Chemometrics and Intelligent Laboratory Systems, 15 (1992) 143–157
Elsevier Science Publishers B.V., Amsterdam

Three-way analyses
Problems and prospects

Age K. Smilde

Research Group Chemometrics, University Centre for Pharmacy, Groningen (Netherlands)

(Received 17 June 1991; accepted 12 November 1991)

Abstract

Smilde, A.K., 1992. Three-way analyses. Problems and prospects. *Chemometrics and Intelligent Laboratory Systems*, 15:143–157.

Three-way methods are multivariate data analysis tools that compress and visualize simultaneous variation of combinations of variables and objects. Unfolding, Tucker and PARAFAC models are examples of such methods. The interest in such tools is increasing in chemometrics. The history of three-way methods is sketched and the basic theory is presented, together with chemical applications. Problems in the three-way analyses area are summarized. Some prospects for future use of three-way methods are given.

INTRODUCTION

The increasing amount of data originating from advanced instrumentation has clearly revealed the need for statistical tools to handle this amount of data. In analytical chemistry, for instance, the growth of hyphenated analysis instruments is such an example. Such developments can be seen in other areas of chemistry, too. Multivariate analysis is particularly suitable for handling large amounts of data, since the problem for which the data are collected usually has a multivariate nature: it can be formulated in terms of the simultaneous influence of a (large) number of variables. Numerous successful applications of multivariate analysis in chemistry are available, showing the power of these methods.

A relatively new branch in the field of multivariate analyses is the analysis of multi-way data. In order to acquaint the reader with multi-way data, some examples of this kind of data will be given later on, in the section on theory and application. For the moment it will suffice to say that multi-way data are data that can be arranged according to more than two categories of variables and/or objects. For example, three-way data are a collection of numbers which can meaningfully be arranged in a data table with three directions, or modes. Most of the readers will be familiar with two-way data: these can be arranged in a two-way table, a matrix. A three-way data table is the logical extension of a matrix.

Correspondence to: Dr. A.K. Smilde, Research Group Chemometrics, University Centre for Pharmacy, A. Deusinglaan 2, 9713 AW Groningen, Netherlands.

The purpose of this paper is to present the ideas behind multi-way analyses, to position multi-way analyses in relation to other multivariate tools, to discuss problems of multi-way analyses, and to sketch the prospects of such methods. The discussion is restricted to three-way analyses since this part of multi-way analysis has been developed to the greatest degree.

Some theory will be presented first, illustrated by a selection of published applications. The problems will be discussed and finally the prospects will receive attention.

THEORY AND APPLICATIONS

This section will be divided into separate parts according to two aspects of data analysis. The first aspect pertains to whether the data are arranged in a two-way or a three-way setup. The second aspect describes whether the data are arranged in one block, two blocks or more than two blocks, respectively. Although not explicitly mentioned in the separate parts of this section, there should be an intimate relation between problem formulation and the ordering and kind of data obtained [1].

Obviously, other aspects of the data are also important: are the measurements made on a ratio, an interval, an ordinal or perhaps on a nominal scaled level? Are the data discrete or continuous? For ease of presentation, ratio-scaled continuous data are presupposed.

Bold, lower case italic characters refer to vectors, bold roman capitals refer to two-way matrices, and bold italic, underlined, capitals refer to three-way matrices. Some capitals have a special meaning: G, I, J, K, L, M, and N are the number of levels in different modes or the number of components in the models (the exact meaning is clarified in the text); g, i, j, k, l, m, and n are used as running indices.

The reader is probably familiar with most of the methods presented for two-way data. In order to make the connection with three-way methods more clear, the two-way methods receive explicit attention.

Methods for two-way one-block data

The most important method in this class is probably the singular value decomposition (SVD), or the closely related principal component analysis (PCA). These methods have been explained in detail elsewhere [2,3]. Suppose the data are collected in the $I \times J$ matrix $\mathbf{X}$, with typical element x_{ij}. The SVD of $\mathbf{X}$ can be written as

$$\mathbf{X} = \mathbf{UDV}' \tag{1}$$

where $\mathbf{U}$ $(I \times G)$ and $\mathbf{V}$ $(J \times G)$ have the property $\mathbf{U}'\mathbf{U} = \mathbf{V}'\mathbf{V} = \mathbf{I}$ and $\mathbf{D}$ $(G \times G)$ is a diagonal matrix with non-negative diagonal elements arranged in decreasing order and g is the rank of $\mathbf{X}$. The power of the SVD becomes clear when (1) is written in the spectral decomposition form:

$$\mathbf{X} = \boldsymbol{u}_1 \boldsymbol{d}_1 \boldsymbol{v}_1' + \boldsymbol{u}_2 \boldsymbol{d}_2 \boldsymbol{v}_2' + \cdots + \boldsymbol{u}_G \boldsymbol{d}_G \boldsymbol{v}_G' \tag{2}$$

where $\boldsymbol{u}_i$ $(i = 1, \ldots, G)$ and $\boldsymbol{v}_j$ $(j = 1, \ldots, G)$ are the ith and jth column vectors of $\mathbf{U}$ and $\mathbf{V}$, respectively, and $\boldsymbol{d}_g$ $(g = 1, \ldots, G)$ is the gth diagonal element of $\mathbf{D}$. Note that each individual matrix $\boldsymbol{u}_i \boldsymbol{d}_i \boldsymbol{v}_i'$ has rank one. It can be shown [4] that $\boldsymbol{u}_1 \boldsymbol{d}_1 \boldsymbol{v}_1'$, is the best rank-one approximation of $\mathbf{X}$ in the sense that $\boldsymbol{u}_1 \boldsymbol{d}_1 \boldsymbol{v}_1'$ minimizes the Frobenius norm (sum of squared entries) of $\mathbf{X} - \boldsymbol{u}_1 \boldsymbol{d}_1 \boldsymbol{v}_1'$, the residuals. Likewise, $\boldsymbol{u}_2 \boldsymbol{d}_2 \boldsymbol{v}_2'$ is the best rank-one approximation of $\mathbf{X} - \boldsymbol{u}_1 \boldsymbol{d}_1 \boldsymbol{v}_1'$ with the restriction that $\boldsymbol{u}_2$ is orthogonal to $\boldsymbol{u}_1$ and $\boldsymbol{v}_2$ is orthogonal to $\boldsymbol{v}_1$.

The individual vectors $\boldsymbol{u}$ are called principal components and can be regarded as latent vectors [5,6] which carry the hidden structure of $\mathbf{X}$. Various kinds of plots will make interpretation of the latent structure more easy [7,8]. The best rank-two approximation of $\mathbf{X}$ is shown in Fig. 1.

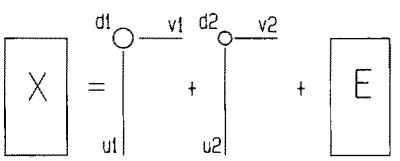


Fig. 1. The singular value decomposition (SVD). The matrix $\mathbf{X}$ is decomposed as a summation of scores ($\boldsymbol{u}_1$ and $\boldsymbol{u}_2$) times loadings ($\boldsymbol{v}_1$ and $\boldsymbol{v}_2$). $\mathbf{E}$ is the matrix of residuals. The circles represent the singular values $\boldsymbol{d}_1$ and $\boldsymbol{d}_2$. In practical applications $\boldsymbol{d}_1$ and $\boldsymbol{d}_2$ are usually absorbed into $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$.

The circles in the figure represent the singular values (the d values). Usually the singular value is absorbed in the u vectors. The squares of the singular values measure the importance of the individual components. The number of components should be kept as low as possible and can be chosen according to different kinds of criteria [9,10].

Data preprocessing is an important step. Data centering is usually performed, e.g. column centering the **X** matrix. This is not always an obvious choice: in spectral map analysis [11] double centering (over columns and rows) is performed for good reasons. Scaling the data is also not straightforward. Usually, when the **X** matrix contains measurements made on different kinds of variables, i.e. not measured in the same units, scaling the columns to unit length (or variance) is recommended. In case of variables that are all measured in the same unit, it is less clear whether or not to scale.

A particular kind of centering and scaling is performed in correspondence analysis [12]. This kind of analysis is best suited for contingency tables, i.e. when the data represent counts.

The way PCA is presented above is a typical data-analytical approach. A more statistical approach, in terms of linear combinations of the original variables, is given by Anderson [13]. An interesting justification of the use of PCA in the analysis of data tables is given by Wold [14]. He shows that PCA can be regarded as an extension of Taylor's approximation theorem.

For ease of discussion, let the rows of **X** represent objects and the columns of **X** the variables measured on those objects. When the analysis of the differences between the objects is wanted, other methods than PCA are available. Grouping

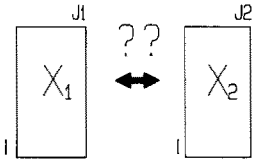


Fig. 2. $\mathbf{X}_1$ and $\mathbf{X}_2$ are matrices between which a connection is sought. One mode is common, e.g. I objects. The number of levels in the other modes (e.g. variables) are indicated by J_1 and J_2.

of the objects can be analyzed by discriminant analysis [3], SIMCA [15], ALLOC [16] or CLASSY [17], when the grouping structure is known beforehand (supervised learning). The objective is then to see whether the grouping structure represents itself in the measured variables. Unsupervised methods, for example cluster analysis [3], are needed when the grouping structure is not known and should be discovered.

When interest focuses on relations between variables in the block other methods come to mind. The principal variables approach [18] tries to select the most important variables. PCA and factor analysis, perhaps with factor rotation options, give also insight in relations between variables.

Methods for two-way two-block data

The two-block problem is visualized in Fig. 2: the problem is formulated in such a way that the connection between the matrices $\mathbf{X}_1$ and $\mathbf{X}_2$ is sought. Usually, $\mathbf{X}_1$ and $\mathbf{X}_2$ share the same objects (have one mode in common). Stated otherwise, two different blocks of variables are measured on the same objects.

The most apparent problems that translate into this situation are the regression problems: $\mathbf{X}_2$ contains the **Y** variables (the dependent variables) and $\mathbf{X}_1$ contains the independent variables. Response surface methodology [19] is an outstanding example of this type of data analysis. Another area is multivariate calibration [20]; however, there is one difference with response surface methodology: in multivariate calibration the distinction between independent and dependent variables is not made, or is not always clear.

Typical statistical methods used in the two-block case are multiple linear regression [3], partial least squares (PLS) regression [20], principal component regression [2,3,20] and ridge regression [21]. All these methods seek to make good predictions of the $\mathbf{X}_2$ block variables, using the $\mathbf{X}_1$ block variables as predictors. In this respect, the two blocks are not treated symmetrically.

Methods that do not treat the two blocks asymmetrically are, for instance, canonical correlation [3] and procrustes analysis [22]. Both meth-

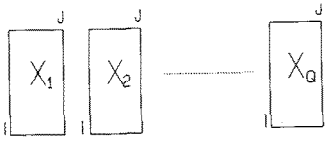


Fig. 3. $\mathbf{X}_1$ to $\mathbf{X}_Q$ are matrices between which similarities are sought. One mode is common, e.g. I objects. The number of levels in the other modes are indicated by J.

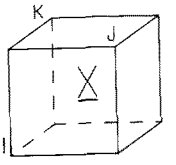


Fig. 5. A three-way data matrix $\underline{X}$ with number of levels I, J, and K in the different modes.

ods aim at understanding the connection between both blocks: canonical correlation tries to find correlated linear combinations between both blocks and procrustes analysis tries to rotate one of the blocks in order to resemble more the other block after the (optimal) rotation.

Methods for two-way multi-block data

The general setup of the data matrices is depicted in Fig. 3. Again, there is usually one mode in common: the object mode. An example of such a problem is sensory data: a number of food varieties (I) is tested on J qualities by Q panel members. Is there agreement between the panel members with respect to certain quality variables? To what extent do the panel members disagree?

Consensus PLS [23] tries to find a common underlying structure for all $\mathbf{X}$ matrices. Generalized procrustes analysis [24] tries to deal with the same question. This method tackles the problem of investigating the differences between the matrices $\mathbf{X}_i$ by rotating simultaneously all these matrices.

Another type of multi-block problem is shown in Fig. 4. The matrices are placed in a certain hierarchy, as a direct result of model assumptions

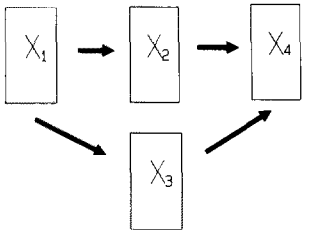


Fig. 4. The four matrices $\mathbf{X}_1$ to $\mathbf{X}_4$ are placed in a certain hierarchy and analyzed as a path model.

associated with the problem at hand. Path analysis [25] and LISREL [26] are developed to handle this situation. The original PLS method was also designed to solve these kind of problems [26].

Methods for three-way one-block data

The data matrix $\underline{X}$ that should be analyzed is visualized in Fig. 5. Three modes can be distinguished: mode one with index $i = 1, \ldots, I$, mode two with index $j = 1, \ldots, J$ and mode three with index $k = 1, \ldots, K$. The general term 'mode' is used here since the difference between object and variable is not always clear in multi-way data. In applications of three-way methods in image analysis [27] the first mode can consist of different wavelengths, the second and third modes describing the spatial arrangement of the surface of which the image is taken at the different wavelengths. In this case, the second and third modes may be called objects. In another application [28] in which, for different samples, chromatograms are taken with a diode-array detector, the samples may be called the objects. Yet, in the case of a three-way matrix, in which one mode consists of different stationary phases, the second mode consists of different mobile phase compositions and the third mode consists of different test solutes [29,30], it is not at all clear what the objects are. From now on, the term mode is used: no attempts are made to classify modes as objects or variables.

In order to analyze the three-way data matrix $\underline{X}$, generalizations of the SVD would be convenient. Psychometricians have already put a lot of effort into this area and have come up with interesting methods. The first generalization was put forward by Tucker [31] and is called the

Tucker3 model. This model is shown in Fig. 6 and, written out formally, it is

$$x_{ijk} = \sum_{l=1}^{L} \sum_{m=1}^{M} \sum_{n=1}^{N} a_{il} b_{jm} c_{kn} z_{lmn} + e_{ijk} \qquad (3)$$

where a_{il} is the typical element of the $(I \times L)$ loading matrix **A** of the first mode, b_{jm} of the $(J \times M)$ second and c_{kn} of the $(K \times N)$ third mode loading matrices, **B** and **C**, respectively. The element z_{lmn} is the typical element of the three-way core matrix $\underline{Z}$ and e_{ijk} is the typical element of the three-way residual matrix $\underline{E}$. If the numbers L, M and N are small a considerable data reduction has taken place. The core matrix $\underline{Z}$ represents the magnitude of and the interactions between the latent factors. Plots of **A**, **B** and **C** may help the interpretation of the latent factors. The matrices **A**, **B** and **C** are column-wise orthonormal [32,33] and **A**, **B**, **C** and $\underline{Z}$ are chosen in such a way that the sum of squared residuals is minimized. The three-way Model 3 does not have a unique solution: rotations can give other solutions with the same sum of squared residuals as the one found.

Two points are worth noticing about the Tucker3 model. Firstly, interaction between latent factors of different modes is provided for by the off-super diagonal elements of the core matrix. Secondly, the number of factors distracted for each mode need not be equal, i.e. L, M, and N are not necessarily equal. Both observations have repercussions on the kind of problem to which the Tucker3 model can be applied. Extensions to the case of four-mode matrices [34] and n-mode matrices [35] have been reported.

One of the earliest applications of the Tucker3 model in chemistry is reported by De Ligny et al. [36]. In this study the Tucker3 model is used to fit normal-phase retention values where the different modes are solutes, adsorbents and eluents. Another application is in the field of air polution chemistry [37] where the modes correspond to chemical elements, sampling sites and time periods. An attempt is made to give a physical interpretation of the Tucker3 model.

The second generalization takes up the idea of the SVD as giving a bilinear model of **X** and provides a trilinear model of $\underline{X}$. Both the CANDECOMP [38] and the PARAFAC [32,39] model come down to a trilinear model. This model is depicted in Fig. 7 and is formally

$$x_{ijk} = \sum_{g=1}^{G} a_{ig} b_{jg} c_{kg} + e_{ijk} \qquad (4)$$

where x_{ijk} and e_{ijk} have the same meaning as in Eqn. 3 and a_{ig}, b_{jg}, c_{kg} are typical elements of the loading matrices **A** $(I \times G)$, **B** $(J \times G)$ and **C** $(K \times G)$. **A**, **B** and **C** are chosen in such a way that the sum of squared residuals is minimized.

Model 4 is seen to be a simplification of Model 3. Firstly, the number of components in all modes

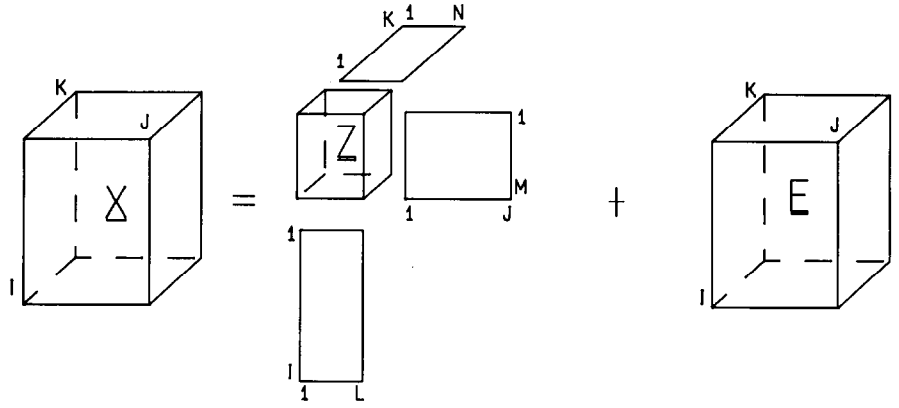


Fig. 6. The Tucker3 model. $\underline{Z}$ is the $(L \times M \times N)$ core matrix and $\underline{E}$ is the matrix of residuals. L, M, and N are the number of components for the first, second, and third mode, respectively.

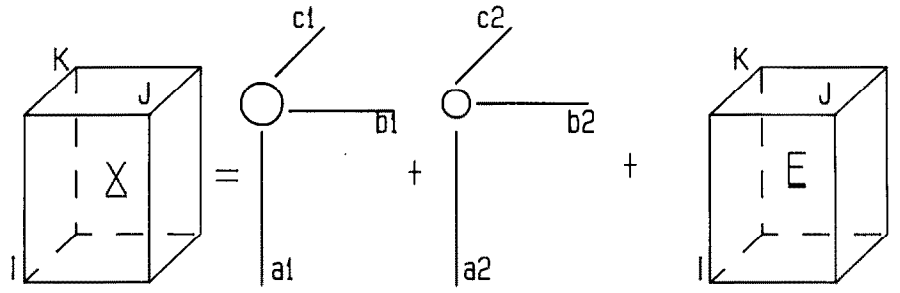


Fig. 7. The PARAFAC model. The circles indicate the sizes of the components, as in the SVD case.

are equal and, secondly, there is no interaction allowed between the latent factors of different modes. Stated otherwise, the core matrix of Eqn. 3 has simply become a three-way identity matrix with ones on the main superdiagonal. It is of course possible to rewrite Eqn. 4 and extract a size component (a kind of eigenvalue) in order to obtain insight in the importance of a specific component. This size component is visualized in Fig. 7 with a circle.

In the PARAFAC model the components can be chosen in such a way that they are not necessarily orthogonal. Fig. 8 shows a more general PARAFAC decomposition for not-necessarily orthogonal components. Note that in the core matrix of Fig. 8 only the super diagonal is non-zero, which makes it clear that the PARAFAC model is a simplified version of the Tucker3 model.

One of the advantages of the PARAFAC model (Model 4) is that the solution is rotation-dependent. Every rotation destroys the minimum sum of squares of the optimal solution. Hence, the solution is unique (apart from sign differences) and is directly interpretable.

The generalized rank annihilation method can be regarded as a PARAFAC model [40] and is used for calibration purposes with liquid chromatography–ultraviolet (LC–UV) measurements.

The final generalization of the SVD is given by the principle of unfolding [41]:

$$x_{ijk} = \sum_{g=1}^{G} t_{ig} v_{jkg} + e_{ijk} \qquad (5)$$

where x_{ijk} and e_{ijk} have the same meaning as before, t_{ig} is the typical element of the $(I \times G)$ score matrix $\mathbf{T}$ and v_{jkg} is the typical element of the loading matrix $\mathbf{V}_g$ $(J \times K)$. The loading matrices $\mathbf{V}_1$ to $\mathbf{V}_G$ can be stacked and the three-way loading matrix in $\underline{V}$ $(J \times K \times G)$ is obtained. Model 5 is shown in Fig. 9, where Eqn. 5 is rewritten in order to provide for a size compo-

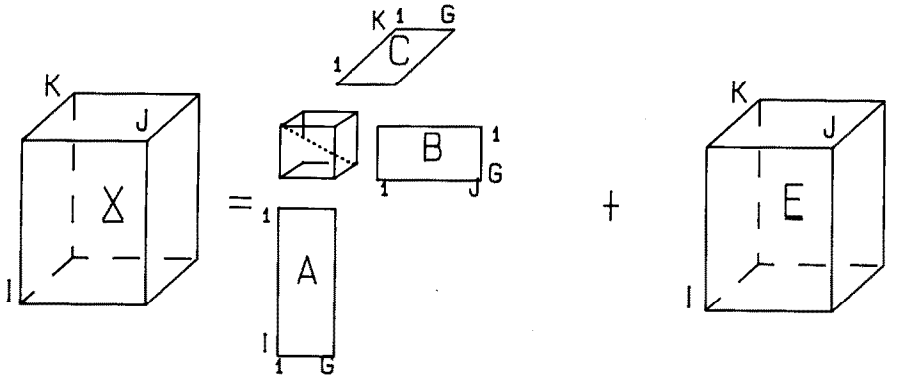


Fig. 8. The PARAFAC model rewritten. The PARAFAC model is seen to be a special case of the Tucker3 model where the number of components in each mode is equal and the core matrix is superdiagonal by (- - - - -).

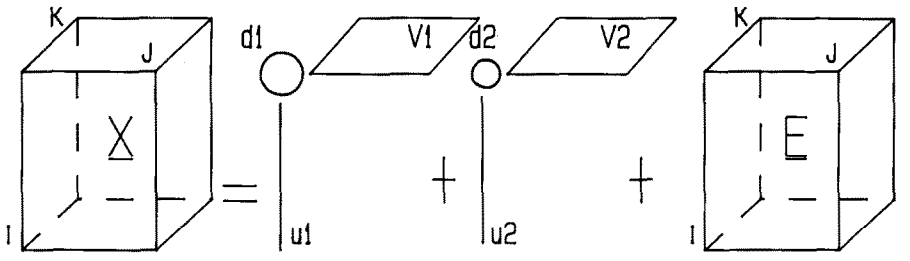


Fig. 9. The unfolded PCA model. The matrix $\underline{X}$ is written as a summation of score vectors (u_1 and u_2) times loading matrices (V_1 and V_2). The circles represent the singular values d_1 and d_2. Usually, the ds are absorbed in the us.

nent (the circles), again indicating the importance of a factor. The principle of unfolding is explained elsewhere, it suffices to say here that **T** and $\underline{V}$ are calculated in such a way that the residual sum of squares is minimized. Rotation of the solution does not destroy the minimum sum of squares. The solution is, therefore, not unique. Note that the treatment of the modes is asymmetric. Hence three different model specifications are possible for a given three-way matrix. Applications of this kind of generalization are, e.g., in image analysis [27,41] and in round-robin tests [42]. In the latter application the three-way data set ($40 \times 4 \times 4$) consists of measurements on four tests with synthetic solutions of Cd, Cu, Zn, and Ni made by eight laboratories, each of which provided five repeat estimations. The laboratories were tested for stable and reliable operations.

There are a number of differences between the three generalizations (Models 3, 4 and 5). This can already be inferred from the equations. Rotation dependency of the solutions gives important differences. The idea of a latent factor differs between the Tucker3 and the PARAFAC model [32]. The PARAFAC model is trilinear, whereas the other two models are not.

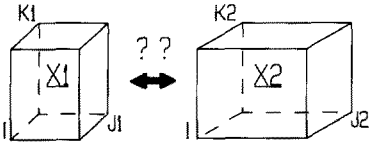


Fig. 10. Three-way matrices $\underline{X}_1$ and $\underline{X}_2$ between which a connection is sought. One mode is common (e.g. I objects).

Methods for three-way two-block data

The general problem is depicted in Fig. 10. What is the connection between the two three-way matrices? Again the idea should be stressed that there is a close relationship between a problem and the ultimate way it is formulated. At first glance, Fig. 10 might seem a little far-fetched, but serious problems which can be formulated in a Fig. 10 structure have already been published [29,30,44,45].

Both the PARAFAC model and the unfold approach have been generalized to solve the question posed in Fig. 10. The unfold solution is discussed first (see ref. 41).

It is supposed that one of the modes is common to both three-way matrices, for instance the object mode with running index $i = 1, \ldots, I$. Fig. 11 shows the unfolding of both three-way matrices leaving the common object mode intact. The PLS method can be used to calculate t_1, u_1, w_1, p_1 and q_1 [20]. It can be shown that these parameters are calculated in such a way that the covariance between t_1 and u_1 is maximized [46]. It is

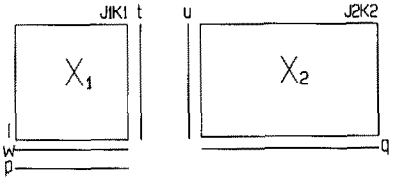


Fig. 11. The two matrices $\underline{X}_1$ and $\underline{X}_2$ of Fig. 10 are unfolded in such a way that the common mode is left intact. A common PLS procedure can be used to estimate the scores t and u, the loadings p and q, and the weights w.

possible to fold p_1 back, in order to obtain a loading matrix $\mathbf{P}_1$. The number of components can, for instance, be established by cross-validation. For a new object, the associated $\mathbf{X}_1$ values have to be known, then the associated $\mathbf{X}_2$ values can be predicted using the PLS model coefficients. Applications are given in the field of chromatography [29,30], pharmacology [44] and with high-performance liquid chromatographic–diode array data [28].

Ståhle [44] gives an interesting algorithm to solve the trilinear three-way decomposition of two-block data. This is a generalization of the question posed in Fig. 10 with a PARAFAC approach. Ståhle assumes that $\underline{X}_1$ and $\underline{X}_2$ have at least one mode in common and he gives a trilinear decomposition of $\underline{X}_1$ that can be used to predict $\underline{X}_2$. An example from the field of pharmacology is given where in $\underline{X}_1$ the first mode consists of rats treated and untreated with amphetamine, the second mode consists of different kinds of neurochemicals and the final mode is the time at which samples are taken from the rats by microdialysis. The data in $\underline{X}_1$ are measurements of concentrations of the neurochemicals. The $\underline{X}_2$ matrix is simply one column with zeros and ones representing whether the rat is treated or not.

Another approach to use the PARAFAC model to help solving the problem presented in Fig. 10 is due to Smilde et al. [29]. They make use of the fact that there is a common mode, for instance (again) the object mode $i = 1, \ldots, I$, and the variables in the other modes have the same background. Stated otherwise, the levels $j_1 = 1, \ldots, J_1$ of $\underline{X}_1$ and $j_2 = 1, \ldots, J_2$ of $\underline{X}_2$ can be pooled in a meaningful way to one set of levels $j = 1, \ldots, J$. The same holds for the levels in the third mode. An example is given from the field of chromatography. The common (first) mode is the stationary phase, the second mode consists of mobile phase compositions at which retention measurements of a set of solutes (the third mode) is made. Measurements of a subset of the solutes at a subset of the mobile phases are put in $\underline{X}_2$. The PARAFAC model is used to predict retention values on a new stationary phase (a new object). For a new stationary phase, only measurements of the solutes at mobile phases that

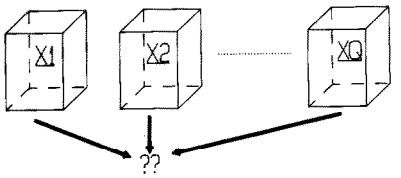


Fig. 12. The connection is sought between three-way matrices $\underline{X}_1 - \underline{X}_Q$ which share at least one mode.

are not in $\underline{X}_2$ suffice to predict the retention values of the subset of solutes at the subset of mobile phases in $\underline{X}_2$. In practice this means that, for instance, for any new stationary phase only three chromatograms of each four solutes have to be taken to predict retention values of a (much) larger set of solutes at a larger set of mobile phase compositions.

Methods for three-way multiple-block data

The multiple-block situation is depicted in Fig. 12. There are as yet no methods for dealing with this type of problem. Are there really no problems that can be meaningfully made operational in such a dataset structure?

Software for the three-way methods

Different software packages are available to calculate the parameters in the three-way models (see the references for addresses). The unfolding solution can be calculated with any program that can handle ordinary PCA or PLS. Usually, missing values can be handled by such software and personal computer versions are available. For the PARAFAC model a program written by Harshman [47], based on an iteration scheme, is available [47], but convergence is sometimes a problem and it is not always possible to impose orthogonality constraints [48]. Missing values can be handled by the Harshman software and a personal computer version is available. A closed solution of the trilinear (PARAFAC) model is published which works well if the factors are linearly independent in at least two of the modes [49]. The Tucker model can be handled by GEPCAM (generalized principal components with missing values) [50]. An alternative for the per-

sonal computer is the program TUCKALS3 [51] which can handle small and medium problems. An algorithm taking care of missing data is being developed [52].

An interactive matrix language (IML) is incorporated in the software package SAS [53]. The NIPALS algorithm, which is often the basis for PCA and PLS calculations, can easily be implemented in SAS/IML.

PROBLEMS

Notation

A very important topic is the kind of notation used in the three-way context. This is perhaps not of fundamental importance, but is surely so from a didactic perspective. Psychometricians have an idiom completely different from chemometricians. Even among the latter, there are two 'schools of notation': the tensor school (Sanchez and Kowalski [40]) and the school of Wold et al. [41]. Some agreement in this respect would be helpful.

As yet unrealized generalizations

Some of the methods presented in the previous sections have not yet been generalized to the three-way case. It is tempting to make all kinds of generalizations, but GLOGA (general lack of good applications) should be avoided [54]. In the Prospects section some ideas for future applications of three-way methods are given, pointing towards meaningful generalizations.

One of the first generalizations which should be tackled is the 'multi-mode variable selection' problem. In the two-way case, methods are available to select within a set of p variables a small number of important ones using the two-way matrix $\mathbf{X}$ where n observations are made on the p variables [2,18]. The selected (important) variables explain as much of the variation in $\mathbf{X}$ as possible. When these important variables are called markers [6,55], projections on these marker variables may give insight into the data set at hand. Analogously, marker object plots can be used.

For the three-way case, where two of the modes are regarded as variables, this generalizes to the question of how combinations of variables in the two variable modes can be chosen in such a way that they explain as much of the variation in $\underline{X}$ as possible. Perhaps the idea of marker variable projections can be generalized, where projections are made on the markers from the two different modes. A 'quick and dirty' method of the multi-mode variable selection problem is described by Smilde et al. [29]. For the multivariate calibration of chromatographic systems, multi-mode variable selection is a problem [29,30].

When a problem can be formulated in terms of the connection between two three-way matrices and when the underlying structure of the two matrices is best described by a Tucker3 model then no method is available for connecting the two blocks. A predictive two-block Tucker3 model might be worth developing. Another method that has not yet been developed for connecting two three-way matrices is three-way procrustes analysis.

When multiple blocks are involved, no three-way methods are available connecting these blocks in a presupposed hierarchy. Generalized three-way procrustes analysis might be worthwhile.

Error propagation and diagnostics

The parameters calculated in the three-way models are, of course, subject to error. When the three-way model is used to make predictions — e.g. the prediction of new retention values [29,30], or the prediction of concentration levels [40] — these predicted values carry an error. For the two-way case some work has been done for calculating this error [56]. For the three-way case little is known: only one theoretical paper has appeared [57].

Diagnostics are particular values that can be calculated and interpreted in order to gain insight into the performance of a method or a model. In the area of regression, diagnostics like Cook's distance and leverages are especially important [58].

Some diagnostics have already been developed for two-way methods. For PCA, for instance, leverages can be calculated [20]. Also, residual variances of fitted objects and variables in PCA and PLS may serve as diagnostic tools.

Residual variances are also useful tools in three-way methods. It would be convenient to have available diagnostics with which one can make a choice between the different three-way models. Is the underlying structure of a three-way matrix trilinear or actually bilinear? Should the more complicated Tucker model be used or the simpler PARAFAC model?

Local rank, contextuality and homogeneity

Local rank is the rank of a part of a two-way matrix [59]. A local rank map can be created in which each part of the two-way matrix is given a rank. A requirement when creating a local rank map is that the modes of the matrix have some metric, e.g. increasing time. This map gives an idea of the homogeneity of a data set. Local rank maps can be generalized to three-way matrices, but different generalizations are possible. The three-way local rank map might give information about the homogeneity of the three-way data set. Supposing that the three-way local rank map shows large rank fluctuations, does this have implications for the kind of (global) methods used to analyze the three-way data set? Perhaps a three-way local rank map can help in solving the multi-mode variable selection problem mentioned earlier.

In contrast to the local rank map, most of the methods for analyzing a three-way matrix do not reckon with contextuality. Interchanging different levels within one mode does not change the outcome of the method. This feature is not always desirable. A simple two-way example from reversed-phase liquid chromatography will illustrate this.

Suppose that capacity factors of p test solutes are measured at n mobile phase compositions. It is known from chromatographic theory [60] that the logarithm of the capacity factor of a solute can be written as a smooth function of the fractions of the mobile phase constituents. When the p solutes are understood as variables and the n mobile phases as objects, the matrix $\mathbf{X}$ $(n \times p)$ can be formed. Summaries of the data (first and second order moments) are the $(p \times 1)$ vector of variable mean values and $\mathbf{X}_c'\mathbf{X}_c$, where the subscript c stands for column-centered. Actually, $\mathbf{X}_c'\mathbf{X}_c$ is a poor summary of the data set since it does not take account of the abovementioned smooth functionality. Yet a lot of multivariate methods (e.g. PCA) work on the basis of this $\mathbf{X}_c'\mathbf{X}_c$ matrix. Similar examples of contextuality problems can be found in spectroscopy.

This contextuality problem is of course also present in the three-way cases. Methods which take account of contextuality would be helpful, for instance in solving the problem of predicting capacity factors over a continuous range of mobile phase compositions [29]. Creating methods that are 'designed context dependent' is really a challenge.

Homogeneity of the data set is another topic which deserves attention and is related to local rank maps and contextuality. Perhaps a three-way data set should be split into more homogeneous parts and analyzed separately. What kind of measures are suitable to detect heterogeneity?

Nonlinearities, bilinearity, and trilinearity

A topic which has received much attention lately in the chemometric literature is the effect of nonlinearities and how to handle them [61,62]. Harmful effects of nonlinearities within a three-way matrix on the above described three-way methods can be expected. When using the PARAFAC method for prediction, nonlinearities can cause large prediction errors [29]. What kind of nonlinearities can be expected? How can they be detected? And how can the damage they cause be prevented?

Ståhle [44] concludes that the trilinear decomposition is perhaps too restricted. Yet any constraint that can be introduced in the unfolding model may improve its prediction properties. Smilde et al. [29] argue along the same lines. Analogously to continuum regression [63], it is perhaps worthwhile to develop a continuum bi–trilinear decomposition where a parameter α gives

either a bilinear ($\alpha = 0$) or a trilinear ($\alpha = 1$) model. The parameter α can be assessed by cross-validation.

Geometrical interpretations

A lot of the two-way methods described have a geometrical interpretation. Much work has been done, particularly in the area of PCA and PLS [5,6]. The interpretation of latent variables by means of object and loading plots [7] gives valuable insights. The biplot technique [8] can serve as a useful tool to visualize the connection between objects and variables.

For three-way methods it is a fact that plots of loadings for different components in the same mode (single-mode loading plots) are analogous to the usual two-way scores or loading plots and can be used in a similar fashion for interpretation. For instance, for a three-way matrix which is decomposed into two components with the PARAFAC model, three single-mode loading plots can be created, visualizing information in the three different modes.

In the TUCKALS3 program [51] multi-mode loading plots can be created, combining loadings in two different modes. These plots can be used to interpret the connected information in two modes. Further insight into the merits of the single- and multi-mode plots and the informative power of these plots towards the hidden structure in the three-way matrix is necessary.

Preprocessing of data

By preprocessing we mean the operation of centering and scaling the data. In two-way data the combination of these two operations is called autoscaling. For two-way data, column-centering is usually a correct thing to do. Column and row centering (double centering) is useful for particular applications. The spectral map method [11] uses this kind of centering.

For three-way data the kind of centering to be used is not obvious at all. When the three-way matrix is understood as a collection of one-way arrays, centering can be performed over these arrays. This already defines three different centering operations. When the three-way array is understood as a collection of stacked planes, the average of each plane can serve to center the whole plane. Again, this defines three centering operations. There is also the possibility to center the whole three-way array at once, using the overall centroid of the three-way matrix. Some combinations of the above-mentioned centering operation give yet further centered three-way matrices.

When scaling is considered, things get even more complicated. There are also different kinds of scaling and, in combination with the centering operations, the number of possibilities for preprocessing the data is large.

The problem is complicated indeed, and some rules of thumb concerning how to preprocess the data in a particular case would be helpful.

Physicochemical interpretations

A physicochemical interpretation of the calculated components is not easy. In the three-way case, in particular, this is usually difficult. Yet some physicochemical information may perhaps be conceived when interpreting the components correctly and after performing various rotation methods; however, problems do arise.

In the case of a PARAFAC model, is it always necessary to choose orthogonal latent vectors? In the PARAFAC model this is a matter of choice. Or should the latent vectors be uncorrelated? (This is not the same as orthogonal; the equivalence depends on the kind of centering.) For the PARAFAC method it holds that rotating the components gives an essentially different solution.

Rotations are allowed for the components of the Tucker3 model and can facilitate interpretation of the underlying physicochemical model [37]. Yet, knowledge of the underlying physicochemical reality has to be available for a good interpretation. Is it possible to use Tucker3 models for exploratory reasons, i.e. to obtain insight into physicochemical realities?

PROSPECTS

Some of the prospects are already given in the Problems section, i.e. the new three-way methods which can be developed. It is very hard to say in general what the prospects of three-way methods are. Progress in the area of applying multivariate methods can perhaps be compared with progress made in the area of experimental design and optimization. The one-factor-at-a-time method has been replaced by the multi-factor approach [64]. The central idea of the multi-factor approach is to vary all factors simultaneously (but in a systematic fashion) in order to optimize some response depending on these factors. There is a growing tendency to use multivariate analysis, but thinking in terms of multi-way data is still uncommon. Yet multi-way methods are flexible tools for studying the simultaneous variation within a system described by different kinds of variables and/or objects. In the following, different kinds of future applications in different areas are sketched. It is by no means exhaustive; rather, it is meant to trigger ideas in the minds of the readers.

Multivariate statistical process control (MSPC)

Statistical process control is an area of considerable importance in industry. A new development is multivariate statistical process control. Three-way methods can play a role in this area. Take, for example, samples at different points in a process, measuring a number of features of that sample at regularly spaced time intervals. A three-way matrix results that can be analyzed and used to monitor the process.

Hyphenated analysis methods

There is a growing interest in hyphenated analysis techniques. Combining liquid chromatography with diode array detection on different samples [28] is an example of such a technique. Other combinations like gas chromatography–mass spectrometry (GC–MS), LC–LC, LC–MS, GC–IR and inductively coupled plasma–atomic emission spectroscopy are available. It is clear that three-way methods are very suitable to handle problems with interpreting and reducing the data obtained with hyphenated analysis methods.

Chromatography

An attempt is already being made to use three-way methods in the multivariate calibration of chromatographic systems [29,30]. When the problem of contextuality (see the Problems section) is solved, an important step forward will have been taken. Ordinary PCA has proved to give insight into the mechanisms that control retention [65]. It is expected that incorporating the stationary phase as a separate mode will give insight into retention mechanisms as they depend on the kind of stationary phase.

Pharmacology

An example of the use of a trilinear decomposition in pharmacology has already been given (see Theory and Applications and ref. 44). It is not difficult to imagine other applications, e.g. by incorporating biological activities. The screening of potential compounds can also be tackled by three-way methods. Suppose a number of compounds are tested, on different tests with the use of different rats. A three-way matrix results, which can be analyzed to solve the question of which compound is the most potent and selective. Combining in vivo and in vitro experiments may give rise to methods of analyzing two-block, three-way matrices.

Quantitative structure–activity relationships (QSAR)

With the development of sophisticated software which calculates the electron density distribution of the minimum energy conformation of a compound [66], a three-way matrix is available that can serve as a blueprint of the compound. How should this three-way matrix be treated? Should it be unfolded or strung out, or used as such? It is clear that incorporating this kind of data in QSAR studies calls for three-way meth-

ods, where the problem sketched in Fig. 12 is no longer far-fetched.

Simulation and random noise testing

Since the underlying structure of real data is generally unknown, simulations can serve as a surrogate to investigate the effect of properties of such an underlying structure. The effect of the degree of bi- and trilinearity of the underlying structure may be simulated. For example, generate elements $x_{ijk} = \Sigma t_{ig} p_{jkg}$ where the summation runs from $g = 1$ to G. These elements can form a three-way matrix $\underline{X}$ and this matrix can be analyzed by three-way models. Varying the rank of the generating loading matrices $\mathbf{P}_1$ to $\mathbf{P}_G$ gives varying bi- and trilinear properties of X (e.g. rank-one matrices $\mathbf{P}_1$–$\mathbf{P}_G$ give trilinearity). Of course, random noise should be added to the generating ts and $\mathbf{P}$s in order to mask the varying bi- and trilinearity. The consequences of the different properties of the underlying structure and of the noise level for the performance of the different three-way models can be evaluated.

The consequences of different kinds and different degrees of nonlinearities is another area in which simulation may be helpful. A systematic approach in which designed nonlinearities are simulated and tested for their consequences on three-way methods can be illuminating. Again, noise levels can be varied to investigate the effect of the noise.

For both of the applications of simulations in three-way data analysis described above, it holds that not only the consequences on the descriptive and predictive power of the three-way methods should be assessed. The validation of diagnostic values attached to the three-way methods is very important. For example, is it possible to develop diagnostic tools to check for bilinearity or nonlinearity, and how do these diagnostics behave under the varying simulated conditions?

Process analytical chemistry

One of the latest developments in analytical chemistry is process analytical chemistry [67]. Inline sensors give complex and noisy data that can perhaps be analyzed with three-way methods in order to measure concentration levels of different substituents in a flow.

ACKNOWLEDGEMENTS

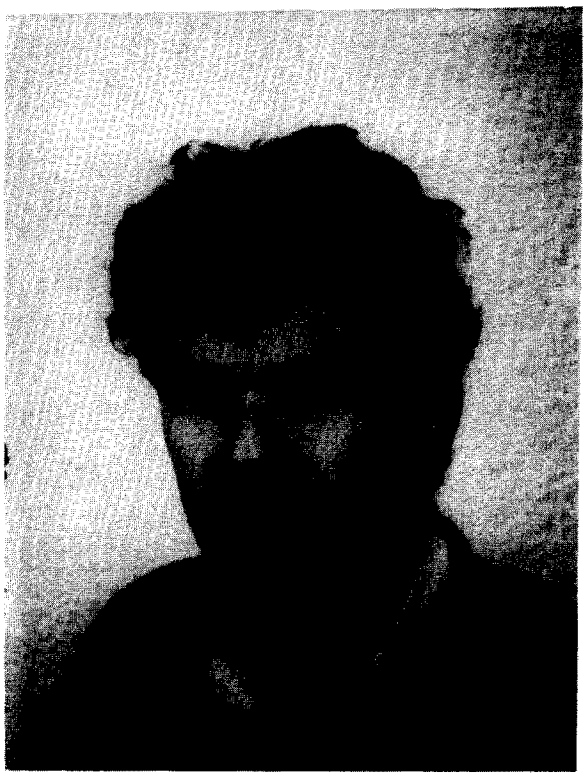
Thanks go to Paul Geladi, University of Umeå, Sweden, for helpful comments on an earlier version of this paper.

REFERENCES

1 K.H. Esbensen, S. Wold and P. Geladi, Relationship between higher-order data array configurations and problem formulations in multivariate data analysis, *Journal of Chemometrics*, 3 (1988) 33–48.
2 I.T. Jolliffe, *Principal Component Analysis*, Springer, New York, 1986.
3 K.V. Mardia, J.T. Kent and J.M. Bibby, *Multivariate Analysis*, Academic Press, London, 1979.
4 K.R. Gabriel and S. Zamir, Lower rank approximation of matrices by least squares with any choice of weights, *Technometrics*, 21(4) (1979) 489–498.
5 S. Wold, K. Esbensen and P. Geladi, Principal component analysis, *Chemometrics and Intelligent Laboratory Systems*, 2 (1987) 37–52.
6 O.M. Kvalheim, Latent-structure decompositions (projections) of multivariate data, *Chemometrics and Intelligent Laboratory Systems*, 2 (1987) 283–290.
7 O.M. Kvalheim, Interpretation of latent-variable regression models, *Chemometrics and Intelligent Laboratory Systems*, 7 (1989) 39–51.
8 K.R. Gabriel, The biplot graphic display of matrices with applications to principal component analysis, *Biometrika*, 58 (1971) 453–467.
9 S. Wold, Cross-validatory estimation of the number of components in factor and principal component analysis, *Technometrics*, 20(4) (1978) 397–405.
10 J.L. Horn, A rationale and test for the number of factors in factor analysis, *Psychometrika*, 30 (1965) 179–186.
11 P.J. Lewi, Spectral map analysis: factorial analysis of contrasts, especially from log ratios, *Chemometrics and Intelligent Laboratory Systems*, 5 (1989) 105–116.
12 M.J. Greenacre, *Theory and Applications of Correspondence Analysis*, Academic Press, New York, 1984.
13 T.W. Anderson, *An Introduction to Multivariate Statistical Analysis*, Wiley, 2nd edn., 1984.
14 S. Wold, A theoretical foundation of extrathermodynamic relationships, *Chemica Scripta*, 5 (1974) 97–106.
15 S. Wold, Pattern recognition by means of disjoint principal component models, *Pattern Recognition*, 8 (1976) 127–139.

16 D. Coomans and I. Broeckaert, *Potential Pattern Recognition in Chemical and Medical Decision Making*, Wiley, New York, 1986.
17 H. van der Voet, P.M.J. Coenegracht and J.B. Hemel, New probabilistic versions of the SIMCA and CLASSY classification methods. Part 1. Theoretical description, *Analytica Chimica Acta*, 192 (1987) 63–75.
18 G.P. McCabe, Principal variables, *Technometrics*, 26 (1984) 137–144.
19 G.E.P. Box and N.R. Draper, *Empirical Model-Building and Response Surfaces*, Wiley, New York, 1987.
20 H. Martens and T. Næs, *Multivariate Calibration*, Wiley, Chichester, 1989.
21 A.E. Hoerl and R.W. Kennard, Ridge regression: biased estimation for non-orthogonal problems, *Technometrics*, 12(1) (1970) 55–67.
22 W.J. Krzanowski, *Principles of Multivariate Analysis: A User's Perspective*, Clarendon, Oxford, 1988.
23 M. Martens and H. Martens, Partial least squares regression, in *Statistical Procedures in Food Research*, Elsevier Applied Science, London, 1986, pp. 293–360.
24 J.C. Gower, Generalized procrustes analysis, *Psychometrika*, 40(1) (1975) 33–51.
25 J.B. Lohmuller; *Latent Variable Path Modeling with PLS*, Springer, New York, 1989.
26 K.G. Jöreskog and H. Wold (Editors), *Systems Under Indirect Observation*, Parts I and II, North-Holland, Amsterdam, 1982.
27 P. Geladi, H. Isaksson, L. Lindqvist, S. Wold and K. Esbensen, Principal component analysis on multivariate images, *Chemometrics and Intelligent Laboratory Systems*, 5 (1988) 209–220.
28 J. Öhman, P. Geladi and S. Wold, Residual bilinearization. Part 2. Application to HPLC–DAD data and comparison with rank annihilation factor analysis, *Journal of Chemometrics*, 4 (1990) 135–146.
29 A.K. Smilde, P.H. van der Graaf, D.A. Doornbos, T. Steerneman and A. Sleurink, Multivariate calibration of reversed-phase chromatographic systems. Some designs based on three-way data analysis, *Analytica Chimica Acta*, 235 (1990) 41–51.
30 A.K. Smilde and D.A. Doornbos, Three-way methods for the calibration of chromatographic systems: comparing PARAFAC and three-way PLS, *Journal of Chemometrics*, 5 (1991) 345–360.
31 L. Tucker, Implications of factor analysis of three-way matrices for measurement of change, in C. Harris (Editor), *Problems of Measuring Change*, University of Wisconsin Press, Madison, WI, 1963, pp. 122–137.
32 H.G. Law, C.W. Snyder, J.A. Hattie and R.P. McDonald (Editor), *Research Methods for Multimode Data Analysis*, Praeger, New York, 1984.
33 P. Kroonenberg, *Three-Mode Principal Component Analysis: Theory and Applications*, DSWO Press, Leiden, 1983.
34 J.L. Lastovicka, The extension of component analysis to four-mode matrices, *Psychometrika*, 46(1) (1981) 47–57.
35 A. Kapteyn, H. Neudecker and T.J. Wansbeek, An approach to *n*-mode component analysis, *Psychometrika*, 51 (1986) 269–275.
36 C. de Ligny, M. Spanjer, J. van Houwelingen and H. Weesie, Three-mode factor analysis of data on retention in normal-phase high-performance liquid chromatography, *Journal of Chromatography*, 301 (1984) 311–324.
37 Y. Zeng and P.K. Hopke, Methodological study applying three-mode factor analysis to three-way chemical data sets, *Chemometrics and Intelligent Laboratory Systems*, 7 (1990) 237–250.
38 J. Carroll and J.J. Chang, Analysis of individual differences in multidimensional scaling via an *N*-way generalization of the Eckart–Young decomposition, *Psychometrika*, 35 (1970) 283–319.
39 R.A. Harshman, Foundations of the PARAFAC procedure: models and conditions for an exploratory multimodal factor analysis, *UCLA Working Papers in Phonetics*, 16 (1970) 1–84.
40 E. Sanchez and B.R. Kowalski, Tensorial calibration. II. Second order calibration, *Journal of Chemometrics*, 2 (1988) 265–280.
41 S. Wold, P. Geladi, K. Esbensen and J. Öhman, Multi-way principal components and PLS analysis, *J. Chemometrics*, 1 (1987) 41–56.
42 R. Henrion, G. Henrion, P. Heininger and G. Steppuhn, Three-way principal component analysis for multivariate evaluation of round robin tests, *Fresenius' Zeitschrift für Analytische Chemie*, 336 (1990) 37.
43 K. Esbensen and P. Geladi, Strategy of multivariate image analysis (MIA), *Chemometrics and Intelligent Laboratory Systems*, 7 (1989) 67–86.
44 L. Ståhle, Aspects of the analysis of three-way data, *Chemometrics and Intelligent Laboratory Systems*, 7 (1989) 95–100.
45 P. Geladi and K. Esbensen, Regression on multivariate images: principal component regression for modeling, prediction and visual diagnostic tools, *Journal of Chemometrics*, 5 (1991) 97–111.
46 A. Höskuldsson, PLS regression methods, *Journal of Chemometrics*, 2 (1988) 211–220.
47 R. Harshman, *PARAFAC*, Department of Psychology, University of Western Ontario, London, Ontario N6A 5C2, Canada.
48 A.K. Smilde, unpublished results, 1991.
49 E. Sanchez and B.R. Kowalski, Tensorial resolution: a direct trilinear decomposition, *Journal of Chemometrics*, 4 (1990) 29–45.
50 J. Weesie and J.H. van Houwelingen, *GEPCAM User's Manual*, Institute of Mathematical Statistics, State University of Utrecht, Boedapestlaan 6, 3584 CD Utrecht, Netherlands.
51 P. Kroonenberg and P. Brouwer, *User's guide to TUCKALS3, Version 4.0*, Section W.E.P., Department of Education, University of Leiden, P.O. Box 9507, 2300 RA Leiden, Netherlands.

52 P. Kroonenberg, personal communication, 1990.
53 SAS Institute Inc., SAS Circle, Box 8000, Cary, NC 27512-8000, USA.
54 P. Geladi, Analysis of multi-way (multi-mode) data, *Chemometrics and Intelligent Laboratory Systems*, 7 (1989) 11–30.
55 A.K. Smilde, C.H.P. Bruins, P.M.J. Coenegracht and D.A. Doornbos, Characterization of stationary/mobile phase combinations by using markers. Part I. Preliminary results for the prediction of RP retention on various phases, *Analytica Chimica Acta*, 212 (1988) 95–104.
56 A. Lorber and B.R. Kowalski, Estimation of prediction error for multivariate calibration, *Journal of Chemometrics*, 2 (1988) 93–109.
57 P.M. Bentler and S.Y. Lee, Statistical aspects of a three-mode factor analysis model, *Psychometrika*, 43(3) (1978) 343–352.
58 S. Weisberg, *Applied Linear Regression*, Wiley, New York, 1985.
59 P. Geladi and S. Wold, Local principal component models, rank maps and contextuality for curve resolution and multi-way calibration inference, *Chemometrics and Intelligent Laboratory Systems*, 2 (1987) 273–281.
60 P.J. Schoenmakers, *Optimization of Chromatographic Selectivity*, Elsevier, Amsterdam, 1986.
61 S. Wold, N. Kettaneh-Wold and B. Skagerberg, Nonlinear PLS modeling, *Chemometrics and Intelligent Laboratory Systems*, 7 (1989) 53–65.
62 I. Frank, A non-linear PLS model, *Chemometrics and Intelligent Laboratory Systems*, 8 (1990) 109–119.
63 M. Stone and R.J. Brooks, Continuum regression: Cross-validated sequentially constructed prediction embracing ordinary least squares, partial least squares and principal component regression, *Journal of the Royal Statistical Society B*, 52(2) (1990) 237–269.
64 G.E.P. Box, W.G. Hunter and J.S. Hunter, *Statistics for Experimenters, An Introduction to Design, Data Analysis, and Model Building*, Wiley, New York, 1978.
65 P.M.J. Coenegracht, A.K. Smilde, H. Benak, C.H.P. Bruins, H.J. Metting, H. de Vries and D.A. Doornbos, Multivariate characterization of solvent strength and solvent selectivity in RP–HPLC, *Journal of Chromatography*, 550 (1991) 397–410.
66 R.D. Cramer III, D.E. Patterson and J.D. Bunce, Comparative molecular field analysis (COMFA) 1. Effect of shape on binding of steroids to carrier proteins, *Journal of the American Chemical Society*, 110 (1988) 5959–5967.
67 M.T. Riebe and D.J. Eustace, Process analytical chemistry, an industrial perspective, *Analytical Chemistry*, 62(2) (1990) 65A–71A.



BIOGRAPHICAL SKETCH

Age Smilde received his Master's Degree in Econometrics from the University of Groningen (Netherlands) in 1984. At the same University he received his Ph.D. in Analytical Chemistry in 1990. At present he is staff-member at the University Centre for Pharmacy in Groningen at the Department of Analytical Chemistry & Toxicology, Research Group Chemometrics, where his daily activities include: supervising Ph.D. students, teaching courses on (multivariate) statistics and optimization, and doing research on higher-order (multi-way) methods. Currently, he is a temporarily visiting scientist at CPAC, Seattle, WA, doing research on multi-way methods. His leisure activities are a.o. running, playing the piano and listening to music.