JOURNAL OF CHEMOMETRICS, VOL. 12, 55–75 (1998)

MULTIVARIATE CURVE RESOLUTION APPLIED TO THREE-WAY TRILINEAR DATA: STUDY OF A SPECTROFLUORIMETRIC ACID–BASE TITRATION OF SALICYLIC ACID AT THREE EXCITATION WAVELENGTHS

R. TAULER,* I. MARQUÉS AND E. CASASSAS
Department of Analytical Chemistry, University of Barcelona, Avda. Diagonal 647, E-08028 Barcelona, Spain

SUMMARY

Application of multivariate curve resolution (MCR) is shown for three-way data obtained in acid–base titrations of salicylic acid monitored using emission spectrofluorimetry at three different excitation wavelengths. Rank analysis of augmented column- and row-wise matrices showed that the experimental fluorescent data have a trilinear data structure. MCR allows the unambiguous recovery of the species profiles in both orders: the fluorescent emission spectrum at each excitation wavelength and the pH distribution profile of the two fluorescent species formed during the acid–base titration. Results are compared with those obtained using the trilinear decomposition and generalized rank annihilation methods. Using the mass balance equation for the total concentration of salicylic acid, the third pH distribution profile of the non-fluorescent fully protonated salicylic acid species is deduced. The values of the two acid dissociation constants of salicylic acid, for the carboxylic and hydroxylic groups, are estimated without using the mass action law and are compared with previous estimations in the literature and with estimations using traditional mass action law hard-modelling-based methods. © 1997 John Wiley & Sons, Ltd.

J. Chemometrics, Vol. **12,** 55–75 (1998)

KEY WORDS multivariate curve resolution; three-way data analysis; spectrofluorimetry; salicylic acid; dissociation constants

INTRODUCTION

When a titration of a sample is monitored using emission spectrofluorimetry at different excitation wavelengths, a set of different correlated data matrices is obtained. Grouping these correlated data matrices gives a three-way data array or third-order tensor. This situation is similar to that encountered in the analysis of a set of related samples (at different concentrations) by a hyphenated chromatographic method, e.g. liquid chromatography with multiwavelength diode array detection,[1] or when several runs of a chemical process are monitored spectroscopically at multiple wavelengths under different initial concentrations of the reagents.[2] However, the internal data structure of all these three-way data arrangements can be different.

A data matrix contains information about two vector spaces or measurement orders: one is the row space or row order (vector space spanned by the rows of the data matrix) and the other is the column space or column order (vector space spanned by the columns of the data matrix). In both vector spaces or orders of measurement the data contain a structural part and noise; the former is usually described by a reduced set of linearly independent vectors. Every chemical species is described by a species

* Correspondence to: R. Tauler. Email: roma@quimio.qui.ub.es

CCC 0886–9383/98/010055–21 $17.50

Received 15 December 1996
Accepted 11 August 1997

profile (vector) in each of the two orders of measurement, usually a species spectrum and a concentration profile (changes in the concentration of the different chemical constituents). In many chemical situations, when several individual data matrices are compared, at least one of the two orders of measurement in the simultaneously analysed data matrices has a common structure, i.e. a common subset of basis vectors spans the row or column spaces (or subspaces) of the different data matrices. This common structure is present many times in the spectral order, i.e. common species with equal spectra in the different experiments (data matrices). In a smaller number of cases the common structure is present in both orders of the simultaneously analysed matrices, i.e. the two vector spaces (rows and columns) of all the simultaneously analysed data matrices are spanned by the same set of basis vectors. This particular three-way data structure is called 'trilinear' and a well-known example of such a situation is the case of excitation–emission fluorescence spectroscopy where two spectral orders are available.[3] It can also be the case for some types of kinetic or evolutionary chemical reactions or processes monitored spectrometrically.[4] For three-way data arrays, trilinearity[5] is a desired situation, since under such a circumstance the rotational factor analysis ambiguities[6] inherent to the factor analysis decomposition of two-way data arrays are totally solved and recovery of species profiles in both orders of measurement can be achieved. This has important consequences in mixture analysis where the problem of quantitation of an analyte in a real sample, in the presence of unknown and uncalibrated interferences, can be solved without sample pretreatment or analyte separation (i.e. second-order advantage[7]). Total trilinear data structure is, however, a highly demanding requirement. For instance, it is not normally fulfilled in most practical applications of hyphenated chromatography, where usually the elution time order is not common between runs, since the shapes of elution profiles of coeluted components vary between different chromatographic runs.[8] In kinetic reactions only under special conditions (first- and pseudo-first-order reactions) and for some of the component species is the trilinearity constraint totally fulfilled.[9] One case where trilinearity is easily achieved is in fluorescence spectroscopy, since in that case there are two highly reproducible spectral orders, excitation and emission; spectral wavelength values and species spectra under the same physical conditions (temperature, ionic strength, solvent composition) are highly reproducible.

As is proved in the present work, a chemical reaction monitored using fluorescent emission spectra obtained at different excitation energies is also a good example of highly structured trilinear data. In the case of an acid–base titration monitored by fluorescent emission spectrometry at different excitation wavelengths, each excitation wavelength provides a data matrix of fluorescent emission spectra recorded at every pH. The information about the changes in concentration of the chemical species during the acid–base titration is modulated in the spectral order by the different excitation wavelengths. More importantly, each particular species yields, at the different excitation wavelengths, fluorescent spectra which show the same spectral bands with different relative intensities. The changes in band intensity depend on both the excitation wavelength and the chemical nature of the particular species. All this agrees with a trilinear data structure.

Several methods have been proposed to solve the components present in unknown mixtures from trilinear three-way data arrays. Among these methods are the generalized rank annihilation method (GRAM)[7] and its extension the trilinear decomposition (TLD) method.[5] These are eigenvalue–eigenvector decomposition-based methods and provide fast non-iterative solutions; they require, however, strict fulfilment of the trilinearity constraints. Alternative methods based on Tucker decomposition[10] have also been proposed; they do not require such strong trilinearity requirements but are iterative and slower to achieve convergence. Multivariate curve resolution (MCR) based on an alternating least squares (ALS) optimization algorithm has been shown to be a powerful tool for resolving two- and three-way data arrays.[11] The main advantage of this method is that it is easily adapted to data sets of different complexity and structure, trilinear or non-trilinear, providing optimal least squares solutions. Although iterative, when the appropriate initial estimations and constraints are

applied, the ALS implementation of the MCR method is also reasonably fast. In the present work the results obtained by ALS MCR are compared with those obtained by GRAM and TLD.

The study of changes in the fluorescence properties of salicylic acid along an acid–base titration is interesting because of its possible application in the study of other related compounds of great environmental importance, such as humic substances, which show a very complex structure and behaviour and contain salicylic-like chemical groupings.[12, 13] Salicylic acid shows two acid–base equilibria: the deprotonation of a carboxylate group and the deprotonation of a hydroxyl group. Three species, H_2sal, $Hsal^-$ and sal^{2-}, can be present in the pH range 0–14. The study of the acid–base equilibria of salicylic acid using fluorescence is complicated, however, because the fully protonated species H_2sal, which participates in the first deprotonation equilibrium, is not a fluorescent one and because of the very weak acidity of the hydroxyl group participating in the second equilibrium.[14] It is not easy to obtain a trustworthy pK_a value for the deprotonation of the hydroxyl group in aqueous medium, as is evident from the lack of agreement among published values of this pK_a. Its high value makes impossible the use of the very precise potentiometric method for its determination in aqueous media. In previous work[15] a continuous spectrophotometric titration of ligand solutions with absorbance readings at a single analytical wavelength (326 nm) and a traditional least squares curve-fitting approach for the treatment of data were employed to find this value. In the present work a continuous spectrofluorimetric titration method was preferred, because the differences between the fluorescent emission spectra of the individual species are sharper than those shown by the UV absorption spectra and also because the recorded multiwavelength fluorescent emission spectra yield a richer amount of information than readings at a single wavelength as obtained in previous work. The pK_a values determined previously:[14] at 25 °C and 0·1 M ionic strength were 2·85 and 13·4 respectively for the two acid dissociation equilibria.

Traditional methods for stability constant determination[14, 15] make use of the mass action law. In the present work an alternative way is proposed which does not use the mass action law constraint. Instead, the stability constants are estimated from the concentration profiles of the species in equilibrium obtained through the MCR method, since for data trilinear in nature the rotational ambiguities can be totally solved without making use of the mass action law. The obtained values are compared with previous estimations in the literature and with estimations obtained in this work by using traditional mass action law-based methods.

EXPERIMENTAL

In the acid–base titration of salicylic acid, three chemical species are formed according to the pH value: at pH below 3 the fully protonated species (H_2sal), at intermediate pH the monoprotonated species ($Hsal^-$) and at pH above 12 the dianion (sal^{2-}) are predominant. According to previous studies,[16] only two of these species are fluorescent, namely those containing the carboxylate group together with a protonated or a deprotonated hydroxyl group, i.e. $Hsal^-$ and sal^{2-} respectively. The titration is monitored spectrofluorimetrically; that is, at each measured pH, three fluorescent emission spectra are recorded using three different excitation wavelengths. In order to test the proposed method, the experimental data set was compared with a data set obtained from a simulation process.

Experimental data

A solution containing $9{\cdot}96\times10^{-5}$ mol l^{-1} salicylic acid at constant ionic strength from 0·1 mol l^{-1} $NaNO_3$ was titrated by adding small amounts of acid or base to change the pH in the range 1·17–13·60. Thirty-six titration points were selected for data treatment. The titration vessels were thermostatted at $25{\cdot}0\pm0{\cdot}1$ °C. At each pH, fluorescent emission spectra were recorded between 250

and 650 nm (0·5 nm resolution, scan rate 200 nm min^{-1}, 15 nm excitation and emission slit width) at three excitation wavelengths: 280, 297 and 314 nm. From these spectra, 351 wavelengths between 350 and 525 nm were selected for data treatment. With the experimental data a matrix is built for each excitation wavelength, with as many rows as pH values measured (*NP*=36) and as many columns as wavelengths monitored in the fluorescent spectra (*NW*=351). In Figure 1 the mesh 3D plots of the three analysed data matrices ($\mathbf{D}_4$ for the excitation wavelength $\lambda_{exc}=280$ nm, $\mathbf{D}_5$ for $\lambda_{exc}=297$ nm and $\mathbf{D}_6$ for $\lambda_{exc}=314$ nm) are given.

These matrices were used to estimate the real fluorescent emission species spectra, pH species profiles and acid–base dissociation constants of the salicylic acid (pK_{a_1} and pK_{a_2}) as well as to test the assumptions about the nature and structure of the experimental data.

Simulated data

A data set was simulated using the previous knowledge of the system in order to validate the MCR method, i.e. to investigate how well the proposed method recovers fluorescent emission spectra and pH distribution profiles of the species, as well as the pK_a values postulated initially. Three simulated data matrices were obtained from a linear model

$$\mathbf{D}_k=\mathbf{S}_k\mathbf{C}^T+\mathbf{N} \tag{1}$$

where $\mathbf{D}_k(NW,NP)$ is the simulated data matrix at one of the three excitation wavelengths under study: $\lambda=280$ nm for k=1, $\lambda=297$ nm for k=2 and $\lambda=314$ nm for k=3; *NW* is the number of wavelengths of the pure species spectra selected for analysis (*NW*=351 as in the experimental data); *NP* is the number of pH values in the titration, equal to the number of spectra simulated for each excitation wavelength, selected for analysis (*NP*=131, whereas in the experimental treatment only 36 points were selected).

Matrices $\mathbf{D}_1$, $\mathbf{D}_2$ and $\mathbf{D}_3$ were evaluated from the simulation of spectral matrices $\mathbf{S}_k$ and concentration matrices $\mathbf{C}^T$ plus a noise term $\mathbf{N}$. $\mathbf{S}_k(NW,NS)$ is the matrix of the *NS* fluorescent emission spectra of the pure species at excitation wavelength k, prepared from theoretical spectra (roughly estimated from a preliminary analysis of the experimental data). For trilinear data,

$$\mathbf{S}_k=\mathbf{S}\mathbf{K}_k \tag{2}$$

where $\mathbf{K}_k$ is a diagonal matrix with the relative intensities of the emission spectra at each excitation wavelength k in its diagonal and $\mathbf{S}$ is the matrix of the unitary pure emission species spectra (common shape).

$\mathbf{C}^T(NS,NP)$ is the concentration matrix of the *NS* fluorescent species, obtained from the pH species distribution for preselected pK_a values of deprotonation equilibria. The same matrix $\mathbf{C}^T$ is used to build the three matrices $\mathbf{D}_k$. The species distribution was calculated from the solution of the mass balance equations using the non-linear Newton–Raphson method with a damping constraint[17] and assuming fulfilment of the mass action law for values of the logarithm of acid dissociation constants equal to 2·85 and 13·4. The matrix $\mathbf{N}$ is a matrix of Gaussian noise, of zero mean and a standard deviation of 0·5% of the maximum fluorescent intensity.

Reagents

Sodium salicylate (Carlo Erba) and other analytical reagent grade chemicals were used. All solutions were made with deionized CO_2-free water and adjusted to an ionic strength of 0·1 mol l^{-1} with $NaNO_3$.

Titrant solutions were prepared from sodium hydroxide and hydrochloric acid (Merck).

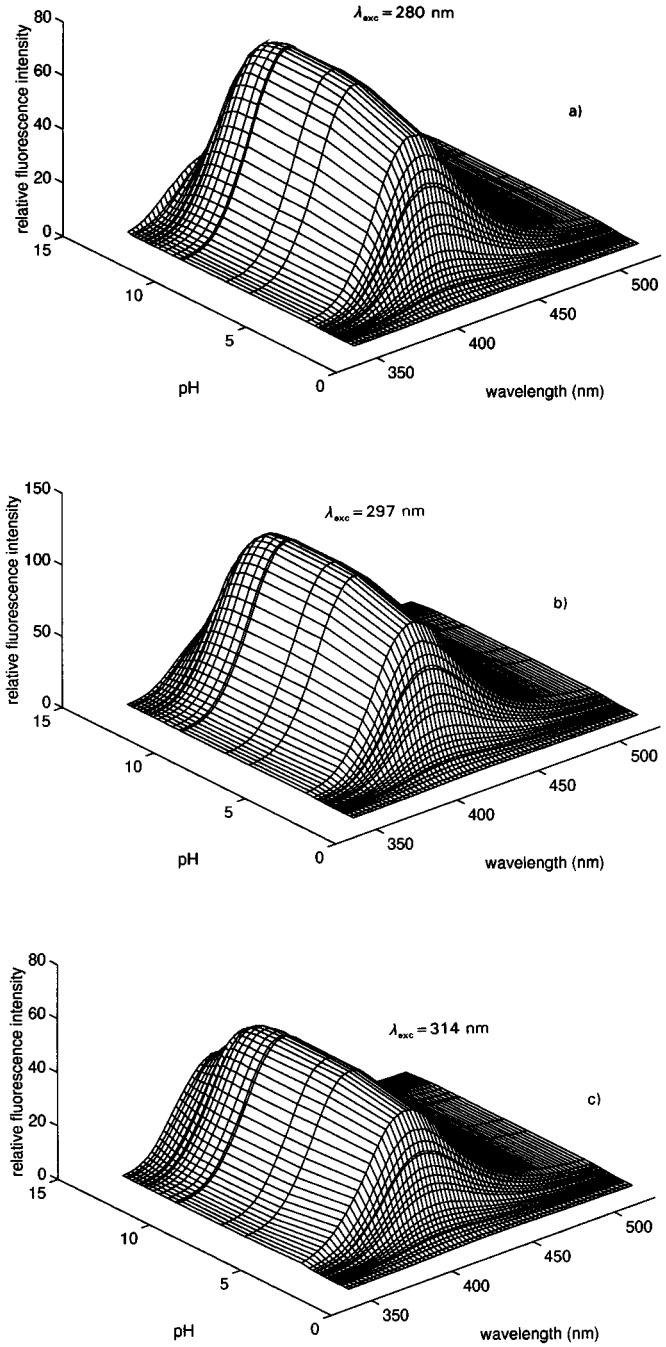

Figure 1. 3D plots of experimental fluorecent emission spectrometric titration data of a salicylic acid solution ([salicyclic acid] = $9{\cdot}96\times10^{-5}$ M) at excitation wavelengths of (a) 280, (b) 297 and (c) 314 nm (pH range 1·17–13·60)

Apparatus

The fluorescence spectra were recorded at each pH using a Perkin Elmer LS 50 luminescence spectrometer equipped with a Hellma 176·752 fused silica flow cell (25 μl inner volume). Acid–base titrations were performed through capillary additions of very small amounts of titrant in order to maintain essentially constant the analytical concentration of the titrand along the whole titration. The EMF readings leading to pH values were made through an Orion 720 pH-meter and a Ross–Orion combined electrode.

The system was organized in a closed loop circuit with continuous flow of the 25 °C thermostatted solutions using a Gilson Minipuls II peristaltic pump.

METHOD

Data structure

Every individual data matrix ($\mathbf{D}_1$, $\mathbf{D}_2$ and $\mathbf{D}_3$ in the simulated study; $\mathbf{D}_4$, $\mathbf{D}_5$ and $\mathbf{D}_6$ in the experimental study) has *NW* rows (number of wavelengths) and *NP* columns (number of pH values or spectra measured). The three data matrices obtained in the simulated study or the three in the experimental study can be arranged in a three-way data array structure (Figure 2). Each of these three-way data arrays can also be unfolded in an augmented row-wise or column-wise two-way data matrix (Figure 2). In reduced notation, $[\mathbf{D}_1;\mathbf{D}_2;\mathbf{D}_3]$ and $[\mathbf{D}_4;\mathbf{D}_5;\mathbf{D}_6]$ are the two augmented column-wise data matrices which have $3\times NW$ rows and *NP* columns. They are built from individual data matrices by setting one on top of the other and keeping the column vector space in common. $[\mathbf{D}_1, \mathbf{D}_2, \mathbf{D}_3]$ and $[\mathbf{D}_4, \mathbf{D}_5, \mathbf{D}_6]$ are the two augmented row-wise matrices which have NW rows and $3\times NP$ columns. They are built from individual data matrices by setting one beside the other and keeping the row vector space in common.

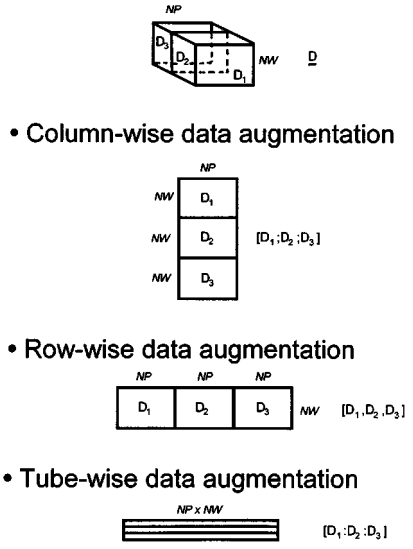

Figure 2. Simultaneous analysis of a set of correlated data matrices by matrix augmentation

Rank analysis and data structure

If the data are trilinear, the data matrices obtained at different excitation wavelengths must have row and column vector spaces in common. The pH distributions of the species in the three data matrices are exactly the same since they belong to the same acid–base titration at the same pH values. The pure emission spectra for the same fluorescent species at the three different excitation wavelengths should have the same shape and only differ in a scaling factor. The ratio of the scaling factors of these fluorescent emission spectra for each species in the different data matrices is unique for each species. Under these assumptions the condition of trilinearity is fulfilled and the singular value decomposition of the augmented data matrices, row- and column-wise, must give the same number of principal components, which should also be equal to the total number of species present in the system. In the case of no rank increase by matrix augmentation a trilinear data structure[8] is confirmed.

Evolving factor analysis (EFA)[18] and derived methods[19] provide information about the local rank[20] of a data matrix. This information is especially useful to check the conditions for complete resolution of a data matrix, as has been pointed out by Manne.[21] Moreover, EFA of a data matrix can provide an initial estimation of the distribution (concentration profiles) of the fluorescent species.

ALS MCR of individual data matrices

Experimental data are described by a bilinear model such as given by equation (1). The complete resolution of individual data matrices depends mostly on the presence of pure variables[22] or selectivity and on the local rank structure of the data matrix.[21] Equation (1) is solved iteratively for $\mathbf{S}_k$ and $\mathbf{C}^{\mathrm{T}}$ by means of an alternating least squares algorithm using natural constraints:[11] such as non-negativity, closure and unimodality to decrease the possibility of rotational ambiguities and to provide a physically reliable optimization path. These constraints are applied externally after each new pseudoinverse estimation of $\mathbf{S}_k$ and $\mathbf{C}^{\mathrm{T}}$. Therefore during the optimization the constraints modify the least squares pseudoinverse estimations and the constrained solutions are not truly least squares solutions. However, in practice, when the optimization advances and becomes closer to convergence, the constraints are not active any more and the estimations become truly least squares solutions. In Table 1 the steps of the ALS procedure are summarized.

ALS MCR of augmented data matrices (Three-way Data Analysis)

The bilinear model derived from a generalized linear equation is simultaneously applied to a set of correlated matrices using

$$\begin{bmatrix} \mathbf{D}_1 \\ \mathbf{D}_2 \\ \dots \\ \dots \\ \mathbf{D}_k \end{bmatrix} = \begin{bmatrix} \mathbf{S}_1 \\ \mathbf{S}_2 \\ \dots \\ \dots \\ \mathbf{S}_k \end{bmatrix} \mathbf{C}^{\mathrm{T}} + \begin{bmatrix} \mathbf{N}_1 \\ \mathbf{N}_2 \\ \dots \\ \dots \\ \mathbf{N}_k \end{bmatrix} \tag{3a}$$

or

$$[\mathbf{D}_1;\mathbf{D}_2;\ldots;\mathbf{D}_k]=[\mathbf{S}_1;\mathbf{S}_2;\ldots\mathbf{S}_k]\mathbf{C}^{\mathrm{T}}+\mathbf{N} \tag{3b}$$

where $[\mathbf{D}_1;\mathbf{D}_2;\ldots;\mathbf{D}_k]$, $[\mathbf{S}_1;\mathbf{S}_2;\ldots;\mathbf{S}_k]$, $\mathbf{C}^{\mathrm{T}}$ and $\mathbf{N}$ are respectively the augmented data, the augmented species spectra, the common concentration profiles (species distribution) and noise matrices. Because of the data arrangement for column-wise augmented data matrices $[\mathbf{D}_1;\mathbf{D}_2;\ldots;\mathbf{D}_k]$, the same set of

Table 1. Steps of alternating least squares multivariate curve resolution procedure

Step	Comments
1	Building up of data matrices $\mathbf{D}_k(NW,NP)$, k=1,2,3 (simulated) and k=4,5,6 (experimental). Building up of augmented column-wise matrices $\mathbf{D}_{aug}$ (either $[\mathbf{D}_1;\mathbf{D}_2;\mathbf{D}_3]$ simulated or $[\mathbf{D}_4;\mathbf{D}_5;\mathbf{D}_6]$ experimental) (see Figure 2 and Experimental for notation).
2	Investigation of the chemical rank of the individual $\mathbf{D}_k$ and augmented $\mathbf{D}_{aug}$ matrices by SVD and PCA. Estimation of the number of fluorescent species. Reproduction of the experimental data matrices for the considered number of principal components.
3	Investigation of pure variables[22] and local rank structure of data matrices by EFA-derived methods.[18,19] Initial estimation of concentration profiles (species distribution) of active fluorescent species.
4	ALS optimization of species profiles in both orders of measurement.[2,4,8,11,12] The basic equations used during the ALS optimization when aplied either to individual or augmented matrices are $\mathbf{C}^T=\mathbf{S}_k^+\mathbf{D}_k^*$ or $\mathbf{C}^T=\mathbf{S}_{aug}^+\mathbf{D}_{aug}^*$; $\mathbf{S}_k=\mathbf{D}_k^*(\mathbf{C}^T)^+$ or $\mathbf{S}_{aug}=\mathbf{D}_{aug}^*(\mathbf{C}^T)^+$ where D_k^* and $\mathbf{D}_{aug}^*$ are the reproduced individual data matrix and the reproduced augmented data matrix respectively for the considered species number. $\mathbf{C}^T$ is the matrix of pH distribution profiles in both type of treatments (analysis of individual and augmented data matrices). $\mathbf{S}_k$ and $\mathbf{S}_{aug}$ are the matrices of fluorescent emission spectra of the pure species in the analysis of the individual k data matrix and in the analysis of the augmented data matrix respectively. The superscript + in the matrices $\mathbf{C}^{T+}$, $\mathbf{S}_k^+$ and $\mathbf{S}_{aug}^+$ refers to the corresponding pseudoinverse matrices.[26] These equations are solved iteratively; at each iteration the following set of natural constraints can be applied. (i) *Selectivity* — In the pH range where only one species is detected by local rank analysis,[11,18–20] the concentration of the other species is forced to be zero. (ii) *Non-negativity* — The concentration and/or spectra should be non-negative for all species. (iii) *Unimodality* — The concentration profiles of all the detected species should be unimodal. An initial unimodality tolerance parameter r is set on starting the optimization; r=0 for non-unimodality; r=1 for strict total unimodality; r=1·1 for 10% of local unimodality departure allowed. When during the optimization the profile becomes non-unimodal, then the profile is corrected. (iv) *Closure* — This constraint could not be applied in all the pH range because of the presence of a non-fluorescent species (H_2sal) in equilibrium with the other fluorescent species ($Hsal^-$ and sal^{2-}). (v) *Trilinearity* — This constraint is shown in Figure 3 and explained in Method. Whereas in the analysis of the individual data matrices $\mathbf{D}_k$ the constraints (i)–(iv) can be applied, in the analysis of the augmented data matrices $\mathbf{D}_{aug}$ the trilinearity constraint (v) can also be optionally applied.
5	Convergence is tested from the root mean square of residuals (σ) between the PCA-reproduced data matrices (for the particular number of species considered) and the ALS-reproduced data matrices during the optimization. When the relative changes in σ are smaller than a certain initially proposed convergence criterion (e.g. 0·01%), local convergence is considered to be achieved.
6	Data fitting is evaluated from the percentage of lack of fit experimental data (see footnotes to Table 3).
7	Validation of the ALS solutions is performed, for simulated data or for a real system with known solutions, by calculation of the correlation or similarity between the ALS-recovered species profiles and the true species profiles (see Results and Table 4).

basis vectors in $\mathbf{C}^T$ spans the common column space of all the individual data matrices $\mathbf{D}_k$ included in the augmented data matrix **D**. This is in agreement with the fact that the same chemical species are present in the different individual data matrices and that they should have exactly the same pH profiles. On the other hand, and in accordance with molecular luminescence spectroscopy theoretical considerations,[16] the emission spectra of the same species obtained at different excitation wavelengths should have the same shape and differ only in a scaling factor. This means that the row space of each data matrix $\mathbf{S}_k$ is also spanned by the same set of basis vectors (see equation (2) and therefore the data should be trilinear.

According to previous results obtained by several authors,[1,3,5,7,23,24] the analysis of trilinear data is optimally performed by methods based on the resolution of a generalized eigenvalue–eigenvector problem, such as GRAM[7] and TLD.[5] Although TLD and GRAM are fast and provide unique solutions, they cannot be used for non-trilinear data and do not allow the input of a different type of data structure nor of the known chemical information to prevent meaningless solutions. On the other side, iterative ALS methods have been usually proposed for the analysis of non-trilinear three-way data[2,4,9–11] which cannot be analysed by TLD or GRAM. In this work the possibilities of the ALS MCR method[2,4,8,9,11] when it is applied to trilinear data are explored and the results obtained are compared with those obtained by GRAM and TLD when they are applied to the same data set.

The ALS analysis of augmented column-wise matrices (three-way data) involves not only the adaptation of the algorithm and constraints used in the ALS analysis of individual data matrices (see steps in Table 1) to column-wise augmented matrices, but also the inclusion of a trilinearity constraint. This is of special relevance in the present work to check the performance of the proposed ALS MCR method when applied to trilinear data. A graphical description of this constraint is shown in Figure 3 for the case of the simultaneous analysis of the simulated data; the same holds for the experimental data in the case of trilinearity. The implementation of this constraint in the ALS MCR algorithm is totally different from that in other ALS-based methods.[10] Every column of the augmented resolved species spectra matrix $[\mathbf{S}_1;\mathbf{S}_2;\mathbf{S}_3]$ $(3\times NW,NS)$ is folded in a new reduced matrix **SR**$(NW,3)$. This matrix is decomposed by principal component analysis (PCA). Since the three resolved species spectra found in **SR** should have the same shape, only the first principal component is considered, giving one score vector **u**$(NW,1)$ and one loading vector $\mathbf{v}^T(1,3)$. The score **u** has the common shape structure of the ALS spectra currently estimated and the loading $\mathbf{v}^T$ has the relative intensity ratios of the spectra in the three different matrices obtained at the three different excitation wavelengths. The product of the score and loading vectors gives a new estimation of the folded matrix **SR*** containing the three filtered (only common shape) resolved species spectra. Finally, this matrix **SR*** is unfolded to restore

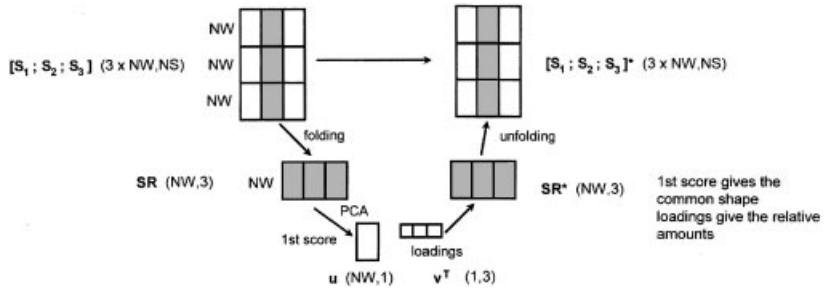

Figure 3. Implementation of trilinearity constraint. The three spectra in the second column of the augmented species spectra matrix $[\mathbf{S}_1;\mathbf{S}_2;\mathbf{S}_3]$ are folded to give the matrix **SR**. Application of PCA to **SR** gives a first score **u** and loading $\mathbf{v}^T$, whose product gives the filtered matrix **SR***. Finally, **SR*** is unfolded to give a new estimation of the three spectra in the second column of the augmented species spectra matrix $[\mathbf{S}_1;\mathbf{S}_2;\mathbf{S}_3]$*. The process is repeated for all columns of $[\mathbf{S}_1;\mathbf{S}_2;\mathbf{S}_3]$, i.e. for all species spectra of the system

each column of the new augmented matrix $[\mathbf{S}_1;\mathbf{S}_2;\mathbf{S}_3]^*$. This is a very convenient and fast way to implement the trilinearity constraint during the ALS optimization procedure. The equality of shapes for the fluorescent emission spectra of the same species at the three excitation wavelengths is therefore guaranteed. When this is the case and the real data structure is trilinear, the application of the proposed method converges quickly to the correct solution giving the common vector profiles defining the row and column common vector spaces of the simultaneously analysed data matrices.

The different steps in Table 1 have been implemented in a set of MATLAB[25] programme files. The execution of the successive steps previously described is fast. As with other non-linear optimization methods, it requires active user interaction for initial estimate selection (step 4) and application of more appropriate constraints. Previous knowledge of the general features of the system is extremely helpful in this context.

GRAM and TLD method

The set of three simulated ($\mathbf{D}_1$, $\mathbf{D}_2$, $\mathbf{D}_3$) or three experimental ($\mathbf{D}_4$, $\mathbf{D}_5$, $\mathbf{D}_6$) data matrices is folded in a cube of data $\underline{\mathbf{D}}$ (third-order tensor, Figure 2) and analysed using GRAM and the TLD method proposed by Sanchez and Kowalski[5] and others.[23, 24] These methods are considered the standard methods for the analysis of three-way trilinear data in chemistry. In this work, results obtained by application of ALS MCR, GRAM and TLD are compared. A brief description of GRAM and TLD is given below.

GRAM allows the simultaneous analysis of two data matrices. Therefore, in our case, three combinations of two data matrices are possible for the three simulated data matrices $\mathbf{D}_1$, $\mathbf{D}_2$ and $\mathbf{D}_3$: $\mathbf{D}_1$ and $\mathbf{D}_2$; $\mathbf{D}_1$ and $\mathbf{D}_3$; and $\mathbf{D}_2$ and $\mathbf{D}_3$. The same holds for the three experimental data matrices $\mathbf{D}_4$, $\mathbf{D}_5$ and $\mathbf{D}_6$. GRAM performs the generalized eigenvalue decomposition of each set of two matrices (e.g. $\mathbf{D}_1$ and $\mathbf{D}_2$) using the QZ simultaneous diagonalization algorithm,[26] solving for the common eigenvectors spanning the common vector spaces of the two simultaneously analysed matrices. In order to use the QZ algorithm, rectangular data matrices should be projected first to square matrices and this transformation must preserve the rank of the two simultaneously analysed matrices. From the common eigenvectors obtained by the QZ method, matrices of emission species spectra ($\mathbf{S}$), of concentration profiles ($\mathbf{C}^T$, species distribution) and of relative intensities of emission spectra at different excitation wavelengths ($\mathbf{K}_k$) are deduced. TLD performs the decomposition of the three-way data set $\underline{\mathbf{D}}$ (Figure 2) in steps. First, the singular value decomposition of row-wise ($[\mathbf{D}_1,\mathbf{D}_2,\mathbf{D}_3]$ in Figure 2), column-wise ($[\mathbf{D}_1;\mathbf{D}_2;\mathbf{D}_3]$ in Figure 2) and tube-wise ($[\mathbf{D}_1:\mathbf{D}_2:\mathbf{D}_3]$ in Figure 2) unfolded augmented matrices gives the row space scores ($\mathbf{U}$), the column space scores ($\mathbf{V}$) and the first two vectors of the tube space scores ($\mathbf{W}$). Second, two representative pseudosamples $\mathbf{G}_1$ and $\mathbf{G}_2$ are obtained by projection of the original tensor $\underline{\mathbf{D}}$ (Figure 2) onto the ($\mathbf{U}$, $\mathbf{V}$, $\mathbf{W}$) basis. Third, the emission species spectra, the concentration profiles and the relative intensities of the emission spectra at different wavelengths are obtained from the resolution of the generalized eigenvalue decomposition of matrices $\mathbf{G}_1$ and $\mathbf{G}_2$ (as in GRAM using the QZ algorithm; see above). All the described steps are also valid for the decomposition of the experimental three-way data set built from the experimental data matrices $\mathbf{D}_4$, $\mathbf{D}_5$ and $\mathbf{D}_6$.

A more detailed description of both GRAM and TLD can be found elsewhere.[1,5,7,23,24] In this work, GRAM and TLD are mostly used to validate the results obtained by ALS MCR.

pK_a evaluation and species distribution of system in whole pH range

The equilibrium reactions and their respective constants for the deprotonation of salicylic acid are

$$H_2sal \rightleftharpoons H^+ + Hsal^-, \qquad K_{a_1} = [Hsal^-][H^+]/[H_2sal] \tag{4}$$

$$Hsal^- \rightleftharpoons H^+ + sal^{2-}, \qquad K_{a_2} = [sal^{2-}][H^+]/[Hsal^-] \tag{5}$$

The dissociation constants (pK_a) and species distribution of the system were calculated from the concentration of $Hsal^-$ and sal^{2-} species obtained from the ALS MCR procedure. The concentration of H_2sal species was calculated from the mass balance of salicylic acid. The hydrogen ion concentration was measured experimentally.

The STAR programme[27] developed for the study of ionic equilibria from spectrophotometric data, is a non-linear least squares regression programme for the refinement of stability constants. The programme is based on fulfilment of the mass action law, mass balance equations and Beer's law. Species fluorescence spectra are obtained by solving the Beer-like linear model from calculated species concentration (for a set of stability constants) and experimental fluorescence spectra.

RESULTS AND DISCUSSION

Rank analysis

Rank analysis of the individual simulated data matrices $\mathbf{D}_1$, $\mathbf{D}_2$ and $\mathbf{D}_3$ using the singular value decomposition (SVD) method[26] gives (Table 2) first and second singular values much larger than the following ones, but the third singular value is already at the same level as the fourth and subsequent ones (see s_3/s_4 ratios). This is in agreement with the detection of two fluorescent species. Rank analysis of the individual experimental data matrices $\mathbf{D}_4$, $\mathbf{D}_5$ and $\mathbf{D}_6$ also gives similar trends in the first singular values (see Table 2). Real experimental noise, however, is not uniformly and randomly distributed, giving some systematic contributions and larger s_3/s_4 ratios. From the obtained results it is assumed that only two of the three species involved in the acid–base equilibria of salicylic acid are fluorescent. Matrices $\mathbf{D}_2$ and $\mathbf{D}_5$ give the largest singular values s_1 and s_2, showing that the excitation wavelength of 290 nm gives a slightly better sensitivity than the other excitation wavelengths used to obtain the matrices $\mathbf{D}_1$, $\mathbf{D}_3$, $\mathbf{D}_4$ and $\mathbf{D}_6$. On the other hand, the lower the ratio s_1/s_2, the more similar are the two singular values and the more orthogonal are the corresponding species spectra. From this point of view the species spectra recovered from matrices $\mathbf{D}_3$ and $\mathbf{D}_6$ ($\lambda = 314$ nm) will be the more orthogonal ones. Considering that the noise level is described by the third singular value, the ratio s_2/s_3 gives an estimation of the signal-to-noise ratio.

Rank analysis of augmented row-wise and column-wise data matrices confirms data trilinearity for

Table 2. Rank analysis of data matrices by SVD

	Singular values				Singular value ratios		
Matrix	s_1	s_2	s_3	s_4	s_1/s_2	s_2/s_3	s_3/s_4
$\mathbf{D}_1$	6425	79	10	10	82	8	1
$\mathbf{D}_2$	10638	154	16	16	69	10	1
$\mathbf{D}_3$	4912	131	7	7	37	19	1
$[\mathbf{D}_1;\mathbf{D}_2;\mathbf{D}_3]$	13361	316	18	18	42	17	1
$[\mathbf{D}_1, \mathbf{D}_2, \mathbf{D}_3]$	13363	222	17	17	60	13	1
$\mathbf{D}_4$	2352	109	16	6	22	7	3
$\mathbf{D}_5$	3928	153	18	6	26	8	3
$\mathbf{D}_6$	1891	133	17	6	14	8	3
$[\mathbf{D}_4;\mathbf{D}_5;\mathbf{D}_6]$	4947	334	35	27	15	9	2
$[\mathbf{D}_4, \mathbf{D}_5, \mathbf{D}_6]$	4953	236	32	11	21	7	3

both simulated and experimental data matrices. There is no rank augmentation when the three different fluorescent emission spectra matrices are augmented both row- and column-wise. This shows that the three matrices have the same row and column vector spaces, defined by common vector profiles, i.e. the same pure species fluorescent emission spectra and species distribution (concentration profiles of the two species at varying pH). The number of components deduced from the number of larger singular values is in all cases also equal to two. Column-wise data matrix augmentation yields in both simulated and experimental cases a lower s_1/s_2 ratio than row-wise data matrix augmentation and similar or better s_1/s_2 and s_2/s_3 ratios than the best of the individual data matrices. From these results the best data matrix set-up that can be selected for optimal signal-to-noise ratio (largest s_2/s_3 ratio), sensitivity (larger values of s_1 and s_2) and selectivity (lower s_1/s_2 ratio) is the column-wise augmented data matrices $[\mathbf{D}_1;\mathbf{D}_2;\mathbf{D}_3]$ and $[\mathbf{D}_4;\mathbf{D}_5;\mathbf{D}_6]$. This column-wise data augmentation was selected for the rest of the calculations.

ALS MCR of individual data matrices

Table 3 gives a summary of the results obtained by application of PCA and ALS MCR to the analysis of individual and augmented data matrices for a two-component model. Data fitting using ALS was

Table 3. Comparison of results (lack of fit (%)[a]) obtained by analysis of individual data matrices and augmented data matrices (experimental and simulated data) using different methods[b]

Matrix	PCA	ALS^{nt}	ALS^{t}	GRAM	TLD
$\mathbf{D}_1$	1·093	1·306	—	—	—
$\mathbf{D}_2$	1·085	1·273	—	—	—
$\mathbf{D}_3$	1·059	1·083	—	—	—
$[\mathbf{D}_1;\mathbf{D}_2]$	1·090	1·089	1·092	1·092	1·092
$[\mathbf{D}_1;\mathbf{D}_3]$	1·085	1·082	1·086	1·088	1·086
$[\mathbf{D}_2;\mathbf{D}_3]$	1·084	1·082	1·085	1·086	1·085
$[\mathbf{D}_1;\mathbf{D}_2;\mathbf{D}_3]$	1·088	1·086	1·090	—	1·090
$\mathbf{D}_4$	0·869	0·872	—	—	—
$\mathbf{D}_5$	0·571	0·619	—	—	—
$\mathbf{D}_6$	0·992	1·016	—	—	—
$[\mathbf{D}_4;\mathbf{D}_5]$	0·901	0·918	1·131	0·938	0·932
$[\mathbf{D}_4;\mathbf{D}_6]$	1·326	1·325	1·627	1·590	1·384
$[\mathbf{D}_5;\mathbf{D}_6]$	0·822	0·822	0·838	0·997	0·838
$[\mathbf{D}_4;\mathbf{D}_5;\mathbf{D}_6]$	1·021	1·021	1·189	—	1·182

[a] Fitting errors are expressed as

$$\text{lack of fit (\%)} = \frac{\sum (d_{ij} - d_{ij}^{*})^2}{\sum (d_{ij})^2} \times 100$$

where d_{ij} are the experimental data points for spectrum i at wavelength j and d_{ij}^{*} are the reproduced data points for spectrum i at wavelength j by PCA, ALS optimization, GRAM or TLD.

[b] Methods used: PCA is principal component analysis; ALS^{nt} is ALS Multivariate curve resolution without the trilinearity constraint; ALS^{t} is ALS multivariate curve resolution with the trilinearity constraint; GRAM is the generalized rank annihilation method;[7] TLD is the trilinear decomposition method.[5] ALS^{t}, GRAM and TLD cannot be applied in the analysis of the individual data matrices. GRAM cannot be applied to the augmented matrices $[\mathbf{D}_1;\mathbf{D}_2;\mathbf{D}_3]$ and $[\mathbf{D}_4;\mathbf{D}_5;\mathbf{D}_6]$.

close to data fitting obtained by PCA. As the ALS optimization gives more constrained solutions than the PCA solutions, the lack of fit obtained by ALS was slightly larger than that obtained by PCA. The lack-of-fit values were around or lower than 1% in all cases. Initial estimations of pH species distribution (concentration profiles) were obtained from EFA of individual data matrices.[17] The constraints applied in the analysis of the individual simulated data matrices were non-negativity, closure and unimodality and those in the analysis of the individual experimental data matrices were non-negativity and unimodality. Since total selectivity (only one of the two species present) does not exist in the pH range under study, this constraint cannot be applied without losing data fitting. Closure could not be applied to the experimental data matrices, since at the beginning of the titration the first species H_2sal exists at high concentrations and is not fluorescent. Too severe a unimodality constraint over the pH distribution profiles in the analysis of the more noisy data had undesired effects, cutting the pH distribution profiles in unreasonable positions. Therefore this constraint was accordingly relaxed.

In Table 4 the resolution power of the ALS method for simulated data is given. This is evaluated from the recovery of the pH distribution profiles (C1 and C2) and species spectra (S1 and S2) of the two detected species ($Hsal^-$ and sal^{2-}). The recovery between the calculated and theoretical profiles was evaluated as the cosine of the angle between the two vectors associated with the two profiles. A

Table 4. Recovery of profiles (cosine values[a]) from analysis of individual data matrices and augmented data matrices (simulated experiments)

Method[b]	Matrix	C1	C2	S1	S2
ALS	$\mathbf{D}_1$	0·9828	0·9995	1·0000	0·9755
ALS	$\mathbf{D}_2$	0·9819	0·9990	1·0000	0·9772
ALS	$\mathbf{D}_3$	0·9861	0·9993	1·0000	0·9864
ALS^{nt}	$[\mathbf{D}_1;\mathbf{D}_2]$	0·9859	0·9953	1·0000	0·9775
ALS^{nt}	$[\mathbf{D}_1;\mathbf{D}_3]$	0·9863	0·9956	1·0000	0·9726
ALS^{nt}	$[\mathbf{D}_2;\mathbf{D}_3]$	0·9863	0·9979	1·0000	0·9825
ALS^{t}	$[\mathbf{D}_1;\mathbf{D}_2]$	0·9999	0·9995	1·0000	0·9989
ALS^{t}	$[\mathbf{D}_1;\mathbf{D}_3]$	1·0000	0·9999	1·0000	1·0000
ALS^{t}	$[\mathbf{D}_2;\mathbf{D}_3]$	1·0000	0·9999	1·0000	0·9999
ALS^{nt}	$[\mathbf{D}_1;\mathbf{D}_2;\mathbf{D}_3]$	0·9863	0·9980	1·0000	0·9806
ALS^{t}	$[\mathbf{D}_1;\mathbf{D}_2;\mathbf{D}_3]$	1·0000	0·9999	1·0000	0·9999
GRAM	$[\mathbf{D}_1;\mathbf{D}_2]$	1·0000	0·9998	1·0000	0·9999
GRAM	$[\mathbf{D}_1;\mathbf{D}_3]$	1·0000	0·9999	1·0000	1·0000
GRAM	$[\mathbf{D}_2;\mathbf{D}_3]$	1·0000	0·9998	1·0000	0·9999
TLD	$[\mathbf{D}_1;\mathbf{D}_2]$	1·0000	0·9995	1·0000	0·9999
TLD	$[\mathbf{D}_1;\mathbf{D}_3]$	1·0000	0·9999	1·0000	1·0000
TLD	$[\mathbf{D}_2;\mathbf{D}_3]$	1·0000	0·9999	1·0000	0·9999
TLD	$[\mathbf{D}_1;\mathbf{D}_2;\mathbf{D}_3]$	1·0000	0·9999	1·0000	0·9999

[a] Recovery of profiles by ALS curve resolution, GRAM or TLD measured as the cosine of the angle between the two vectors associated with the calculated and theoretical profiles (see second subsection of Results).
[b] ALS^{nt} is the ALS method without the trilinearity constraint; ALS^{t} is the ALS method with the trilinearity constraint; GRAM is the generalized rank annihilation method;[7] TLD is the trilinear decomposition method.[5]
[c] Average of the values obtained for the S2 ALS-recovered profiles at each excitation wavelength when the trilinearity constraint is not applied. The S1 profile is always totally recovered (cosine equal to one).

high similarity (correlation) between the calculated and theoretical profiles gives a very small angle between the two vectors representing the two profiles, i.e. a cosine of this angle close to one. From the results in Table 4 it is seen that the recovery of the C2 and S1 profiles was very good, whereas that of the C1 and S2 profiles was not as good. This shows that in the analysis of the individual data matrices the factor analysis ambiguities for profiles C2 and S1 were practically solved, whereas those for profiles C1 and S2 were not. These results are in agreement with previous investigations[11] and with resolution conditions based on local rank.[21] As the first species was practically (not totally) the only existing species at the beginning of the simulated titration, its spectrum S1 is unambiguously recovered. As shown by Malinowsky[28] and others,[29, 30] the recovery of the concentration profile of a particular species is possible if there is a local region or window where this species is absent or at very low concentration (practically non-existent). This condition was fulfilled for the pH distribution profile C2 (Figure 4(b)), since this species practically did not exist at the beginning of the simulated acid–base titration. In contrast, it was not possible to recover correctly the C1 and S2 profiles, because the deprotonation reaction could not be pushed forward to its end in aqueous medium ($pK_w = 13{\cdot}74$) and no selectivity exists for species 2. Linear combinations of the C1 and S2 profiles were then possible, fitting equally well the data, even when other constraints such as non-negativity, closure or unimodality were applied. This situation was very similar to that also found in chromatography for resolution of embedded peaks.[21, 30] Figure 4 shows graphically the agreement between the theoretical distribution and spectra profiles and those obtained by the ALS MCR method in the analysis of the individual simulated data matrix $\mathbf{D}_2$. Rotational ambiguities were solved (equal shape) for the S1 and C2 profiles but were not solved (different shapes) for the S2 and C1 profiles. Additionally, because the system was closed and there was almost selectivity at the beginning of the titration, the intensity ambiguity was also solved for the S1 spectra profile. However, for the C2 profile, although the rotational ambiguity had been practically solved (see Table 4), the intensity ambiguity persisted. This was a consequence of the lack of selectivity at the end of the titration, which allowed still ambiguous solutions for intensity to be present for the concentration profiles C1 and C2 and also for the S2 spectra profiles.

Results obtained for experimental data matrices showed the same trends as for simulated data matrices. The same conclusions can be extracted in relation to the resolution of the species profiles. Lack-of-fit values obtained for experimental data matrices were similar to those obtained for simulated data matrices, which again allows the extrapolation of the resolution considerations given for simulated data. The experimental data structure was very similar to the simulated data structure.

ALS MCR of augmented data matrices

Initial estimations of pH distribution profiles for the simultaneous analysis of correlated data matrices were taken from EFA of any of the individual data matrices. In any case, identical solutions were recovered after the ALS optimization. Table 3 gives the ALS lack of fit obtained in the analysis of the column-wise augmented matrices for both simulated and experimental data. As for individual data matrices, the values of the ALS lack of fit for both augmented simulated and experimental data matrices were similar. When the trilinearity constraint was not applied, the ALS lack-of-fit values were practically identical to those obtained by PCA. On applying the trilinearity constraint, very small increases in the ALS lack-of-fit values were observed; for example, for the augmented experimental data matrix $[\mathbf{D}_4;\mathbf{D}_5;\mathbf{D}_6]$ a small increase from 1·021% (ALS with no trilinearity) to 1·189% (ALS with trilinearity) was observed, probably due to very small experimental departures from the postulated trilinear model.

Table 4 shows the results achieved in the recovery of pH distribution profiles and spectra profiles of the two detected species using the proposed ALS MCR method when applied to the simulated

augmented data matrices [$\mathbf{D}_1$;$\mathbf{D}_2$], [$\mathbf{D}_1$;$\mathbf{D}_3$], [$\mathbf{D}_2$;$\mathbf{D}_3$] amd [$\mathbf{D}_1$;$\mathbf{D}_2$;$\mathbf{D}_3$]. The best results were obtained when the trilinearity constraint was applied during the ALS optimization. Other constraints applied were the non-negativity and closure constraints for simulated data matrices and the non-negativity constraint for experimental data matrices. The unimodality constraint was applied under the same conditions as in the analysis of individual data matrices (see above). As can be deduced from the high

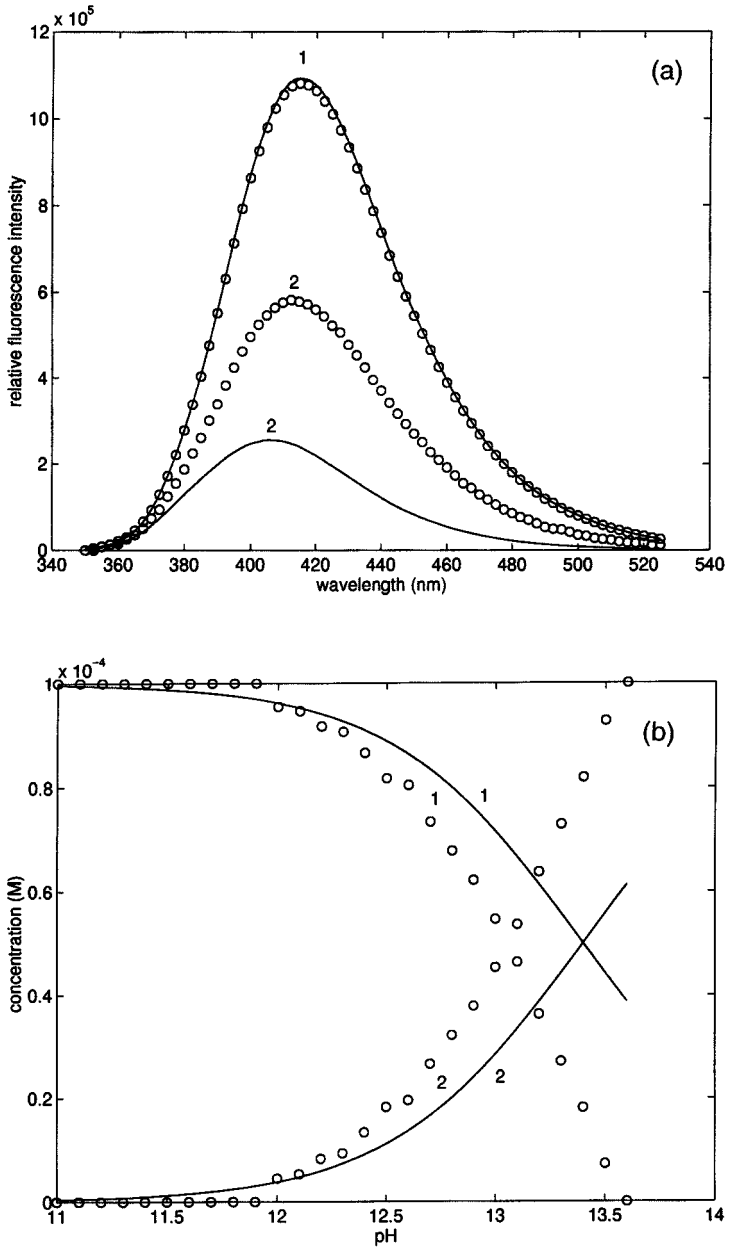

Figure 4. Simulated (lines) and recovered (circles) (a) spectra and (b) species distribution for $Hsal^-$ species (number 1) and sal^{2-} species (number 2) in analysis of individual data matrix $\mathbf{D}_2$ (pH range 11–13·6)

cosine values in Table 4, all profiles, C1, C2, S1 and S2, were well recovered when the trilinearity constraint was applied. Figure 5 shows graphically the perfect agreement between the simulated and ALS-estimated profiles. The rotational and intensity ambiguities associated with the factor analysis of the individual data matrices were completely solved in the analysis of the column-wise augmented matrices with the proposed method. As seen in Figure 5(b), the shapes of the fluorescent emission spectra at the three excitation wavelengths were identical (trilinearity), as can be easily proved from

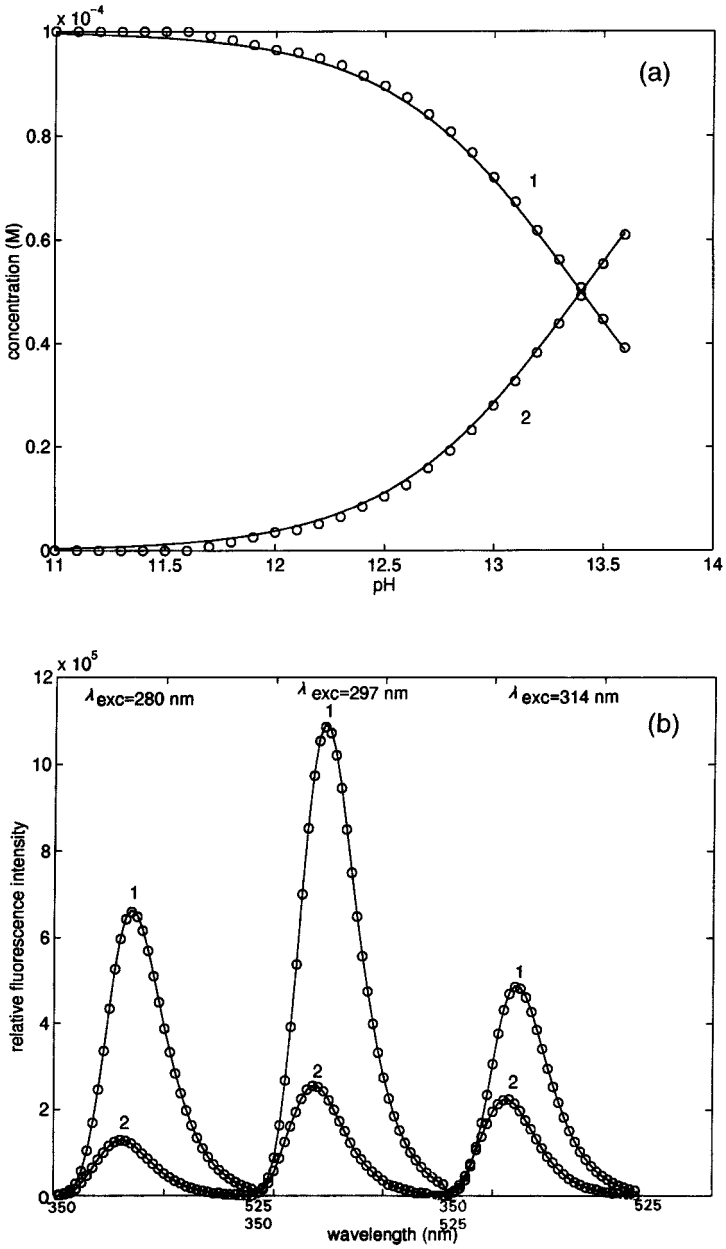

Figure 5. Simulated (lines) and recovered (circles) (a) species distributions and (b) spectra for $Hsal^-$ species (number 1) and sal^{2-} species (number 2) in analysis of augmented data matrix $[\mathbf{D}_1;\mathbf{D}_2;\mathbf{D}_3]$ (pH range 11–13·6)

singular value analysis of the three fluroescent emission spectra of every species: only one significant singular value was obtained. When the trilinearity constraint was not applied, although the ALS lack of fit was even better than when it was applied, some rotational ambiguities persisted for profiles C1 and S2 (see Table 4). Trilinearity is therefore a necessary condition to solve completely the rotational ambiguities, i.e. to solve a system such as the present one where the selectivity and/or local rank resolution conditions[21] were not present.

Figure 6 shows the recovered fluorescent emission spectra from the analysis of the augmented experimental data matrix $[\mathbf{D}_4;\mathbf{D}_5;\mathbf{D}_6]$. The non-negativity, unimodality and trilinearity constraints were applied as in the analysis of the simulated data. The shape of the three spectra of the same species at different excitation wavelengths was the same, but the intensity changed. The intensity ratio of the three fluorescent emission spectra of the same species at each excitation wavelength was specific for each species, as it should be from theoretical considerations.

Comparison between GRAM, TLD method and ALS MCR method

For simulated three-way trilinear data (simultaneous analysis of $\mathbf{D}_1$, $\mathbf{D}_2$ and $\mathbf{D}_3$ the results obtained by GRAM, TLD and ALS MCR (using the trilinearity constraint) were very close (Tables 3 and 4). In all cases, nearly perfect agreement was found between the recovered and true spectra and pH distribution profiles. These results confirm that the proposed MCR method with the trilinearity constraint gives similar results to those obtained by GRAM and TLD when applied to three-way trilinear data systems.

In Table 3 the lack-of-fit values obtained by GRAM and TLD are also given for the analysis of experimental data matrices $\mathbf{D}_4$, $\mathbf{D}_5$ and $\mathbf{D}_6$. Small differences are shown between different treatments which can be caused by small departures from trilinearity in the experimental data and their effect on the different algorithms. In general, very good agreement is obtained between GRAM, TLD and ALS^t (with the trilinerity constraint) for the C1, C2, S1 and S2 species profiles, with cosine values close to one (not given in Table 4). Larger differences, although still small, are observed between these values

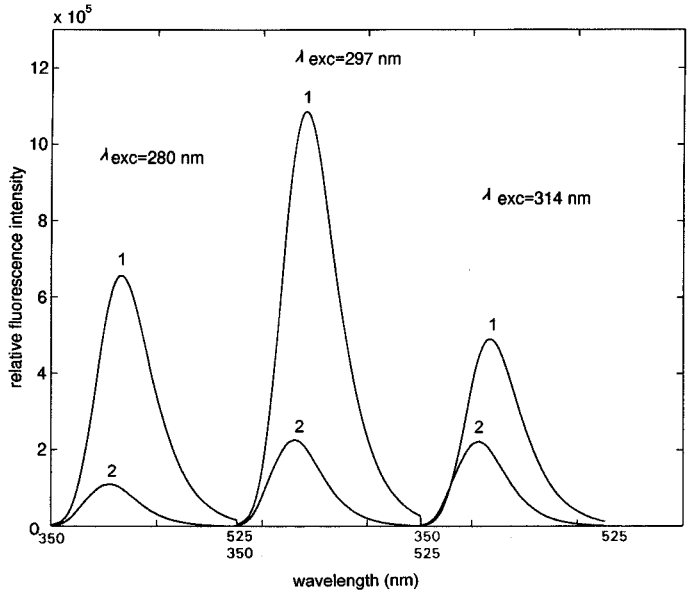

Figure 6. Fluorescent emission spectra of $Hsal^-$ species (number 1) and sal^{2-} species (number 2) resolved through ALS method from experimental augmented data matrix $[\mathbf{D}_4;\mathbf{D}_5;\mathbf{D}_6]$

and those obtained by ALS^{nt} (without the trilinearity constraint), which also confirm the results obtained using simulated data (given in Table 4).

Apart from data fitting (measured by lack-of-fit values) and recovery (measured by cosine values), GRAM, TLD and ALS MCR can also be compared in terms of speed and ease of use. Giving exact figures on the speed of these methods is largely dependent on hardware and software implementations; therefore only qualitative estimation and interpretation of the differences observed between the different methods are given. Our personal evaluation of speeds for the analysis of data sets under study yielded the following results (given in parentheses in relative times): GRAM (1) < SVD or PCA(10) < TLD(40) < ALS (5–200). The fastest method was GRAM. It was faster than TLD because in its algorithm the singular value decomposition was limited to ten components, whereas in TLD the singular value decomposition of the three augmented matrices (column-wise, row-wise and tube-wise) was performed for all the components. TLD is not restricted, however, to the analysis of two data sets as GRAM is; GRAM and TLD were implemented in a single programme and were fast and easy to use. MCR, on the other hand, involves several steps (see Table 1) and requires some previous learning on how to use the ALS method. However, the interpretation of results was easier in MCR than in GRAM or TLD, since the profiles obtained by MCR were always real positive numbers, whereas those obtained by GRAM or TLD must be transformed appropriately.[24] In MCR there are two steps which determine the speed of the analysis. The first step is the calculation of initial estimates to be used as starting values of the optimization. In the present work this was performed by means of EFA. The speed of EFA is determined by the speed of the method for eigenanalysis or singular value decomposition. Using the MATLAB[25] SVD method, this is usually performed in seconds or a few minutes, depending on the size of the data matrices. The second step is the ALS optimization. The speed of this optimization depends on various factors, among which the 'quality' of the initial estimates, the constraints applied during the optimization and the convergence criterion are important. Initial estimates were obtained from the EFA (see above) and proved good enough in all cases studied in this work. It was found that application of the right constraints during the ALS optimization increased the speed of convergence, but this depended on which algorithms were used for the implementation of these constraints. A more detailed study of the effect of constraints on the speed of the optimization is outside the scope of this work. The convergence criterion also affected the speed of the calculations; a convergence criterion of 0·01% change between the standard deviations of the residuals in two consecutive iterations was used in all calculations. A looser convergence criterion of 0·1% would have given faster results, with practically identical solutions to those shown in Tables 3–5 and Figures 4–6 for a convergence criterion of 0·01%. In any case, none of the optimizations performed using the ALS MCR method lasted more than 5 min on an IBM RISC 590 AIX computer.

In the present work the ALS MCR method is validated for the analysis of trilinear three-way data; other examples of application to trilinear and non-trilinear data have already been published[2,8,9,11,12] or are in progress at present. Compared with GRAM and TLD, the ALS MCR method has the favourable feature that it can be used for both two- and three-way data analysis, while within three-way data it can be used for both trilinear and non-trilinear data. GRAM and TLD are much more restricted than the proposed ALS MCR method because they can be only used for three-way trilinear data. Moreover, GRAM can only be used in the analysis of two data sets. The ALS method can take advantage of additional data constraints, which can be essential in the case of three-way data with departures from the trilinear model.

pK_a evaluation and species distribution of system in whole pH range

In Table 5 a summary of the pK_a estimations performed for the different data matrices (see Method) is given. The values obtained using the ALS MCR method are compared with those obtained by the

Table 5. pK_a values of salicylic acid obtained by ALS, TLD or GRAM and a traditional least squares curve-fitting programme (STAR) for experimental and simulated data

Equilibrium	Matrix	ALS (TLD or GRAM)[a]	STAR	Value in literature
$H_2sal \rightleftharpoons Hsal^- + H^+$	$\mathbf{D}_4$	2·83(2)[b]	2·842(4)	2·83[c]
	$\mathbf{D}_5$	2·82(2)	2·830(4)	
	$\mathbf{D}_6$	2·82(6)	2·80(1)	
	$[\mathbf{D}_4;\mathbf{D}_5;\mathbf{D}_6]$	2·82(2)	2·829(4)	
$Hsal^- \rightleftharpoons sal^{2-} + H^+$	$\mathbf{D}_1$	13·21(8)	13·394(5)	13·39[d]
	$\mathbf{D}_2$	13·21(9)	13·396(5)	
	$\mathbf{D}_3$	13·22(8)	13·391(7)	
	$[\mathbf{D}_1;\mathbf{D}_2;\mathbf{D}_3]$	13·43(2)	13·398(5)	
	$\mathbf{D}_4$	13·23(6)	13·122(4)	
	$\mathbf{D}_5$	13·09(5)	13·167(4)	
	$\mathbf{D}_6$	13·09(5)	13·19(1)	
	$[\mathbf{D}_4;\mathbf{D}_5;\mathbf{D}_6]$	13·09(6)	13·146(4)	

[a] In the case of simultaneous analysis of several data matrices, results are analogous for GRAM, TLD and ALS multivariate curve resolution with the trilinearity constraint.
[b] The values of standard deviations are given in parentheses.
[c] Obtained through a potentiometric method.[17]
[d] Obtained through a spectrophotometric method.[17]

STAR computer programme. Whereas the latter is a hard-modelling method using the mass action law, the former is a soft-modelling method which does not use the mass action law constraint.

The estimation of pK_{a_2} is firstly shown for the individual simulated data matrices $\mathbf{D}_1$, $\mathbf{D}_2$ and $\mathbf{D}_3$. As expected, the results obtained using the STAR programme have recovered correctly the pK_{a_2} value (13·39) used in the simulations. On the other hand, the values obtained by ALS in the individual analysis of the simulated data matrices differ from the theoretical value used, in agreement with the fact that in the individual analysis the rotational ambiguities were not completely solved and therefore the recovery of the pH distribution profiles (see Figure 4) was not fully correct. The effect of these remaining ambiguities and of the closure constraint was to push forward the formation of the deprotonated species more than allowed if the mass action law were applied, and at the same time the intensity of the corresponding species spectra was changed (Figure 4). On the other hand, as already pointed out, the simultaneous analysis of the simulated matrices $\mathbf{D}_1$, $\mathbf{D}_2$ and $\mathbf{D}_3$ solved the rotational ambiguities and recovered totally the pH distribution profiles (Figures 5 and 6). Consequently, the value of pK_{a_2} calculated by ALS was in this case much closer to that used in the simulations. Also, the pK_a values obtained in the simultaneous analysis of several data matrices by ALS MCR are very similar to those obtained by GRAM and TLD if the appropriate similarity transformations are applied,[24] i.e. TLD and GRAM negative and complex profiles are converted to their positive real values.

The real value of pK_{a_2} was estimated from the experimental data matrices $\mathbf{D}_4$, $\mathbf{D}_5$ and $\mathbf{D}_6$ and from the column-wise augmented matrix $[\mathbf{D}_4;\mathbf{D}_5;\mathbf{D}_6]$, both with the STAR programme[27] and with the ALS method (Figure 7). In this case the differences between the ALS and STAR individual analyses were smaller than those found previously for simulated data and within the estimated standard deviations. It appears from these results that for real data the effect of rotational ambiguities is less important than for simulated data. One possible explanation of this can be found in the fact that the recovered pK_{a_2} value, 13·1, is lower than that used for the simulation, 13·4. A lower pK_{a_2} means that more of the

deprotonated species is formed at lower pH, giving better conditions for the recovery of its pH distribution profile and species spectra. Another possibility is that the effect of non-randomly distributed experimental error constrains the number of possible ambiguous solutions. Whereas rotational ambiguities appear clearly for simulated data with no error or with small randomly distributed errors, they would not be so clearly deduced from real data with non-random error distributions. This would mean a possible advantage of the factor analysis of real data compared with the factor analysis of simulated data. This subject, however, deserves further investigation.

Evaluation of the first pK_a was also possible from the experimental data matrices. The results obtained using the two different approaches were in this case very similar (Table 5) for both individual data matrix analysis and augmented data matrix analysis. The agreement in the evaluation of the first pk_a between the ALS estimations obtained from both kinds of data matrices and those obtained using the STAR programme proves that no rotational ambiguities are present in the pH region where this equilibrium predominates. The value of $pK_{a_1} = 2{\cdot}82 \pm 0{\cdot}02$ obtained under experimental conditions of 25 °C and 0·1 M ionic strength agrees with that proposed in the literature.[13, 14]

From the pH distribution profiles recovered from the analysis of experimental data, the second pK_a of salicylic acid at 0·1 M ionic strength and 25 °C is finally estimated (see Method) to be equal to 13·1. This value differs somewhat from our previous estimation of 13·4 using a single-wavelength spectrophotometric method under much more limited experimental conditions.[15]

CONCLUSIONS

A set of fluorescent emission correlated data matrices obtained in a spectrofluorimetric acid–base titration monitored at different excitation wavelengths was shown to have a three-way trilinear data

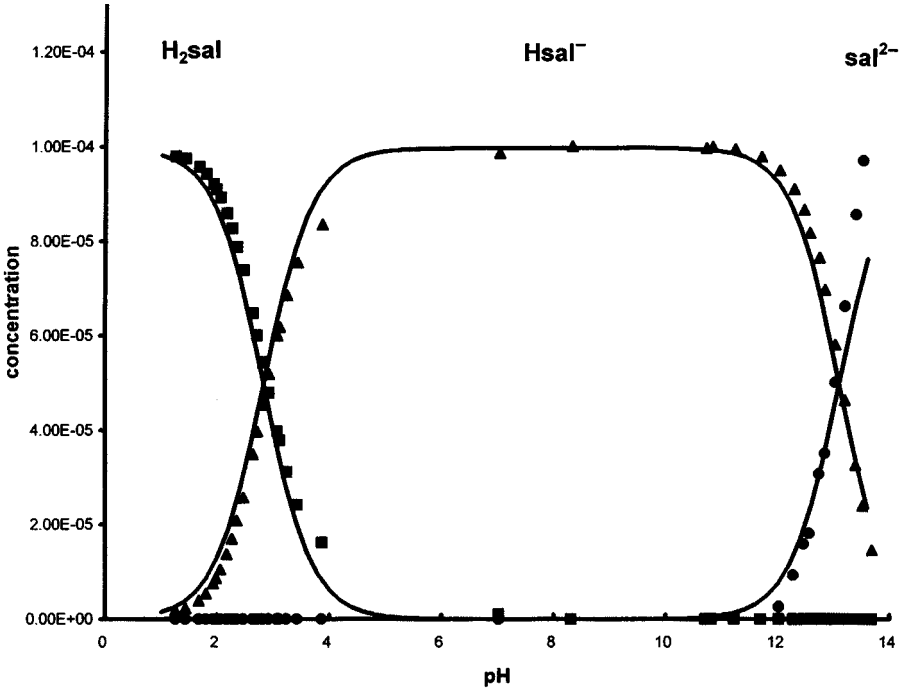

Figure 7. Theoretical (lines) and calculated through ALS MCR method (symbols) species distributions from experimental augmented data matrix [$\mathbf{D}_4$;$\mathbf{D}_5$;$\mathbf{D}_6$]: ■, H_2sal; ▲, $Hsal^-$; ●, sal^{2-} ([salicylic acid] = $9{\cdot}96 \times 10^{-5}$, pH range 1·17–13·60, $pK_{a_1} = 2{\cdot}82$, $pK_{a_2} = 13{\cdot}09$)

structure. The inclusion of a trilinearity constraint during the curve resolution of a set of correlated data matrices allowed optimal recovery of the species profiles, concentration pH distribution profiles and spectra profiles, solving those factor analysis ambiguities present in the factor analysis of individual data matrices using the same curve resolution method. From the optimal pH distribution profiles recovered by the proposed alternating least squares multivariate curve resolution method, it was possible to get an improved estimation of the second pK_a of the hydroxyl group of salicylic acid, equal to 13·1 at 25 °C and 0·1 M ionic strength.

In the present work the alternating least squares multivariate curve resolution method was validated for the analysis of trilinear three-way data and showed similar results to those obtained by application of the generalized rank annihilation and trilinear decomposition methods when applied to the same data sets.

REFERENCES

1. E. Sanchez, L. S. Ramos and B. R. Kowalski, *J. Chromatogr.* **385**, 151 (1987).
2. R. Tauler, B. R. Kowalski and S. Flemming, *Anal. Chem.* **65**, 2040 (1993).
3. C. N. Ho, G. D. Christian and E. R. Davidson, *Anal. Chem.* **50**, 1108 (1978).
4. R. Tauler, A. K. Smilde, J. M. Henshaw, L. W. Burgess and B. R. Kowalski, *Anal. Chem.* **66**, 3337 (1994).
5. E. Sanchez and B. R. Kowalski, *J. Chemometrics*, **4**, 29 (1990).
6. E. Spjotvoll, H. Martens and R. Volden, *Technometrics*, **3**, 173 (1982).
7. E. Sanchez and B. R. Kowalski, *J. Chemometrics*, **2**, 265 (1988).
8. R. Tauler, *Chemometrics Intell. Lab. Syst.* **30**, 133 (1995).
9. J. Saurina, S. Hernández-Cassou and R. Tauler, *Anal. Chem.* **69**, 2329 (1997).
10. A. Smilde and D. A. Doornbos, *J. Chemometrics*, **5**, 345 (1991).
11. R. Tauler, A. Smilde and B. R. Kowalski, *J. Chemometrics*, **9**, 31 (1995).
12. E. Casassas, R. Tauler and I. Marqués, *Anal. Chim. Acta*, **310**, 473 (1995).
13. J. Buffle, *Complexation Reactions in Aquatic Systems. An Analytical Approach*, Ellis Horwood/Wiley, New York (1988).
14. A. E. Martell and R. J. Smith, *Critical Stability Constants*, Vol. 5, Plenum, New York (1992).
15. E. Casassas and R. Tauler, *J. Chim. Phys.* **81**, 233 (1984).
16. S. G. Schulman, *Molecular Luminescence Spectroscopy. Methods and Applications. Part I*, Wiley, New York (1985).
17. R. Tauler and E. Casassas, *Anal. Chim. Acta*, **206**, 189 (1988).
18. H. Gampp, M. Maeder, Ch. Meyer and A. D. Zuberbuhler, *Talanta*, **32**, 1133 (1985).
19. R. Keller and D. L. Massart, *Anal. Chim. Acta*, **246**, 379 (1991).
20. P. Geladi and S. Wold, *Chemometrics Intell. Lab. Syst.* **4**, 11 (1988).
21. R. Manne, *Chemometrics Intell. Lab. Syst.* **27**, 89 (1995).
22. W. Windig and J. Guilment, *Anal. Chem.* **63**, 1425 (1991).
23. M. Gui, S. C. Rutan and A. Agbodjan, *Anal. Chem.* **67**, 3293 (1995).
24. S. Li, J. C. Hamilton and P. J. Gemperline, *Anal. Chem.* **64**, 599 (1992).
25. *MATLAB Version 4.2*, The MathWorks, Inc., Natick, MA (1994).
26. G. H. Golub and Ch. F. Van Loan, *Matrix Computations*, 2nd edn, Johns Hopkins University Press, London (1989).
27. J. L. Beltrán, R. Codony and M. D. Prat, *Anal. Chim. Acta*, **276**, 441 (1993).
28. E. R. Malinowski, *J. Chemometrics*, **6**, 29 (1992).
29. M. Maeder, *Anal. Chem.* **59**, 527 (1987).
30. O. M. Kvalheim and Y. Z. Liang, *Anal. Chem.* **64**, 936 (1992).

