ELSEVIER

Analytica Chimica Acta 398 (1999) 237–251

ANALYTICA CHIMICA ACTA

www.elsevier.com/locate/aca

# Calibration methods for complex second-order data

Age K. Smilde [a,*], Roma Tauler [b], Javier Saurina [b], Rasmus Bro [c]

[a] *Laboratory for Analytical Chemistry, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands*
[b] *Department of Analytical Chemistry, University of Barcelona, Diagonal 647, Barcelona 08028, Spain*
[c] *Chemometrics Group, Food Technology, The Royal Vet. & Agri. University, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark*

Received 11 September 1998; received in revised form 15 April 1999; accepted 26 April 1999

**Abstract**

In this paper, different three-way methods are tested for their power and shortcomings to solve complex second-order calibration problems. The generic calibration problem is quantifying for an analyte in the presence of an unknown interferent: a second-order calibration problem. Due to rank restrictions of the data, standard second-order calibration methods like Generalized Rank Annihilation cannot be used to solve the type of complex second-order calibration problems shown in this paper. Different real examples are tested in which it is shown that the three-way methods can, to a certain extent, deal with the complex calibrations. This stresses the fact that all second-order calibration methods should be regarded as three-way methods, and when put in this framework, can be compared with respect to their performance. 

*Keywords:* Flow injection analysis; Rank deficiency; Multivariate curve resolution; PARATUCK2; Restricted Tucker3

## 1. Introduction

Calibration is an important topic in analytical chemistry. The purpose of calibration is to determine the amount of analyte(s) of interest in unknown samples which might contain (unknown) interferents. Moreover, as a side goal, it is sometimes convenient to obtain confirmative information about the identity of the analytes and the interferents, if present.

The area of calibration can be divided in zero, first and second-order calibration, etc. A nice summary is given by Booksh and Kowalski [1]. Zero-order calibration is used with zero-order data. If measurements on standards and unknown samples give a single instrumental read-out per standard or unknown sample measurement, the resulting data is called zero-order data. Of course, a standard can be measured several times, but the mean value is used in the calibration, and in essence, it is a single point measurement. The term 'zero-order' is borrowed from tensor algebra [2]: a number is a zero-order tensor. The assumption in zero-order calibration is that only the analyte(s) of interest contribute to the measured signal. This is often a severe limitation in practice.

First-order calibration is used for first-order data. An example of first-order data is an NIR-spectrum taken from a standard or a sample. Essentially, a vector of measurements is generated per standard and unknown sample. This vector (a first-order tensor) is then used in the calibration. This is known in the analytical chemistry and chemometric literature as multivariate

* Corresponding author. Tel.: +31-20-5255062; fax: +31-20-5256638
*E-mail addresses:* asmilde@anal.chem.uva.nl (A.K. Smilde), roma@quimio.qui.ub.es (R. Tauler), rb@kvl.dk (R. Bro)

0003-2670/99/$ – see front matter 
PII: S0003-2670(99)00374-8

calibration [3,4]. Considerable effort has been applied on the development of multivariate calibration tools [5–7]. An advantage of first-order calibration is that it is possible to quantify the amounts of analytes of interest in unknown samples when these samples contain interferents. The only requirement is that these interferents are also contained in the standards used to calibrate the system.

Second-order calibration is used for second-order data. Such data is produced by instruments that give a matrix of responses for a single measured standard or unknown sample. Examples of such instruments are: hyphenated instruments [8] (GC-IR, LC-UV, GC-MS, MS-MS, etc.); fluorescence emission/excitation instruments [9]; and second-order chemical sensors [10,11]. Clearly, the number of second-order instruments is growing and, therefore, the need for good second-order calibration tools is obvious.

Several methods for second-order calibration exist nowadays. In the early days, Rank Annihilation Factor Analysis (RAFA) was used [12]; this was later generalized to Generalized Rank Annihilation (GRAM) [13]. Both methods work well under certain conditions of the second-order data: the pseudo-rank of a pure analyte response is 1 and this pure analyte response has the same form in the standard as in the unknown mixture.

Pseudo-rank is defined as the rank of a measurement if this measurement contains no experimental error. Since all measurements are noisy, the pseudo-rank has to be estimated and this estimate can be used to check the condition for applying RAFA and/or GRAM. Note that the theory of instrumentation allows in some cases an a priori assessment of the pseudo-rank of a measurement, e.g., LC–UV usually gives a pseudo-rank 1 response per analyte due to the specific properties of the LC and UV instruments. The condition of the response of the pure analyte having the same form in the standard as in the unknown mixture implies that trilinear models can be used to model the joint variation in the standard and the mixture. This is exactly what GRAM does.

Under the above-mentioned conditions of second-order data, second-order calibration using GRAM has several distinct advantages. The most pronounced advantage is that it makes possible the quantification of the analyte(s) of interest in samples containing interferents. Those interferents can be completely unknown, that is, it is not necessary that they have been incorporated in the standards used for calibration. This is called the second-order advantage. Another advantage is that it is possible to recover the individual instrument responses of the analyte(s) of interest, e.g., the LC-chromatograms and UV-spectra can be obtained for the individual analyte(s). This makes it possible to perform qualitative analysis and to check the identity of the analytes. Hence, second-order calibration has also a resolution aspect: resolving individual instrumental profiles. GRAM and RAFA also have a serious drawback: they can only be used in cases of a single standard and a single unknown mixture.

Second-order calibration in situations where the pseudo-rank 1 restriction does not hold is considerably more complicated. The direct generalization of GRAM for situations with pure analyte responses of pseudo-rank higher than 1 is Nonbilinear Rank Annihilation (NBRA) [14]. Quantification is still possible, but resolving individual profiles is not possible anymore. It becomes even more complicated when the property of rank linear additivity [15] does not hold. Rank linear additivity means that if analyte 1 gives a rank $r_1$ response and analyte two a rank $r_2$ response, then the mixture of the two analytes gives a rank $r_1 + r_2$ response. This is not always the case as will be shown below in an example, and also analytes and interferents can fail to have the rank linear additivity property. Second-order calibration using GRAM or NBRA breaks down if the rank linear additivity property does not hold anymore. A mathematical treatment of the properties of GRAM, NBRA and the relation to rank linear additivity is given by Kiers and Smilde [16].

Several methods have been described for complicated second-order calibration, that is, cases where the rank 1 property and perhaps even rank linear additivity do not hold. One of these methods is multivariate curve resolutions (MCR) with restrictions [10,17,18]. Another method uses restricted Tucker3 (RT3) models to calibrate the complicated second-order system [11,19]. Both methods have been applied to a chemical sensor [10,11] with reasonable success. The purpose of the present paper is to use several examples varying in complexity to show in a systematic way the behavior of MCR and RT3 and that of a completely new method based on PARATUCK2 (PT2) models.

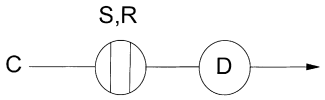


Fig. 1. The experimental set-up: C is the carrier stream (with pH 4.5); S is the injected sample; R is the injected reagent stream (with pH 11.4); D is the detector.

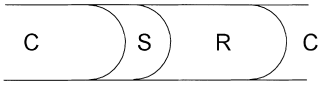


Fig. 2. The flow through the flow-injection channel indicating the sample subjected to the pH gradient (for a legend of C, S and R, see Fig. 1).

This paper is organized as follows. The examples are all taken from the same flow injection analysis measurements. The system and the examples taken from this system will be described. The theory of the three different second-order calibration methods (MCR, RT3, PT2) will be explained and outlined. Then, the results will be presented.

## 2. Experimental

### *2.1. Flow injection analysis system*

The flow injection analysis (FIA) system used in this investigation [20] 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) between the carrier and the reagent stream, as shown in Fig. 2. 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.

The sample is led into an 8 μl flow cell and measured on a HP 8452A photodiode array spectrophotometer. The sample is measured for 88 s with 1 s intervals 20 s after injection from 250 to 450 nm with 2 nm intervals. The second-order data obtained from each sample is thus of size $89 \times 101$. Since spikes were detected in one of the channels of the photodiode-array, the readings of this channel were removed from the data. Hence, the actual data set for each sample is $89 \times 100$.

The duration of detection and the pH gradient are sufficient to ensure that the analytes are both in their acidic and basic forms 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 dihydrogenphosphate, boric acid, and tri-(hydroxymethyl)aminomethane (TRIS) according to Perrin and Dempsey [21]. 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 with sodium hydroxide.

The test solutes 2-hydroxybenzaldehyde (2-HBA), 3-hydroxybenzaldehyde (3-HBA) and 4-hydroxybenzaldehyde (4-HBA) show different absorption spectra, depending on whether they are in their acidic or basic form. 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 thus the same as for the sample as such, but due to the pH-gradient, the first part of the sample plug is dominated by deprotonated solutes, while the end of the sample plug is dominated by protonated solutes. Depending on the p$K_a$ of a given solute, it will show up with different acidic and basic profiles in the sample plug. The p$K_a$ values of the three solutes 2-HBA, 3-HBA and 4-HBA are 8.37, 8.98 and 7.61, respectively [20]. Fig. 3 shows the raw data for 2-HBA. This figure also explains graphically what the structure of the data matrix is.

### *2.2. Description of the data*

Three-way arrays are written as boldface underlined uppercase characters, e.g., $\underline{\mathbf{A}}$, $\underline{\mathbf{B}}$. Matrices are written as boldface uppercase characters, e.g., **A**, **B**. Vectors are written as bold lowercase characters, e.g., **a**, **b**. Vectors are always column vectors. Scalars are written as lowercase characters, e.g., a, b.

The spectra of the acidic forms of 2-HBA, 3-HBA and 4-HBA are called, respectively, $\mathbf{sa}_2$, $\mathbf{sa}_3$ and $\mathbf{sa}_4$;

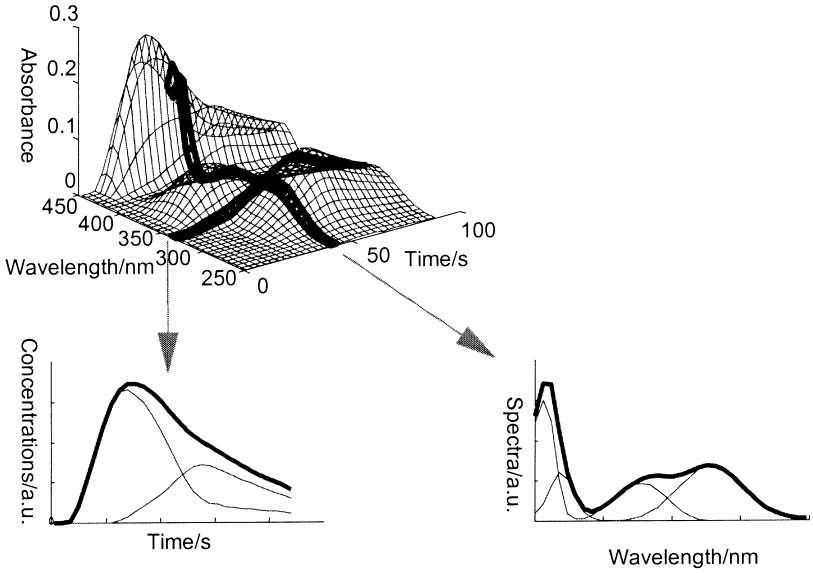


Fig. 3. Landscape obtained from a sample containing only 2-HBA (top). The measured profile at 340 nm is shown below at the left (thick line). This profile is the sum of the profile of the (unknown) acidic and basic profiles (thinner lines). Below, at the right the measured spectrum at time 43 s is shown (thick line). This spectrum is the sum of the (unknown) acidic and basic spectra (thin lines).

and the spectra of the basic forms $\mathbf{sb}_2$, $\mathbf{sb}_3$ and $\mathbf{sb}_4$, respectively. The subscripts 2, 3 or 4 refer to 2, 3 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 called, respectively, $\mathbf{ca}_2$, $\mathbf{ca}_3$ and $\mathbf{ca}_4$; and of the basic forms $\mathbf{cb}_2$, $\mathbf{cb}_3$ and $\mathbf{cb}_4$, respectively. The concentration profiles and spectra belong to unit concentrations of the solutes unless stated otherwise.

The response of the pure solutes 2-HBA, 3-HBA and 4-HBA will be called $\mathbf{N}_{2\text{-HBA}}$, $\mathbf{N}_{3\text{-HBA}}$ and $\mathbf{N}_{4\text{-HBA}}$, respectively. These matrices have the spectra measured at different points in time in their rows. Their order is, therefore, $89 \times 100$, being the number of time points versus 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

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

where $\mathbf{E}_{2\text{-HBA}}$ stands for the experimental error in $\mathbf{N}_{2\text{-HBA}}$. Similar formulas hold for 3-HBA and 4-HBA.

Using local rank analysis techniques (e.g., fixed size moving window evolving factor analysis [22]), it can be concluded that the rank of the matrices consisting of the first 10–20 rows and all the columns of the response matrices of the pure solutes is of rank 1. This means that there is only one UV-absorbing species in the beginning of the response matrices. This is reasonable, given the set-up of the experiment: in the beginning, only the basic species is absorbing. Likewise, local rank analysis shows that the last 10–20 rows also of the response matrices make up a rank 1 matrix. This means that using these local rank 1 windows, reasonable estimates of the pure acidic and basic spectra of all the solutes can be obtained. A non-negative least squares step using these estimated spectra and Eq. (1) is then sufficient to calculate the concentration profile estimates.

In order to check the obtained estimates of the spectra and concentration profiles, an auxiliary measurement was performed for each of the solutes. UV-spectra of the pure solutes were taken at different pH-values, corresponding to the acidic and basic conditions in the above-described FIA experiments. The agreement between the estimates obtained with local rank analysis and the measured ones were excellent. From now on, these measured spectra and estimated concentration profiles will be referred to as 'true' spectra and profiles, respectively.

Closely examining the spectra and concentration profiles, as obtained from the auxiliary experiment, it appears that the pure spectra and concentration profiles of 2-HBA and 3-HBA are very similar (see also Nørgaard and Ridder [20]). Hence, calibrations including both these solutes will be difficult.

Assessment of the linearity of the responses was performed. The FIA system showed good linearity. The reproducibility of repeated measurements was tested and resulted in a relative error of approximately 5%, which includes instrumental error and sampling error.

### 2.3. Restriction on the data

During an analysis, the total concentration profile of a solute (acidic and basic form) is $\mathbf{ca} + \mathbf{cb}$. 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.

The shape of the total concentration profile is defined by the diffusion properties of the solutes. Since the solutes resemble each other very much, it can be expected that the diffusion behavior is equal for all three. Hence, the shape of the total concentration profiles is equal: $\mathbf{ctot}_2 = \alpha \cdot \mathbf{ctot}_3 = \beta \cdot \mathbf{ctot}_4$, where $\alpha$ and $\beta$ are constants. This phenomenon puts a restriction on the calibration problem and destroys the rank linear additivity of the system. To show this, consider two analytes 2-HBA and 3-HBA and a mixture of the two (where, for convenience, the terms representing errors are dropped):

$$\mathbf{N}_{2\text{-HBA}} = \mathbf{ca}_2 \cdot \mathbf{sa}_2^{\mathrm{T}} + \mathbf{cb}_2 \cdot \mathbf{sb}_2^{\mathrm{T}}$$

$$\mathbf{N}_{3\text{-HBA}} = \mathbf{ca}_3 \cdot \mathbf{sa}_3^{\mathrm{T}} + \mathbf{cb}_3 \cdot \mathbf{sb}_3^{\mathrm{T}}$$

$$\begin{aligned}\mathbf{M} &= \mathbf{ca}_2 \cdot \mathbf{sa}_2^{\mathrm{T}} + \mathbf{cb}_2 \cdot \mathbf{sb}_2^{\mathrm{T}} + \mathbf{ca}_3 \cdot \mathbf{sa}_3^{\mathrm{T}} + \mathbf{cb}_3 \cdot \mathbf{sb}_3^{\mathrm{T}} \\ &= (\alpha \cdot \mathbf{ctot}_3 - \mathbf{cb}_2)\,\mathbf{sa}_2^{\mathrm{T}} + \mathbf{cb}_2 \cdot \mathbf{sb}_2^{\mathrm{T}} + (\mathbf{ctot}_3 - \mathbf{cb}_3) \\ &\quad \times \mathbf{sa}_3^{\mathrm{T}} + \mathbf{cb}_3 \cdot \mathbf{sb}_3^{\mathrm{T}} = \mathbf{ctot}_3 \left(\alpha \cdot \mathbf{sa}_2^{\mathrm{T}} + \mathbf{sa}_3^{\mathrm{T}}\right) \\ &\quad + \mathbf{cb}_2 \left(-\mathbf{sa}_2^{\mathrm{T}} + \mathbf{sb}_2^{\mathrm{T}}\right) + \mathbf{cb}_3 \left(-\mathbf{sa}_3^{\mathrm{T}} + \mathbf{sb}_3^{\mathrm{T}}\right) \qquad (2)\end{aligned}$$

The last line of Eq. (2) shows that the rank of **M** is 3 and not 4. Hence, rank linear additivity does not hold.

The restriction on the data has another consequence which is related to the observation on rank linear additivity given above. Defining the following matrices:

$$\mathbf{A} = [\,\mathbf{ca}_2 \quad \mathbf{cb}_2 \quad \mathbf{ca}_3 \quad \mathbf{cb}_3\,] \qquad \text{and}$$

$$\mathbf{B} = [\,\mathbf{sa}_2 \quad \mathbf{sb}_2 \quad \mathbf{sa}_3 \quad \mathbf{sb}_3\,] \qquad (3)$$

then, obviously, **A** is rank deficient and **B** has full rank. Hence, all standard second-order calibration methods (e.g., GRAM [13]) cannot be used immediately [16]. This rank deficiency problem is solved by the methods proposed in this paper in different ways.

Summarizing, the FIA system represents a really difficult second-order system: it is not a rank 1 system and rank linear additivity does not hold. The FIA system is, therefore, an ideal system for testing several calibration methods for complex second-order data.

## 3. Theory

### 3.1. Restricted Tucker3 models (RT3 models)

The RT3 models will be explained with the problem of quantifying 2-HBA in the presence of an unknown interferent. Mathematically, this problem comes down to

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

$$\begin{aligned}\mathbf{M} &= \gamma \cdot \mathbf{ca}_2 \cdot \mathbf{sa}_2^{\mathrm{T}} + \gamma \cdot \mathbf{cb}_2 \cdot \mathbf{sb}_2^{\mathrm{T}} \\ &\quad + \mathbf{ca}_{\mathrm{u}} \cdot \mathbf{sa}_{\mathrm{u}}^{\mathrm{T}} + \mathbf{cb}_{\mathrm{u}} \cdot \mathbf{sb}_{\mathrm{u}}^{\mathrm{T}} + \mathbf{E}_M \qquad (4)\end{aligned}$$

where the **E** matrices refer to the measurement and model error always present in real measurements. The purpose is to estimate $\gamma$ and to obtain as much information of the unknown interferent as possible, i.e. to obtain $\mathbf{ca}_{\mathrm{u}}$, $\mathbf{sa}_{\mathrm{u}}$, $\mathbf{cb}_{\mathrm{u}}$ and $\mathbf{sb}_{\mathrm{u}}$ (the subscript 'u' stands for 'unknown interferent'). The concentration of the interferent in **M** is absorbed in $\mathbf{ca}_{\mathrm{u}}$ and $\mathbf{cb}_{\mathrm{u}}$, without loss of generality.

Due to the above mentioned restriction on the total concentration profiles, it can be assumed that

$$\mathbf{ca}_2 + \mathbf{cb}_2 = \mathbf{ctot} \qquad \text{and} \qquad \mathbf{ca}_{\mathrm{u}} + \mathbf{cb}_{\mathrm{u}} = \alpha\mathbf{ctot} \quad (5)$$

where $\alpha$ is an arbitrary constant. There are several ways to rewrite Eq. (4) according to Eq. (5). One way of doing this is by writing

$$\mathbf{cb}_2 = \mathbf{ctot} - \mathbf{ca}_2 \qquad \text{and} \qquad \mathbf{ca}_{\mathrm{u}} = \alpha\mathbf{ctot} - \mathbf{cb}_{\mathrm{u}} \quad (6)$$

and in that case, $\mathbf{cb}_2$and $\mathbf{ca}_{\mathrm{u}}$ can be eliminated from the model, resulting in

$$\mathbf{N}_{2\text{-HBA}} = \mathbf{ca}_2 \cdot \mathbf{sa}_2^{\mathrm{T}} + \mathbf{ctot} \cdot \mathbf{sb}_2^{\mathrm{T}} - \mathbf{ca}_2 \cdot \mathbf{sb}_2^{\mathrm{T}} + \mathbf{E}_{2\text{-HBA}}$$

$$\mathbf{M} = \gamma \cdot \mathbf{ca}_2 \cdot \mathbf{sa}_2^{\mathrm{T}} + \gamma \cdot \mathbf{ctot} \cdot \mathbf{sb}_2^{\mathrm{T}} - \gamma \cdot \mathbf{ca}_2 \cdot \mathbf{sb}_2^{\mathrm{T}} + \alpha \cdot \mathbf{ctot} \cdot \mathbf{sa}_{\mathrm{u}}^{\mathrm{T}} - \mathbf{cb}_{\mathrm{u}} \cdot \mathbf{sa}_{\mathrm{u}}^{\mathrm{T}} + \mathbf{cb}_{\mathrm{u}} \cdot \mathbf{sb}_{\mathrm{u}}^{\mathrm{T}} + \mathbf{E}_M \quad (7)$$

and upon collecting the proper vectors, the following matrices can be defined:

$$\mathbf{A}_{\mathrm{RT3}} = \begin{bmatrix} \mathbf{ca}_2 & \mathbf{ctot} & \mathbf{cb}_{\mathrm{u}} \end{bmatrix}$$

$$\mathbf{B}_{\mathrm{RT3}} = \begin{bmatrix} \mathbf{sa}_2 & \mathbf{sb}_2 & \mathbf{sa}_{\mathrm{u}} & \mathbf{sb}_{\mathrm{u}} \end{bmatrix}$$

$$\mathbf{C}_{\mathrm{RT3}} = \begin{bmatrix} 1 & 0 \\ \gamma & 1 \end{bmatrix} \quad (8)$$

where the choice of Eq. (6) generates a matrix with the acidic profile of the analyte and the basic profile of the unknown interferent.

The two matrices $\mathbf{N}_{2\text{-HBA}}$ and $\mathbf{M}$ can be stacked on top of each other to form the three-way matrix $\underline{\mathbf{X}}$. Using Eqs. (7) and (8), this three-way matrix $\underline{\mathbf{X}}$ can be modelled with a Restricted Tucker model after finding the proper core-array. Closely examining Eqs. (7) and (8) gives the structure of the unfolded core-array as

$$\mathbf{G}_{\mathrm{RT3}} = \left[\begin{array}{cccc|cccc} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & \alpha & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{array}\right] \quad (9)$$

where each block of $3 \times 4$ refers to the two column vectors in $\mathbf{C}_{\mathrm{RT3}}$; each row refers to a column vector in $\mathbf{A}_{\mathrm{RT3}}$; each column within a block refers to a column vector in $\mathbf{B}_{\mathrm{RT3}}$ and the zeroes are indicating whether a core-array element is forced to be zero. For example, the (1,1)th element in $\mathbf{G}_{\mathrm{RT3}}$ describes that the first column vector in $\mathbf{A}_{\mathrm{RT3}}$ ($\mathbf{ca}_2$), the first column vector in $\mathbf{B}_{\mathrm{RT3}}$ ($\mathbf{sa}_2$) and the first column vector in $\mathbf{C}_{\mathrm{RT3}}$ have to be combined, as can be checked in formula (7). Now, using the matrices $\mathbf{A}_{\mathrm{RT3}}$, $\mathbf{B}_{\mathrm{RT3}}$, $\mathbf{C}_{\mathrm{RT3}}$ and $\mathbf{G}_{\mathrm{RT3}}$ of Eqs. (8) and (9), it can be shown [19,23] that the following holds:

$$[\mathbf{N}_{2\text{-HBA}} | \mathbf{M}] = \mathbf{A}_{\mathrm{RT3}} \mathbf{G}_{\mathrm{RT3}} \left( \mathbf{C}_{\mathrm{RT3}}^{\mathrm{T}} \otimes \mathbf{B}_{\mathrm{RT3}}^{\mathrm{T}} \right) + [\mathbf{E}_{2\text{-HBA}} | \mathbf{E}] \quad (10)$$

where the symbol $\otimes$ is used to indicate the Kronecker matrix product.

Eq. (10) gives a combined model of the measured data: standard and unknown sample. The unknown parameters in the model are $\mathbf{A}_{\mathrm{RT3}}$, $\mathbf{B}_{\mathrm{RT3}}$ and the scalars $\alpha$ and $\gamma$ in $\mathbf{G}_{\mathrm{RT3}}$ and $\mathbf{C}_{\mathrm{RT3}}$, respectively. These parameters have to be estimated. An important question is whether these parameters can be determined uniquely. This is a very difficult issue, but it can be proved [23] that for the model described in Eq. (10), uniqueness is obtained for $\mathbf{C}_{\mathrm{RT3}}$ and **ctot**. This is an important result since a unique $\mathbf{C}_{\mathrm{RT3}}$ provides that the result obtained in $\mathbf{C}_{\mathrm{RT3}}$ can be used for quantification. Unfortunately, Eq. (10) does not give unique estimates for the spectra and concentration profiles of the unknown interferent. However, the subspace spanned by the spectra of the unknown interferent can be estimated uniquely [23]. This gives room for curve resolution techniques to resolve those spectra, but that is not pursued here.

The model parameters are estimated by means of a constrained alternating least squares (ALS) algorithm, where non-negativity constraints are put on $\mathbf{A}_{\mathrm{RT3}}$, $\mathbf{B}_{\mathrm{RT3}}$, $\mathbf{C}_{\mathrm{RT3}}$ using the non-negative least squares algorithm of Lawson and Hanson [24]; applying the constraints on $\mathbf{G}_{\mathrm{RT3}}$ and using the fact that part of $\mathbf{C}_{\mathrm{RT3}}$ is known. Starting values are provided to the algorithm and the error-sum-of-squares (the summed norms of the $\mathbf{E}$ matrices) is iteratively minimized.

Note that $\mathbf{A}_{\mathrm{RT3}}$ has full rank now. The rank deficiency as mentioned in the previous section is solved by explicitly removing the linear dependency between the columns of $\mathbf{A}$ using the restriction.

Since ALS is an iterative procedure it has to be provided with starting values. After convergence of the algorithm, the estimates of $\mathbf{A}_{\mathrm{RT3}}$, $\mathbf{B}_{\mathrm{RT3}}$, $\mathbf{C}_{\mathrm{RT3}}$ and $\mathbf{G}_{\mathrm{RT3}}$ are obtained. For quantification, the estimated $\mathbf{C}_{\mathrm{RT3}}$ is important since it contains the estimated concentrations. The estimated concentration profiles and spectra are obtained in $\mathbf{A}_{\mathrm{RT3}}$ and $\mathbf{B}_{\mathrm{RT3}}$, respectively.

### 3.2. Multivariate curve resolution (MCR)

The MCR method is explained using the same example as above for the RT3 models; detailed explanations are given elsewhere [17,18]. The starting equations are the same as above, i.e., Eq. (4) defining the calibration problem.

The concentration profiles $\mathbf{ca}_2$, $\mathbf{cb}_2$, $\mathbf{ca}_{\mathrm{u}}$ and $\mathbf{cb}_{\mathrm{u}}$ are collected in two matrices $\mathbf{A}_{2\text{-HBA}}$ and $\mathbf{A}_M$, which are defined as follows:

$$\mathbf{A}_{2\text{-HBA}} = \begin{bmatrix} \mathbf{ca}_2 & \mathbf{cb}_2 & 0 & 0 \end{bmatrix}$$

$$\mathbf{A}_M = \begin{bmatrix} \gamma \cdot \mathbf{ca}_2 & \gamma \cdot \mathbf{cb}_2 & \mathbf{ca}_{\mathrm{u}} & \mathbf{cb}_{\mathrm{u}} \end{bmatrix} \quad (11)$$

where the '**0**' in $\mathbf{A}_{2\text{-HBA}}$ stands for a vector of zeroes and the concentration of the interferent is again absorbed in $\mathbf{ca}_u$ and $\mathbf{cb}_u$, as before. Likewise, the spectra are collected in $\mathbf{B}_{MCR}$:

$$\mathbf{B}_{MCR} = \begin{bmatrix} \mathbf{sa}_2 & \mathbf{sb}_2 & \mathbf{sa}_u & \mathbf{sb}_u \end{bmatrix} \tag{12}$$

Note that $\mathbf{B}_{MCR} = \mathbf{B}_{RT3}$. The calibration problem of Eq. (4) can now be written as

$$\begin{bmatrix} \mathbf{N}_{2\text{-HBA}} \\ \mathbf{M} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{2\text{-HBA}} \\ \mathbf{A}_M \end{bmatrix} \mathbf{B}_{MCR}^T + \mathbf{E} \tag{13}$$

and, upon defining

$$\mathbf{X} = \begin{bmatrix} \mathbf{N}_{2\text{-HBA}} \\ \mathbf{M} \end{bmatrix} \quad \text{and} \quad \mathbf{A}_{MCR} \begin{bmatrix} \mathbf{A}_{2\text{-HBA}} \\ \mathbf{A}_M \end{bmatrix},$$

estimation proceeds by alternating between

$$\min_{\hat{\mathbf{B}}_{MCR}} \left\| \mathbf{X} - \hat{\mathbf{A}}_{MCR} \cdot \hat{\mathbf{B}}_{MCR}^T \right\|^2 \tag{14}$$

and

$$\min_{bf(A)_{MCR}} \left\| \mathbf{X} - \hat{\mathbf{A}}_{MCR} \cdot \hat{\mathbf{B}}_{MCR}^T \right\|^2 \tag{15}$$

where the notation $\hat{\mathbf{B}}_{MCR}$ means that an estimate of $\hat{\mathbf{B}}_{MCR}$ is calculated and the shapes of the estimated analyte concentration profiles are forced to be the same in $\hat{\mathbf{A}}_{2\text{-HBA}}$ and $\hat{\mathbf{A}}_M$ using the trilinearity assumption (see Section 1). The problems (14) and (15) are solved under the restrictions of non-negativity for $\hat{\mathbf{A}}_{MCR}$, $\hat{\mathbf{B}}_{MCR}$ (using Lawson and Hansons's algorithm or by forcing negative values to zero) and unimodality for the concentration profiles ($\hat{\mathbf{A}}_{MCR}$). After each pair of steps (14) and (15), the estimated $\mathbf{X} = \hat{\mathbf{A}}_{MCR} \cdot \hat{\mathbf{B}}_{MCR}^T$ is compared with the measured $\mathbf{X}$. If the difference between $\mathbf{X}$ and $\hat{\mathbf{X}}$ (the norm of the error matrix $\mathbf{E}$) is small enough, then the algorithm is converged. The restriction of equal total concentration profile shapes (Eq. (5)) is not built in explicitly.

In Appendix A, it is shown that $\mathbf{A}_{MCR}$ has full column rank under mild conditions making regression problem (14) well defined. Hence, the problem of rank deficiency is solved by augmenting the rank deficient matrix $\mathbf{A}$ (=$\mathbf{A}_M$) with $\mathbf{A}_{2\text{-HBA}}$. This is similar to the procedure proposed by Amrhein et al. [25] for rank deficient factor analysis of process data.

After convergence of the algorithm, the final estimates of $\mathbf{A}_{MCR}$ and $\mathbf{B}$ contain the quantitative and the qualitative information. With the algorithm, the spectra and concentration profiles of the analyte 2-HBA can be recovered. The spectra and concentration profiles of the interferent can be estimated, but they are not uniquely determined (see Appendix A). Of course, the concentration of the analyte in $\mathbf{M}$ (the parameter $\gamma$) can also be found.

### 3.3. PARATUCK2 (PT2) models

The calibration problem of Eq. (4) will also be used to explain the PT2 approach. Concentrating first on a model of the mixture $\mathbf{M}$, by defining $\mathbf{m}_i^T$ as the $i$th row of $\mathbf{M}$, it holds that

$$\mathbf{m}_i^T = \gamma \cdot \mathrm{ca}_{2,i} \cdot \mathbf{sa}_2^T + \gamma \cdot \mathrm{cb}_{2,i} \cdot \mathbf{sb}_2^T + \mathrm{ca}_{u,i} \cdot \mathbf{sa}_u^T + \mathrm{cb}_{u,i} \mathbf{sa}_u^T \tag{16}$$

where the subscript '$i$' of $\mathbf{ca}_2$, $\mathbf{cb}_2$, $\mathbf{ca}_u$ and $\mathbf{cb}_u$ refers to the $i$th elements of those vectors, and, again, the concentration of the interferent is absorbed in $\mathbf{ca}_u$ and $\mathbf{cb}_u$. Defining the following vectors and matrices

$$\mathbf{c}_M^T = \begin{bmatrix} \gamma & 1 \end{bmatrix}$$

$$\mathbf{G}_{PT2} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

$$\mathbf{B}_{PT2} = \begin{bmatrix} \mathbf{sa}_2 & \mathbf{sb}_2 & \mathbf{sa}_u & \mathbf{sb}_u \end{bmatrix}$$

$$\mathbf{A}_{PT2,i} = \begin{bmatrix} \mathrm{ca}_{2,i} & 0 & 0 & 0 \\ 0 & \mathrm{cb}_{2,i} & 0 & 0 \\ 0 & 0 & \mathrm{ca}_{u,i} & 0 \\ 0 & 0 & 0 & \mathrm{cb}_{u,i} \end{bmatrix} \tag{17}$$

then

$$\mathbf{m}_i^T = \mathbf{c}_M^T \mathbf{G}_{PT2} \mathbf{A}_{PT2,i} \mathbf{B}_{PT2}^T \tag{18}$$

which can easily be checked by using matrix multiplication rules. In a similar fashion, the $i$th row of $\mathbf{N}_{2\text{-HBA}}$ ($\mathbf{n}_i^T$) can be written as

$$\mathbf{n}_i^T = \mathbf{c}_N^T \mathbf{G}_{PT2} \mathbf{A}_{PT2,i} \mathbf{B}_{PT2}^F \quad \text{with} \quad \mathbf{c}_N^T = \begin{bmatrix} 1 & 0 \end{bmatrix} \tag{19}$$

and hence,

$$\mathbf{X}_i = \begin{bmatrix} \mathbf{n}_i^T \\ \mathbf{m}_i^T \end{bmatrix} = \begin{bmatrix} \mathbf{c}_N^T \\ \mathbf{c}_M^T \end{bmatrix} \mathbf{G}_{PT2} \mathbf{A}_{PT2,i} \mathbf{B}_{PT2}^T \quad i = 1, \ldots, I = 89 \tag{20}$$

The general formulation of the PT2 model is [26]

$$\mathbf{X}_i = \mathbf{U}^{\mathrm{L}}\mathbf{D}_i\mathbf{H}^{\mathrm{R}}\mathbf{D}_i\mathbf{V}^{\mathrm{T}} \quad \text{with} \quad {}^{\mathrm{L}}\mathbf{D}_i \text{ and } {}^{\mathrm{R}}\mathbf{D}_i \text{ being diagonal;} \quad i = 1, ..., I \tag{21}$$

and by defining $\mathbf{U} = \begin{bmatrix} \mathbf{c}_N^T \\ \mathbf{c}_M^T \end{bmatrix} = \mathbf{C}_{PT2}$, ${}^{L}\mathbf{D}_i = \mathbf{I}$, $\mathbf{H} = \mathbf{G}_{\mathrm{PT2}}$, ${}^{R}\mathbf{D}_i = \mathbf{A}_{\mathrm{PT2},i}$ and $\mathbf{V} = \mathbf{B}_{\mathrm{PT2}}$, Eq. (20) is a special case of Eq. (21). Hence, the calibration problem (4) can be written as a special case of the PT2 model.

Estimation of the PT2 model is performed with an ALS scheme: fix ${}^{\mathrm{L}}\mathbf{D}_i$, $\mathbf{H}$, ${}^{\mathrm{R}}\mathbf{D}_i$, $\mathbf{V}$ and minimize $\sum_i \left\|\mathbf{X}_i - \mathbf{U}^{\mathrm{L}}\mathbf{D}_i\mathbf{H}^{\mathrm{R}}\mathbf{D}_i\mathbf{V}^{\mathrm{T}}\right\|^2$ by estimating $\mathbf{U}$; then fix ${}^{\mathrm{L}}\mathbf{D}_i$, $\mathbf{H}$, ${}^{\mathrm{R}}\mathbf{D}_i$, $\mathbf{U}$ and estimate $\mathbf{V}$ by solving the corresponding least squares problem, etc. This process is continued until convergence is achieved, in the same way as for the MCR and RT3 methods. In this ALS scheme, restrictions are built in. First, non-negativity is enforced in the estimation of $\mathbf{C}_{\mathrm{PT2}} = [\mathbf{c}_N\mathbf{c}_M]^{\mathrm{T}}$, $\mathbf{B}_{\mathrm{PT2}}$ and $\mathbf{A}_{\mathrm{PT2},i}$ by using Lawson and Hanson's algorithm. Secondly, since $\mathbf{G}_{\mathrm{PT2}}$, ${}^{\mathrm{L}}\mathbf{D}_i$ and part of $\mathbf{C}_{\mathrm{PT2}}$ are known, these parameters do not have to be estimated. Thirdly, since the concentration profiles should be unimodal, a unimodality constraint was built in for the concentration profiles.

The restriction of equal shapes of concentration profiles (Eq. (5)) is built in using a penalty approach. To illustrate this, suppose the least squares problem is

$$\min_b \|\mathbf{y} - \mathbf{Z}\mathbf{b}\|^2 \quad \text{subject to} \quad \mathbf{t} = \mathbf{X}\mathbf{b} \tag{22}$$

where $\mathbf{y}$, $\mathbf{Z}$, and $\mathbf{t}$, $\mathbf{X}$ are given. A solution of this problem is to reformulate it as

$$\min_b \left[ \left\| \begin{matrix} \lambda\,(\mathbf{t} - \mathbf{X}\mathbf{b}) \\ \mathbf{y} - \mathbf{Z}\mathbf{b} \end{matrix} \right\|^2 \right] = \min_b \left\| \begin{bmatrix} \mathbf{t}\lambda \\ \mathbf{y} \end{bmatrix} - \begin{bmatrix} \mathbf{X}\lambda \\ \mathbf{Z} \end{bmatrix} b \right\|^2 \quad \text{given a certain } \lambda \tag{23}$$

where the tuning parameter $\lambda$ is used to penalize on the restriction: if $\lambda$ is small, then the restriction of $\mathbf{t} = \mathbf{X}\mathbf{b}$ is 'softly' imposed, whereas for a large $\lambda$, this restriction is imposed severely. This penalty procedure is used to impose the constraint of Eq. (5) on the estimation of $\mathbf{A}_{\mathrm{PT2},i}$ ($i = 1,...,89$) in the PT2 model. A high value of $\lambda$ is chosen, thereby imposing the equality restriction in a hard way.

In Appendix A, it is shown that the PT2 model equals the RT3 model if, in both models, the equality constraint is imposed. Hence, the PT2 model too solves the rank deficiency by explicitly accounting for this constraint.

### *3.4. Similarities and differences between the methods*

Assume that the data are ideal: no measurement error, Lambert-Beer's law holds and the equality of the total concentration profiles holds exactly. If the equality of the total concentration profiles is imposed on the PT2 and RT3 models, then these models can be shown to be mathematically identical. This is proved in Appendix A. Moreover, if the equality as formulated in Eq. (5) is present in the data, then, MCR is mathematically equivalent to RT3; this is also shown in Appendix A.

Summarizing, all three methods (RT3, MCR and PT2) make a model of the measurements. All models should fit exactly under ideal circumstances, that is, with no noise present in the data and the model being correct. Under these circumstances, the models are mathematically equivalent.

In real practical situations, measurement noise is present in the data, and Lambert-Beer's law and the equality of the total concentration profiles do not hold exactly. The differences between the methods show up during the estimation process and in the way the constraints are built in:

1. all three methods use the non-negativity constraints on the estimated spectra and concentration profiles, but these are applied differently. PT2 puts non-negativity on all concentration profiles, whereas RT3 does only apply these constraints on the profiles in $\mathbf{A}_{\mathrm{RT3}}$;
2. the constraint of unimodality of the concentration profiles are used in MCR and PT2. In PT2, this is done in a least squares sense, while in MCR, parameters violating the unimodality constraint are cut off to obey unimodality;
3. the most obvious difference of the methods with respect to the handling of constraints is the way the constraint of equal total concentration profiles is imposed. In MCR, this constraint is not imposed at all; in RT3, this constraint is imposed stringently

by making it, explicitly, a part of the model; and in PT2, the constraint is imposed by choosing the penalizing parameter λ; and

4. due to the way the constraints are imposed, PT2 and RT3 give least squares solutions, whereas MCR gives an approximate least squares solution.

All the estimation methods use starting values and an iterative scheme for estimation. Thus, they all encounter, to a certain extent, the problems of iterative estimation: local minima, convergence problems, slowness in the iterations, etc. Yet, there are ways to deal with these problems, e.g., starting with different starting values, speeding-up mechanisms, etc.

Summarizing, this is not a study of models but a study of methodologies. The methodology consists basically of the following ingredients: the model, the way constraints are imposed, estimation method and initialization of the estimation method.

## 4. Description of the different examples

### 4.1. General

All the following examples will have the same structure. There is always a standard response matrix containing the response of a pure analyte, and this standard is used to quantify the analyte in a mixture which, besides the standard, also contains an unknown interferent. Due to the set-up of the experiments, this interferent is known, but in the calibration, it is treated as unknown.

The calibrations will be performed as a full second-order calibration, i.e. an attempt will be made to quantify the amount of the analyte in the presence of unknown interferents (the second-order advantage). Both quantification for the analyte of interest and qualitative information are important. The first aspect is of primary interest of course, but the second aspect gives room for diagnostic checking of the results. Note that these calibrations really test the different methods: there is only one standard, the mixture contains unknown interferents and the solutes are very similar.

The different examples that follow have different degrees of complexity. They will serve to illustrate the power and shortcoming of the methods.

#### 4.1.1. Example 1

In the first example, the standard is 2-HBA with concentration 0.15 mM. The mixture contains 0.10 mM 2-HBA and 0.05 mM 3-HBA as interferent. A priori, this example is expected to be difficult since 2-HBA and 3-HBA are very similar in response. Mathematically, this calibration problem can be stated as in Eq. (4).

#### 4.1.2. Example 2

In this example, 4-HBA is the analyte of interest and a mixture of 4-HBA and 2-HBA is the unknown, where 2-HBA serves as the unknown interferent. The standard $\mathbf{N}_{4\text{-HBA}}$ is the response of the analyte at a concentration of 0.07 mM, and the mixture contained 0.06 mM 4-HBA and 0.05 mM 2-HBA. This example is expected to be not too difficult since 2-HBA and 4-HBA have dissimilar responses [20]. Mathematically, this calibration problem comes down to

$$\mathbf{N}_{4\text{-HBA}} = \mathbf{ca}_4 \cdot \mathbf{sa}_4^{\mathrm{T}} + \mathbf{cb}_4 \cdot \mathbf{sb}_4^{\mathrm{T}} + \mathbf{E}_{4\text{-HBA}}$$

$$\mathbf{M} = \gamma \cdot \mathbf{ca}_4 \cdot \mathbf{sa}_4^{\mathrm{T}} + \gamma \cdot \mathbf{cb}_4 \cdot \mathbf{sb}_4^{\mathrm{T}} + \mathbf{ca}_{\mathrm{u}} \cdot \mathbf{sa}_{\mathrm{u}}^{\mathrm{T}} + \mathbf{cb}_{\mathrm{u}} \cdot \mathbf{sb}_{\mathrm{u}}^{\mathrm{T}} + \mathbf{E}_M \tag{24}$$

where the **E** matrices refer again to measurement error and the concentration of the interferent is absorbed in $\mathbf{ca}_{\mathrm{u}}$ and $\mathbf{cb}_{\mathrm{u}}$. The restrictions on the concentration profiles in the RT3 model become

$$\mathbf{ca}_4 + \mathbf{cb}_4 = \mathbf{ctot} \quad \text{and} \quad \mathbf{ca}_{\mathrm{u}} + \mathbf{cb}_{\mathrm{u}} = \alpha\mathbf{ctot} \tag{25}$$

where, again, $\alpha$ is an arbitrary constant. Assuming

$$\mathbf{ca}_4 = \mathbf{ctot} - \mathbf{cb}_4 \quad \text{and} \quad \mathbf{cb}_{\mathrm{u}} = \alpha\mathbf{ctot} - \mathbf{ca}_{\mathrm{u}} \tag{26}$$

and rewriting Eq. (24) according to Eq. (26) by defining

$$\mathbf{A} = \begin{bmatrix} \mathbf{ctot} & \mathbf{cb}_4 & \mathbf{ca}_{\mathrm{u}} \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} \mathbf{sa}_4 & \mathbf{sb}_4 & \mathbf{sa}_{\mathrm{u}} & \mathbf{sb}_{\mathrm{u}} \end{bmatrix}$$

$$\mathbf{C} = \begin{bmatrix} 1 & 0 \\ \gamma & 0 \end{bmatrix} \tag{27}$$

and

$$\mathbf{G} = \left[\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & \alpha \\ -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{array}\right] \tag{28}$$

the RT3 model for this calibration is

$$[\mathbf{N}_{4\text{-HBA}}|\mathbf{M}] = \mathbf{AG}\left(\mathbf{C}^{\mathrm{T}} \otimes \mathbf{B}^{\mathrm{T}}\right) + [\mathbf{E}_{4\text{-HBA}}|\mathbf{E}_M] \tag{29}$$

and it can be shown that **C** can be determined uniquely. The uniqueness properties of this model are discussed elsewhere [23]. A consequence of the non-uniqueness of the spectra and profiles of the estimated interferent is that during the estimation, the acidic and basic spectra of the interferent can be interchanged simultaneously with the interchanging of the acidic and basic concentration profiles. To be specific, the pair of matrices **A** and **B** cannot be distinguished from $\mathbf{A}^* = [\mathbf{ctot} \quad \mathbf{cb}_4 \quad \mathbf{cb}_u]$ and $\mathbf{B}^* = [\mathbf{sa}_4 \quad \mathbf{sb}_4 \quad \mathbf{sb}_u \quad \mathbf{sa}_u]$. This is shown in Appendix A and it holds actually for all three calibration methods and is a restatement of the fact that the rank 2 interferent cannot be decomposed uniquely in its basic and acidic spectra/profiles (see also Section 3). However, as stated earlier, the space spanned by the interferent spectra can be estimated uniquely.

The MCR model of the calibration problem of example is

$$\begin{bmatrix} \mathbf{N}_{4\text{-HBA}} \\ \mathbf{M} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{4\text{-HBA}} \\ \mathbf{A}_M \end{bmatrix} \mathbf{B}^{\mathrm{T}} + \mathbf{E} \tag{30}$$

for the proper matrices $\mathbf{A}_{4\text{-HBA}}$, $\mathbf{A}_M$ and **B** defined analogously as in Eq. (11).

The PT2 model of the calibration problem is very similar as given in Eq. (20), the only difference being that, now, 4-HBA is the standard. Changes have to be made accordingly, but an analogous model as in Eq. (20) is obtained and not repeated here.

Table 1
The quantitative results of the three different calibration methods. The symbols used in the table are explained in the text

| Examples | RT3 | MCR | PT2 |
|---|---|---|---|
| *Example 1: 2-HBA* | | | |
| True concentration (mM) | 0.10 | 0.10 | 0.10 |
| Estimated concentration (mM) | 0.0902 | 0.1052 | 0.1093 |
| Relative error (%) | 9.8 | 5.2 | 9.3 |
| $R^2$ | 0.9993 | 0.9990 | 0.9993 |
| *Example 2: 4-HBA* | | | |
| True concentration (mM) | 0.06 | 0.06 | 0.06 |
| Estimated concentration (mM) | 0.0654 | 0.0610 | 0.0626 |
| Relative error (%) | 9.1 | 1.8 | 4.3 |
| $R^2$ | 0.9996 | 0.9994 | 0.9996 |
| *Example 3: 2-HBA* | | | |
| True concentration (mM) | 0.10 | 0.10 | 0.10 |
| Estimated concentration (mM) | 0.1017 | 0.0953 | 0.1026 |
| Relative error (%) | 1.7 | 4.7 | 2.6 |
| $R^2$ | 0.9995 | 0.9995 | 0.9996 |

#### 4.1.3. Example 3

In this example, the roles of 2-HBA and 4-HBA are reversed compared to the previous example. Now, 2-HBA will be quantified in a mixture with 4-HBA serving as the unknown interferent. The standard contains 2-HBA at a concentration of 0.15 mM, and the mixture contains 2-HBA at a concentration of 0.10 mM and the interferent 4-HBA at a concentration of 0.06 mM.

The mathematical formulation of the calibration problem in terms of RT3, MCR and PT2 is very similar to the ones described in example 2 and will, therefore, not be repeated. Again, for RT3, the concentration matrix **C** can be determined uniquely.

## 5. Results and discussion

### 5.1. Quantitative analysis

The quantitative results are summarized in Table 1. This table shows that there is no 'king' method. The numbers in this table have to be compared with the reproducibility of 5%. Example 1 was a priori expected to be the most difficult one, and indeed, the results for this example are consistently the worst of all examples for all calibration methods. Nevertheless, it can be concluded that, for all the examples, the quantitative results of all calibration methods are reasonable, keeping in mind the presence of unknown interferents and the 5% reproducibility error. The fit of all models (RT3, MCR and PT2) to the data was calculated (Table 1). This can be judged by the $R^2$ values. The $R^2$ value has a maximum of 1, meaning a perfect fit. All $R^2$ values are above 0.999, indicating a very good fit of the data. The models differ mainly due to the way the restrictions are imposed, but they do not differ much in structure: they all fit the data very well. The fact that the models behave similarly shows that, for this particular application, the differences between methodologies do not dominate the solution.

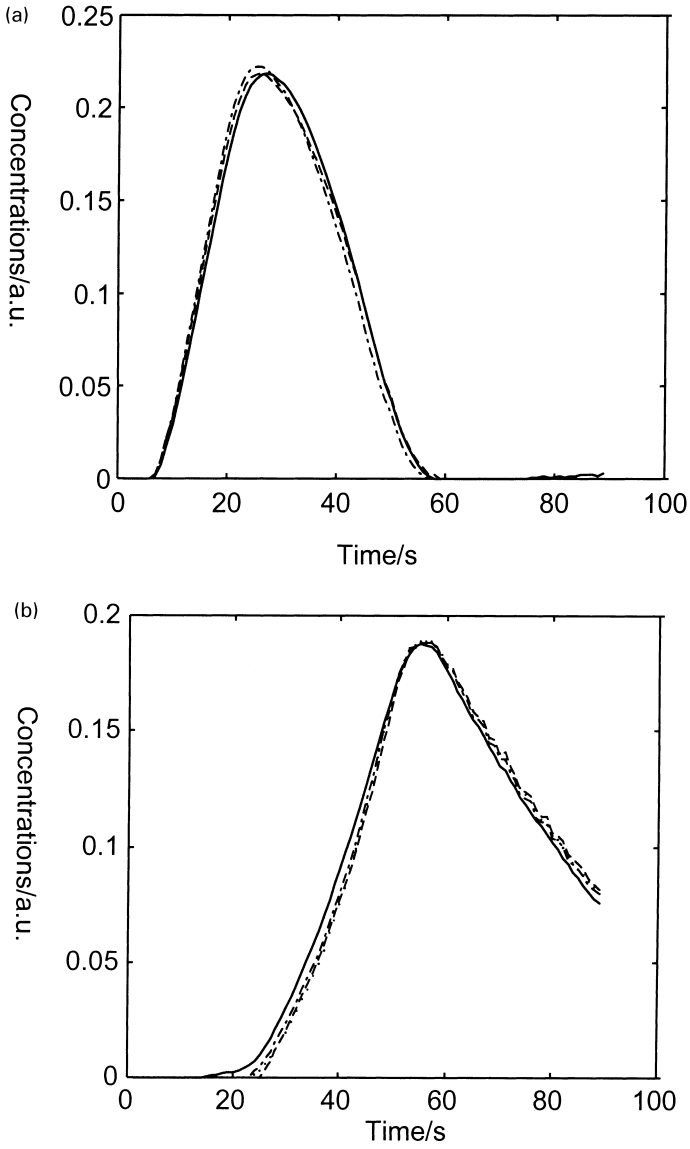


Fig. 4. (a) The estimated acidic concentration profile of the unknown interferent with the different methods. The abbreviation 'a.u.' stands for 'arbitrary units'. True (-), MCR (-.) and PT2 (- -) estimated profiles. (b) The estimated basic concentration profile of the unknown interferent with the different methods. True (-), MCR (-.), PT2 (- -) and RT3 (..) estimated profiles. (c) The estimated total concentration profile of the unknown interferent with RT3. True (-) and RT3 (..) estimated profiles.

### *5.2. Qualitative analysis*

In Section 2, it has already been mentioned that reasonable estimates of the true spectra and concentration profiles were available from the auxiliary experiment. Hence, the estimated spectra and concentration profiles can be compared with those.

The qualitative results of example 2 are shown because these are representative of examples 1–3. All methods give good estimates of the concentration pro-

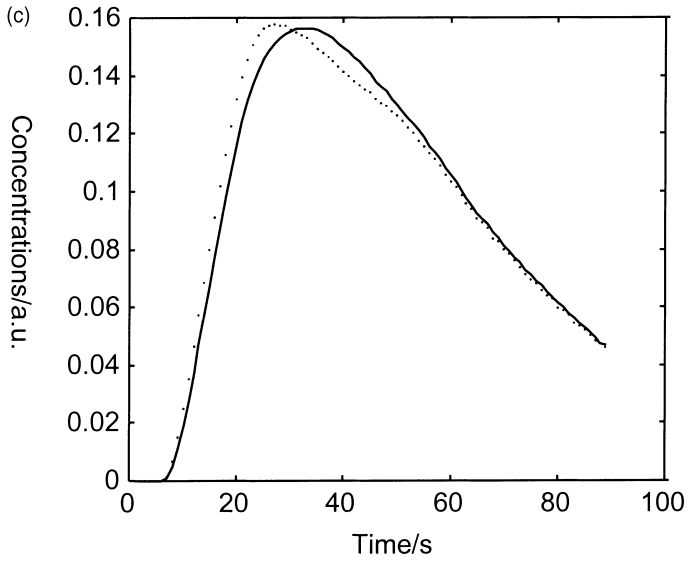


Fig. 4. (*Continued*).

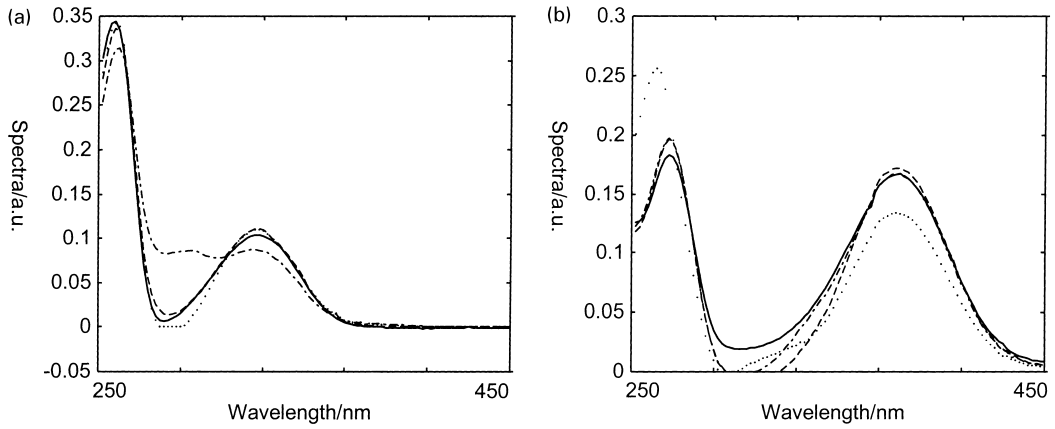


Fig. 5. (a) The estimated acidic spectrum of the unknown interferent with the different methods. True (-), MCR (-.), PT2 (- -) and RT3 (..) estimated spectra. (b) The estimated basic spectrum of the unknown interferent with the different methods. True (-), MCR (-.), PT2 (- -) and RT3 (..) estimated spectra.

files and spectra of the acidic and basic form of the analyte 4-HBA (not shown). This is not surprising since the problem is well-posed with respect to these parameters.

After proper matching of the true and estimated acidic and basic spectra and concentration profiles of the interferent, Fig. 4 and Fig. 5 could be made. Fig. 4 (a) shows the estimates of the acidic concentration profile of the unknown interferent (2-HBA). Clearly, MCR and PT2 do give a good estimate. Likewise, Fig. 4 (b) and (c) give the estimates of $\mathbf{cb}_u$and **ctot**, respectively. These estimates are also very reasonable.

Fig. 5 (a) and (b) show the quality of the estimates of the unknown acidic and basic spectra. All methods perform reasonably well, although MCR seems to have some trouble with the acidic spectrum of the interferent, and likewise, RT3 for the basic spectrum of the interferent. Note, however, that all three methods

are not capable of estimating uniquely the acidic and basic spectra of the interferent, as has been mentioned before.

## 6. Conclusions

In this paper, alternative methods are given and compared for complex second-order calibration situations. The methods give similar results for the quantitative analysis, making it clear that there is no king method. The quantitative results are good, keeping in mind the complexity of the problem, which shows that, even for such complex data, the second-order advantage is obtainable.

The qualitative results are less impressive. This indicates that, going from the straightforward second-order calibration cases (rank 1 pure analyte responses) to more complex situations, a price has to be paid. This price is in the qualitative analyses and not in the quantitative analysis.

Scrutinizing the structural bases for the three models, it is clear that they are equal under ideal conditions, i.e., no measurement error, Lambert-Beer's law holds exactly and the total concentration profiles are exactly equal. As stated before in Section 3, the whole methodology should be taken into account.

There are no clear recommendations for the method of choice. It is merely a matter of taste, interpretability and availability of software. The flexibility of imposing constraints to a certain degree (as was done in the PT2 model) is appealing. Such a flexibility can also be combined with the other two approaches (MCR and RT3). When more than one unknown interferent is present, the situation becomes more difficult. This is a subject of further investigation.

## Acknowledgements

Lars Nørgaard and Carsten Ridder are thanked for making available the FIA data. Hans Meurs is thanked for performing some of the experiments; Gjalt Feenstra for helping in formatting the figures; and Henk Kiers is thanked for stimulating discussions.

## Appendix A. Proof that $\mathbf{A}_{\mathrm{MCR}}$ has full column rank

Consider the matrix $\mathbf{A}_{\mathrm{MCR}}$ $(2N \times 4)$

$$\mathbf{A}_{\mathrm{MCR}} = \begin{bmatrix} \mathbf{ca}_2 & \mathbf{cb}_2 & \mathbf{0} & \mathbf{0} \\ \gamma \cdot \mathbf{ca}_2 & \gamma \cdot \mathbf{cb}_2 & \mathbf{ca}_{\mathrm{u}} & \mathbf{cb}_{\mathrm{u}} \end{bmatrix} \tag{A.1}$$

It is assumed that $2N \geq 4$; $\mathbf{ca}_2 \neq 0$ and $\mathbf{cb}_2 \neq 0$ are not proportional; and $\mathbf{ca}_{\mathrm{u}} \neq 0$ and $\mathbf{cb}_{\mathrm{u}} \neq 0$ are not proportional, otherwise $\mathbf{A}_{\mathrm{MCR}}$, trivially, does not have full column rank. These requirements are fulfilled in practice (e.g., in the examples of this paper). In order to be of full rank, the columns of $\mathbf{A}_{\mathrm{MCR}}$ must satisfy

$$\mu_1 \begin{bmatrix} \mathbf{ca}_2 \\ \gamma \cdot \mathbf{ca}_2 \end{bmatrix} + \mu_2 \begin{bmatrix} \mathbf{cb}_2 \\ \gamma \cdot \mathbf{cb}_2 \end{bmatrix} + \mu_3 \begin{bmatrix} 0 \\ \mathbf{ca}_{\mathrm{u}} \end{bmatrix} + \mu_4 \begin{bmatrix} 0 \\ \mathbf{cb}_{\mathrm{u}} \end{bmatrix} = 0 \Rightarrow \mu_1 = \mu_2 = \mu_3 = \mu_4 = 0 \tag{A.2}$$

Consider the upper part of Eq. (A.2): $\mu_1 \mathbf{ca}_2 + \mu_2 \mathbf{cb}_2 = 0$. If $\mu_1 \neq 0$, then $\mu_2 \neq 0$ (since $\mathbf{ca}_2 \neq 0$); but this implies that $\mathbf{ca}_2$ and $\mathbf{cb}_2$ are proportional. Since it is assumed that $\mathbf{ca}_2$ and $\mathbf{cb}_2$ are not proportional, it holds that $\mu_1 = 0$. Hence, since $\mathbf{cb}_2 \neq 0$, it holds that $\mu_2 = 0$. Consider now the lower part of Eq. (A.3): $\mu_3 \mathbf{ca}_{\mathrm{u}} + \mu_4 \mathbf{cb}_{\mathrm{u}} = 0$. Following the same reasoning, it holds that $\mu_3 = \mu_4 = 0$. Hence, the implication in Eq. (A.2) holds and $\mathbf{A}_{\mathrm{MCR}}$ has full rank under the assumptions made in the beginning of this section.

## Appendix B. Proof that RT3 and PT2 are equal under ideal conditions

The RT3 model is given by Eqs. (8), (9) and (10). For ideal conditions (no measurement error, Lambert-Beer's law holds exactly and the total concentration profiles are exactly equal), the error term in Eq. (10) should be dropped. For any Tucker model (and consequently, also for RT3 models), there are alternative ways to write such a model [27]. An alternative way of writing the (noiseless) Eq. (10) is

$$\mathbf{X}_i = \mathbf{C}_{\mathrm{RT3}} \left[ \sum_{p=1}^{3} \mathbf{a}_{ip} \mathbf{G}_p \right] \mathbf{B}_{\mathrm{RT3}}^{\mathrm{T}}; \quad i = 1, \ldots, 89 \tag{A.3}$$

where $\mathbf{a}_{ip}$ is the $ip$th typical element of $\mathbf{A}_{\mathrm{RT3}}$, $\mathbf{X}_i$ is the $i$th slice with dimensions $100 \times 2$ of the three-way

array $\mathbf{X}$ formed by stacking $\mathbf{N}_{2\text{-HBA}}$ and $\mathbf{M}$ on top of each other; $\mathbf{G}_1$, $\mathbf{G}_2$ and $\mathbf{G}_3$ are rearrangements of $\mathbf{G}_{RT3}$:

$$\mathbf{G}_1 = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{G}_2 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & \alpha & 0 \end{bmatrix}$$

$$\mathbf{G}_3 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix} \tag{A.4}$$

Writing out the summation term in Eq. (A.3), this becomes

$$\begin{bmatrix} \mathbf{ca}_{2,i} & \mathbf{ctot}_i - \mathbf{ca}_{2,i} & 0 & 0 \\ 0 & 0 & \alpha\mathbf{ctot}_i - \mathbf{cb}_{u,i} & \mathbf{cb}_{u,i} \end{bmatrix} \tag{A.5}$$

The PT2 model is shown in Eq. (20). The matrices $\mathbf{X}_i$, $\mathbf{C}_{PT2}$, and $\mathbf{B}_{PT2}$ equal their counterparts in Eq. (A.3) ($\mathbf{X}_i$, $\mathbf{C}_{RT3}$ and $\mathbf{B}_{RT3}$, respectively). Hence, it remains to be shown that $\mathbf{G}_{PT2}\mathbf{A}_{PT2,i}$ equals Eq. (A.5). $\mathbf{A}_{PT2,i}$is defined in Eq. (17) and by substituting $\mathbf{cb}_{2,i} = \mathbf{ctot}_i - \mathbf{ca}_{2,i}$ and $\mathbf{ca}_{u,i}$=$\alpha\mathbf{ctot}_i - \mathbf{cb}_{u,i}$, $\mathbf{A}_{PT2,i}$ becomes

$$\mathbf{A}_{PT2,i} = \begin{bmatrix} \mathrm{ca}_{2,i} & 0 & 0 & 0 \\ 0 & \mathrm{ctot}_i - \mathrm{ca}_{2,i} & 0 & 0 \\ 0 & 0 & \alpha \cdot \mathrm{ctot}_i - \mathrm{cb}_{u,i} & 0 \\ 0 & 0 & 0 & \mathrm{cb}_{u,i} \end{bmatrix} \tag{A.6}$$

and if this $\mathbf{A}_{PT2,i}$ is multiplied by the $\mathbf{G}_{PT2}$ of Eq. (17), it results in Eq. (A.5). Hence, RT3 and PT2 are equal if the equality constraint is applied.

## Appendix C. Proof that MCR and RT3 are equal under ideal conditions

In this part, it is shown that under ideal conditions (as defined above), the MCR and RT3 models are equal. Note that the equality constraint is explicitly accounted for in the RT3 model, whereas this constraint is not imposed on the MCR model. Nevertheless, it is assumed that the constraint holds exactly in the data.

The MCR model is stated in Eq. (13). In Appendix A, the error term $\mathbf{E}$ is assumed to be zero. Eq. (13) can be rewritten as

$$\begin{aligned} \begin{bmatrix} \mathbf{N}_{2\text{-HBA}}^T & \mathbf{M}^T \end{bmatrix} &= \mathbf{B}\begin{bmatrix} \mathbf{A}_{2-\text{HBA}}^T & \mathbf{A}_M^T \end{bmatrix} \\ &= \mathbf{B}\begin{bmatrix} \mathbf{ca}_2^T & \gamma \cdot \mathbf{ca}_2^T \\ \mathbf{ctot}^T - \mathbf{ca}_2^T & \gamma \cdot \mathbf{ctot}^T - \gamma \cdot \mathbf{ca}_2^T \\ 0 & \alpha \cdot \mathbf{ctot}^T - \mathbf{cb}_u^T \\ 0 & \mathbf{cb}_u^T \end{bmatrix} \\ &= \mathbf{B} \cdot \mathbf{A}_{MCR}^T \end{aligned} \tag{A.7}$$

where Eq. (6) is used to account for the equality constraint in the data.

The RT3 model of Eq. (10) can be rewritten (skipping the error term) as

$$\begin{bmatrix} \mathbf{N}_{2\text{-HBA}}^T & \mathbf{M}^T \end{bmatrix} = \mathbf{B}\tilde{\mathbf{G}}_{RT3}(\mathbf{A}_{RT3}^T \otimes \mathbf{C}^T) \tag{A.8}$$

where

$$\mathbf{G}_{RT3} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \alpha & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{A.9}$$

and, clearly, MCR equals RT3 if $\mathbf{A}_{MCR}^T = \tilde{\mathbf{G}}_{RT3}(\mathbf{A}_{RT3}^T \otimes \mathbf{C}^T)$.Writing out the Kronecker product of this last term, this becomes

$$\begin{aligned} \mathbf{A}_{RT3}^T \otimes \mathbf{C}^T &= \begin{bmatrix} \mathbf{ca}_2^T \\ \mathbf{ctot}^T \\ \mathbf{cb}_u^T \end{bmatrix} \otimes \begin{bmatrix} 1 & \gamma \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{ca}_2^T & \gamma \cdot \mathbf{ca}_2^T \\ \mathbf{ctot}^T & \gamma \cdot \mathbf{ctot}^T \\ \mathbf{cb}_u^T & \gamma \cdot \mathbf{cb}_u^T \\ 0 & \mathbf{ca}_2^T \\ 0 & \mathbf{ctot}^T \\ 0 & \mathbf{cb}_u^T \end{bmatrix} \end{aligned} \tag{A.10}$$

and upon multiplying Eq. (A.10) with $\tilde{\mathbf{G}}_{RT3}$using the rules of partitioned matrices, the result is $\mathbf{A}_{MCR}^T$. Hence, MCR equals RT3 when the equality constraint is present in the data and applied to the RT3 model.

## Appendix D. Confusion of interferent spectra and profiles in the RT3 model

In this part of Appendix A, it is shown that the pure spectra and concentration profiles of the interferents can be permuted without changing the solution. This will be shown for the RT3 model in example 2, but it holds for all models in all examples.

Using the restrictions in Eq. (25) to reformulate $\mathbf{N}_{4\text{-HBA}}$ and $\mathbf{M}$ gives

$$\mathbf{N}_{4\text{-HBA}} = \mathbf{ctot} \cdot \mathbf{sa}_4^{\mathrm{T}} - \mathbf{cb}_4 \cdot \mathbf{sa}_4^{\mathrm{T}} + \mathbf{cb}_4 \cdot \mathbf{sb}_4^{\mathrm{T}}$$

$$\begin{aligned}\tilde{\mathbf{M}} = {} & \gamma \cdot \mathbf{ctot} \cdot \mathbf{sa}_4^{\mathrm{T}} - \gamma \cdot \mathbf{cb}_4 \cdot \mathbf{sa}_4^{\mathrm{T}} + \gamma \cdot \mathbf{cb}_4 \cdot \mathbf{sb}_4^{\mathrm{T}} \\ & + \mathbf{ca}_{\mathrm{u}} \cdot \mathbf{sa}_{\mathrm{u}}^{\mathrm{T}} + \alpha \cdot \mathbf{ctot} \cdot \mathbf{sb}_{\mathrm{u}}^{\mathrm{T}} - \mathbf{ca}_{\mathrm{u}} \cdot \mathbf{sb}_{\mathrm{u}}^{\mathrm{T}}\end{aligned} \tag{A.11}$$

Simultaneous changing $\mathbf{ca}_{\mathrm{u}}$, $\mathbf{cb}_{\mathrm{u}}$ and $\mathbf{sa}_{\mathrm{u}}$, $\mathbf{sb}_{\mathrm{u}}$ gives

$$\mathbf{N}_{4\text{-HBA}} = \mathbf{ctot} \cdot \mathbf{sa}_4^{\mathrm{T}} - \mathbf{cb}_4 \cdot \mathbf{sa}_4^{\mathrm{T}} + \mathbf{cb}_4 \cdot \mathbf{sb}_4^{\mathrm{T}}$$

$$\begin{aligned}\tilde{\mathbf{M}} = {} & \gamma \cdot \mathbf{ctot} \cdot \mathbf{sa}_4^{\mathrm{T}} - \gamma \cdot \mathbf{cb}_4 \cdot \mathbf{sa}_4^{\mathrm{T}} + \gamma \cdot \mathbf{cb}_4 \cdot \mathbf{sb}_4^{\mathrm{T}} \\ & + \mathbf{cb}_{\mathrm{u}} \cdot \mathbf{sb}_{\mathrm{u}}^{\mathrm{T}} + \alpha \cdot \mathbf{ctot} \cdot \mathbf{sa}_{\mathrm{u}}^{\mathrm{T}} - \mathbf{cb}_{\mathrm{u}} \cdot \mathbf{sa}_{\mathrm{u}}^{\mathrm{T}}\end{aligned} \tag{A.12}$$

and using $\mathbf{cb}_{\mathrm{u}} = \alpha\mathbf{ctot} - \mathbf{ca}_{\mathrm{u}}$ to rewrite $\tilde{\mathbf{M}}$ gives

$$\begin{aligned}\tilde{\mathbf{M}} = {} & \gamma \cdot \mathbf{ctot} \cdot \mathbf{sa}_4^{\mathrm{T}} - \gamma \cdot \mathbf{cb}_4 \cdot \mathbf{sa}_4^{\mathrm{T}} + \gamma \cdot \mathbf{cb}_4 \cdot \mathbf{sb}_4^{\mathrm{T}} \\ & + (\alpha \cdot \mathbf{ctot} - \mathbf{ca}_{\mathrm{u}})\mathbf{sb}_{\mathrm{u}}^{\mathrm{T}} + \alpha \cdot \mathbf{ctot} \cdot \mathbf{sa}_{\mathrm{u}}^{\mathrm{T}} \\ & - (\alpha \cdot \mathbf{ctot} - \mathbf{ca}_{\mathrm{u}})\mathbf{sa}_{\mathrm{u}}^{\mathrm{T}} = \tilde{\mathbf{M}}\end{aligned} \tag{A.13}$$

where the last line of Eq. (A.13) can be deduced by proper collecting of the terms. Hence, the calibration models where both $\mathbf{sa}_{\mathrm{u}}$, $\mathbf{sb}_{\mathrm{u}}$, and simultaneously, $\mathbf{ca}_{\mathrm{u}}$, $\mathbf{cb}_{\mathrm{u}}$ are permuted cannot be distinguished from each other.

## References


[1] K.S. Booksh, B.R. Kowalski, Anal. Chem. 66 (1994) 782A.
[2] E. Sanchez, B.R. Kowalski, J. Chemometrics 2 (1988) 247.
[3] P.J. Brown, J.R. Statist. Soc. B 44 (1982) 287.
[4] H. Martens and T. Naes, Multivariate Calibration, Wiley, Chichester, 1991.
[5] A. Höskuldsson, J. Chemometrics 2 (1988) 211.
[6] S. De Jong, H.A.L. Kiers, Chemometrics and Intelligent Laboratory Systems 14 (1992) 155.
[7] D.M. Haaland, E.V. Thomas, Anal. Chem. 60 (1988) 1193.
[8] T. Hirschfeld, Anal. Chem. 52 (1980) 297A.
[9] S. Leurgans, R. Ross, Statistical Sci. 7 (1992) 289.
[10] R. Tauler, A.K. Smilde, J.M. Henshaw, L.W. Burgess, B.R. Kowalski, Anal. Chem. 66 (1994) 3337.
[11] A.K. Smilde, R. Tauler, J.M. Henshaw, L.W. Burgess, B.R. Kowalski, Anal. Chem. 66 (1994) 3345.
[12] C-N Ho, G.D Davidson, E.R. Davidson, Anal. Chem. 50 (1978) 1108.
[13] E. Sanchez, B.R. Kowalski, Anal. Chem. 58 (1986) 496.
[14] B.E. Wilson, W. Lindberg, B.R. Kowalski, J. Am. Chem. Soc. 111 (1989) 3797.
[15] Y. Wang, O. Borgen, B.R. Kowalski, M. Gu, F. Turecek, J. Chemometrics 7 (1993) 117.
[16] H.A.L. Kiers, A.K. Smilde, J. Chemometrics 9 (1995) 179.
[17] R. Tauler, A.K. Smilde, B.R. Kowalski, J. Chemometrics 9 (1995) 31.
[18] R. Tauler, Chemometrics and Intelligent Laboratory Systems 30 (1995) 133.
[19] A.K. Smilde, Y. Wang, B.R. Kowalski, J. Chemometrics 8 (1994) 21.
[20] L. Nørgaard, C. Ridder, Chemometrics and Intelligent Laboratory Systems 23 (1994) 107.
[21] D.D. Perrin and B. Dempsey, Buffers for pH and Metal Ion Control, Chapman & Hall, London, 1974.
[22] H.R. Keller, D.L. Massart, Analytica Chimica Acta 246 (1991) 379.
[23] H.A.L. Kiers, A.K. Smilde, J. Chemometrics 12 (1998) 125.
[24] C.L. Lawson and R.J. Hanson, Solving Least Squares Problems, Prentice Hall, Englewood Cliffs, 1974.
[25] M. Amrhein, B. Srinivasan, D. Bonvin, M.M. Schumacher, Chemometrics and Intelligent Laboratory Systems 33 (1996) 17.
[26] R.A. Harshman, M.E. Lundy, Psychometrika 61 (1996) 133.
[27] P.M. Kroonenberg, Three-mode Principal Component Analysis, DSWO Press, Leiden, 1983.