, University of Amsterdam, Netherlands, 1998.
13 R. Bro, R. Harshman and N. D. Sidiropoulos, *J. Chemom.*, in press.
14 M. M. Reis, S. P. Gurden, A. Smilde and M. M. C. Ferreira, *Anal. Chim. Acta*, 2000, **422**, 21–36.
15 H. C. Goicoechea and A. C. Olivieri, *Appl. Spectrosc.*, 2005, **59**, 67–74.
16 For parafac, http://www.models.kvl.dk/source/, for MCR-ALS, http://www.ub.es/gesq/mcr/mcr.htm.
17 R. Combes and R. Haveland-Smith, *Mutat. Res.*, 1982, **98**, 101.
18 European Union, Directive 94/36/EC on colours for use in foodstuffs, 1994.
19 M. S. García Falcón and J. Simal Gándara, *Food Control*, 2005, **16**, 293–297.
20 Q. Chen, S. Mou, X. Hou, J. M. Riviello and Z. Ni, *J. Chromatogr. A*, 1998, **827**, 73–81.
21 M. Holcapek, P. Bandera and P. Zderadicka, *J. Chromatogr. A*, 2001, **926**, 175–186.
22 C. Scarpi, F. Ninci, M. Centini and C. Anselmo, *J. Chromatogr. A*, 1998, **796**, 319–325.

23 H. Huang, Y. Shih and Y. Chen, *J. Chromatogr. A*, 2002, **959**, 317–325.
24 M. Perez-Urquiza and J. L. Beltrán, *J. Chromatogr. A*, 2000, **898**, 271–275.
25 E. Dinc, E. Baydan, M. Kanpur and F. Onur, *Talanta*, 2002, **58**, 579–594.
26 J. J. Berzas Nevado, J. Rodríguez Flores, M. Villaseñor Llerena and N. Rodríguez Fariñas, *Talanta*, 1999, **48**, 895–903.
27 S. Chanlon, L. Joly-Pottuz, M. Chatelut, O. Vittori and J. L. Cretier, *J. Food Compos. Anal.*, 2005, **18**, 503–515.
28 J. J. Berzas Nevado, J. Rodríguez Flores and M. J. Villaseñor Llerena, *Talanta*, 1997, **44**, 467–474.
29 K. Harada, K. Masuda, M. Suzuki and H. Oka, *Biol. Mass Spectrosc.*, 1991, **20**, 522–528.
30 M. Perez-Urquiza and J. L. Beltrán, *J. Chromatogr. A*, 2001, **917**, 331–336.
31 N. M. Faber, J. Ferré, R. Boqué and J. H. Kalivas, *Chemom. Intell. Lab. Syst.*, 2002, **63**, 107.
32 A. G. González, M. A. Herrador and A. G. Asuero, *Talanta*, 1999, **48**, 729–736.