PSYCHOMETRIKA—VOL. 56, NO. 3, 449–470
SEPTEMBER 1991

HIERARCHICAL RELATIONS AMONG THREE-WAY METHODS

HENK A. L. KIERS

UNIVERSITY OF GRONINGEN

A number of methods for the analysis of three-way data are described and shown to be variants of principal components analysis (PCA) of the two-way supermatrix in which each two-way slice is "strung out" into a column vector. The methods are shown to form a hierarchy such that each method is a constrained variant of its predecessor. A strategy is suggested to determine which of the methods yields the most useful description of a given three-way data set.

Key words: three-way data, longitudinal data, principal components analysis.

Introduction

Three-way data are data that are classified in three ways. Longitudinal data, for example, are three-way, because of repeated observations of the same variables on the same objects. In this example the three ways pertain to three *different* classes of entities (modes) so they can be called three-mode three-way data (Carroll & Arabie, 1980). In other examples, like similarities among stimuli as observed by different observers, the data are classified in three ways using only two modes (stimuli and observers) one of which is used twice. Such data are called two-mode three-way data. Related to three-way data are scores on different sets of observation units (objects) on the same variables, or conversely, different sets of variables observed on the same objects. The latter will be referred to as multiple data sets. Multiple data sets cannot be considered three-way data because the three modes (sets, observation units and variables) are not fully crossed, but it is often possible to derive three-way data from multiple data sets by aggregating over one mode. For instance, if observations on different sets of objects are made on the same variables, a set of cross-product matrices between the variables is a two-mode three-way data set with two of the ways referring to the same mode. The analysis of multiple data sets is often performed using such derived two-mode three-way data.

Before discussing methods for the analysis of multiple data sets and three-way data, the notation used in the present paper will be described. For ordinary three-way data, the three ways refer to three different modes, where modes refer to objects, variables, and occasions. Of course, the modes could pertain to any other classes of entities, but the classes mentioned are the most common. The elements x_{ijk} denote entries in the three-way array, where $i = 1, \ldots, n$ is the object subscript, $j = 1, \ldots, m$ is the variable subscript, and $k = 1, \ldots, p$ is the occasion subscript. Matrix X_k is defined as the n by m matrix containing the elements of the k-th frontal slice of the three-way array (i.e., observations on the n objects and m variables at occasion k). For multiple data sets, only the case where observations are made on different sets of objects on the same variables is treated. Matrix X_k again contains observations on a set

The Netherlands organization for scientific research (NWO) is gratefully acknowledged for funding this project. This research was conducted while the author was supported by a PSYCHON-grant (560-267-011) from this organization. The author is obliged to Jos ten Berge and Pieter Kroonenberg.

Requests for reprints should be sent to Henk A. L. Kiers, Department of Psychology, Grote Kruisstraat 2/1, 9712 TS Groningen, THE NETHERLANDS.

0033-3123/91/0900-88024$00.75/0

of objects and variables at occasion k, but because different sets of objects with possibly different numbers of objects are observed at different occasions, the matrices X_k do not necessarily have the same orders. If n_k denotes the number of objects observed at occasion k, the order of matrix X_k is n_k by m. For both three-way data and multiple data sets, the m by m matrix of sums of cross-products among the variables for occasion k (called the cross-product matrix) is defined as $S_k \equiv X_k'X_k$. Obviously, when X_k is centered columnwise, S_k is proportional to a covariance matrix; when the columns of X_k are standardized, S_k is proportional to a correlation matrix.

A number of methods will be described for the analysis of three-way and multiple data sets, and since for the latter, only methods that are applied to derived three-way data will be treated, all methods considered here can be referred to as three-way. Of the many different methods for the analysis of three-way data, some treat the three modes symmetrically; others first provide an optimal representation of the entities of one mode, and base the representations of the entities of the other modes on this initial (incomplete) representation only. A special circumstance where this asymmetrical approach is often applied occurs for three-way data where one mode is time; for example, we may have repeated measurement data or multiple data sets resulting from measurements of different objects at different occasions. One such asymmetrical approach for the analysis of time dependent data, STATIS (L'Hermier des Plantes, 1976), proceeds as follows. First, data are compared over occasions by a principal components analysis (PCA) of the matrices $X_1, \ldots, X_p$ strung out into column vectors (considered as variables) belonging to different occasions. This is called the inter analysis because it describes relations between (inter) occasions. Next, a more detailed analysis of the three-way data is performed describing relations among the objects and variables (called the intra analysis) associated with the most strongly related occasions (summarized in the first principal component over occasions). Symmetrical approaches, like CANDECOMP/PARAFAC (Carroll & Chang, 1970; Harshman, 1970) or TUCKALS (Kroonenberg & de Leeuw, 1980), pay equal attention to the analysis of inter and intra relations. Symmetrical approaches generally yield less optimal representations for the first mode but more optimal representations for the others. Finally, asymmetrical techniques like analyse factorielle multiple (AFM; Escofier & Pagès, 1983) mainly stress the intra relations.

In the present paper, an asymmetrical view is adopted, stressing interest in inter analysis. Methods are described in terms of their adequacy to represent relations between occasions, thus focussing on stability versus change over occasions. Of course, by reordering the three-way data this focus can be shifted to a different mode. In PCA, such an objective is generally described as finding that subspace for a class of entities (in our case, occasions) that yields the best description of these entities, and which is generally quantified by minimizing the sum-of-squared Euclidean distances between the coordinates for the entities and their projections on that subspace. In the sequel, the objective of finding a subspace representation of the occasions will be replaced by the more general fitting of a model for the three-way data. Minimizing the sum-of-squared distances between observed coordinates and model coordinates will be referred to as fitting the data to a model in the least-squares sense.

Apart from the distinction between symmetric and asymmetric approaches to fitting three-way data to a model, methods may differ in terms of which matrices are fitted. Some fit a model directly to the three-way data (direct fitting). Alternatively, there are methods that fit a model to derived three-way matrices (e.g., cross-product matrices), which may be referred to as fitting derived data. The term fitting derived data is preferred over the practically equivalent term indirect fitting (Harshman & Lundy, 1984b, pp. 137–139) because it clarifies that fitting derived data is not restricted to fitting the derived three-way matrices to the models which are analogously derived from the direct fitting models, but may also pertain to fitting the derived data to other three-way models. In the following, the direct and derived data fitting approaches will be treated separately. A number of three-way methods will be described in terms of fitting a model to three-way data, which may be a rather unusual way of presenting the method, but such a description will facilitate comparison.

The main purpose of the present paper is to show that both the direct and derived data fitting methods can be related in a simple way because both types form a hierarchy. Explicitly, the methods discussed optimize the same criterion function, but those lower in the hierarchy optimize this function subject to more severe constraints than those higher in the hierarchy, which results in poorer model fits. On the other hand, the more severely constrained methods fit a simpler model, yielding an easier interpretation. On the basis of these hierarchical relations between the methods, a strategy is suggested to find which three-way method in the hierarchy yields the most useful representation of a given data set. Similar, but slightly different hierarchical relations have been described by Carroll and Chang (1970, p. 312), Carroll and Wish (1974, pp. 92–96), Kroonenberg (1983, pp. 49 ff), Harshman and Lundy (1984b), and Lundy, Harshman, and Kruskal (1989, p. 128). The hierarchies given here differ from those given elsewhere in that they are based on describing all methods as a PCA of occasions. In addition, the incorporation of some French three-way methods and orthogonally constrained versions of CANDECOMP/PARAFAC and INDSCAL in these hierarchies appears new.

The main part of the paper consists of the description of a number of three-way methods, stressing technical aspects on which comparisons between the methods can be based. For more substantive and interpretational aspects, the reader is referred to the sources mentioned. The methods will be treated in a simplified way, and although some methods to be described allow for adaptation of the metrics for the objects and the variables, it will be assumed that the matrices defining the metric for variables and objects are all equal to the identity. This does not reduce generality because these metric matrices may be assumed built into the data matrices $X_1, \ldots, X_p$. Also, the matrices $X_1, \ldots, X_p$ may be centered in various ways before being submitted to a three-way analysis. In the descriptions of the three-way methods to be treated, such preprocessing procedures are assumed incorporated in the matrices $X_1, \ldots, X_p$. In the next two sections, the direct and derived data fitting methods will be described. Within each section, the order of description corresponds to the order in which the methods from a hierarchy from least to most restricted.

Direct Fitting

PCA of a Derived Two-Way Supermatrix

The first method to be described is PCA of a derived two-way supermatrix (PCA-SUP). One of the steps in analyse triadique, a method proposed by Jaffrennou (1978), consists of a PCA over occasions. Thus, this step is described as PCA on the matrices $X_1, \ldots, X_p$, strung out rowwise into p column vectors of order nm. This PCA is equivalent to the method Tucker (1966) uses for finding an approximate solution for the occasion components in his three-mode factor analysis model.

To provide the mathematical description of this method, it is useful to describe the model on which the representation for occasions is based. For PCA, this model for the three-way data is given by the projection of the variables Vec $(X_1), \ldots,$ Vec (X_p) on the subspace spanned by the principal components Vec $(F_1), \ldots,$ Vec (F_r), where Vec $(\cdot)$ denotes the vector containing all the elements of the matrix strung out rowwise into a column vector. All vectors are of order nm, corresponding to the p known $n \times m$ matrices $X_1, \ldots, X_p$, and the r unknown $n \times m$ matrices $F_1, \ldots, F_r$, respectively. The projection coordinates for the variables on the principal components are denoted by c_{kl} (also called loadings for the variables on the principal components), $k = 1, \ldots, p$; $l = 1, \ldots, r$. Hence, the PCA model for the three-way data can be expressed as

$$\hat{x}_{ijk} = \sum_{l=1}^{r} f_{ijl} c_{kl}, \tag{1}$$

where f_{ijl} denotes the element (i, j) of F_l. PCA-SUP reduces to fitting the PCA model in (1) to the observed three-way data, which implies minimizing the loss function

$$\begin{aligned} \text{PCASUP}(C, F_1, \ldots, F_r) &= \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{p} \left(x_{ijk} - \sum_{l=1}^{r} f_{ijl} c_{kl} \right)^2 \\ &= \sum_{k=1}^{p} \left\| \text{Vec}\,(X_k) - \sum_{l=1}^{r} c_{kl}\ \text{Vec}\,(F_l) \right\|^2, \end{aligned} \tag{2}$$

over arbitrary matrices C and $F_1, \ldots, F_r$, where C is a $p \times r$ matrix of component loadings for occasions, and $F_1, \ldots, F_r$ $(n \times m)$ contain component coordinates for the nm object-variable combinations.

If the columns Vec $(X_1), \ldots,$ Vec (X_p) are collected in a matrix X $(nm \times p)$, and the columns Vec $(F_1), \ldots,$ Vec (F_r) in a matrix F $(nm \times r)$, the loss function can be rewritten as

$$\text{PCASUP}(C, F) = \|X - FC'\|^2. \tag{3}$$

Obviously, minimizing (3) over arbitrary matrices F and C of appropriate orders is simply a PCA on the $nm \times p$ matrix X. In PCA-SUP, the columns of this data matrix refer to occasions and the rows to object-variable combinations.

TUCKALS-3

Tucker (1966) proposed a model for three-mode factor analysis based on reducing the dimensionality of all three modes to describe the information in the three-way data through a limited number of factors. The three modes are treated symmetrically. Tucker's model can be given as

$$\hat{x}_{ijk} = \sum_{u=1}^{r_1} \sum_{v=1}^{r_2} \sum_{l=1}^{r} a_{iu} b_{jv} c_{kl} g_{uvl}, \tag{4}$$

where r_1, r_2, and r are the (reduced) dimensionalities of the component spaces for the three modes: a_{iu}, $i = 1, \ldots, n$, $u = 1, \ldots, r_1$, is the component score of object i on the u-th object-component (idealized object); b_{jv}, $j = 1, \ldots, m$, $v = 1, \ldots, r_2$, is the loading of variable j on the v-th variable-component (idealized variable); c_{kl}, $k = 1, \ldots, p$, $l = 1, \ldots, r$, is the loading of occasion k on the l-th occasion-component (idealized occasion), and the three-way array G is the so-called core of order $r_1 \times r_2 \times r$ with elements g_{uvl} denoting relations between idealized objects, idealized variables, and idealized occasions. Note that C in (1) and C in (4) are not the same matrices in general. The same symbol is used because in both cases C denotes a $p \times r$ matrix of occasion coordinates on r dimensions. The model given by (4) is called the. Tucker-3 model because it specifies component scores for all three modes in reduced dimensionalities.

Kroonenberg and de Leeuw (1980) have provided a method for the least squares fitting of the Tucker-3 model (called TUCKALS-3), which minimizes the function

$$\text{TUCKALS3}(A, B, G_1, \ldots, G_r, C) = \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{p} \left(x_{ijk} - \sum_{u=1}^{r_1} \sum_{v=1}^{r_2} \sum_{l=1}^{r} a_{iu} b_{jv} c_{kl} g_{uvl} \right)^2$$

$$= \sum_{k=1}^{p} \left\| X_k - A \sum_{l=1}^{r} c_{kl} G_l B' \right\|^2, \tag{5}$$

over (columnwise orthonormal) matrices $A(n \times r_1)$, $B(m \times r_2)$, and $C(p \times r)$, and arbitrary matrices $G_1, \ldots, G_r$ of order $r_1 \times r_2$. The phrase columnwise orthonormal is put between parentheses because this constraint can be imposed on the matrices without loss of generality. That is, minimizing the TUCKALS-3 loss function yields the same minimum whether or not the matrices A, B, and C are constrained to be columnwise orthonormal. For computational and interpretational convenience, it seems useful to impose these inactive constraints.

To align the description of TUCKALS-3 to that of PCA-SUP, the data matrices in (5) and their model description are strung out into column vectors and collected in one matrix of order $nm \times p$. We then have

$$\text{TUCKALS3}(A, B, \bar{G}, C) = \sum_{k=1}^{p} \left\| \text{Vec}\,(X_k) - \text{Vec}\left(A \sum_{l=1}^{r} c_{kl} G_l B' \right) \right\|^2$$

$$= \sum_{k=1}^{p} \left\| \text{Vec}\,(X_k) - (A \otimes B)\,\text{Vec}\left(\sum_{l=1}^{r} c_{kl} G_l \right) \right\|^2$$

$$= \|(\text{Vec}\,(X_1)| \ldots | \text{Vec}(X_p)) - (A \otimes B)$$

$$(\text{Vec}\,(G_1)| \ldots | \text{Vec}\,(G_r))C'\|^2$$

$$= \|X - (A \otimes B)\bar{G}C'\|^2, \tag{6}$$

where $\otimes$ denotes the Kronecker product, matrix $\bar{G}(r_1 r_2 \times r)$ is defined by $\bar{G} \equiv (\text{Vec}\,(G_1)| \ldots | \text{Vec}(G_r))$, and $A(n \times r_1)$, $B(m \times r_2)$, and $C(p \times r)$, are columnwise orthonormal matrices. Minimizing (6) over A, B, C, and $\bar{G}$ is equivalent to minimizing the PCA-SUP loss function (3) over arbitrary C and over all matrices F that can be written as $F = (A \otimes B)\bar{G}$, for certain matrices A, B, and $\bar{G}$ of the orders mentioned. Obviously, the set of all such matrices F is a subset of $\mathbb{R}^{nm \times r}$. In PCA-SUP, function (3) is minimized over *all* matrices F in $\mathbb{R}^{nm \times r}$. Because TUCKALS-3 minimizes the PCA-SUP loss function (3) over matrices F in a subset of $\mathbb{R}^{nm \times r}$, TUCKALS-3 can be seen as a constrained variant of PCA on the $nm \times p$ matrix X. As a consequence, TUCKALS-3 never yields a better fit for X than does PCA-SUP. This description of TUCKALS-3 as a constrained variant of PCA has been provided more or less implicitly by several authors (e.g., Weesie & Van Houwelingen, 1983). Moreover, Kroonenberg and de Leeuw (1980) remark that reducing only one of the three modes in TUCKALS-3 leads to a model that they call the Tucker-1 model, which in fact is equivalent to the model fitted by PCA-SUP. Obviously, reducing only the third mode in the Tucker-3 model is equivalent to setting $r_1 = n$ and $r_2 = m$, from which it follows that fitting the Tucker-3 model with $r_1 < n$ and $r_2 < m$ is a constrained version of fitting the Tucker-1 model.

Considering fit, constrained PCA (TUCKALS-3) is less useful than unconstrained PCA because constrained PCA yields a poorer fit of the data than unconstrained PCA. Explaining the relations between objects and variables, however, TUCKALS-3 yields more useful information than unconstrained PCA because TUCKALS-3 provides coordinates for the objects and the variables, and these coordinates for the objects and the variables are linked to the matrix F, which gives coordinates for object-variable combinations. In TUCKALS-3 these coordinates for objects and variables are given by matrices A and B. In addition, TUCKALS-3 provides measures that indicate the interaction-relations between different components for the objects and variables (given in the matrices $G_1, \ldots, G_r$). The latter, however, are difficult to interpret because they are relations between idealized objects and idealized variables considered in relation to the idealized occasions (indicated by the subscripts of the matrices $G_1, \ldots, G_r$).

CANDECOMP/PARAFAC

Carroll and Chang (1970) and Harshman (1970) independently developed a model that decomposes a three-way array in a very simple manner. Harshman called his model PARAFAC (PARAllel FACtor analysis), whereas Carroll and Chang christened their method CANDECOMP (CANonical DECOMPosition). The CANDECOMP/PARAFAC model is based on a very simple rationale. The model expression for any entry in the three-way array X is

$$\hat{x}_{ijk} = \sum_{l=1}^{r} a_{il} b_{jl} c_{kl}. \tag{7}$$

The elements a_{il}, b_{jl}, and c_{kl} are component coordinates of the objects, variables, and occasions, respectively, on the l-th CANDECOMP/PARAFAC component. According to the model, there are only proportional differences with respect to each of the components, between subjects, variables and occasions.

The interpretational difficulties in TUCKALS-3 concerning the matrices $G_1, \ldots, G_r$ are overcome by the CANDECOMP/PARAFAC method. For CANDECOMP/PARAFAC, the matrices $G_1, \ldots, G_r$ lose their function of relating components for different ways to each other, because the model does not allow for relations between components for different modes. In fact, CANDECOMP/PARAFAC provides only one set of components instead of three as does TUCKALS-3. These components can be interpreted as components for all modes simultaneously, which makes interpreting the results much easier than for a TUCKALS-3 analysis because in the latter, interpretations of components for different modes can only be linked through the elements of G_l, $l = 1, \ldots, r$. The interpretation of results from TUCKALS-3 is not merely complicated in itself, it is further complicated by the fact that the solutions of TUCKALS-3 have rotational freedom. The CANDECOMP/PARAFAC model does not allow for rotation of its components and provides unique axes.

The CANDECOMP/PARAFAC model is fitted to the data by minimizing

$$\text{CP}(A, B, \Lambda_1, \ldots, \Lambda_p) = \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{p} \left(x_{ijk} - \sum_{l=1}^{r} a_{il} b_{jl} c_{kl} \right)^2$$

$$= \sum_{k=1}^{p} \|X_k - A\Lambda_k B'\|^2, \tag{8}$$

over arbitrary matrices A, B, and $\Lambda_1, \ldots, \Lambda_p$, where matrices A and B are of order n by r, and m by r, respectively, and Λ_k is a diagonal matrix of order r, with the elements of the k-th row of C on its diagonal (i.e., $\lambda_{kl} \equiv c_{kl}$, for $k = 1, \ldots, p$ and $l = 1, \ldots, r$).

To relate CANDECOMP/PARAFAC to PCA-SUP, (8) is rewritten as

$$\text{CP}(A, B, C) = \sum_{k=1}^{p} \| \text{Vec}(X_k) - \text{Vec}(A\Lambda_k B')\|^2$$

$$= \sum_{k=1}^{p} \left\| \text{Vec}(X_k) - \text{Vec}\left(\sum_{l=1}^{r} \mathbf{a}_l \mathbf{b}_l' \lambda_{kl} \right) \right\|^2$$

$$= \sum_{k=1}^{p} \left\| \text{Vec}(X_k) - \sum_{l=1}^{r} \text{Vec}(\mathbf{a}_l \mathbf{b}_l') \lambda_{kl} \right\|^2$$

$$= \|X - (\text{Vec}(\mathbf{a}_1 \mathbf{b}_1') | \ldots | \text{Vec}(\mathbf{a}_r \mathbf{b}_r'))C'\|^2, \tag{9}$$

where X is the $nm \times p$ matrix defined above, $\mathbf{a}_l$ and $\mathbf{b}_l$ denote the l-th columns of A and B, respectively. Minimizing (9) is equivalent to minimizing the PCA-SUP loss function (3) over C and F, subject to the constraint that F can be written as $F = (\text{Vec}(\mathbf{a}_1 \mathbf{b}_1') | \ldots | \text{Vec}(\mathbf{a}_r \mathbf{b}_r'))$.

In CANDECOMP/PARAFAC the constraints imposed on PCA-SUP are stronger than those for TUCKALS-3, provided that the numbers of dimensions for the objects (r_1) and for the variables (r_2) are larger than or equal to the number of dimensions for occasions. This can be seen by verifying that CANDECOMP/PARAFAC can be considered as TUCKALS-3 with the matrices $G_1, \ldots, G_r$ constrained such that $g_{uvl} = 1$ when $u = v = l$, and 0 otherwise, provided that $r_1 \geq r$, and $r_2 \geq r$ (see Carroll & Chang, 1970, p. 312). Of course, the advantages of the stronger and simpler model fitted by CANDECOMP/PARAFAC are offset by the expected loss of fit for this more heavily constrained version of PCA-SUP.

As in TUCKALS-3, matrices A and B can be constrained to be columnwise orthonormal. However, in the CANDECOMP/PARAFAC model this cannot be done without loss of generality; thus, by constraining A and/or B to be columnwise orthonormal, the fit for the CANDECOMP/PARAFAC model will usually decrease. The method that fits the CANDECOMP/PARAFAC model by minimizing (8) while A is orthogonally constrained will be called ORTCP-A; the variant in which B is orthogonally constrained, ORTCP-B; and the variant in which both A and B are orthogonally constrained, ORTCP. Imposing orthogonality constraints on A and/or B is sometimes done to avoid degenerate CANDECOMP/PARAFAC solutions (see Harshman and Lundy, 1984a; Lundy et al., 1989).

Obviously, any variant of CANDECOMP/PARAFAC that requires A or B to be columnwise orthonormal is a variant of PCA-SUP that is even more constrained than CANDECOMP/PARAFAC. Therefore, ORTCP-A, ORTCP-B, and ORTCP are constrained variants of CANDECOMP/PARAFAC, and ORTCP is a constrained variant of ORTCP-A and of ORTCP-B. In contrast to imposing orthogonality constraints on the matrices A and B, imposing orthogonality on C is not discussed here because the resulting method does not have hierarchical relations with the method to be discussed next, SUMPCA.

SUMPCA

There seems to be no name for the method that consists of simply performing a PCA on the sum of matrices $X_1, \ldots, X_p$, which can be described mathematically as minimizing the function

$$\begin{aligned}
\text{SUMPCA}(A, B, \Lambda) &= \left\| \sum_{k=1}^{p} X_k - A\Lambda B' \right\|^2 \\
&= \text{tr} \left(\sum_{k=1}^{p} X_k \right)' \left(\sum_{k=1}^{p} X_k \right) - 2 \text{ tr} \left(\sum_{k=1}^{p} X_k \right)' A\Lambda B' \\
&\quad + \text{tr } A\Lambda B'B\Lambda A' \\
&= \underline{c + p} \sum_{k=1}^{p} \text{tr } X_k'X_k - 2 \sum_{k=1}^{p} \text{tr } X_k'A\Lambda B' \\
&\quad + \text{tr } A\Lambda B'B\Lambda A' \\
&= p \sum_{k=1}^{p} \|X_k - A(p^{-1}\Lambda)B'\|^2 + c, \qquad (10)
\end{aligned}$$

over the columnwise orthonormal matrices $A(n \times r)$ and $B(m \times r)$, and the diagonal matrix Λ of order r (c is a constant). Matrix A contains object coordinates and B contains variable loadings. At the minimum of SUMPCA, matrix Λ will contain the r singular values that correspond to the r principal components of $\Sigma_k^p X_k$.

From (10), SUMPCA can be described as fitting the three-way data to a model given by

$$\hat{x}_{ijk} = \sum_{l=1}^{r} p^{-1} a_{il} b_{jl} \lambda_l, \qquad (11)$$

where λ_l is the l-th element of Λ. SUMPCA can be related to PCA-SUP, in an admittedly forced way, by describing it as minimizing the function

$$\begin{aligned}
\text{SUMPCA}^*(A, B, C) &= \sum_{k=1}^{p} \|X_k - A\Lambda^* B'\|^2 \\
&= \underline{\sum_{k=1}^{p}} \left\| \text{Vec }(X_k) - \sum_{l=1}^{r} \text{Vec }(\mathbf{a}_l \mathbf{b}_l') \lambda_l^* \right\|^2
\end{aligned}$$

$$= \|X - (\text{Vec}\ (\mathbf{a}_1\mathbf{b}_1') \mid \ldots \mid \text{Vec}\ (\mathbf{a}_r\mathbf{b}_r'))C'\|^2, \qquad (12)$$

where $X(nm \times p)$ is defined as above, $A^* = p^{-1}A$, and $c_{kl} \equiv \lambda_l^*$ (see (10)). Minimizing (12) reduces to minimizing the PCA-SUP loss function (3) over C and F, subject to the constraints that C has equal rows, and F can be written as $F = (\text{Vec}\ (\mathbf{a}_1\mathbf{b}_1')| \ldots | \text{Vec}\ (\mathbf{a}_r\mathbf{b}_r'))$ for certain columnwise orthonormal matrices A and B. Clearly, SUMPCA is a variant of PCA-SUP that is even more constrained than ORTCP.

It is somewhat forced to consider SUMPCA as a constrained variant of PCA over occasions because its constraints have caused any resemblance with PCA over occasions to disappear. Whereas all previously treated constrained variants of PCA-SUP had at least the matrix of loadings for occasions (C) in common with unconstrained PCA-SUP, even this correspondence is lost for SUMPCA. Nevertheless, it is treated here because it is an interesting extreme of a series of variants of PCA. It provides the simplest model for PCA over occasions because it prescribes all occasions to have the same loadings (on each of the components). This is of interest when data are (approximately) stable over occasions.

Fitting Derived Data

A number of methods for fitting derived data will be described in this section. As mentioned, fitting derived data implies the use of a set of derived matrices, such as $S_1, \ldots, S_p$, instead of fitting the three-way data consisting of $X_1, \ldots, X_p$. Which derived matrices are fitted depends on the presence of more than one set of objects (and only one set of variables) or more than one set of variables (and only one set of objects). In the first case, the derived data fitting approach is applied to matrices of cross-products between variables (often covariances). In the second, the matrices of scalar products between objects are fitted. Given one set of objects and one set of variables, both approaches can be applied. Technically, the second case is analogous to the first, and therefore, in the description of the methods given here, one set of variables is assumed and the cross-product matrices $S_1, \ldots, S_p$ will be fitted to their model descriptions.

In the derived data fitting methods treated, the models for the matrices $S_1, \ldots, S_p$ are essentially the same as those for $X_1, \ldots, X_p$. Therefore, the description in the present section is analogous to that of the preceding. The only important difference between the models for S_k and those for X_k is that because S_k is symmetric, the model is also chosen to be symmetric by replacing any matrix A in the preceding section by B in the present (i.e., the loadings for variables as row entries (in A) must be equal to the loadings for variables as column entries (in B)). This replacement of B by A in fact constrains the models. For SUMPCA, this constraint is inactive because when applied to symmetrical matrices $S_1, \ldots, S_p$, minimizing the loss function over A and B automatically yields equal matrices A and B. However, for TUCKALS-3 and CANDECOMP/PARAFAC, this is not necessarily the case. For CANDECOMP/PARAFAC this was pointed out by ten Berge and Kiers (1991) who give examples with solutions where A and B differ.

STATIS

STATIS was developed by L'Hermier des Plantes (1976), based on Escoufier (1973), as a method for performing PCA on a set of matrices as if the matrices were variables. STATIS is defined by a three-step procedure. The first consists of performing PCA on a set of matrices, considered as variables, which are usually the cross-product matrices $S_1, \ldots, S_p$. This first step will be called STATIS-1, to denote it as a separate method not necessarily followed by the next two steps of STATIS.

STATIS-1 is in fact PCA-SUP applied to cross-product matrices, and can be described as minimizing the sum of squared differences between the coordinates of the occasions (scores s_{ijk}) and the coordinates of their projection on a low-dimensional subspace ($\hat{s}_{ijk} = \Sigma_l^r f_{ijl} c_{kl}$). This reduces to minimizing

$$\text{STATIS1 } (F, C) = \sum_{k=1}^{p} \left\| \text{Vec } (S_k) - \sum_{l=1}^{r} c_{kl} \text{ Vec } (F_l) \right\|^2, \tag{13}$$

over arbitrary matrices $C(p \times r)$, and $F_1, \ldots, F_r$ $(m \times m)$. Note that F_l, $l = 1, \ldots, r$, is symmetric, which is convenient because F_l is a matrix where the row and column entities are the same. Vec (F_l) is the l-th principal component of the variables Vec $(S_1), \ldots,$ Vec (S_p), and C contains the loadings c_{kl} of the occasions on the components.

The PCA-SUP model $\hat{x}_{ijk} = \Sigma_l^r f_{ijl} c_{kl}$ has been described by Kroonenberg and de Leeuw (1980) as the Tucker-1 model because it is a variant of the Tucker-3 model in which only one mode is reduced. Because STATIS-1 fits the symmetric variant of the PCA-SUP model, STATIS-1 is also equivalent to fitting the Tucker-1 model. Kroonenberg and de Leeuw also mention that fitting the Tucker-1 model to similarity matrices (i.e., STATIS-1) "is identical to the procedure developed by Tucker & Messick (1963, pp. 336 ff)" (pp. 78–79). This statement lacks some precision, however, because although the procedures are similar indeed, the Tucker and Messick approach is applied to vectors containing the *dissimilarities* (instead of similarities) among the $n(n - 1)/2$ *different* stimulus pairs (instead of all n^2 possible pairs, including double pairs, and pairs containing the same stimulus), and would therefore yield different results.

Like PCA-SUP, STATIS-1 can be described as PCA on the $m^2 \times p$ matrix S containing the column vectors Vec $(S_1), \ldots,$ Vec (S_p). If F is the $m^2 \times r$ matrix containing the column vectors Vec $(F_1), \ldots,$ Vec (F_r), then (13) can be rewritten

$$\text{STATIS1 } (F, C) = \|S - FC'\|^2. \tag{14}$$

The second step of STATIS consists of defining the compromise matrix as the first principal component (F_1) of the matrices $S_1, \ldots, S_p$. That is, assuming that α_k gives the first principal component weight for matrix S_k, the compromise is given by $F_1 = \Sigma_k^p \alpha_k S_k = \Sigma_k^p \alpha_k X_k' X_k$. The third step, which will be called STATIS-3, consists of PCA on the compromise matrix defined in the second step. To make this step comparable to the methods described above, it will be phrased in terms of minimizing a loss function. It is readily verified that PCA on matrix F_1 is equivalent to minimizing

$$\text{STATIS3}(B, \Lambda) = \left\| \sum_{k=1}^{p} \alpha_k S_k - B\Lambda B' \right\|^2$$

$$= p \sum_{k=1}^{p} \|\alpha_k S_k - B(p^{-1}\Lambda)B'\|^2 + c, \tag{15}$$

over diagonal matrices Λ, and $B(m \times r)$ subject to $B'B = I_p$ (in (15), c does not depend on B and Λ). At the minimum of the STATIS-3 function, B will contain the compromise component scores for the objects, and Λ will contain the corresponding eigenvalues. Procedures for interpreting these and other parts of the solution of a STATIS analysis are given by, for instance, Glaçon (1981), Lavit (1985, 1988), and Lechevallier (1987), but will not be treated here.

Three-Mode Scaling

The Tucker-3 model has been described for the matrices $X_1, \ldots, X_p$, but obviously, the model can also be formulated and slightly adapted for the S_k matrices (Tucker, 1972). Fitting this model is called three-mode scaling (see Kroonenberg, 1983, pp. 52–53), and reduces to the minimization of the function

$$\text{TUCKALS3*}(B, G_1, \ldots, G_r, C) = \sum_{k=1}^{p} \left\| S_k - B \sum_{l=1}^{r} c_{kl} G_l B' \right\|^2, \qquad (16)$$

over matrices $B(m \times r')$, $G_1, \ldots, G_r$ $(r' \times r')$, and $C(p \times r)$, where B and C can be constrained to be column-wise orthonormal. In analogy to the case of directly fitting the TUCKALS-3 model, three-mode scaling can be described as a constrained version of STATIS-1 on the matrices $S_1, \ldots, S_p$. To show this, the loss function (16) is rewritten as

$$\text{TUCKALS3*}(B, G_1, \ldots, G_p, C) = \|(\text{Vec}\,(S_1)| \ldots | \text{Vec}\,(S_p)) - (B \otimes B)$$
$$(\text{Vec}\,(G_1)| \ldots | \text{Vec}\,(G_r))C'\|^2. \qquad (17)$$

Clearly, minimizing (17) over arbitrary matrices $B(m \times r')$, $G_1, \ldots, G_r$ $(r' \times r')$, and $C(p \times r)$ is equivalent to minimizing the STATIS-1 loss function (14) over C and F, subject to the constraint that $F(m^2 \times r)$ can be written as $F = (B \otimes B)(\text{Vec}\,(G_1)| \ldots |\text{Vec}\,(G_r))$, for certain matrices B and $G_1, \ldots, G_r$ of appropriate orders. In this way, three-mode scaling can be seen as constrained STATIS-1, where the matrix C yields loadings for the occasions.

INDSCAL

When the CANDECOMP/PARAFAC model is applied to $S_1, \ldots, S_p$ and slightly modified, we find the INDSCAL model for scalar products as described by Carroll and Chang (1970). Fitting the INDSCAL model reduces to minimizing

$$\text{INDSCAL}\,(B, \Lambda_1, \ldots, \Lambda_p) = \sum_{k=1}^{p} \|S_k - B\Lambda_k B'\|^2, \qquad (18)$$

over B and the diagonal matrices $\Lambda_1, \ldots, \Lambda_p$, where $\lambda_{kl} \equiv c_{kl}$. Usually, λ_{kl} is required to be nonnegative. It is of interest to mention here that INDSCAL is not only a method for derived data fitting, but also for indirectly fitting the ORTCP-A model. The method for indirectly fitting the unconstrained CANDECOMP/PARAFAC model, PARAFAC-2 (Harshman, 1972) is not treated because it cannot be seen as a constrained variant of STATIS-1 over occasions.

INDSCAL can be described in terms of PCA on an $m^2 \times p$ data matrix S in a similar way as CANDECOMP/PARAFAC (see (9)) by rewriting (18) as

$$\text{INDSCAL}\,(B, C) = \|S - (\text{Vec}\,(\mathbf{b}_1\mathbf{b}_1')| \ldots | \text{Vec}\,(\mathbf{b}_r\mathbf{b}_r'))C'\|^2. \qquad (19)$$

Obviously, minimizing (19) over arbitrary B and (nonnegative) C is equivalent to minimizing the STATIS-1 loss function (14) over C and F, subject to the constraint that $F = (\text{Vec}\,(\mathbf{b}_1\mathbf{b}_1')| \ldots | \text{Vec}\,(\mathbf{b}_r\mathbf{b}_r'))$ for some $m \times r$ matrix B (and C is nonnegative). In the same way as CANDECOMP/PARAFAC is a constrained variant of TUCKALS-3 provided that $r_1 \geq r$ and $r_2 \geq r$, INDSCAL is a constrained version of three-mode scaling provided that $r' \geq r$.

The method that minimizes (18) subject to the constraint $B'B = I_r$ is called orthogonally constrained INDSCAL, denoted by the acronym INDORT. Obviously, the INDORT method is a constrained variant of STATIS-1.

SUMPCA for Cross-Product Matrices

Levin (1966) developed a method for the simultaneous factor analysis of multiple data sets based on PCA of the sum of cross-product matrices $S_1, \ldots, S_p$. His method is equivalent to one of the stages in Tucker's three-mode principal components analysis (Tucker, 1966). As has been shown by Jaffrennou (1978), this in turn is equivalent to Jaffrennou's method for analyzing a three-mode array. Finally, Gower's principal coordinates analysis (Gower, 1966) can be seen to be equivalent to this method as well, when one considers the cross-product matrices as similarity matrices.

The method independently invented by Levin, Tucker, Jaffrennou, and Gower will be called SUMPCA for cross-product matrices, or simply SUMPCA_C, if it cannot be confused with the method described earlier. It can be described mathematically as minimizing the function

$$\text{SUMPCA}_C(B, \Lambda) = \left\| \sum_{k=1}^{p} S_k - B\Lambda B' \right\|^2$$

$$= p \sum_{k=1}^{p} \|S_k - B(p^{-1}\Lambda)B'\|^2 + c, \tag{20}$$

over $B(m \times r)$, subject to $B'B = I_r$, and over the diagonal matrix Λ (in (20), c does not depend on B and Λ). Clearly, SUMPCA for cross-product matrices is SUMPCA applied to the matrices $S_1, \ldots, S_p$ instead of to $X_1, \ldots, X_p$.

Comparison of (15) and (20) shows that STATIS-3 is a weighted variant of SUMPCA_C. When all weights $\alpha_1, \ldots, \alpha_p$ in STATIS-3 are (taken) equal, SUMPCA_C and STATIS-3 coincide. The weighting by α_k in STATIS-3 can also be considered a form of data preprocessing, for example, by applying this weighting to X_k (then using the square root of α_k). Viewed in this way, the methods are equivalent but imply a different preprocessing.

SUMPCA for cross-product matrices can be described as a constrained variant of STATIS-1 in a similar, somewhat forced, way as SUMPCA has been described as a constrained variant of PCA-SUP. SUMPCA for cross-product matrices can be described as minimizing

$$\text{SUMPCA}_C^*(B, \Lambda^*) = \sum_{k=1}^{p} \|S_k - B\Lambda^* B'\|^2, \tag{21}$$

over B and Λ^*, where $\Lambda^* = p^{-1}\Lambda$, and subject to the constraint that B is columnwise orthonormal. Defining $c_{kl} \equiv \lambda_l^*$, we have analogously to (12),

$$\text{SUMPCA}_C^*(B, \Lambda^*) = \|S - (\text{Vec}\,(\mathbf{b}_1\mathbf{b}_1') | \ldots | \text{Vec}\,(\mathbf{b}_r\mathbf{b}_r'))C'\|^2. \tag{22}$$

Obviously, SUMPCA_C is a constrained variant of PCA, in that it minimizes (14) over F subject to the constraint that F can be written as $F = (\text{Vec}\,(\mathbf{b}_1\mathbf{b}_1') | \ldots | \text{Vec}\,(\mathbf{b}_r\mathbf{b}_r'))$, for a certain columnwise orthonormal matrix B, like INDORT, and over C subject to the constraint that all rows of C are equal. Clearly, SUMPCA_C is a constrained variant of STATIS-1 that is even more heavily constrained than INDORT; when in INDORT C is constrained to be nonnegative, SUMPCA_C is a constrained version of INDORT only if in SUMPCA_C the elements of C are nonnegative also, as is always the case when $S_1, \ldots, S_p$ are positive semi-definite. Again, the additional constraint corresponds to a simplification of the interpretation of the analysis of the matrices $S_1, \ldots, S_p$. The results imply that all occasions have the same loadings (on every component), and the components do not differentially weight occasions. Therefore, this analysis only provides an adequate description of the relations between the occasions when there are minimal changes in score patterns over occasions.

Analyse Factorielle Multiple

Escofier and Pagès (1983, 1984) developed analyse factorielle multiple (AFM) for the simultaneous analysis of a number of data sets with the same objects and different variables as an alternative to generalized canonical analysis (Carroll, 1968), among others. However, AFM can just as well be used when there is only one set of variables and different sets of objects (Escofier, personal communication, September 30, 1987). AFM is treated only in the latter case even though this may not be the most familiar description. It should be noted, however, that objects and variables can be interchanged everywhere.

AFM consists of two steps. In the first, the data sets $X_1, \ldots, X_p$ are normalized such that all their first principal components explain the same amount of inertia. This reduces to using $\beta_k S_k$ instead of S_k, where β_k is the inverse of the largest eigenvalue of S_k. The second step consists of a (two-way) PCA on the total of all sets of objects, considered as one set of objects with scores on the set of variables. This is equivalent to SUMPCA_C when the S_k matrices are replaced by $\beta_k S_k$. Thus, like STATIS-3, AFM can be seen as a weighted variant of SUMPCA_C. Again, the particular choice of weights (β_k) is a form of data preprocessing, and is a strategy that might be useful in other three-way methods as well.

Hierarchical Relations Between Three-Way Methods

Two sets of three-way methods have been described. The first consists of methods for directly fitting the three-way data; the second set parallels the first but is meant for fitting derived data. For both sets it has been shown that all methods are constrained versions of a PCA over occasions, and are treated in such an order that each method is a constrained version of its predecessor. Thus, for the direct fitting methods, we have the following hierarchy: SUMPCA is a constrained version of ORTCP; ORTCP is a constrained version of ORTCP-A and ORTCP-B; ORTCP-A and ORTCP-B are constrained versions of CANDECOMP/PARAFAC; CANDECOMP/PARAFAC is a constrained version of TUCKALS-3 when $r_1 \geq r$ and $r_2 \geq r$; and TUCKALS-3 is a constrained version of PCA-SUP.

In Table 1 an overview is given of the direct fitting methods that form a hierarchy as described. For all of these methods, there is an indication of which coordinate and parameter sets form the solution. Analogously, for the derived data fitting methods we have: SUMPCA for cross-products is a constrained version of INDORT; INDORT is a constrained version of INDSCAL; INDSCAL is a constrained version of three-mode scaling provided $r' \geq r$, and three-mode scaling is a constrained version of STATIS-1. Because STATIS-3 and AFM can be seen as SUMPCA_C applied to weighted matrices $S_1, \ldots, S_p$, they are not mentioned separately in the hierarchy given in Table 2. The weights in STATIS-3 and AFM might be incorporated in the matrices $S_1, \ldots, S_p$ themselves, so that the weighting is considered part of the preprocessing of the three-

TABLE 1

Hierarchy of Direct Fitting Methods

method	occasion loadings	object/variable combinations	object coordinates	variable coordinates
PCA-SUP	C	F	-	-
TUCKALS-3	C	$(A \otimes B)\bar{G}$	A	B
CANDECOMP/PARAFAC	C	$(\mathbf{a}_1\mathbf{b}_1'..\mathbf{a}_r\mathbf{b}_r')$	A	B
ORTCP-A, ORTCP-B	C	$(\mathbf{a}_1\mathbf{b}_1'..\mathbf{a}_r\mathbf{b}_r')$	A(*col-orth?*)	B(*col-orth?*)
ORTCP	C	$(\mathbf{a}_1\mathbf{b}_1'..\mathbf{a}_r\mathbf{b}_r')$	A(*col-orth*)	B(*col-orth*)
SUMPCA	$\mathbf{1c}'$	$(\mathbf{a}_1\mathbf{b}_1'..\mathbf{a}_r\mathbf{b}_r')$	A(*col-orth*)	B(*col-orth*)

way data. This preprocessing might thus be used in all other methods in the hierarchy as well. It should be noted, however, that using the STATIS-3 weights for the matrices $S_1, \ldots, S_p$, and subsequently performing a variant of PCA on these matrices strung out as vectors, reduces to a kind of double PCA, the usefulness of which is not immediately clear. The AFM weights, on the other hand, do seem useful in all variants of PCA.

TABLE 2

Hierarchy of Derived Data Fitting Methods

method	occasion loadings	variable/variable combinations	variable coordinates
STATIS-1	C	F	-
three-mode scaling	C	$(B \otimes B)\bar{G}$	B
INDSCAL	C	$(\mathbf{b}_1\mathbf{b}_1'..\mathbf{b}_r\mathbf{b}_r')$	B
INDORT	C	$(\mathbf{b}_1\mathbf{b}_1'..\mathbf{b}_r\mathbf{b}_r')$	B(*col-orth*)
SUMPCA for cross-product matrices	$\mathbf{1c}'$	$(\mathbf{b}_1\mathbf{b}_1'..\mathbf{b}_r\mathbf{b}_r')$	B(*col-orth*)

Suggestions for an Eclectic Approach to Three-Way Analysis

It has been shown that a number of well-known three-way methods can be ordered in a hierarchy. The higher the position a method takes, the better the fit of the model for PCA over occasions. Simultaneously, the higher the position, the more parameters are involved, and hence, the more complex the model is (i.e., in terms of interpretation, not necessarily computationally). The latter statement may not be obvious for the methods that take the highest positions in the hierarchies. It may seem that PCA-SUP and STATIS-1 in fact use less parameters than TUCKALS-3 and three-mode scaling, respectively. However, recalling that PCA-SUP can be seen as a version of TUCKALS-3 with $r_1 = n$ and $r_2 = m$, it is clear that PCA-SUP fits a model with more parameters than TUCKALS-3 with $r_1 < n$ and $r_2 < m$. Analogously, STATIS-1 fits a more complex model than three-mode scaling. A similar description for the hierarchy for the derived data fitting methods could be given. In the sequel, only methods for direct fitting will be treated while the same reasoning could be made for methods for fitting derived data.

To perform PCA over occasions, one might choose from all of the methods above, and usually it is not at all clear which of these methods should be selected to meet the representational detail demanded by the researcher. No general statement as to which method is the *best* can be made, and therefore, a researcher has no tool to determine which method would be most useful for the analysis of a given data set. However, the hierarchies described above might be used to find empirically which method is the most *useful* for data description by means of a PCA over occasions. Obviously, the *best* PCA over occasions is provided by PCA-SUP. For the purpose of solution interpretation, however, this method is a bit poor because it yields a rather complicated (and not at all parsimonious) representation for the objects and variables. Therefore, the following strategy is proposed:

We start by determining the number of principal components needed to describe the relations between occasions. Therefore, a PCA-SUP is performed, and the smallest dimensionality that yields a useful description is determined in the usual way. Note that the dimensionality may not be completely clear, and several dimensionalities might have to be tried. This analysis also provides the fit value for the best approximation of the matrices $X_1, \ldots, X_p$ for the chosen number of dimensions. The latter helps to establish how well a constrained variant of PCA-SUP approximates the best possible fit for the present dimensionality.

Having thus chosen the dimensionality of the solution, the data are then fitted by methods that have a simpler model and interpretation. We start with the simplest (and most restricted) model, SUMPCA, because if SUMPCA yields an adequate fit of the matrices $X_1, \ldots, X_p$, this model would be the most favored. Because in SUMPCA all occasions get the same loadings on each dimension, one would then conclude that the data are quite stable over occasions. When the model does not fit adequately, the next method in the hierarchy may be applied, ORTCP. It will always yield a fit that is at least as good as the fit by means of SUMPCA, but the representation of the data is a little less simple because it provides *different* loadings for the occasions on each dimension. The A and B matrices are required to be orthonormal. If this model does not fit well, its restrictions may be weakened by dropping one or both of the orthogonality restrictions (and thus introducing a little more complexity into the model). The resulting methods are ORTCP-A, ORTCP-B, or CANDECOMP/PARAFAC.

When CANDECOMP/PARAFAC still does not adequately fit the matrices $X_1, \ldots, X_p$, there is only one more method, TUCKALS-3, available that might provide a solution consisting of coordinates for objects and variables that is linked to the principal components for occasions, albeit in a more complex way. When the TUCKALS-3 fit is not adequate either, one should either be satisfied with the global description provided by PCA-SUP with combined representations for objects and variables linked to each of the occasion components, or choose different dimensionality for the occasion components.

The sketch for how a researcher might empirically determine which method yields the most useful description of one's three-way data favors no method in general, and depending on the data at hand, a choice must be made from a set of methods. Therefore, this strategy may be referred to as "eclectic".

Example of Direct Fitting

To illustrate the hierarchical relations between the direct fitting methods, and the eclectic approach described above, an empirical data set was analyzed by all the methods presented for direct fitting. Because the methods have amply been illustrated in the literature, only very little attention is paid to the interpretation of the solutions. The data, taken from Doledec and Chessel (1987), see also Sabatier (1987, p. 160), consist of a number of measurements of pollution at several locations of the river Meaudret in France. At six stations along the river (1 is most down-river; 6 most up-river), measurements were carried out at four occasions (February, June, August, and November) on ten biochemical variables: temperature in °C, flux: liters of water passing the station per second; acidity (PH); conductivity in μOhm/s (Cond.); dissolved oxygen in % (O_2); biochemical oxygen demand in mg/l O_2 (BOD); chemical oxygen demand in mg/l O_2 (COD); quantity of NH_4 in mg/l (NH_4); quantity of NO_3 in mg/l (NO_3); and quantity of PO_4 in mg/l (PO_4).

Before analyzing the data, the variables were centered and normalized across stations and occasions to a sum-of-squares of one to maintain differences in means over occasions. This option has been advocated by Kroonenberg (1989). The data have been analyzed earlier by Thioulouse and Chessel (1987) using analyse triadique, and by Kroonenberg (1989) using TUCKALS-3, but in both analyses one variable and one station were deleted. The complete data set is analyzed here by the methods of PCA-SUP, TUCKALS-3, CANDECOMP/PARAFAC, ORTCP-A, ORTCP-B, ORTCP, and SUMPCA, for three different dimensionalities (r = 1, 2, and 3, respectively). In TUCKALS-3, r_1 and r_2 were set equal to r. The amount of inertia explained was computed, defined by the total sum-of-squares of the model estimates. It is readily verified that for each of the methods, the amount of explained inertia equals the total sum-of-squares of the data (which is 10) minus the value of the loss function at its minimum. Table 3 provides the explained inertia for each of the methods for each of the dimensionalities. Note that when $r = 1$, the explained inertia is equal for the methods TUCKALS-3 through ORTCP because then these methods are equivalent. Table 3 clearly illustrates the hierarchical relations among the direct fitting methods. Note that no such relations hold among ORTCP-A and ORTCP-B; these methods have been ordered on the basis of the observed function values.

Applying the eclectic approach sketched above to the present data set, one might arrive at the following line of conclusions. First, the dimensionality is determined. On the basis of the amounts of explained inertia of the PCA-SUP solution alone, one would tend to choose the two- or three-dimensional solution. First, the two-dimensional solutions are compared. At the bottom of the hierarchy, SUMPCA explained only 39.5% of the inertia, which by no means approximates that of PCA-SUP. Hence, it seems worthwhile to ascend the hierarchy and see whether a considerably better fit might be attained by one of the other methods. There is a large increase in explained inertia while

TABLE 3

Explained Inertia of Meaudret Data for $r = 1$, 2, and 3

method	explained inertia (in percentages)		
	$r = 1$	$r = 2$	$r = 3$
PCA–SUP	57.4	75.3	89.9
TUCKALS–3	44.4	62.5	77.6
CANDECOMP/PARAFAC	44.4	62.5	73.5
ORTCP–A	44.4	61.8	69.1
ORTCP–B	44.4	55.5	67.1
ORTCP	44.4	55.0	63.0
SUMPCA	33.4	39.5	43.5

ascending from SUMPCA to ORTCP, hardly no increase when going from ORTCP to ORTCP-B, but ORTCP-A is considerably better again. The subsequent methods, CANDECOMP/PARAFAC and TUCKALS-3, hardly improve over this solution which explains almost 62% of the inertia. Therefore, the ORTCP-A solution seems the best compromise between parsimony and fit for $r = 2$. This solution will be described in a little more detail.

Table 4 gives the (normalized) coordinates for the variables and stations, and the loadings for the occasions, on the first dimension only. These coordinates are ordered with respect to their values to facilitate interpretation. Clearly, the first dimension contrasts variables measuring "health" of the water versus pollutedness as expressed by the variables conductivity, PO_4, NH_4, BOD, and COD. From the coordinates of the stations on this dimension, it is clear that the ordering of the stations along the river is almost reproduced exactly: the stations are ordered from least polluted (Station 6) to most polluted (Station 2), where only Station 1 is located incorrectly in this ordering. This can be explained by considering that the more down-river a station is situated, the more pollution one can expect to be collected. Station 1 (the most down-river) takes an unexpected position in this order (i.e., station 1 seems to be very little polluted, although it is the most down-river). An explanation might be based on the peculiar position this station has. It is situated just behind a point where two tributaries have joined the mainstream. The first axis shows that all through the year there is a gradual increase in pollutedness of the river from February until November. During the winter months pollutedness seems to decrease rapidly.

The second dimension, for which the results are given in Table 5, is much more difficult to interpret. It seems to focus on details that would require a deeper insight in the material under study. It has been verified, however, that this second dimension does not merely reflect peculiarities due to the orthogonality constraint on mode A (the

TABLE 4

Coordinates for Variables and Stations, and Loadings for Occasions for the first ORTCP–A Dimension

variables		stations		occasions	
O_2	–.34	no.6	–.31	February	.32
PH	–.28	no.5	–.31	June	.63
NO_3	–.20	no.1	–.24	August	.97
Flux	–.07	no.4	–.22	November	1.63
Temp.	.00	no.3	.00		
Cond.	.32	no.2	.83		
PO_4	.33				
NH_4	.41				
BOD	.42				
COD	.46				

stations) because the second CANDECOMP/PARAFAC dimension yields practically the same results. On the other hand, the interpretational difficulties of the second dimension might also indicate that a two-dimensional solution is in fact too limited and that the second dimension captures several aspects that had better be described by more dimensions. In that case, a three-dimensional solution is called for. Upon inspection of the amounts of explained inertia for $r = 3$ as reported in Table 3, the TUCKALS-3 solution seems most useful because it explains considerably more of the inertia than does CANDECOMP/PARAFAC. The interpretation of this solution is beyond the scope of the present paper.

Discussion

Methods for direct and for derived data fitting have been described. Although fitting derived data can be used both on multiple data sets and on three-way data, the use of these methods for ordinary three-way data is not favored. To describe a given three-way data set as well as possible, a useful model should be provided and fitted to the three-way data themselves, not to their cross-product matrices because fitting derived data typically does not fit the original data as well as direct fitting methods do. In this respect, PCA methods for three-way data differ from PCA for two-way data where fitting derived data (of a correlation or covariance matrix) yields the same solution as direct fitting of (standardized or deviation score) data.

Other methods might exist that could be placed in the hierarchies mentioned. Recently, Lundy et al. (1989) have proposed a method PFCORE as an extension of

TABLE 5

Coordinates for Variables and Stations, and Loadings for Occasions for the second ORTCP-A Dimension

variables		stations		occasions	
PH	.42	no.3	.66	February	.32
Flux	.33	no.4	.57	June	.60
O_2	.31	no.2	.35	November	−.55
COD	−.17	no.5	.29	August	−1.15
Temp.	−.18	no.6	.16		
BOD	−.26	no.1	.11		
NO_3	−.33				
NH_4	−.33				
Cond.	−.34				
PO_4	−.40				

CANDECOMP/PARAFAC to be used when this method yields so-called degenerate solutions. Degenerate solutions refer to solutions with very large positive or negative correlations between some columns of the matrices A, B, and/or C. To avoid these degenerate solutions, PFCORE imposes orthogonality constraints on at least one of the modes (assume A or B), and next computes a TUCKALS-3 core matrix that is the least squares approximation of the core for matrices A, B, and C obtained from ORTCP-A (or -B). In this way PFCORE combines the unique axes property of CANDECOMP/PARAFAC with the use of a core matrix as in TUCKALS-3. The fit value for these parameters can be shown to lie between those of TUCKALS-3 and ORTCP-A (or -B). Because it is not clear how the PFCORE loss function value compares to that of CANDECOMP/PARAFAC, it should be located partly outside the hierarchy for direct fitting methods, in branches parallel to the main hierarchy. In our eclectic approach such branches should be followed only in cases where CANDECOMP/PARAFAC turns out to yield a degenerate solution.

Orthogonality constraints are an essential part of the PFCORE method when it is used for explaining and avoiding degenerate solutions. However, application of PFCORE need not be limited to these situations, and as a consequence orthogonality constraints need not be imposed. Thus, one may compute a TUCKALS-3 core based on the A, B, and C matrices from the *unconstrained* CANDECOMP/PARAFAC solution. The loss function for the resulting method lies precisely between those of TUCKALS-3 and CANDECOMP/PARAFAC. Hence, this method can be placed in the hierarchy for direct fitting methods in between TUCKALS-3 and CANDECOMP/

PARAFAC, with an interpretation that is as simple as that of the latter complemented by information on its misfit.

It has been mentioned that the hierarchical relations between TUCKALS-3 and CANDECOMP/PARAFAC only hold when $r_1 \geq r$ and $r_2 \geq r$. It is not a priori clear whether for certain $r_1 < r$ and/or $r_2 < r$, TUCKALS-3 still yields a better fit than CANDECOMP/PARAFAC. This need not affect our eclectic approach, since this is based on first applying CANDECOMP/PARAFAC, and only if this solution is not useful does one determine the TUCKALS-3 solution with $r_1 \geq r$ and $r_2 \geq r$, which never yields a poorer fit. If this solution is useful, one may decrease the values for r_1 and r_2 as long as this does not seriously affect the fit. However, these values of r_1 and r_2 should not be decreased so much that the TUCKALS-3 fit becomes equal or even poorer than that of CANDECOMP/PARAFAC because then the latter has a better fit *and* only r components to interpret while the former has $r + r_1 + r_2$ components and their relations to interpret. Obviously, similar arguments hold for the comparison of three-mode scaling and INDSCAL.

There are other methods for direct fitting that have not been treated here, for example, LONGI (Pontier, Pernin, & Pagès, 1985). LONGI has been designed for a purpose that differs from the methods described here, which perform PCA on occasions considered as variables, and are directed at determining which patterns are common over occasions. LONGI, on the other hand, is directed toward a description that emphasizes the differences between occasions or objects.

Three-way data often consist of repeated measurements at different occasions. The three-way methods presented in no way use the ordering of the occasions in the time. The ordering of the occasions is used as an interpretative feature after the analysis, as is done in the example. The relations between the occasions are explored rather than imposed. This can be done sensibly only if the data are preprocessed so differences over occasions are maintained, as discussed for instance by Bentler (1973), and mentioned by Kroonenberg (1989) in discussing the analysis of the Meaudret data by Thioulouse and Chessel (1987).

Instead of analyzing repeated measurement data by focusing on a PCA of the occasions, a useful alternative might be to focus on a PCA of the variables. For the symmetric methods CANDECOMP/PARAFAC and TUCKALS-3 this shift of focus does not make any difference, but the results of PCA-SUP might be more revealing: Matrix C would give a description of the variables, and F would give a joint representation of the objects and time points, which when plotted would provide trajectories for the individuals displaying (global) trends over time. At the other extreme of the hierarchy, this focus on PCA of variables would change SUMPCA into a probably less interesting method analyzing the means over the variables.

A similar eclectic procedure to choose from a (different) set of methods has been proposed by Lundy et al. (1989). They do not merely suggest that comparing fit values may help one to choose from different methods, but also give a "yardstick against which to compare the increases in fit" (p. 129), based on a comparison with synthetic data arrays. A similar procedure might be useful here as well, but has not yet been tested.

References

Bentler, P. M. (1973). Assessment of developmental factor change at the individual and group level. In J. A. Nesselroade & H. W. Reese (Eds.), *Life-span developmental psychology: Methodological issues* (pp. 145–174). New York: Academic Press.

Carroll, J. D. (1968). Generalization of canonical correlation analysis to three or more sets of variables. *Proceedings of the 76th Convention of the American Psychological Association 3*, 227–228.

Carroll, J. D., & Arabie, P. (1980). Multidimensional scaling. *Annual Review of Psychology, 31*, 607–649.

Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition. *Psychometrika, 35*, 283–319.

Carroll, J. D., & Wish, M. (1974). Models and methods for three-way multidimensional scaling. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.) *Contemporary developments in mathematical psychology, Vol. II: Measurement, psychophysics, and neural information processing* (pp. 57–105). San Francisco: Freeman & Co.

Doledec, S., & Chessel, D. (1987) Rythmes saisonniers et composantes stationnelles en milieu aquatique. I.—Description d'un plan d'observation complet par projection de variables [Seasonal rhythms and stationary components in water environments. I.—Description of a complete observation scheme by projection of variables]. *Acta Oecologia/Oecologia Generalis, 8*, 403–406.

Escofier, B., & Pagès, J. (1983). Méthode pour l'analyse de plusieurs groupes de variables—Application a la caractérisation de vins rouges du Val de Loire [Method for the analysis of several groups of variables—Application to the characterization of red wines from the Loire valley]. *Revue de Statistique Appliquée, 31*, 43–59.

Escofier, B., & Pagès, J. (1984). *L'analyse factorielle multiple* [Multiple factor analysis]. Cahiers du bureau universitaire de recherche opérationnelle, no. 42, Université Pierre et Marie Curie, Paris.

Escoufier, Y. (1973). Le traitement des variables vectorielles [The treatment of vector-valued variables]. *Biometrics, 29*, 751–760.

Glaçon, F. (1981). *Analyse conjointe de plusieurs matrices de données* [Simultaneous analysis of several data matrices]. Unpublished doctoral dissertation, University of Grenoble, France.

Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika, 53*, 325–338.

Harshman, R. A. (1970). Foundations of the PARAFAC procedure: models and conditions for an "explanatory" multi-mode factor analysis. *UCLA Working Papers in Phonetics, 16*, 1–84.

Harshman, R. A. (1972). PARAFAC2: Mathematical and technical notes. *UCLA Working Papers in Phonetics, 22*, 31–44.

Harshman, R. A., & Lundy, M. E. (1984a). Data preprocessing and the extended PARAFAC model. In H. G. Law, C. W. Snyder, J. A. Hattie, & R. P. McDonald (Eds.), *Research methods for multimode data analysis* (pp. 216–284). New York: Praeger.

Harshman, R. A., & Lundy, M. E. (1984b). The PARAFAC model for three-way factor analysis and multidimensional scaling. In H. G. Law, C. W. Synder, J. A. Hattie, & R. P. McDonald (Eds.), *Research methods for multimode data analysis* (pp. 122–215). New York: Praeger.

Jaffrennou, P. A. (1978). *Sur l'analyse des familles finies de variables vectorielles* [On the analysis of finite families of vector-valued variables]. University of Saint-Étienne, France.

Kroonenberg, P. M. (1983). *Three mode principal component analysis: Theory and applications*. Leiden: DSWO press.

Kroonenberg, P. M. (1989). The analysis of multiple tables in factorial ecology: III Three-mode principal component analysis: "Analyse triadique complète". *Acta Oecologia/Oecologia Generalis, 10*, 245–256.

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

Lavit, C. (1985). Application de la méthode STATIS [Application of the STATIS method]. *Statistique et Analyse des Données, 10*, 103–116.

Lavit, C. (1988). *Analyse conjointe de tableaux quantitatifs* [Simultaneous analysis of quantitative data sets]. Paris: Masson.

Lechevallier, F. (1987). *L'Analyse de l'evolution dans STATIS: Une solution et des généralisations* [The analysis of evolution in STATIS: A solution and some generalizations]. Document de travail, UFR Sciences Economiques et Sociales, Villeneuve-d'Ascq, France.

Levin, J. (1966). Simultaneous factor analysis of several gramian matrices. *Psychometrika, 31*, 413–419.

L'Hermier des Plantes, H. (1976). *Structuration des tableaux à trois indices de la statistique* [Structuring statistical three-way data matrices]. Unpublished doctoral dissertation, University of Montpellier, France.

Lundy, M. E., Harshman, R. A., & Kruskal, J. B. (1989). A two-stage procedure incorporating good features of both trilinear and quadrilinear models. In R. Coppi & S. Bolasco (Eds.), *Multiway data analysis* (pp. 123–130). Amsterdam: Elsevier Science Publishers.

Pontier, J., Pernin, M. O., & Pagès, M. (1985, May). *LONGI: Une méthode d'analyse de données longitudinales multivariées* [LONGI: A method for the analysis of longitudinal multivariate data]. Paper presented at the Journées de Statistique, Pau, France.

Sabatier, R. (1987). *Méthodes factorielles en analyse des données: Approximations et prise en compte de variables concomitantes* [Factorial methods in data analysis: Approximations and accounting for instrumental variables]. Unpublished doctoral dissertation, University of Montpellier, France.

ten Berge, J. M. F., & Kiers, H. A. L. (1991). Some clarifications of the CANDECOMP algorithm applied to INDSCAL. *Psychometrika, 56*, 317–326.

Thioulouse, J., & Chessel, D. (1987). Les analyses multitableaus en écologie factorielle. I.—De la typologie d'état à la typologie de fonctionnement par l'analyse triadique [The analysis of multiple tables in factorial ecology. I.—From state typology to function typology by means of triadique analysis]. *Acta Oecologia/Oecologia Generalis, 8*, 463–480.

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

Tucker, L. R. (1972). Relations between multidimensional scaling and three-mode factor analysis. *Psychometrika, 37*, 3–27.

Tucker, L. R., & Messick, S. (1963). An individual differences model for multidimensional scaling. *Psychometrika, 28*, 333–367.

Weesie, J., & Van Houwelingen, H. (1983). *GEPCAM Users' manual: Generalized principal components analysis with missing values*. Institute of Mathematical Statistics, University of Utrecht, The Netherlands.

Manuscript received 6/17/88
Final version received 8/14/90