JOURNAL OF CHEMOMETRICS, VOL. 12, 125–147 (1998)

CONSTRAINED THREE-MODE FACTOR ANALYSIS AS A TOOL FOR PARAMETER ESTIMATION WITH SECOND-ORDER INSTRUMENTAL DATA

HENK A. L. KIERS[1] AND AGE K. SMILDE[2]
[1]*Department of Psychology, University of Groningen, Grote Kruisstraat 2/1, NL-9712 TS Groningen, The Netherlands*
[2]*Laboratory for Analytical Chemistry, University of Amsterdam, Nieuwe Achtergracht 166, NL-1018 WV Amsterdam, The Netherlands*

SUMMARY

Three-mode factor analysis models are often used in exploratory analysis of three-way data. However, in some situations it is a priori known that a particular constrained three-mode factor analysis (C3MFA) model describes an underlying process exactly. In such situations, fitting a C3MFA model to a data set can be used for both quantitative analysis (e.g. estimating concentrations of a chemical substance in a mixture) and qualitative analysis (e.g. on the basis of certain subsets of parameters one can identify the substances present in a mixture). In this paper a general algorithm for fitting a range of such C3MFA models is proposed. Whether C3MFA is used for qualitative or quantitative analysis, in both cases it is crucial that the relevant parameter estimates are uniquely determinable. In the present paper it is discussed how and to what extent uniqueness of certain model parameters can be assessed. © 1998 John Wiley & Sons, Ltd.

J. Chemomerics, Vol. **12,** 125–147 (1998)

KEY WORDS multiway methods; calibration; uniqueness; rotational ambiguity

1. INTRODUCTION

1.1 Multiway methods

Multiway methods are statistical methods that deal with multiway data. Multiway data is the generic term for two-way data (matrices), three-way data (a cube or a block of data), four-way data and so on. In general, multiway data of order *N* are data that can be meaningfully arranged in an *N*-way matrix. The majority of multivariate analysis methods work with two-way data. In the following the discussion will be focused on three-way or three-mode data, but it should be kept in mind that extensions to higher order are possible.

In chemistry and chemical engineering it was recognized by several researchers that some problems and questions generate data that can be arranged in a three-way set-up.[1,2] This has led to an

* Correspondence to: H. A. L. Kiers, Department of Psychology, University of Groningen, Grote Kruisstraat 2/1, NL-9712 TS Groningen, The Netherlands.
Contract grant sponsor: Royal Netherlands Academy of Arts and Sciences.

CCC 0886–9383/98/020125–23 $17.50

Received 18 August 1997
Accepted 10 November 1997

introduction of three-way methods in chemistry and chemical engineering, and the research in this area is growing. Different potential application areas can be recognized, amongst them second-order calibration,[3] identification of compounds in complex biochemical systems,[4] image analysis[5] and multivariate statistical process control.[6,7] In most of these application areas the power of three-way methods was clearly demonstrated.

1.2 Three-mode factor analysis and (non-)uniqueness

One of the oldest three-way methods is three-mode factor analysis[8,9] (3MFA). This method has mostly been used as an exploratory technique for the analysis of three-mode data. The method is thus used for describing the most important aspects of the data in terms of components for all three modes. An important problem of the method is that, as in two-way factor analysis, the components are by no means unique.

In the two-way case the uniqueness problem is typically solved by one of the following approaches. One approach is to rotate the loadings to a simple structure; another approach is to constrain some loadings to be zero in such a way that the model becomes identified. In chemical applications the problem of rotational freedom is tackled by imposing constraints such as non-negativity and unimodality on the solution. Although this helps in constraining the solution, it does not necessarily make the solution unique.[10] If selectivity is present, then the rotational problem can be solved in certain cases.[11] For 3MFA the uniqueness problem has been approached in analogy to the two-way case, and proposals for simple structure rotation have been made as well.

The above approaches are meant for exploratory analysis of three-mode data. An alternative approach, which is not fully exploratory, is based on constraining certain parameters in the 3MFA model to zero.[12,13] In certain situations in chemistry, particular constrained 3MFA (C3MFA) models are known to conform exactly to the process underlying the measured data values.[14] In such instances, fitting the C3MFA model to a data set is not meant for assessing which model should be used to describe the data (as in exploratory analysis), but for determining which parameter values govern the known underlying model. This parameter estimation in turn can be used for quantitative analysis (e.g. estimating concentrations of a chemical substance in a mixture) as well as for qualitative analysis (e.g. on the basis of certain subsets of the estimated parameters one can identify the substances present in a mixture). When parameter estimation is the main purpose of the analysis, uniqueness of the parameters at stake becomes crucial. This is because the data analyst will not accept non-unique parameter values even if they describe the data well. In the present context the parameters actually govern observed processes which refer to (a unique) reality. Hence the parameters themselves must be unique.

For some C3MFA models, uniqueness of the parameters has been proven,[15] but most C3MFA models encountered in practice do not belong to this limited class of unique C3MFA models. Therefore in most practical cases the analyst cannot rely on uniqueness results from the literature, and uniqueness has to be assessed for each model separately. Moreover, so far only C3MFA methods have been described where certain elements of the core are constrained to be zero. In certain applications, however, a more flexible approach is desired. For instance, in addition to the core elements, also elements of the component matrices may have to be constrained. The first purpose of the present paper is to propose a general C3MFA algorithm which, in addition to zero constraints on the core, allows for a variety of constraints, both on the core and on the component matrices. Furthermore, such constraints will aid the identification of the model and in certain cases lead to models where at least certain subsets of parameters can be identified uniquely. The second purpose of this paper is to demonstrate that in certain chemical applications such partial uniquenesses can be proven to hold, thus extending earlier results,[15] and can be exploited in the actual estimation of

concentrations. Before considering these topics, we briefly describe a chemical application which will serve as leading example, on which the main developments in the present paper will be illustrated right after their introduction.

1.3. Calibration of a second-order sensor

As a leading example we consider the calibration of a reaction-based second-order sensor for chlorinated hydrocarbons. Such a sensor can be used for quick assessment of hazardous compounds in the environment. An extensive description of the sensor is given elsewhere.[16,17] In this paper we shall only give a short introduction to the sensor and focus on calibration issues.

The sensor is a small reaction chamber sealed with a membrane. If the sensor is placed in the sample, analytes can diffuse through the membrane. A reaction (the so-called Fujiwara reaction) starts in which the analytes react with the reagent to form UV/VIS absorbing species in time. At equidistant points in time a UV/VIS spectrum is obtained. This generates a matrix of responses (time versus wavelength), hence the name second-order sensor. Assuming first-order kinetics, the concentrations of the analytes are linearly related to the absorbances of the species. The membrane creates a certain amount of selectivity, ensuring that only analytes and interferents of a similar structure diffuse through the membrane.

Here we consider the problem of measuring trichloroethylene in a mixture of trichloroethylene (TCE) and an unknown interferent. Owing to the construction of the sensor, the interferent is known to be a member of the class of small trichlorinated compounds (such as TCE and chloroform ($CHCl_3$)), because only these compounds pass through the membrane and give spectroscopically active products in the reaction. The standard with which the calibration is performed contains only TCE. This is a difficult calibration problem because the analyte and the unknown interferent are very similar. Since there is no possibility to avoid interferences by, for instance, a different sensor design or a different type of reaction, a good calibration procedure which is capable of quantitating in the presence of highly similar interferences is of much practical value.

There are two aspects related to the calibration of the sensor: a quantitative and a qualitative aspect. Quantitation means that the sensor should be able to quantitate in an unknown sample the amount of the analyte with a certain precision. Qualitative analysis pertains to the possibility to identify unknown interferents (if present) in the sample. This aspect is also relevant since identification of an unknown interferent using its spectra and getting insight into the kinetics of the reaction using the temporal profiles are important.

2. CONSTRAINED THREE-MODE FACTOR ANALYSIS

2.1. Theory

The three-mode factor analysis (3MFA) model can be described as follows. Let **X** denote a three-way array with elements x_{ijk}, $i = 1,\ldots, I$, $j = 1,\ldots,J$,$k = 1,\ldots,K$. Then the 3MFA model is given by

$$x_{ijk} = \sum_{p=1}^{P}\sum_{q=1}^{Q}\sum_{r=1}^{R} a_{ip}b_{jq}c_{kr}g_{pqr} + e_{ijk} \qquad (1)$$

where a_{ip}, b_{jq} and c_{kr} are elements of the three component matrices **A, B** and **C** of orders $I \times P$, $J \times Q$ and $K \times R$ respectively, **G** is a $P \times Q \times R$ three-way array denoted as the *core*, and e_{ijk} denotes the error term for observation x_{ijk}. The matrices **A, B** and **C** can be considered component weights for 'A-mode components' (in **A**), 'B-mode components' (in **B**) and 'C-mode components' (in **C**) respectively. The

elements of the core indicate how the components from the different modes interact. The 3MFA model is fitted to a data array by minimizing the sum of squared error terms, using an alternating least squares (ALS) algorithm.[9]

To provide some insight into the role of the core array, we give the following (simplified) tensorial description of the 3MFA model. Considering **x** as a vector containing all elements of the three-way array **X** (ordered such that the first index runs slowest and the last fastest), and **e** as the vector of corresponding error terms, the 3MFA model can be written as

$$\begin{aligned}\mathbf{x} = {} & g_{111}(\mathbf{a}_1 \otimes \mathbf{b}_1 \otimes \mathbf{c}_1) + g_{112}(\mathbf{a}_1 \otimes \mathbf{b}_1 \otimes \mathbf{c}_2) + g_{113}(\mathbf{a}_1 \otimes \mathbf{b}_1 \otimes \mathbf{c}_3) + \ldots \\ & + g_{211}(\mathbf{a}_2 \otimes \mathbf{b}_1 \otimes \mathbf{c}_1) + \ldots + g_{PQR}(\mathbf{a}_P \otimes \mathbf{b}_Q \otimes \mathbf{c}_R) + \mathbf{e} \end{aligned} \tag{2}$$

where $\otimes$ denotes the (right) Kronecker product and $\mathbf{a_i} \otimes \mathbf{b_j} \otimes \mathbf{c_k}$ denotes the vector with all triple products between elements of column i of **A**, column j of **B** and column k of **C**. It is clear from (2) that the elements of the core serve as indices for these vectors of triple products.

A third way of describing the model is in terms of the frontal planes of the data array. Let $\mathbf{X_k}$, $k = 1,\ldots,K$, denote the kth frontal plane of **X**. Then the 3MFA model can be described as

$$\mathbf{X}_k = \mathbf{A}\sum_{r=1}^{R} c_{kr}\mathbf{G}_r\mathbf{B}^{\mathrm{T}} + \mathbf{E}_k \tag{3}$$

where $\mathbf{E_k}$ contains the error terms for the kth frontal plane. Formulation (3) relates the 3MFA model to well-known two-way models. In fact, (3) shows that the structural part of the frontal plane $\mathbf{X_k}$ is decomposed into the product of three low-rank matrices, namely **A, B** and $\Sigma_r c_{kr}\mathbf{G}_r$.

2.2. C3MFA model for second-order sensor calibration

When data are obtained in a second-order sensor calibration study (see Section 1.3), we first have to check how many interferents we have, as can be done by assessing the approximate rank of the mixture matrix **M**. Here we consider the case where the number of interferents turns out to be one; when more than one interferent is present, a similar but more extensive model can be used. Here the data for the calibration sample (in **N**) and for the mixture (in **M**) can be described mathematically as

$$\mathbf{N} = z_1\mathbf{x}_2\mathbf{y}_2^{\mathrm{T}} + z_1\mathbf{x}_3\mathbf{y}_3^{\mathrm{T}} + \mathbf{E}_N \tag{4a}$$

$$\mathbf{M} = z_2\mathbf{x}_2\mathbf{y}_2^{\mathrm{T}} + z_2\mathbf{x}_3\mathbf{y}_3^{\mathrm{T}} + z_3\mathbf{x}_4\mathbf{y}_4^{\mathrm{T}} + z_3\mathbf{x}_5\mathbf{y}_5^{\mathrm{T}} + \mathbf{E}_M \tag{4b}$$

where $\mathbf{E_N}$ and $\mathbf{E_M}$ contain error terms for the calibration sample and the mixture respectively and

$\mathbf{x}_2$ temporal profile of species 1 formed by the analyte (TCE)
$\mathbf{y}_2$ UV/VIS spectrum of species 1 formed by the analyte
$\mathbf{x}_3$ temporal profile of species 2 formed by the analyte
$\mathbf{y}_3$ UV/VIS spectrum of species 2 formed by the analyte
$\mathbf{x}_4$ temporal profile of species 1 formed by the interferent
$\mathbf{y}_4$ UV/VIS spectrum of species 1 formed by the interferent
$\mathbf{x}_5$ temporal profile of species 2 formed by the interferent
$\mathbf{y}_5$ UV/VIS spectrum of species 2 formed by the interferent
z_1 concentration of the analyte in the standard
z_2 concentration of the analyte in the mixture
z_3 concentration of the interferent in the mixture

in which the first species is an intermediate product in the reaction and the second species is an end-product.

As stated earlier, the interferents are similar to the analyte in chemical structure. For this particular sensor, all compounds diffusing through the membrane are small trichlorinated compounds such as TCE and $CHCl_3$. According to the mechanisms proposed for the Fujiwara reaction, it is reasonable to assume that these small trichlorinated compounds form the same end-product, i.e. the second species.[18,19] Note that the temporal profiles do not need to be equal: different reaction rates can be involved in the formation of the second species by either the analyte or the interferent. Formally, the above knowledge implies that $\mathbf{y}_5$ equals $\mathbf{y}_3$. When, for convenience, the relative concentration z_2/z_1 is replaced by γ, z_1 is absorbed in $\mathbf{x}_2$ and $\mathbf{x}_3$, and z_3 is absorbed in $\mathbf{x}_4$ and $\mathbf{x}_5$, (4a,b) reduce to

$$\mathbf{N} = \mathbf{x}_2\mathbf{y}_2^{\mathrm{T}} + \mathbf{x}_3\mathbf{y}_3 + \mathbf{E}_N \tag{5a}$$

$$\mathbf{M} = \gamma\mathbf{x}_2\mathbf{y}_2^{\mathrm{T}} + \gamma\mathbf{x}_3\mathbf{y}_3^{\mathrm{T}} + \mathbf{x}_4\mathbf{y}_4^{\mathrm{T}} + \mathbf{x}_5\mathbf{y}_3^{\mathrm{T}} + \mathbf{E}_M \tag{5b}$$

To further clarify the equivalence of (4a,b) and (5a,b), it should be noted that if the ratio of the analyte present in **M** and **N** (i.e. γ) is known, then the absolute amount is also available because the amount of the analyte giving the response **N** is known. Note that in fact any numbers can be used for z_1 and z_3 as long as z_2 is such that $z_2/z_1 = \gamma$. Hence in the following the alternative model formulations (5a,b) will be used.

Several comments are appropriate. First, owing to the fact that the response matrix of a single analyte has pseudorank (i.e. the rank of the response matrix without noise) two, it is not possible to obtain from such a matrix the pure spectra and kinetic profiles of the two absorbing species. This is known as the curve resolution problem and has attracted considerable interest in the literature.[11] Secondly, although the second species being formed by the analyte and the interferent are equal, the kinetic profiles are different when being formed by the analyte or the interferent. This gives some selectivity. Thirdly, since the analyte and the interferent are very similar (owing to the membrane!), the difference in responses from the two is very small. This makes calibration difficult but not impossible.

The model can be reformulated[14] as the C3MFA model given in (3), with the frontal planes of the $I \times J \times 2$ array **X** taken to be **N** and **M** respectively, the component matrices defined as

$$\mathbf{A} \equiv (\mathbf{x}_2 \vdots \mathbf{x}_3 \vdots \mathbf{x}_4 \vdots \mathbf{x}_5), \qquad \mathbf{B} \equiv (\mathbf{y}_2 \vdots \mathbf{y}_3 \vdots \mathbf{y}_4), \qquad \mathbf{C} \equiv \begin{pmatrix} 1 & 0 \\ \gamma & 1 \end{pmatrix}$$

and the core **G** with frontal planes

$$\mathbf{G}_1 \equiv \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \mathbf{G}_2 \equiv \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

which is therefore completely specified. Thus the unknown parameters in this model can be estimated by fitting the above C3MFA model to the data. In fact, this C3MFA model cannot be fitted directly by the core-constrained 3MFA algorithms[12] because, in addition to the zero constraints on many core elements, a zero constraint is also imposed on element c_{12} of **C**. Moreover, two other elements in **C** and all non-zero elements in the core have also been specified and can hence be fixed to the specified values (although these values can also be obtained by rescalings of the solution). Therefore in Section 4 a general algorithm is proposed for fitting the C3MFA model with various kinds of constraints. First, however, in the next section we focus on a different issue related to C3MFA models: to what extent do these models lead to *unique* parameter estimates?

3. UNIQUENESS OF C3MFA MODELS

3.1. Some general results

When C3MFA methods are used for parameter estimation, it is very important that the estimated parameters are uniquely determinable. This is because if different parameter values lead to the same model fit, the data do not provide adequate information to pinpoint a useful estimate of the parameter at hand. To assess uniqueness, it is necessary to study the following problem. Suppose two sets of parameters lead to the same estimate $\hat{\mathbf{X}}$ for $\mathbf{X}$, i.e. suppose

$$\hat{\mathbf{X}}_k = \mathbf{A}\sum_{r=1}^{R} c_{kr}\mathbf{G}_r\mathbf{B}^{\mathrm{T}} = \tilde{\mathbf{A}}\sum_{r=1}^{R} \tilde{c}_{kr}\tilde{\mathbf{G}}_r\tilde{\mathbf{B}}^{\mathrm{T}}, \quad k = 1,\ldots,K. \tag{6}$$

Then a parameter is considered to be uniquely, determinable if its 'tilded' version is necessarily equal to its 'untilded' version. In the unconstrained model this is by no means the case. For instance, (6) is satisfied for $\tilde{\mathbf{A}} = \mathbf{AT}$ and $\tilde{\mathbf{G}}_r = \mathbf{T}^{-1}\mathbf{G}_r$, $r = 1,\ldots,R$, for any non-singular $\mathbf{T}$. However, certain patterns of zero constraints on $\mathbf{G}$ imply 'essential uniqueness' for the component matrices **A, B** and **C**.[15] Here 'essential uniqueness' means that **A, B** and **C** are unique up to scaling, reflection and sometimes permutation of the columns of **A, B** and **C**. Because this freedom in scale and permutation can be undone rather easily, the parameters are effectively unique and can be considered as uniquely determinable parameters.

In practice the constraints imposed in a C3MFA model usually do not have a pattern that belongs to the class for which uniqueness has been established.[15] Therefore a procedure is needed to study the uniqueness for the particular model at hand. In the Appendix some results are given to help assess uniqueness. These results will be used to establish partial uniqueness in the practical examples discussed in the present paper.

3.2. Partial uniqueness in model for second-order sensor calibration example

For the second-order sensor calibration example it can be proven (see Appendix, Section A2) that the model in (5) is partially unique. Specifically, for this model, γ, $\mathbf{x_2}$, $\mathbf{x_4}$ and $\mathbf{y_3}$ are uniquely determinable and $\mathbf{x_3}$, $\mathbf{x_5}$, $\mathbf{y_2}$ and $\mathbf{y_4}$ are not. Therefore we may expect reasonable estimates for the relative concentration parameter γ and for the temporal profiles of species 1 formed by the analyte ($\mathbf{x_2}$) and species 1 formed by the interferent ($\mathbf{x_4}$), as well as for the UV/VIS spectrum of species 2 formed by the analyte ($\mathbf{y_3}$). For the other profiles and spectra we do not expect to get reasonable estimates.

4. ALGORITHMS

4.1. General

Above we have discussed C3MFA models and their uniqueness properties. Little has been said about fitting such models to empirical data. In the current algorithms for C3MFA,[12,13] only certain elements of the core are constrained to zero, and in one of them,[13] also the component matrices **A, B** and **C** are constrained to be columnwise orthonormal. Here it will be described how models incorporating *additional* constraints (e.g. on certain specific *elements* of **A, B** and **C**, or non-negativity constraints) can be fitted. Moreover, attention will be paid to fitting such models in the presence of severe multicollinearity.

Both current C3MFA methods[12,13] are based on ALS algorithms. The ALS algorithm where **A, B** and **C** are unconstrained[12] will serve as our starting point here. In this algorithm the loss function

$$f(\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{G}) = \left\| \mathbf{x} - \sum_{p=1}^{P}\sum_{q=1}^{Q}\sum_{r=1}^{R} g_{pqr}(\mathbf{a}_p \otimes \mathbf{b}_q \otimes \mathbf{c}_r) \right\|^2 \tag{7}$$

is decreased by alternately updating the matrices **A, B** and **C** and the core **G**. The algorithm can be described schematically as follows.

1. Initialize **A, B, C** and **G.**
2. Compute $f_0 = f$(**A,B,C,G**).
3. Update **A.**
4. Update **B.**
5. Update **C.**
6. Update **G.**
7. Compute $f = f$(**A,B,C,G**).

If $f_0 - f > f_0\varepsilon$ (where ε is a prespecified small value), set $f_0 = f$ and return to step 2; else consider the algorithm converged.

In step 1 all parameter matrices are initialized, e.g. by taking random values for all unconstrained parameters or by taking the first P principal components of $(\mathbf{X}_1 \vdots \ldots \vdots \mathbf{X}_K)$ for **A** and analogous initializations for **B** and **C**. Steps 3–6 are the iterative steps for updating the parameter matrices. Each update is chosen such that it minimizes the function over one parameter set, while the other parameters are considered fixed. In this way it is guaranteed that the function value decreases monotonically. Because the function value is bounded below by zero, it is guaranteed that the algorithm will converge to a stable function value. Incidentally, it may be noted that the algorithm is easily adjusted for handling missing data:[20] before each step in which the function value is computed, the values of **x** that are missing are replaced by the estimates for these values on the basis of the current values for **A, B, C** and **G**.

The update for **A** (step 3) is found by minimizing

$$\begin{aligned} f(\mathbf{A},*,*,*) &= \sum_{k=1}^{K} \left\| \mathbf{X}_k - \mathbf{A}\sum_{r=1}^{R} c_{kr}\mathbf{G}_r\mathbf{B}^{\mathrm{T}} \right\|^2 \\ &= \left\| (\mathbf{X}_1 \vdots \ldots \vdots \mathbf{X}_K) - \mathbf{A}\left(\sum_r c_{1r}\mathbf{G}_r\mathbf{B}^{\mathrm{T}} \vdots \ldots \vdots \sum_r c_{Kr}\mathbf{G}_r\mathbf{B}^{\mathrm{T}} \right) \right\|^2 \end{aligned} \tag{8}$$

which is a series of linear regression problems for the rows of **A**. Defining the $\boldsymbol{I}\times\boldsymbol{JK}$ matrix $\mathbf{Y} \equiv (\mathbf{X}_1 \vdots \ldots \vdots \mathbf{X}_k)$ and the $\boldsymbol{P}\times\boldsymbol{JK}$ matrix $\mathbf{Z} \equiv (\Sigma_r c_{1r}\mathbf{G_r}\mathbf{B}^{\mathrm{T}} \vdots \ldots \vdots \Sigma_r c_{Kr}\mathbf{G_r}\mathbf{B}^{\mathrm{T}})$, the problem reduces to minimizing

$$\|\mathbf{y}_i^{\mathrm{T}} - \mathbf{a}_i^{\mathrm{T}}\mathbf{Z}\|^2 = \|\mathbf{y}_i - \mathbf{Z}^{\mathrm{T}}\mathbf{a}_i\|^2 \tag{9}$$

where $\mathbf{y}_i{}^{\mathrm{T}}$ and $\mathbf{a}_i{}^{\mathrm{T}}$ are the ith rows of **Y** and **A** respectively, $i = 1,\ldots,I$. The updates for **B** and **C** (in steps 4 and 5) are found in a fully analogous way. The update for **G** (in step 6) is found by minimizing

$$f(*,*,*,\mathbf{G}) = \|\mathbf{x} - (\mathbf{A}\otimes\mathbf{B}\otimes\mathbf{C})\mathbf{g}\|^2 \tag{10}$$

where **x** and **g** are vectors with elements of **X** and **G** respectively, ordered such that the third index runs fastest and the first index runs slowest. The solution for unconstrained **g** can be found from the regression of **x** on $\mathbf{F} \equiv \mathbf{A}\otimes\mathbf{B}\otimes\mathbf{C}$. When certain elements in **g** are constrained to be zero,[12,13] the

update for **g** is still obtained by means of regression of **x** on columns of **F**, but now this regression is performed only on those columns of **F** that correspond to the unconstrained elements of **g**. Specifically, let **w** denote a binary vector of the same order as **g** that has unit elements at the positions of the unconstrained elements in **g** and zeros elsewhere. Furthermore, let **F(w)** denote the submatrix of **F** containing the columns corresponding to the unit elements in **w** and let **g(w)** denote the vector with elements of **g** corresponding to the unit elements in **w**; hence **g(w)** contains the unconstrained elements of **g**. Then the problem is to minimize $||\mathbf{x} - \mathbf{F(w)g(w)}||^2$, and the **g(w)** solving this regression problem gives the updated values for the unconstrained elements of **g**. In the following subsections it will be described how the above general algorithm can handle a variety of other constraints in addition to the core constraints.

4.2. Zero constraints on A, B and/or C

Suppose certain elements of **A** are constrained to be zero. Then the update for $\mathbf{a_i}^T$, the *i*th row of **A**, can still be found by regressing $\mathbf{y_i}$ on columns of $\mathbf{Z}^T$, but now on those columns of $\mathbf{Z}^T$ that correspond to the unconstrained elements of $\mathbf{a_i}$. Specifically, when **w** is the indicator vector indicating (by unit elements) which elements of $\mathbf{a_i}$ are unconstrained and which are constrained, then the regression problem is to minimize $||\mathbf{y_i} - \mathbf{Z}^T\mathbf{(w)a_i(w)}||^2$, where the notation $\mathbf{Z}^T\mathbf{(w)}$ and $\mathbf{a_i(w)}$ is analogous to that of **F(w)** and **g(w)** respectively above. Of course, constrained updates of (the rows of) **B** and **C** can be found in a fully analogous way. In fact, the above approach is fully analogous to the approach to update **G** subject to zero constraints.

4.3. Non-zero constraints on A, B, C and/or G

In certain situations one wishes to constrain certain elements of **G** to be equal to a non-zero but prespecified value, e.g. one. To update the other elements of **G**, we can use basically the same approach as above. When **w** still indicates unconstrained elements and when we define **v** as the vector indicating constrained elements, we have to minimize

$$\|\mathbf{x} - \mathbf{F}(\mathbf{v})\mathbf{g}(\mathbf{v}) - \mathbf{F}(\mathbf{w})\mathbf{g}(\mathbf{w})\|^2 \tag{11}$$

Hence we have to solve the problem of regressing the columns of $\mathbf{x} - \mathbf{F(v)g(v)}$ on **F(w)**. The resulting vector of regression weights, **g(w)**, gives the updates for the unconstrained values of **g**. The procedures to update the rows of **A, B** and/or **C** subject to non-zero constraints can be derived analogously. It should be noted that as far as in different rows of a matrix the same elements are constrained to fixed values (e.g. when complete columns of a matrix are constrained to fixed values), these rows employ the same regression and regressand matrices, which hence need not be recomputed for every row.

4.4. Equality constraints on A, B, C, and/or G

Sometimes it is desirable to constrain certain (non-zero) parameters to be equal to each other. Such constraints have been found useful when applied to certain elements of the core, but may also be interesting for certain elements in **A, B** and/or **C**. When such constraints are imposed on elements of the core or on elements in the same row of one of the component matrices, the updates can be found straightforwardly as follows. Let $\mathbf{W_s}$ indicate a subset of elements of **g** that are constrained to be equal, $s = 1,\ldots,S$. Then the update for the S different subsets of elements of **g** (denoted by $g_1,\ldots,g_s$) can

be found by minimizing

$$\left\| \mathbf{x} - (\mathbf{F}(\mathbf{w_1})\mathbf{1} \vdots \ldots \vdots \mathbf{F}(\mathbf{w}_S)\mathbf{1}) \begin{pmatrix} g_1 \\ \vdots \\ g_S \end{pmatrix} \right\|^2 \tag{12}$$

over $g_1,\ldots,g_s$, which is again a linear regression problem; here **1** is a vector of unit elements. To give a little more insight into why the problem of updating **g** subject to equality constraints amounts to minimizing (12), consider the following simple example. Let $\mathbf{g} = (g_{111} \quad g_{211} \quad g_{121} \quad g_{221})^{\mathrm{T}}$ and denote the columns of **F** by $\mathbf{f_1},\ldots,\mathbf{f_4}$. Suppose that the constraints specify that $g_{111} = g_{121} \equiv g_1$ and $g_{211} = g_{221} \equiv g_2$. Then the problem is to minimize

$$\|\mathbf{x} - (\mathbf{f}_1 \vdots \mathbf{f}_2 \vdots \mathbf{f}_3 \vdots \mathbf{f}_4)(g_1 \quad g_2 \quad g_1 \quad g_2)^{\mathrm{T}}\|^2 = \|\mathbf{x} - (\mathbf{f}_1 + \mathbf{f}_3 \vdots \mathbf{f}_2 + \mathbf{f}_4)(g_1 \quad g_2)^{\mathrm{T}}\|^2$$

Clearly, this expression is in agreement with (12), because $\mathbf{w_1} = (1 \quad 0 \quad 1 \quad 0)^{\mathrm{T}}$ implies $\mathbf{F(w_1)} = (\mathbf{f_1} \vdots \mathbf{f_3})$ and $\mathbf{F(w_1)1} = \mathbf{f_1} + \mathbf{f_3}$, and similarly, $\mathbf{F(w_2)1} = \mathbf{f_2} + \mathbf{f_4}$.

Obviously, updates for rows of **A, B** or **C** subject to equality constraints can be found analogously to the above.

4.5. Non-negativity constraints on A, B, C and/or G

When the (unconstrained) elements of **A, B, C** and/or **G** are required to be non-negative, we have to solve the corresponding regression problems subject to the constraint that the regression weights are non-negative. For this we can use the non-negative least squares (NNLS) procedure.[21] In fact, it is not necessary that *all* elements of **A, B** or **C** are constrained to be non-negative. The NNLS procedure can be used for updating separate rows; hence it is possible to constrain only certain rows of a matrix to be non-negative, whereas others are left unconstrained.

4.6. Combination of constraints

The above four types of constraints on **A, B, C** and **G** can be imposed simultaneously as long as equality constraints do not pertain to parameters in different regression problems. This leaves a broad range of possible applications of combinations of constraints and thus makes our general algorithm very flexible.

4.7. A procedure for handling severe multicollinearity

An algorithm incorporating most of the possibilities above has been programmed in MATLAB 4.02.[22] When analysing the example data sets studied in the present paper, it turned out that the algorithm was very slow, and results of different randomly started runs tended to differ only slightly in terms of loss function value but considerably in terms of obtained parameter estimates. These problems can be attributed to the large size of the arrays under study and especially to severe multicollinearities in the data. This motivated the development of a procedure for dealing with these problems simultaneously.

To deal with the large size of the problem, we used a compression technique[23–25] which searches a low-dimensional approximation of the data and consequently fits a compressed version of the model to the compressed data. After this compressed fitting procedure the full data model can be recomputed from the fitted compressed model. For the practical examples studied here, we used compression only for the two large modes (A and B) and thus left the C-mode intact. Therefore in the analysis of the compressed data the constraints on **C** and **G** are the same as they are in the analysis of the full data.

We found that an effective procedure to deal with the severe multicollinearity in these data was to use a new kind of compression basis matrix, namely such that the resulting compressed array no longer showed high multicollinearity. In fact, we took bases that themselves captured the multicollinearity, and transferred collinearity from the compressed data to the basis matrices, thus 'regularizing' the data, as follows. For the mode for which the core had the largest size (r_1 say), we took as basis the matrix with the first r_1 principal components of the associated supermatrix, having sums of squares equal to the associated eigenvalues. The now obtained compressed array (compressed over one mode) was used to determine the next basis matrix. The next basis matrix was obtained for the mode with the second largest size (r_2) and consisted of the first r_2 principal components of the supermatrix associated with the current compressed array. In our examples the third mode had only two entries, so compression was not sensible. It turned out that analysing the thus compressed data led to reasonable estimates for the parameters of interest.

The above procedure worked reasonably well, but it turned out that better estimates were obtained with a procedure combining 'regularized' compression with ordinary compression (i.e. as described in Reference 23) as follows. From an analysis of the regularized data (possibly using a multistart procedure) we constructed a solution for the full data set by premultiplications with the associated basis matrices. From this solution we derived a start for a run of the analysis of data compressed in the original way.[23] From the solution obtained by this run, a new solution for the full data set was obtained (by premultiplication with the present basis matrices), which in turn was used as input for a final analysis of the complete data set. Thus we used a three-step procedure consisting of

(1) analysis of regularized compressed data
(2) analysis of ordinary compressed data
(3) analysis of full data.

This three-step procedure was a prototype for a procedure[26] developed, shortly after the present developments, for fitting the PARAFAC model. Because Reference 26 is fully devoted to this type of compression-based procedure for handling multicollinearity, the reader is referred to that paper for more details on the general idea of the three-step method proposed here.

5. ANALYSIS OF SECOND-ORDER SENSOR CALIBRATION DATA

To test the calibration procedure using the sensor described in Section 1.3, a mixture was prepared containing 0.976 ppm TCE and 0.995 ppm $CHCl_3$ (which is one of the possible interferents that can pass through the membrane). Three physically different standards were prepared and measured, each containing 0.488 ppm TCE. Experimental details have been given in Reference 16, the only difference being that the present data were measured at 10 °C whereas the data analysed in Reference 16 were measured at 20 °C. Hence the circumstances are not exactly comparable. The three different standards can be used to calculate the reproducibility of the measurements with the procedure mentioned in Reference 14. This results in a coefficient of variation (CV) or a relative standard deviation of 5.1% in the measured absorbances. This is a measure of the reliability of the measurements. For the final calibration of the sensor the average of the data for the three original standards was used as the standard (**N**).

Two different kinds of error are present in the data. The first error is due to imperfections in the experimental conditions. Although all measurements should be performed under equal conditions (e.g. temperature, start of data acquisition, etc.), small deviations are unavoidable. The resulting error will be called experimental error. The second kind of error is pure instrumental error. Even with perfect repeatability of the experimental conditions of the measurements, there will always exist differences between repeated measurements. Those differences are due to instrumental noise and will

be called instrumental error.

The influence of the combination of experimental error and instrumental error can be established by using the three original standards separately in three different calibrations. The differences between the estimated concentrations, expressed as a coefficient of variation, can be attributed to the aforementioned errors. Note that the thus obtained standard deviation gives an upper bound to the expected standard deviation of the estimated concentration of the final calibration, because in this final calibration the average of three standards is used and this filters out some error.

The influence of the pure instrumental error can be established by using a jack-knife procedure on the data set used for the final calibration. In ten steps several rows and columns of **N** and **M** are deleted from **N** and **M**, where in each step different rows and columns are deleted. This generates ten matrices **N** and **M** and, upon using the C3MFA model for each set of **N** and **M**, this generates ten different concentration estimates. The differences between these estimates (again expressed as a standard deviation) are due to instrumental noise. Note that the thus obtained standard deviation gives a lower bound to the standard deviation of the concentration estimates of the final calibration model, because instrumental error is always present in the data and will influence the quality of the concentration estimate.

To all data sets we fitted the C3MFA model, with the core elements fixed to one or zero, as indicated in Section 2.2. In **C**, only the value c_{21} (representing γ) was left unconstrained, the other elements being fixed to one or zero, as indicated in Section 2.2. Furthermore, because the value c_{21} represents a positive value, we deemed it useful to constrain it to be non-negative, as a means to steer the solution in the proper direction; because all other parameters also pertain to non-negative entities, they could, in principle, have been constrained in the same way, but the compression procedures cannot handle such constraints. The ensuing model was fitted to the data by means of the three-step procedure described in Section 4.7, using five randomly started runs for the first step (and choosing the best solution of these as input for the second step); in all steps a run of an algorithm was considered converged if the function value did not change by more than 0.0001%.

It is worth noting that the model and estimation procedure are severely tested: only one standard is used and, to predict the concentration of TCE in **M**, the model is used to extrapolate. This is an extrapolation because the unknown concentration lies outside the interval [0,0.488] defined by the two calibration points used here. In practice, multiple standards should be used to obtain more stable results and to avoid extrapolation.

The purpose of the calibration is twofold: quantitative analysis, i.e. estimating the relative amount of the analyte present in the mixture, and qualitative analysis, i.e. estimating the instrumental response of the unknown interferents, thereby enabling the researcher to identify that unknown compound. These issues will be discussed in two separate subsections.

5.1. Quantitative results

The analysis of the full data set led to an estimate of the concentration of TCE in the mixture of 0.997 (obtained by dividing the estimate for γ by 0.488). Compared with the true value of 0.976, this represents a relative error of only 2.2%. A similar figure of merit was reported for the, at first sight, very similar data analysis reported in Reference 27, but it should be noted that the experimental conditions for those data were essentially different from those for our data as well those for the data analysed in Reference 16. This embarrasses a thorough comparison of the results reported in the earlier papers with our results. We can, however, compare our results with that obtained by the conventional method for such analysis, RAFA.[28] With RAFA the estimate was 2.045, which is 110% in error.

To study the reliability of these results, we also analysed the three original data matrices. The three

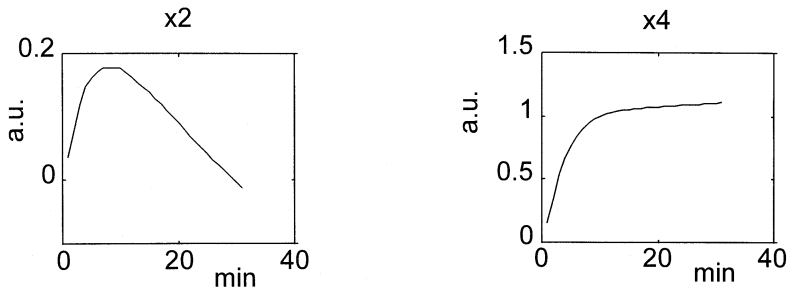

Figure 1. Estimated temporal profiles $\mathbf{x_2}$ and $\mathbf{x_4}$ of sensor example. The abbreviation 'min' stands for time in minutes and 'a.u.' stands for arbitrary units. The meaning of $\mathbf{x_2}$ and $\mathbf{x_4}$ is explained in the text.

concentration estimates were 0.923, 1.133 and 1.022 (average 1.026 and CV of 10.3%), which show some variation, but not more than could be expected on the basis of variation due to experimental error. As a second procedure to assess the reliability of these results, we did a small jack-knife study (see earlier) on the full data set. We created ten data sets from the original data set by leaving out parts of the data. Specifically, in each such jack-knife data set, data on 10% of the wavelengths and time points were left out using the Venetian blind method (leaving out 1,5,9,…; 2,6,10,…; etc.). The ten analyses led to an average concentration estimate of 0.995 and had a standard deviation of 0.006, thus indicating very little variation over the different jack-knives. We therefore conclude that the concentration estimate reported above is not only quite accurate but also quite reliable.

5.2. Qualitative results

Qualitative information is obtained in the calibration for the analyte and the interferent. It has been shown in the Appendix that $\mathbf{x_2}$, $\mathbf{x_4}$, $\mathbf{y_3}$ and the concentration of the analyte in $\mathbf{M}$ are uniquely determinable. The estimated spectra and profiles can serve as starting point for further analyses.

Figure 1 shows the estimated temporal profiles of $\mathbf{x_2}$ and $\mathbf{x_4}$. The estimated temporal profile of $\mathbf{x_2}$ is reasonable and shows that the analyte behaves according to the theory of reaction kinetics. The reaction kinetics dictate that the temporal profile of $\mathbf{x_4}$ should also go down during the reaction (like $\mathbf{x_2}$) since it is an intermediate product, but it is not known at what point in time. Hence, if the time window which is used to follow the reaction is too short, then no decrease can be observed.

The estimated spectrum for $\mathbf{y_3}$ (i.e. the UV/VIS spectrum of species 2 formed by the analyte) is shown in Figure 2. This spectrum resembles the one obtained in an earlier analysis,[17] but comparison

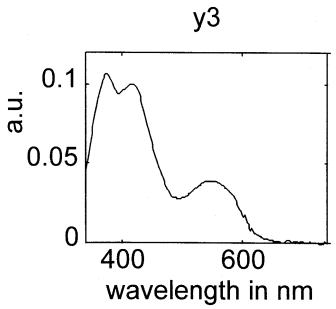

Figure 2. Estimated spectrum $\mathbf{y_3}$ of sensor example. The abbreviation 'nm' stands for nanometres and 'a.u' stands for arbitrary units. The meaning of $\mathbf{y_3}$ is explained in the text.

is difficult since the measuring conditions were slightly different.

Summarizing the qualitative results, it can be stated that the obtained results are in agreement with the theory. Only the behaviour of $\mathbf{x_4}$ is perhaps not completely according to the theoretical results, but the possible reason for this behaviour is explained above.

6. A DIFFERENT EXAMPLE: SECOND-ORDER CALIBRATION USING FIA–UV DATA

In this section an example is discussed which differs considerably from the previous example as far as the measurement set-up is concerned. It also seems that this calibration problem is even more difficult than the one discussed before. Since we could perform an auxiliary experiment, we were able to obtain the true underlying profiles and spectra. This makes this example ideally suited to test and illustrate the C3MFA method.

6.1. Introduction

This example pertains to a flow injection analysis (FIA) system coupled with UV diode array detection (250–450 nm). Experimental details are given elsewhere.[29] The system consists of a flow injection channel with a diode array detector at the outlet. The carrier stream is a buffer of pH 4.5. The sample S is caught in between the reagent plug (a buffer with pH 11.4) and the carrier stream. This generates a reproducible pH gradient over the sample plug. If the sample contains different analytes, selectivity is obtained owing to differences between pK_a values of these analytes. The pK_a value of a solute is that pH value at which half of the solute molecules are in the acidic form (protonated) and half in the basic form (deprotonated). The scanning starts 20 s after injection and continues for 88 s with a 1 s interval. At each scanning, absorbances at the indicated wavelengths are measured. Hence the result is an 89 time points by 100 wavelengths matrix of absorbances. No physical separation takes place.

In practice the analyte in the calibration sample is of course known, and it is also known to which class of chemical substances possible interferents in the mixture belong. Again, on the basis of the approximate rank of the mixture matrix, one assesses how many interferents there are. Here we consider the situation with one interferent. Specifically, we study the situation where the calibration sample is 3-hydroxybenzaldehyde (3-HBA) and the interferent is 2-hydroxybenzaldehyde (2-HBA) or a different substance in the class of possible interferents, which can all be assumed to have the same behaviour as 2-HBA. For convenience, we only consider the situation where this interferent actually *is* 2-HBA, but neither the calibration procedure nor the model would change if it were a different substance in this class.

The pK_a values of 2-HBA and 3-HBA are 8.37 and 8.98 respectively. Therefore both the acidic and basic forms of a solute will be present along the pH gradient. Hence the UV spectra of the acidic and basic forms of each solute produce the signal together with the concentration profiles of the acidic and basic forms. If the spectra of the acidic forms are called $\mathbf{sa_2}$ and $\mathbf{sa_3}$ for 2-HBA and 3-HBA respectively and $\mathbf{ca_2}$ and $\mathbf{ca_3}$ are the concentration profiles of the acidic forms of 2-HBA and 3-HBA, and if analogous definitions are used for the basic forms ($\mathbf{sb_2}$, $\mathbf{sb_3}$, $\mathbf{cb_2}$, $\mathbf{cb_3}$), then the responses of the pure solutes 2-HBA and 3-HBA can be written as

$$\mathbf{N_{2\text{-HBA}}} = z_2 \cdot \mathbf{ca_2} \cdot \mathbf{sa_2}^{\mathrm{T}} + z_2 \cdot \mathbf{cb_2} \cdot \mathbf{sb_2}^{\mathrm{T}} \tag{13a}$$

$$\mathbf{N_{3\text{-HBA}}} = z_3 \cdot \mathbf{ca_3} \cdot \mathbf{sa_3}^{\mathrm{T}} + z_3 \cdot \mathbf{cb_3} \cdot \mathbf{sb_3}^{\mathrm{T}} \tag{13b}$$

where z_2 and z_3 are the concentrations of 2-HBA and 3-HBA in the standards respectively.

During an analysis the total concentration of a solute (acidic and basic forms) at a given time and place in the FIA channel is $\mathbf{ca}+\mathbf{cb}$. This gives two total concentration profiles: $\mathbf{ctot_2}$ ($=\mathbf{ca_2}+\mathbf{cb_2}$) and

$\mathbf{ctot_3}$ (=$\mathbf{ca_3}$+$\mathbf{cb_3}$) for 2-HBA and 3-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 their diffusion behaviour is equal. Hence the shape of the total concentration profiles is equal: $\mathbf{ctot_2} = \alpha.\mathbf{ctot_3}$, where α is a constant. This phenomenon puts a restriction on the calibration problem which translates into a C3MFA model. Moreover, the mixtures made of 2-HBA and 3-HBA are not rank-additive.[30] This can be seen by adding (13a) and (13b) in certain amounts; if, for simplicity, amounts of 2-HBA and 3-HBA at unit concentration are added to form **M**, then the result is

$$\mathbf{M} = \mathbf{ca_2} \cdot \mathbf{sa_2}^{\mathrm{T}} + \mathbf{cb_2} \cdot \mathbf{sb_2}^{\mathrm{T}} + \mathbf{ca_3} \cdot \mathbf{sa_3}^{\mathrm{T}} + \mathbf{cb_3} \cdot \mathbf{sb_3}^{\mathrm{T}} = (\mathbf{ca_2}\vdots\mathbf{cb_2}\vdots\mathbf{ca_3}\vdots\mathbf{cb_3}) \begin{pmatrix} \mathbf{sa_2} \\ \mathbf{sb_2} \\ \mathbf{sa_3} \\ \mathbf{sb_3} \end{pmatrix} \tag{14}$$

with

$$\mathbf{ca_2} + \mathbf{cb_2} = \alpha(\mathbf{ca_3} + \mathbf{cb_3})$$

It can easily be seen that the rank of the response of an individual solute is two, whereas the rank of **M** is three and not four, owing to the restriction on the columns of **M**. Of course, all these rank considerations hold for the case without measurement noise. In practice, measurement noise is present and the above reasoning should be understood in terms of pseudorank.

To test our procedure, we used data from a mixture for which the concentrations were known.[29] Specifically, we analysed data for a mixture containing 0.05 mM of 3-HBA and 0.10 mM of the interferent (2-HBA). Again the actual analysis data set was obtained as the average of three original data sets. The three repeated standards can again be used to calculate the coefficient of variation due to experimental and instrumental error, as explained above for the sensor example. This results in a value of 1.1%. This value is lower than in the example of the sensor, indicating that the FIA experimental set-up is more stable than the set-up for the sensor, as expected. This value can serve as a threshold value for evaluating the predictive performance of the C3MFA models. In the original paper, RAFA was used for the calibration, which resulted in some cases in relative prediction errors of 60%–80%. The reasons for such high prediction error are discussed in Reference 30.

An auxiliary experiment was carried out in which the spectra of the solutes 2-HBA and 3-HBA were measured in a solution of pH 4.5 and a solution of pH 11.4. Hence the true spectra of the acidic and basic species of 2-HBA and 3-HBA were available. Using these spectra, a simple regression step of the standard **N** on these spectra gives the true temporal profiles of 2-HBA. This can be done in a similar way for 3-HBA. Hence, owing to this auxiliary experiment, we have the true spectra and temporal profiles of 2-HBA and 3-HBA, which is helpful in assessing the potential usefulness of the C3MFA model.

In order to illustrate the working of the C3MFA models, we will show the results of a calibration in which the standard contains only 3-HBA and the mixture contains 3-HBA and the (in this test case) known interferent 2-HBA. Note that this is a difficult calibration: there is only one analyte in the standard, while in the mixture there is an interferent which strongly resembles the analyte.

6.2. Mathematical formulation of calibration problem

In this calibration, 3-HBA is the standard and the mixture contains 3-HBA as an analyte together with

an unknown interferent (which can be 2-HBA or a similar substance). The formal calibration model is

$$\mathbf{N}_{3\text{-HBA}} = \mathbf{ca_3} \cdot \mathbf{sa_3}^{\mathrm{T}} + \mathbf{cb_3} \cdot \mathbf{sb_3}^{\mathrm{T}} \tag{15a}$$

$$\mathbf{M} = \gamma \cdot \mathbf{ca_3} \cdot \mathbf{sa_3}^{\mathrm{T}} + \gamma \cdot \mathbf{cb_3} \cdot \mathbf{sb_3}^{\mathrm{T}} + \mathbf{ca}_{\mathrm{u}} \cdot \mathbf{sa}_{\mathrm{u}}^{\mathrm{T}} + \mathbf{cb}_{\mathrm{u}} \cdot \mathbf{sb}_{\mathrm{u}}^{\mathrm{T}} \tag{15b}$$

where the unknown interferent is indicated with the subscript 'u' and the concentration of this interferent is absorbed in $\mathbf{ca}_{\mathrm{u}}$ and $\mathbf{cb}_{\mathrm{u}}$. Assuming that the total concentration profiles of the interferent and the analyte are proportional gives

$$\mathbf{ca}_{\mathrm{u}} + \mathbf{cb}_{\mathrm{u}} = \alpha(\mathbf{ca_3} + \mathbf{cb_3}) = \alpha \cdot \mathbf{ctot_3} \tag{16}$$

where α is an unknown constant and not of direct importance. The calibration problem of (15a,b) can be rewritten using $\mathbf{cb}_{\mathrm{u}}$=α.$\mathbf{ctot}_3$-$\mathbf{ca}_{\mathrm{u}}$ and $\mathbf{ca}_3$=$\mathbf{ctot}_3$-$\mathbf{cb}_3$. This results in

$$\mathbf{N}_{3\text{-HBA}} = \mathbf{ctot_3} \cdot \mathbf{sa_3}^{\mathrm{T}} - \mathbf{cb_3} \cdot \mathbf{sa_3}^{\mathrm{T}} + \mathbf{cb_3} \cdot \mathbf{sb_3}^{\mathrm{T}} \tag{17a}$$

$$\mathbf{M} = \gamma \cdot \mathbf{ctot_3} \cdot \mathbf{sa_3}^{\mathrm{T}} - \gamma \cdot \mathbf{cb_3} \cdot \mathbf{sa_3}^{\mathrm{T}} + \gamma \cdot \mathbf{cb_3} \cdot \mathbf{sb_3}^{\mathrm{T}} + \mathbf{ca}_{\mathrm{u}} \cdot \mathbf{sa}_{\mathrm{u}}^{\mathrm{T}} + \alpha \cdot \mathbf{ctot_3} \cdot \mathbf{sb}_{\mathrm{u}}^{\mathrm{T}} - \mathbf{ca}_{u} \cdot \mathbf{sb}_{\mathrm{u}}^{\mathrm{T}} \tag{17b}$$

where the parameter γ and the spectra and time profiles have to be estimated. If the matrices **A, B** and **C** are defined as

$$\mathbf{A} = (\mathbf{ctot_3} \vdots \mathbf{cb_3} \vdots \mathbf{ca}_{\mathrm{u}}), \qquad \mathbf{B} = (\mathbf{sa_3} \vdots \mathbf{sb_3} \vdots \mathbf{sa}_{\mathrm{u}} \vdots \mathbf{sb}_{\mathrm{u}}), \qquad \mathbf{C} = \begin{pmatrix} 0 & 1 \\ 1 & \gamma \end{pmatrix} \tag{18}$$

and the matrices $\mathbf{N}_{3\text{-HBA}}$ and **M** are stacked to form a three-way array **X** (which for the present data has order $89 \times 100 \times 2$), then calibration model (15a,b) can be written as a C3MFA model with loading matrices **A, B** and **C** as given in (3) and the $3 \times 4 \times 2$ core array **G** with elements g_{141}, g_{331}, g_{341}, g_{112}, g_{212} and g_{222} fixed to one. All other elements in **G** are forced to be zero. This specific pattern of zeros and ones can be deduced by looking carefully at the calibration problem (17) and keeping track of which triple products of elements of columns of **A, B** and **C** should be incorporated in the C3MFA model (see also equation (2)). As shown in the Section A3 of the Appendix, the present C3MFA model leads to unique estimates of γ and $\mathbf{ctot_3}$ and of the subspaces of $(\mathbf{sa_3} \vdots \mathbf{sb_3})$ and $(\mathbf{sa}_{\mathrm{u}} \vdots \mathbf{sb}_{\mathrm{u}})$.

6.3. Analyses

In the analyses of these data, as in the analyses of the second-order sensor data, the three-step approach (with five random starts for the first step) was employed and the value of γ in **C** is constrained to be non-negative. In addition to the full data set, we analysed the three original data sets and ten jack-knife data sets. These additional analyses were meant to gain insight into the reliability of our results for the full data set.

6.3.1. Quantitative results

The analysis of the full data set led to an estimate of the concentration of 3-HBA of 0.048 (upon dividing the obtained value for γ by 0.15). Compared with the true value of 0.05, this represents a relative error of 4.2%. In comparison, the estimate obtained by RAFA was 0.081 (62.6% relative error). It can be seen that our procedure leads to a considerable improvement over the RAFA procedure.

Our procedure also turned out to be reliable: the three original data sets led to concentration estimates of 0.046, 0.051 and 0.046 (average 0.048 and CV of 6%), which clearly varied only slightly. The jack-knives (constructed by leaving out observations on 5% of the wavelengths and time points)

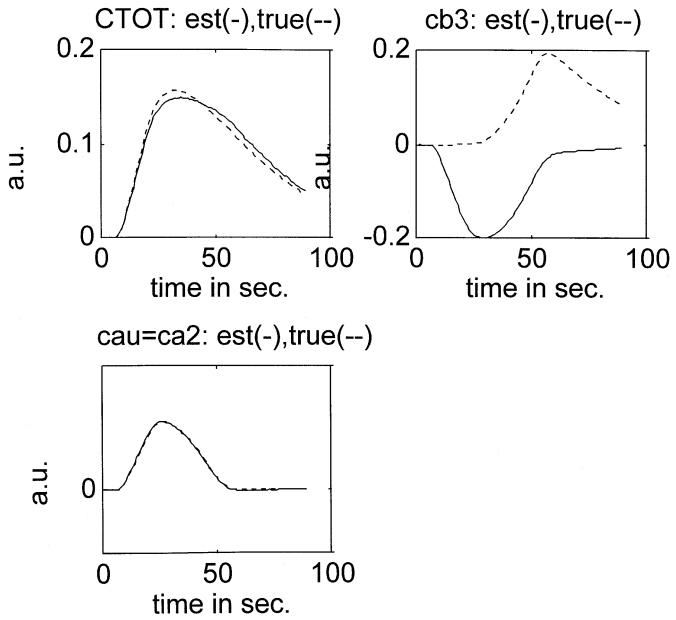

Figure 3. Concentration profiles of FIA–UV example: —, estimated profiles; - - -, true profiles. The abbreviation 'sec' stands for seconds and 'a.u.' stands for arbitrary units. The meaning of **ctot,** $\mathbf{cb_3}$ and $\mathbf{ca_u}$ is explained in the text.

led to an average value of 0.048 with a CV of 6%, again indicating very little variation due to changes in the data. It can be concluded that the concentration estimate is again both quite accurate and quite reliable.

6.3.2. Qualitative results of FIA calibration

The estimated concentration profiles (Figure 3) and spectra (Figure 4) as a result of the C3MFA model clearly demonstrate the consequences of the non-uniqueness of the model. Except for the **ctot** profile, none of the profiles and spectra can be estimated uniquely, and as a consequence, nonsensical values (e.g. negative absorbances) can be found for these parameters. Negative values could of course have been prevented by imposing non-negativity constraints, but even then results may not be unique; moreover, non-negativity constraints can only be used in analyses of the full rather than the compressed data, and our experience with such analyses is that results become very unreliable and start-dependent.

The estimated concentration profiles confirm our theoretical results: the first concentration profile (**ctot**) is recovered very well, as can be seen upon comparison with the true profile, established using the auxiliary experiment as explained earlier, which is also plotted. No other concentration profile can be obtained reliably, as is clearly illustrated for the profile $\mathbf{cb_3}$ which assumes negative values throughout; surprisingly, but coincidentally, profile $\mathbf{ca}_u$ is very close to the actual profile.

The results for the estimated spectra are less easily explained. The estimated spectrum for the basic species of the unknown interferent is surprisingly good. As was stated in the theory section, the C3MFA model should (only) be able to identify the subspace spanned by $\mathbf{sa_3}$ and $\mathbf{sb_3}$. This can be checked by a simple calculation regressing $\mathbf{S} = (\mathbf{sa_3} \vdots \mathbf{sb_3})$ on the first two columns of **B**. The residuals

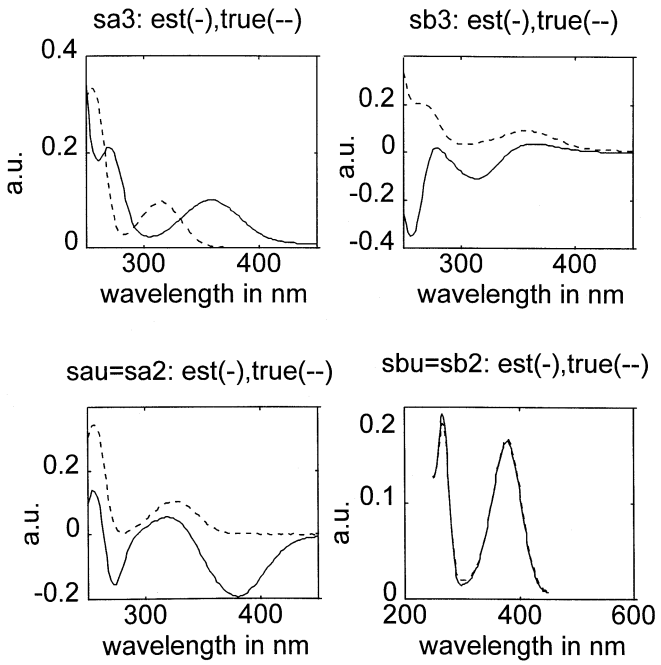

Figure 4. Spectra of FIA–UV example: —, estimated spectra; - - -, true spectra. The abbreviation 'nm' stands for nanometres and 'a.u'. stands for arbitrary units. The meaning of $\mathbf{sa_3}$, $\mathbf{sb_3}$, $\mathbf{sa_u}$ and $\mathbf{sb_u}$ is explained in the text.

of that regression are very small, indicating that the first two estimated spectra do indeed span the range of **S**. Using conventional curve resolution techniques, a rotation can be found of the first two columns of **B** in such a way that the resulting rotated spectra obey the non-negativity constraint and give better estimates of the true spectra. This topic is not pursued further.

In a similar fashion it can be shown that the last two columns of **B** give a good estimate of the subspace generated by $\mathbf{sa_2}$ (=$\mathbf{sa}_u$) and $\mathbf{sb_2}$ (=$\mathbf{sb}_u$). This is in agreement with the theoretical derivation. Hence, by applying again conventional curve resolution techniques, reasonable estimates for the spectra of the unknown interferent can be obtained. The transformations implied by such curve resolution techniques are to be compensated in a transformation of the profiles in matrix **A**, which will, however, only affect the second and third profiles.

The present exemplary analysis shows that the results of the calculations are in full agreement with the theory. Only one concentration profile can be obtained directly (**ctot**), whereas the other concentration profiles and spectra cannot be obtained directly. Yet, curve resolution techniques should be able to give better estimates of the latter. In addition, and most importantly, the concentration can be determined uniquely, and it is indeed found here that the estimated concentration is good and reliable.

7. A SMALL SIMULATION STUDY COMPARING OUR ALGORITHMS AND RAFA

To test our three-step algorithm and compare it with RAFA and two variants of our algorithm, we conducted a small simulation study. For this we created data resembling the FIA data (Section 6) as follows. The true spectra of the solutes 2-HBA and 3-HBA were obtained from the auxiliary experiment described in Section 6.1. These spectra and profiles were used to construct matrices $\mathbf{B}_0$ and $\mathbf{A}_0$ respectively. Together with $\mathbf{C}_0$ and **G** (defined in Section 6.2), we are able to construct data

that resemble the FIA data and exactly fit the C3MFA model. By adding some noise to such data, we created a realistic situation, with which we tested our algorithm and compared it with RAFA and other algorithms. As in Section 6, the concentration of 3-HBA was chosen to be 0.05, and in the construction of the data we used the matrix

$$\mathbf{C}_0 = \begin{pmatrix} 0 & 0.15 \\ 0.10 & 0.05 \end{pmatrix}$$

We constructed the 'perfect' three-way data array $\mathbf{X_0}$ from $\mathbf{A_0}$, $\mathbf{B_0}$, $\mathbf{C_0}$ and the core specified in Section 6.2 according to the 3MFA model. We constructed 30 data sets as $\mathbf{X} = \mathbf{X_0} + \mathbf{N_1} + \mathbf{N_2}$, where $\mathbf{N_1}$ and $\mathbf{N_2}$ represent three-way arrays containing different kinds of instrumental noise. The noise in $\mathbf{N_1}$ pertained to uniformly distributed random values with mean zero, the expected absolute value of which was 0.5% of the mean value of $\mathbf{X_0}$, so that this noise term can be considered as 0.5% uniform offset noise, a percentage which resembles that occurring in practice. The noise in $\mathbf{N_2}$ was taken *proportional* to the size of the values in $\mathbf{X_0}$, at three different levels. Specifically, this noise was chosen such that the expected absolute value of $(\mathbf{X} - \mathbf{X_0})/\mathbf{X_0}$ (where the solidus denotes elementwise division) was α, where α was taken equal to 0.002, 0.005 and 0.01 respectively, thus creating noise levels of 0.2%, 0.5% and 1% respectively. These amounts of proportional noise were considered reasonable amounts of measurement error. For each error level we constructed ten data sets.

The 30 data sets were analysed by RAFA and the three-step approach, as well as by two less complex variants: method 1—five randomly started analyses of the regularized compressed data, the best solution of which was used as input for a final analysis of the full data; method 2—five randomly started analyses of the ordinary compressed data, the best solution of which was used as input for a final analysis of the full data. All iterative algorithms used $\varepsilon = 10^{-6}$ as convergence criterion.

For all four methods we recorded the number of iterations required, the computation time, the number of local optima encountered in the initial analyses (with five runs) and, most importantly, the absolute deviation of the estimated concentration parameter from the real value (0.05).

The absolute deviations, averaged within each error level, are reported in Table 1. It can be seen that the three-step approach performs considerably better than RAFA (which in the last condition had relative errors of 24% on average). Method 1 was outperformed consistently by the three-step approach. The best method turned out to be method 2, closely followed by the three-step approach. As already anticipated in Section 4.7, the latter is considerably more efficient than method 2: the three-step approach used 99 s on average, whereas method 2 (based on the ordinary compression approach) required 2216 s on average. This difference can be explained upon inspecting the numbers of iterations required (which in our analyses was not allowed to exceed 20 000): the five runs in method 2 required 10 155 iterations each on average, while the final run of the full data required 182 iterations on average; the three-step approach required 24 iterations for each of the five initial runs, 1454 iterations for the second step and 40 iterations for the final analysis, all averaged over 30 replications. It was also seen that the five runs in method 2 very often led to widely differing values, of which at least 50% were clearly suboptimal; the five runs starting the three-step approach, on the other hand, consistently led to the same function value. Finally, we inspected the loss function values obtained in the final, full data, analyses by methods 1 and 2 and the three-step approach. It turned out that these values only differed in the sixth decimal place (which implies differences smaller than 0.15%); thus, as far as approximation of the actual data is concerned, no clear preference can be given to any of the methods.

It can be concluded from this small simulation study that method 2 and the three-step approach give the best concentration estimates. The three-step approach is much more efficient than method 2 and is much less sensitive to local optima, and for this reason can be expected to be more stable as well.

Table 1. Average absolute deviations of true concentration estimates

Error level (%)	RAFA	Three-step	Method 1	Method 2
0·02	0·0036	0·0013	0·0013	0·0016
0·05	0·0061	0·0024	0·0026	0·0018
1·0	0·0121	0·0026	0·0042	0·0017

8. GENERAL DISCUSSION

Clearly, the success of the estimation of concentrations depends on how well the model fits the data. If measurements are not very accurate, the parameter estimates will usually not be very accurate either. However, as indicated above, an even more important issue is the uniqueness of the parameter estimate in the model. If the parameter cannot be estimated uniquely, C3MFA will fail to give correct parameter estimates, even with perfectly accurate measurements. Therefore proving uniqueness is of paramount importance in parameter-estimating applications of C3MFA. The present paper illustrates how uniqueness can be proven in practice.

In the present paper a flexible algorithm for fitting a variety of C3MFA models has been proposed. Furthermore, for cases with severe multicollinearity it is proposed to use a three-step procedure consisting of applications of the algorithm to compressed versions of the data array. In the analyses reported in the present paper, the three-step approach turned out to be very efficient and stable (in the sense of sensitivity to local optima as well as in the sense of stability over repeated trial data); moreover, it performed well in terms of concentration estimates. In all cases the concentration estimates were much better than those obtained by the RAFA method. Thus it has been demonstrated that the C3MFA model, especially when fitted by the three-step approach, is a useful tool for obtaining concentration estimates when data on a mixture and one or more calibration sets are available.

ACKNOWLEDGEMENTS

This research has been made possible by a fellowship from the Royal Netherlands Academy of Arts and Sciences to the first author. The authors are obliged to Jos ten Berge for useful comments on an earlier version of this manuscript and to Lars Nørgaard and Carsten Ridder for generously making available the FIA data. We thank John Henshaw for performing the sensor experiments and Hans Meurs for performing the auxiliary FIA experiment.

APPENDIX

A1. Some results useful for assessing uniqueness

Result 1

If **A, B** and **C** have full column rank and the supermatrices formed by concatenating the frontal, lateral and horizontal planes of **G** respectively have full row rank (which implies at least $QR \geq P$, $PR \geq Q$ and $PQ \geq R$), then (6) holds if

$$\tilde{\mathbf{G}}_k = \mathbf{S} \sum_{r=1}^{R} u_{kr} \mathbf{G}_r \mathbf{T}^{\mathrm{T}}, \quad k = 1, \ldots, K \tag{19}$$

and $\mathbf{A} = \tilde{\mathbf{A}}\mathbf{S}$, $\mathbf{B} = \tilde{\mathbf{B}}\mathbf{T}$ and $\mathbf{C} = \tilde{\mathbf{C}}\mathbf{U}$ for certain non-singular matrices **S, T** and **U**.

Proof If (6) holds, we obtain, upon concatenating the frontal planes of $\hat{\mathbf{X}}$ into $\hat{\mathbf{X}}_\mathbf{F}$ and of $\mathbf{G}$ into $\mathbf{G}_\mathbf{F}$ that $\hat{\mathbf{X}}_\mathbf{F} = (\hat{\mathbf{X}}_\mathbf{1} \vdots \ldots \vdots \hat{\mathbf{X}}_\mathbf{K}) = \mathbf{A}\mathbf{G}_\mathbf{F}$ $(\mathbf{C}^\mathrm{T} \otimes \mathbf{B}^\mathrm{T}) = \tilde{\mathbf{A}}\tilde{\mathbf{G}}_\mathbf{F}(\tilde{\mathbf{C}}^\mathrm{T} \otimes \tilde{\mathbf{B}}^\mathrm{T})$. Here $\mathbf{C} \otimes \mathbf{B}$ has full column rank (QR) and $\mathbf{G}_\mathbf{F}$ has full row rank *(P)*. It follows that $\mathbf{A} = \hat{\mathbf{X}}_\mathbf{F}(\mathbf{C}^\tau \otimes \mathbf{B}^\tau)^+\mathbf{G}_\mathbf{F}{}^+$, where $^+$ denotes the Moore–Penrose inverse. Because $\mathbf{A}$ has rank P, $\hat{\mathbf{X}}_\mathbf{F}$ must have at least rank P and hence $\tilde{\mathbf{A}}$ must also have rank P. Because $\mathbf{A}$ and $\tilde{\mathbf{A}}$ span the same column space, it follows that $\mathbf{A} = \tilde{\mathbf{A}}\mathbf{S}$ for a certain non-singular matrix $\mathbf{S}$. In a completely analogous fashion we find that $\mathbf{B} = \tilde{\mathbf{B}}\mathbf{T}$ and $\mathbf{C} = \tilde{\mathbf{C}}\mathbf{U}$ for non-singular matrices $\mathbf{T}$ and $\mathbf{U}$. Hence it follows from (6) that

$$\tilde{\mathbf{A}} \sum_{r=1}^{R} \tilde{c}_{kr} \tilde{\mathbf{G}}_r \tilde{\mathbf{B}}^\mathrm{T} = \tilde{\mathbf{A}}\mathbf{S} \sum_{r=1}^{R} [\tilde{\mathbf{C}}\mathbf{U}]_{kr} \mathbf{G}_r \mathbf{T}^\mathrm{T} \tilde{\mathbf{B}}^\mathrm{T}, \quad k = 1, \ldots, K \tag{20}$$

It can be verified[15] that (20) is equivalent to (19). Conversely, suppose (19) holds and $\mathbf{A} = \tilde{\mathbf{A}}\mathbf{S}$, $\mathbf{B} = \tilde{\mathbf{B}}\mathbf{T}$ and $\mathbf{C} = \tilde{\mathbf{C}}\mathbf{U}$; then (6) follows trivially upon multiplication of the left- and right-hand sides of (19) by **A, B** and **C** (in the proper directions). □

Result 2

When the conditions of Result 1 are satisfied and a component matrix (**A, B** or **C**) is constrained such that a subset of rows forms a non-singular lower (upper) triangular matrix, then the corresponding transformation matrix is lower (upper) triangular.

Proof Suppose that a subset of rows of **A** (collected in $\mathbf{A}_\mathbf{s}$) is constrained such that $\mathbf{A}_\mathbf{s}$ and $\tilde{\mathbf{A}}_\mathbf{s}$ are non-singular and lower (upper) triangular. From Result 1 it follows that $\mathbf{A} = \tilde{\mathbf{A}}\mathbf{S}$ hence $\mathbf{A}_\mathbf{s} = \tilde{\mathbf{A}}_\mathbf{s}\mathbf{S}$. Therefore $\mathbf{S} = \tilde{\mathbf{A}}_\mathbf{s}^{-1}\mathbf{A}_\mathbf{s}$ which is lower (upper) triangular. □

A2. Proof for partial uniqueness in second-order sensor calibration example

To assess (partial) uniqueness for model (5) in Section 2.2 (under the assumption that **A, B** and **C** have full column rank), we first use Results 1 and 2 of Section A1. Hence we have $\tilde{\mathbf{G}}_\mathbf{k} = \mathbf{S}\Sigma_r u_{kr}\mathbf{G}_r\mathbf{T}^\mathrm{T}$, $k = 1,2$, and $\mathbf{A} = \tilde{\mathbf{A}}\mathbf{S}$, $\mathbf{B} = \tilde{\mathbf{B}}\mathbf{T}$ and $\mathbf{C} = \tilde{\mathbf{C}}\mathbf{U}$ with

$$\mathbf{U} = \begin{pmatrix} 1 & 0 \\ u_{21} & 1 \end{pmatrix}$$

Next we use

$$\tilde{\mathbf{G}}_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} = \mathbf{S} \begin{pmatrix} u_{21} & 0 & 0 \\ 0 & u_{21} & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \mathbf{T}^\mathrm{T} \tag{21}$$

from which it follows that

$$\begin{pmatrix} u_{21} & 0 & 0 \\ 0 & u_{21} & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

has rank two. This implies that $u_{21} = 0$, hence $\mathbf{U = I}$. Now we have $\tilde{\mathbf{G}}_\mathbf{1} = \mathbf{SG_1T}^{\mathrm{T}}$ and $\tilde{\mathbf{G}}_\mathbf{2} = \mathbf{SG_2T}^{\mathrm{T}}$ hence

$$\tilde{\mathbf{G}}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \mathbf{s}_1\mathbf{t}_1{}^{\mathrm{T}} + \mathbf{s}_2\mathbf{t}_2{}^{\mathrm{T}} \tag{22}$$

$$\tilde{\mathbf{G}}_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} = \mathbf{s}_3\mathbf{t}_3{}^{\mathrm{T}} + \mathbf{s}_4\mathbf{t}_2{}^{\mathrm{T}} \tag{23}$$

From (22) and the non-singularity of **S** and **T** it follows that the last two elements of $\mathbf{s_1}$ and $\mathbf{s_2}$ and the third elements of $\mathbf{t_1}$ and $\mathbf{t_2}$ are zero. From (23) it follows that the first two elements of $\mathbf{s_3}$ and $\mathbf{s_4}$ and the first elements of $\mathbf{t_2}$ and $\mathbf{t_3}$ are zero. Furthermore, for the element (2,1) in (22) we have $0 = s_{21}t_{11} + s_{22}t_{12} = s_{21}t_{11}$. Because $t_{11} = 0$ would imply that the first row of **T** is zero and hence **T** would be singular (violating the assumption that **B** has full column rank), it follows that $s_{21} = 0$. Similarly, from (23) it follows for element (4,3) that $s_{43}t_{33} = 0$; hence, because $t_{33} = 0$ would imply that **T** would be singular, it follows that $s_{43} = 0$. To sum up, we have

$$\mathbf{S} = \begin{pmatrix} s_{11} & s_{12} & 0 & 0 \\ 0 & s_{22} & 0 & 0 \\ 0 & 0 & s_{33} & s_{34} \\ 0 & 0 & 0 & s_{44} \end{pmatrix}, \qquad \mathbf{T} = \begin{pmatrix} t_{11} & 0 & 0 \\ t_{21} & t_{22} & t_{23} \\ 0 & 0 & t_{33} \end{pmatrix}, \qquad \mathbf{U = I} \tag{24}$$

To see if we have used all the information available, we elaborate the right-hand sides of the expressions for $\mathbf{G_k = S}\Sigma_r u_{kr}\mathbf{G_rT}^{\tau}$, $k = 1,2$, in terms of the present expressions for **S, T** and **U**. Thus we find

$$\mathbf{G}_1 = \begin{pmatrix} s_{11} & s_{12} & 0 & 0 \\ 0 & s_{22} & 0 & 0 \\ 0 & 0 & s_{33} & s_{34} \\ 0 & 0 & 0 & s_{44} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} t_{11} & t_{21} & 0 \\ 0 & t_{22} & 0 \\ 0 & t_{23} & t_{33} \end{pmatrix} = \begin{pmatrix} s_{11}t_{11} & s_{11}t_{21} + s_{12}t_{22} & 0 \\ 0 & s_{22}t_{22} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\mathbf{G}_2 = \begin{pmatrix} s_{11} & s_{12} & 0 & 0 \\ 0 & s_{22} & 0 & 0 \\ 0 & 0 & s_{33} & s_{34} \\ 0 & 0 & 0 & s_{44} \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} t_{11} & t_{21} & 0 \\ 0 & t_{22} & 0 \\ 0 & t_{23} & t_{33} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & s_{34}t_{22} + s_{33}t_{23} & s_{33}t_{33} \\ 0 & s_{44}t_{22} & 0 \end{pmatrix}$$

Because $\mathbf{G_1}$ and $\mathbf{G_2}$ are completely specified, it follows that $s_{11}t_{11} = s_{22}t_{22} = s_{33}t_{33} = s_{44}t_{22} = 1$, $s_{11}t_{21}+s_{12}t_{22} = 0$ and $s_{34}t_{22}+s_{33}t_{23} = 0$. These restrictions can be satisfied for arbitrary choices of the elements of **T** involved (provided that **T** is non-singular), as long as the elements of **S** are chosen such that we have $s_{11} = {t_{11}}^{-1}$, $s_{22} = s_{44} = {t_{22}}^{-1}$, $s_{33} = {t_{33}}^{-1}$, $s_{12} = -{t_{11}}^{-1}{t_{22}}^{-1}t_{21}$ and $s_{34} = -{t_{22}}^{-1}{t_{33}}^{-1}t_{23}$. Hence $\mathbf{G_k = S}\Sigma_r u_{kr}\mathbf{G}_r\mathbf{T}^{\tau}$, $k = 1,2$, is satisfied for matrices **S, T** and **U** of the form specified in (24), where the elements of **S** are expressed in the elements of **T** by the above relations. It follows that the first and third columns of **A** and the second column of **B** are uniquely determinable (up to scaling and reflection), whereas the other columns of **A** and **B** are not uniquely determinable. Thus two of the x-profiles ($\mathbf{x_2}$ and $\mathbf{x_4}$) and one y-profile ($\mathbf{y_3}$) can be identified. In addition, it follows from $\mathbf{U = I}$ that **C** is uniquely determinable, hence the relative concentration value γ is uniquely determinable.

A3. Proof for partial uniqueness in FIA–UV example

The C3MFA model employed in the example in Section 6 involves a constrained core array **G** which can be written as

$$\mathbf{G}_1 = \tilde{\mathbf{G}}_1 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}, \qquad \mathbf{G}_2 = \tilde{\mathbf{G}}_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Furthermore, the matrix **C** is constrained as

$$\mathbf{C} = \begin{pmatrix} 0 & 1 \\ 1 & \gamma \end{pmatrix}$$

To prove partial uniqueness for this model (under the assumption that **A, B** and **C** have full column rank), we first use Results 1 and 2 again, giving $\tilde{\mathbf{G}}_\mathbf{k} = \mathbf{S}\Sigma_r u_{kr}\mathbf{G}_r\mathbf{T}^T$, $k = 1,2$, and $\mathbf{A}=\tilde{\mathbf{A}}\mathbf{S}$, $\mathbf{B}=\tilde{\mathbf{B}}\mathbf{T}$ and $\mathbf{C}=\tilde{\mathbf{C}}\mathbf{U}$ with

$$\mathbf{U} = \begin{pmatrix} 1 & u_{12} \\ 0 & 1 \end{pmatrix}$$

Next we use

$$\mathbf{G}_1 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} = \mathbf{S}\begin{pmatrix} u_{12} & 0 & 0 & 1 \\ u_{12} & u_{12} & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}\mathbf{T}^T \tag{25}$$

from which it follows that

$$\begin{pmatrix} u_{12} & 0 & 0 & 1 \\ u_{12} & u_{12} & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

has rank two. This implies that $u_{12} = 0$, hence $\mathbf{U} = \mathbf{I}$.

Using $\mathbf{U} = \mathbf{I}$, from $\tilde{\mathbf{G}}_\mathbf{k} = \mathbf{S}\Sigma_r u_{kr}\mathbf{G}_\mathbf{r}\mathbf{T}^T$ we have $\tilde{\mathbf{G}}_\mathbf{1} = \mathbf{S}\mathbf{G}_\mathbf{1}\mathbf{T}^T$ and $\tilde{\mathbf{G}}_\mathbf{2} = \mathbf{S}\mathbf{G}_\mathbf{2}\mathbf{T}^T$, hence

$$\mathbf{G}_1 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} = \mathbf{s}_1\mathbf{t}_4{}^T + \mathbf{s}_3\mathbf{t}_3{}^T + \mathbf{s}_3\mathbf{t}_4{}^T \tag{26}$$

$$\mathbf{G}_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \mathbf{s}_1\mathbf{t}_1{}^T + \mathbf{s}_2\mathbf{t}_1{}^T + \mathbf{s}_2\mathbf{t}_2^T \tag{27}$$

From (26) and the non-singularity of **S** and **T** it follows that the first two elements of $\mathbf{t}_3$ and $\mathbf{t}_4$ and the second elements of $\mathbf{s}_1$ and $\mathbf{s}_3$ are zero. From (27) it follows that the last two elements of $\mathbf{t}_1$ and $\mathbf{t}_2$ and the third elements of $\mathbf{s}_1$ and $\mathbf{s}_2$ are zero. Thus, we have

$$\mathbf{S} = \begin{pmatrix} s_{11} & s_{12} & s_{13} \\ 0 & s_{22} & 0 \\ 0 & 0 & s_{33} \end{pmatrix}, \qquad \mathbf{T} = \begin{pmatrix} t_{11} & t_{12} & 0 & 0 \\ t_{21} & t_{22} & 0 & 0 \\ 0 & 0 & t_{33} & t_{34} \\ 0 & 0 & t_{43} & t_{44} \end{pmatrix}$$

It can be shown that the zeros in **S** and **T** above are the only elements that can always be taken zero. It can be concluded that **C** is uniquely determinable (since $\tilde{\mathbf{C}} = \mathbf{C}$) and that the first column of **A** is

identified uniquely, since $\mathbf{A}$ and $\tilde{\mathbf{A}}$ are related by $\mathbf{A} = \tilde{\mathbf{A}}\mathbf{S}$ and hence their first columns are proportional. Furthermore, the subspaces of the first two columns and of the last two columns of $\mathbf{B}$ are uniquely identified, as follows from $\mathbf{B} = \tilde{\mathbf{B}}\mathbf{T}$. The second and third columns of $\mathbf{A}$ are not uniquely determinable, since they are confounded with the first column of $\mathbf{A}$.

REFERENCES

1. P. Geladi, *Chemometrics Intell. Lab. Syst.* **7**, 11 (1989).
2. A. K. Smilde, *Chemometrics Intell. Lab. Syst.* **15**, 143 (1992).
3. E. Sanchez and B. R. Kowalski, *Anal. Chem.* **58**, 496 (1986).
4. S. Leurgans and R. T. Ross, *Statist. Sci.* **7**, 289 (1992).
5. P. Geladi, H. Isaksson, L. Lindqvist, S. Wold and K. Esbensen, *Chemometrics Intell. Lab. Syst.* **5**, 209 (1989).
6. P. Nomikos and J. F. MacGregor, *Technometrics*, **37**, 41 (1995).
7. K. A. Kosanovich, K. S. Dahl and M. J. Piovoso, *Ind. Engng. Chem. Res.* **35**, 138 (1996).
8. L. R. Tucker, *Psychometrika*, **31**, 279 (1966).
9. P. M. Kroonenberg and J. de Leeuw, *Psychometrika*, **45**, 69 (1980).
10. W. H. Lawton and E. A. Sylvestre, *Technometrics*, **13**, 617 (1971).
11. R. Tauler, A. K. Smilde and B. R. Kowalski, *J. Chemometrics*, **9**, 31 (1995).
12. H. A. L. Kiers, *Statist. Appl.* **4**, 659 (1992).
13. R. Rocci, *J. Italian Statist. Soc.* **3**, 413 (1992).
14. A. K. Smilde, R. Tauler, J. M. Henshaw, L. W. Burgess and B. R. Kowalski, *Anal. Chem.* **66**, 3345 (1994).
15. H. A. L. Kiers, J. M. F. ten Berge and R. Rocci, *Psychometrika*, **62**, 349 (1997).
16. J. M. Henshaw, L. W. Burgess, K. S. Booksh and B. R. Kowalski, *Anal. Chem.* **66**, 3328 (1994).
17. R. Tauler, A. K. Smilde, J. M. Henshaw, L. W. Burgess and B. R. Kowalski, *Anal. Chem.* **66**, 3337 (1994).
18. G. A. Lugg, *Anal. Chem.* **38**, 1532 (1966).
19. J. F. Reith, W. C. van Ditmarsch and Th. de Ruiter, *Analyst*, **99**, 652 (1974).
20. J. Weesie and H. van Houwelingen, *GEPCAM Users' Manual: Generalized Principal Components Analysis with Missing Values*, Institute of Mathematical Statistics, University of Utrecht (1983).
21. C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*, Prentice-Hall, Englewood Cliffs, NJ (1974).
22. *MATLAB*, Mathworks Inc., Natick, MA (1992).
23. C. J. Appellof and E. R. Davidson, *Anal. Chem.* **53**, 2053 (1981).
24. B. K. Alsberg and O. M. Kvalheim, *Chemometrics Intell. Lab. Syst.* **24**, 43 (1994).
25. H. A. L. Kiers and R. A. Harshman, *Chemometrics Intell. Lab. Syst.* **36**, 31 (1997).
26. H. A. L. Kiers, *J. Chemometrics*, in press.
27. K. S. Booksh, J. M. Henshaw, L. W. Burgess and B. R. Kowalski, *J. Chemometrics*, **9**, 263 (1995).
28. C. -N. Ho, G. D. Christian and E. R. Davidson, *Anal. Chem.* **50**, 1108 (1978).
29. L. Nørgaard and C. Ridder, *Chemometrics Intell. Lab. Syst.* **23**, 107 (1994).
30. H. A. L. Kiers and A. K. Smilde, *J. Chemometrics*, **9**, 179 (1995).

