*Chemometrics and Intelligent Laboratory Systems*, 7 (1989) 11–30
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

# Analysis of Multi-Way (Multi-Mode) Data

PAUL GELADI

*Research Group for Chemometrics, Umeå University, S-901 87 Umeå (Sweden)*

(Received 5 October 1988; accepted 20 February 1989)

CONTENTS



## ABSTRACT

Geladi, P., 1989. Analysis of multi-way (multi-mode) data. *Chemometrics and Intelligent Laboratory Systems*, 7: 11–30.

The concepts data matrix and multivariate data analysis are rapidly becoming popular and well-known words in chemistry. Many methods used in the laboratory can produce data arrays of a greater complexity than the data matrix. The broad picture easily gets lost here, not least because of the confusing nomenclature. There is a need for systematization and generalization.

0169-7439/89/$03.50 

Methods available in psychometrics and methods used in chemical research are described and compared in this paper. The goal is to provide a systematic overview and a simple introduction to the subject. References are made to more detailed descriptions in the literature.

## 1 INTRODUCTION

Data analysis is an essential part of all experimentation. Data are measured on objects, samples or cases. For the selection of objects a problem definition has to be present; experiments and data analysis are then adapted to the problem definition. The goal is to look for systematic variation in the data and try to relate this variation to the original problem definition. In some cases the original problem definition can be modified: sometimes it is discarded, sometimes it is improved and sometimes a whole new problem definition arises.

For example, if the objects are lakes the formulation of the problem could be: "Are some lakes more acid than the others?" or "Have these lakes been less acid in the past?". The experiment may then consist of obtaining samples and measuring their pH. The data analysis stage only starts when a table of lake names and their corresponding pH values can be studied.

Data analysis is univariate when only one variable is measured systematically for many objects. Univariate statistics has been applied to the study of this type of data analysis for a long time and all introductory courses in statistics contain a multitude of univariate methods.

In the example of the lakes, one could use a *t*-test for comparing lakes or groups of lakes, or one could carry out out a trend analysis over time.

A question to be asked here is whether univariate methods really exist. Is it not a fact that scientists, even when they are using only one measured variable, also have a large amount of external information in their heads and use this information for producing their final conclusions?

In recent decades, multivariate data analysis has been introduced into chemistry and many chemists are starting to understand and use multivariate ideas. Almost everybody knows what a data matrix is and how to extract some useful information from it. One of the most interesting techniques here is bilinear decomposition (eigenanalysis, principal component analysis, singular value decomposition) [1]. This method requires no a priori information or constraints or conditions, except that the residual has a minimal sum of squares: structure originates in the data matrix itself. The structures found with bilinear methods are associated with the geometry of multivariate space. This makes it easy to visualize the results obtained. In the case of the lakes, more variables than just the pH could be measured. Examples are ammonia, sulphate and phosphate contents. Knowledge about chemical equilibria tells us that these are not uncorrelated; neither are they uncorrelated with pH. Multivariate data analysis depends heavily on the correlation between variables. Bilinear decomposition of a data matrix produces latent variables [2]. These can be viewed as summary variables derived from the original ones, which are called manifest variables. Sometimes a meaning can be attributed to the latent variables. One could for example obtain eutrophication or pollution latent variables that would mean more than each of the manifest variables alone.

The question that can now be asked is: "What next?". It is possible to create arrays that are more complex than just numbers spread out in a planar array. Volumes in three dimensions, and also in more dimensions, can be used as data arrays [3]. Most of these contain data matrices as subunits. Techniques in analytical chemistry for producing these types of data arrays are becoming more common. Methods for analyzing multiway data arrays have also been proposed. Many of them are still based on those used for a 2-way data matrix. It is of course necessary to retain a downward compatibility with data matrix methods. Techniques as general as bilinear decomposition of a data matrix are still unavailable for higher-order arrays. General trilinear decomposition seems to pose great problems except in trivial cases.

In the case of the lakes, one could combine the

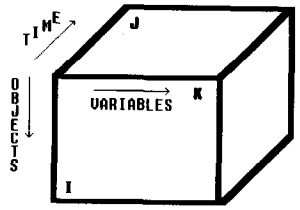


Fig. 1. The 3-way or 3-mode data array represented by a parallelepiped of size $I \times J \times K$. Parallelepipeds of this type were first proposed for data analysis by Cattell in 1952 [20].

many variables and their variations over time in a 3-way array (see Fig. 1). This array would have as axes: lakes (objects), variables (pH, chemicals, ions) and time. The following 2-way arrays can be extracted from this 3-way array:
- objects (lakes) × variables at each specific time
- variables × time for each specific object (lake)
- objects (lakes) × time for each variable

The purpose of this article is to present the current situation regarding such methods as are available, and the laboratory techniques that require these methods. At the present moment there seems to be a tendency to diversification: methods are proposed for specific types of multi-order arrays, but this opens the way to more generalized methods for the future. The emphasis is on 3-way data arrays, with excursions into 4-way and higher generalizations. Most of the techniques presented are based on linear models.

The methods studied are: 3-way factor analysis from psychometrics, 3-way generalizations of principal component analysis and regression, rank annihilation methods, consensus factor analysis and multivariate image analysis. The usual nomenclature, boldface lower case italic for vectors and boldface uppercase roman for matrices, is used. Three-way arrays are shown in boldface uppercase with underscore. The terms 'way' and 'mode' are given their standard meanings. A matrix is a 2-way or 2-mode array. Other nomenclature will be defined in the text. The term 'elements of a mode' is very useful. For an object mode, the elements of the mode are individual objects, samples or cases. For a variable mode, the elements of the mode are individual variables. For a time mode, as in Fig. 1, time intervals are elements of the time mode.

## 2 MANY-WAY DATA IN CHEMISTRY

It has already been mentioned that measurement and data analysis in analytical chemistry are becoming increasingly complex. The name hyphenated methods was added to the vocabulary some years ago [4]. The methods described by this name all provide at least 3-way data with one sample mode and two variable modes. This is an $OV^2$ array. Most popular is the combination of some type of chromatography with a multivariate, often spectral detector. More about this can be found in refs. 5–7.

The analysis of $OV^2$ arrays can have many different goals. One may be interested in calculating concentrations of overlapping chemical constituents for every object. This is a calibration situation. The goal may be to find pure spectra or chromatograms. This is curve resolution. Some methods combine the calibration and curve resolution aspect, while others specialize in either calibration or curve resolution. The sheer amount of data causes complications, but the special properties of the 3-way arrays make some aspects of analysis easier. More about this can be found in Section 7.

In physical chemistry important progress was made in NMR, with the step from 1D- to 2D- [8,9] and recently 3D-NMR [10]. In NMR studies it is often not the concentration of chemical constituents that is desired. The goal is to assign peaks to atoms, often in the presence of disturbing background, and to calculate the structure of molecules. Three-way arrays are also obtained in theoretical modeling in chromatography [11].

Image analysis in its simplest form is a technology producing large 2D arrays of data. More complex data arrays are 3D images and multivariate 2D or 3D images. Multivariate 2D images are satellite images ($O^2V$ arrays). Tomographic techniques using X-rays or NMR can produce 3D images. With NMR tomography, it is easy to produce multivariate 2D and 3D images ($O^2V$ and $O^3V$ arrays). Refs. 12 and 13 describe the use of image analysis as an analytical instrument able to provide chemical data interpretation. The goal of image analysis is often classification of parts of the image. Another goal may be calibration against

external data with subsequent prediction. More about this can be found in Section 5.

The relation between psychometric data and chemical data may be found in pharmaceutical and food chemistry. In this case, it is normal for the psychometric data to be 3-way. For food studies this was described in ref. 14. More about this can be found in Section 6.

## 3 THREE-WAY AND MULTI-WAY ANALYSIS IN PSYCHOMETRICS

In mathematics, multilinear algebra has existed for many years, but the first scientists to use data arranged in a multi-way array for data analysis were psychometricians.

### *3.1 On the hierarchy of 3-way methods*

There are different ways of looking at a 3-way array. A hierarchy is given in Fig. 2. It is possible to look at the elements of a mode as treatments in a multi-way experimental design. A good explanation of this was given by Lohmöller in 1979 [15]. Also, in Chapter 2 of ref. 16, Kruskal describes 3-way analysis as an experimental design model. ANOVA allows one to estimate the importance of main effects (modes), interactions (mode × mode) and interactions (model × mode2 × mode3) [17]. The method may be used for a hypothetical example, comparing laboratories that have analysed the same chemical compounds in the same set of samples. The elements of the ANOVA table for this hypothetical example are given in Table 1.

In this example, one would expect the main effect of the laboratories to be small since they are all assumed to measure the same results. If this is not the case then there are some deviations. These then also show up as interactions of laboratories × analytes or laboratories × samples.

Experimental designs are useful when some a priori knowledge about the system under study is available and a design can be carried out.

When a more detailed knowledge of the relationship between the elements of a mode is required, explatory 3-way analysis is used. Here the distinction can be made between using scaling methods and using natural scales. In the natural sciences, natural scales are the usual way to measure things. Differences in pH, temperature, or concentrations are tangible. But this is not the case in the life sciences, where a variable may be a score from 1 to 5 indicating a preference, or else an answer to a question with just a yes/no answer.

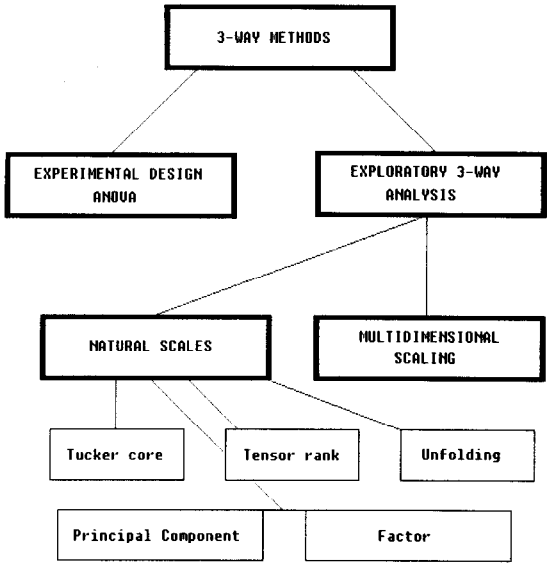


Fig. 2. The hierarchy of 3-way methods. Experimental designs and multidimensional scaling are not dealt with in this paper. An important subdivision for natural scales is in principal component or factor analysis. The other subdivision emphasized most in this paper is that in 'Tucker core', 'tensor rank' or 'canonical decomposition' and 'unfolding' models.

TABLE 1

Main effects and interactions in multi-way ANOVA

| Name | Chemical example |
|---|---|
| mode 1: observers | laboratories |
| mode 2: attributes | analytes |
| mode 3: objects | samples |
| mode 1 × mode 2 | labs × analytes |
| mode 2 × mode 3 | analytes × samples |
| mode 3 × mode 1 | samples × labs |
| mode 1 × mode 2 × mode 3 | labs × analytes × samples |

The apparatus of multidimensional scaling was developed in order to find useful scales for life-science data. This paper will not go into these methods: it is assumed that meaningful natural scales are available and that the measured values can be expressed as reals. In the natural sciences, a small number of binary or categorical variables is sometimes used together with naturally scaled variables. This does not seem to pose any great problem.

When natural scales are used, one has the choice between three fundamentally different models: the 'Tucker core' model, the 'tensor rank' or 'canonical decomposition' model, and the 'unfolding' model. These will be explained further on in the text. It will be shown that a mental transition between them is possible.

Another distinction that can be made is that between principal component analysis and factor analysis [18,19]. In principal component analysis, structures are found in the data itself. In factor analysis the relation between the data and a priori known (or assumed) structures is studied. Also, the distribution of error residuals is often assumed to be known a priori in factor analysis. The distinction between these two models is not always very clear in the scientific literature.

### *3.2 On the way equations are written*

Equations can be expressed in three ways:

- as matrix and vector equations, which is a simplification that makes comprehension easier. There is, however, no 'right' notation where many-way arrays are concerned.
- as summations of indexed quantities. This notation is extremely confusing to look at when many indices are involved but it is rather strict and it reduces the chances of misinterpretation.
- as figures in which the vectors are lines, matrices are rectangles and 3-way arrays are parallelepipeds. This is a very instructive notation, but is restricted to maximal 3 ways.

Two new symbols are introduced: $*$ the Kronecker product and $.*.$, the Kronecker summation product. These are explained in the Appendix, part A.1 and also in ref. 47.

### *3.3 Some historical psychometric work: Cattell, Tucker*

The volume of psychometric literature on many-way tables is very large. Many papers do not describe material that is useful outside psychology. Some key papers only are therefore highlighted here.

As early as 1952, an article by Raymond Cattell appeared [20] describing factor analysis as a multi-way activity. According to some authors, Cattell had started thinking about these problems since the middle' forties. Cattell starts by observing that "All scientific method deals with observations of covariation,...". He then gives the five modes of factor analysis as:

(1) circumstances; such as place or time of the observation.
(2) attribute; the thing measured, response.
(3) object; on which the attribute is measured.
(4) scale; this mode refers to a typical problem in the life sciences where different scales have to be tested and compared. When a natural scale is available this mode is nullified.
(5) observer; when more than one observer is active.

For practical purposes Cattell reduces this to a 3-way table with objects or persons, occasions or circumstances and attributes or variables as ways. This gives six ways of studying correlations called *R, P, Q, O, T* and *S* and they are described in the rest of the article [20]. This early article shows a very confusing nomenclature and the mathematics is not very well explained. The Cattell parallelepiped is shown in Fig. 1.

Some early systematization in multi-way psychology was done by Ledyard Tucker [21,22]. He gives an artificial example of 12 individuals $\times$ 9 traits $\times$ 5 raters and places the data in a parallelepiped. He also mentions the problem with the double use of the word 'dimension'. He suggests the words 'mode' and 'mode of classification' for the ways. The indices used are: $i = 1$ to $I$ for individuals, $j = 1$ to $J$ for traits (variables) and $k = 1$ to $K$ for raters (observers).

The 3-mode data array $\underline{\mathbf{X}}$ has the size $I \times J \times K$. It is decomposed into smaller arrays:

| | |
|---|---|
| $\mathbf{U}$ $I \times A$ | loadings (or scores) for individuals $A < I$ |
| $\mathbf{V}$ $J \times B$ | loadings (or scores) for traits $B < J$ |
| $\mathbf{W}$ $K \times C$ | loadings (or scores) for raters $C < K$ |
| $\underline{\mathbf{G}}$ $A \times B \times C$ | the Tucker core matrix: a 3-way array. |

The symbols U and V were chosen for their analogy with singular value decomposition notation, see eq. 2, below. It is then natural to use W for a third way. $\underline{\mathbf{G}}$ was proposed by Tucker.

The model can be written as:

$$x_{ijk} = \sum_{a=1}^{A} \sum_{b=1}^{B} \sum_{c=1}^{C} u_{ia} v_{jb} w_{kc} g_{abc} + e_{ijk} \qquad (1a)$$

or:

$$\underline{\mathbf{X}} = \mathbf{U}\underline{\mathbf{G}}\mathbf{V}\mathbf{W} + \underline{\mathbf{E}} \qquad (1b)$$

See also Fig. 3.

$\underline{\mathbf{E}}$ is a 3-way array of size $I \times J \times K$ of residuals.

The sequential way of writing equations such as eq. 1b is confusing. A better version would be:

$$\underline{\mathbf{X}} = \mathbf{U}\underline{\mathbf{G}}\overset{\mathbf{V}}{\mathbf{W}} + \underline{\mathbf{E}} \qquad (1c)$$

Tucker calls this the 'Fundamental Model'. Although eq. 1c poses some typographical problems, it is good to remember it for didactical reasons.

The Tucker decomposition can be used to separate $\underline{\mathbf{X}}$ into a structural model part and a noise part. One can also choose the model so that $\underline{\mathbf{E}}$ becomes filled with zeros. A very important aspect is that $A$, $B$ and $C$ do not have to be equal. This is a general decomposition.

The first Tucker articles are very confusing, because nomenclature and matrix algebra are not well-defined. The same decomposition appeared in an article by Levin in 1965 [23]. The decomposition as proposed by Tucker and Levin is shown in Fig. 3. A general program for this decomposition is called TUCKALS3 [16,24,25]. A program called TUCKALS2 [16,25,26] is often used for the same decomposition where the dimensionality in one of the modes is not reduced (in the core matrix). This can be useful in these

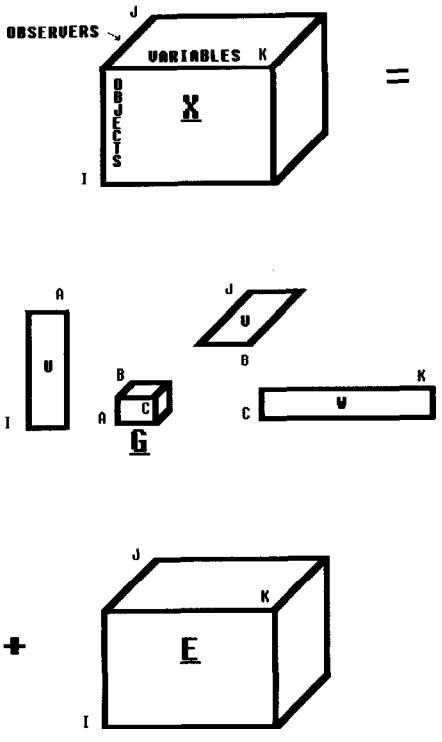


Fig. 3. The general decomposition given by Tucker in 1963 and by Levin in 1965. $\underline{\mathbf{X}}$ can be viewed as the sum of a structure part and a noise part $\underline{\mathbf{E}}$. The matrices U, V and W contain loadings explaining the ideal elements of their modes as functions of the natural elements. $\underline{\mathbf{G}}$ is called the core matrix. It describes the relations between the idealized elements of the modes. One of the goals of the model is data reduction.

circumstances when one of the modes is $I$ time slices. In this case it may not be useful to make a reduction from $I$ (eq. 1) to something less [25]. Note that the names TUCKALS3 and TUCKALS2 have no strict relationship to the names Method I, Method II and Method III given by Tucker. TUCKALS3 is based on a more advanced version of the Tucker Method I algorithm [24], TUCKALS2 is a more advanced version resembling the Tucker method III algorithm.

Perhaps the most intriguing part of the Tucker decomposition is the core matrix. There is an analog to be found in the data analysis of 2-way arrays. This is in singular value decomposition (SVD):

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}' \qquad (2)$$

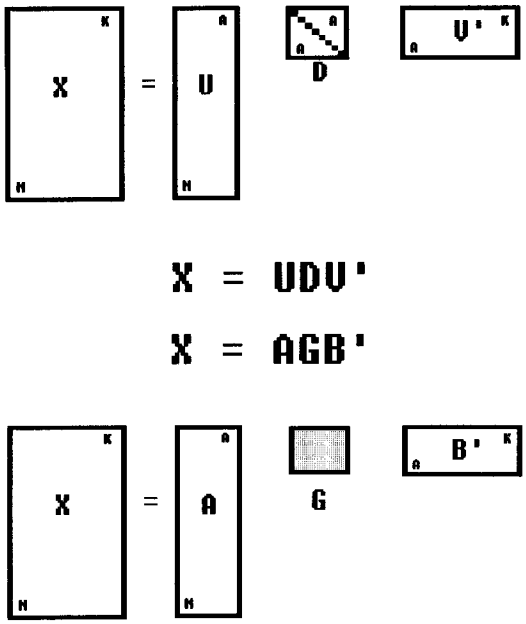


Fig. 4. The singular value decomposition, also called Eckart–Young decomposition in the psychometrical literature, of a matrix **X** of $K$ variables measured on $N$ objects. The diagonal matrix **D** becomes a core matrix **G** when **U** and **V** are rotated towards **A** and **B** respectively.

Here **X** is a data matrix of $N$ objects and $K$ variables; **U** is an orthonormal matrix of scores, size $N \times A$; **D** is a diagonal matrix of size coefficients, size $A \times A$; and **V** is an orthonormal matrix of loadings, size $K \times A$.

The singular value decomposition is also shown in Fig. 4. Here, **D** can be seen as a kind of core matrix. In the psychometric literature the singular value decomposition is also called the Eckart–Young decomposition, after an article that appeared in 1936 [27].

Levin explains how a simple rotation can transform **D** into a non-diagonal core matrix, using rotation matrices $\mathbf{S'S} = \mathbf{I}$ and $\mathbf{T'T} = \mathbf{I}$.

$$\mathbf{X} = \mathbf{U(S'S)D(T'T)V'} = \mathbf{(US')(SDT')(TV')} = \mathbf{AGB} \quad (3)$$

where **A** is a matrix of rotated factor scores; **B** is a matrix of rotated factor loadings; and **G** is a non-diagonal core matrix.

This can be seen in Fig. 4. A look at this figure shows that the matrix **A** gives the relation between the natural objects and some idealized objects, measured in latent variables. The matrix **B** gives the relations between the natural variables and the idealized, latent variables. The core matrix **G** gives the relation between idealized objects and idealized variables.

With this knowledge, one can go back to Fig. 3 and eq. 1 and explain the Tucker decomposition. Levin [23] calls the elements of **U**, **V** and **W** 'loadings' and the elements of $\underline{\mathbf{G}}$ 'scores'. The matrix **U** gives relations between the natural and idealized objects, the matrix **V** does this for variables and the matrix **W** does this for observers. The core matrix **G** gives the 3-way relations between idealized objects, idealized observers and idealized variables. The term idealized is of course interchangeable with 'latent'.

### *3.4 Simplifications of the Tucker model*

An important aspect of the Tucker core matrix is how it can be simplified and what kinds of methods and ideas correspond to these simplifications. The following simplifications can be made:

1. $A = B = C$. The core matrix becomes a cube of size $A \times A \times A$.
2. as in point 1 and the core matrix becomes diagonal.
3. as in point 2 but only ones on the diagonal, an identity core matrix:

$$\begin{aligned} g_{abc} = d_{abc} &= 1 \quad \text{for } a = b = c \qquad a = 1 \ldots A \\ &= 0 \quad \text{otherwise} \qquad b = 1 \ldots A \\ & \qquad\qquad\qquad\quad c = 1 \ldots A \end{aligned} \quad (4)$$

$d_{abc}$ is the 3-way Kronecker delta.

4. The core matrix is not calculated, but left multiplied with two of the loading matrices. Only one of the loading matrices is calculated explicitly. This is the same as keeping some of the indices in the combined mode. The technique has also been called unfolding, because it is the same as slicing $\underline{\mathbf{X}}$ along one of the modes and putting the obtained 2-way arrays side by side to form a large 2-way array. More about this can be found further on in the text (Section 4).

These simplifications are the basis for a general interpretation of multi-way methods as Tucker models. The Tucker fundamental model can be called type zero simplification.

### *3.5 Three-mode factor analysis: the Tucker–Levin algorithm*

Levin [23] defines three products of the array $\underline{\mathbf{X}}$ with its own transposes as follows:

$$\underline{\mathbf{X}}\underline{\mathbf{X}}' = \mathbf{M} = \sum_{j=1}^{J}\sum_{k=1}^{K} x_{ijk}x_{ijk} \tag{5}$$

$$\underline{\mathbf{X}}'\underline{\mathbf{X}}'' = \mathbf{N} = \sum_{k=1}^{K}\sum_{i=1}^{I} x_{ijk}x_{ijk} \tag{6}$$

$$\underline{\mathbf{X}}''\underline{\mathbf{X}} = \mathbf{P} = \sum_{i=1}^{I}\sum_{j=1}^{J} x_{ijk}x_{ijk} \tag{7}$$

where $\mathbf{M}$ is a symmetric matrix of size $I \times I$; $\mathbf{N}$ is a symmetric matrix of size $J \times J$; and $\mathbf{P}$ is a symmetric matrix of size $K \times K$.

Principal components, eigenvectors or factors of $\mathbf{M}$, $\mathbf{N}$ and $\mathbf{P}$ can then be used to fill $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{W}$, e.g.:

$$\mathbf{M} = \mathbf{UDU}' + \mathbf{E} \tag{8}$$

where $\mathbf{U}$ is a matrix of eigenvectors of $\mathbf{M}$, size $I \times A$ and $\mathbf{D}$ is a diagonal matrix of eigenvalues, size $A \times A$.

$A$ can be chosen to have a value between 1 and $I$. The choice influences the final solution for $\underline{\mathbf{G}}$ (eq. 1). One of the main goals is data reduction, so the value of $A$ would be chosen so that $\underline{\mathbf{E}}$ in eq. 8 describes mainly noise.

Equations similar to eq. 8 can be written for $\mathbf{N}$ and $\mathbf{P}$ to find values for $\mathbf{V}$ and $\mathbf{W}$. $B$ and $C$ can be chosen in a similar way to $A$. Besides that, rotated versions of $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{W}$ can be used [22]. This makes the decomposition very flexible, a true factor analysis type decomposition. Once $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{W}$ are chosen, one can calculate $\underline{\mathbf{G}}$. This is explained in Part A.2 of the Appendix.

An important property is the limitation of the number of factors. Eqs. 5, 6 and 7 show that $A$ is limited by $I$, $B$ by $J$ and $C$ by $K$.

A more advanced and detailed artible by Tucker appeared in 1966 [28]. In this article Tucker gives a better definition of the nomenclature. The Kronecker product and some of its properties are introduced as an aid for writing 3-mode equations.

Kroonenberg [25] gives some properties of the abovementioned algorithm, also called Tucker's method I. Tucker's method II is just a numerical trick to save computer memory when the matrices in eqs. 5, 6 or 7 become too large. When $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{W}$ in eq. 9 represent the full ranks of the matrices $\mathbf{M}$, $\mathbf{N}$ and $\mathbf{P}$ (eqs. 5, 6 and 7) then the core matrix $\underline{\mathbf{G}}$ obtained is a least squares estimator of the true $\underline{\mathbf{G}}$, with property:

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} x_{ijk} = \sum_{a=1}^{A}\sum_{b=1}^{B}\sum_{c=1}^{C} g_{abc} \tag{9}$$

If lower ranks are used to partition the data into a structural part and a noise part (eq. 8) the $\underline{\mathbf{G}}$ obtained is not the least squares estimator. Kroonenberg [25] gives an algorithm for obtaining a least squares estimate:

$$\mathrm{SS}\left(\underline{\mathbf{X}} - \mathbf{U}\,\underline{\mathbf{G}}\,\overset{\mathbf{V}}{\mathbf{W}}\right) = \mathrm{SS}(\underline{\mathbf{E}}) \tag{10}$$

This least squares estimate minimizes the sum of squares (SS) in eq. 10.

A four-mode Tucker model was described by Lastovicka [29] in 1981. It is similar to the 3-mode model, but has a 4-way core matrix. The algorithm used is similar to that of Tucker and Levin described above.

A chemical example of the use of Tucker's model for 3-way analysis of chromatographic data was described by De Ligny et al. [11] and in Spanjer's Ph.D. thesis [30].

### *3.6 Scatter plots for visual interpretation*

The Tucker's 'fundamental model' gives orthogonal $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{W}$. This allows the construction of scatter plots, as in principal component analysis; the loading plots or score plots. One could plot $\boldsymbol{u}_1$ against $\boldsymbol{u}_2$, $\boldsymbol{u}_1$ against $\boldsymbol{u}_3$, $\boldsymbol{u}_2$ against $\boldsymbol{u}_3$ etc. The same holds for the vector elements of $\mathbf{V}$ and $\mathbf{W}$. These scatter plots allow a visual interactive analysis of data structures, since outliers

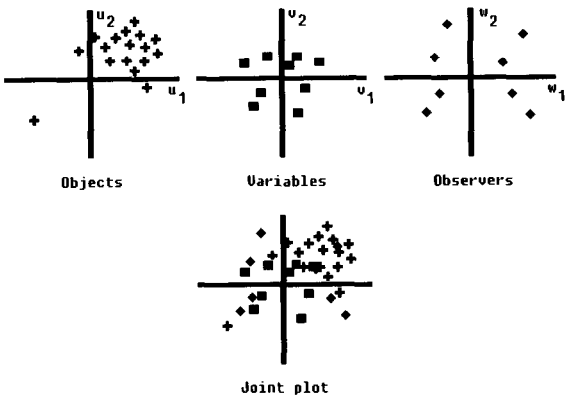


Fig. 5. The loadings of the Tucker decomposition can be represented visually as scatter plots. Inspection of U, V and W, respectively, gives information about objects, variables or observers. In this hypothetical figure, there are 17 objects, 9 variables and 8 observers. One could also attempt to show these loading plots superimposed on each other, with appropriate scaling. This is called a joint plot.

and clusterings are easily detected. 3D scatter plots are also possible. One could also plot 2D scatter plots on top of each other e.g. $u_1$ against $u_2$ on top of $v_1$ against $v_2$ on top of $w_1$ against $w_2$, if appropriate scaling is used. This would give a very good overview of the data. Fig. 5 demonstrates a conceptual view of how this could be done.

### *3.7 Canonical decomposition: the Caroll and Chang algorithm*

Carroll and Chang have written an article in 1970 [31] with an interesting paragraph on 'canonical decomposition of 3-way tables', but which is otherwise rather unreadable because of psychometric jargon they used. They limit themselves to an identity core matrix. This is simplification 3 of the Tucker model. Canonical decomposition and tensor rank studies have some similarities, but are of a different genetic origin; they are therefore dealt with in separate parts of the article. The canonical decomposition is stated as having been programmed for up to seven-way tables. Kruskal [32] mentions a similar decomposition proposed by Harshman in 1970. The canonical decomposition is based on the principles of NIPALS [1]. It is an alternating least squares procedure.

The model is:

$$x_{ijk} = \sum_{a=1}^{A} u_{ia} v_{ja} w_{ka} + e_{ijk} \tag{11a}$$

or

$$\underline{\mathbf{X}} = \mathbf{U}\,\underline{\overset{\mathbf{V}}{\mathbf{I}}}\,\mathbf{W} + \underline{\mathbf{E}} = \mathbf{U}.*.\underline{\mathbf{H}}_{vw} + \underline{\mathbf{E}} \tag{11b}$$

The algorithm consists of taking some starting values for the matrices V and W and calculating a least squares value for U, where $\underline{\mathbf{H}}_{vw}$ is an auxiliary (help) array:

$$(1)\quad \mathbf{U} = \underline{\mathbf{X}}\mathbf{H}^{-}_{vw} \tag{12}$$

This least squares value for U is then used in calculating a least squares value for V. Then the least squares value for W is calculated before returning to step one. With some rewriting of indices:

$$(2)\quad \mathbf{V} = \mathbf{H}^{-}_{uw}\underline{\mathbf{X}} \tag{13}$$

$$(3)\quad \mathbf{W} = \mathbf{H}^{-}_{uv}\underline{\mathbf{X}} \tag{14}$$

(4) go back to 1 if no convergence.

$\mathbf{H}^{-}_{vw}$, $\mathbf{H}^{-}_{uw}$ and $\mathbf{H}^{-}_{uv}$ are generalized inverses, just as those used for the 2-way NIPALS, Appendix Part A.3.

A comparison with some steps of the NIPALS algorithm is in order. The NIPALS algorithm for 2-way arrays is given briefly in the Appendix, Part A.3.

The difference for 3-way data is that there are three pseudoinverse multiplications: eqs. 12, 13 and 14, as expected. The authors point out another important difference with the usual NIPALS. This is that one cannot calculate one set of vectors $u$, $v$ and $w$, subtract them from $\underline{\mathbf{X}}$ and then calculate a next set from the residual of $\underline{\mathbf{X}}$. The whole of the sets U, V and W have to be used all the time. This makes it necessary to define a rank for $\underline{\mathbf{X}}$ in advance. This is a serious complication.

In the NIPALS algorithm for 2-way arrays, the residual after a number of components lies in a space orthogonal to that of the calculated components. That fact makes it possible to calculate the components one-by-one. For the Carroll and Chang algorithm this is not true. Orthogonality in

U, V and W would almost always give a core matrix different from identity.

A less important difference between NIPALS for 2-way arrays and the canonical decomposition of Carroll and Chang is that of the rotation problem [31]. This makes it necessary to keep the size of $t$ or $p$ fixed in the NIPALS algorithm. This requirement falls away in the canonical decomposition of 3-way arrays.

The computer program for the Carroll and Chang decomposition is called CANDECOMP [16]. An independent development of the same algorithm has led to the PARAFAC program of Harshman [16,33]. The Harshman PARAFAC program is considered as being very advanced [16]. It uses cross-validation for determining the validity of the model of eqs. 11a and 11b.

### 3.8 Tensor rank

One simplification of the Tucker model is studied in mathematics as multilinear algebra or tensor algebra. Tensor algebra [34] has a lot of uses outside data analysis, but some of its aspects can be useful even for many-way analysis. Here only the principles of tensor rank and tensorial decomposition will be taken up, since they fit in well with the other topics of this article, especially the Carroll and Chang canonical decomposition.

A 3-way array is also called an order 3 tensor. From the middle' seventies a lot of papers appeared, mainly in the journal *Linear Algebra and its Applications*, describing the Tucker decomposition with identity core matrix [35–40]. First of all some definitions are in order: a 2-way data matrix is an order 2 tensor. The tensor rank is its usual matrix rank. The matrix can be decomposed into outer products of 2 vectors, also called dyads. The minimum number of dyads that can represent a matrix is its rank or tensor rank:

$$\mathbf{X} = \sum_{a=1}^{A} t_a p_a' + \mathbf{E} = \mathbf{TP}' + \mathbf{E} \qquad (15)$$

Here $t_a p_a'$ is a dyad, outer product of vectors $t_a$ and $p_a'$, and $A$ is the rank of the matrix **X**.

For third order tensors, triads are defined as outer products of three vectors (a generalization of the outer product of two vectors; this can best be seen in Fig. 6):

$$\underline{\mathbf{X}} = \sum_{a=1}^{A} u_a v_a w_a + \underline{\mathbf{E}} = \mathbf{U}\,\underline{\mathbf{I}}\,\mathbf{V}\,\mathbf{W} + \underline{\mathbf{E}} \qquad (16)$$

The tensor rank of $\underline{\mathbf{X}}$ is the minimum number $A$ of triads necessary to describe $\underline{\mathbf{X}}$, meaning that $\underline{\mathbf{E}}$ becomes a tensor of zeros.

In eq. 16: $\underline{\mathbf{X}}$ is a third order tensor, size $I \times J \times K$; $u_a v_a w_a$ is a triad of the vectors $u_a$, $v_a$ and $w_a$; $\underline{\mathbf{E}}$ is a residual tensor; $\underline{\mathbf{I}}$ is the identity tensor size $A \times A \times A$; **U** is a matrix, size $I \times A$ with the $u_a$ as columns; **V** is a matrix, size $J \times A$ with the $v_a$ as columns; and **W** is a matrix, size $K \times A$ with the $w_a$ as columns. See also Figs. 6 and 7.

The study of tensor rank is rather complex. One result that is generally accepted is that of the maximum rank of an order 3 tensor of size $I \times J \times K$ [37,38]:

$$\max \text{ rank } (\underline{\mathbf{X}}) = \min (I \times J,\ J \times K,\ K \times I) \qquad (17)$$

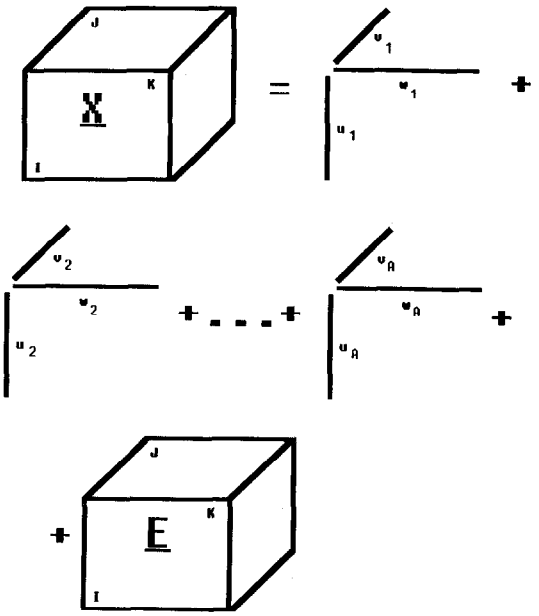


Fig. 6. The decomposition of a 3rd order tensor $\underline{\mathbf{X}}$ (size $I \times J \times K$) into triads. The residual tensor $\underline{\mathbf{E}}$ can be a tensor filled with zeros. If that is the case and $A$, the number of triads, is minimal then $A$ is the tensor rank of $\underline{\mathbf{X}}$.

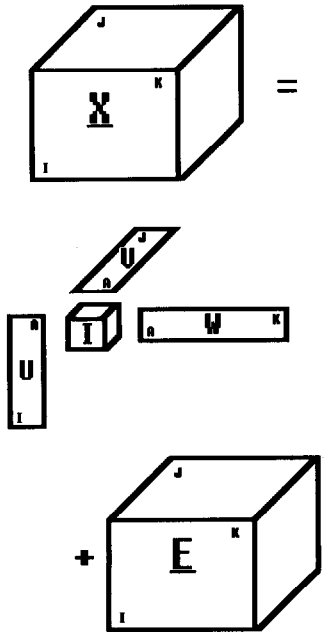


Fig. 7. The tensorial decomposition of $\underline{\mathbf{X}}$ represented as a Tucker decomposition with an identity core matrix $\underline{\mathbf{I}}$ of size $A \times A \times A$. Compare this with Figs. 3 and 6.

This is different from the rank of an order 2 tensor of size $N \times K$:

$$\text{max rank } (\mathbf{X}) = \min (N, K) \tag{18}$$

Just as in matrix algebra, it is not necessarily the full rank cases that are the most useful. Kruskal [37] has pointed out that the result of eq. 16 is valid for the canonical decomposition of Carroll and Chang. The number of 'canonical components' ($A$ in eq. 11 can be higher than each of the sizes ($I$, $J$, $K$) of the array $\underline{\mathbf{X}}$. This is different from the Tucker–Levin algorithm in eqs. 5–10. This difference is a very important one.

## 4 PRINCIPAL COMPONENTS, PARTIAL LEAST SQUARES REGRESSION

Principal component analysis (PCA) is a very convenient tool for the analysis of 2-way arrays. The technique is very visually oriented. It is desirable to transfer some of the properties of PCA to arrays of more than 2 ways. In many situations in chemistry one can calculate principal components [41] on an unfolded [42] multi-way array. Unfolding (shown in Fig. 8) creates a hierarchy of modes. This hierarchy can be meaningful in many situations. For a 3-way array, the usual situations are: (1) one object mode and two variable modes ($OV^2$ array) and (2) two object modes and one variable mode ($O^2V$ array). In each case, the modes that are most related are kept together.

It has been found that it is always possible to decompose an $N$-way array into a sum of Kronecker products of $(N-n)$-way arrays and $n$-way arrays [43]:

$$\mathbf{X}^N = \sum_{a=1}^{A} \mathbf{T}_a^{(N-n)} * \mathbf{P}_a^n + \mathbf{E}^N \tag{19}$$

Here $\mathbf{X}^N$ is an $N$-way array; $\mathbf{T}_a^{(N-n)}$ is an $(N-n)$-way score; $\mathbf{P}_a^n$ is an $n$-way loading; $\mathbf{E}^N$ is an

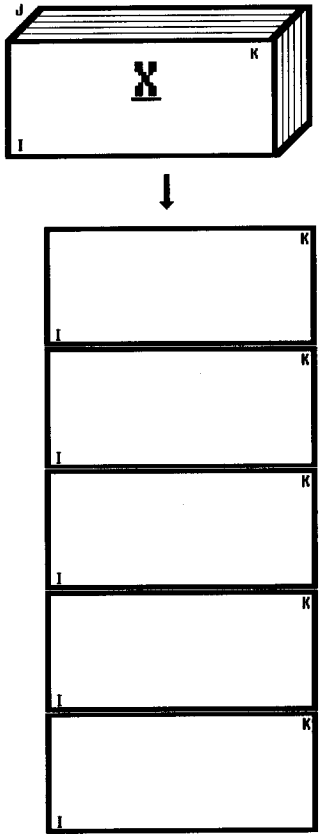


Fig. 8. A 3-way array $\underline{\mathbf{X}}$ can be sliced along one of the ways and the slices can be rearranged into a matrix $\mathbf{X}$. There are three possibilities: $I \times (J \times K)$, $J \times (I \times K)$ and $K \times (I \times J)$.

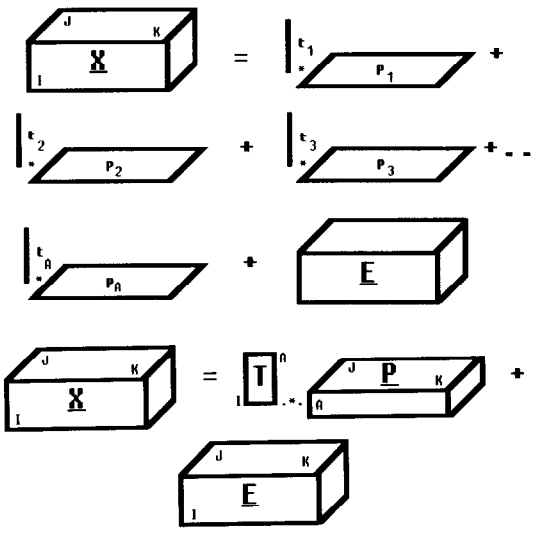

Fig. 9. The 3-way array $\underline{\mathbf{X}}$ is decomposed into a sum of Kronecker products of score vectors and loading matrices and a residual array $\underline{\mathbf{E}}$. Below that is a simplified representation.

$N$-way residual; and * is the symbol for the Kronecker product. See also Fig. 9.

For the 3-way decomposition, this can be seen as a type 4 simplification of the Tucker model, e.g.:

$$\underset{\mathbf{V}}{\underline{\mathbf{X}}} = \mathbf{U}\,\underline{\mathbf{G}}\,\mathbf{W} + \underline{\mathbf{E}} = \mathbf{U} \,.*.\, \underline{\mathbf{H}} + \underline{\mathbf{E}} \tag{20a}$$

$$x_{ijk} = \sum_{a=1}^{A} u_{ia} h_{jka} + e_{ijk} \tag{20b}$$

$\underline{\mathbf{H}}$ is an $A \times J \times K$ array; $\mathbf{V}$, $\mathbf{W}$ and $\underline{\mathbf{G}}$, as in eq. 20a, are not calculated separately. .*. is the symbol for a summation of Kronecker products, represented as a product.

For practical reasons, one can unfold $\underline{\mathbf{X}}$ to a matrix $\mathbf{X}$ of size $I \times (J \times K)$. $\mathbf{U}$ then becomes a matrix of size $I \times A$ and $\mathbf{H}$ one of size $A \times (J \times K)$. This allows the use of principal component analysis, as used for 2-way matrices.

With some rearrangement of indices, this model can also be written with $\mathbf{V}$ or $\mathbf{W}$ as separate matrices:

$$\underline{\mathbf{X}} = \underset{\mathbf{V}}{\underline{\mathbf{H}}} \,.*.\, + \underline{\mathbf{E}} \tag{21}$$

$$\underline{\mathbf{X}} = \underline{\mathbf{H}} \,.*.\, \mathbf{W} + \underline{\mathbf{E}} \tag{22}$$

Eqs. 20, 21 and 22 represent unfolding in three different directions. There is no guarantee that the $\underline{\mathbf{E}}$ is the same for all three equations (eqs. 20, 21, 22).

With the same ideas in mind, one can build up models for regression based on latent variables: partial least squares regression and principal component regression. Fig. 10 gives an idea of how different regression situations can be viewed. The fact that regression is proposed gives a hierarchy to the modes. Only when modes in $X$ and $Y$ blocks are common can a regression be defined. More about PLS regression with multi-way arrays can be found in ref. 43. This is an area that was largely ignored by the psychometric literature, and where for chemical examples there is still much to be done.

## 5 MULTIVARIATE IMAGE ANALYSIS

Multivariate image analysis in its simplest form is carried out on $O^2V$ arrays. Model-based studies of multivariate image analysis have been carried out by statisticians. A good example with literature references is the 1986 article by Besag [44].

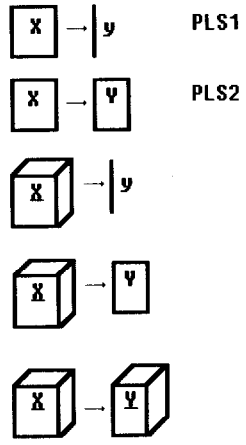


Fig. 10. Different ways of writing the regression between blocks of variables as PLS or PCR regressions. The two uppermost regressions are the known PLS1 and PLS2. The fact that a regression is proposed creates a hierarchy in the modes. Latent variables are created along a mode and the inner relation of PLS is the linear regression along a latent variable for an $X$-block and that for a $Y$-block.

A multivariate image is a stack $\underline{\mathbf{X}}$ of congruent images, each measured for a different variable. Its special feature is the enormous amount of data in the geometrical ($O^2$) modes [12,13,45,46].

Principal component analysis can be carried out as described in eq. 19:

$$\underline{\mathbf{X}} = \sum_{a=1}^{A} \mathbf{T}_a * \boldsymbol{p}'_a + \underline{\mathbf{E}} = \underline{\mathbf{T}} .*. \mathbf{P}' + \underline{\mathbf{E}} \qquad (23)$$

This is the same as in Fig. 9, but with the **T**s and $\boldsymbol{p}$s taking each other's place [46,47].

Because of the similarity of the two object modes, a hierarchical decomposition (Tucker simplification 4) is in order. The score matrices obtained are score images (sometimes also called eigenimages). They can be viewed on a monitor screen. Further analysis of score images is usually done by classical univariate image analysis operations. It is often the context in local neighbourhoods that determines how the data are interpreted.

Regression can also be carried out on images [46]. The regression methods based on latent variables have the advantage that they tolerate collinear data in a multivariate image. A graphical representation is given in Fig. 11. The difference from the regression methods in Fig. 10 is that the latent variables are 2-way arrays instead of vectors.

Image analysis is a case where decomposition of the 3-way array into vectors (tensorial decomposition) seems meaningless. It is an advantage to get images that are of a rank higher than one if they have to get a meaningful interpretation. In the future, however, ways may also be found to use the tensorial decomposition on multivariate images.

Chemical applications may be found in all kinds of microscopy, remote sensing and medical tomographic imaging [47,48].

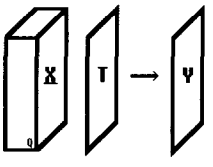


Fig. 11. A regression model can be built between a multivariate image $\underline{\mathbf{X}}$ of $Q$ slices and external information available for all or some pixels (array **Y**). It is advantageous to let the relation go via the latent variable image(s) **T**. This in contrast with the situation in Fig. 9, where the latent variables are vectors.

## 6 CONSENSUS METHODS IN CHEMISTRY

Consensus methods for comparing object × variable 2-way matrices obtained for different judges have existed for some time [49]. An interesting variant, based on the NIPALS iterations method, was proposed by Harald Martens, Svante Wold and Magni Martens in 1984 [50]. The method was primarily used by Magni Martens [14] in work concerning the relation between sensory measurement of foods and the chemical composition of those foods.

Sensory data have a relation to psychometric data. The general scheme is that a panel of judges (observers) tests a number of food samples (objects, cases) and makes a statement about quality parameters (variables) [51]. This is shown in Fig. 12. The 3-way array can be sliced up judge-wise to give a stack of object × variable matrices. Each of these separate judge matrices can produce a score, but the goal of the method is to provide a consensus score, a linear combination of individual judge scores:

$$\boldsymbol{t}_c = w_1\boldsymbol{t}_1 + w_2\boldsymbol{t}_2 + \ldots + w_J\boldsymbol{t}_J \qquad (24)$$

where $w_j$ is the weight for judge $j$, $j = 1 \ldots J$; $\boldsymbol{t}_j$ is the individual score for judge $j$; and $\boldsymbol{t}_c$ is the consensus score.

Comparing the consensus scores with PCA scores can be done via eigenvalue–eigenvector equations of the judge-wise unfolded array $\underline{\mathbf{X}}$:

$$\mathbf{X}\mathbf{X}'\boldsymbol{t} = l\boldsymbol{t} \qquad (25)$$

where **X** is the unfolded array $\underline{\mathbf{X}}$ of size $I \times (J \times K)$; $\boldsymbol{t}$ is a principal component (PC) score; and $l$ is a PC eigenvalue. This can be done $A$ times to create $A$ eigenvalues and eigenvectors. Note that these are not indexed in the equations.

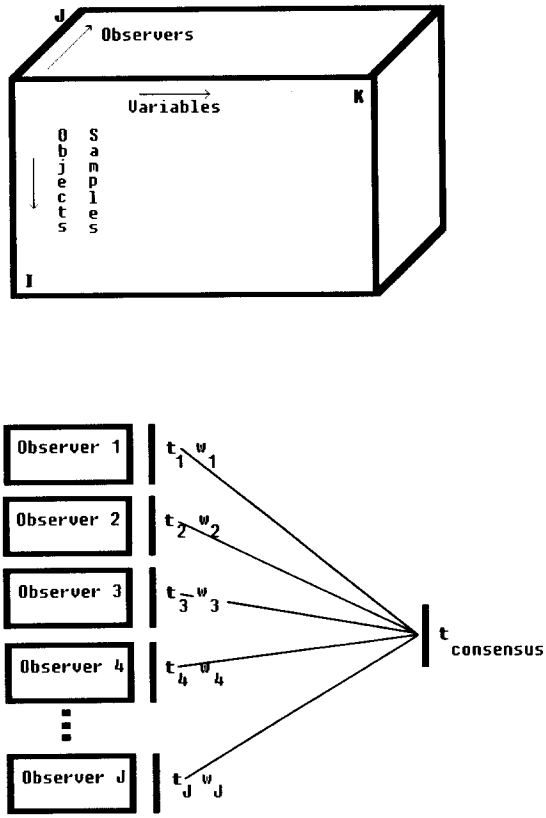


Fig. 12. A three-way array for $K$ variables measured on $I$ objects by $J$ observers can be decomposed to give a consensus score. The consensus score is a weighted average of the local scores $t_j$ with weights $w_j$.

The eigenvalue–eigenvector equation for the consensus method is:

$$(w_1\mathbf{X}_1\mathbf{X}_1' + w_2\mathbf{X}_2\mathbf{X}_2' + \ldots + w_J\mathbf{X}_J\mathbf{X}_J')t_c = l_c t_c \tag{26}$$

where $\mathbf{X}_j$ is the $I \times K$ data matrix for judge $j$, $j = 1 \ldots J$; $t_c$ is a consensus score; $l_c$ is a consensus eigenvalue; and $w_j$ is a weight for judge $j$.

Again there are $A$ possible eigenvectors and eigenvalues, but the index $a$ has been omitted in order to simplify the equations. One can easily see that, by setting the weights $w_j$ all to 1, one obtains the usual PC solution for the unfolded array $\underline{\mathbf{X}}$ as in eq. 25.

The calculation of the consensus score and of the judge weights is performed in an iterative fashion. Judge weights give the importance of each judge in building up the consensus score. More than one consensus component can be extracted from the data. The method has also been proposed for interlaboratory comparisons, where a number of laboratories (observers) measure a number of variables on the same set of objects [52,53].

The basic decomposition of consensus methods is [54]:

$$\underline{\mathbf{X}} = \sum_{a=1}^{A} t_{ca} * \mathbf{P}_a + \underline{\mathbf{E}} = \mathbf{T}_c \,.*\, \underline{\mathbf{P}} + \underline{\mathbf{E}} \tag{27}$$

This can be seen as a type 4 simplification of the Tucker model.

## 7 RANK ANNIHILATION ANALYSIS

Rank annihilation was introduced by Ho, Christian and Davidson in 1978 [55]; see also refs. 56 and 57. It is used for calibration with bilinear data. If one considers a hyphenated technique such as LC–UV, the result obtained when injecting a sample in the instrument is a matrix. This matrix $\mathbf{M}$ can be thought of as a composite. Consider first the case of only one analyte being present:

$$\mathbf{M} = xcy' \tag{28}$$

here $x$ is the standardized chromatogram of the analyte ($J \times 1$); $c$ is the unknown concentration of the analyte; and $y$ is the standardized spectrum of the analyte ($K \times 1$).

Ho et al. [55] first did this for fluorescence data, where $x$ contains an excitation spectrum and $y$ an emission spectrum. The total fluorescence intensity is of course linearly dependent on the concentration $c$ of the analyte studied.

When many compounds are present, the matrix $\mathbf{M}$ is a linear combination of their respective contributions:

$$\mathbf{M} = \sum_{a=1}^{A} x_a c_a y_a' \tag{29}$$

which may be written as:

$$\mathbf{M} = \mathbf{X}\mathbf{C}_u\mathbf{Y}' \tag{30}$$

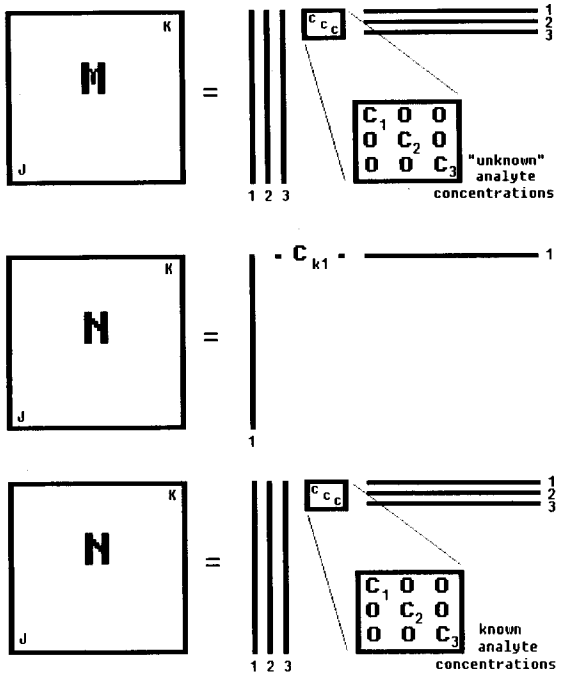


Fig. 13. The matrix **M** for a bilinear method with unknown analyte concentrations. The matrix **N** can be for a known concentration of one analyte. This is 'classical' rank annihilation. In generalized rank annihilation the matrix **N** contains results for more analytes of known concentration at the same time. The method only works when some of the analytes in **N** are also present in **M**.

where **X** is the matrix containing the chromatograms as columns ($J \times A$); $\mathbf{C}_u$ is a diagonal matrix of unknown concentrations ($A \times A$); and $\mathbf{Y}'$ is the matrix containing the spectra as rows ($A \times K$). See also Fig. 13.

When there is no overlap in at least one of the variable modes, it is easy to partition the matrix **M** into smaller segments that are unique for a certain analyte. But there are always situations where this is not possible.

An early technique of analysis was based on using the matrix **N** for a known concentration for one analyte $c_k$:

$$\mathbf{N} = x_k c_k y_k' \tag{31}$$

It is assumed that **M** and **N** have one pair of $x$ and $y$ in common. One can find the ratio of concentrations in **M** and **N** of a common analyte by considering the expression:

$$\mathbf{M} - b_k\mathbf{N} \tag{32}$$

This expression has a rank that is one lower than that of **M** for a certain value of $b_k$:

$$\text{rank}(\mathbf{M} - b_k\mathbf{N}) = \text{rank}(\mathbf{M}) - 1 \tag{33}$$

The old method consisted of trying different values of $b_k$ until the condition in eq. 33 was fulfilled. More recent methods can estimate $b_k$ directly. They were proposed by Kim [58], Lorber [59,60] and Sanchez, Kowalski and Ramos [61–65]. More details can be found in ref. 61. The method not only allows calculation of unknown concentrations, but also the calculation of spectra and chromatograms. It is insensitive to other compounds having overlapping spectral or chromatographic information that may be present in **M**. The problem of the determination of rank (eq. 33) remains. The more generalized method uses a matrix **N** with many analytes of known concentration:

$$\mathbf{N} = \mathbf{X}\mathbf{C}_k\mathbf{Y}' \tag{34}$$

where $\mathbf{C}_k$ is a diagonal matrix of known analyte concentrations. The matrices **M** and **N** form two planes in a 3-way array. The most generalized rank annihilation can be seen as a Tucker decomposition for the simplified case of $\underline{\mathbf{G}}$ being an identity matrix. See Fig. 14. This makes rank annihilation a canonical decomposition method. It has all the properties of such a method. It has the property of eq. 16 and the rotation indeterminacy is also absent.

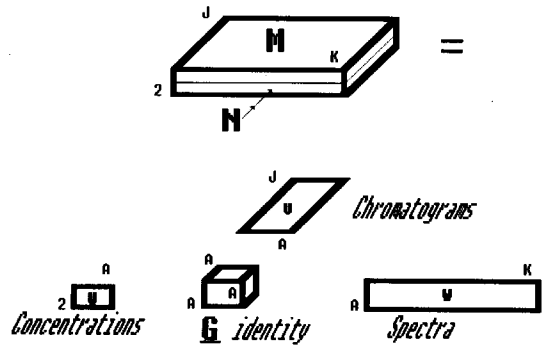


Fig. 14. The pair of matrices **M** and **N** can be seen as a 3-way array of two layers. Rank annihilation can then be explained as a Tucker decomposition with identity core matrix. This decomposition has the same model as a Carroll and Chang canonical decomposition.

There has been some criticism voiced against the rank annihilation method because it relies on only one calibration standard, which means that any error in this standard propagates to all results with no chance of its being discovered. One easy way to get around this problem is to use many calibration matrices $\mathbf{N}_1 \ldots \mathbf{N}_J$. This allows the calculation of $J$ values for the unknown concentration in $\mathbf{M}$. A robust mean of these $J$ values would then certainly be a good estimate of the true concentration.

Appellof and Davidson introduced 3-way rank annihilation in 1983 [66] by considering fluorescence measurements at the exit of a chromatograph. The equations then become:

$$\underline{\mathbf{M}} = \sum_{a=1}^{A} c_a x_a y_a z_a = \mathbf{X}\underline{\mathbf{C}}\mathbf{Y}\mathbf{Z} \qquad (35)$$

where $\underline{\mathbf{M}}$ is a 3-way array; $\mathbf{X}$ is a matrix with chromatograms as columns; $\mathbf{Y}$ is a matrix with excitation spectra as rows; $\mathbf{Z}$ is a matrix with emission spectra as rows; and $\underline{\mathbf{C}}$ is a diagonal 3-way array with concentrations of the analytes on the diagonal. The product $x_a y_a z_a$ is a triad. The expression to be studied then becomes:

$$\text{rank}\,(\underline{\mathbf{M}} - b_k\underline{\mathbf{N}}) = \text{rank}\,(\underline{\mathbf{M}}) - 1 \qquad (36)$$

where $\underline{\mathbf{N}}$ contains some chemical constituents present in $\underline{\mathbf{M}}$, but for known concentrations.

## 8 CONCLUSION

It is important to be aware of the hierarchy of multi-way methods. For well-defined problems with enough a priori knowledge available, experimental design and ANOVA analysis may be sufficient. If this is not the case then an exploratory data analysis has to be performed. In the natural sciences many variables have a reliable natural scale and the problems of multidimensional scaling are not as relevant as in psychometrics, sociology etc.

The Tucker 3-way decomposition is general and many other 3-way methods can be seen as simplifications of the Tucker model. Simplifications that can be useful for chemistry are:

type 1 = the Tucker core model, with cubic core.
type 3 = the Carroll and Chang, tensor rank model.
type 4 = unfolding when hierarchical modes are present.

The type zero simplification of the Tucker model with a non-cubic core matrix does not seem to make too much sense for chemical applications, where the assumption of an additive or multiplicative model is often made and the tensor rank is an expression of a number of chemical constituents. The determination of rank is one of the crucial problems in data analysis. This is well-known in principal component analysis and latent variable regression methods, but the problems become even more complex when 3-way data and tensor rank are considered.

A hierarchical treatment of the modes can simplify multi-way analysis. In some cases the relationships between the elements of a mode are so loose that it is possible to slice the multi-way array. The slices can then be analyzed and their results interpreted independently.

Calibration using regression equations has not been treated as well as exploratory data analysis in the psychometric literature. In the chemical literature, the emphasis is more on calibration (regression equations).

The methods treated in this paper deal only with linear models. How well they handle nonlinearities is not very well known and nonlinear methods are not available.

In many cases there is no 'best' choice, but the problem's definition determines which method is most useful. The important step of preprocessing the data has not been mentioned in this paper, which emphasizes an understanding of the models and algorithms. Neither was the statistical approach, where a priori models are assumed for the residuals (ref. 16, Chapter 4) been treated.

I conclude with a general comment about the development of methods. Methods are usually based on a theory and this theory is based on general mathematical principles or concepts. From this theory an algorithm has to be derived. This

algorithm has to be written as a computer program that is error-free and has a certain degree of user-friendliness. It then has to be demonstrated that this program can be used to solve an example problem. From then on applications may be found for which the technique becomes useful in science, industry and society. The last step is a very important one. Many publications in the scientific literature only go as far as the example problem. This can be called GLOGA (General Lack Of Good Applications). GLOGA and a lack of understanding of multi-mode mathematics are a hindrance to the common use and acceptance of multi-way methods of data analysis.

## 9 ACKNOWLEDGEMENTS

Kim Esbensen of the Norwegian Computing Center, Oslo, Norway, Svante Wold, Umeå University, Sweden, Bruce Kowalski, Avi Lorber and Eugenio Sanchez of the University of Washington, Seattle are thanked for useful comments. A special thanks goes to Age Smilde of the University of Groningen, The Netherlands for accurate proofreading of the manuscript and valuable comments on the final version. The Peat Characterization Project of Umeå, Sweden is thanked for logistic support.

## APPENDIX

### *A.1 On notation and operations for multi-way algebra*

Multi-way arrays are given as boldface uppercase underlined symbols. An example for a 3-way array is:

$$\underline{\mathbf{X}} = (x_{ijk}) \tag{A1.1}$$

The underline is used to distinguish the 3-way array from a 2-way array $\mathbf{X}$, the well-known data matrix.

The symbol $\underline{\mathbf{X}}$ represents a 3-way array with indices given as $i$ ($i = 1, I$), $j$ ($j = 1, J$) and $k$ ($k = 1, K$). This array can be represented as a parallelepiped with sizes $I$, $J$ and $K$. This makes it much easier to understand some of the operations. It is necessary to make some operations fit together with usual matrix and vector algebra.

Ledyard Tucker introduced the analysis of multi-way arrays in psychometrics [22]. He suggested the use of Kronecker products as useful operations in multi-way algebra. A simple operation is the product of a matrix (2-way, $I \times J$) $\mathbf{T}$ with a vector (1-way, $1 \times K$) $\boldsymbol{p}$ to give a 3-way array $\underline{\mathbf{X}}$:

$$\underline{\mathbf{X}} = \mathbf{T} * \boldsymbol{p} \tag{A1.2}$$

where $*$ means the Kronecker product. This means that the matrix $\mathbf{T}$ is multiplied by each scalar element in $\boldsymbol{p}$ ($p_k$) and that these products are stacked to form the 3-way array $\underline{\mathbf{X}}$. One can also express this with indices:

$$x_{ijk} = t_{ij} \cdot p_k \tag{A1.3}$$

This operation is carried out for all $i$, $j$ and $k$. The symbol . * . is used to replace a summation of Kronecker products:

$$\underline{\mathbf{X}} = \sum_{a=1}^{A} \mathbf{T}_a * \mathbf{p}_a = \underline{\mathbf{T}} \,.*.\, \mathbf{P} \tag{A1.4}$$

Or, shown as indexed operations:

$$x_{ijk} = \sum_{a=1}^{A} t_{ija} \cdot p_{ka} \tag{A1.5}$$

This is done for all $i$, $j$ and $k$.

### *A.2 Calculation of the core matrix according to Levin*

After a choice of $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{W}$ has been made (eqs. 5–8), a least squares estimator for $\underline{\mathbf{G}}$ can be calculated by least squares regression. Assuming that $\underline{\mathbf{E}}$ in eq. 1 can be written as an array of zeros:

$$\underline{\mathbf{X}} = \overset{\mathbf{V}}{\mathbf{U}\underline{\mathbf{G}}\mathbf{W}} \tag{A2.1}$$

left-multiply by $\mathbf{U}'$

$$\mathbf{U}'\underline{\mathbf{X}} = (\mathbf{U}'\mathbf{U})\overset{\mathbf{V}}{\underline{\mathbf{G}}\mathbf{W}} \tag{A2.2}$$

If **U** is well chosen, the product **U'U** equals the identity matrix **I**.

$$\mathbf{U}'\underset{\mathbf{V}}{\underline{\mathbf{X}}} = \mathbf{I}\underline{\mathbf{G}}\mathbf{W} \qquad \text{(A2.3)}$$

If the same is done for **V** and **W**:

$$\underline{\mathbf{G}} = \mathbf{U}'\underset{\mathbf{V}'}{\underline{\mathbf{X}}}\mathbf{W}' \qquad \text{(A2.4)}$$

If **U**, **V** and **W** are not chosen to be square and orthogonal, pseudoinverses $\mathbf{U}^-$, $\mathbf{V}^-$ and $\mathbf{W}^-$ can provide a solution:

$$\underline{\mathbf{G}} = \mathbf{U}^-\underset{\mathbf{V}^-}{\underline{\mathbf{X}}}\mathbf{W}^- \qquad \text{(A2.5)}$$

### *A.3 The NIPALS algorithm for 2-way arrays*

The NIPALS algorithm gives the principal component model:

$$\mathbf{X} = \sum_{a=1}^{A} t_a p_a' + \mathbf{E} = \mathbf{T}\mathbf{P}' + \mathbf{E} \qquad \text{(A3.1)}$$

where **X** is a matric of size $N \times K$; **T** is a matrix with the $t_a$ as columns, size $N \times A$; **P'** is a matrix with the $p_a'$ as rows, size $A \times K$; and **E** is a residual matrix size $N \times K$. If $A$ is the rank of **X** then **E** is a matrix of zeros.

The following is done for each dimension (factor, component), the index $a$ is omitted for ease of reading. NIPALS starts by choosing an initial value for $t$ and then multiplying **X** with the pseudoinverse of $t$ to get $p'$:

$$(1) \quad p' = (t't)^{-1}t'\mathbf{X} \qquad \text{(A3.2)}$$

This step can be compared with eq. 16. The estimate of $p$ so obtained is then used to form a pseudoinverse to calculate $t$:

$$(2) \quad t = \mathbf{X}_p(p'p)^{-1} \qquad \text{(A3.3)}$$

This step can be compared with eq. 17 or 18.

(3) Alternate these 'least squares' estimates until convergence. The complete NIPALS algorithm can be found in ref. 1.

REFERENCES


1 S. Wold, K. Esbensen and P. Geladi, Principal component analysis, *Chemometrics and Intelligent Laboratory Systems*, 2 (1987) 37–52.
2 F. Bookstein, The elements of latent variable models: a cautionary lecture, in M. Lamb, A. Brown and B. Rogoff (Editors), *Advances in Developmental Psychology*, Vol. 4, Lawrence Erlbaum Associates, Hillsdale, NJ, 1986, pp. 203–230.
3 K. Esbensen, S. Wold and P. Geladi, Relationships between higher-order data array configurations and problem formulations in multivariate data analysis, *Journal of Chemometrics*, 3 (1988) 33–48.
4 T. Hirschfeld, The hy-phen-ated methods, *Analytical Chemistry*, 52 (1980) 297A–305A.
5 C. Wilkins, Linked gas chromatography–infrared–mass spectrometry, *Analytical Chemistry*, 59 (1987) 571A–581A.
6 L. Davies, M. Raynor, J. Kithinji, K. Bartle, P. Williams and G. Andrews, SFE–GC–LC–SFC interfacing, *Analytical Chemistry*, 60 (1988) 683A–702A.
7 R. Clement, F. Onuska, G. Eiceman and H. Hill, Gas chromatography, *Analytical Chemistry*, 60 (1988) 279R–294R.
8 A. Derome, *Modern NMR Techniques for Chemistry Research*, Pergamon Press, Oxford, 1987.
9 R. Ernst, G. Bodenhausen and A. Wokaun, *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Clarendon Press, Oxford, 1987.
10 G. Bodenhausen and R. Ernst, The accordion experiment, a simple approach to three-dimensional NMR spectroscopy, *Journal of Magnetic Resonance*, 45 (1981) 367–373.
11 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.
12 P. Geladi and K. Esbensen, Can image analysis provide information useful in chemistry?, *Journal of Chemometrics*, 3 (1989) 419–429.
13 P. Geladi, K. Esbensen and S. Wold, Image analysis and chemical information in images, *Analytica Chimica Acta*, 191 (1986) 473–480.
14 M. Martens, Determining sensory quality of vegetables: a multivariate study, Ph.D. thesis, Agricultural University of Norway, Ås, Norway, 1986.
15 J.-B. Lohmöller, Die trimodale Faktorenanalyse von Tucker: Skalierungen, Rotationen, andere Modelle, *Archiv für Psychologie*, 131 (1979) 137–166.
16 H. Law, C. Snyder, J. Hattie and R. McDonald (Editors), *Research Methods for Multimode Data Analysis*, Praeger, New York, 1984, Ch. 1–4.
17 L. Pallesen, Statistical assessment of PCDD and PCDF emission data, *Waste Management and Research*, 5 (1987) 367–379.
18 R. Sands and F. Young, Component models for three-way data: an alternate least squares algorithm with optimal scaling features, *Psychometrika*, 45 (1980) 39–67.

19 R. Harshman and S. Berenbaum, Basic concepts underlying the PARAFAC-CANDECOMP three-way factor analysis and its application to longitudinal data, in D. Eichorn, P. Mussen, J. Clausen, N. Haan and M. Honzik (Editors), *Present and Past in Middle Life*, Academic Press, New York, 1981, pp. 435–459.
20 R. Cattell, The three basic factor-analytic research designs — their interrelations and derivatives, *Psychological Bulletin*, 49 (1952) 499–520.
21 L. Tucker, Implications of factor analysis of three-way matrices for measurement of change, in C. Harris (Editor), *Problems of Measuring Change*, The University of Wisconsin Press, Madison, WI, 1963, pp. 122–137.
22 L. Tucker, The extension of factor analysis to three-dimensional matrices, in N. Frederiksen and H. Gulliksen (Editors), *Contributions to Mathematical Psychology*, Holt, Rinehart and Winston, New York, 1964, pp. 110–182.
23 J. Levin, Three-mode factor analysis, *Psychological Bulletin*, 64 (1965) 442–452.
24 P. Kroonenberg and P. Brouwer, User's guide to TUCKALS3, version 4.0, Report WR-85-09-RP, Rijksuniversiteit Leiden, The Netherlands, 1985.
25 P. Kroonenberg, *Three-Mode Principal Component Analysis: Theory and Applications*, DSWO Press, Leiden, 1983.
26 P. Kroonenberg and P. Brouwer, User's guide to TUCKALS2, version 4.0, Report WR-85-12-RP, Rijksuniversiteit Leiden, The Netherlands, 1985.
27 C. Eckart and C. Young, The approximation of one matrix by another of lower rank, *Psychometrika*, 1 (1936) 211–218.
28 L. Tucker, Some mathematical notes on three-mode factor analysis, *Psychometrika*, 31 (1966) 279–311.
29 J. Lastovicka, The extension of component analysis to four-mode matrices, *Psychometrika*, 46 (1981) 47–57.
30 M. Spanjer, Substituent interaction effects and mathematical-statistical description of retention in liquid chromatography, Ph.D. thesis, University of Utrecht, The Netherlands, 1984.
31 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.
32 J. Kruskal, More factors than subjects, tests and treatments: an indeterminacy theorem for canonical decomposition and individual differences scaling, *Psychometrika*, 41 (1976) 281–293.
33 PARAFAC, Richard Harshman, Dept. of Psychology, University of Western Ontario, London, Ontario N6A 5C2, Canada.
34 *Mc-Graw-Hill Encyclopedia of Science and Technology*, 5th ed., Part 2, McGraw-Hill, New York, 1982, pp. 524–527.
35 V. Strassen, Vermeidung von Divisionen, *Journal für Reine und Angewandte Mathematik*, 264 (1973) 184–202.
36 J.-C. Lafon, Optimum computation of $p$ bilinear forms, *Linear Algebra and its Applications*, 10 (1975) 225–240.
37 J. Kruskal, Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics, *Linear Algebra and its Applications*, 18 (1977) 95–138.
38 T. Howell, Global properties of tensor rank, *Linear Algebra and its Applications*, 22 (1978) 9–23.
39 M. Atkinson and S. Lloyd, Bounds on the ranks of some 3-tensors, *Linear Algebra and its Applications*, 31 (1980) 19–31.
40 T. Lickteig, Typical tensorial rank, *Linear Algebra and its Applications*, 69 (1985) 95–120.
41 J.-B. Lohmöller and H. Wold, Three-mode path models with latent variables and partial least squares (PLS) parameter estimation, presented at the European meeting of the Psychometrics Society, Groningen, The Netherlands, June 1980.
42 K. Esbensen and S. Wold, SIMCA, MACUP, SELPLS, GDAM, SPACE & UNFOLD, the way towards regionalized principal components analysis and subconstrained $N$-way decomposition – with geological illustrations, in O. Christie (Editor), *Proceedings Nordic Symposium on Applied Statistics*, Stokkand Forlag, Stavanger, 1983, pp. 11–36.
43 S. Wold, P. Geladi, K. Esbensen and J. Öhman, Multi-way principal components and PLS-analysis, *Journal of Chemometrics*, 1 (1987) 41–56.
44 J. Besag, On the statistical analysis of dirty pictures, *Journal of the Royal Statistical Society B*, 48 (1986) 259–302.
45 K. Esbensen, P. Geladi and S. Wold, BAMID — Bilinear Analysis of Multivariate Image Data, *Proc. 3rd Nordic Symposium on Applied Statistics, Copenhagen, January 27–29, 1986*, pp. 279–297.
46 P. Geladi, K. Esbensen and S. Wold, Soft modeling on multivariate images, Technical Report SAND/23/1987, Norwegian Computing Center, Oslo, Norway, Presented at the meeting: Partial Least Squares: Theory and Application, Frankfurt am Main, September 23–25, 1987.
47 P. Geladi, H. Isaksson, L. Lindqvist, S. Wold and K. Esbensen, Principal component analysis on multivariate images, *Chemometrics and Intelligent Laboratory Systems*, 5 (1989) 209–220.
48 H. Grahn, N.M. Szeverenyi, M.W. Roggenbuck, F. Delaglio and P. Geladi, Data analysis of multivariate magnetic resonance images. I. A principal component analysis approach, *Chemometrics and Intelligent Laboratory Systems*, 5 (1989) 311–322.
49 J. Gower, Generalized procrustes analysis, *Psychometrika*, 40 (1975) 33–51.
50 H. Martens, personal communication.
51 M. Martens and H. Martens, Partial least squares regression, in J. Piggott (Editor), *Statistical Procedures in Food Research*, Elsevier Applied Science, London, 1986, pp. 293–359.
52 K. Esbensen, M. Martens, P. Geladi and H. Martens, Multivariate inter-laboratory calibration and rapid quality control, in P. Baltes, R. Baardseth, K. Norang and K. Söyland (Editors), *Proceedings EURO FOOD CHEM IV, Loen, Norway, 1–4 June 1987*, pp. 505–508.
53 P. Geladi, H. Martens, M. Martens, S. Kalvenes and K.

Esbensen, Multivariate comparison of laboratory measurements, in P. Thorbøl (Editor), *Proceedings Symposium in Applied Statistics 1988, Copenhagen, 25–27 January, 1988*, Danmarks edb-center for forskning og uddannelse, Copenhagen, 1988, pp. 15–30.

54 H. Martens, S. Wold and M. Martens, Finding the main common information in multi-table sensory data by consensus principle component analysis, *Journal of the Science of Food and Agriculture*, in preparation.

55 C. Ho, G. Christian and E. Davidson, Applications of the method of rank annihilation to quantitative analyses of multicomponent fluorescence data from the video fluorimeter, *Analytical Chemistry*, 50 (1978) 1108–1113.

56 C. Ho, G. Christian and E. Davidson, Application of the method of rank annihilation to fluorescent multicomponent mixtures of polynuclear aromatic hydrocarbons, *Analytical Chemistry*, 52 (1980) 1071–1079.

57 C. Ho, G. Christian and E. Davidson, Simultaneous multicomponent rank annihilation and applications to multicomponent fluorescent data acquired by the video fluorometer, *Analytical Chemistry*, 53 (1981) 92–98.

58 R. Kim, Matrix algorithms for bilinear estimation problems in chemometrics, Ph.D. Thesis, MIT, Cambridge, MA, 1985.

59 A. Lorber, Quantifying chemical composition from two-dimensional data arrays, *Analytica Chimica Acta*, 164 (1984) 293–297.

60 A. Lorber, Features of quantifying chemical composition from two-dimensional data arrays by the rank annihilation factor analysis method, *Analytical Chemistry*, 57 (1985) 2395–2397.

61 E. Sanchez and B. Kowalski, Generalized rank annihilation factor analysis, *Analytical Chemistry*, 58 (1986) 496–499.

62 E. Sanchez, Tensorial calibration: the generalized rank annihilation method, Ph.D. Thesis, University of Washington, Seattle, 1987.

63 E. Sanchez, L.S. Ramos and B.R. Kowalski, Generalized rank annihilation method. I. Application to liquid chromatography diode array ultraviolet detection data, *Journal of Chromatography*, 385 (1987) 151–164.

64 L.S. Ramos, E. Sanchez and B.R. Kowalski, Generalized rank annihilation method. II. Analysis of bimodal chromatographic data, *Journal of Chromatography*, 385 (1987) 165–180.

65 E. Sanchez and B. Kowalski, Tensorial calibration: II second-order calibration, *Journal of Chemometrics*, 2 (1988) 265–280.

66 C. Appellof and E. Davidson, Three-dimensional rank annihilation for multi-component determinations, *Analytica Chimica Acta*, 146 (1983) 9–14.