*Anal. Chem.* **1994**, *66*, 3345–3351

# Multicomponent Determination of Chlorinated Hydrocarbons Using a Reaction-Based Chemical Sensor. 3. Medium-Rank Second-Order Calibration with Restricted Tucker Models

**A. K. Smilde,[†] R. Tauler,[‡] J. M. Henshaw,[§] L. W. Burgess, and B. R. Kowalski***

*Center for Process Analytical Chemistry, University of Washington, BG-10, Seattle, Washington 98195*

**The calibration of a chemical sensor for chlorinated hydrocarbon analytes based on the Fujiwara reaction is described. This sensor generates a particular type of data: medium-rank second-order data. With this type of data it is possible to calibrate the sensor in such a way that quantitation for the analytes in the presence of unknown interferents is possible. The calibration method developed is a new approach based on so-called restricted Tucker models that utilize all available chemical information.**

In order to be able to monitor in situ multiple chlorinated hydrocarbons in contaminated soil, different approaches are available.[1] One of the approaches is based on the Fujiwara reaction of chlorinated hydrocarbons with pyridine and a base.[2] In this reaction, intermediates and final products are formed which can be detected with UV/visible spectroscopy.[3] One way to use this reaction is to measure the absorbance at a particular wavelength of the species being formed after the analyte permeates a membrane of a chemical sensor and encounters the Fujiwara reagent. Since the measurement is done at one wavelength, this is a zero-order measurement.[4] Therefore, calibration for the chlorinated hydrocarbons using this chemical sensor is called zero-order calibration. The disadvantage of using zero-order calibration is that the selectivity of such a calibration is low. There are usually more absorbing species at the particular wavelength and these species may be present in unknown concentrations in unknown samples. This puts a severe constraint on the applicability of zero-order calibration: an unknown interference cannot be detected and accounted for. This is called the background problem.

If the measurements are not performed at one particular wavelength but at multiple wavelengths simultaneously, the result is a vector of measurements acquired at a certain point in time. Calibration with a chemical sensor based on multiple wavelength absorbances at a certain point in time is called first-order calibration. For the chlorinated hydrocarbons chloroform, trichloroethylene, and 1,1,1-trichloroethane, calibration based on first-order measurements with a chemical sensor is described in the first part of this series.[5] Although less severe, the background problem is still present: an unknown interference can be detected but calibration for the analyte of interest is impossible.

The approach taken in this paper and in the second part of this series[6] uses not only the full spectrum at a certain time point but also temporal information of the successive UV/visible spectra. At regularly spaced time intervals, a UV/visible spectrum is collected. At each point in time such a spectrum has contributions from a mixture of the intermediate and final species. The data obtained from measuring one sample (containing a mixture of analytes of interest and unknown interferences) can be ordered in a matrix with dimensions equal to the number of time points by the number of wavelengths. This type of data is called second-order data.[7] Under certain conditions, calibration with second-order data yields the second-order advantage: the ability to quantify for analytes of interest in the presence of unknown interferents. Of course, it is possible to obtain first-order data using this matrix of measurements by stringing out the matrix in one long vector. The first-order calibration methods—like partial least squares—can then be used. This is also shown in the first part of this series. This stringing out and using the data in a first-order way however destroys the second-order advantage.

The analyte of interest enters the reaction chamber of the chemical sensor through a membrane. The Fujiwara reagent is added, the reaction starts, and intermediate and final species are formed. The development or vanishing in time of a particular intermediate or final species will be referred to as the time profile of that species. This particular species has a UV/visible spectrum which will be referred to as the species spectrum. The time profiles and the spectra of the species are not assumed to be known a priori. Estimates of these time profiles and spectra can be obtained from multivariate curve resolution methods.[8] This is shown in the second part of this series.[6]

In order to solve the background problem, a full second-order method should be used.[7] In case of response matrices

[†] Present address: Laboratory for Analytical Chemistry, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands.

[‡] Present address: Department of Analytical Chemistry, University of Barcelona, Diagonal 647, Barcelona 08028, Spain.

[§] Present address: ORS Environmental Systems, 7 Barnabas Rd., Marion, MA 02738.

(1) Liebman, K. C.; Hindman, J. D. *Anal. Chem.* **1964**, *36*, 348.
(2) Fujiwara, K. *Sitzungsber. Abh. Naturforsch Ges. Rostock* **1916**, *6*, 33.
(3) Reith, J. F.; van Ditsmarsch, W. C.; de Reuter, T. *Analyst* **1974**, *99*, 652.
(4) Sanchez, E.; Kowalski, B. R. *J. Chemom.* **1988**, *2*, 247–264.
(5) Henshaw, J. M.; Burgess, L. W.; Booksh, K. S.; Kowalski, B. R., *Anal. Chem.*, first paper of three in this issue.
(6) Tauler, R.; Smilde, A. K.; Henshaw, J. M.; Burgess, L. W.; Kowalski, B. R. *Anal. Chem.* second paper of three in this issue.
(7) Sanchez, E.; Kowalski, B. R. *J. Chemom.* **1988**, *2*, 265–283.
(8) Tauler, R.; Kowalski, B. R.; Fleming, S. *Anal. Chem.* **1993**, *65*, 2040.

0003-2700/94/0366-3345$04.50/0

generated by a pure analyte with rank one, the generalized rank annihilation method can be used.[7] This method is able to quantitate for specific analytes in the presence of unknown interferences.

For pure analyte responses of rank higher than one (caused by multiple absorbing species) the calibration becomes more complicated. The challenge is, of course, to retain the second-order advantage. Recently, a second-order calibration method suited for medium-rank cases was developed.[9] The calibration of the reagent-based chemical sensor fits exactly in the framework of medium-rank second-order calibration. This medium-rank second-order calibration method is based on restricted Tucker models.

The first paper in this series deals with the description of the sensor and zero- and first-order calibration.[5] The second paper describes the curve resolution of the response matrices, i.e., the estimation of the underlying spectra and time profiles.[6] The second paper also describes details concerning the kinetics of the applied reactions and a calibration procedure based on the full-response matrices. This third paper describes the results of applying medium-rank second-order calibration to the Fujiwara reaction-based data. A short description of the theory will be given.

## THEORY

**Medium-Rank Second-Order Calibration with Restricted Tucker Models.** In the following, italic lower and upper case characters refer to scalars, bold lower case characters refer to vectors, bold uppercase characters refer to matrices and bold underlined uppercase characters refer to three-way matrices (also called third-order tensor). (Note: scalars, vectors, and matrices designated in this third paper of the series do not directly correspond to those of the same name in the first and second parts of the series.)

A short description of the theory will be given as more detailed information is given elsewhere.[9] One calibration example is used as an illustration in order to explain the theory. Suppose that the analyte 1,1,1-trichloroethane (TCA) generates one unique absorbing species, the analyte trichlororethylene (TCE) generates two absorbing species, and the analyte chloroform ($CHCl_3$) also generates two absorbing species. The number of species generated by the analytes can be established with rank analysis of the response matrices. Principal component analysis or curve resolution can be used for this purpose. Assume that the second species being formed by TCE and $CHCl_3$ has the same spectrum; i.e., it is the same species but with a different time profile for analytes TCE and $CHCl_3$. The calibration problem can be stated as follows:

$$\mathbf{N}_1 = z_{11}\mathbf{x}_1\mathbf{y}_1^t + \mathbf{E}_1 \quad (1a)$$

$$\mathbf{N}_2 = z_{22}\mathbf{x}_2\mathbf{y}_2^t + z_{22}\mathbf{x}_3\mathbf{y}_3^t + \mathbf{E}_2 \quad (1b)$$

$$\mathbf{N}_3 = z_{33}\mathbf{x}_4\mathbf{y}_4^t + z_{33}\mathbf{x}_5\mathbf{y}_5^t + \mathbf{E}_3 \quad (1c)$$

$$\mathbf{M} = z_{41}\mathbf{N}_1^* + z_{42}\mathbf{N}_2^* + z_{43}\mathbf{N}_3^* + \mathbf{U}_m + \mathbf{E}_m \quad (1d)$$

where $z_{lk}$ represents the true concentration of analyte $k$ in mixture $l$. The $\mathbf{x}$ vectors are the true time profiles of the species, the $\mathbf{y}$ vectors the true spectra of the species, $\mathbf{U}_m$ is the interferent in the mixture sample $\mathbf{M}$, and the $\mathbf{E}$ matrices are error matrices. The rank of $\mathbf{U}_m$ can vary between 0 (no interferent) and any positive number. From now on $\mathbf{N}_1$ refers to TCA, $\mathbf{N}_2$ refers to TCE, and $\mathbf{N}_3$ refers to $CHCl_3$. $\mathbf{N}_k^*$ is the unit concentration response of analyte $k$ = 1, 2, 3 (TCA, TCE, and $CHCl_3$, respectively); i.e., $\mathbf{N}_k^* = \mathbf{N}_k/z_{kk}$. With NBRA[10] the presence of $\mathbf{N}_1$, $\mathbf{N}_2$, and $\mathbf{N}_3$ in $\mathbf{M}$ can be established. The rank of $\mathbf{U}_m$ can be established by analyzing the rank of $\mathbf{M}$, using the information of the presence of the analyte, and with knowledge about the level of the measurement error. It is assumed that the unit time profiles and spectra of the species are equal in the pure samples and in the mixtures. It is important to stress the fact that the time scale of the measurements represented in $\mathbf{N}_1$, $\mathbf{N}_2$, $\mathbf{N}_3$, and $\mathbf{M}$ is the same; i.e. the recording of the spectra starts at time zero and evolves at the same points in time for each analyte and mixture.

The quantitation of TCA (=$\mathbf{N}_1$) can be found with the generalized annihilation method (GRAM[7]) since the rank of $\mathbf{N}_1$ (eliminating the noise) is one. Hence, the contribution of $\mathbf{N}_1$ can be substracted from $\mathbf{M}$.

For illustrative purposes it is assumed that there is no interferent present. Examples of other calibration models which take care of interferents will be discussed later in this paper.

Assuming absence of TCA and no interferent in M the calibration problem is

$$\mathbf{N}_2 = z_{12}\mathbf{x}_2\mathbf{y}_2^t + z_{12}\mathbf{x}_3\mathbf{y}_3^t + \mathbf{E}_2 \quad (2a)$$

$$\mathbf{N}_3 = z_{23}\mathbf{x}_4\mathbf{y}_4^t + z_{23}\mathbf{x}_5\mathbf{y}_5^t + \mathbf{E}_3 \quad \text{with} \quad \mathbf{y}_5 = \mathbf{y}_3 \quad (2b)$$

$$\mathbf{M} = z_{32}\mathbf{N}_2^* + z_{33}\mathbf{N}_3^* + \mathbf{E}_m \quad (2c)$$

where the values of $z_{32}$ and $z_{33}$ are desired (quantitation) together with estimations of the $x$ and $y$ vectors (qualitative information). Calibration proceeds by stacking $\mathbf{N}_2$, $\mathbf{N}_3$, and $\mathbf{M}$ to form the three-way matrix $\underline{\mathbf{D}}$ ($I \times J \times 3$), where $I$ and $J$ are the dimensions of $\mathbf{N}_2$, $\mathbf{N}_3$, and $\mathbf{M}$. Let $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ be the matrices consisting of the $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ vectors. The estimated counterparts of these matrices will be called $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$, respectively.

Two different models for analyzing three-way data are useful in this calibration. The first model is the PARAFAC or trilinear model.[11,12] An element $d_{ijk}$ of $\underline{\mathbf{D}}$ is modeled as

$$d_{ijk} = \sum_{q=1}^{Q} a_{iq}b_{jq}c_{kq} + e_{ijk} \quad (3)$$

where $Q$ is chosen in such a way that $\underline{\mathbf{E}}$ with elements $e_{ijk}$ has a small norm (i.e., $\sum e_{ijk}^2$ is small). Define the vectors $\mathbf{a}_q$, $\mathbf{b}_q$, and $\mathbf{c}_q$ as vectors with elements $a_{iq}$, $b_{iq}$, and $c_{iq}$, respectively. The outer tensor product of $\mathbf{a}_q$, $\mathbf{b}_q$, and $\mathbf{c}_q$ is called the $q$th triad in the PARAFAC decomposition and is obtained by multiplying $a_{iq}$, $b_{jq}$, and $c_{kq}$ for each $i$, $j$, and $k$. The PARAFAC model as represented in eq 3 is visualized in Figure 1.

(9) Smilde, A. K.; Wang, Y.; Kowalski, B. R. *J. Chemom.* **1994**, *8*, 21.

(10) Wang, Y; Borgen, O.; Kowalski, B. R.; Gu, M.; Turecek, F. *J. Chemom.* **1993**, *7*, 117.

(11) Harshman, R. A; Lundy, M. In *Research Methods for Multimode Analysis*; Law, H. G., Snyder, C. W., Hattie, J. A., McDonald, R. P., Eds.; Prager: New York, 1984.

(12) Smilde, A. K. *Chemom. Intell. Lab. Syst.* **1992**, *15*, 143–163.

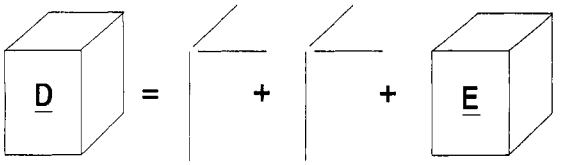


**Figure 1.** PARAFAC model. The three-way matrices **D** and **E** are explained in the text. The sticks to the right of the equal sign represent the vectors **a**, **b**, and **c** of the PARAFAC model.

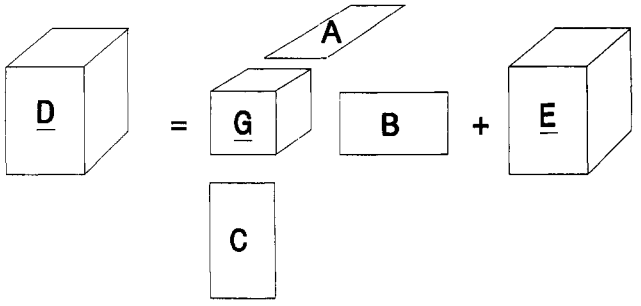


**Figure 2.** Tucker model. The matrices **A**, **B**, **C**, **D**, **E**, and **G** are explained in the text.

Another way to decompose $\underline{\mathbf{D}}$ is the Tucker model.[12,13] Again, let $d_{ijk}$ be an element of $\underline{\mathbf{D}}$, then

$$d_{ijk} = \sum_{p=1}^{P}\sum_{q=1}^{Q}\sum_{r=1}^{R} a_{ip}b_{jq}c_{kr}g_{pqr} + e_{ijk} \qquad (4)$$

is the Tucker decomposition of $\underline{\mathbf{D}}$. The elements $g_{pqr}$ are the typical elements of the three-way array $\underline{\mathbf{G}}$ ($P \times Q \times R$) which has a much smaller size than $\underline{\mathbf{D}}$. The three-way array $\underline{\mathbf{G}}$ is called the core array. $P$, $Q$, and $R$ are parameters to be chosen in such a way that $\underline{\mathbf{E}}$, with elements $e_{ijk}$, has a small norm. The Tucker model, as represented in eq 4, is visualized in Figure 2. From this figure it is obvious that the vectors $\mathbf{a}_p$, $\mathbf{b}_q$, and $\mathbf{c}_r$ (defined in the same way as above) are connected (multiplied) with each other by the core element $g_{pqr}$. In eq 4 this is accomplished by the triple summation.

From eqs 3 and 4 it is clear that PARAFAC is a special case of the Tucker model. Taking $P = Q = R$ and $g_{pqr} = 1$, if and only if $p = q = r$, transforms the Tucker model into the PARAFAC model. For the PARAFAC model the core array is a $P \times P \times P$ cube with ones on the superdiagonal. The Tucker model is more general due to the possibility of an unequal number of loading vectors in each dimension and because of a more general core array.

One of the latest developments in three-way analysis is the restricted Tucker model.[14] The Tucker model can be restricted by setting specific core array elements to zero. The result is a much more parsimonious model in terms of the number of parameters to be estimated. Only particular multiplications of $\mathbf{a}_p$, $\mathbf{b}_q$, and $\mathbf{c}_r$ are allowed. The zero elements in the core array can be selected by theory or chemical knowledge about the system.

Other restrictions also apply to the Tucker model. Since the pure underlying time profiles, spectra, and concentrations

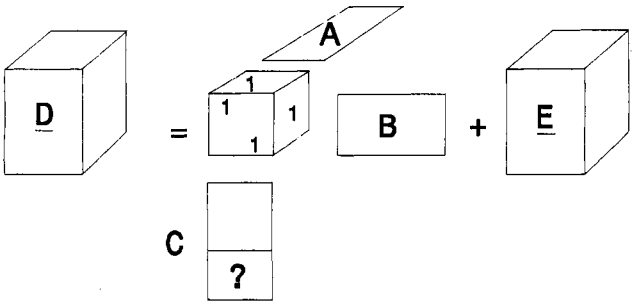


**Figure 3.** Restricted Tucker model. This model is a special case of the Tucker model of Figure 2. The core array has ones (or nonzero's) at particular places. The **C** matrix contains the concentrations, and part of **C** (the upper part) is known. The other part of **C** is unknown (?) and has to be estimated together with **A** and **B**.

**Table 1. Unfolded Core Array of the PARAFAC Four-Component Model of the Theory Example**

| | 1 | 2 | 3 | Q | 1 | 2 | 3 | Q | 1 | 2 | 3 | Q | 1 | 2 | 3 | Q |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| *P* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | | | 1 | | | | 2 | | | | 3 | | | | *R* | |

**Table 2. Unfolded Core Array of the Theory Example after Applying $y_3 = y_5$**

| | 1 | 2 | Q | 1 | 2 | Q | 1 | 2 | Q | 1 | 2 | Q |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| *P* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| | | 1 | | | 2 | | | 3 | | | *R* | |

are nonnegative, this can also be assumed for the estimated counterparts **A**, **B**, and **C**. Moreover, part of **C** is known: the part describing the concentrations in the standard solutions. All of these constraints can be incorporated in the restricted Tucker model. The estimation procedure becomes a combination of an alternating least squares and a nonnegative least squares algorithm.[15]

Counting the number of vectors in the calibration problem represented by eqs 2a–c present in the $X$, $Y$, and $Z$ domain, the maximum is four (in the $X$ domain). In order to provide "space" for the $X$-domain vectors a four-component PARAFAC model must be used to model eqs 2a–c simultaneously. The core array of this four-component PARAFAC model is a 4 × 4 × 4 three-way array with ones on the superdiagonal. If each horizontal slice of this core array is placed next to each other (unfolded or strung out[12]), the result is presented in Table 1.

In Table 1, $P = 4$, $Q = 4$, and $R = 4$ stand for the number of components in the $X$, $Y$, and $Z$ domain, respectively. More specifically, $\mathbf{x}_2$–$\mathbf{x}_5$ correspond to $P = 1$ to $P = 4$, respectively, and likewise $\mathbf{y}_2$–$\mathbf{y}_5$ correspond to $Q = 1$ to $Q = 4$. The $z$ vectors are represented by $R$ = 1–4. Applying the knowledge that $\mathbf{y}_3 = \mathbf{y}_5$ gives the unfolded core array of Table 2.

In Table 2 the column associated with $q = 4$ ($\mathbf{y}_5$) is added to the column associated with $q = 2$ ($\mathbf{y}_3$) of Table 1. Therefore, the value of $Q$ changes from 4 to 3 in Table 2. The first two

(13) Kroonenberg, P. M. *Three-mode Principal Component Analysis*; DSWO Press: Leiden, The Netherlands, 1983.
(14) Kiers, H. A. L. *Stat. App.* **1992**, *4*, 659–667.
(15) Lawson, Ch. L.; Hanson, R. J. *Solving Least Squares Problems*; Prentice-Hall: Englewood Cliffs, NJ, 1974.

**Table 3. Final Unfolded Core Array of the Theory Example**

| | 1 | 2 | $Q = 3$ | 1 | 2 | $Q = 3$ |
|---|---|---|---|---|---|---|
| 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 1 | 0 | 0 | 0 | 0 |
| 3 | 0 | 0 | 0 | 0 | 0 | 1 |
| *P* | 0 | 0 | 0 | 0 | 1 | 0 |
| | | 1 | | | $R = 2$ | |

columns of **Z** in the four-component PARAFAC model are equal because both dyads $\mathbf{x}_2\mathbf{y}_2^t$ and $\mathbf{x}_3\mathbf{y}_3^t$ represent the contribution of $N_2$. Both $\mathbf{z}_1$ and $\mathbf{z}_2$ will describe the concentration of $N_2$ in $N_2$, $N_3$, and **M**. Obviously, $\mathbf{z}_1 = \mathbf{z}_2$, and likewise, $\mathbf{z}_3 = \mathbf{z}_4$, where $\mathbf{Z} = [\mathbf{z}_1\ \mathbf{z}_2\ \mathbf{z}_3\ \mathbf{z}_4]$. In order to accommodate these relationships the $r = 2$ ($\mathbf{z}_2$) block is added to the $r = 1$ ($\mathbf{z}_1$) block and the $r = 4$ ($\mathbf{z}_4$) block is added to the $r = 3$ ($\mathbf{z}_3$) block of Table 2. The resulting unfolded core array is shown in Table 3.

The core array shown in Table 3 has exactly the right size ($4 \times 3 \times 2$) since the model has to support four $x$ vectors, three $y$ vectors, and two concentration vectors ($\mathbf{z}$). The restricted Tucker model is estimated using the core array of Table 3, applying nonnegativity constraints to the estimation of **A**, **B**, and **C**, and using the known concentrations in **C**.

Estimation of the restricted Tucker models starts with starting values for **A**, **B**, and **C**. Multivariate curve resolution (MCR) of the pure analyte responses gives estimates of the true time profiles (**A**) and spectra (**B**).[6,8] This is explained in detail in the second paper of this series. These estimated time profiles and spectra will be used as starting values. The first two rows of **C** are ($z_{22}$ 0) and (0 $z_{33}$), respectively. In the third row of **C**, arbitrary positive constants are used as starting values. After convergence of the iterative estimation procedure, the final row in **C** gives the estimates of the concentrations of $N_2$ and $N_3$ in **M**.

**Complexity of the Calibration Problems.** The calibration method developed will be illustrated with two data sets. The experimental setup of both data sets is described in part 1 of this series. The difference between both data sets is that the first one is measured at room temperature and the second one at 10 °C. It is expected that the same absorbing species are being formed, but of course at different reaction rates. The complexity of the first data set will be described in detail, and the complexity of the second data set differs mainly with respect to the temporal profiles and will be discussed shortly.

In part 2 of this series, the estimated unit spectra (spectra at unit concentration) and time profiles of the different species formed by the analytes in the Fujiwara reaction were presented. These spectra and time profiles can be helpful for assessing the complexity of the calibration problems.

If each spectrum is understood as a vector in the 100-dimensional space (the number of wavelengths at which absorbances are measured equal 100), then the cosine of the angle between the vectors serves as a measure of similarity of the spectra. These cosines are presented in Table 4.

The cosine varies between 0 and 1, where a 0 means that the spectra are very dissimilar and a 1 means that the spectra are very similar. From this table it is clear that there is a high overlap between the second species formed by TCE (TCE2) and both species formed by $CHCl_3$ ($CHCl_3$1 and $CHCl_3$2). Also, the spectra of both species formed by $CHCl_3$ are very

**Table 4. Spectral Similarities**

| | TCA | TCE1 | TCE2 | $CHCl_3$1 | $CHCl_3$2 |
|---|---|---|---|---|---|
| TCA | 1 | 0.86 | 0.47 | 0.49 | 0.40 |
| TCE1 | 0.86 | 1 | 0.70 | 0.76 | 0.62 |
| TCE2 | 0.47 | 0.70 | 1 | 0.98 | 0.98 |
| $CHCl_3$1 | 0.49 | 0.76 | 0.98 | 1 | 0.97 |
| $CHCl_3$2 | 0.40 | 0.62 | 0.98 | 0.97 | 1 |

**Table 5. Time Profile Similarities**

| | TCA | TCE1 | TCE2 | $CHCl_3$1 | $CHCl_3$2 |
|---|---|---|---|---|---|
| TCA | 1 | 0.15 | 0.93 | 0.30 | 0.94 |
| TCE1 | 0.15 | 1 | 0.36 | 0.95 | 0.37 |
| TCE2 | 0.93 | 0.36 | 1 | 0.56 | 0.999 |
| $CHCl_3$1 | 0.30 | 0.95 | 0.56 | 1 | 0.56 |
| $CHCl_3$2 | 0.94 | 0.37 | 0.999 | 0.56 | 1 |

similar (0.97) to each other. Since the second species formed by TCE and $CHCl_3$ are both final products in the Fujiwara reaction and have a very high spectral overlap, it is assumed that these species are identical.

The analyte TCA forms one species that has spectral differences with respect to the ones formed by TCE and $CHCl_3$. This makes the calibration for TCA easier, and also the calibration of TCE and/or $CHCL_3$ in the presence of TCA should be possible.

A similar approach can be taken with respect to the time profiles. Again a cosine is calculated between the different time profiles understood as vectors in the 61 dimensional space (measurements were taken at 61 regularly spaced time points). Note that the cosine gives information about the angle between time profiles and is therefore a measure of similarity of the unit time profiles: the time profiles scaled to unit concentration. In other words, the cosine is a measure of similarity in shape. This holds, of course, also for the spectral similarities.

In Table 5, the time profile similarities are shown. The first species formed by TCE and $CHCl_3$ overlap; the second species overlaps even more. There is also some overlap in time profile shapes of the species formed by TCA and the second species formed by TCE and $CHCl_3$. This overlap counteracts the spectral differences between the species formed by TCA and the ones formed by TCE and $CHCl_3$ (see Table 4).

In the calibrations, the full response matrices of the analytes are used. It is therefore useful to look at the full response matrices of each species and of the analytes. The response matrices of the five different species can be approximated by taking the outer product of the spectrum and the time profile for each species. Thus, five different response matrices are obtained, assuming five different species. Each response matrix is therefore of rank one. A congruence coefficient between these matrices can be calculated according to

$$\frac{\|\mathbf{A}^t\mathbf{B}\|}{\|\mathbf{A}^t\mathbf{A}\|^{1/2}\|\mathbf{B}^t\mathbf{B}\|^{1/2}} \tag{5}$$

where $\|\mathbf{A}\|$ is the Frobenius norm of a matrix.[16] This congruence coefficient varies between 0 and 1, where 0 means no similarity and 1 means perfect similarity between the matrices **A** and

(16) Golub, G. H.; van Loan, Ch. F. *Matrix Computations*; John Hopkins University Press: Baltimore, MD, 1989.

**Table 6. Pure Species Similarities**

| | TCA | TCE1 | TCE2 | $CHCl_3$1 | $CHCl_3$2 |
|---|---|---|---|---|---|
| TCA | 1 | 0.02 | 0.86 | 0.09 | 0.88 |
| TCE1 | 0.02 | 1 | 0.13 | 0.90 | 0.14 |
| TCE2 | 0.86 | 0.13 | 1 | 0.32 | 0.997 |
| $CHCl_3$1 | 0.09 | 0.90 | 0.32 | 1 | 0.32 |
| $CHCl_3$2 | 0.88 | 0.14 | 0.997 | 0.32 | 1 |

**Table 7. Analyte Similarities**

| | TCA | TCE | $CHCl_3$ |
|---|---|---|---|
| TCA | 1 | 0.71 | 0.68 |
| TCE | 0.71 | 1 | 0.998 |
| $CHCl_3$ | 0.68 | 0.998 | 1 |

**B**. In Table 6, the similarities between the different species are presented. It is clear that the first species formed by TCE (TCE 1) and formed by $CHCl_3$ ($CHCl_3$1) are similar (0.90) and the two second species ($TCE_2$ and $CHCl_3$2) are assumed to be equal, which is supported by the similarity of 0.997.

There is no direct relationship between the values in either Table 4 or 5 and Table 6. However, some conclusions can be drawn. The high similarity between the time profile shapes of TCA and those of TCE2 and $CHCl_3$2 is counteracted to some extent by a low spectral agreement. Still, the similarity of the total TCA species and the second TCE and $CHCl_3$ species is moderate (0.86 and 0.88, respectively). The first species of TCE and $CHCl_3$ are overlapped (0.90). This is a consequence of a high spectral similarity.

As a final step, the total similarity between analytes is calculated. In order to do this, the two total species responses TCE1 and TCE2 are added. The same is done for $CHCl_3$. For TCA, the total species response is the same as the total analyte response since TCA generates only one species. The congruence between the total response matrices generated by the pure analytes can be calculated using eq 5. The results are reported in Table 7.

Table 7 clearly shows the difficulty of the calibration problem. There is a high similarity between TCE and $CHCl_3$. This means that the quantitation of, for example, TCE in the presence of the unknown interferent $CHCl_3$, or vice versa, is a very difficult task. The analyte TCA as an interferent should not present too much difficulty.

If the experiment is performed at a lower temperature (10 °C; second data set) a difference in time profiles is expected. This is exactly what happens, and the time profile similarities show less overlap than in the first data set and result in lower species similarities. The analyte similarity of TCE and $CHCl_3$ is 0.46, which is low compared to the analogous similarity in the first data set (0.998).

While it is true that thus far only three analytes are dealt with which would yield good results if they are the only ones present, the real aim is to do the analysis in the presence of unexpected interferents. Therefore, one should consider the analysis of, for example, TCE in a case where TCA, $CHCl_3$, and other interferents are present. In order to illustrate the calibration methods, TCE will be used as an interferent in the calibration of $CHCl_3$ in a mixture and vice versa. These cases represent the hardest calibration cases, since the responses of TCE and $CHCl_3$ overlap severely. The results extend easily to cases of more unknown interferents.

**Table 8. Unfolded Core Array of Example 2**

| | 1 | 2 | $Q = 3$ | 1 | 2 | $Q = 3$ |
|---|---|---|---|---|---|---|
| 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 1 | 0 | 0 | 0 | 0 |
| 3 | 0 | 0 | 0 | 0 | 0 | 1 |
| $P$ | 0 | 0 | 0 | 0 | 1 | 0 |
| | | 1 | | | $R = 2$ | |

## EXPERIMENTAL SECTION

In this section different calibration examples are described which are used to test the restricted Tucker calibration methodology. The calibrations differ with respect to the presence of an interferent. Results for both data sets will be presented, but the structures of the calibration problems are the same for both data sets. Throughout it will be assumed that TCE and $CHCl_3$ have the second (final) species in common: the spectra of the second species being formed by TCE and $CHCl_3$ are equal, but their time profiles are not equal since the reaction kinetics differ. The structure of the calibration problems will be discussed in this section, whereas the results for the two different data sets will be presented in the Results section.

*Example 1.* The first calibration is already formalized in eqs 2a–c and will be called example 1. The model is estimated by using the constraints that **A**, **B**, and **C** are nonnegative and the first two rows in **C** are ($z_{12}$ 0) and (0 $z_{23}$), respectively. Starting values for **A** and **B** are obtained from multivariate curve resolution.

*Example 2.* In the second example, the analyte $CHCl_3$ is used as an unknown interferent. The standard $N_2$ is the response of pure TCE, and the mixture consists of TCE and $CHCl_3$. This is the real second-order challenge: is it possible to quantify for TCE in the presence of an unknown? Of course, it must first be determined whether TCE is in the mixture. There are several ways of doing this.[10] The equations that describe example 2 are

$$\mathbf{N}_2 = z_{12}\mathbf{x}_2\mathbf{y}_2^t + z_{12}\mathbf{x}_3\mathbf{y}_3^t \quad (6a)$$

$$\mathbf{M} = z_{12}\mathbf{N}_2^* + \mathbf{U}_m \quad (6b)$$

$$\mathbf{U}_m = z_{mu}\mathbf{x}_4\mathbf{y}_4^t + z_{mu}\mathbf{x}_5\mathbf{y}_5^t \text{ with } \mathbf{y}_5 = \mathbf{y}_3 \quad (6c)$$

where $z_{mu}$ stands for the concentration of the unknown $\mathbf{U}_m$ in **M**. The error matrices are left out for convenience. The unfolded core array associated with example 2 is presented in Table 8.

In Table 8, $P = 4$ ($X$ domain), $Q = 3$ ($Y$ domain), and $Q = 2$ (concentration domain). The first row of **C** is ($z_{12}$ 0). Since $CHCl_3$ acts as an interferent in example 2, the associated starting columns in **A** and **B** are chosen randomly. The columns in **A** and **B** associated with TCE have starting values obtained from multivariate curve resolution.

*Example 3.* In this example, a mixture of TCE and $CHCl_3$ is used and only $CHCl_3$ is used as a standard. Hence, it is the reverse of the previous example. The calibration model is nearly the same; the core arrays associated with examples 2 and 3 are equal. Of course, the proper arrangements have to be made in matrices **A**, **B**, and **C**.

## RESULTS AND DISCUSSION

**Quality of the First Data Set (Room Temperature).** In part 1 of this series, a general description of the experimental

**Table 9. Results of Example 1 Data Set 1**

| | TCE conc | max abs | $CGCl_3$ conc | max abs | SD |
|---|---|---|---|---|---|
| *N* | 7.31 | 0.24 | 1.48 | 0.24 | |
| *M* | 7.31 | 0.24 | 1.48 | 0.24 | |
| | 7.14 (2.3%) | | 1.58 (7.0%) | | 0.0027 |

**Table 10. Results of Example 2 Data Set 1**

| | TCE conc | max abs | $CHCl_3$ conc | max abs | SD |
|---|---|---|---|---|---|
| *N* | 7.31 | 0.24 | | | |
| *M* | 7.31 | 0.24 | 1.48 | 0.24 | |
| | 8.26 (13.2%) | | | | 0.0018 |

**Table 11. Results of Example 3 Data Set 1**

| | TCE conc | max abs | $CHCl_3$ conc | max abs | SD |
|---|---|---|---|---|---|
| *N* | | | 4.45 | 0.71 | |
| *M* | 2.93 | 0.09 | 4.45 | 0.71 | |
| | | | 3.84 (13.7%) | | 0.0025 |

**Table 12. Results of Example 1 Data Set 2**

| | TCE conc | max abs | $CHCl_3$ conc | max abs | SD |
|---|---|---|---|---|---|
| *N* | 0.488 | 0.055 | 0.995 | 0.28 | |
| *M* | 1.464 | 0.165 | 0.995 | 0.28 | |
| | 1.520 (3.8%) | | 0.975 (2.0%) | | 0.00065 |

setup is given. The pure experimental error of the data in this data set is between 3.2 and 3.4% (relative error) for TCA and between 2.4 and 2.5% for $CHCl_3$. For TCE, a similar number is expected. These values are calculated as averages over time for absorbances at particular wavelengths. Another way to approach reproducibility is by calculating

$$100\frac{\|\mathbf{A}-\mathbf{B}\|}{\|0.5\mathbf{A}+0.5\mathbf{B}\|} \tag{7}$$

where $\|\mathbf{A}\|$ denotes the Frobenius norm of **A** and the matrices **A** and **B** are responses of reproduced measurements. For TCE at a concentration of 7.31 ppm this number is 3.04%. A similar number is expected for $CHCl_3$. These numbers give an indication of the range of relative errors in concentrations due to measurement error.

In a similar way, a standard deviation value can be calculated by dividing $\|\mathbf{A}-\mathbf{B}\|$ by $61 \times 100$ (number of time-points times number of wavelengths) and taking the square root. These numbers are 0.0059 and 0.0113 for the above-mentioned TCE and TCA standards, respectively. The SD values given in Tables 9–11 can be compared with the range 0.0059–0.0113.

**Quality of the Second Data Set (10 °C).** The only difference between the second and first data set is the temperature. Experimental details are given in part 1 of this series.

For $CHCl_3$ the results of using eq 7 was 3–5%, and for TCE this number was 5–8%. These values give an indication of the experimental error in the second data set. A standard deviation value, analogous to the first data set, was calculated. These values were 0.0063 and 0.0048 for $CHCl_3$ and TCE, respectively. The SD values reported in Tables 12–14 can be compared with these values.

**Results for Data Set 1.** *Example 1.* It is clear from Table 7 that the analytes TCE and $CHCl_3$ have a very similar response. In the mixture of example 1, the analytes TCE and $CHCl_3$ are present in the same amount as in the standards. Table 9 shows the results of the calibration. The relative prediction errors are low, and the SD value indicates a very good fit of the model. Hence, despite the similar response of the analytes TCE and $CHCL_3$, calibration is possible provided that a mixture behaves well: linear additivity should hold, and each analyte has to contribute to a certain extent to the signal in the mixture. The latter can be checked by comparing the maximum absorbance contribution of the analytes to the mixture. For both analytes this maximum was 0.24, indicating that both analytes contribute in the same order of magnitude to the overall absorbance of the mixture. The results of this example are better than the corresponding ones in part 2 (Table 2, part 2).[6]

**Table 13. Results of Example 2 Data Set 2**

| | TCE concn | max abs | $CHCl_3$ conc | max abs | SD |
|---|---|---|---|---|---|
| *N* | 0.488 | 0.055 | | | |
| *M* | 1.464 | 0.165 | 0.995 | 0.28 | |
| | 1.418 (3.1%) | | | | 0.00068 |

**Table 14. Results of Example 3 Data Set 2**

| | TCE conc | max abs | $CHCl_3$ conc | max abs | SD |
|---|---|---|---|---|---|
| *N* | | | 0.995 | 0.28 | |
| *M* | 1.464 | 0.165 | 0.995 | 0.28 | |
| | | | 0.742 (25.4%) | | 0.00055 |

*Example 2.* In this example, the real second-order challenge is obtained: the mixture contains an unknown interferent. This is important since it tests the ability of the method to calibrate in the presence of an unknown interferent. The results are reported in Table 10. The interferent has approximately the same contribution to the mixture response signal as the analyte. The quantitation for TCE is performed with a relative prediction error of 13.2%. Given the high overlap between the analyte and the interferent this error is very reasonable, also compared to an experimental error in the order of 3–4%. The prediction errors are of the same order of magnitude as in part 2 (Table 2, part 2).[6] The fit of the model to the data is good again.

An analogous case of predicting the concentration of TCE in the presence of the interferent $CHCl_3$ was performed in part 1.[5] This resulted in a prediction error of 105%. The difference between 13.2% on the one hand and 105% on the other hand clearly shows the second-order advantage of the restricted Tucker model approach.

*Example 3.* This example is the reverse of the previous one. Now TCE acts as an interferent and the results are reported in Table 11. The interferent in this example has a lower contribution to the mixture response signal than the analyte. The relative prediction error in the concentration of $CHCl_3$ is 13.7%. This error is reasonable, also compared with the error of the same calibration in part 2 (Table 2, part 2)[6] and an experimental error of 3–4%. The SD value is low, indicating a good fit of the model.

The error of predicting the concentration of $CHCl_3$ in the presence of TCE in part 1[5] is 42.7%. Example 3 clearly gives a better result.

**Results for Data Set 2.** *Example 1.* The setup of the calibration is the same as for the first example of data set 1.

The results of the calibration are shown in Table 12. The prediction errors are very low, and the predictions can be considered as being very good. The model fits the data very well.

*Example 2.* This example shows the second-order advantage: calibration of TCE in the presence of an unknown interferent ($CHCl_3$). Table 13 presents the result of this calibration. The prediction error of TCE in the mixture is very low (3.1%). The prediction is very good: on the order of magnitude of the experimental error. The model fits the data very well.

*Example 3.* In this example, TCE serves as the unknown interferent in a mixture also containing the analyte $CHCl_3$. Table 14 presents the result. The prediction error of $CHCl_3$ is higher than in example 2. This is because the two species being formed by $CHCl_3$ have very similar spectra (a cosine value of 0.99); this causes instability in the estimated model, especially in the **B** matrix. Since the model fits the data very well, the conclusion is that the error surface is rather flat in the neighborhood of the minimum error: relatively small changes in parameter values (e.g., the estimated concentration) do not change the error very much.

**General Comments.** The starting values of the algorithm for the unknown parts in **C** are important. Calculations have shown that different starting values give slightly different calibration results. Good starting values can perhaps be found using nonbilinear rank annihilation.[10] More research is needed in this area.

The problem encountered in example 3 of data set 2 can be circumvented by designing better algorithms. Such algorithms should be more robust to similarities in spectra (estimation of **B**) and time profiles (estimation of **A**).

In all the above examples, only one standard was used to quantitate for an analyte in a mixture. The restricted Tucker model approach can be augmented easily to support more than one standard. The results should be better when multiple standards that span a range of concentrations are used.

## GENERAL CONCLUSIONS

A complete new class of calibration models is described for second-order calibration. It is shown with experimental data that quantitation of analytes in mixtures containing unknown interferences is possible. The restricted Tucker model approach makes it possible to develop fully selective reaction-based chemical sensors.


## ACKNOWLEDGMENT

A.K.S. acknowledges a grant from the Netherlands Organization for Scientific Research (NWO) for his stay at the Laboratory for Chemometrics at the University of Washington. R.T. acknowledges the financial support from the DGICYT (Spain) for his stay at the Laboratory for Chemometrics at the University of Washington.

Received for review March 15, 1994. Accepted June 30, 1994.[⊗]

[⊗] Abstract published in *Advance ACS Abstracts,* August 15, 1994.