ELSEVIER

Calibration and detailed analysis of second-order flow injection analysis data with rank overlap

Marlon M. Reis [b,1], Stephen P. Gurden [a], Age K. Smilde [a,*], Márcia M.C. Ferreira [b]

[a] *Department of Chemical Engineering, Process Analysis and Chemometrics, University of Amsterdam,*
Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands
[b] *Department of Physical Chemistry, Chemistry Institute, State University of Campinas – UNICAMP, Campinas – SP, Brazil*

Abstract

With the current popularity of second-order (or hyphenated) instruments, there now exists a number of chemometric techniques for the so-called second-order calibration problem, i.e. that of quantifying an analyte of interest in the presence of one (or more) unknown interferent(s). Second-order instruments produce data of varying complexity, one particular phenomenon sometimes encountered being that of rank overlap (or rank deficiency), where the overall rank of the data is not equal to the sum of the ranks of the contributing species. The purpose of the present work is to evaluate the performance of two second-order calibration methods, a least squares-based and an eigenvalue-based solution, in terms of their quantitative ability and stability, as applied to flow injection analysis (FIA) data which exhibits rank overlap. In the presence of high collinearity in the data, the least squares methods is found to give a more stable solution. Two-mode component analysis (TMCA) is used to investigate the reasons for this difference in terms of the chemical properties of the species analysed. The success of second-order calibration of this data is found to depend strongly on the collinearity between the acidic and basic time profiles and the reproducibility of the pH gradient in the FIA channel, both of which are shown to be related to the pK_a values of the species. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Flow injection analysis; Second-order calibration methods; Two-mode component analysis; Rank overlap

1. Introduction

The development of second-order (or hyphenated) instruments [1] such as LC-UV, GC-MS and MS-MS has brought several advantages to the analytical chemist in terms of quantification and identification of compounds within a mixture system. The data generated by these instruments have necessitated the development of an area of chemometric calibration techniques called second-order methods [2]. Although these methods have similar general aims, i.e. the quantification of a known analyte in the presence of unknown interferents, there are also significant differences between the algorithms reported in the literature. For example, in terms of the type of data to which they have been applied, some methods have been used to solve problems of rank-one (also known as 'complexity-one' or 'bilinear') data and others used to solve higher-than-rank-one (also known as 'mixed-complexity' or 'non-bilinear') data. In terms of their algebraic formulation, some methods use a

* Corresponding author. Tel.: +31-20-525-5062;
fax: +31-20-525-5604.
E-mail address: asmilde@its.chem.uva.nl (A.K. Smilde).
[1] On leave.

least squares (LS) approach and others are eigenvalue (EV) based. Among the least squares-based methods are residual bilinearisation (RBL) [3], PARAFAC [4] and restricted Tucker3 models [5], whilst among the eigenvalue-based methods are rank annihilation factor analysis (RAFA) [6], the generalised rank annihilation method (GRAM) [7] and non-bilinear rank annihilation (NBRA) [8,9].

Second-order instruments produce data of varying complexity depending upon the nature of the analytical techniques being combined [10]. One particular phenomenon sometimes found is that of rank overlap (or rank deficiency), where the overall rank of the measured data is not equal to the sum of the ranks of the individual species contributions to the data, i.e. $X = A + B$ but $\text{rank}(X) < \text{rank}(A) + \text{rank}(B)$. Examples of data where this can occur are flow injection analysis in the presence of a pH gradient [11]; flow injection analysis where a kinetic decomposition takes place [12]; the use of an optical-fibre based chemical sensor for determination of halogenated molecules based on the Fujiwara reaction [13].

The purpose of the present work is to evaluate the performance of two second-order calibration methods, a least squares-based and an eigenvalue-based solution, in terms of their quantitative ability and stability, as applied to a data set having rank overlap. The problem of rank overlap was chosen because this kind of data is sometimes encountered in practice and represents non-ideal behaviour in measured data which is not easily eliminated. Sometimes rank overlap is intrinsic to the instrumentation being used, e.g. using similar solutes during a flow injection analysis leads to similar dispersion properties, and therefore, closure due to the solutes having equal total elution profiles. Alternatively (or additionally), rank overlap can be found simply where components have similar or equal spectral profiles, e.g. using a chemical sensor to follow different kinetic pathways with the same end-product can lead to rank overlap as the spectrum of the end-product is the same for the different pathways. In the presence of low experimental reproducibility and/or a low signal-to-noise ratio, high collinearity between component profiles can be a source of rank overlap.

The methods chosen for the comparison are a least squares-based method, RBL, and an eigenvalue-based method, NBRA. These can be considered generally applicable methods as no prior information about the data is used during the calibration. A comparison on simulation data has been carried out previously by Wang et al. [14] who found that whilst the two methods could be considered as mathematically equivalent, they have different noise propagation properties. The data used here are real laboratory measurements of the flow injection analysis (FIA) of hydroxybenzaldehyde isomers [11], which in the presence of a pH gradient are dissociated into basic and acidic forms.

It should be noted here that better results for this data set can be obtained by least squares-based methods which take advantage of constraints (non-negativity, unimodality) on the estimated profiles, prior information about the analyte (spectra and time profiles found by curve resolution in a prior analysis) and techniques which allow the reduction of the collinearity influence over the performance of the method [15,16]. These methods do, however, require knowledge about the chemical system which may often not be available.

2. Experimental

The FIA system used in this investigation has been described previously in the literature [11] and is shown schematically in Fig. 1. Polypropylene tubes (0.70 cm internal diameter) were used throughout. The carrier stream was a Britton–Robinson buffer with a pH of 4.5 and the reagent stream a Britton–Robinson buffer with pH 11.4.

The sample was injected by an ABU 80 autoburette (0.375 ml min^{-1}) between the carrier and the reagent stream as shown in Fig. 1. As the sample volume is small (77 μl) compared to the carrier stream and the reagent (770 μl), a smooth pH gradient is created over the sample plug due to dispersion of the carrier (low pH) and the reagent (high pH) stream.
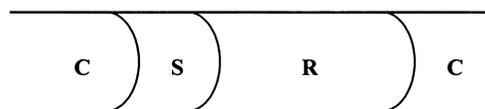

Fig. 1. Schematic for the flow injection analysis system. C = carrier stream (Britton–Robinson buffer, pH = 4.5); S = sample (77 μl); and R = reagent (770 μl, Britton–Robinson buffer, pH = 11.4).
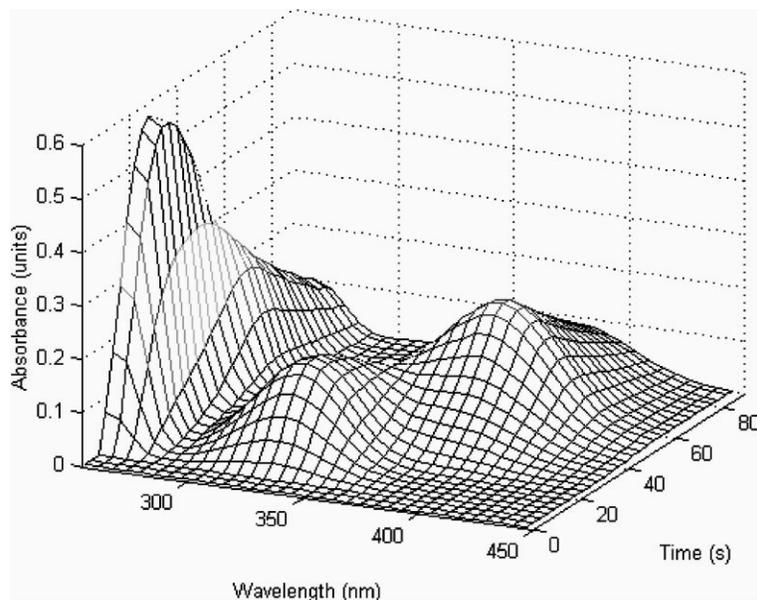
Fig. 2. Example of flow injection analysis data (2-HBA standard).

The sample is led into an 8 μl flow cell and then measured using a HP 8452A photodiode array spectrophotometer. The sample is measured for 88 s at 1 s intervals, 20 s after injection, and from 254 to 450 nm at 2 nm intervals. The second-order data obtained from each sample is, thus, of size 89 × 99, an example (the response of the pure solute 2-HBA, see below) being shown in Fig. 2.

The duration of detection and the pH gradient is sufficient to ensure that the analyte is present in both its acidic and basic form during detection. Ethanol–water solutions were used in preparing the carrier, reagent, and standards so that the final solutions were 1:9 ethanol–water (v/v). The Britton–Robinson buffer contained citric acid, potassium dihydrogenphospate, boric acid, and tri-(hydroxymethyl)amino-methane (TRIS) according to Perrin and Demsey [17]. TRIS was used instead of 5,5-diethylbarbituric acid to prevent absorption of the buffer in the ultraviolet region. The buffer concentration was 1.788 mM and the pH of the reagent solution was adjusted using sodium hydroxide.

The test solutes 2-hydroxybenzadehyde (2-HBA), 3-hydroxybenzadehyde (3-HBA) and 4-hydroxybenzadehyde (4-HBA) show different absorption spectra depending on whether they are in their acidic or basic form. Fig. 3 illustrates this for the example of 2-HBA. Theoretically there is no separation of the constituents of the sample since FIA is not a chromatographic system but a transportation system. The shape of the concentration profile of a specific solute is the same as for the sample as such, but due to the pH gradient, the first part of the sample plug is dominated by protonated solutes and the last part by deprotonated solutes. Depending on the pK_a of a given solute it will exhibit different acidic and basic profiles in the sample plug, as shown in Fig. 4 for all three solutes. The pK_a values of 2-HBA, 3-HBA and 4-HBA are 8.37, 8.98 and 7.61, respectively [18].

3. Description of the data

The spectra of the acidic forms of 2-HBA, 3-HBA and 4-HBA are denoted as sa_2, sa_3 and sa_4, respectively, and the spectra of the basic forms denoted as sb_2, sb_3 and sb_4, respectively. The subscript '2', '3' or '4' refers to 2-HBA, 3-HBA or 4-HBA and the extension 'a' or 'b' to the acidic or basic form. Likewise, the concentration profiles of the acidic forms of 2-HBA, 3-HBA and 4-HBA are denoted by ca_2, ca_3 and ca_4, and the concentration profiles of the basic
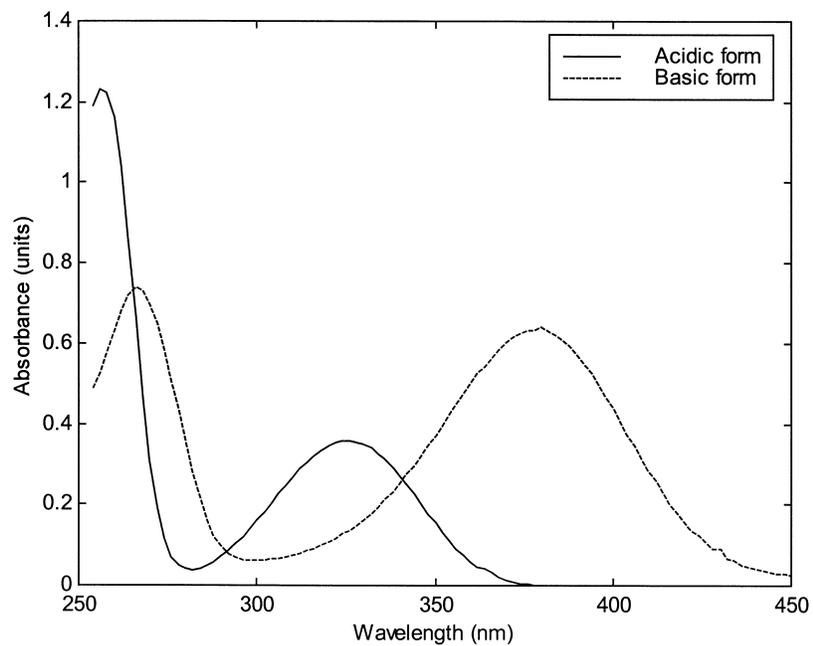
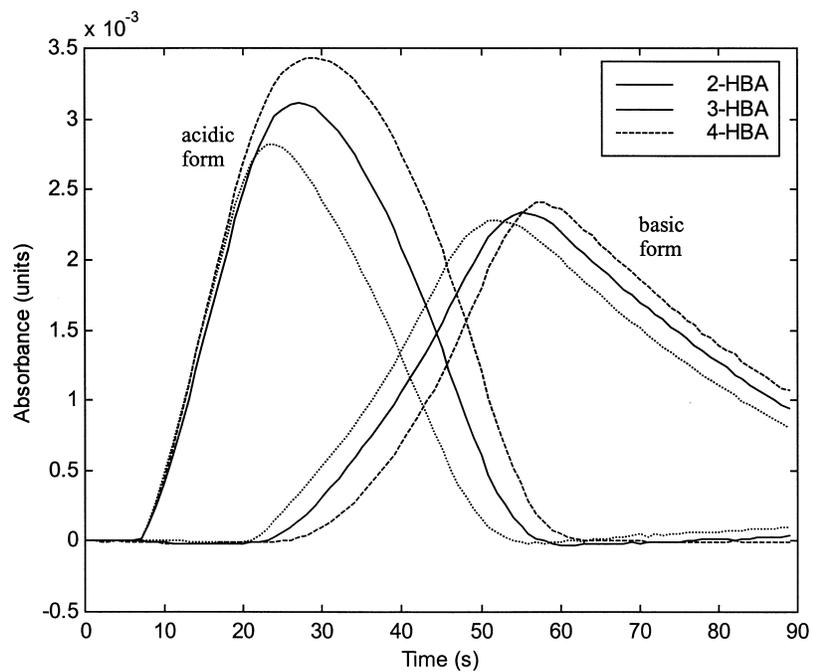Fig. 3. Spectra of 2-HBA in its acidic and basic forms.

Fig. 4. Acidic (on left) and basic time profiles for 2-HBA, 3-HBA and 4-HBA.

forms by $\mathbf{cb}_2$, $\mathbf{cb}_3$ and $\mathbf{cb}_4$. These spectra and concentration profiles refer to unit concentrations of the solutes unless stated otherwise.

The measured responses of the pure solutes 2-HBA, 3-HBA and 4-HBA are denoted as $\mathbf{N}_{\text{2-HBA}}$, $\mathbf{N}_{\text{3-HBA}}$ and $\mathbf{N}_{\text{4-HBA}}$, respectively. These matrices have dimensions 89×99, this being the number of time points by the number of wavelengths. Assuming Beer's law for the measurements, the response of the pure solute 2-HBA at unit concentration can be written as follows:

$$\mathbf{N}_{\text{2-HBA}} = \mathbf{ca}_2 \cdot \mathbf{sa}_2^{\text{T}} + \mathbf{cb}_2 \cdot \mathbf{sb}_2^{\text{T}} + \mathbf{E}_{\text{2-HBA}} \qquad (1)$$

where $\mathbf{E}_{\text{2-HBA}}$ is the measurement error.

4. Calibration

The aim of the calibration is to quantify an analyte of interest in the presence of one (or more) unknown interferent(s) (the so-called 'second-order advantage' [2]). In this work, only the quantification of an analyte in a binary mixture is considered. For that, two matrices are used: a standard, $\mathbf{N}$, where only the analyte is present with a known concentration, and a binary mixture, $\mathbf{M}$, consisting of the analyte at an unknown concentration (to be found by the calibration method) and an unknown interferent.

5. Rank overlap

During the analysis, the total concentration profile of a compound (acidic and basic form) is given by $\mathbf{ca} + \mathbf{cb}$. This gives three total concentration profiles: $\mathbf{ctot}_2$ ($= \mathbf{ca}_2 + \mathbf{cb}_2$), $\mathbf{ctot}_3$ ($= \mathbf{ca}_3 + \mathbf{cb}_3$) and $\mathbf{ctot}_4$ ($= \mathbf{ca}_4 + \mathbf{cb}_4$) for 2-HBA, 3-HBA and 4-HBA, respectively. These are shown in Fig. 5. The shape of the total concentration profile is defined by the dispersion properties of the FIA channel (and is independent of molar absorptivity of the species or the type of detector used). Since the solutes resemble each other very much, it can be expected that the dispersion behaviour is equal for all three [11]. Hence, the shape of the total concentration profiles is equal (i.e. $\mathbf{ctot}_2 = \alpha \cdot \mathbf{ctot}_3 = \beta \cdot \mathbf{ctot}_4$, where α and β are constants). This phenomenon puts a restriction on the calibration problem and destroys the rank linear additivity of the system, resulting in a rank overlap between the time profiles.
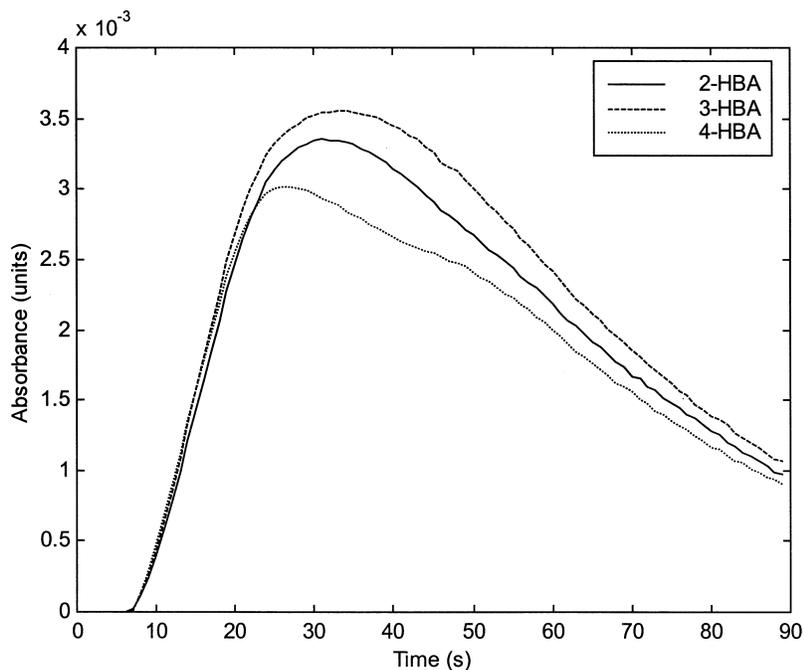

Fig. 5. Total concentration profiles for 2-HBA, 3-HBA and 4-HBA.

For an algebraic description of the rank overlap, an example of a calibration of 2-HBA as standard and a mixture containing 2-HBA as analyte in presence of an unknown interferent (e.g. 3-HBA or 4-HBA) is used. The mixture can be written as follows:

$$M = \gamma \cdot \mathbf{ca}_2 \cdot \mathbf{sa}_2^T + \gamma \cdot \mathbf{cb}_2 \cdot \mathbf{sb}_2^T + \mathbf{ca}_u \cdot \mathbf{sa}_u^T$$
$$+ \mathbf{cb}_u \cdot \mathbf{sb}_u^T + E_M = \tilde{M} + E_M \qquad (2)$$

where M is the experimental response in presence of measurement noise, $\tilde{M}$ the experimental response in the ideal case without measurement error, E_M the measurement error, γ the concentration of the analyte in the mixture and the subscript 'u' indicates the unknown interferent. The concentration information for the interferent is absorbed in $\mathbf{ca}_u$ and $\mathbf{cb}_u$ without loss of generality.

Given that the total concentration profiles for analyte 2-HBA and the unknown interferent have the same shape, as expressed in Eqs. (3) and (4):

$$\mathbf{ctot}_2 = \mathbf{ca}_2 + \mathbf{cb}_2 \qquad (3)$$

$$\alpha \cdot \mathbf{ctot}_2 = \mathbf{ca}_u + \mathbf{cb}_u \qquad (4)$$

where α is a scalar constant, then the standard and mixture responses given by Eqs. (1) and (2) may be rewritten by eliminating $\mathbf{ca}_2$ and $\mathbf{cb}_u$ using Eqs. (3) and (4) as follows:

$$N = \mathbf{ctot}_2 \cdot \mathbf{sa}_2^T + \mathbf{cb}_2(\mathbf{sb}_2^T - \mathbf{sa}_2^T) + E_N \qquad (5)$$

$$
\begin{aligned}
M &= \gamma \cdot \mathbf{ctot}_2 \cdot \mathbf{sa}_2^T + \gamma \cdot \mathbf{cb}_2(\mathbf{sb}_2^T - \mathbf{sa}_2^T) \\
&\quad + \alpha \cdot \mathbf{ctot}_2 \cdot \mathbf{sb}_u^T + \mathbf{ca}_u(\mathbf{sa}_u^T - \mathbf{sb}_u^T) + E_M \\
&= \mathbf{ctot}_2(\gamma \cdot \mathbf{sa}_2^T + \alpha \cdot \mathbf{sb}_u^T) + \gamma \cdot \mathbf{cb}_2(\mathbf{sb}_2^T - \mathbf{sa}_2^T) \\
&\quad + \mathbf{ca}_u(\mathbf{sa}_u^T - \mathbf{sb}_u^T) + E_M = [\mathbf{ctot}_2|\mathbf{cb}_2|\mathbf{ca}_u] \\
&\quad \times \begin{bmatrix} \gamma \cdot \mathbf{sa}_2^T + \alpha \cdot \mathbf{sb}_u^T \\ \gamma \cdot (\mathbf{sb}_2^T - \mathbf{sa}_2^T) \\ \mathbf{sa}_u^T - \mathbf{sb}_u^T \end{bmatrix} + E_M \qquad (6)
\end{aligned}
$$

Eq. (6) shows that the pseudo-rank of M, which corresponds to the actual rank of $\tilde{M}$ (rank(M) > pseudo-rank(M) due the measurement error), is actually three, because $\tilde{M}$ can be described in terms of matrices of rank 3. $\tilde{M}$ is said to have a rank overlap of one, since it is the result of a sum of two matrices both having pseudo-rank 2 and its psedu-rank in 3, because the rank linear additivity of the system does not hold.

6. Calibration methods

The two methods used to quantify the analyte in the presence of an unknown interferent were chosen for their algebraic characteristics, one being eigenvalue-based (NBRA) and the other least squares-based (RBL). The aim is to verify if these characteristics provide stable estimations despite the non-ideal behavior of the experimental data (i.e. rank overlap, collinearity, reproducibility, spectral noise). For this purpose, the methods must present a general treatment of the problem, with no prior information about the data being used during the calibration. For NBRA, the only assumption is that the rank of the mixture must be known. This can be found by rank analysis techniques [19], although in this work the rank is known using prior knowledge of the data. For RBL, the rank of the interferent can be determined within the RBL algorithm [3], although again, here external knowledge of the rank of the interferent is used so as to increase the speed of the algorithm. Including the rank values as prior information does not affect the generality of the methods, but reduces the computational time spent for each analysis.

6.1. NBRA

NBRA is explained here by considering the case where the pseudo-rank of the mixture is equal to three and the analyte equal to two, without losing the generality of the method which can be applied to mixtures having different number of interferents. Following the general representation of the rank overlap problem suggested by Kiers and Smilde [10], Eqs. (5) and (6) can be put in terms of Eqs. (10) and (11) via Eqs. (7)–(9).

$$X_r = [\mathbf{ctot}_2|\mathbf{cb}_2] \qquad (7)$$

$$Y_r = [\mathbf{sa}_2|(\mathbf{sb}_2 - \mathbf{sa}_2)] \qquad (8)$$

$$D_r = \begin{bmatrix} \gamma & 0 \\ 0 & \gamma \end{bmatrix} \qquad (9)$$

$$N = X_r Y_r^T + E_N \qquad (10)$$

$$M = X_r D_r Y_r^T + X_s Y_s^T + X_t Y_t^T + E_M \qquad (11)$$

where X_r ($I \times 2$) and Y_r ($J \times 2$) give the profiles for the analyte; D_r gives the concentration ratio of the

analyte in the mixture with respect to the analyte in the standard; X_s $(I \times 1) = \mathbf{ctot}_2$ and Y_s $(J \times 1) = \alpha \cdot \mathbf{sb}_u$ denote the profiles for the analyte which have rank overlap with the interferent (the column of X_s equals the first column of X_r); finally, X_t $(I \times 1) = \mathbf{ca}_u$ and Y_t $(J \times 1) = \mathbf{sa}_u - \mathbf{sb}_u$ denote the profiles for analytes that have no rank overlap with previous analytes.

The NBRA algorithm used in this work was formulated in terms of the GRAM method described by Sanchez and Kowalski [7], where the following eigenvalue problem must be solved:

$$\bar{U}^{\mathrm{T}} M \bar{V} \bar{S}^{-1} Z = Z \Lambda \tag{12}$$

where $\bar{U}$, $\bar{S}$ and $\bar{V}$ are the results, truncated to three components, from an SVD of W, where

$$W = M + N \tag{13}$$

and so

$$\bar{W} = \bar{U} \bar{S} \bar{V}^{\mathrm{T}} \tag{14}$$

where the pseudo-rank of W is equal to the pseudo-rank of the mixture, M (given that the standard, N, is also present in the mixture). The solution to this eigenvalue problem is shown in Appendix A. Three eigenvalues are found, of which the smallest is the eigenvalue used for finding the analyte concentration (see Eq. (A.7)).

Summary of NBRA:
1. Solve the eigenvalue problem given in Eq. (12) — only the pseudo-rank of M is required.
2. Use the smallest eigenvalue to find γ, and thus, the analyte concentration.

6.2. RBL

Residual bilinearization is a least squares method that considers the residual signal in the data (i.e. after the analyte has been subtracted) to have a bilinear structure. In this case, the residual signal is the sum of the interferent(s) and experimental noise. The aim of RBL is to minimize the loss function given by Eq. (15):

$$E = ||M - \gamma N - PQ^{\mathrm{T}}||^2 \tag{15}$$

where P and Q are the scores and loadings from a principal component decomposition of $(M - \gamma N)$. The rank of the interferent must be known, but can be found by increasing the estimated rank of the interferent, starting at one, until the variance of residuals equals the measurement noise [3], or by using prior information as done in this work. For the data described here, the rank of the interferent in the binary mixtures is two. The algorithm, given in Appendix B, converges when the principal components P span the column space of the interferent and components Q span the row space of the interferent, i.e.

$$P_{\mathrm{spans}}([X_s X_t]), \qquad Q_{\mathrm{spans}}([Y_s Y_t]) \tag{16}$$

7. Validation

Validation of the calibrations was performed using a resampling procedure [20]. Subsets of the full 89×99 matrices were generated systematically, 16 submatrices (67×75) being generated for each full matrix. The first data set of the 16 corresponded to the wavelengths $2, 3, 4, 6, 7, 8, \ldots$ (the wavelengths $1, 5, 9, \ldots$ being left out) and the times $2, 3, 4, 6, 7, 8, \ldots$ (the times $1, 5, 9, \ldots$ being left out). For each set of wavelengths, four sets of different times were constructed and validated. In total, 16 concentration values were calculated for each mixture being calibrated, the final concentration value being the mean of these 16.

8. Calibration results

The calibration results for six binary mixtures formed by combinations of the three isomers (2-HBA, 3-HBA and 4-HBA) using both NBRA and RBL are shown in Table 1. This table presents the relative error percent of the calibration found by using the formula: error% $= 100 \times (c_{\mathrm{est}} - c_{\mathrm{true}})/c_{\mathrm{true}}$, where c_{est} is the estimated concentration and c_{true} the true concentration. Note that each row gives the results of *two* calibrations. For example, the row for mixture 1 considers both the case where 3-HBA is the analyte and 4-HBA the interferent, and the case where 4-HBA is the analyte and 3-HBA the interferent.

Each estimation is the mean value of 16 concentrations found by the validation step. The standard deviation around the mean of 16 concentrations predicted in each validation set varied between 0.05 and 0.87% of the concentration value being predicted for all but

Table 1
Calibration of six binary mixtures of 2-HBA, 3-HBA and 4-HBA using NBRA and RBL[a]

No.	Mixture (μM)			Relative error (%)					
	2HBA	3HBA	4HBA	NBRA			RBL		
				2HBA	3HBA	4HBA	2HBA	3HBA	4HBA
1	0	100	40	–	−1.09	10.94	–	−0.87	13.32
2	0	100	60	–	15.50	−3.83	–	17.39	−5.50
3	50	0	60	−0.54	–	−5.97	−13.09	–	7.84
4	100	0	60	2.22	–	−7.26	0.86	–	12.47
5	50	100	0	79.94	68.86	–	20.29	19.44	–
6	100	50	0	15.53	288.59	–	13.55	74.54	–

[a] A negative value indicates that the predicted concentration was smaller than its real value.

two of the calibrations, these being 1.43% for the calibration of 3-HBA (50 μM) in the presence of 2-HBA (100 μM) using NBRA and 4.44% for the same calibration using RBL. The small values of these standard deviations shows that spectral noise does not have a large influence on the calibration results, as the results are almost equal for calibrations using different wavelengths.

The results for the calibrations of 4-HBA as an analyte in the presence of 2-HBA or 3-HBA are generally good and similar for both methods. For the calibration of 2-HBA or 3-HBA as the analyte with 4-HBA as the interferent the results do not differ much between the methods. However, for the calibration of mixtures consisting of 2-HBA and 3-HBA, the results are not good, although RBL does give better results than NBRA. A detailed explanation for this is given in the following sections.

9. Exploratory analysis

The purpose of the exploratory analysis described here is to investigate why the mixtures of 2-HBA and 3-HBA yielded poor calibration results. The time profiles of the acidic and basic forms of 2-HBA and 3-HBA in both the standards and mixtures are found using two mode-component analysis (TMCA) [21] and the reason that rank overlap affects the estimation of these time profiles is considered.

The principle of TMCA is to decompose a matrix L into three full rank matrices and a matrix of residuals:

$$L = WKZ^T + E \tag{17}$$

where the matrix dimensions are L ($n \times m$), W ($n \times r$), K ($r \times q$), Z ($m \times q$) and E ($n \times m$). Note that a special case of TMCA is the singular value decomposition (SVD), where W and Z are orthogonal matrices and K is a square ($q = r$), diagonal matrix with non-negative elements. A set of restrictions are used here for the TMCA, in which the K and Z matrices are restricted to equal known parameters.

For the FIA data, the matrix containing the spectral response of the experiment (time versus wavelength) is decomposed into a time profile matrix, a concentrations matrix and a spectral profile matrix, where the last two matrices are known a priori. The time profiles of a standard, for example, 2-HBA, are found by solving the following least squares problem:

$$\min ||N_{2\text{-HBA}} - X_{2\text{-HBA}} D_{2\text{-HBA}} Y_{2\text{-HBA}}^T||^2 \tag{18}$$

where

$$X_{2\text{-HBA}} = [\mathbf{ca_2} | \mathbf{cb_2}] \tag{19}$$

$$D_{2\text{-HBA}} = \begin{bmatrix} \gamma_{2\text{-HBA}} & 0 \\ 0 & \gamma_{2\text{-HBA}} \end{bmatrix} \tag{20}$$

where $\gamma_{2\text{-HBA}}$ is the actual concentration of 2-HBA, and

$$Y_{2\text{-HBA}} = [\mathbf{sa_2} | \mathbf{sb_2}] \tag{21}$$

For a mixture of, for example, 2-HBA and 3-HBA, the time profiles are found by solving

$$\min ||M - XDY^T||^2 \tag{22}$$

where

$$X = [\mathbf{ca_2} | \mathbf{cb_2} | \mathbf{ca_3} | \mathbf{cb_3}] \tag{23}$$

$$D = \begin{bmatrix} \gamma_{\text{2-HBA}} & 0 & 0 & 0 \\ 0 & \gamma_{\text{2-HBA}} & 0 & 0 \\ 0 & 0 & \gamma_{\text{3-HBA}} & 0 \\ 0 & 0 & 0 & \gamma_{\text{3-HBA}} \end{bmatrix} \quad (24)$$

and

$$Y = [\mathbf{sa}_2 | \mathbf{sb}_2 | \mathbf{sa}_3 | \mathbf{sb}_2] \quad (25)$$

The UV-spectra of the pure compounds used in the TMCA, Y, were obtained by an auxiliary experiment performed for each solute at different pH values corresponding to the acidic and basic conditions in the FIA experiment [15]. The actual concentrations of the solutes contained in C are also known in advance.

Using TMCA, the time profiles for 2-HBA and 3-HBA in both the standards and a mixture (number 6 in Table 1) were calculated. These are shown in Fig. 6 in terms of the total concentration profiles (i.e. basic and acidic time profiles added). The shape of the curves found for the standards agree with the curve expected for the FIA system. However, the curves found for 2-HBA and 3-HBA in the mixture present shapes that do not have physical meaning in terms of the FIA system.

The reason that finding the X matrix by solving the Eq. (22) results in the wrong time profiles can be explained as follows. The M matrix has a rank >3 due to the presence of measurement noise (rank(M) > pseudo-rank(M)). Using matrices D and Y, which are of full rank (and have, therefore, a rank 4) for solving the least squares problem in Eq. (22) implies that the estimated X also has rank 4. However, as the true X has rank 3 (like the pseudo-rank of M — see Eq. (6)), then fitting X with rank 4 results in noise modelling.

The correct TMCA model should be formulated taking into account the rank overlap present in the data. Eq. (6) can be reformulated as follows:

$$M = [\mathbf{ctot}_2 | \mathbf{cb}_2 | \mathbf{ca}_\text{u}] \times \begin{bmatrix} \gamma \cdot \mathbf{sa}_2^\text{T} + \alpha \cdot \mathbf{sb}_\text{u}^\text{T} \\ \gamma \cdot (\mathbf{sb}_2^\text{T} - \mathbf{sa}_2^\text{T}) \\ \mathbf{sa}_\text{u}^\text{T} - \mathbf{sb}_\text{u}^\text{T} \end{bmatrix} + E_\text{M}$$

$$= [\mathbf{ctot}_2 | \mathbf{cb}_2 | \mathbf{ca}_\text{u}] \times \begin{bmatrix} \gamma & 0 & 0 & \alpha \\ -\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

$$\times \begin{bmatrix} \mathbf{sa}_2^\text{T} \\ \mathbf{sb}_2^\text{T} \\ \mathbf{sa}_\text{u}^\text{T} \\ \mathbf{sb}_\text{u}^\text{T} \end{bmatrix} + E_\text{M} = X^* D^* Y^\text{T} + E_\text{M} \quad (26)$$

which can be checked by matrix multiplication. Note that the matrices X and D have been updated and that X^* has rank 3, made explicit now because X^* has three columns, instead of X, which had four linearly independent columns.

The TMCA model is estimated using the known spectra, Y, and known concentration ratio, γ. This results in X^* and α. Fig. 7 shows the new total concentration profiles, which are almost equal when calculated from either the standards or the mixture. This shows that the property of linear additivity, in accordance with the Lambert–Beer law, holds good for the data.

10. Discussion

The basic profiles of 2-HBA and 3-HBA in the standards and in the mixture as calculated by the reformulated TMCA are shown in Fig. 8, where it is seen that the profiles in the mixture are highly collinear. To study the relationship between the collinearity among the time profiles of the dissociated species and the isomer dissociation constants, the dissociation of the isomers represented by Reaction (1) and the dissociation constants $k_{\text{a}-2}$ of 2-HBA and $k_{\text{a}-3}$ of 3-HBA given by Eqs. (27) and (28) are considered:

$$\text{AH} + \text{H}_2\text{O} \overset{k_\text{a}}{\rightleftharpoons} \text{A}^- + \text{H}_3\text{O}^+ \qquad \text{Reaction (1)}$$
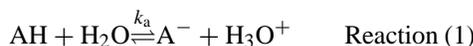
$$k_{\text{a}-2} = \frac{[\text{A}^-]_{2,t}[\text{H}_3\text{O}^+]_t}{[\text{AH}]_{2,t}} \quad (27)$$

$$k_{\text{a}-3} = \frac{[\text{A}^-]_{3,t}[\text{H}_3\text{O}^+]_t}{[\text{AH}]_{3,t}} \quad (28)$$

where $[\text{A}^-]_{2,t}$ and $[\text{A}^-]_{3,t}$ are the concentrations of the basic forms of 2-HBA and 3-HBA at time t (or, alternatively, at a certain point in the FIA channel) and $[\text{AH}]_{2,t}$ and $[\text{AH}]_{3,t}$ are the concentrations of the acidic forms of 2-HBA and 3-HBA at time t.

Using the fact that the total concentrations of 2-HBA and 3-HBA are proportional

$$\text{ct}_t = [\text{A}^-]_{2,t} + [\text{AH}]_{2,t} \quad (29)$$

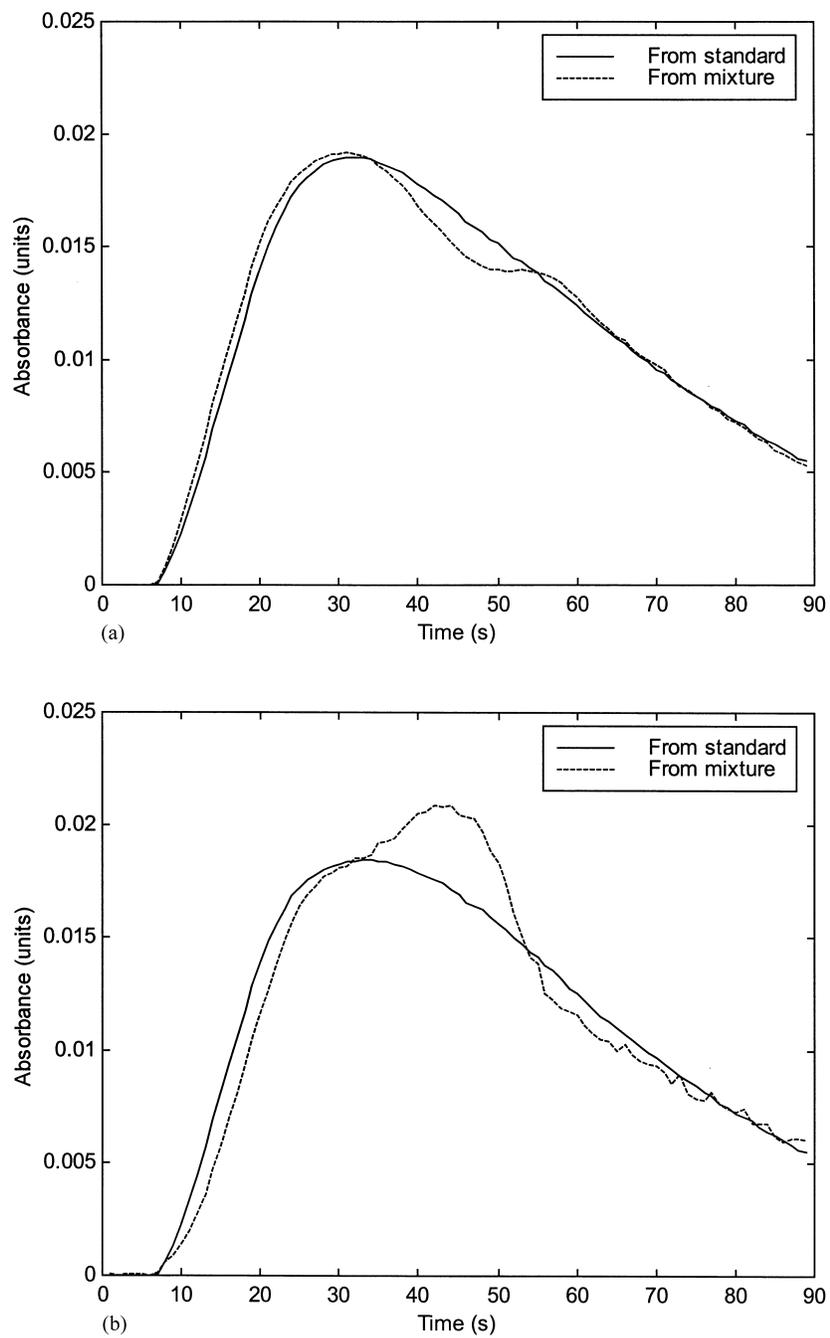$$\alpha \cdot \text{ct}_t = [\text{A}^-]_{3,t} + [\text{AH}]_{3,t} \quad (30)$$

Fig. 6. Total concentration profiles from a standard and a mixture as found by TMCA for (a) 2-HBA ($\mathbf{ca}_2 + \mathbf{cb}_2$) and (b) 3-HBA ($\mathbf{ca}_3 + \mathbf{cb}_3$).
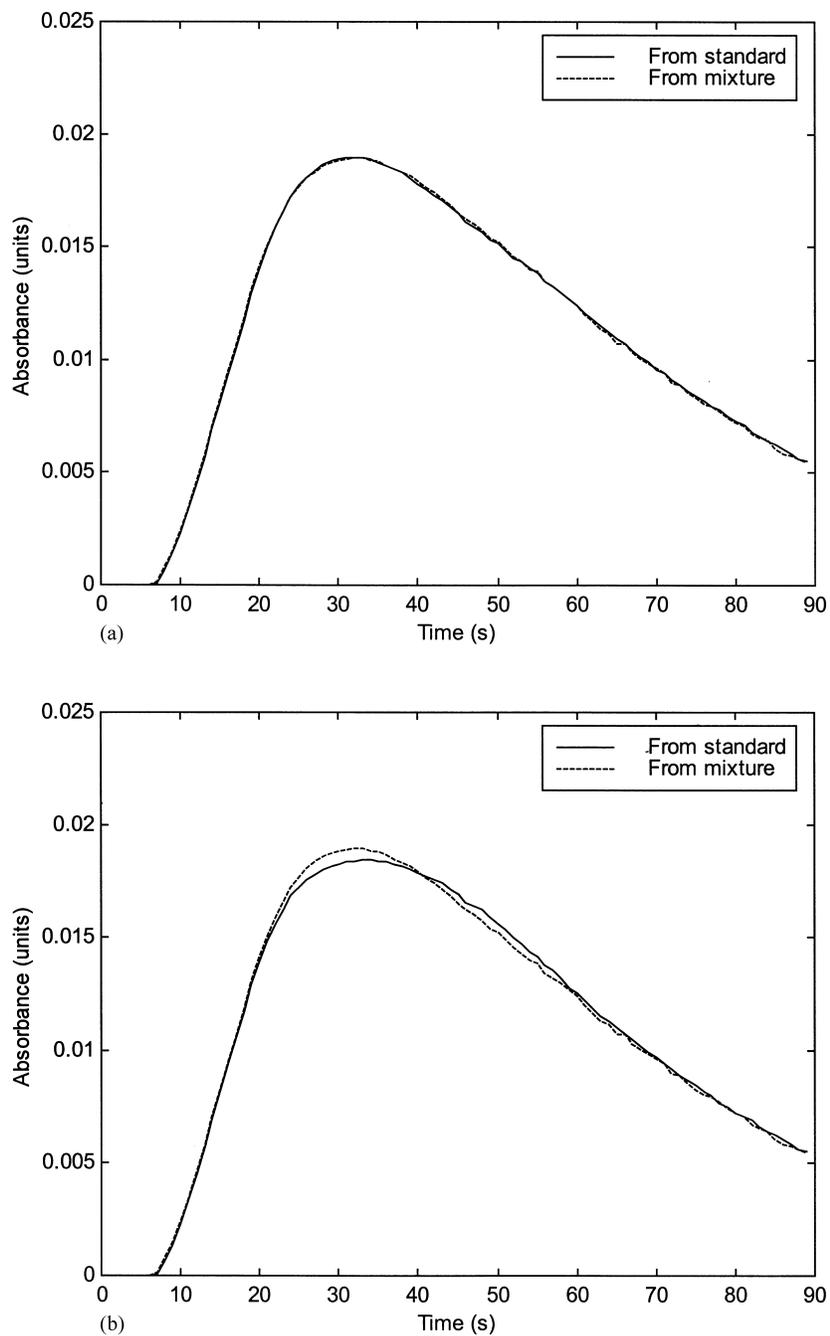
Fig. 7. Total concentration profiles from a standard and a mixture as found by the reformulated TMCA for (a) 2-HBA ($\mathbf{ca}_2 + \mathbf{cb}_2$) and (b) 3-HBA ($\mathbf{ca}_3 + \mathbf{cb}_3$).
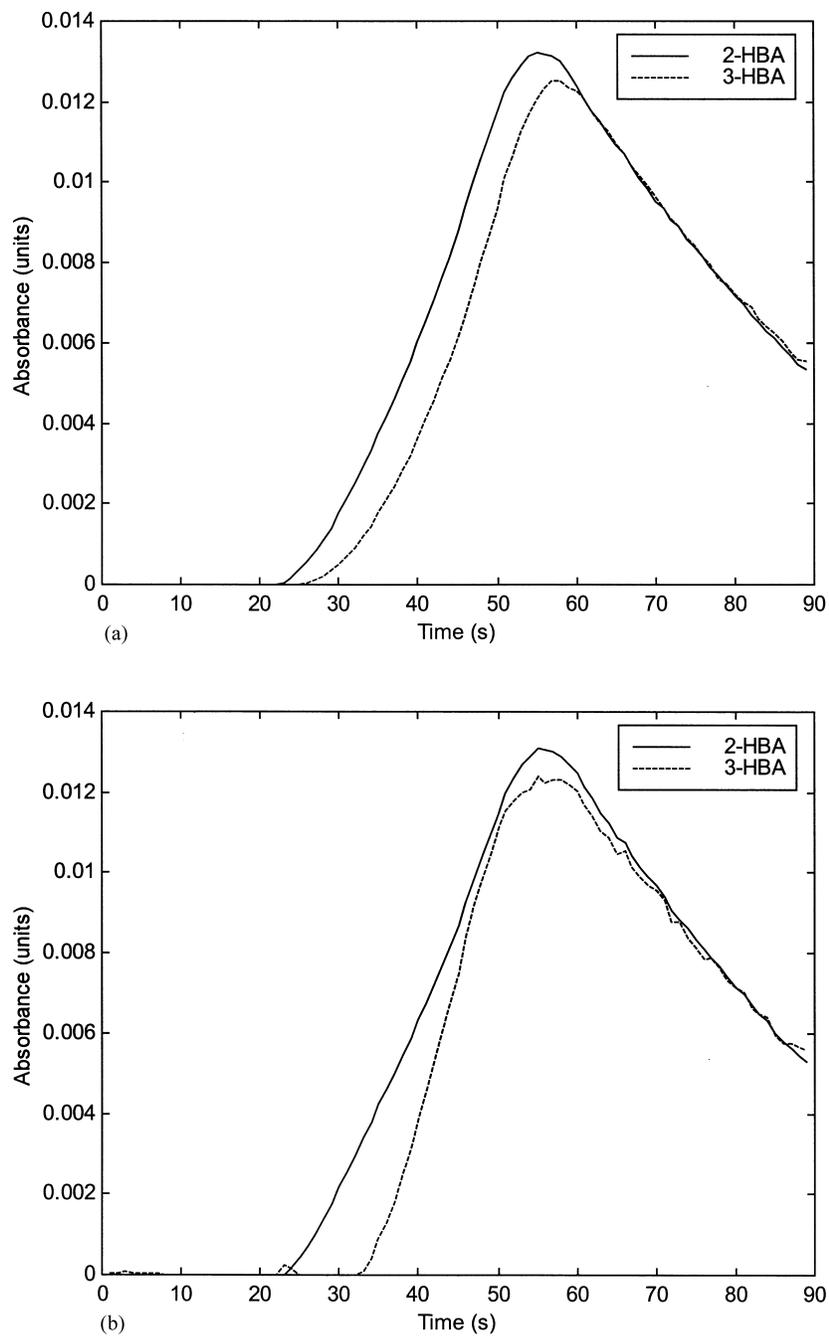
Fig. 8. Time profiles of the basic species of 2-HBA and 3-HBA as found by the reformulated TMCA (a) in the standards and (b) in the mixture.

where ct_t is the total concentration at time t and α the scalar constant, the relationship between the basic profiles of 2-HBA and 3-HBA can be found as follows:

$$[H_3O^+]_t = \frac{k_{a-2}(ct_t - [A^-]_{2,t})}{[A^-]_{2,t}}$$

$$[H_3O^+]_t = \frac{k_{a-3}(\alpha \cdot ct_t - [A^-]_{3,t})}{[A^-]_{3,t}}$$

$$\frac{k_{a-2}(ct_t - [A^-]_{2,t})}{[A^-]_{2,t}} = \frac{k_{a-3}(\alpha \cdot ct_t - [A^-]_{3,t})}{[A^-]_{3,t}}$$

$$[A^-]_{3,t} = \frac{(\alpha \cdot ct_t) \cdot [A^-]_{2,t} k_{a-3}}{k_{a-2}ct_t + [A^-]_{2,t}(k_{a-3} - k_{a-2})} \tag{31}$$

Eq. (31) shows that the difference between the basic profiles of the isomers depends upon the values of the dissociation constants, k_{a-2} and k_{a-3}. In cases where k_a for the two isomers are very similar, the term '$k_{a-3} - k_{a-2}$' in Eq. (31) is close to zero and the time profiles of 2-HBA and 3-HBA become collinear (note that a similar analysis holds for the acidic profiles). Considering the values of k_a for the isomers ($k_{a-2} = 4.27 \times 10^{-9}$, $k_{a-3} = 1.05 \times 10^{-9}$ and $k_{a-4} = 24.55 \times 10^{-9}$), it is possible to explain the high collinearity between profiles of 2-HBA and 3-HBA in the mixture, in other words, the mixtures of 2-HBA and 3-HBA are more difficult to calibrate due to the collinearity between the time profiles of interferent and analyte which can be attributed to the similarity between the k_a values of 2-HBA and 3-HBA.

The basic time profile of 3-HBA in the mixture is shifted in comparison to the same profiles in the 3-HBA standard (compare Fig. 8(a) and (b)). This shift can be explained by taking as an example the dissociation shown in Reaction (1) where the concentration of the basic species at time t is calculated in terms of the dissociation constant given in Eq. (28), the total concentration at time t as given in Eq. (30) and the pH at time t given as

$$pH_t = -\log_{10}([H_3O^+]_t) \tag{32}$$

The shape of the pH gradient as present in the sample plug depends on the length of the sample zones in the FIA system. The sample zones for duplicates of an FIA experiment for the same sample are illustrated in Fig. 9, where w_1 is the length of the sample
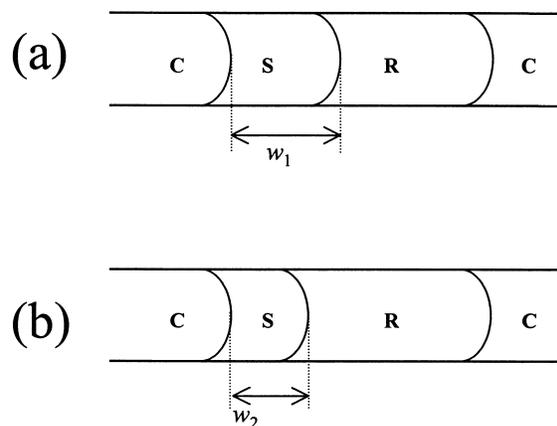

Fig. 9. (a) Duplicate one of the FIA experiment where w_1 is the length of the sample zone; (b) duplicate two of the FIA experiment where w_2 is the length of the sample zone; C = carrier stream (Britton–Robinson buffer, pH = 4.5); S = sample (77 μl); and R = reagent (770 μl, Britton–Robinson buffer, pH = 11.4).

zone for the first duplicate and w_2 for the second duplicate ($w_1 > w_2$). Considering that it is not possible to reproduce the exact length of the sample zones, the shape of the pH gradient will induce a small variation between the two duplicates as illustrated by Fig. 10. This variation in the pH at the time t for the two duplicates is represented by

$$[H_3O^+]_{\text{dupl-2},t} = [H_3O^+]_{\text{dupl-1},t} + \delta_t \tag{33}$$

where the subscripts 'dupl-1' and 'dupl-2' indicate the first and second duplicate, respectively, and δ_t is the difference between the $[H_3O^+]$ concentration of the duplicates at time t.

New expressions for the concentration at time t of the basic form of 3-HBA for the two duplicates can now be written:

$$[A^-]_{3,\text{dupl-1},t} = \frac{\alpha \cdot ct_t}{10^{pK_a - pH} + 1} \tag{34}$$

$$[A^-]_{3,\text{dupl-2},t} = \frac{\alpha \cdot ct_t}{10^{pK_a - pH_t} + \delta_t 10^{pK_a} + 1} \tag{35}$$

Eqs. (34) and (35) show that shift in the time profiles due to small changes in the shape of the pH gradient depends on the pK_a of the isomers, so that the higher the pK_a value, the larger the shifts in the time profiles. Given the problem of exactly duplicating the FIA sample zone length, it is apparent that solutes with higher
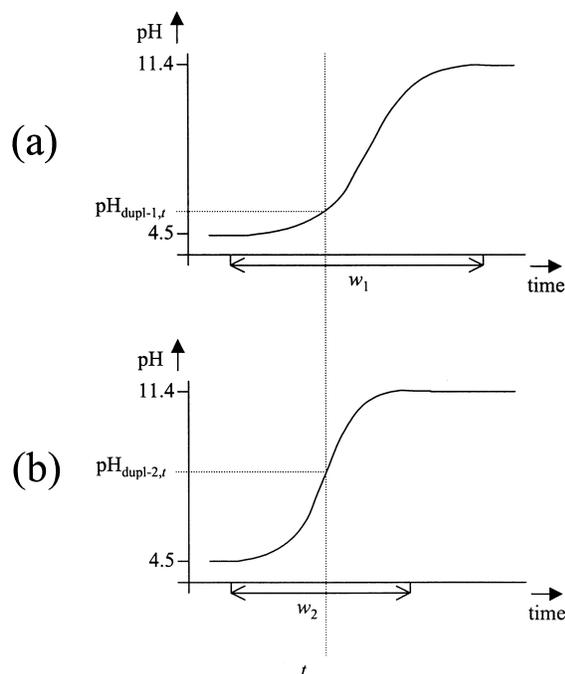
Fig. 10. The pH gradient shape and the time window of FIA measurement for (a) duplicate one of the FIA experiment, where w_1 is the length of the sample zone; (b) duplicate two of the FIA experiment where w_2 is the length of the sample zone.

pK_a values have a higher associated experimental error in terms of reproducibility. Considering the pK_a values of the three solutes (2-HBA = 8.37, 3-HBA = 8.98 and 4-HBA = 7.61), this explains why the calibration of 3-HBA, which has the highest pK_a among the three isomers, generally gives the worse results.

11. Conclusions

Using the second-order FIA system to analyse mixtures of solutes with similar dispersion properties (such as for the HBA isomers described here) introduces a rank overlap problem. Both second-order calibration methods described here, NBRA and RBL, have been shown to be capable of handling the rank overlap problem.

It is shown that similar solutes which also have similar pK_a values give collinear basic and acidic time profiles. Under these conditions, the eigenvalue-based method, NBRA, yielded poor results, because the

eigenvalue used to find the concentration of the analyte in the mixture depends upon how the eigenvectors are placed in the vector space of the mixture matrix. The collinearity between analyte and interferent results in an unstable subspace being defined by the eigenvectors and, thus, the subspace is more sensitive to the presence of experimental noise. The least squares solution, RBL, seemed to be more stable in the presence of collinearity, which is an important aspect if collinearity cannot be eliminated from the experimental data. This finding is in accordance with the work of Wang et al. [9] who suggested that RBL will generally have a better noise-filtering capability than NBRA.

Finally, it has been shown that the basic and acidic time profiles can experience a shift due to the difficulty of exactly reproducing the FIA sample zone length. Solutes with a high pK_a exhibit a larger shift and, therefore, a larger experimental error in terms of reproducibility. This shift will cause a lack of synchronisation between the time profiles within the standard and mixture matrices [22] and, therefore, a breakdown in the trilinear structure assumed by second-order calibration methods such as NBRA and RBL. This is seen in the calibration of 3-HBA in the presence of 2-HBA (mixture 6 in Table 1), for which both NBRA and RBL perform badly.

The success of second-order calibration of this data is found to depend strongly on the collinearity between the acidic and basic time profiles and the reproducibility of the pH gradient. By comparing the results obtained here with the results from using restricted Tucker models, multivariate curve resolution and PARATUCK2 presented in the literature [15], it is seen that for cases where collinearity and shift in the time profiles is not significant, both NBRA and RBL work comparably well, with NBRA having the advantage of being computationally faster.

pitality during the period in which this work was carried out.

Appendix A. Non-bilinear rank annihilation (NBRA)

To solve the eigenvalue problem posed by NBRA, as described in the main text, the first step is to express matrices $\boldsymbol{M}$ and $\boldsymbol{W}$ as follows:

$$\begin{aligned} \boldsymbol{M} &= \boldsymbol{X}_r \boldsymbol{D}_r \boldsymbol{Y}_r^T + \boldsymbol{X}_s \boldsymbol{Y}_s^T + \boldsymbol{X}_t \boldsymbol{Y}_t^T + \boldsymbol{E}_M \\ &= \boldsymbol{X}_r (\boldsymbol{Y}_r \boldsymbol{D}_r + [\boldsymbol{Y}_r | \boldsymbol{0}])^T + \boldsymbol{X}_t \boldsymbol{Y}_t^T + \boldsymbol{E}_M \end{aligned} \quad (A.1)$$

where the matrices are the same as defined for Eq. (11) in the main text and the column of $\boldsymbol{X}_s$ equals the first column of $\boldsymbol{X}_r$.

$$\begin{aligned} \boldsymbol{W} &= \boldsymbol{M} + \boldsymbol{N} = \boldsymbol{X}_r (\boldsymbol{Y}_r \boldsymbol{D}_r + [\boldsymbol{Y}_s | \boldsymbol{0}])^T \\ &\quad + \boldsymbol{X}_t \boldsymbol{Y}_t^T + \boldsymbol{E}_M + \boldsymbol{X}_r \boldsymbol{Y}_r^T + \boldsymbol{E}_N \\ &= \boldsymbol{X}_r (\boldsymbol{Y}_r (\boldsymbol{D}_r + \boldsymbol{I}) + [\boldsymbol{Y}_s | \boldsymbol{0}])^T \\ &\quad + \boldsymbol{X}_t \boldsymbol{Y}_t^T + \boldsymbol{E}_M + \boldsymbol{E}_N \end{aligned} \quad (A.2)$$

The solution for the eigenvalue problem shown in Eq. (A.3a) is found by solving the determinant in Eq. (A.3b), which can be rewritten in terms of Eq. (A.3c) by using $\bar{\boldsymbol{S}} = \bar{\boldsymbol{U}}^T \bar{\boldsymbol{W}} \bar{\boldsymbol{V}}$ from Eq. (14):

$$\bar{\boldsymbol{U}}^T \boldsymbol{M} \bar{\boldsymbol{V}} \bar{\boldsymbol{S}}^{-1} \boldsymbol{Z} = \boldsymbol{Z} \boldsymbol{\Lambda} \quad (A.3a)$$

$$|\bar{\boldsymbol{U}}^T \boldsymbol{M} \bar{\boldsymbol{V}} - \lambda \bar{\boldsymbol{S}}| = 0 \quad (A.3b)$$

$$|\bar{\boldsymbol{U}}^T (\boldsymbol{M} - \lambda \bar{\boldsymbol{W}}) \bar{\boldsymbol{V}}| = 0 \quad (A.3c)$$

The determinant shown in Eq. (A.3c) is reduced to the form shown in Eq. (A.5). This is done first by replacing $\boldsymbol{M}$ and $\boldsymbol{W}$ in Eq. (A.3c) by Eqs. (A.1) and (A.2), which results in Eq. (A.4).

$$\begin{aligned} \Big| \boldsymbol{U}^T [\boldsymbol{X}_r | \boldsymbol{X}_t] [\boldsymbol{Y}_r (\boldsymbol{D}_r - \lambda (\boldsymbol{D}_r + \boldsymbol{I})) \\ + [(1-\lambda)\boldsymbol{Y}_s | \boldsymbol{0}] | (1-\lambda)\boldsymbol{Y}_t]^T \boldsymbol{V} \Big| = 0 \end{aligned} \quad (A.4)$$

As the square matrix $\boldsymbol{U}^T (\boldsymbol{X}_r | \boldsymbol{X}_t)$ is non-singular, it can be removed from Eq. (A.4), which is thus reduced to Eq. (A.5).

$$\begin{aligned} |[\boldsymbol{Y}_r (\boldsymbol{D}_r - \lambda (\boldsymbol{D}_r + \boldsymbol{I})) + [(1-\lambda)\boldsymbol{Y}_s | \boldsymbol{0}] | (1-\lambda)\boldsymbol{Y}_t]^T \boldsymbol{V}| \\ = 0 \end{aligned} \quad (A.5)$$

The determinant in Eq. (A.5) has three eigenvalues as solutions (it is the determinant of a 3×3 matrix) which can found by setting the columns of the matrix $[\boldsymbol{Y}_r (\boldsymbol{D}_r - \lambda (\boldsymbol{D}_r + \boldsymbol{I})) + [(1-\lambda)\boldsymbol{Y}_s | \boldsymbol{0}] | (1-\lambda)\boldsymbol{Y}_t]$ equal to zero. The results are shown in Eqs. (A.6b), (A.7b) and (A.8):

$$\boldsymbol{y}_1 (\gamma - \lambda_1 (\gamma + 1)) + \boldsymbol{y}_3 (1 - \lambda_1) = 0 \quad (A.6a)$$

$$\frac{\boldsymbol{y}_3^T \boldsymbol{y}_1 \gamma + \boldsymbol{y}_3^T \boldsymbol{y}_3}{\boldsymbol{y}_3^T \boldsymbol{y}_3 + \boldsymbol{y}_3^T \boldsymbol{y}_1 (\gamma + 1)} = \lambda_1 \quad (A.6b)$$

$$\boldsymbol{y}_2 (\gamma - \lambda_2 (\gamma + 1)) = 0 \quad (A.7a)$$

$$\lambda_2 = \frac{\gamma}{\gamma + 1} \quad (A.7b)$$

$$\lambda_3 = 1 \quad (A.8)$$

where $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ are the columns of $\boldsymbol{Y}_r$ and $\boldsymbol{y}_3$ is the column of $\boldsymbol{Y}_s$.

The eigenvalue λ_2 from Eq. (A.7) is used to find γ the ratio of the concentration of the analyte in the mixture to that in the standard. λ_2 is the smallest eigenvalue among the three found as solutions to the eigenvalue problem of NBRA, as can be shown by considering the inequalities given by Eqs. (A.9) and (A.10):

$$\frac{\boldsymbol{y}_3^T \boldsymbol{y}_1 \gamma + \boldsymbol{y}_3^T \boldsymbol{y}_3}{\boldsymbol{y}_3^T \boldsymbol{y}_3 + \boldsymbol{y}_3^T \boldsymbol{y}_1 (\gamma + 1)} > \frac{\gamma}{\gamma + 1}, \quad \gamma > 0 \quad (A.9)$$

$$\frac{\gamma}{\gamma + 1} < 1, \quad \gamma > 0 \quad (A.10)$$

Appendix B. Residual bilinearization (RBL)

The RBL algorithm is given as an alternating least squares procedure with the following minimization function:

$$\min_{\gamma, \boldsymbol{P}, \boldsymbol{Q}} ||\boldsymbol{M} - \gamma \boldsymbol{N} - \boldsymbol{P} \boldsymbol{Q}^T||^2 \quad (B.1)$$

where γ is minimised using a vectorised regression and $\boldsymbol{P}$ and $\boldsymbol{Q}$ are minimised using the NIPALS algorithm.

B.1. Initialization

The procedure is begun by making an initial estimate for γ:

$$\gamma_0 = \text{vec}(\boldsymbol{N})^+ \text{vec}(\boldsymbol{M}) \quad (B.2)$$

where the vectorisation function is given as

$$\mathrm{vec}(\boldsymbol{x}_1 \ldots \boldsymbol{x}_i \ldots \boldsymbol{x}_I) = \begin{pmatrix} \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_i \\ \vdots \\ \boldsymbol{x}_I \end{pmatrix} \tag{B.3}$$

and $\boldsymbol{x}_i$ is a column vector.

B.2. Main loop

$\boldsymbol{P}$ and $\boldsymbol{Q}$ are estimated using the NIPALS algorithm, in alternation with the estimation of γ, until convergence is reached:

$$\boldsymbol{P}_k \boldsymbol{Q}_k^{\mathrm{T}} = \boldsymbol{M} - \gamma_k \boldsymbol{N} \tag{B.4}$$

$$\gamma_{k+1} = \mathrm{vec}(\boldsymbol{N})^+ (\mathrm{vec}\,(\boldsymbol{M}) - \mathrm{vec}(\boldsymbol{P}_k \boldsymbol{Q}_k^{\mathrm{T}})) \tag{B.5}$$

If $\gamma_{k+1} - \gamma_k$ is greater than a given convergence criterion, then $k = k + 1$ and another iteration is performed.

References

[1] T. Hirschfeld, Anal. Chem. 52 (1980) A297.

[2] K.S. Booksh, B.R. Kowalski, Anal. Chem. 66 (1994) A782.

[3] J. Öhman, P. Geladi, S. Wold, J. Chemometrics 4 (1990) 79.

[4] R. Bro, Chemometrics Intell. Lab. Syst. 38 (1997) 149.

[5] A.K. Smilde, Y. Wang, B.R. Kowalski, J. Chemometrics 8 (1994) 21.

[6] C.N. Ho, G.D. Christian, E.R. Davidson, Anal. Chem. 50 (1978) 1108.

[7] E. Sanchez, B.R. Kowalski, Anal. Chem. 58 (1986) 496.

[8] B.E. Wilson, W. Lindberg, B.R. Kowalski, J. Am. Chem. Soc. 111 (1989) 3797.

[9] Y. Wang, O.S. Borgen, B.R. Kowalski, M. Gu, F. Turecek, J. Chemometrics 7 (1993) 117.

[10] H.A.L. Kiers, A.K. Smilde, J. Chemometrics 9 (1995) 179.

[11] L. Nørgaard, C. Ridder, Chemometrics Intell. Lab. Syst. 23 (1994) 107.

[12] J. Saurina, S. Hernández-Cassou, R. Tauler, A. Izquierdo-Ridorsa, J. Chemometrics 12 (1998) 183.

[13] A.K. Smilde, R. Tauler, J.M. Henshaw, I.W. Burgess, B.R. Kowalski, Anal. Chem. 66 (1994) 3345.

[14] Y. Wang, O.S. Borgen, B.R. Kowalski, J. Chemometrics 7 (1993) 439.

[15] A.K. Smilde, R. Tauler, J. Saurina, R. Bro, Anal. Chim. Acta 398 (1999) 237.

[16] H.A.L. Kiers, A.K. Smilde, J. Chemometrics 12 (1998) 125.

[17] D.D. Perrin, B. Demsey, Buffers for pH and Metal Ion Control, Chapman & Hall, London, 1974.

[18] E.P. Serjeant, B. Demsey, Ionisation Constants of Organic Acids in Aqueous Solution, Pergamon Press, Oxford, 1979.

[19] N.M. Faber, L.M.C. Buydens, G. Kateman, Chemometrics Intell. Lab. Syst. 25 (1994) 203.

[20] B. Efron, R.J. Tibshirani, An Introduction to the Bootstrap, CRC Press, London, 1994.

[21] J.R. Magnus, H. Neudecker, Matrix Differential Calculus with Applications in Statistics and Econometrics, Wiley, Chichester, 1988, p. 361.

[22] R. Tauler, Chemometrics Intell. Lab. Syst. 30 (1995) 133.