ELSEVIER

Chemometrics and Intelligent Laboratory Systems 25 (1994) 1–23

Chemometrics and intelligent laboratory systems

Tutorial

N-way principal component analysis Theory, algorithms and applications

René Henrion

Institute of Applied Mathematics, Humboldt University of Berlin, 10099 Berlin, Germany

Received 22 March 1993; accepted 13 December 1993

Abstract

Due to sophisticated experimental designs and to modern instrumental constellations the investigation of *N*-dimensional (or *N*-way or *N*-mode) data arrays is attracting more and more attention. Three-dimensional arrays may be generated by collecting data tables with a fixed set of objects and variables under different experimental conditions, at different sampling times, etc. Stacking all the tables along varying conditions provides a cubic arrangement of data. Accordingly the three index sets or modes spanning a three-way array are called objects, variables and conditions. In many situations of practical relevance even higher-dimensional arrays have to be considered. Among numerous extensions of multivariate methods to the three-way case the generalization of principal component analysis (PCA) has central importance. There are several simplified approaches of three-way PCA by reduction to conventional PCA. One of them is unfolding of the data array by combining two modes to a single one. Such a procedure seems reasonable in some specific situations like multivariate image analysis, but in general combined modes do not meet the aim of data reduction. A more advanced way of unfolding which yields separate component matrices for each mode is the Tucker 1 method. Some theoretically based models of reduction to two-way PCA impose some specific structure on the array. A proper model of three-way PCA was first formulated by Tucker (so-called Tucker 3 model among other proposals). Unfortunately the Tucker 1 method is not optimal in the least squares sense of this model. Kroonenberg and De Leeuw demonstrated that the optimal solution of Tucker's model obeys an interdependent system of eigenvector problems and they proposed an iterative scheme (alternating least squares algorithm) for solving it. With appropriate notation Tucker's model as well as the solution algorithm are easily generalized to the *N*-way case ($N > 3$). There are some specific aspects of three-way PCA, such as complicated ways of data scaling or interpretation and simple-structure-transformation of a so-called core matrix, which make it more difficult to understand than classical PCA. An example from water chemistry serves as an illustration. Additionally, there is an application section demonstrating several rules of interpretation of loading plots with examples taken from environmental chemistry, analysis of complex round robin tests and contamination analysis in tungsten wire production.

Contents

SSDI 0169-7439(93)E0086-J

1. Introduction

During the last few decades the investigation of so-called 'three-way data sets' (more generally: *N*-way data sets) has become a rapidly growing branch within multivariate data analysis. Originally such data sets were primarily considered in psychometry. In 1963 Tucker was the first to establish a general three-way model of principal component analysis [1]. Since then, numerous extensions of conventional methods to the three-way case have been given. For instance, in ref. [2] Bloxom formulated a factor-analytic version of Tucker's model. In ref. [3] Carroll and Chang

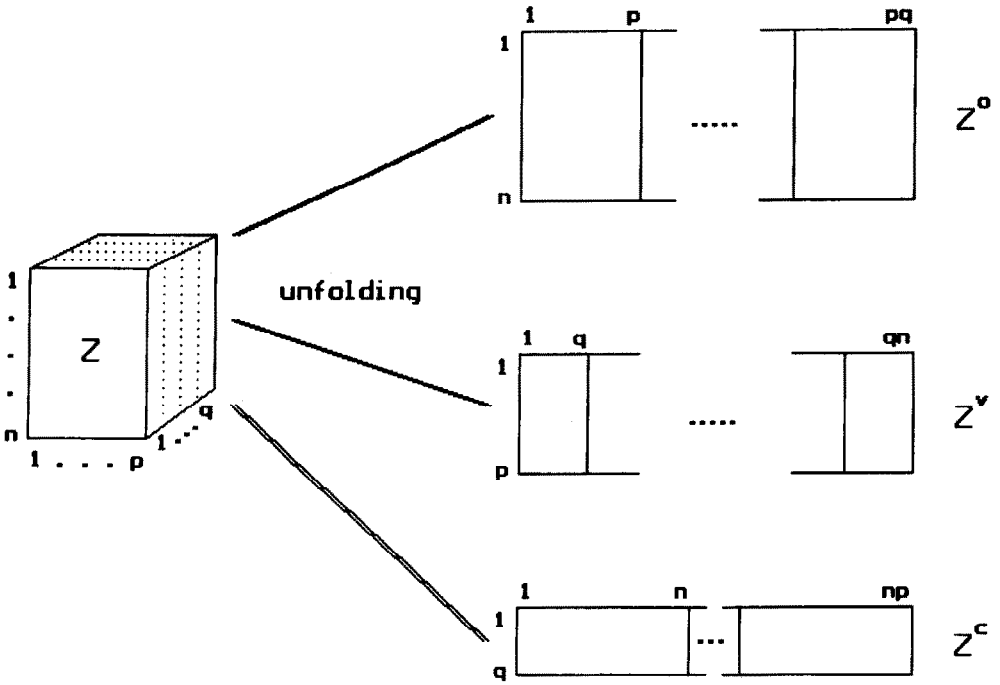

Fig. 1. A three-mode data set and the three (essential) kinds of its unfolding.

generalized the treatment of classical multidimensional scaling to individual differences scaling, where a stack of similarity matrices was obtained due to differing subjective judgments of several individuals. Another aspect of generalization is given in the context of three-mode path analysis (see Lohmöller and Wold [4]). An extensive annotated bibliography of early psychometric papers in this area may be found in ref. [5]. Furthermore, many relevant contributions to N-way analysis are contained in ref. [6]. At this point it is interesting to note that N-way procedures are not restricted to data analysis. In ref. [7] Kruskal described an application to arithmetic complexity, where the computational effort of matrix multiplication can be reduced via a minimum rank decomposition of a three-way array which is related to this algebraic operation.

In the eighties, chemometrics started to pick up this field [8–10] and to become a driving force of its development, as stated by Sanchez and Kowalski in ref. [11]: "Chemistry, a latecomer in the third-order world is perhaps much more suitable for third-order factor analysis than many other branches of science owing to the abundance of instruments that can automatically collect precise third-order arrays in a short period of time". Overviews in the chemometric literature relating to three-way methods may be found in refs. [12–15]. The present paper will be mainly concerned with N-way principal component analysis.

1.1. Examples of N-way data arrays

The increasing complexity of experimental designs and instrumental constellations leads to a natural generalization of the concept 'data set' from conventional tables to higher-dimensional arrays. Frequently, a fixed classical objects/variables setting is observed several times under different conditions yielding a separate data table for each condition. Putting all these together, one arrives at a cubic or three-dimensional data array as depicted in Fig. 1. For a concrete example we refer to ref. [16] where samples of natural water (Niger Delta) had been characterized by the repeated observation of 15 physicochemical parameters (variables) at 10 different sampling stations (objects) 22 times in the course of one year (conditions). In the sequel this example shall serve as an illustration. Clearly, the identification of objects, variables and conditions in a data set is arbitrary and not necessarily justified in all constellations. If, for instance, different chemical samples are analyzed by hyphenated methods like high-performance liquid chromatography with diode array detection (HPLC–DAD) [17,18] then one would rather speak of an (objects, variables, variables) type of data set. If, instead, one single sample yields three-dimensional chemical information [8], then a (variables, variables, variables) type occurs, while in image analysis the (objects, objects, variables) type is relevant (compare ref. [19]). From the abstract point of view there is no need for such terminology and one simply speaks of the three different modes or ways of a three-dimensional data array. In a mathematical setting these modes are index sets $I = \{1, \ldots, n\}$, $J = \{1, \ldots, p\}$, $K = \{1, \ldots, q\}$ and the data array is a mapping $\vec{\mathbf{Z}}$: $I \times J \times K \to \mathbb{R}$ into the reals (for metric data) assigning to each triple (i, j, k) the observation z_{ijk}. Of course it is more convenient to imagine a three-way array as a parallelepiped as in Fig. 1 but the formal notation has the advantage to be easily extended to more than three modes while imagination fails. Thus an N-way data array is a real valued mapping defined on the cartesian product of N index sets $I_1 = \{1, \ldots, n_1\}, \ldots, I_N = \{1, \ldots, n_N\}$. The linear space of all N-way arrays of the corresponding order shall be denoted by $\mathscr{A}^{n_1, n_2, \ldots, n_N}$.

The first investigation of a four-way array was undertaken by Lastovicka [20] in the context of a sophisticated psychological study of television advertisements. Measuring sciences, on the other hand, can produce higher-dimensional data sets in quite a natural manner. For instance, the analysis of spatio-temporal concentration change of some chemical substance in air, water or soil will immediately yield a five-way array (three spatial coordinates + time + concentrations). Even more directly, N-way arrays are obtained by suitable combination of analytical instruments (compare ref. [21]), such as three-dimensional image analysis combined with liquid chromatography and

spectrometry. It should be noted that in multivariate statistics more general concepts of N-way arrays are introduced by not requiring fixed elements in each mode. As an example, under several conditions there might be measured varying sets of variables for a constant set of objects. A systematic collection of such types was provided by Coppi [22].

1.2. Unfolding

The easiest way of handling a higher-dimensional array is to reduce it to a data table by so called unfolding and subsequently analyzing it by conventional methods. Fig. 1 illustrates the three essentially different possibilities of unfolding a three-way data set. In each situation two of the original modes are combined to yield a single one while the third is kept unchanged. The matrices obtained will be denoted by $\mathbf{Z}^{o}$, $\mathbf{Z}^{v}$, $\mathbf{Z}^{c}$, respectively, where the upper index refers to the mode which remains unchanged and defines the rows (objects, variables or conditions, respectively). For instance, $\mathbf{Z}^{o}$ results from horizontally glueing together the frontal planes (condition slices) of the array $\vec{\mathbf{Z}}$. In a strict sense, one could interchange once more the way of generating combination modes in each unfolding, but this has no substantial effect. One imagines that for arrays of dimension $N > 3$ the general principle of unfolding is to subdivide the modes into two subsets and to combine all modes within each subset to give two new larger modes. A more precise definition is as follows: Given any partition of the N modes into two subsets J_1, J_2 the unfolding operator *, assigns to each N-way array $\vec{\mathbf{Z}} \in \mathscr{R}^{n_1,\ldots,n_N}$ a matrix $\mathbf{Z}^{*} \in \mathscr{R}^{N_1,N_2}$ with $z^{*}_{\alpha_1,\alpha_2} = z_{i_1,\ldots,i_N}$ as its general element, where

$$N_i = \prod_{j \in J_i} n_j \ (i = 1, 2) \text{ and}$$

$$\alpha_p = i_{\gamma_p} + \sum_{j \in J_p} (i_j - 1) \prod_{k \in J_p, k < j} n_k \ (p = 1, 2)$$

with γ_p being the index of the last mode in J_p. From a programming point of view, unfolding of $\vec{\mathbf{Z}}$ may be achieved by implementing a sequence of for–do loops, where the outer loops change (slowly) the indices of the first group J_1 of modes and the inner loops change (quickly) the indices of the second group J_2 of modes. Increasing α_1, α_2, respectively, whenever the index of the most inner loop of the first group or second group, respectively progresses, one arrives at the coordinates α_1, α_2 of the current array element in the resulting large matrix.

Under some special circumstances such unfolding followed by standard methods, such as principal component analysis (PCA), might be sufficient for an adequate analysis. This is true, e.g., when parts of the array are built up by spatial coordinates. In a typical application of image analysis the first two modes correspond to x and y coordinates of points in the screen plane while the third mode defines several spectroscopical wavelengths, for instance. Combining the first two modes would then produce a new mode the elements $x_i y_j$ of which have a substantial meaning, namely the position of points in the plane. As a consequence, factor scores computed by PCA are directly interpretable as images by suitably transforming the scores into grey values of corresponding screen pixels (compare ref. [23]). In most problems, however, the combination modes obtained by unfolding have an artificial character and are hardly interpretable. In the example of water analysis introduced above, any one of the three possible combinations of two modes, such as sampling stations with physicochemical parameters, can be carried out formally but does not make much sense. Furthermore, combination modes usually have very large dimension which contradicts the aim of data reduction.

Still the problem can be solved by performing separate PCA analyses for each of the data matrices obtained by the different ways of unfolding (see Fig. 1). Then, ignoring the loadings of combination modes and saving merely the scores of the respective isolated modes, one obtains separate PCA representations for the items of each mode. In fact, this method (among two others) was proposed first by Tucker [24] and is frequently called the Tucker 1 method. This procedure is simply generalized to N dimensions by restricting to specific types of unfolding where, in the setting given above, the first group J_1 consists of single

modes. The advantage of the Tucker 1 method is its simple implementation by reduction to standard PCA, while its drawback is that it only gives an approximate solution in the least squares sense.

To provide an illustration let us return to the water analysis example introduced above. Fig. 2 shows the resulting Tucker 1 plots. The set of sampling stations, where measurements were carried out, was subdivided into five stations (label 'a') which are equally distributed around an effluent discharge point of an oil refinery and five non-effluent-receiving stations (label 'b') for comparative purposes. The two subgroups are clearly recognized in the objects plot (Fig. 2a). The set of physicochemical parameters subdivides into a factor which is somehow related to salinity (determined by conductivity, chloride, hardness, alkalinity and NO_3–N) and a factor which seems to be determined by pH and NH_3–N (Fig. 2b). Finally the 22 sampling times, reaching from February to December of one year with a frequency of two times a month, clearly arrange according to seasons, early times to the left, late times to the right in the plot (Fig. 2c). Such existence of a seasonal factor among physicochemical parameters is strongly confirmed by plotting the loadings of sampling times along the horizontal axis of Fig. 2c over the times themselves (Fig. 2d). The resulting curve, which is similar to some climatic function although only physicochemical parameters are considered, is mainly related to the first factor of parameters

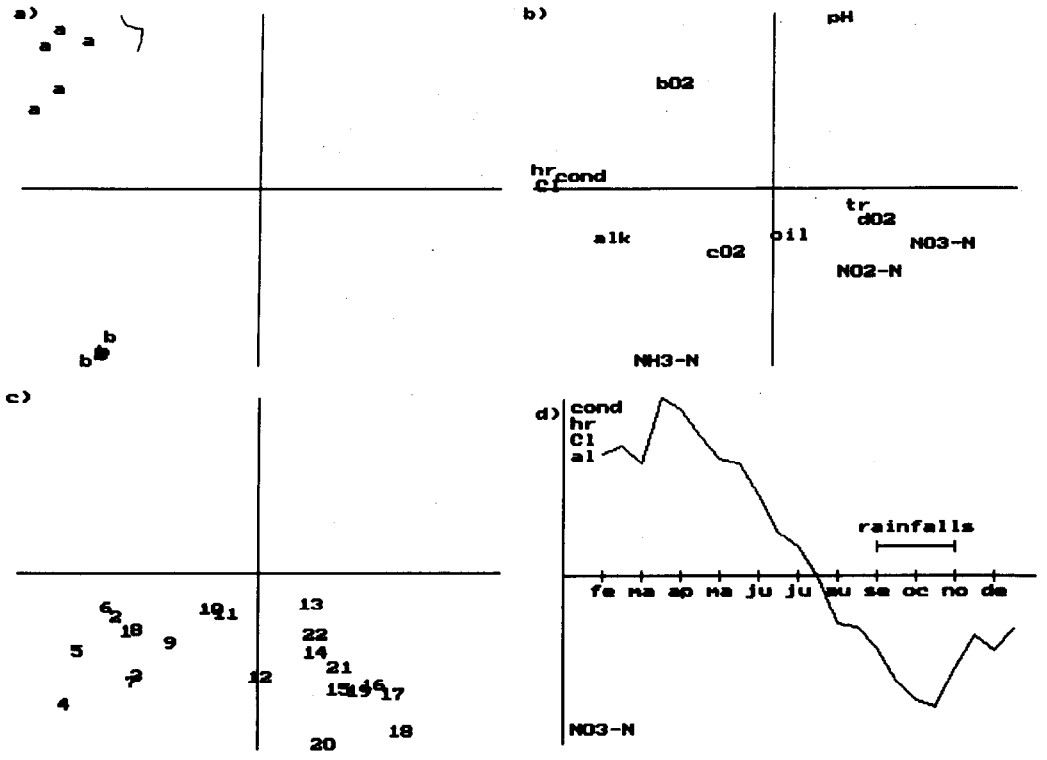

Fig. 2. PC plots (first PC vs. second) via the Tucker 1 method for a data set from water chemistry. The diagrams show the loadings for (a) objects (sampling stations), (b) variables (physico-chemical parameters) and (c) 'conditions' (sampling times). In (d) the loadings of sampling times with respect to the first PC (horizontal axis in (c)) are plotted over the times themselves. Labels for physicochemical parameters are: tr (transparency); pH (pH); chloride (Cl); hardness (hr); conductivity (cond); total alkalinity (al); dissolved oxygen (dO_2); biological oxygen demand (bO_2); chemical oxygen demand (cO_2); oil and grease (oil); ammonia–nitrogen (NH_3–N); nitrate–nitrogen (NO_3–N); nitrite–nitrogen (NO_2–N).

which was identified above as salinity. In fact, the curve reflects a negative peak of salinity during the period of strong rainfalls from September to November while a positive peak occurs during the 'dry months'.

It should be remarked here, that data sets which are partially spanned by 'ordered' modes like time, spatial coordinates or wavelength can be investigated by more sophisticated statistical methods such as time series or longitudinal data analysis [25]. This, however, requires more theoretical insight in the present model and a sufficient number of observation points in these modes. Lacking these conditions, more explorative methods of data analysis such as PCA may still provide satisfactory hypotheses on the main data structure (see ref. [26], p. 287).

2. Three-way principal component analysis

From a data-analytic point of view, principal component analysis is frequently formulated in the 'Eckart–Young' sense [27]. Accordingly, a best least squares factorization.

$$\mathbf{Z} \approx \mathbf{A}\mathbf{H}^{\mathrm{T}} \tag{1}$$

of an (n, p) data matrix $\mathbf{Z}$ into an (n, r) score matrix $\mathbf{A}$ and a (p, r) loading matrix $\mathbf{H}$ with $r < p$ is given by choosing $\mathbf{H}$ to have in its columns the normalized eigenvectors corresponding to the r leading eigenvalues of $\mathbf{Z}^{\mathrm{T}}\mathbf{Z}$ and by defining $\mathbf{A} = \mathbf{Z}\mathbf{H}$. Sometimes a symmetric version, called singular value decomposition, of (1) is preferred:

$$\mathbf{Z} \approx \mathbf{G}\Lambda\mathbf{H}^{\mathrm{T}} \tag{2}$$

where Λ is a diagonal (r, r) matrix containing the square roots of the r greatest eigenvalues of $\mathbf{Z}^{\mathrm{T}}\mathbf{Z}$ and $\mathbf{G} = \mathbf{A}\Lambda^{-1}$. Frequently, singular value decomposition is followed by a rotation of scores and loadings to simple structure. Postmultiplication by orthogonal (r, r) matrices $\mathbf{P}$, $\mathbf{T}$ provides new scores $\bar{\mathbf{G}} = \mathbf{G}\mathbf{P}$ and new loadings $\bar{\mathbf{H}} = \mathbf{H}\mathbf{T}$ which might be more interpretable than the old ones. Inserting these expressions, (2) transforms to

$$\mathbf{Z} \approx \bar{\mathbf{G}}\mathbf{C}\bar{\mathbf{H}}^{\mathrm{T}};\ z_{ij} \approx \sum_{u=1}^{r}\sum_{v=1}^{r} \bar{g}_{iu}\bar{h}_{jv}c_{uv} \tag{3}$$

where $\mathbf{C} = \mathbf{P}^{\mathrm{T}}\Lambda\mathbf{T}$ denotes the so-called core matrix which is not necessarily symmetric. The squared elements c_{uv}^2 of the core matrix indicate the variance contribution explained by the combination of the uth column of scores with the vth column of loadings. Model (3) is also called the two-way PCA model. The main difference between the two-way PCA model and singular value decomposition is that in the former — due to independent rotation of score and loading matrices — there may occur interactions between unequal components, e.g., between the first components of scores and the second component of loadings. Such effects are precluded in singular value decomposition where the 'core matrix' Λ is diagonal, see (2). In the general two-way PCA model even unequal numbers of components for scores and loadings may be admitted ([28], p. 361), but further discussions will be postponed to the derivation of the three-way PCA model.

In contrast to the quite heuristic unfolding techniques described above, several attempts have been made for theoretically based reductions of three-way data analysis to conventional PCA. This was done by assuming some specific structure of the data array and by imposing certain restrictions on the model. Hakstian [29], for instance, investigated the problem of jointly decomposing the condition slices $\mathbf{Z}_1, \ldots, \mathbf{Z}_q$ (frontal planes of the three-way array). He distinguished four cases:

$$\mathbf{Z}_i \approx \mathbf{A}\mathbf{H}^{\mathrm{T}};\quad \mathbf{Z}_i \approx \mathbf{A}_i\mathbf{H}^{\mathrm{T}};\quad \mathbf{Z}_i \approx \mathbf{A}\mathbf{H}_i^{\mathrm{T}};\quad \mathbf{Z}_i \approx \mathbf{A}_i\mathbf{H}_i^{\mathrm{T}}$$

$$(i = 1, \ldots, q).$$

In the first case it is assumed that neither scores nor loadings change under different conditions $i = 1, \ldots, q$. Then a least squares solution relates to the eigenanalysis of the covariance matrix corresponding to the average $\mathbf{Z}_1 + \cdots + \mathbf{Z}_q$ of condition slices. If, as in the second case, only loadings are assumed to remain fixed while scores may change, then the eigenanalysis has to be performed on the matrix $\mathbf{Z}_1^{\mathrm{T}}\mathbf{Z}_1 \cdots \mathbf{Z}_q^{\mathrm{T}}\mathbf{Z}_q$, which for $q = 2$ is closely related to two-block methods with a constant set of variables (such as linear characteristics [30] or transposed PLS [31]). The third case is symmetric to the second one by assuming non-changing scores instead. Here, an

average covariance matrix $\mathbf{Z}_1^T\mathbf{Z}_1 + \cdots + \mathbf{Z}_q^T\mathbf{Z}_q$ has to be considered. Finally, both scores and loadings are allowed to change. This last case is treated under different criteria in ref. [29].

In the above-mentioned models different slices of the three-way array have been separately decomposed under some joint criterion (minimal sum of residuals over all slices). A less restrictive approach is simultaneous decomposition of the data array by symmetrically treating all modes. The two most important techniques for doing so are Tucker's model of three-way PCA and trilinear decomposition (to be briefly discussed later).

2.1. Tucker's model

The derivation of Tucker's three-way model [1] is most easily explained by adopting the psychometric point of view (refs. [32] or [26], p. 17) which assumes that the data variation within an array is generated by the action of certain non-measurable idealized items, all measurable real items being linear combinations thereof. It is not assumed that the number of idealized items be equal in each mode, hence there may be r idealized objects, s idealized variables and t idealized conditions. In psychometry one frequently observes different test results of different persons under different conditions and one attempts to reduce these observations to the effect of idealized subjects (e.g., sociable persons or anxious children), idealized tests (e.g., verbal or mathematical school tests) and idealized conditions (e.g., various stress situations). In a chemometric context idealized objects might be classes of chemical substances (e.g., polar or viscous solvents). Among a set of water-characterizing parameters there might occur idealized variables called 'organic' or 'inorganic exposure'. Finally, if the third mode coincides with the direction of time then periods, seasons and trends can be understood as idealized conditions.

Adopting this viewpoint, it is sufficient to know the scores of all idealized items in order to explain the scores of all real items. Denote by c_{uvw} the score of idealized object u for idealized variable v under idealized condition w. The scores of the real object i will then be linear combinations (weightings) of the scores of all r idealized objects. Particularly, object i scores for idealized variable v and idealized condition w as

$$\alpha_{vw}^{i} = \sum_{u=1}^{r} g_{iu} c_{uvw} \tag{4}$$

Next, the score of real object i for real variable j under idealized condition w will be a linear combination of the corresponding scores (fixed i and w) for the s different idealized variables:

$$\beta_{w}^{ij} = \sum_{v=1}^{s} h_{jv}\alpha_{vw}^{i} = \sum_{u=1}^{r}\sum_{v=1}^{s} g_{iu}h_{jv}c_{uvw} \tag{5}$$

Finally, the observation of real object i for real variable j under real condition k results as a linear combination of the corresponding scores of fixed object i and fixed variable j under the t different idealized conditions:

$$\gamma^{ijk} = \sum_{w=1}^{t} e_{kw}\beta_{w}^{ij} = \sum_{u=1}^{r}\sum_{v=1}^{s}\sum_{w=1}^{t} g_{iu}h_{jv}e_{kw}c_{uvw} \tag{6}$$

This last equation gives the error-free reproduction of real observations from the idealized scores. Including experimental error ϵ, one gets Tucker's model for decomposing a three-way array $\vec{\mathbf{Z}}$:

$$z_{ijk} = \sum_{u=1}^{r}\sum_{v=1}^{s}\sum_{w=1}^{t} g_{iu}h_{jv}e_{kw}c_{uvw} + \epsilon_{ijk} \tag{7}$$

Obviously, this model is a straightforward extension of (3), but now with different numbers of components retained in each mode. It can also be understood as a generalization of singular value decomposition (2) but here the differences become much more severe as a consequence of admitting interaction effects between unequal components of different modes. Chemometric references for practical applications of model (7) are refs. [9,16,33].

For deriving theoretical results it is more convenient to use a matrix notation. An extremely useful tool for treating N-way arrays is the Kronecker product of two matrices [34]. The Kronecker product is a mapping $\otimes: \mathscr{M}^{\alpha,\beta} \times \mathscr{M}^{\delta,\gamma} \rightarrow \mathscr{M}^{\alpha\delta,\beta\gamma}$, assigning to each pair **A**, **B** of matrices of

arbitrary order a new (large) matrix $\mathbf{A} \otimes \mathbf{B}$ of corresponding product order in the following way:

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} b_{11}\mathbf{A} & \cdots & b_{1\gamma}\mathbf{A} \\ \vdots & \ddots & \vdots \\ b_{\delta 1}\mathbf{A} & \cdots & b_{\delta\gamma}\mathbf{A} \end{pmatrix} \tag{8}$$

An important feature of the Kronecker product is the following simple formula of multiplication: $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD})$, whenever the occurring matrix multiplications are allowed.

Now we are ready to state the Tucker model in matrix formulation. Rearrangement of (7) yields:

$$z_{ijk} = \sum_{u=1}^{r} g_{iu} \left(\sum_{v=1}^{s} \sum_{w=1}^{t} h_{jv} e_{kw} c_{uvw} \right) + \epsilon_{ijk} \tag{9}$$

For the following line of argumentation one is advised to consult Fig. 3. Assume that all occurring entities in this equation are elements of corresponding matrices **G**, **H**, **E**, or three-way arrays $\vec{\mathbf{Z}}$, $\vec{\mathbf{C}}$. Then z_{ijk} is the element which is located in row i and column $(k-1)p+j$ of the unfolded array $\mathbf{Z}^o$. On the other hand, from the definition of the Kronecker product in (8) one easily verifies that $h_{jv}e_{kw}$ is the element being located in row $(w-1)s+v$ and column $(k-1)p+j$ of the matrix $\mathbf{H}^T \otimes \mathbf{E}^T$. Similarly, c_{uvw} is the element in row u and column $(w-1)s+v$ of the unfolded core array $\mathbf{C}^o$. Consequently, the expression in parentheses in (9) is the element in row u and column $(k-1)p+j$ of the matrix $\mathbf{C}^o(\mathbf{H}^T \otimes \mathbf{E}^T)$, and, finally, the entire non-error term in (9) turns out to be the element in row i and column $(k-1)p+j$ of the matrix $\mathbf{GC}^o(\mathbf{H}^T \otimes \mathbf{E}^T)$. Now, (7) or (9) get the matrix formulation

$$\mathbf{Z}^o \approx \mathbf{GC}^o(\mathbf{H}^T \otimes \mathbf{E}^T) \tag{10}$$

where the error array ϵ has been omitted in favour of the $\approx$ symbol. Analogously, the two other modes of unfolding (see Fig. 1) can be used to obtain the symmetric expressions $\mathbf{Z}^v \approx \mathbf{HC}^v(\mathbf{E}^T \otimes \mathbf{G}^T)$ and $\mathbf{Z}^c \approx \mathbf{EC}^c(\mathbf{G}^T \otimes \mathbf{H}^T)$, which are equivalent to (10).

In analogy to conventional PCA the decomposition (10) is exact when using component matrices of maximal order (i.e. $r=n$, $s=p$, $t=q$). In fact, let **G**, **H**, **E** be arbitrary orthogonal matrices of orders (n, n), (p, p), (q, q), respectively. Then, defining the core array via $\mathbf{C}^o = \mathbf{G}^T\mathbf{Z}^o(\mathbf{H} \otimes \mathbf{E})$, (10) is fulfilled as an equality

$$\mathbf{GC}^o(\mathbf{H}^T \otimes \mathbf{E}^T) = \mathbf{GG}^T\mathbf{Z}^o(\mathbf{HH}^T \otimes \mathbf{EE}^T)$$
$$= \mathbf{I}_n\mathbf{Z}^o\mathbf{I}_{pq} = \mathbf{Z}^o$$

due to orthogonality of the component matrices. In the context of data reduction, however, much smaller numbers r, s, t of components are desired. The heuristic idea behind the Tucker 1 method (compare preceding section) is, that the

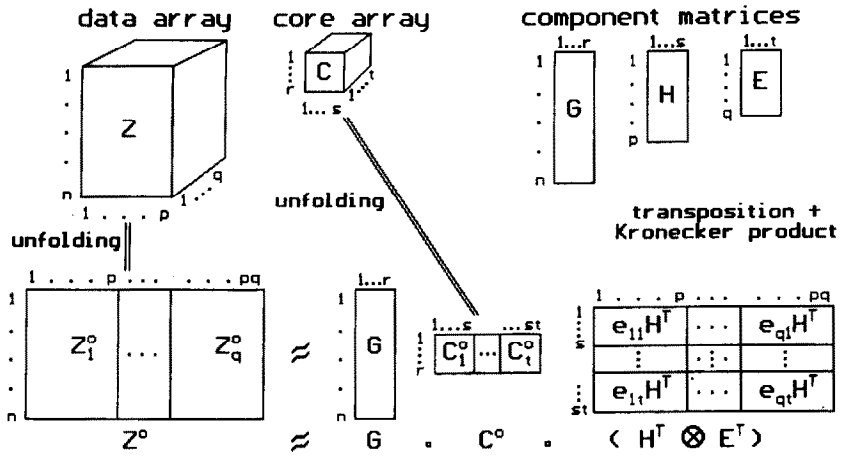

Fig. 3. Illustration of Tucker's three-way PCA decomposition of a data array into three-component matrices which are linked by a core array.

loss of fit in (10) by reducing the order of component matrices will be small if **G**, **H**, **E** are taken to be matrices of leading eigenvectors of the corresponding cross product matrices $\mathbf{Z}^{o}\mathbf{Z}^{oT}$, $\mathbf{Z}^{v}\mathbf{Z}^{vT}$, $\mathbf{Z}^{c}\mathbf{Z}^{cT}$, respectively, compare Fig. 1. As it is well known, such a procedure yields a least squares approximation in conventional PCA, but this does not hold in three-way PCA. Hence, the Tucker 1 method produces approximate solutions which are not optimal in the least squares sense. An algorithm overcoming this drawback was first discussed by Kroonenberg and De Leeuw [35].

2.2. A solution algorithm

In ref. [35] first-order conditions were derived to solve the least squares problem (10) for given numbers of components r, s, t under the restriction that all component matrices be orthonormal, i.e.,

$$\mathbf{G}^{T}\mathbf{G} = \mathbf{I}_r; \quad \mathbf{H}^{T}\mathbf{H} = \mathbf{I}_s; \quad \mathbf{E}^{T}\mathbf{E} = \mathbf{I}_t;$$

The mentioned first-order conditions may be reformulated in terms of a system of mutually dependent eigenproblems. More precisely, the optimal matrix $\mathbf{G}^*$ of object components has to contain the r leading normalized eigenvector of the (n, n)-matrix

$$\mathbf{Z}^{o}(\mathbf{H}^*\mathbf{H}^{*T} \otimes \mathbf{E}^*\mathbf{E}^{*T})\mathbf{Z}^{oT} \qquad (11)$$

where $\mathbf{H}^*$, $\mathbf{E}^*$ refer to the optimal matrices of variable and condition components. Similarly, $\mathbf{H}^*$ is related to the eigenvectors of $\mathbf{Z}^{v}(\mathbf{E}^*\mathbf{E}^{*T} \otimes \mathbf{G}^*\mathbf{G}^{*T})\mathbf{Z}^{vT}$ and $\mathbf{E}^*$ to those of $\mathbf{Z}^{c}(\mathbf{G}^*\mathbf{G}^{*T} \otimes \mathbf{H}^*\mathbf{H}^{*T})\mathbf{Z}^{cT}$. Obviously, none of these optimal matrices can be obtained without knowledge of the others. Therefore application of conventional eigenanalysis fails and, instead, an iterative procedure is unavoidable. On the other hand, once optimal component matrices have been found by any method, the optimal core array in the decomposition (10) may be computed as in the Tucker 1 method via

$$\mathbf{C}^{o*} = \mathbf{G}^{*T}\mathbf{Z}^{o}(\mathbf{H}^* \otimes \mathbf{E}^*) \qquad (12)$$

In fact, the core array is not an independent parameter in (10) and the relation (12) is easily found by applying a simplified version of a lemma by Penrose, see ref. [36] and ref. [35], p. 71.

The structure of the interdependent eigensystem stated above suggests the alternating least squares (ALS) algorithm introduced in ref. [35]: Given any initial component matrices $\mathbf{G}_0$, $\mathbf{H}_0$, $\mathbf{E}_0$, which are orthonormal and of corresponding order, compute a new iterate $\mathbf{G}_1$ for object components as the normalized (n, r) eigenvector matrix corresponding to (11) with $\mathbf{H}^*$, $\mathbf{E}^*$ replaced by $\mathbf{H}_0$, $\mathbf{E}_0$. Similarly, in the second step, a new iterate $\mathbf{H}_1$ of variable components is obtained as the analogous eigenvector matrix based on the initial $\mathbf{E}_0$ and on the just-obtained $\mathbf{G}_1$ (see the corresponding expression below (11)). Finally, $\mathbf{E}_1$ results from $\mathbf{G}_1$, $\mathbf{H}_1$. Now, one entire iteration consisting of three substeps is terminated and the whole procedure is repeated until convergence. The detailed solution algorithm can be formulated as follows:

1. Initialization by Tucker 1 method:
   Compute $\mathbf{G}_0$ as the (n, r) matrix the columns of which are the normalized eigenvectors related to the r largest eigenvalues of $\mathbf{Z}^{o}\mathbf{Z}^{oT}$ (see Fig. 1). Similarly, compute $\mathbf{H}_0$, $\mathbf{E}_0$ as the (p, s), (q, t), respectively, eigenvector matrices associated with $\mathbf{Z}^{v}\mathbf{Z}^{vT}$, $\mathbf{Z}^{c}\mathbf{Z}^{cT}$, respectively.
2. ALS algorithm:
   (a) Initialization of the iteration counter:
   Put $it = 0$
   (b) G substep:
   Define the auxiliary (n, st) matrix $\mathbf{A}^{o} = \mathbf{Z}^{o}(\mathbf{H}_{it} \otimes \mathbf{E}_{it})$. Let $\mathbf{G}_{it+1}$ be the (n, r) matrix having as columns the normalized eigenvectors associated with the r largest eigenvalues of $\mathbf{A}^{o}\mathbf{A}^{oT}$.
   (c) H substep:
   Define the (p, tr) matrix $\mathbf{A}^{v} = \mathbf{Z}^{v}(\mathbf{E}_{it} \otimes \mathbf{G}_{it+1})$. Let $\mathbf{H}_{it+1}$ be the analogous (p, s) eigenvector matrix of $\mathbf{A}^{v}\mathbf{A}^{vT}$.
   (d) E substep:
   Define the (q, rs) matrix $\mathbf{A}^{c} = \mathbf{Z}^{c}(\mathbf{G}_{it+1} \otimes \mathbf{H}_{it+1})$. Let $\mathbf{E}_{it+1}$ be the analogous (q, t) eigenvector matrix of $\mathbf{A}^{c}\mathbf{A}^{cT}$.
   (e) Test of convergence:
   If convergence has not been obtained yet then put $it = it + 1$ and repeat steps 2(b)–(d).

3. Optimal core matrix:
   Compute the optimal core matrix from the optimal component matrices via (12).

There are several remarks on this method. In principle, one could start the algorithm with any orthonormal component matrix, but initialization by the Tucker 1 method is preferred since it transfers the iterates much faster into a region around the optimum, although the optimum itself cannot be attained. For computing the eigenvectors in the ALS algorithm, Kroonenberg and De Leeuw suggest application of the Bauer–Rutishauser method [37] which is appropriate if a predetermined small number of eigenvectors has to be extracted. The test of convergence should not be based on the change of component matrices since these may happen to be merely sign changes in the course of iterations. Rather, the degree of fit in (10) obtained by the current iterates should be used. As a rule, three complete main iterations seem to be sufficient in most cases.

Finally, for the purpose of interpretation it is suggested [38] (but see also the critical notes in ref. [26], p. 162), not to use the raw elements of the orthonormal component matrices but to scale the columns according to their contributions to data fit. In conventional PCA the symmetric way of such scaling is multiplication of components by the square root of corresponding eigenvalues ([39], p. 458). The extension to three-way PCA runs as follows: Putting

$$\lambda_u = \sum_{v=1}^{s}\sum_{w=1}^{t} c_{uvw}^2; \quad \mu_v = \sum_{u=1}^{r}\sum_{w=1}^{t} c_{uvw}^2;$$

$$\nu_w = \sum_{u=1}^{r}\sum_{v=1}^{s} c_{uvw}^2 \qquad (13)$$

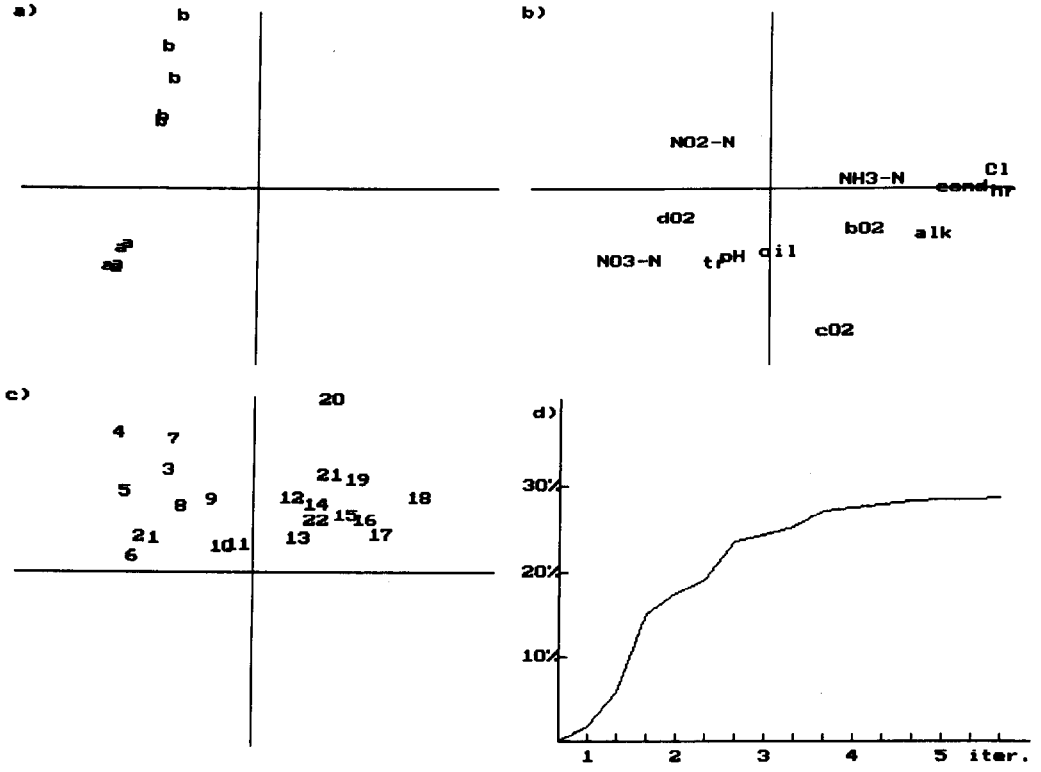

Fig. 4. Three-way PCA plots for the example of water analysis. (a) = Sampling stations, (b) = parameters, (c) = sampling times, (d) = relative improvement of fit compared to the Tucker 1 solution.

scale the columns g_u, h_v, e_w of **G**, **H**, **E**, respectively, according to:

$$\bar{g}_u = g_u\sqrt{\lambda_u}\,; \quad \bar{h}_v = h_v\sqrt{\mu_v}\,; \quad \bar{e}_w = e_w\sqrt{\nu_w} \qquad (14)$$

This kind of scaling is based on the fact that the squared core array elements represent the variance contribution of corresponding components (see interpretation of the core array below).

For the purpose of illustration inspect Fig. 4 which shows the first two (scaled) component loadings for all modes (i.e., $r = s = t = 2$) in the water analysis example. Fig. 4d maps the relative gain in data fit compared to the Tucker 1 solution (initial component matrices). After five main iterations this gain was around 30%. In other applications we could observe improvements of up to 70% confirming that the Tucker 1 solution might be far from optimal. Sometimes, however, there is no substantial difference between both procedures. It is worth noting that even quite small increases in fit may generate significant changes in point constellations. Comparing Figs. 4 and 2 this is most obvious for the display of variables (b). There is a loss of loadings for pH and NH_3–N in favour of cO_2. Also the arrangement of comparative stations in the object displays (part a of figures) becomes linear and, incidentally, agrees with their geographical position along the considered branch of the delta. In contrast, the positions of sampling times are almost constant up to sign alternation.

The percentage of total data variation covered by the PC plots in Fig. 4 is around 30%. This number seems to be small but one should keep in mind that the data structure itself is much more complicated than for conventional tables. As a rule, variance portions for a fixed number of included components clearly decrease when passing to three-way arrays. For instance, conventional PCAs on each of the unfolded tables $\mathbf{Z}^o$, $\mathbf{Z}^v$, $\mathbf{Z}^c$ yield much higher percentages than 30%, but, as was stated before, such a procedure does not make too much sense. It is clear that for a complete interpretation of data variation one has to increase the number of components. This, however, may become very complicated, since even unequal numbers of components in each mode must be considered. Here, for simplicity, the discussion of examples shall be reduced to diagrams as they are common in two-way PCA. This still seems the optimal way of quickly representing the essential data structure.

2.3. Specific aspects of three-way PCA

Tucker's formulation of three-way PCA differs from conventional PCA in a non-trivial way. Perhaps the most severe difference is that solutions need not be nested. This means, that restricting the optimal component matrices **G**, **H**, **E** of orders (n, r), (p, s), (q, t), to the first $r' < r$, $s' < s$, $t' < t$ columns, respectively, does not provide, in general, optimal component matrices of orders (n, r'), (p, s'), (q, t'). Accordingly, an optimal one-component solution in the example of water analysis does not coincide with the respective projections onto the horizontal axis of the constellations in Fig. 4a–c. The non-hierarchic character of solutions implies a growing computational effort when comparing different component constellations. In particular, it is not possible sequentially to extract component solutions one dimension at a time, as may be done in two-way PCA using the NIPALS method [10,40]. A related aspect is that no single eigenproblem is available for solving (10). In fact, the ALS algorithm described in the last section solves auxiliary eigenproblems in each substep of its iterations.

Appropriate data scaling is an important basis of three-way PCA. This problem being crucial even in two-way PCA (e.g., use of raw data, centering or standardization of data [41]) becomes much more complex now. Roughly speaking, one has the possibilities of non-scaling, overall-scaling, scaling of single modes and scaling of combination modes, where scaling refers to centering and/or additional standardization of data. Care must be taken for an appropriate choice since scaling heavily influences the solutions obtained and their interpretations afterwards. Discussions of this subject are contained, for instance, in refs. [26] and [42]. Frequently, measured variables have strongly different mean values and standard deviations. In general, this less interesting part of data variation is removed in

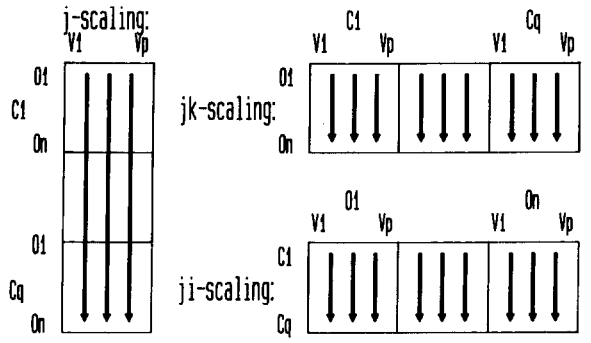

Fig. 5. Illustration of different kinds of data scaling in three-way PCA. Labels o, v and c refer to objects, variables and conditions of the data array.

order to focus on differences between objects and conditions. Therefore, a quite natural way of data pre-treatment is scaling of variables over the whole data array, i.e., along the combination mode of objects and conditions. One also speaks of '*j*-scaling' with reference to the usual index *j* for the variable mode. After this kind of scaling the differences of means and dispersions among conditions and among objects are retained in the data structure. Hence, *j*-scaling is preferred if object and condition differences are to be investigated. Sometimes, however, one is primarily interested in object differences while absolute differences between conditions are to be removed. Then, standardization should be carried out along objects only. This is called *jk*-scaling with reference to the indices of modes being subject to scaling. In a symmetric manner one can also remove absolute differences between objects and focus instead on differences between conditions (*ji*-scaling). These three typical ways of data scaling are illustrated in Fig. 5 where the arrows indicate columns of the unfolded array which get zero mean and unit standard deviation. The different effects of scaling on the interpretation of solutions will be discussed in the application section.

Perhaps the most difficult consequence of Tucker's model is an interpretation of the core matrix elements, from which essential information on the interaction between modes can be gained.

3. The core matrix

In ref. [35] it was shown that, with reference to Tucker's model (10), the sum of squared elements of an array may be decomposed into the sum of squared elements of the modeled part of this array and the sum of squared residuals:

$$\operatorname{tr} \mathbf{Z}^{\mathrm{oT}}\mathbf{Z}^{\mathrm{o}} = \operatorname{tr} \overline{\mathbf{Z}}^{\mathrm{oT}}\overline{\mathbf{Z}}^{\mathrm{o}} + \operatorname{tr}(\mathbf{Z}^{\mathrm{o}} - \overline{\mathbf{Z}}^{\mathrm{o}})^{\mathrm{T}}(\mathbf{Z}^{\mathrm{o}} - \overline{\mathbf{Z}}^{\mathrm{o}}) \tag{15}$$

where $\overline{\mathbf{Z}}^{\mathrm{o}} = \mathbf{G}\mathbf{C}^{\mathrm{o}}(\mathbf{H}^{\mathrm{T}} \otimes \mathbf{E}^{\mathrm{T}})$ denotes the modeled part of $\mathbf{Z}^{\mathrm{o}}$ in (10) and tr refers to the trace. Now it is convenient to write this modeled part as the sum of contributions due to all components, compare (10):

$$\begin{aligned}\overline{\mathbf{Z}}^{\mathrm{o}} &= \sum_{u=1}^{r}\sum_{v=1}^{s}\sum_{w=1}^{t} c_{uvw}\boldsymbol{g}_u(\boldsymbol{h}_v^{\mathrm{T}} \otimes \boldsymbol{e}_w^{\mathrm{T}}) \\ &= \sum_{u=1}^{r}\sum_{v=1}^{s}\sum_{w=1}^{t} \overline{\mathbf{Z}}^{\mathrm{o}}_{u,v,w}\end{aligned} \tag{16}$$

where $\boldsymbol{g}_u$, $\boldsymbol{h}_v$, $\boldsymbol{e}_w$ are corresponding columns of the component matrices and $\overline{\mathbf{Z}}^{\mathrm{o}}_{u,v,w}$ is the data array which is generated by their combination. Obviously, the variance modeled by joint action of a fixed triple u, v, w of components is given by

$$\begin{aligned}&\operatorname{tr} \overline{\mathbf{Z}}^{\mathrm{oT}}_{u,v,w}\overline{\mathbf{Z}}^{\mathrm{o}}_{u,v,w} \\ &\quad = \operatorname{tr} c_{uvw}(\boldsymbol{h}_v \otimes \boldsymbol{e}_w)\boldsymbol{g}_u^{\mathrm{T}}c_{uvw}\boldsymbol{g}_u(\boldsymbol{h}_v^{\mathrm{T}} \otimes \boldsymbol{e}_w^{\mathrm{T}}) \\ &\quad = c_{uvw}^2 \operatorname{tr} \boldsymbol{g}_u^{\mathrm{T}}\boldsymbol{g}_u(\boldsymbol{h}_v^{\mathrm{T}} \otimes \boldsymbol{e}_w^{\mathrm{T}})(\boldsymbol{h}_v \otimes \boldsymbol{e}_w) = c_{uvw}^2\end{aligned} \tag{17}$$

due to orthogonality of components. This last relation clarifies the role of core matrix elements in analogy to singular value decomposition: the squared entries of the core array indicate that part of data variance which is modeled by the joint action of component u of the object mode, component v of the variable mode and component w of the condition mode. This fact is crucial for interpretation of results.

3.1. Interpretation of the core matrix

As an illustration, the example of water analysis is reconsidered now. Using optimal two-component matrices from the ALS algorithm, which

are displayed in Fig. 4, the corresponding optimal core matrix is calculated via (12). Its unfolded representation is:

$$\left(\begin{array}{cc|cc} \mathbf{20.35} & -7.05 & 4.25 & 4.86 \\ 4.41 & -1.82 & -15.36 & \mathbf{-8.52} \end{array}\right) \quad (18)$$

The two parts of this matrix refer to the condition slices (frontal planes) of the three-way core array. Thus, the left part is associated with the first component of conditions and the right part with the second component. Rows refer to components of objects while columns within each of the two parts are related to components of variables. In this setting, the entry 4.86, for instance, links the first component of objects with the second component of variables and the second component of conditions. Squaring in mind all these entries according to the explanations given above, one verifies that the greatest part of variance is covered by the combination of the first components of all modes on the one hand (1/1/1) and of the second component of objects with the first component of variables and the second component of conditions on the other hand (2/1/2). Minor contributions are given by the links (1/2/1) and (2/2/2). Reinspecting Fig. 4 the link (1/1/1) describes mainly the seasonal variation of chloride, conductivity, hardness, alkalinity and NO_3–N uniformly at all sampling stations (all items projected onto the horizontal axes). Similarly, the link (2/1/2) reflects a spatial influence of the same physicochemical parameters which separates sampling stations uniformly at all sampling times. Such separation appears to be a natural geographic factor. The minor entries of the core array describe spatio/seasonal variation of the second component of variables which is mainly determined by chemical oxygen demand. A specific difficulty arises from the signs of core matrix elements. Consider, as an example, the entry −15.36 of the link (2/1/2). In this link, as can be seen from Fig. 4, stations 'b' have high positive loadings on the vertical axis, salinity parameters (conductivity, etc.) have high positive loadings on the horizontal axis and all sampling times have more or less positive loadings on the vertical axis. Multiplying the product of all these loadings with the strongly negative core matrix value will give an approximately reproduced value of stations 'b' for salinity parameters at all times according to the decomposition in (7). Carrying out these multiplications in a symbolic manner one gets $+\cdot+\cdot+\cdot-=-$, which means, that over the whole year salinity at stations 'b' is low compared to stations 'a'. Similarly, opposite groups of objects or variables might be discussed by altering corresponding signs in the symbolic product. This complicates interpretation, but in many cases it is possible to give all axes an appropriate orientation by sign change of columns in the component matrices leading to sign change in the core array, see (12). If, for instance, all relevant core matrix elements have a positive sign and most items to be discussed have positive loadings, then considerable simplification can be achieved.

3.2. Transformations of the core matrix

The interpretation of all relevant core array entries may be quite tedious, especially for interaction terms, i.e., those linking unequal components of different modes. This problem may be partially solved by a transformation of the core array to a simple structure. If **P**, **Q**, **T** are arbitrary orthogonal transformation matrices of orders (r, r), (s, s), (t, t), respectively, then the transformed core array $\bar{\mathbf{C}}^{\circ}$ has to be calculated from the transformed component matrices $\bar{\mathbf{G}} = \mathbf{GP}$, $\bar{\mathbf{H}} = \mathbf{HQ}$, $\bar{\mathbf{E}} = \mathbf{ET}$ via (12):

$$\begin{aligned} \bar{\mathbf{C}}^{\circ} &= \bar{\mathbf{G}}^{\mathrm{T}}\mathbf{Z}^{\circ}(\bar{\mathbf{H}} \otimes \bar{\mathbf{E}}) = \mathbf{P}^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\mathbf{Z}^{\circ}(\mathbf{HQ} \otimes \mathbf{ET}) \\ &= \mathbf{P}^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\mathbf{Z}^{\circ}(\mathbf{H} \otimes \mathbf{E})(\mathbf{Q} \otimes \mathbf{T}) = \mathbf{P}^{\mathrm{T}}\mathbf{C}^{\circ}(\mathbf{Q} \otimes \mathbf{T}) \end{aligned} \quad (19)$$

In this simple way the transformed core array directly computes from the original array by merely using the transformation matrices. In a similar manner one verifies that

$$\bar{\mathbf{G}}\bar{\mathbf{C}}^{\circ}(\bar{\mathbf{H}}^{\mathrm{T}} \otimes \bar{\mathbf{E}}^{\mathrm{T}}) = \mathbf{GC}^{\circ}(\mathbf{H} \otimes \mathbf{E})$$

by exploiting orthogonality of the transformation matrices. This means, that the transformed core and component matrices reproduce the same data array in (10) as the original ones. Hence, in analogy to conventional PCA, there is no loss in fit by transformation. In ref. [26] Kroonenberg

discusses both orthogonal and oblique transformations with the aim to get slice-wise diagonal core matrices.

A stronger objective would be to generate a so-called body-diagonal structure. This means that those contributions in the core array which link unequal components of different modes should be small. Since the sum of squared elements in the core matrix remains unchanged by transformations of the type (19), an equivalent task is to maximize the sum of squared body-diagonal elements (bold face numbers in (18)). In the example of water analysis the following transformed core array turns out to be optimal:

$$\left(\begin{array}{cc|cc} \mathbf{21.41} & 5.20 & -1.59 & 2.56 \\ 0.43 & 0.80 & 8.31 & \mathbf{16.49} \end{array}\right) \qquad (20)$$

The degree of body diagonality, which is the percentage of the sum of squared body-diagonal elements with respect to the total sum of squares, has increased from 58.2% in the original array (18) to 87.3% in the new array. As a consequence, the two major contributions are located in the corners and one of originally two minor contributions has disappeared by splitting into other entries. The superimposed PCA diagrams of the corresponding transformed component matrices are shown in Fig. 6. Apart from reflections, which give the axes an appropriate orientation, there are only slight modifications of the original constellations, nevertheless leading to simpler interpretation. Due to the fairly high degree of body diagonality in this example, an interpretation can be made on the basis of the superimposed plot. In analogy to the usual biplot [38] one could here speak of a triplot. Considering first all projections onto the horizontal axis, a seasonal factor of salinity is revealed which is independent of sampling stations. Then, projecting all points onto the vertical axis, a spatial factor may be detected which is independent of sampling times apart from a peak at $t = 20$. This factor is domi-

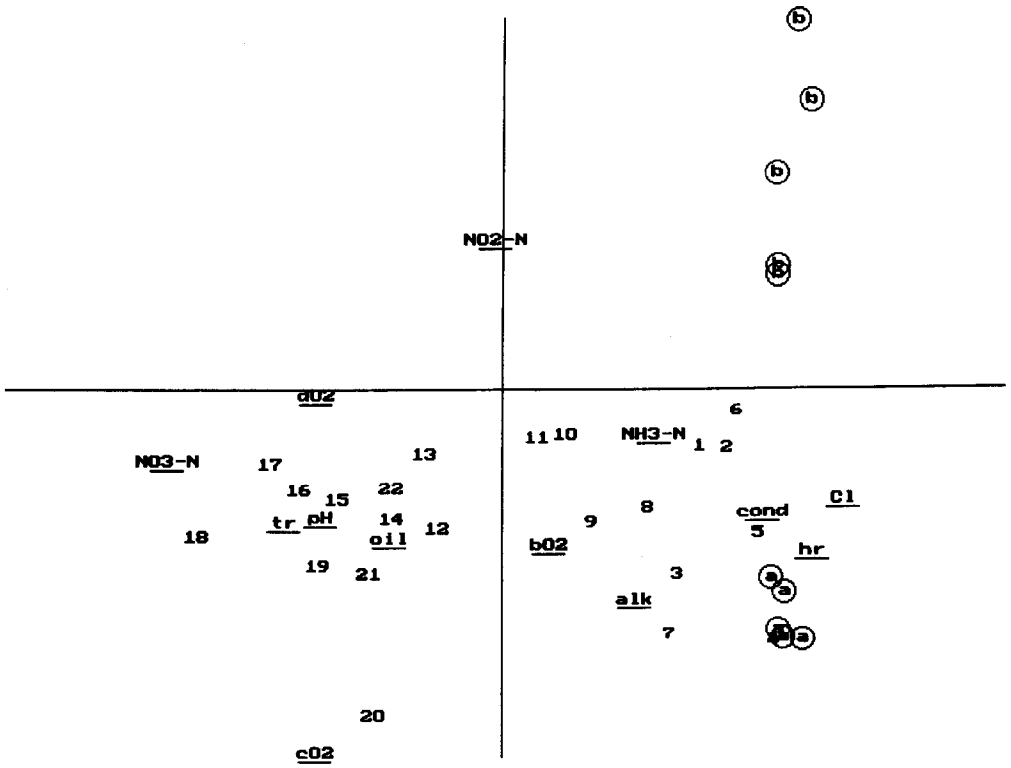

Fig. 6. Superimposed PCA plot for the example of water analysis after appropriate transformation of components.

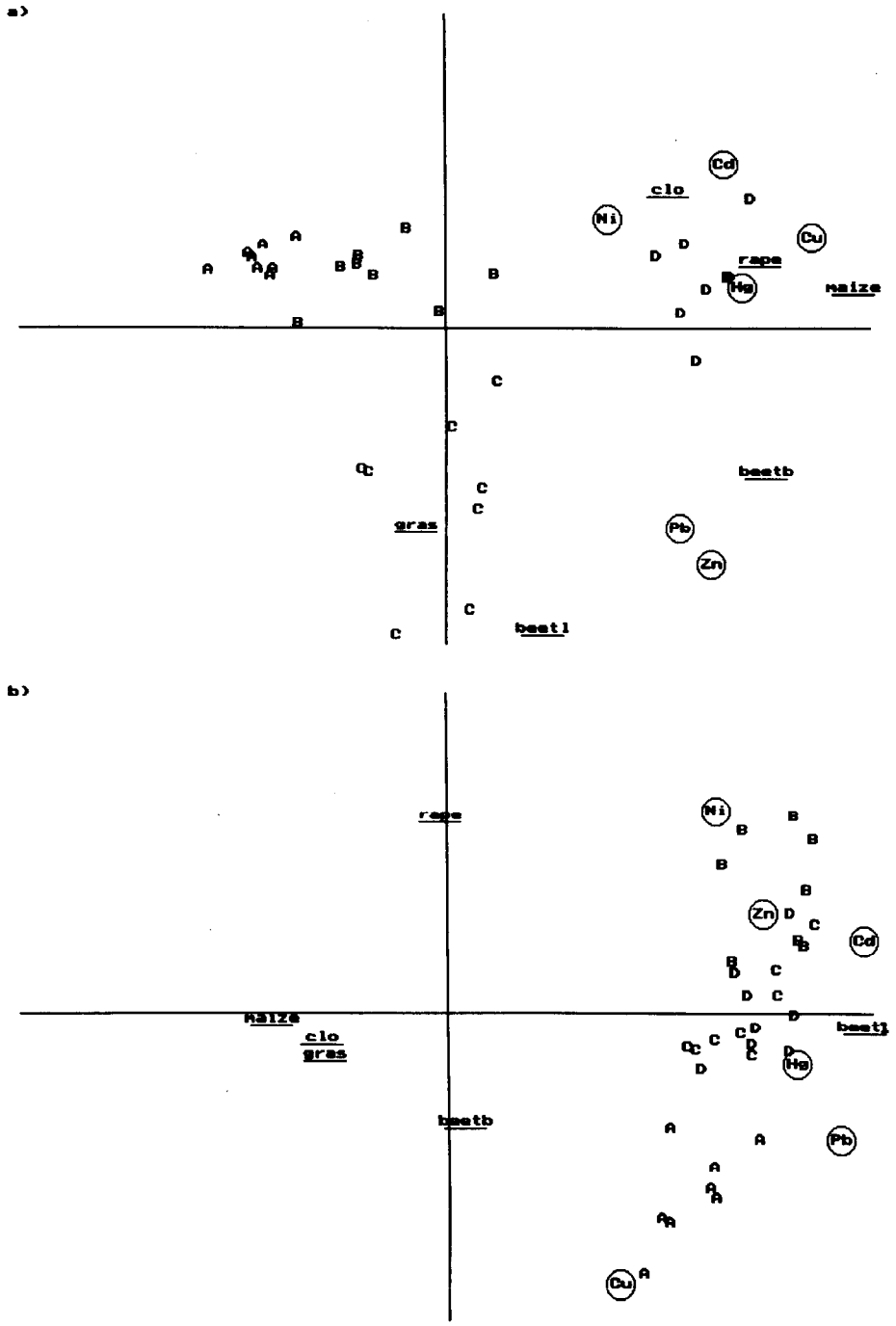

Fig. 7. Superimposed three-way PCA plots for an example of environmental chemistry after (a) *jk*-scaling and (b) *ji*-scaling of data. Repeated measurements of four different sampling regions are labeled by letters. Encircled symbols refer to the measured trace elements. The analysed fodder plants are underlined (clo = clover, beetl + beetb = leaf and body of beet).

nated by chemical oxygen demand and separates effluent-receiving from non-effluent-receiving stations and is thus probably related to environmental influences. Again, the link (2/1/2) in (20) reflects the geographic (spatial) aspect of salinity. It should be noted that the transformed diagrams do not result from simple rotations of the original plots since all components have been scaled afterwards according to (13).

Concerning the generality of the success of transforming the core matrix to a body-diagonal structure it was proved in ref. [43] theoretically as well as by simulation that in the average two-component case, which is of main interest for graphical purposes, a degree of body-diagonality can be achieved which lies in the range 80–90%.

4. Applications

In this section several practical examples demonstrate that three-way PCA is an efficient tool for extreme data reduction and for graphical analysis of the main data variance. In all cases except one a sufficiently high degree of body diagonalization was obtained after appropriate core matrix transformation. Therefore all interpretations are restricted to superimposed diagrams in the sense of discussing only projections onto equal axes for different modes (either all horizontal or all vertical axes at a time, but no mixture corresponding to interactions).

4.1. Environmental chemistry

In a comparative study the heavy-metal exposure of soils with different fodder plant cultivations was analyzed in different East-German regions. The trace concentrations of Ni, Cd, Hg, Cu, Pb, and Zn were determined in grass, rape, maize, clover (clo), body of beet (beetb) and leaf of beet (beetl) consistently at four locations A, B, C and D. Eight parallel samples were collected at each location. As a result a three-way data array with 4 locations by 8 samples = 32 objects, 6 variables (metal concentrations) and 6 'conditions' (fodder plants) could be established.

For a general characterization of different regions the influences of differing mean values and dispersions among elements and plants were removed by means of *jk*-scaling (see Fig. 5). Fig. 7a shows the superimposed PCA plot for this situation. As a consequence of *jk*-scaling objects are centered in contrast to the other modes. There is a clear separation of all four regions taking into account the dispersion of parallel samples. All metals having high positive first coordinates, the horizontal axis may be identified as a factor of general exposure which is mainly reflected in maize, clover, rape and body of beet. Accordingly the regions may be arranged with respect to their horizontal positions: D (high exposure), B, C (medium exposure) and A (low exposure). Apart from this, there occurs an independent second factor along the vertical axis. It is determined by Pb and Zn exposure in grass and beet and it separates the position of region C from the others. Since this variation concerns only a minor part of objects, variables and conditions, it should be identified as a factor of specific exposure.

Alternatively, one might wish to put emphasis on a characterization of fodder plants rather than regions. Then *ji*-scaling is preferred (see Fig. 5). Fig. 7b illustrates the resulting PC plot, now with conditions being centered. The situation appears quite symmetric to the case of regions. Again, the horizontal axis turns out to be a factor of general exposure and plants arrange by their horizontal projections according to their capacity of exposure: leaf of beet (high), body of beet and rape (medium) and maize, clover, grass (low). This characterization is valid for all regions since each object has a high positive loading on the first axis. Once more there appears a specific factor determined by rape which shall not be discussed in more detail here due to the exceptional non-body-diagonal structure of the core matrix in this situation.

4.2. Complex round robin tests

The advantageous application of multivariate data analysis to a graphical evaluation of round robin tests has been pointed out in ref. [44]. This becomes more relevant if it is not just a single test that is carried out but many routine tests

with a stable set of participating laboratories. Fig. 8 relates to an example of water control. In four succeeding routine tests (I, II, III, IV) with intervals of a half year, eight laboratories (A,...,H) determined the concentrations of Cd, Ni, Cu and Zn in synthetic water samples. Each laboratory provided five parallel estimations. Since one is mainly interested in absolute differences between laboratories rather than between elements or tests, *jk*-scaling was applied. Consequently the origin in Fig. 8 corresponds to the multivariate mean (over elements and tests) of all laboratories. At first glance one observes the outlying position of F which is obviously related to strong systematic underestimation of all elements in the first two tests (horizontal axis). To a smaller degree, laboratory C exhibits an opposite position. The vertical axis confirms that during the last two tests systematic under- or overestimations, respectively, arise for laboratories D and E in elements Cd and Zn according to the respective position of points along this axis. The subdivision of tests into two subgroups reveals that some of the laboratories changed their working procedure in the course of time. Additional information is gained if the tests are designed with true results known to the organizers as in the given example. All these 16 true concentrations (four elements in four tests) may be collected to give a multivariate theoretical reference point. Such an additional point has to be projected onto the model in the following way: Denote this point by a column vector $\boldsymbol{z}$ with pq components and further denote by $\boldsymbol{g}$ the column vector containing the desired s coordinates of $\boldsymbol{z}$ after projection onto the model (in the given diagram situation one has $s = 2$). Then, applying the transposed version of (10) to these vectors one arrives at

$$(\mathbf{H} \otimes \mathbf{E})\mathbf{C}^{oT}\boldsymbol{g} \approx \boldsymbol{z}$$

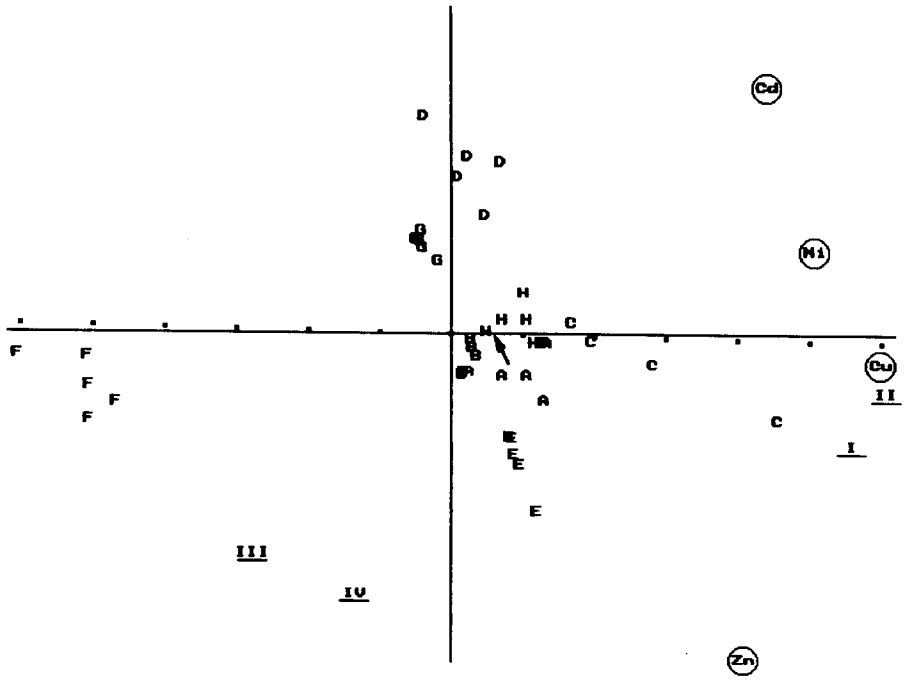

Fig. 8. Graphical three-way PCA analysis of a complex round robin test. Letters refer to repeated measurements of eight laboratories, the measured elements are encircled and underlined roman numbers represent round robin tests at subsequent times (with a distance of nearly a half year). The dotted line is the direction of systematic error.

Obviously, the desired solution g is obtained from the optimal core and component matrices via multiple linear regression:

$$g = \left[\mathbf{C}^{o}(\mathbf{H}^{T} \otimes \mathbf{E}^{T})(\mathbf{H} \otimes \mathbf{E})\mathbf{C}^{oT}\right]^{-1}\mathbf{C}^{o}(\mathbf{H}^{T} \otimes \mathbf{E}^{T})z$$

$$= \left[\mathbf{C}^{o}\mathbf{C}^{oT}\right]^{-1}\mathbf{C}^{o}(\mathbf{H}^{T} \otimes \mathbf{E}^{T})z \qquad (21)$$

due to orthonormality of component matrices. Using this derivation, an additional object is easily located in the PCA display after scaling it in the same way as the model building objects. In Fig. 8 the theoretical point is depicted as an arrow. Now, a direct multivariate evaluation of laboratories becomes possible. Best results were supplied by laboratories B, H and perhaps A. Finally it is possible to include into the plot a direction of general systematic error, i.e., uniform positive or negative deviations in all elements for all tests. Such a direction may be identified as an artificial additional object having constant components, e.g., (1, 1,... 1) (recall that, as a consequence of *jk*-scaling, the multivariate average estimation is (0, 0,..., 0). The position of such an object in the PCA display may be again computed from (21). Joining this position with the origin yields the desired line of general systematic deviations. By chance, in the present example this line is almost identical to the first axis. In general, it allows to detect laboratories with systematic failures in the working procedure. In this way one obtains a logical generalization of the so-called Youden plot [45] which is well-known in bivariate evaluation of round robin tests.

4.3. Quality control

This last example deals with quality control in tungsten wire production. It is well known that trace concentrations of certain elements heavily influence the properties of tungsten wires. Therefore it is desirable to detect sources of contamination and decontamination of such elements during the production process. Since this process is very complicated, considerations will be restricted here to the following segment: ammoniumparatungstate (APT) — tungsten oxide (TO)

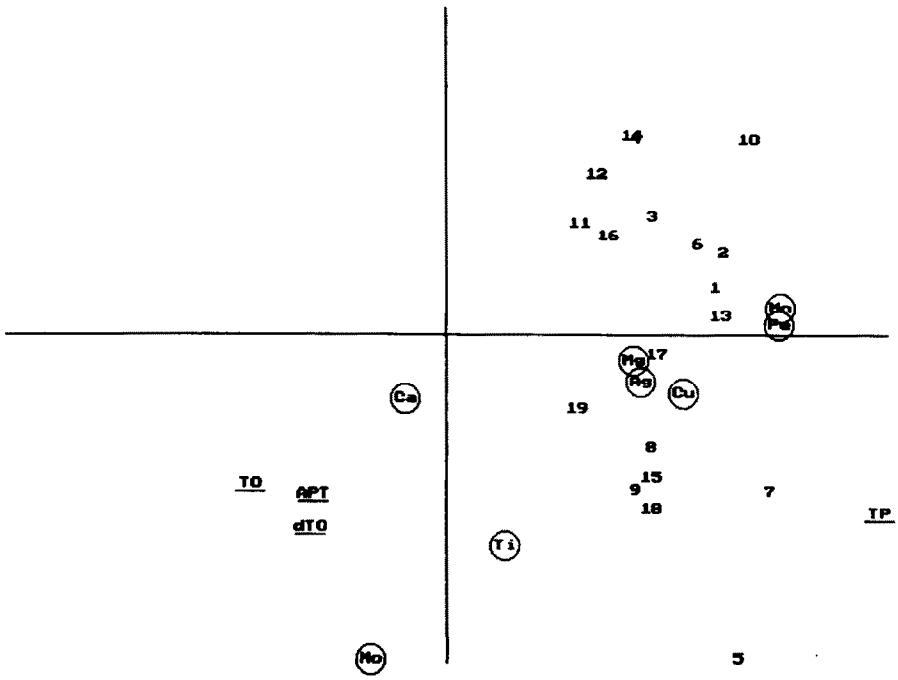

Fig. 9. Superimposed three-way PCA plots for an example from analysis of trace contamination in tungsten wire production. The concentration of eight trace elements (encircled) were measured in twenty samples (numbers) in the course of four production stages (underlined, APT = ammoniumparatungstate, TO = tungsten oxide, dTO = doped tungsten oxide, TP = tungsten powder).

— doped tungsten oxide (dTO) — tungsten powder (TP). Through these stages, 19 samples were monitored with respect to their contaminations by Mo, Ca, Ti, Mg, Ag, Cu, Mn and Fe. In order to make visible both absolute differences between samples and between stages of production, j-scaling was preferred here. The PCA plot in Fig. 9 shows that contamination patterns are stable during the first three stages but then change drastically when passing to tungsten powder. Hence, the horizontal axis reflects uniform contamination of all samples mainly by Fe and Mn in the last stage of production considered here. It has been demonstrated elsewhere [46] that this contamination is caused by the use of steel materials for the reduction of tungsten oxide. The vertical axis determines differences between samples. For Mo and Ti, these differences are stable over the whole process (predetermined by the tungsten ore).

5. N-way PCA

In principle both the model of three-way PCA and the ALS algorithm are generalized to N-way data sets ($N > 3$) in a straightforward manner. A precise definition, however, requires appropriate notation for handling the necessary N-way operations, as was pointed out by Kroonenberg [26]. Kapteyn et al. [47] were the first to provide such a notation and to give an elegant N-way formulation of Tucker's PCA model and of the ALS algorithm (see also ref. [28]). Before going into detail it will be convenient to introduce three operators on N-way data sets explained below and illustrated in Fig. 10:

Transposition operator:
T: $\mathscr{A}^{n_1,n_2,\ldots,n_N} \to \mathscr{A}^{n_N,n_{N-1},\ldots,n_1}$. Each N-way array $\vec{\mathbf{Z}}$ with general element $z_{i_1,\ldots,i_N}$ is assigned the transposed array $\vec{\mathbf{Z}}^{\mathrm{T}}$ with general element $z^{\mathrm{T}}_{i_1,\ldots,i_N} = z_{i_N,\ldots i_1}$.

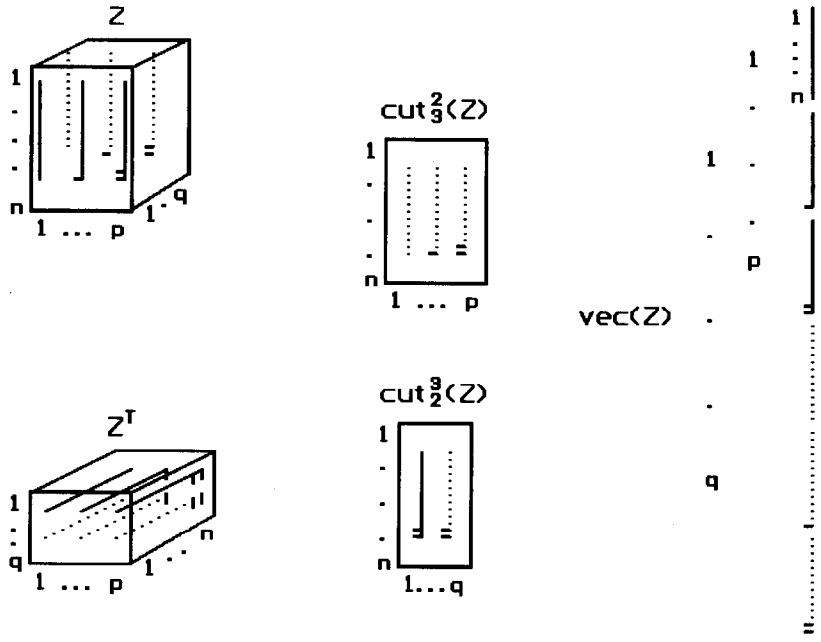

Fig. 10. Illustration of some important operators on N-way data arrays: transposition (index T), cutting of slice j (upper index) from mode k (lower index) and vectorization. For clarity, the columns of the array are marked in different ways.

Cutting operator:
cut_k^j: $\mathscr{A}^{n_1,\ldots,n_N} \to \mathscr{A}^{n_1,\ldots,n_{k-1},n_{k+1},\ldots,n_N}$
$(1 \leqslant j \leqslant n_k)$.
Each N-way array $\vec{\mathbf{Z}}$ with general element $z_{i_1,\ldots,i_N}$ is assigned the $N-1$-way array $\text{cut}_k^j(\vec{\mathbf{Z}})$ with general element $z^*_{i_1,\ldots i_{N-1}} = z_{i_1,\ldots,i_{k-1},j,i_{k+1},\ldots,i_N}$.

Vectorization operator:
vec: $\mathscr{A}^{n_1,\ldots,n_N} \to \mathscr{A}^{n_1 \cdots n_N}$. Each N-way array is assigned a 'long' vector containing all entries of the array in the following recursive arrangement:
- $(N=2)$
  vec: $\mathscr{A}^{n_1,n_2} \to \mathscr{A}^{n_1 n_2}$. Each matrix $\mathbf{Z}$ with general element z_{ij} is assigned a vector $\text{vec}(\mathbf{Z})$ with general element $z^*_{n_1(j-1)+i} = z_{ij}$ by stacking of columns.
- (general N)
  Each N-way array $\vec{\mathbf{Z}}$ with general element $z_{i_1,\ldots,n_N}$ is assigned a vector $\text{vec}(\vec{\mathbf{Z}}) := \text{vec}(\mathbf{R})$, where $\mathbf{R} \in \mathscr{A}^{n_1 \cdots n_{N-1},n_N}$ is a usual matrix having as jth column the vector $\text{vec}(\text{cut}_N^j(\vec{\mathbf{Z}}))$. This last expression is correctly defined since $\text{cut}_N^j(\vec{\mathbf{Z}})$ is an $N-1$-way array.

Transposition means reversing the order in the sequence of all modes and coincides with matrix transposition for $N=2$. The cutting operator serves to cut out the jth $N-1$-slice of the kth mode in a given array. In Fig. 1, for instance, the frontal planes (condition slices, marked by dots) would be denoted by $\text{cut}_3^1(\vec{\mathbf{Z}}), \ldots, \text{cut}_3^q(\vec{\mathbf{Z}})$. The vectorization operator is easily understood from a programming language point of view: Unfold the N-dimensional array by a sequence of N for/do loops with the last index running slowly in the outer most loop and the first index running quickly in the inner most loop.

Using these tools, it is not hard to verify, for instance, that the two-way PCA model in (3) may be equivalently formulated as $\text{vec}(\vec{\mathbf{Z}}) \approx (\mathbf{G} \otimes \mathbf{H})\text{vec}(\vec{\mathbf{C}})$. Similarly Tucker's three-way PCA model (10) rewrites as $\text{vec}(\vec{\mathbf{Z}}) \approx (\mathbf{G} \otimes \mathbf{H} \otimes \mathbf{E})\text{vec}(\vec{\mathbf{C}})$. By direct generalization one obtains an extremely simple formulation of the N-way PCA model:

$$\text{vec}(\vec{\mathbf{Z}}) \approx \left(\bigotimes_{i=1}^{N} \mathbf{A}_i \right) \text{vec}(\vec{\mathbf{C}}) \qquad (22)$$

where $\mathbf{A}_i$ are orthonormal component matrices of orders (n_i, r_i) with $r_i < n_i$ and $\otimes$ denotes the N-fold Kronecker product. One advantage of (22) is the symmetric treatment of all modes. The extension of the ALS algorithm for solving (22) runs as follows:

1. Initialization by Tucker 1 method:
   Compute orthonormal starting component matrices $\mathbf{A}_i^0$ $(i = 1, \ldots, N)$ of orders (n_i, r_i) which contain as columns the r_i leading normalized eigenvectors of the matrix $\mathbf{P}^i$ with general element $p^i_{jk} = [\text{vec}(\text{cut}_i^j(\vec{\mathbf{Z}}))]^{\mathrm{T}}[\text{vec}(\text{cut}_i^k(\vec{\mathbf{Z}}))]$ $(j, k = 1, \ldots, n_i)$.
2. Outer iteration:
   (a) Initialization of the outer iteration counter:
   Put ou = 0.
   (b) Inner iteration:
   (i) Initialization of the inner iteration counter:
   Put in = 1.
   (ii) in sub-step:
   Put $\mathbf{Q} = \mathbf{A}_1^{\text{ou}+1} \otimes \cdots \otimes \mathbf{A}_{\text{in}-1}^{\text{ou}+1} \otimes \mathbf{A}_{\text{in}+1}^{\text{ou}} \otimes \cdots \otimes \mathbf{A}_N^{\text{ou}}$. Compute an iterated component matrix $\mathbf{A}_{\text{in}}^{\text{ou}+1}$ of order (n_i, r_i) for mode 'in' as containing the r_i leading normalized eigenvectors of the matrix $\mathbf{P}$ with general element $p_{jk} = [\text{vec}(\text{cut}_{\text{in}}^j(\vec{\mathbf{Z}}))]^{\mathrm{T}}\mathbf{Q}\mathbf{Q}^{\mathrm{T}}[\text{vec}(\text{cut}_{\text{in}}^k(\vec{\mathbf{Z}}))]$.
   (iii) Inner iteration 'convergence test':
   if in $< N$ then put in := in + 1 and go back to 2bii else proceed.
   (c) Outer iteration convergence test:
   If the gain in fit of (22) by using the current component matrices is significant then put ou := ou + 1 and go back to 2b else proceed.
3. Optimal core array:
   Compute the optimal core array from the final component matrices via

$$\text{vec}(\vec{\mathbf{C}}) = \left(\bigotimes_{i=1}^{N} \mathbf{A}_i \right)^{\mathrm{T}} \text{vec}(\vec{\mathbf{Z}})$$

6. Related methods

Perhaps the most frequently applied alternative to Tucker's three-way PCA model is so-called canonical (or trilinear) decomposition which was

introduced by Carroll and Chang in ref. [3] and, starting from another viewpoint, by Harshman [48] under the name of parallel factor analysis. Since this model is easily extended to N-dimensional arrays, we only give the general formulation:

$$z_{i_1,\ldots,i_N} \approx \sum_{u=1}^{r} a_{i_1 u} a_{i_2 u} \cdots a_{i_N u}$$

or

$$\mathrm{vec}(\vec{\mathbf{Z}}) \approx \sum_{u=1}^{r} \left(\bigotimes_{i=1}^{N} \boldsymbol{a}_u^i \right) \tag{23}$$

where $\boldsymbol{a}_u^i$ refers to column u of the component matrix for mode i. The solution algorithm proposed in ref. [3] is based on a simple iterative scheme: Given starting matrices $\mathbf{A}^i$ for each mode, each of these is adjusted in an inner iteration on the basis of the $N-1$ remaining matrices and the data array by an ordinary least squares estimation (multiple linear regression). If each $\mathbf{A}^i$ has been adjusted once, one outer iteration has been accomplished. Outer iterations are continued until convergence.

There are several essential differences between trilinear decomposition and Tucker's model. First, the number of components is required to be the same in each mode. Secondly, no interactions between unequal components of different modes are taken into account, hence there occurs only one summation sign instead of three in Tucker's model. This simplification leads to a reduction of computational effort. In contrast to Tucker's model, where the component matrices are orthonormal but not uniquely defined (recall the possibility of transformations without affecting the degree of fit), the solution matrices in (23) are unique but need not be orthonormal, which makes interpretation more difficult (compare [26] p. 60). Canonical decomposition (23) appears to be appropriate for many types of instrumental data from hyphenated methods of analysis [11]. Applications and comparative discussions may be found, for instance, in refs. [8,11,49].

Apart from extending PCA a corresponding generalization of the partial least squares method (PLS) has attracted much attention in chemometrics. In the context of this method two multidimensional arrays of not necessarily the same order have to be related such that the second one may be predicted from the first one. In ref. [10] Wold et al. proposed a multiway PLS algorithm based on unfolding both arrays which is followed by standard PLS. In ref. [50] Ståhle investigated a specific three-way/two-block constellation by incorporating trilinear decomposition into the algorithm. Further methods and models being relevant in chemometrics are discussed in refs. [13,15].

Acknowledgements

The author wishes to thank Prof. O.M. Kvalheim (University of Bergen, Norway) for his motivating discussion on the subject of this paper. Thanks go to Dr. G.C. Onuoha (University of Port Harcourt, Nigeria), Dr. P. Heininger (Bundesanstalt für Gewässerkunde, Berlin, Germany), Dr. M. Stoyke (Institut für Veterinärpharmakologie und-toxikologie, Bernau, Germany) and Dr. U. Rassmann for providing data examples.

References

[1] L.R. Tucker, Implications of factor analysis of three-way matrices for measurement of change, in C.W. Harris (Editor), *Problems of Measuring Change*, University of Wisconsin Press, Madison, MI, 1963, pp. 122–137.

[2] B. Bloxom, A note on invariance in three-mode factor analysis, *Psychometrika*, 33 (1968) 347–850.

[3] J.D. Carroll and J.J. Chang, Analysis of individual differences in multidimensional scaling via an N-way generalization of 'Eckart-Young' decomposition, *Psychometrika*, 35 (1970) 283–319.

[4] J.B. Lohmöller and H. Wold, *Three-mode Path Models with Latent Variables and Partial Least Squares (PLS) Parameter Estimation*, Forschungsbericht 80.03, Fachbereich Pädagogik der Bundeswehrhochshule München, Neubiberg, Germany, 1980.

[5] P.M. Kroonenberg, Annotated bibliography of three-mode factor analysis, *British Journal of Mathematical and Statistical Psychology*, 36 (1983) 81–113.

[6] H.G. Law, C.W. Snyder, J.A. Hattie and R.P. McDonald (Editors), *Research Methods for Multimode Data Analysis*, Praeger, New York, 1984.

[7] J.B. 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.

[8] C.J. Appellof and E.R. Davidson, Strategies for analyzing data from video fluorometric monitoring of liquid chromatographic effluents, *Analytical Chemistry*, 53 (1981) 2053–2056.

[9] 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.

[10] S. Wold, P. Geladi, K. Esbensen and J. Öhman, Multi-way principal components and PLS analysis, *Journal of Chemometrics*, 1 (1987) 41–56.

[11] E. Sanchez and B.R. Kowalski, Tensorial resolution: a direct trilinear decomposition, *Journal of Chemometrics*, 4 (1990) 29–45.

[12] H.J.H. MacFie, Simultaneous multivariate analysis of multiple data matrices, in H.L.C. Meuzelaar and T.L. Isenhour (Editors), *Computer-Enhanced Analytical Spectroscopy*, Plenum, New York, 1987, pp. 103–119.

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

[14] P. Geladi, Analysis of multi-way (multi-mode) data, an overview, *Analytical Proceedings*, 27 (1990) 306–308.

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

[16] R. Henrion, G. Henrion and G.C. Onuoha, Multi-way principal components analysis of a complex data array resulting from physico-chemical characterization of natural waters, *Chemometrics and Intelligent Laboratory Systems*, 16 (1992) 87–94.

[17] J. Öhman, P. Geladi and S. Wold, Residual bilinearization. Part 2: Application to HPLC–diode array data and comparison with rank annihilation factor analysis, *Journal of Chemometrics*, 4 (1990) 135–146.

[18] O.M. Kvalheim and Y. Liang, Heuristic evolving latent projections: Resolving two-way multicomponent data. 1. Selectivity, latent-projective graph, datascope, local rank, and unique resolution, *Analytical Chemistry*, 64 (1992) 936–946.

[19] P. Geladi and K. Esbensen, Can image analysis provide information useful in chemistry?, *Journal of Chemometrics*, 3 (1989) 419–429.

[20] J.L. Lastovicka, The extension of component analysis to four mode matrices, *Psychometrika*, 46 (1981) 47–57.

[21] T. Hirschfeld, J.B. Callis and B.R. Kowalski, Chemical sensing in process analysis, *Science*, 226 (1984) 312–318.

[22] R. Coppi, Analysis of three-way data matrices based on pairwise relation measures, in F. De Antoni, N. Lauro and A. Rizzi (Editors), *COMPSTAT 1986 — Proceedings in Computational Statistics*, Physica, Heidelberg, 1986, pp. 129–139.

[23] P. Geladi, H. Isaksson, L. Lindqvist and S. Wold, Principal component analysis of multivariate images, *Chemometrics and Intelligent Laboratory Systems*, 5 (1989) 209–220.

[24] L.R. Tucker, Some mathematical notes on three-mode factor analysis, *Psychometrika*, 31 (1966) 279–311.

[25] K.G. Jöreskog and D. Sörbom, Statistical models and methods for analysis of longitudinal data, in D.J. Aigner and A.S. Goldberger (Editors), *Latent variables in Socio-Economic Models*, North Holland, Amsterdam, 1977, pp. 285–325.

[26] P.M. Kroonenberg, *Three-Mode Principal Component Analysis: Theory and Applications*, DSWO Press, Leiden, 1983.

[27] C. Eckart and G. Young, The approximation of one matrix by another of lower rank, *Psychometrika*, 1 (1936) 211–218.

[28] J.R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley, Chichester, 1988.

[29] R. Hakstian, Procedures for the factor analytic treatment of measures obtained on different occasions, *British Journal of Mathematical and Statistical Psychology*, 26 (1973) 219–239.

[30] M.H. De Groot and C.C. Li, Correlations between similar sets of measurements, *Biometrics*, 22 (1966) 781–790.

[31] N.B. Vogt, Soft modelling and chemosystematics, *Chemometrics and Intelligent Laboratory Systems*, 1 (1987) 213–231.

[32] J. Levin, Three-mode factor analysis, *Psychological Bulletin*, 64 (1965) 442–452.

[33] 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.

[34] R.R. Bellmann, *Introduction to Matrix Analysis*, McGraw-Hill, New York, 1960.

[35] P.M. Kroonenberg and J. de Leeuw, Principal component analysis of three-mode data by means of alternating least squares algorithms, *Psychometrika*, 45 (1980) 69–97.

[36] R. Penrose, On the best approximate solutions of linear matrix equations, *Proceedings of the Cambridge Philosophical Society*, 51 (1955) 406–413.

[37] H. Rutishauser, Computational aspects of F.L. Bauer's simultaneous iteration method, *Numerische Mathematik*, 13 (1969) 4–13.

[38] D. Bartussek, Zur Interpretation der Kernmatrix in der dreimodalen Faktorenanalyse von R.L. Tucker, *Psychologische Beiträge*, 15 (1973) 169–184.

[39] K.R. Gabriel, The biplot graphic display of matrices with application to principal component analysis, *Biometrika*, 58 (1971) 453–467.

[40] H. Wold, Nonlinear estimation by iterative least squares procedures, in F. David (Editor), *Research Papers in Statistics*, Wiley, New York, 1966, pp. 411–444.

[41] I. Noy-Meir, Data transformations in ecological ordination. I. Some advantages of non-centering, *Journal of Ecology*, 61 (1973) 329–341.

[42] R.A. Harshman and M.E. Lundy, Data Preprocessing and the Extended PARAFAC Model, in H.G. Law, C.W. Snyder, J.A. Hattie and R.P. McDonald (Editors), *Research Methods for Multimode Data Analysis*, Praeger, New York, 1984, pp. 216–284.

[43] R. Henrion, Body diagonalization of core matrices in

three-way principal components analysis: Theoretical bounds and simulation, *Journal of Chemometrics*, 7 (1993) 477–494.

[44] R. Henrion, G. Henrion, P. Heininger and G. Steppuhn, Statistical analysis of complex round robin tests, *Acta Hydrochimica et Hydrobiologica*, 19 (1991) 603–614.

[45] W.J. Youden, The sample, the procedure and the laboratory, *Analytical Chemistry*, 32 (1960) 23A–37A.

[46] G. Henrion, H.J. Lunk, A. Henrion and R. Henrion, Klassifizierung von sehr ähnlichen Wolframmaterialien durch multivariate Interpretation umfangreicher Analysenserien der Spurenelemente, *Zeitschrift für Chemie*, 25 (1985) 393–397.

[47] A. Kapteyn, H. Neudecker and T. Wansbeek, An approach to *n*-mode components analysis, *Psychometrika*, 51 (1986) 269–275.

[48] R.A. Harshman, Foundations of the PARAFAC procedure: Models and conditions for an 'explanatory' multimode factor analysis, *UCLA Working Papers in Phonetics*, 16 (1970) 1–84.

[49] 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.

[50] L. Ståhle, Relating multivariate time-series by linear three-way decomposition (LTD) and partial least squares (PLS) analysis, *Journal of Pharmaceutical and Biomedical Analysis*, 9 (1991) 671–678.