JOURNAL OF CHEMOMETRICS
J. Chemometrics 2000; **14**: 151–170

Some procedures for displaying results from three-way methods

Henk A. L. Kiers*†

Heymans Institute (PA), University of Groningen, Grote Kruisstraat 2/1, NL-9712 TS Groningen, The Netherlands

SUMMARY

Three-way Tucker analysis and CANDECOMP/PARAFAC are popular methods for the analysis of three-way data (data pertaining to three sets of entities). To interpret the results from these methods, one can, in addition to inspecting the component matrices and the core array, inspect visual representations of the outcomes. In this paper, first an overview is given of plotting procedures currently in use with three-way methods. Not all of these optimally correspond to the actual approximation of the data furnished by the three-way method at hand. Next it is described how plotting procedures can be designed that do correspond exactly to the low-dimensional description of the data by means of the three-way method at hand, and it is indicated to what extent these correspond to the ones currently in use. Specifically, procedures are described for displaying either one set of entities (e.g. a set of chemical samples) in two- or three-dimensional plots, or a set of combinations of entities (e.g. pertaining to each object at each time point, thus providing 'trajectories' for each object). Furthermore, it is shown how, in these plots, the other entities can be plotted simultaneously (e.g. superimposing the variables on a plot with trajectories for objects). Both procedures are summarized in an appendix. Copyright © 2000 John Wiley & Sons, Ltd.

KEY WORDS: loading plot; biplot; principal component analysis; Tucker analysis; CANDECOMP/PARAFAC

1. INTRODUCTION

Three-way data pertain to measurements related to three entities (modes); for instance, measurements of a number of objects, on a number of variables at several different occasions (which may refer to different points in time or, more generally, different measurement conditions). For the exploratory analysis of three-way data, two methods are particularly suitable. These are Tucker's [1,2] three-mode factor analysis, here called three-way Tucker analysis, and CANDECOMP/PARAFAC [3,4]. Both methods are three-way generalizations of principal component analysis (PCA). Like PCA, both CANDECOMP/PARAFAC and three-way Tucker analysis yield component matrices for the objects and for the variables, but, in contrast to PCA, they also yield a component matrix for the occasions. Also like PCA, both methods yield a low-dimensional representation of the three-way data. For PCA it is customary to display such low-dimensional configurations in the form of a plot of the variables and/or the objects. To do so, it is tempting to simply use the loadings of the variables as co-ordinates to plot the variables as points in a space spanned by the components (drawn as a set of Cartesian axes), and likewise to use the object component scores as co-ordinates for plotting the objects as points in a

* Correspondence to: H. A. L. Kiers, Heymans Institute (PA), University of Groningen, Grote Kruisstraat 2/1, NL-9712 TS Groningen, The Netherlands.
† E-mail: h.a.l.kiers@ppsw.rug.nl

space spanned by the components. However, as will be explained in Section 4, simply using loadings or component scores as co-ordinates may lead to misleading plots, providing a distorted representation of the low-dimensional configuration found by PCA. In such distorted plots the distances between entities do not correspond to their actual similarities according to the modeled data, hence similar objects may inadvertently be plotted farther apart than less similar objects. To give an adequate representation of the low-dimensional PCA configuration, it is necessary to carefully choose the axes with respect to which the variables or objects are to be displayed, and to compute co-ordinates with respect to these axes (see Section 4).

An interesting alternative method for displaying PCA results is offered by the so-called biplot [5–7] in which objects and variables are displayed jointly. This is done in such a way that, for each object and each variable, the inner product between the vectors pointing to an object and a variable approximates the score of this object on the variable. Such plots do not represent the low-dimensional configuration of the objects or of the variables, but focus on reproducing the scores of the objects on the variables.

In three-way generalizations of PCA, both types of plots can, in principle, be made, and in fact, some such procedures have been proposed by Kroonenberg [8,9] for use with three-way Tucker analysis. However, some of the proposed procedures lead to misleading plots. Furthermore, not all possibilities have been explored systematically. In the present paper it will be explained how one can obtain adequate representations of low-dimensional configurations that correspond to the model actually fitted to the data and hence do not give misleading displays. The requirements for such adequate representations are very similar to those that hold for plots of (two-way) PCA solutions. Therefore we will first establish these requirements for PCA, and next, for three-way methods, propose general procedures for adequately plotting configurations of one set of entities, as well as configurations for combinations of entities from two different modes. These procedures will be summarized in an appendix. The present paper will start, however, with a brief introduction to CANDECOMP/PARAFAC and three-way Tucker analysis, and the current techniques for displaying their results.

2. CANDECOMP/PARAFAC

CANDECOMP was proposed by Carroll and Chang [3] as an **N**-way generalization of singular value decomposition (and therefore of PCA) and dubbed 'CANonical DECOMPosition'. PARAFAC, for PARAlel FACtor analysis, was proposed by Harshman [4] as a method that solves the rotational indeterminacy problem of PCA and factor analysis (see Reference [10], pp. 123 and 147–169) by simply using scores of the same objects on the same variables on more than one occasion. Despite their different origins, CANDECOMP and PARAFAC (henceforth denoted as CP) are fully equivalent in that they employ the same model and the same least squares fitting approach. When the elements of an $I \times J \times K$ three-way array $\underline{\mathbf{X}}$ are given as x_{ijk}, $i = 1,\ldots,I$, $j = 1,\ldots,J$ and $k = 1,\ldots,K$, then the model can be described as

$$x_{ijk} = \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr} + e_{ijk} \tag{1}$$

where a_{ir}, b_{jr} and c_{kr} denote elements of the component matrices **A, B** and **C** of orders $I \times R$, $J \times R$ and $K \times R$ respectively and e_{ijk} denotes an error term for element x_{ijk}. The model is fitted to a data set by minimizing the sum of squared error terms, $\sum_{ijk} e_{ijk}^2$, over **A, B** and **C** by means of an alternating least squares algorithm [3,4]. The model generalizes the two-way PCA model, as can be seen immediately upon writing this as

$$x_{ij} = \sum_{r=1}^{R} a_{ir} b_{jr} + e_{ij} \tag{2}$$

where a_{ir} and b_{jr} refer to component scores and component loadings respectively. Clearly, in the case where the three-way array has only one slice (i.e. $K = 1$), the CP model reduces to $x_{ij} = \sum_r a_{ir}(b_{jr}c_{lr}) + e_{ij}$, which is of the same form as (2). Thus PCA can be seen as a special case of CP.

The PCA model is better known in its matrix notation

$$\mathbf{X} = \mathbf{A}\mathbf{B}' + \mathbf{E} \tag{3}$$

and we will likewise describe the CP model in matrix notation. For this purpose we denote by $\mathbf{X}_a$ the $I \times JK$ mode A matricized [11] form of $\underline{\mathbf{X}}$ which has the frontal slices of $\underline{\mathbf{X}}$ next to each other. Similarly, we represent the three-way array of error terms, $\underline{\mathbf{E}}$, by its matricized version $\mathbf{E}_a$. Furthermore, we introduce the $R \times R \times R$ unit superdiagonal array $\underline{\mathbf{I}}$ which has unit elements in the positions (i,i,i), $i = 1,\ldots,R$, and zeros elsewhere. The $R \times R^2$ matricized version is denoted as $\mathbf{I}_a$. In this notation we can write the CP model as

$$\mathbf{X}_a = \mathbf{A}\mathbf{I}_a(\mathbf{C}' \otimes \mathbf{B}') + \mathbf{E}_a \tag{4}$$

where $\otimes$ denotes the Kronecker product. From (4) the similarity of the CP model to the PCA model (3) becomes clear. In fact, now it can be seen at once that CP is a constrained version of PCA applied to $\mathbf{X}_a$: CP is PCA of $\mathbf{X}_a$ subject to the constraints that the PCA loading matrix for the variable/occasion combinations can be written as $(\mathbf{C} \otimes \mathbf{B})\mathbf{I}'_a$ (see also Reference [12]).

In the literature, results from applications of CP to practical data sets are usually given in the form of tables or one-dimensional plots [13], the rationale being that the CP dimensions, which are determined uniquely [10,14], can be inspected one by one, and this should indeed be done in order to correctly relate the component matrices for the three different modes. In CP applications it does not seem customary to plot component scores on different dimensions against each other [15–17], and when this is done [18], it is done without reference to approximation of the higher-dimensional configuration data space: component scores are just plotted against each other, without justifying the choice of (orthogonally drawn) axes with respect to which the entities at hand are plotted. Such plots may be rather misleading, as they incorrectly suggest that the distances between entities are meaningful. In the present paper, procedures for adequately plotting configurations representing CP solutions will be described.

3. THREE-WAY TUCKER ANALYSIS

Three-way Tucker analysis [1,2] is a three-way generalization of PCA which can be seen as an extended version of CP. As in CP, in three-way Tucker analysis, component matrices are employed for each mode, but in contrast to what is the case in CP, different numbers of components can be used in the different modes. Moreover, whereas in CP each component in mode A is related to precisely one component in mode B and one in mode C, in three-way Tucker analysis each component is related to every component of both other modes: the relations between the P components of mode A, the Q components of mode B and the R components of mode C are captured by the $P \times Q \times R$ so-called 'core' array, here denoted as $\underline{\mathbf{G}}$. Specifically, the three-way Tucker model can be written as

$$x_{ijk} = \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} a_{ip} b_{jq} c_{kr} g_{pqr} + e_{ijk} \tag{5}$$

or, in matrix notation, as

$$\mathbf{X}_a = \mathbf{A}\mathbf{G}_a(\mathbf{C}' \otimes \mathbf{B}') + \mathbf{E}_a \tag{6}$$

where **A**, **B** and **C** (with elements a_{ip}, b_{jq} and c_{kr} respectively) are component matrices of orders $I \times P$, $J \times Q$ and $K \times R$ respectively, $\underline{\mathbf{G}}$ is the $P \times Q \times R$ core array (the $P \times QR$ mode A matricized version of which is $\mathbf{G}_a$) with elements g_{pqr}, and $\mathbf{E}_a$ is the matricized version of the array with error terms. Three-way Tucker analysis consists of fitting model (5) to a data array by minimizing the sum of squared error terms over **A, B, C** and $\underline{\mathbf{G}}$; for instance, by means of an alternating least squares algorithm [2].

In contrast to what is the case with the CP model, a solution obtained by three-way Tucker analysis is by no means uniquely determined. In fact, in three-way Tucker analysis the rotational indeterminacy is even more extensive than in two-mode PCA: arbitrary non-singular transformations of all component matrices simultaneously do not affect the model representation, provided that these transformations are compensated by the inverse transformations applied to the core [1]. This follows at once from the fact that $\hat{\mathbf{X}}_a = \mathbf{A}\mathbf{G}_a(\mathbf{C}' \otimes \mathbf{B}') = \tilde{\mathbf{A}}\ \tilde{\mathbf{G}}_a(\tilde{\mathbf{C}}' \otimes \tilde{\mathbf{B}}')$ for $\tilde{\mathbf{A}} = \mathbf{AS}$, $\tilde{\mathbf{B}} = \mathbf{BT}$, $\tilde{\mathbf{C}} = \mathbf{CU}$ and $\tilde{\mathbf{G}}_a = \mathbf{S}^{-1}\mathbf{G}_a((\mathbf{U}')^{-1} \otimes (\mathbf{T}')^{-1})$ for any set of non-singular matrices **S**, **T** and **U**. Therefore, in three-way Tucker analysis, two- or higher-dimensional plots are particularly useful, as they, to some extent, obviate the need for actually carrying out transformations/rotations of the component matrices. Indeed, several procedures for plotting results from a three-way Tucker analysis have been proposed, in particular by Kroonenberg (see Reference [8], Chap. 6). This is not to say that, in three-way Tucker analysis, dimension-by-dimension interpretation is uncommon. On the contrary, the interpretation of three-way Tucker analysis solutions will almost always be based on dimensionwise interpretation of components for each of the three modes, followed by an assessment of the strength of the interaction of such components (as indicated by the core). However, as Kroonenberg has demonstrated frequently, interpretation of components as well as of relations between entities from different modes is considerably facilitated by various plotting procedures.

Kroonenberg (See Reference [8], pp. 154–157) first discussed interpretation and plotting of the values in the component matrices. For plotting the entities of a particular mode, without defining the Cartesian axes to be used, he mentions two possibilities: (1) use as co-ordinates the rows of the columnwise orthonormal component matrix for the mode at hand; (2) first scale the columns of the component matrix such that their sums of squares equal their 'weights', which for mode A denote the eigenvalues of $\mathbf{G}_a\mathbf{G}'_a$ and for modes B and C denote the eigenvalues of analogously defined matrices, and use the ensuing values as co-ordinates. In fact, in applications, both procedures are used (see e.g. Reference [8], pp. 208 and 211 or Reference [9], p. 89). Kroonenberg (see Reference [8], p. 155) mentions that 'Adjusting the components in such a way that their lengths are proportionate to their (standardized) weights has certain advantages for plotting components against one another. Especially when the weights associated with the components are very different, directly plotting them without adjustment might give a wrong impression of their relative importance'. It will be shown in Section 5 of the present paper that, indeed, using the unweighted co-ordinates is misleading, and that the 'certain advantages' of the second procedure pertain to the fact that it does give an adequate representation of the low-dimensional data approximation provided by three-way Tucker analysis.

Interesting data features may also be revealed by plotting results for combinations of entities of different modes; for instance, displaying a configuration of all variables at all occasions. Kroonenberg

(see Reference [8], pp. 165–166) proposed to display co-ordinates of such combined entities (denoted by him as 'component scores') and did so in various applications (see e.g. pp. 42 and 220). Here he plotted such component scores, for each dimension separately, against the entities of the two combined modes. The alternative of plotting such combined scores against each other to obtain so-called 'trajectories' is used by Kroonenberg [19] and mentioned in a more general context by him (see Reference [9], p. 85). As will be shown in Section 6 of the present paper, these trajectories can also be viewed as low-dimensional approximations of trajectories in the actual high-dimensional space, provided that proper scalings are used.

A third way of displaying the results from a three-way Tucker analysis is by means of so-called 'joint plots' that relate the entities of one mode to the entities of one other mode, for each of the components in the third mode (see Reference [8], pp. 164–165). Specifically, for the rth C-mode component a joint plot for A- and B-mode entities is given by first decomposing $\mathbf{G}_r$ (the rth frontal slice of $\underline{\mathbf{G}}$) as $\mathbf{G}_r = \mathbf{UDV}'$ (by means of a singular value decomposition); next the first $S = \min(P,Q)$ singular vectors and values are collected in $\mathbf{U}_S$, $\mathbf{D}_S$ and $\mathbf{V}_S$. Then a joint plot is obtained by simply plotting the A-mode entities vectors (or points) with co-ordinates given by the rows of $\tilde{\mathbf{A}} = (I/J)^{1/4}\mathbf{A}\mathbf{U}_S\mathbf{D}_S{}^{1/2}$ and plotting the B-mode entities (as vectors or points in the same plot) by using as co-ordinates the rows of $\tilde{\mathbf{B}} = (J/I)^{1/4}\mathbf{B}\mathbf{V}_S\mathbf{D}_S{}^{1/2}$, both with respect to Cartesian axes. In these plots, no meaning is to be attached to the axes. Instead, the information in the plot is carried by the inner products between the vectors for the A-mode and B-mode entities. These inner products are the elements of $\tilde{\mathbf{A}}\,\tilde{\mathbf{B}}' = \mathbf{A}\mathbf{U}_S\mathbf{D}_S\mathbf{V}'_S\mathbf{B}' = \mathbf{A}\mathbf{G}_r\mathbf{B}'$, which approximates the data representation as accounted for by the rth C-mode component. In the plot, these inner products can be found exactly by projection of each mode A vector on each mode B vector, and multiplying the length of the projection by the length of the mode B vector (or *vice versa*); a quick grasp of these inner products can be obtained at once from the (non-)closeness of groups of vectors, as follows from the fact that, as long as vectors have similar lengths, the inner product between vectors is strongly correlated with the closeness of their end points. That these lengths have similar magnitudes is to some extent ascertained by the use of the scalar factors $(I/J)^{1/4}$ and $(J/I)^{1/4}$; in fact, the rationale behind the use of these factors is that the ensuing biplot yields the smallest possible sum of distances between row and column points of all possible biplots displaying vectors whose inner products give the elements of $\mathbf{A}\mathbf{G}_r\mathbf{B}'$, as proven by Kroonenberg and De Leeuw [20]. Thus, from (non-)closeness of the plotted points, the joint plots yield a global insight into the modeled relations between A- and B-mode entities, given the C-mode entity. Joint plots as proposed by Kroonenberg are very insightful in the case where the C-mode entities are easily interpretable (see e.g. Reference [8], pp. 218–221). However, in the case where the C-mode entities are difficult to interpret, it may be better to display relations between all three modes simultaneously. This can be done by using biplots displaying relations between entities from one mode together with combined entities from the other two modes, as will be described in Section 6.

Having described the currently used procedures for plotting results from three-way methods, we will now describe a general procedure for displaying configurations that adequately correspond to the data approximations offered by three-way methods. If possible, we will relate these to the ones mentioned above. Before doing so, however, we will start with a detailed explanation of such displaying procedures for the case of (two-way) PCA, thus laying the foundation for the procedures to be used for three-way methods.

4. PLOTTING IN TWO-WAY PCA

In Section 4.1 we will describe, in ample detail, a procedure for plotting the column entities (variables). In Section 4.2 an analogous procedure for plotting row entities will be described briefly, and it will be indicated how, in this plot, the variables can be plotted meaningfully as well.

4.1. Plotting the column entities (the variables)

In PCA of an $I \times J$ objects-by-variables matrix $\mathbf{X}$ a plot of the variable loadings gives an adequate representation of the configuration of the variables in $\mathbb{R}^I$, provided that the loadings have componentwise sums of squares equal to the associated eigenvalues. This can be seen as follows. We consider the variables as points in the high-dimensional space $\mathbb{R}^I$, spanned by I orthogonal axes (one for each object), given by the co-ordinates in the columns $\mathbf{x}_j$ ($j = 1,\ldots,J$) of $\mathbf{X}$. Thus for each variable the co-ordinates with respect to these Cartesian axes are simply the scores of the objects on this variable. The Euclidean distances between these points are directly related to differences in the scores on the variables concerned: the squared distance between the points representing variables j and k in $\mathbb{R}^I$ is $\|\mathbf{x}_j - \mathbf{x}_k\|^2 = (\mathbf{x}'_j\mathbf{x}_j + \mathbf{x}'_k\mathbf{x}_k - 2\mathbf{x}'_j\mathbf{x}_k)$, which, in the case where the variables are centered and normalized, equals $2(1 - r_{jk})$, where r_{jk} denotes the correlation between variables j and k.

It has been seen above that the configuration of the variable points in $\mathbb{R}^I$ captures the information on the relations between the variables. Now in PCA the matrix $\mathbf{X}$ is approximated by a matrix $\hat{\mathbf{X}} = \mathbf{AB}'$ for certain matrices $\mathbf{A}$ and $\mathbf{B}$ of orders $I \times R$ and $J \times R$ respectively. In fact, a special property of PCA is that $\hat{\mathbf{X}}$ is an orthogonal projection of $\mathbf{X}$ onto a particular subspace. This follows from the fact that $\hat{\mathbf{X}} = \mathbf{U}_R\mathbf{D}_R\mathbf{V}'_R$, where $\mathbf{U}_R$, $\mathbf{D}_R$ and $\mathbf{V}_R$ are matrices containing the first R singular vectors and values from the SVD $\mathbf{X} = \mathbf{UDV}'$. Hence $\hat{\mathbf{X}} = \mathbf{U}_R\mathbf{D}_R\mathbf{V}'_R = \mathbf{U}_R\mathbf{U}_R'\mathbf{X}$, where $\mathbf{U}_R\mathbf{U}'_R$ is an orthogonal projector, which shows that $\hat{\mathbf{X}}$ contains projections of the columns of $\mathbf{X}$ onto the subspace spanned by $\mathbf{U}_R$. Thus the columns of $\hat{\mathbf{X}}$ still pertain to points in $\mathbb{R}^I$, but these points all lie in an R-dimensional subspace of $\mathbb{R}^I$. Now in PCA, usually either $\mathbf{A}$ or $\mathbf{B}$ is chosen to have unit sums of squares, hence either $\mathbf{A} = \mathbf{U}_R$ and $\mathbf{B} = \mathbf{D}_R\mathbf{V}_R$ or $\mathbf{A} = \mathbf{U}_R\mathbf{D}_R$ and $\mathbf{B} = \mathbf{V}_R$. Here we assume that the former choice (which is common in psychometrics, but not in chemometrics) is made. Furthermore, we let $\mathbf{A}^\perp$ denote an orthonormal complement matrix such that the square matrix $\mathbf{A}_s = (\mathbf{A}|\mathbf{A}^\perp)$ is orthonormal. We may use the columns of the $I \times I$ matrix $\mathbf{A}_s$ as an alternative set of orthogonal normalized basis vectors for $\mathbb{R}^I$, hence as an alternative set of Cartesian axes, and can now express $\hat{\mathbf{X}} = \mathbf{A}_s\mathbf{A}'_s\hat{\mathbf{X}} = \mathbf{A}_s\mathbf{A}'_s\mathbf{AB}'$ with respect to this alternative basis as $\mathbf{A}'_s\mathbf{AB}'$. Thus, with respect to this basis, the co-ordinates of the variable points are given in the columns of $\mathbf{A}'_s\mathbf{AB}' = (\mathbf{A}|\mathbf{A}^\perp)'\mathbf{AB}' = (\mathbf{B}|\mathbf{0})'$, from which it can be seen that the (approximated) variable points all lie in the column space spanned by the first R axes (with co-ordinates in $\mathbf{B}'$), and that the co-ordinates with respect to the remaining axes are all zero. Thus a plot of the variable loadings given in the rows of $\mathbf{B}$ actually displays the configuration of the variable points given by the low-dimensional approximation $\hat{\mathbf{X}}$ of $\mathbf{X}$, with respect to a set of new orthogonal axes spanning this low-dimensional space. A different way of looking at PCA, hence, is that in PCA we rotate the axes spanning $\mathbb{R}^I$, where the rotation is represented by the orthonormal matrix $\mathbf{A}_s$, in such a way that the I-dimensional configuration is optimally captured within the subspace spanned by the first R dimensions. To visualize this, we may consider that the subspace refers, for example, to the plane that lies closest to all variable points in $\mathbb{R}^3$, and a plot of the loadings refers to the orthogonal projection of these points on this plane. An example of such a situation is given in Figure 1. The original points in $\mathbb{R}^3$ are given, as well as the plane to which they are closest. The approximated configuration of these five points, given by projections of the points on this plane, is indicated as well. To optimally visualize the approximated configuration, we rotate the three-dimensional space such that the plane is captured completely in the space of the present page. Since the projected points fall exactly in this plane, the two-dimensional plot adequately displays the approximated configuration.

The above geometrical interpretation of a loading plot as an approximation of the variable configuration in $\mathbb{R}^I$ only holds if the component matrix $\mathbf{A}$ is columnwise orthonormal. If it is not, the above-mentioned matrix $\mathbf{A}_s$ will no longer be orthonormal, and the configuration based on a plot of the loadings in $\mathbf{B}$ with respect to orthogonally drawn axes no longer pertains to a simple rotated version of the actual approximated configuration of the variables. Instead, it displays a distorted

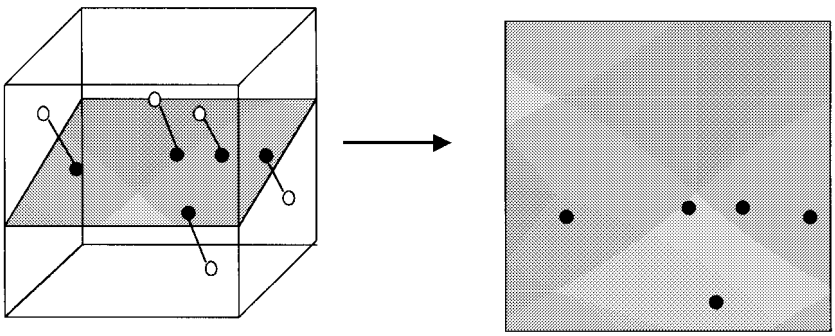

Figure 1. Five points in $\mathbb{R}^3$ (open circles) are projected on the plane to which they are closest (projections given by full circles). The three-dimensional space is rotated such that the plane with the projected points is oriented parallel to the plane of the page; the third dimension, orthogonal to it, need not be visualized, because the projected points lie exactly in the plane.

version of the approximated configuration. To illustrate this, consider the simple example with

$$\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 0{\cdot}5 \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} 1 & 0{\cdot}5 & 0{\cdot}5 & 0 \\ 0 & 0{\cdot}5 & -0{\cdot}5 & 1 \end{pmatrix}'$$

Then

$$\hat{\mathbf{X}} = \mathbf{A}\mathbf{B}' = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 0{\cdot}75 & 0{\cdot}25 & 0{\cdot}5 \end{pmatrix}$$

Plotting the column entities (labeled respectively as A, B, C and D), using as co-ordinates the columns of **B**, gives the left-hand plot in Figure 2. This, however, is not an adequate representation of the configuration based on the estimated data values, given in $\hat{\mathbf{X}}$, as can be seen from the plot of the column entities based on the columns of $\hat{\mathbf{X}}$, which give the co-ordinates of the column entities with respect to the actual Cartesian data axes in $\mathbb{R}^I$, which here is $\mathbb{R}^2$ (middle plot of Figure 2); it should be noted that the present situation is special, because usually the columns of $\hat{\mathbf{X}}$ have more than two elements and hence their location in $\mathbb{R}^I$ cannot be visualized. Both plots suggest that A, B and D are located on a line, but the plots differ completely as far as C is concerned: the left-hand plot suggests

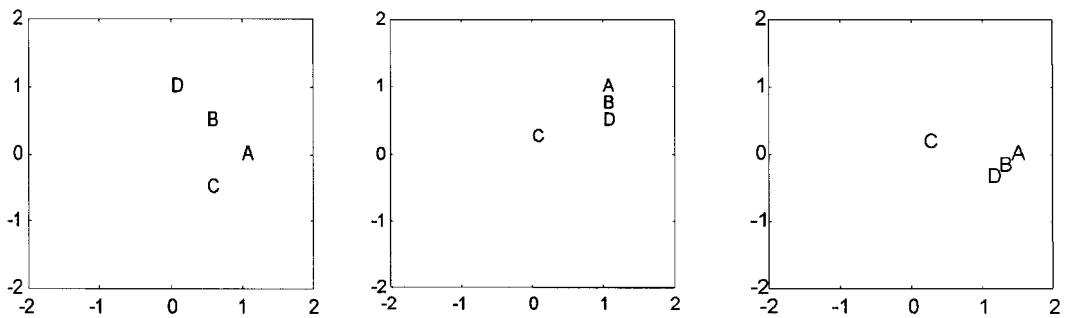

Figure 2. Plots (from left to right) of rows of **B**, columns of $\hat{\mathbf{X}}$ and rows of $\tilde{\mathbf{B}}$.

that C is closer to A than to D, whereas, in the (estimated) data, C is farthest from A and closest to D. Thus the plot based on the co-ordinates in **B** is considerably distorted. This is a consequence of the fact that the axes used when taking the values in **B** as co-ordinates are the two columns in **A**, which are separated by an angle of only 18 ° (and hence far from orthogonal) and are, moreover, of unequal length.

Thus, when the component matrix **A** is not columnwise orthonormal, for giving an undistorted plot of the approximated configuration of the variables with respect to orthogonally drawn axes, we must find an alternative basis matrix for $\hat{\mathbf{X}}$ that *is* columnwise orthonormal. Specifically, if **A** is not columnwise orthonormal, we search for a transformation matrix **T** such that **AT** is columnwise orthonormal (e.g. by Gram–Schmidt orthonormalization), and postmultiply **B** by $(\mathbf{T}')^{-1}$ so that $\tilde{\mathbf{A}}\tilde{\mathbf{B}}'$ (with $\tilde{\mathbf{A}} = \mathbf{AT}$ and $\tilde{\mathbf{B}} = \mathbf{B}(\mathbf{T}')^{-1}$) still equals $\hat{\mathbf{X}}$. Then a plot of the variables with the values in $\tilde{\mathbf{B}}$ as co-ordinates does give an adequate representation of the approximated configuration of the variables. In the above example an orthonormal basis for the columns of $\hat{\mathbf{X}}$ is given by

$$\tilde{\mathbf{A}} = \begin{pmatrix} 0{\cdot}707 & -0{\cdot}707 \\ 0{\cdot}707 & 0{\cdot}707 \end{pmatrix}$$

and

$$\tilde{\mathbf{B}} = \begin{pmatrix} 1{\cdot}41 & 1{\cdot}24 & 0{\cdot}18 & 1{\cdot}06 \\ 0 & -0{\cdot}18 & 0{\cdot}18 & -0{\cdot}35 \end{pmatrix}'$$

gives the associated co-ordinates for the column entities, which are used in the right-hand plot in Figure 2. It can be seen that this configuration is undistorted, being simply a rotation of the actual estimated data configuration, given in the middle plot of Figure 2.

If the low-dimensional approximation uses more than two dimensions, it is hard to display the resulting configuration, and if it uses more than three dimensions, this can only be done by indirect means. In such situations it is possible that there are clusters of points that are far from more than one axis, which hence would not show up in plots where these dimensions are not used simultaneously. To avoid missing such clusters, it is recommended to rotate the loadings to simple structure (e.g. by means of varimax [21]): the aim of simple structure rotations is to rotate such that clusters of points lie close to the axes. If this rotation achieves its goal, as it often does, little insight is lost when the configuration is displayed only with respect to pairs of dimensions.

4.2. Plotting the row entities (the objects) and representing the variables as projected axes

Rather than plotting the approximated configuration of the variables, one may also wish to plot the approximated configuration of the objects, which can themselves be represented by points in $\mathbb{R}^J$. In this situation a plot of the co-ordinates in **A** can be used, provided that **B** is columnwise orthonormal, as follows from the same reasoning as for the variables, with **X** replaced by $\mathbf{X}'$ and **A** by **B** and *vice versa*. Thus, if **B** is not columnwise orthonormal, we find a transformation **T** such that **BT** is columnwise orthonormal (e.g. by Gram–Schmidt orthonormalization), and postmultiply **A** by $(\mathbf{T}')^{-1}$ so that $\tilde{\mathbf{A}}\tilde{\mathbf{B}}'$ (with $\tilde{\mathbf{A}} = \mathbf{A}(\mathbf{T}')^{-1}$ and $\tilde{\mathbf{B}} = \mathbf{BT}$) still equals $\hat{\mathbf{X}}$. Then a plot of the objects with the values in $\tilde{\mathbf{A}}$ as co-ordinates gives an adequate representation of the approximated configuration of the objects.

In practice, a PCA does not give a solution with both **A** and **B** columnwise orthonormal, nor can such a solution be obtained upon transformation. Hence it is impossible to plot objects and variables on the basis of component scores and loadings that are associated with each other. Therefore, with the above-mentioned procedures, the configurations of objects and variables must be considered

independently. An approach to study objects and variables interdependently is by means of the biplot [5–7], as mentioned in Section 1. In fact, given the configuration for the objects, we can plot the variables as vectors in the same plot in such a way that the inner products between (the vectors to) the points for the objects and the vectors for the variables equal the values in $\hat{\mathbf{X}}$. Moreover, the vectors for the variables then represent the projections of the original axes spanning $\mathbb{R}^J$ (each of which pertains to one variable) onto the low-dimension space displayed. Given a configuration for the objects, the axes for the variables are found as follows.

Suppose that $\tilde{\mathbf{B}}$ is a columnwise orthonormal basis matrix for the R-dimensional object configuration. Then, after projection on the subspace, the rows of $\hat{\mathbf{X}} = \mathbf{X}\tilde{\mathbf{B}}\tilde{\mathbf{B}}'$ contain the co-ordinates for each of the objects with respect to the original J axes (which represent the variables), and the rows of the $(I \times R)$ matrix $\mathbf{X}\tilde{\mathbf{B}}$ contain the co-ordinates with respect to the R axes chosen in the R-dimensional subspace. Now, to find the projection of the *original* J axes on this R-dimensional subspace, we consider the J fictitious objects that lie *on* these original axes and have unit length, hence objects that originally have score profiles $\mathbf{e}_1' = (1\,0\,0\ldots0)$, $\mathbf{e}_2' = (0\,1\,0\ldots0)$, …, $\mathbf{e}_J' = (0\,0\,0\ldots1)$, which, respectively, lie on the axes representing the first variable, the second variable, …, the Jth variable. Now, after projection, these objects lie on the *projected* variable axes, hence, by locating the projected location of these fictitious objects and noting that the origin, after projection, remains at the origin, we can assess the orientation of each of the projected axes by simply drawing the vectors associated with the projected variables, $\mathbf{e}_1'\tilde{\mathbf{B}} = \tilde{\mathbf{b}}_1'$, $\mathbf{e}_2'\tilde{\mathbf{B}} = \tilde{\mathbf{b}}_2'$,…, $\mathbf{e}_J'\tilde{\mathbf{B}} = \tilde{\mathbf{b}}_J'$, where $\tilde{\mathbf{b}}_j'$, denotes the jth row of $\tilde{\mathbf{B}}$. The lines through these vectors form the projected variables axes. In the resulting biplot the length of the projection of an object point on a variable axis, multiplied by the length of the variable vector, approximates the actual score of the object on the variable. This is because the length of the projection of object i on axis j is $\tilde{\mathbf{a}}_i'\tilde{\mathbf{b}}_j(\tilde{\mathbf{b}}_j'\tilde{\mathbf{b}}_j)^{-1/2}$, which, multiplied by $(\tilde{\mathbf{b}}_j'\tilde{\mathbf{b}}_j)^{1/2}$, gives $\tilde{\mathbf{a}}_i'\tilde{\mathbf{b}}_j = \mathbf{a}_i'\mathbf{b}_j = \hat{\mathbf{X}}_{ij}$. Thus comparison of scores on the same variable can be done straightforwardly by comparison of projection lengths.

Here we only mentioned projection of the original variable axes onto the low-dimensional space containing the approximate configuration for the objects, because this seems of most practical interest. However, given a configuration for the variables (Section 4.1), we can also project the original 'object axes' onto the low-dimensional space containing the approximate configuration for the variables, by following a fully analogous procedure.

As in PCA, in three-way methods, low-dimensional approximations to the data are obtained. How the above procedures for plotting objects or variables (and, if desired, projected variable or object axes) can be used for the results of three-way methods will be discussed in the following sections.

5. PLOTTING IN THREE-WAY ANALYSIS: CONFIGURATIONS FOR ENTITIES OF ONE MODE

5.1. The three-way Tucker model

In the case of three-way data the entities of each mode can be considered as points in a high-dimensional space, and the three-way Tucker model and the CP model provide low-dimensional approximations to the high-dimensional configurations. Specifically, the A-mode entities are I points in $\mathbb{R}^{JK}$, and when the data are represented by the Tucker model $\hat{\mathbf{X}}_a = \mathbf{A}\mathbf{G}_a(\mathbf{C}'\otimes\mathbf{B}')$ with $\mathbf{A}$ an $I \times P$ matrix, then the rows of $\hat{\mathbf{X}}_a$ give co-ordinates for I points in a P-dimensional subspace of $\mathbb{R}^{JK}$. Hence, to display this approximated configuration, we have to find a basis of this P-dimensional subspace, find the co-ordinates of the A-mode entities with respect to Cartesian axes in this subspace, and plot the points corresponding to these co-ordinates.

To find a basis for the above-mentioned subspace and co-ordinates of the points with respect to Cartesian axes therein, we first rewrite the model expression for $\hat{\mathbf{X}}_a$ as $\hat{\mathbf{X}}_a = \mathbf{A}((\mathbf{C}\otimes\mathbf{B})\mathbf{G}_a')'$ to make it

optimally similar to that for $\hat{\mathbf{X}}$ in Section 4·2, where the matrix $\mathbf{A}$ now has order $I \times P$ and the role of the $J \times R$ matrix $\mathbf{B}$ in Section 4·2 is now played by the $JK \times P$ matrix $\mathbf{F} \equiv (\mathbf{C} \otimes \mathbf{B})\mathbf{G}_a'$. Thus we actually have a situation analogous to that for two-way PCA, and to produce a plot of the A-mode entities, we have to replace $\mathbf{F} = (\mathbf{C} \otimes \mathbf{B})\mathbf{G}_a'$ by an orthonormal basis and adjust the co-ordinates for the A-mode entities accordingly. Hence, as in Section 4·2, we find a transformation $\mathbf{T}$ such that $\mathbf{FT}$ is columnwise orthonormal (e.g. by Gram–Schmidt orthonormalization), and postmultiply $\mathbf{A}$ by $(\mathbf{T}')^{-1}$ so that $\tilde{\mathbf{A}}\tilde{\mathbf{F}}'$ (with $\tilde{\mathbf{A}} = \mathbf{A}(\mathbf{T}')^{-1}$ and $\tilde{\mathbf{F}} = \mathbf{FT}$) still equals $\mathbf{AF}' = \mathbf{A}((\mathbf{C} \otimes \mathbf{B})\mathbf{G}_a')' = \hat{\mathbf{X}}_a$. Then a plot of the objects with the values in $\tilde{\mathbf{A}}$ as co-ordinates gives an adequate representation of the approximated configuration of the objects.

If so desired, the original *JK* axes can be projected into the same plot, but this will give a biplot with, usually, many vectors, representing projected axes for all combinations of B- and C-mode entities (e.g. representing each variable at each different occasion). Such a plot will in many cases be hard to study and is not considered further here.

To obtain plots for the B- and C-mode entities, an analogous procedure can be used, based on the mode B and mode C matricized versions of $\underline{\mathbf{X}}$ and $\underline{\hat{\mathbf{X}}}$. Thus we have $\hat{\mathbf{X}}_b = \mathbf{BG}_b(\mathbf{A}' \otimes \mathbf{C}')$ and $\hat{\mathbf{X}}_c = \mathbf{CG}_c(\mathbf{B}' \otimes \mathbf{A}')$ [11], and co-ordinates for the B- and C-mode entities can be obtained after orthonormalizing $(\mathbf{A} \otimes \mathbf{C})\mathbf{G}_b'$ and $(\mathbf{B} \otimes \mathbf{A})\mathbf{G}_c'$ respectively.

Often the three-way Tucker solution is given in such a way that finding the required orthornormal bases reduces to a simple columnwise scaling. Specifically, for the (unrotated) solution it usually holds that matrix $\mathbf{A}$ contains unit normalized eigenvectors of $\hat{\mathbf{X}}_a\hat{\mathbf{X}}_a'$, $\mathbf{B}$ contains unit normalized eigenvectors of $\hat{\mathbf{X}}_b\hat{\mathbf{X}}_b'$, and $\mathbf{C}$ contains unit normalized eigenvectors of $\hat{\mathbf{X}}_c\hat{\mathbf{X}}_c'$, which is sometimes called the principal axes solution. It follows that $\mathbf{A}'\hat{\mathbf{X}}_a\hat{\mathbf{X}}_a'\mathbf{A} = \mathbf{A}'\mathbf{AG}_a\mathbf{G}_a'\mathbf{A}'\mathbf{A} = \mathbf{G}_a\mathbf{G}_a' = \mathbf{\Lambda}_a$, where $\mathbf{\Lambda}_a$ denotes the diagonal matrix with eigenvalues of $\hat{\mathbf{X}}_a\hat{\mathbf{X}}_a'$. Hence in this case the basis matrix $\mathbf{F} = (\mathbf{C} \otimes \mathbf{B})\mathbf{G}_a'$ is columnwise orthogonal, because $\mathbf{F}'\mathbf{F} = \mathbf{G}_a(\mathbf{C}'\mathbf{C} \otimes \mathbf{B}'\mathbf{B})\mathbf{G}_a' = \mathbf{G}_a\mathbf{G}_a' = \mathbf{\Lambda}_a$, so to orthonormalize $\mathbf{F}$ it suffices to divide the columns by the square roots of the associated eigenvalues; this scaling is to be compensated in $\mathbf{A}$ by multiplying the columns of $\mathbf{A}$ by the square roots of the associated eigenvalues. The thus scaled matrix $\mathbf{A}$ is said to contain principal *co-ordinates*. This scaling is exactly the scaling suggested by Kroonenberg (see Reference [8], p. 155) which 'has certain advantages' when plotting the entities for the mode at hand. The present explanation clarifies what these advantages are: the scaling at hand ensures that the plot corresponds exactly to the projection of the original configuration on the low-dimensional space used in the three-way Tucker representation. This is because, in that case, $\mathbf{F}$ does not only have orthogonal columns, but owing to the scaling by means of the inverse square roots of the eigenvalues, these columns have unit sums of squares as well. Some further comments are in order here. First, of course, the same reasoning can be applied for plotting B- and C-mode entities. Second, it should be noted that, after rotation of a three-way Tucker solution, the component matrices no longer contain eigenvectors, and the above reasoning breaks down, *unless* an orthogonal rotation was applied to *scaled* component matrices: for instance, if the rotated A-mode component matrix is found as an orthogonal rotation of the principal co-ordinates (eigenvectors with sums of squares equal to the associated eigenvalues), then both before and after rotation we have $\mathbf{F}'\mathbf{F} = \mathbf{I}_p$ and hence the basis matrix remains columnwise orthonormal.

5.2. The CP model

The above procedure has been described for the three-way Tucker model, but exactly the same procedure can be used for plotting configurations for A-, B- or C-mode entities based on the approximation given by an *R*-dimensional CP model. To this end, it suffices to replace $\underline{\mathbf{G}}$ in Section 5.1 by $\underline{\mathbf{I}}$. However, such plotting procedures would ignore the fact that the CP model gives (usually oblique) *unique* axes, which may have an intrinsic meaning. Thus one may like to visualize these

unique axes in their actual oblique orientation. For this purpose one might use a procedure for plotting with respect to oblique axes. However, since all standard plotting procedures are based on orthogonal Cartesian axes, we will here, as before, find an (auxiliary) orthogonal basis, and on the resulting plot we will project the oblique CP axes, as follows.

Suppose we have plotted the A-mode entities, with respect to orthonormal axes $\tilde{\mathbf{F}} = \mathbf{FT}$, in which $\mathbf{F} = (\mathbf{C} \otimes \mathbf{B})\mathbf{I}_a' = \mathbf{C} \odot \mathbf{B}$ gives the original oblique axes, and $\odot$ denotes the columnwise Kronecker (or Khatri–Rao) product [11]. To locate the direction of the original axes (columns of $\mathbf{F}$) in the current plot, we simply use that $\mathbf{F} = \tilde{\mathbf{F}}\mathbf{T}^{-1}$. Hence, with respect to the current axes (the columns of $\tilde{\mathbf{F}}$), the locations of the columns of $\mathbf{F}$ are given by the co-ordinates in the columns of $\mathbf{T}^{-1}$. Thus the original axes can be plotted through the vectors given by the rows of $(\mathbf{T}')^{-1}$. It can be verified that the location of the plotted objects can now be found as the linear combination of the vectors representing the oblique axes, using as weights the (original) scores of the objects (in $\mathbf{A}$), which, in fact, in the original CP representation, give the co-ordinates of the objects with respect to the oblique CP axes.

In order to let $\mathbf{A}$ carry all the relevant scale size information, one may normalize both $\mathbf{B}$ and $\mathbf{C}$ to have unit sums of squares columnwise (which can always be arranged in the CP model, without loss of fit, by multiplying the columns of $\mathbf{A}$ by the square roots of the sums of squares of the associated columns in $\mathbf{B}$ and $\mathbf{C}$). Then $\mathbf{F}$ has unit column sums of squares as well. If, furthermore, $\tilde{\mathbf{F}}$ is chosen such that $\tilde{\mathbf{F}}'\tilde{\mathbf{F}} = \mathbf{I}$, then the rows of $(\mathbf{T}')^{-1}$ have unit sums of squares and hence all vectors representing the CP axes are unit-length vectors; this follows from the fact that $\mathbf{F} = \tilde{\mathbf{F}}\mathbf{T}^{-1}$ and Diag $(\mathbf{F}'\mathbf{F}) = \text{Diag}((\mathbf{T}')^{-1}\tilde{\mathbf{F}}'\tilde{\mathbf{F}}\mathbf{T}^{-1}) = \text{Diag}((\mathbf{T}')^{-1}\mathbf{T}^{-1})$, hence the row sums of squares $(\mathbf{T}')^{-1}$ equal the column sums of squares of F, which in this case are all unity.

5.3. Illustration

The procedure for plotting entities of one mode can be useful if one wishes to get an overview of the similarities among (subsets of) such entities. Sometimes such entities have no intrinsic interest. For instance, if one or two of the modes pertain to different (emission/absorption) wavelengths, there is little point in showing similarities between such wavelengths. Hence in such instances one will be mainly interested in the similarities between entities of the other mode, which may pertain to samples consisting of different mixtures. In other examples one may be interested in similarities between all entities of all modes. A case in point is the following.

To illustrate the procedure for plotting the entities of one mode, we use the results of a Tucker analysis of a three-way data set [22,23] consisting of ten measures of pollution (temperature, flux, pH, conductivity, O_2, biochemical O_2 demand, chemical O_2 demand, NH_4, NO_3, PO_4), taken at six different 'stations' along the river Meaudret, at four different occasions (February, June, August and November). Thus we have a $6 \times 10 \times 4$ data set. It was assumed that to model these data by means of a three-way Tucker model, we need additional (unknown) additive terms for the variables ('intercepts'), as described by $\mathbf{X}_a = \mathbf{1}(\mathbf{1} \otimes \boldsymbol{\mu})' + \mathbf{AG}_a(\mathbf{C} \otimes \mathbf{B})' + \mathbf{E}_a$, where $\mathbf{1}$ denotes a vector with unit elements and μ denotes a vector with unknown intercepts for the variables. Rather than estimating the intercepts explicitly, we eliminate these by centering across mode $\mathbf{A}$ and fit the model $\tilde{\mathbf{A}}\mathbf{G}_a(\mathbf{C} \otimes \mathbf{B})'$, where $\tilde{\mathbf{A}}$ represents the centered version of $\mathbf{A}$. To account for differences in scale size of the different variables, we normalize the thus centered data within mode B; that is, for each of the variables, across all scores on it. The fit values for the thus centered and normalized data for several solutions with different dimensionalities are given in Table I. The bold entities pertain to the best-fitting solution within a class of solutions with the same total number of dimensions. We choose the solution for which the highest increase of fit was found upon increasing the sum of dimensions one by one [24], which in this case leads us to the choice $P = 3$, $Q = 3$ and $R = 2$.

A full interpretation of the results would involve inspection of all component matrices and the core,

Table I. Three-way Tucker fit values for several combinations of dimensions

P	*Q*	*R*	Fit (%)
1	2	2	58·0
2	1	2	57·3
2	**2**	**1**	**65·3**
2	**2**	**2**	**68·6**
2	2	3	69·5
2	**3**	**2**	**71·9**
3	2	2	69·8
2	3	3	74·2
3	2	3	71·8
3	**3**	**2**	**76·9**
3	**3**	**3**	**80·5**

and would, *in addition*, make use of displays of some sets of entities. Here we do not consider interpretation of the component matrices and the core, but only illustrate the procedure for displaying one set of entities, namely the stations. In the left-hand panel of Figure 3 the stations are plotted with respect to an orthonormal basis in the three-dimensional subspace used to represent these data. The co-ordinates used in this plot are obtained after first transforming $(\mathbf{C}\otimes\mathbf{B})\mathbf{G}_a'$ to columnwise orthonormality and applying the inverse transformation to $\tilde{\mathbf{A}}$. To see what would happen without this transformation, the right-hand panel of Figure 3 gives a plot based on simply using the rows of $\tilde{\mathbf{A}}$ as co-ordinates. In both cases the centering ensured that the configuration is given with the centroid of the stations as origin.

Stations 1–5 are ordered downwards along the river Meaudret, and station 6 is located in a confluent at, as far as pollution is concerned, a similar location as station 5. The location of station 1 is special in that it is located just before the point where two other confluents join the Meaudret. This causes, as far as pollution is concerned, station 1 to be more similar to the stations further down the

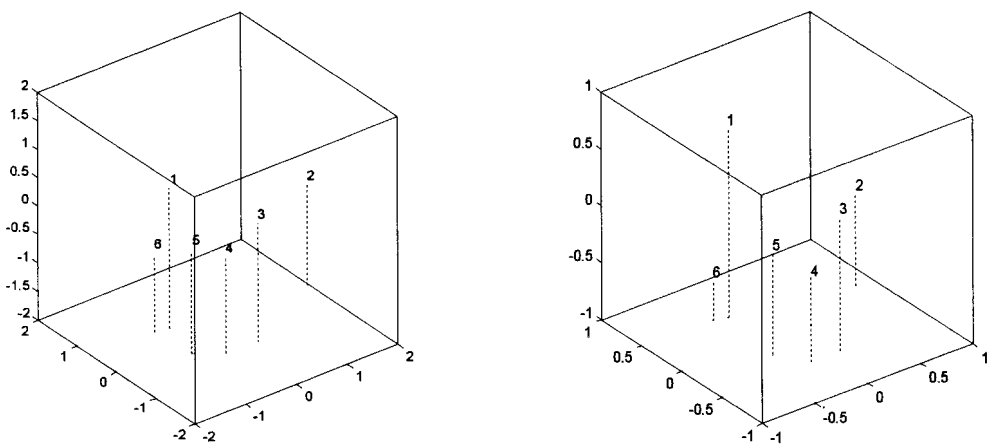

Figure 3. The left-hand plot displays the low-dimensional approximation for the original configuration for the stations; the right-hand plot is based on simply using the rows of the component matrix as co-ordinates for the stations.

Table II. Full data—based distances between stations

	Station 1	Station 2	Station 3	Station 4	Station 5	Station 6
Station 1	—	2·82	1·81	1·73	1·45	1·52
Station 2	2·82	—	2·14	2·64	2·94	2·94
Station 3	1·81	2·14	—	0·99	1·55	2·02
Station 4	1·73	2·64	0·99	—	0·97	1·52
Station 5	1·45	2·94	1·55	0·97	—	1·34
Station 6	1·52	2·94	2·02	1·52	1·34	—

river than the ones topographically closest to it. It can be seen that the configurations differ considerably with respect to the location of this special station 1. In the left-hand plot in Figure 3 it is close to stations 5 and 6, whereas in the right-hand plot it is far off from all other stations. 'In reality', station 1 is not very far from stations 5 and 6, as we verified on the actual data, as follows. We computed Euclidean distances between the stations, as they are located in the original high-dimensional space (based on centered and normalized data in $\underline{\tilde{\mathbf{X}}}$). Specifically, the distance between stations h and i is expressed by

$$d_{hi} = \sqrt{\sum_{j=1}^{J}\sum_{k=1}^{K}\left(\tilde{x}_{hjk} - \tilde{x}_{ijk}\right)^2}$$

These distances are given in Table II. It can be seen that station 1 is much closer to stations 5 and 6 than to station 2. Furthermore, it can be seen that stations 2 and 3 are farther apart than stations 1 and 5, whereas the right-hand plot in Figure 3 clearly displays the reverse. As a final remark on these displays, it can be seen that station 2, which, according to Table II, is clearly most different from the other stations, is indeed an outlier in the left-hand plot, whereas this is by no means as clear in the right-hand plot. Thus the left-hand plot in Figure 3, based on the actual projected configuration, is considerably more realistic than the right-hand plot.

6. PLOTTING IN THREE-WAY ANALYSIS: CONFIGURATIONS FOR COMBINATIONS OF ENTITIES OF TWO MODES (TRAJECTORIES)

In Section 5 we discussed how one can plot the entities of one mode. However, sometimes one is interested in visualizing the location of combinations of entities of two modes. This is especially interesting in cases where we have a sequence of measurements. For instance, suppose we have repeated measurements of a number of objects on a set of variables (e.g. in batch MSPC studies, where a number of batches is measured on a number of variables, at a sequence of time points). Then it can be of interest to see how the location of the objects in the space spanned by the variables changes over time. In other words, we would like to depict the *trajectory* along which the objects 'travel' in the space spanned by the variables, similar to trajectories used in, for instance, STATIS [25]. Since we cannot depict the trajectories in the full-dimensional space, we use the projection of these trajectories on the low-dimensional approximation found by means of the three-way method at hand. For this purpose we basically follow the same procedure as in Section 4·2, as will be explained now.

Suppose we have a three-way data array, organized such that mode A pertains to the objects, mode

B to the variables and mode C to the measurement occasions. Then after CP or three-way Tucker analysis we have $\hat{\mathbf{X}}_a = \mathbf{A}\mathbf{G}_a(\mathbf{C}'\otimes\mathbf{B}')$, but also $\hat{\mathbf{X}}_b = \mathbf{B}\mathbf{G}_b(\mathbf{A}'\otimes\mathbf{C}')$ and $\hat{\mathbf{X}}_c = \mathbf{C}\mathbf{G}_c(\mathbf{B}'\otimes\mathbf{A}')$, as mentioned in Section 5.1. Now, to plot the objects at each time point (i.e. to plot all combinations of A- and C-mode entities), we have to find a representation for the set consisting of the columns of $\hat{\mathbf{X}}_b$, because each column pertains to one object/occasion combination. In $\hat{\mathbf{X}}_b$ the co-ordinates for these columns are given with respect to the full variable space $\mathbb{R}^J$, but they are all located in a low-dimensional subspace, a basis for which is given by **B**. Specifically, the rows of $\mathbf{F} \equiv (\mathbf{A}\otimes\mathbf{C})\mathbf{G}_b'$ give the co-ordinates of each object/occasion combination with respect to the basis spanned by the columns of **B**. In the case of a Tucker analysis the matrix **B** is usually columnwise orthonormal, hence the co-ordinates can be used immediately to plot the object/occasion combinations with respect to Cartesian axes. In other cases (notably in case of CP) we have to transform **B** into an orthonormal basis matrix and apply the inverse transformation to $(\mathbf{A}\otimes\mathbf{C})\mathbf{G}_b'$ to obtain the object/occasion co-ordinates with respect to this basis. Having thus found the location of each object at each occasion, it is often useful to connect points pertaining to the same object at consecutive occasions, so as to find the projected trajectory of each of the objects.

As in the PCA situation, it may be interesting to visualize the location of the original variables after projection on the subspace. The procedure is fully analogous to that described in Section 4.2. Thus, if transformation **T** yields an orthonormal basis matrix $\tilde{\mathbf{B}} = \mathbf{B}\mathbf{T}$, then $\mathbf{F} = (\mathbf{A}\otimes\mathbf{C})\mathbf{G}_b'$ is to be transformed by $(\mathbf{T}')^{-1}$ to find the co-ordinates $\tilde{\mathbf{F}} = \mathbf{F}(\mathbf{T}')^{-1}$ for all object/occasion combinations. Now, in analogy to the derivation in Section 4·2, the rows of $\tilde{\mathbf{B}}$ give the vectors describing the projected variables. Furthermore, we can superimpose the original CP axes on the plot, in an analogous way as in Section 5.2. That is, the original CP axes are given by the columns of **B**. To express them with respect to the columns of $\tilde{\mathbf{B}}$, the currently used axes, we simply use the co-ordinates in the columns of $\mathbf{T}^{-1}$, because $\mathbf{B} = \tilde{\mathbf{B}}\mathbf{T}^{-1}$, and we plot these columns as vectors in the current plot. Note that when **B** has unit column sums of squares, the columns of $\mathbf{T}^{-1}$ also have unit sums of squares (compare Section 5.2), and thus these vectors all have unit length. The CP *axes* can be plotted as the lines running through these vectors. The co-ordinates of the actual object/occasion combinations can now also be found as the linear combination of the vectors representing the oblique axes, using as weights the (original) co-ordinates with respect to these oblique axes. It can be verified that these linear combinations exactly lead to the locations at which the object/occasion combinations have now been plotted. Thus in the case of CP one can display in a single plot: the trajectories of the objects along which they travel over the occasions, the projected axes representing the variables, and the unique CP axes.

Rather than plotting trajectories for objects (combinations of A- and C-mode entities), in some cases one might be interested in plotting combinations of A- and B-mode entities or of B- and C-mode entities. The procedure for finding proper co-ordinates is fully analogous to that described above, but then based on $\hat{\mathbf{X}}_c = \mathbf{C}\mathbf{G}_c(\mathbf{B}'\otimes\mathbf{A}')$ and $\hat{\mathbf{X}}_a = \mathbf{A}\mathbf{G}_a(\mathbf{C}'\otimes\mathbf{B}')$ respectively. We have chosen to describe only the procedure for plotting combinations of A- and C-mode entities, because these will often pertain to trajectories over time for objects, which seems the most interesting application of combined plotting.

The procedure will now be illustrated for a data set described by Nomikos and MacGregor [26]. The data set consists of (simulated) measurements on 52 batches (A-mode) with respect to nine variables (B-mode), at 200 consecutive time points, with 5 min intervals (C-mode). The first 50 batches are more or less 'normally behaving' batches, whereas the 51st and 52nd are known to behave abnormally. The nine variables, as described by Nomikos and MacGregor [26], are: (1) flow rates of styrene, (2) flow rates of butadiene, (3) temperature of the feeds, (4) temperature of the reactor, (5) temperature of the cooling water, (6) temperature of the reactor jacket, (7) density of the latex in the reactor, (8) total conversion and (9) instantaneous rate of energy release.

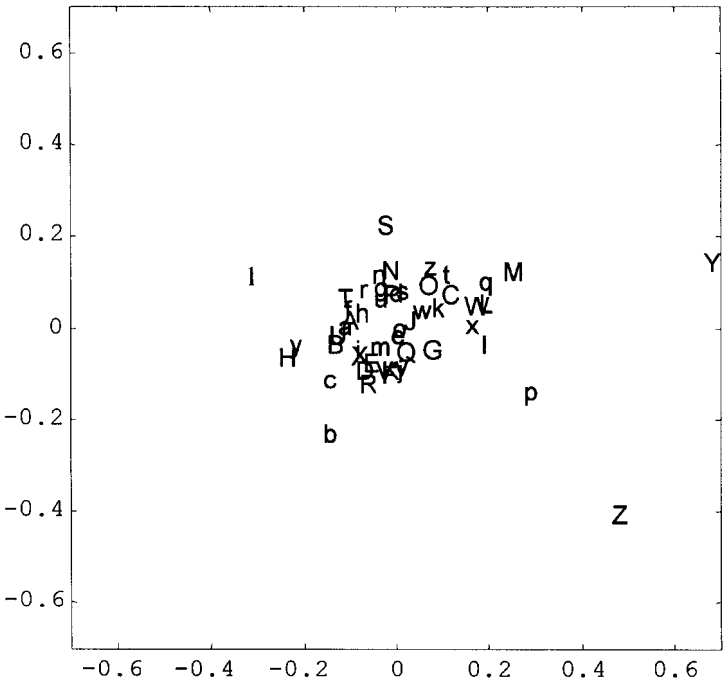

Figure 4. Plot of the 52 batches.

The data were preprocessed by centering across the 52 batches, and by normalizing each of the variables (i.e. across batches *and* time points) such that for each variable the sum of squares was unity. The thus preprocessed data have been analyzed by three-way Tucker analysis using various dimensionalities, and it was soon found that solutions using more than two components for each mode added relatively little to the fit of the model. For instance, the (2,2,2) solution gave 22·0% fit, while (3,2,2), (2,3,2) and (2,2,3) gave 23·3%, 22·4% and 23·3% fit respectively. It was also verified how well the two-component CP model fitted these data, and it was seen that this led to a degenerate, uninterpretable solution. Therefore we chose to use results of the Tucker analysis, using two components for each mode, to illustrate our plotting procedure.

We first produced a plot of the batches (Figure 4), hence of the A-mode entities, in the same way as described in Section 5·1. Batches 1,…,52 were labeled a,…,z,A,…,Z. It can be seen clearly that the abnormal batches (Y and Z) differ considerably from the other batches as well as from each other. The main purpose of the present section, however, is to illustrate a plot of trajectories of batches. As each trajectory pertains to 200 time points and as there are 52 batches, the plot would get cluttered if we were to display all batches simultaneously. Therefore in Figure 5 we give the trajectories for four batches only: batches 2, 12, 51 and 52. These were selected because they could be expected to differ considerably, as follows from Figure 4: batches 2 and 12 are normally behaving batches (in the plot denoted by b and l) and can be seen to be relatively wide apart: batches 51 and 52 are the two abnormal ones and were already noted to be wide apart both from the others as well as from each other. In the plot we also give the projections of the variable axes onto the subspace displayed. It should be realized, however, that variables 1, 2, 3 and 4 are represented rather poorly by this Tucker3 analysis (fit percentages smaller than 4%), and hence interpretation of the plot with respect to these variables may not be very reliable.

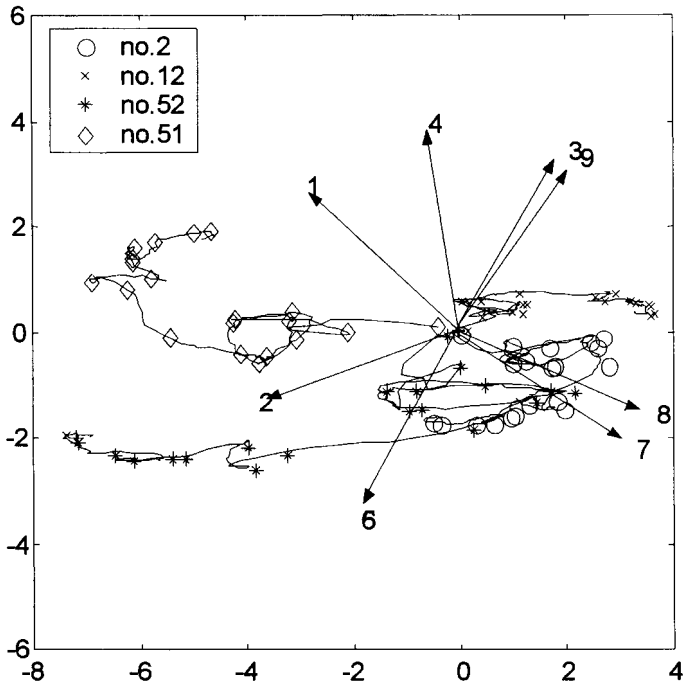

Figure 5. Trajectories for batches 2, 12, 51 and 52. All trajectories start near the origin (marked by '+'); the arrows represent the projected variable axes (5 and 6 coincide).

From Figure 5 we learn quickly that the trajectories for the four batches displayed differ considerably. The trajectories for batches 2 and 12 remain mostly on the right side of the plot, whereas those for batches 51 and 52 run from the right to the left. Batch 52 is known to start being deviant only halfway through the process (see Reference [26] p. 1365); it can be seen to indeed remain in the 'normal area' for a relatively long time, before it starts diverging towards the lower left. Batch 51 diverges right from the start. Furthermore, we can follow each separate trajectory more closely, and see, for instance, that batch 2 starts moving in the direction of arrows 7 and 8 (thus indicating a gradual growth of scores on variables 7 and 8), and eventually bends towards arrow 6 (implying that scores on variable 6 eventually increase considerably). The trajectory for batch 12 is totally different: it moves quite consistently towards the right (mainly indicating increasingly high scores on variables 7, 8 and 9). As said, the trajectories for batches 51 and 52 mainly move from right to left and hence show decreasing scores on variables 7, 8 and 9. It can also be seen from the projection of these trajectories on arrow 6 that batch 52 gets increasingly high scores on this variable, whereas for batch 51 this holds only for the first three-quarters of the trajectory, whilst, towards the end, there is a decrease in scores on variable 6.

For batches 2 and 52 we also plotted the actual (preprocessed) scores on the variables 5, 6, 7 and 9 (Figures 6 and 7) to demonstrate that, indeed, the general tendency displayed in the single plot in Figure 5 conforms to the data. Of course, the plots in Figures 6 and 7 give much more detail, but the general tendencies of increase or decrease conform very well to the data, and it can be concluded that the plot in Figure 5 nicely summarizes the main information in the data. The advantage of Figure 5 is that it captures all the information in a single plot, which allows for easy mutual comparison of trajectories of different batches.

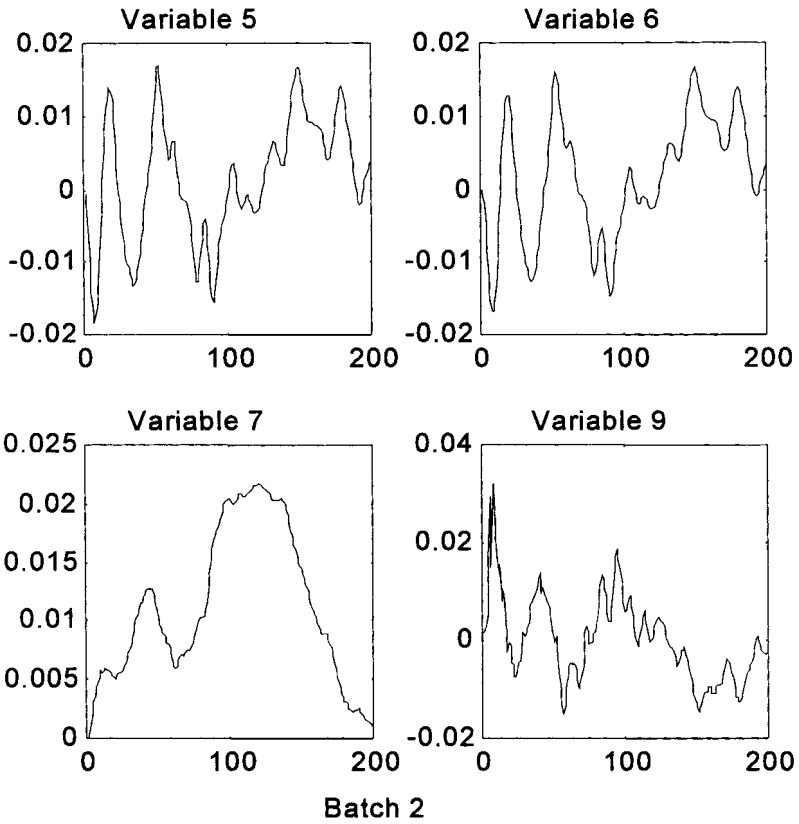

Figure 6. Plots of preprocessed scores of batch 2 on variables 5, 6, 7 and 9.

7. DISCUSSION

The present paper has described some procedures for plotting the low-dimensional configurations that are obtained by means of three-way analysis as approximations to the high-dimensional data configuration for one set of entities or for a combination of two sets of entities. The main rule that should be obeyed is that, to adequately plot one set of entities, one should use co-ordinates with respect to orthonormal basis vectors. In the procedures described, it is derived how the co-ordinates should be computed for such plots and how the original axes can be projected into the same plot. The procedure is described for the general R-dimensional situation, but in the actual practice of producing plots, R is usually restricted to 2 or 3. In cases where higher-dimensional configurations are used, one can display these by either plotting co-ordinates for all combinations of pairs (or triples) of dimensions (which actually comes down to orthogonal projections of the full configuration), or only use consecutive pairs (or triples). The latter may very well suffice, if care has been taken that points are located reasonably close to these planes (or subspaces), as is typically achieved by simple structure rotations. Thus, in cases where more than two or three dimensions are to be used, it is strongly recommended to precede these plottings by procedures for simple structure rotation [27].

The plotting procedures described in the present paper focus on displaying the approximation of the high-dimensional configurations entailed by the original data. Other plotting procedures (e.g. joint plots [8]) focus on plotting 'latent' data configurations, which together, by means of linear

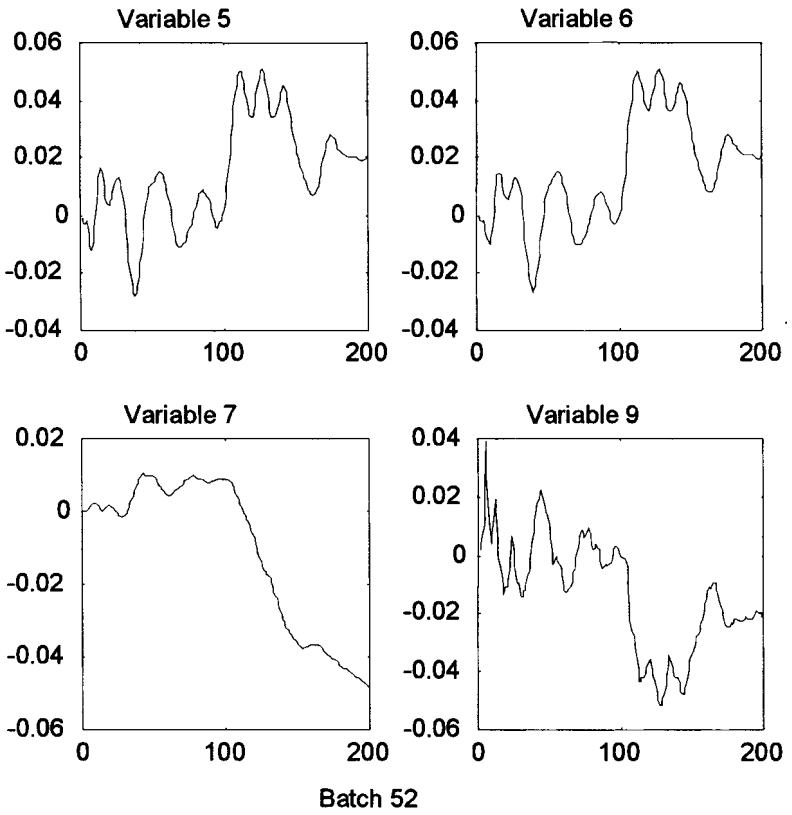

Figure 7. Plots of preprocessed scores of batch 52 on variables 5, 6, 7 and 9.

combinations, approximate the data. Both approaches are mainly meant to supplement the interpretation of the results of one's three-way analysis, and should not completely replace the results that are usually given in tabular form (such as those for the core array in three-way Tucker analysis), nor do they obviate the dimension-by-dimension interpretation used in CP.

In the present paper, only in the illustrative analyses, mention has been made of how the data are treated before analysis. It should be noted, however, that centering the data across one mode seriously affects the meaning of the data elements and should hence be taken into account. Often this centering is mainly meant to eliminate additive constants from the data. In such cases, estimates of these eliminated additive constants could be used when interpreting the complete model, or else, in interpreting the results, the estimated data should be considered as *deviations* from these constants (for instance, deviations from the means). What is most appropriate depends on the particular situation at hand. In cases where additive constants are explicitly modeled with the data, in displays of the data these additive terms could also be taken into account as additional dimensions. For instance, suppose the data are modeled as $\hat{\mathbf{X}}_a = \mathbf{1}\boldsymbol{\mu}' + \mathbf{A}\mathbf{G}_a(\mathbf{C}' \otimes \mathbf{B}')$, where $\mathbf{1}$ denotes an I-vector with unit elements only and $\boldsymbol{\mu}$ denotes a JK-vector containing additive terms for all combinations of B- and C-mode entities. In cases where interest is in plotting the A-mode units, one may choose to simply plot the (usually columnwise centered) matrix $\mathbf{A}$, and consider these co-ordinates as co-ordinates with respect to a translated origin (e.g. translated to the centroid of the configuration). When interest is in plotting the combined B- and C-mode units, it is useful to write this model as

$\hat{\mathbf{X}}_\mathbf{a} = \mathbf{(1|A)(}\boldsymbol{\mu}\mathbf{|(C}\otimes\mathbf{B)G_a}')' = \tilde{\mathbf{A}}\tilde{\mathbf{F}}'$, thus involving one additional dimension for the whole configuration. If interest is in displaying other configurations (e.g. that for the B-mode entities or the combined A- and C-mode entities), it should be noted that the full model does not represent projections of these configurations on a low-dimensional subspace, and hence the present plotting procedures make no sense. In such cases, either the results should be interpreted in terms of configurations of centered data, or an alternative approach to modeling additive terms that *is* in line with the configurations that are to be displayed should be used. Such alternative models would be of the form $\hat{\mathbf{X}}_\mathrm{a} = \boldsymbol{\mu}_\mathrm{a}(\mathbf{1}\otimes\mathbf{1})' + \mathbf{1}(\mathbf{1}\otimes\boldsymbol{\mu}_\mathrm{b})' + \mathbf{1}(\boldsymbol{\mu}_\mathrm{c}\otimes\mathbf{1})' + \mathbf{AG}_\mathrm{a}(\mathbf{C}'\otimes\mathbf{B}')$, possibly with one or two of the additive terms left out. For each mode the latter model gives a projection on a subspace which is at most three dimensions higher than the subspaces involved in $\mathbf{AG}_\mathrm{a}(\mathbf{C}'\otimes\mathbf{B}')$.

The plotting procedures described in the present paper are used for displaying results from three-way methods, but actually only rely on two-way (PCA) models, obtained after rewriting the three-way models at hand. It follows that the plots do not capture all the information available in the results of a three-way method, and they should never be used to *replace* the full results of a three-way analysis. Rather they should be used to complement such results. Results from four- and, in general, *N*-way methods can likewise be displayed by rewriting the models back into two-way PCA form. In this respect the present paper is by no means limited to three-way data. However, for higher-way data it will be not be as feasible to represent the entities of all modes in one plot, as was done in Section 6.

ACKNOWLEDGEMENTS

The author is obliged to Jos ten Berge and Pieter Kroonenberg for comments on an earlier version of this manuscript, and to John MacGregor and Dora Kourti (McMaster, Canada) for making available the data used in Section 6.

APPENDIX. SUMMARY OF MAIN PLOTTING PROCEDURES FOR THREE-WAY METHODS

In Sections 5 and 6, procedures have been described for plotting one or two modes of entities, on the basis of results from three-way methods. These can be summarized as follows.

Given: $\hat{\mathbf{X}}_\mathrm{a} = \mathbf{AG}_\mathrm{a}(\mathbf{C}'\otimes\mathbf{B}')$, where $\mathbf{G}_\mathrm{a} = \mathbf{I}_\mathrm{a}$ in case of CP; define $\mathbf{F} \equiv \mathbf{(C}\otimes\mathbf{B)G_a}'$.

1. *Displaying approximate configuration of A-mode entities*

(a) Find $\mathbf{T}$ such that, for $\tilde{\mathbf{F}} = \mathbf{FT}, \tilde{\mathbf{F}}'\tilde{\mathbf{F}} = \mathbf{I}$ holds (e.g. by Gram–Schmidt orthonormalization).
(b) Plot rows of $\tilde{\mathbf{A}} = \mathbf{A(T')}^{-1}$ to display the I A-mode entities.
(c) Plot rows of $\tilde{\mathbf{F}}$ to display projections of the JK original axes onto subspace.
(d) (In case of CP) Plot rows of $(\mathbf{T}')^{-1}$ to display projection of unique axes (columns of $\mathbf{F}$) onto subspace.

2. *Displaying approximate configuration of combination of B- and C-mode entities*

(a) Find $\mathbf{T}$ such that, for $\tilde{\mathbf{A}} = \mathbf{AT}, \tilde{\mathbf{A}}'\,\tilde{\mathbf{A}} = \mathbf{I}$ holds (e.g. by Gram–Schmidt orthonormalization).
(b) Plot rows of $\tilde{\mathbf{F}} = \mathbf{F(T')}^{-1}$ to display the JK combinations of B- and C-mode entities (forming trajectories).
(c) Plot rows of $\tilde{\mathbf{A}}$ to display projections of the I original axes onto subspace.
(d) (In case of CP) Plot rows of $(\mathbf{T}')^{-1}$ to display projection of unique axes (columns of $\mathbf{A}$) onto subspace.

REFERENCES

1. Tucker LR. Some mathematical notes on three-mode factor analysis. *Psychometrika* 1966; **31**: 279–311.
2. Kroonenberg PM, de Leeuw J. Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika* 1980; **45**: 69–97.
3. Carroll JD, Chang JJ. Analysis of individual differences in multidimensional scaling via an N-way generalization of 'Eckart–Young' decomposition. *Psychometrika* 1970; **35**: 283–319.
4. Harshman RA. Foundations of the PARAFAC procedure: models and conditions for an 'explanatory' multi-mode factor analysis. *UCLA Working Papers Phonet*. 1970; **16**: 1–84.
5. Tucker LR. Intra-individual and inter-individual multidimensionality. In *Psychological Scaling: Theory and Applications*, Gulliksen H, Messick S (eds). Wiley: New York, 1960; 155–167.
6. Gabriel KR. The biplot-graphic display of matrices with applications to principal components analysis. *Biometrika* 1971; **58**: 453–467.
7. Gower JC, Hand DJ. *Biplots.* Chapman and Hall: London, 1996.
8. Kroonenberg PM. *Three Mode Principal Component Analysis: Theory and Applications*. DSWO Press: Leiden, 1983.
9. Kroonenberg PM. The TUCKALS line: a suite of programs for three-way data analysis. *Comput. Statist. Data Anal.* 1994; **18**: 73–96.
10. Harshman RA, Lundy ME. The PARAFAC model for three-way factor analysis and multidimensional scaling. In *Research Methods for Multimode Data Analysis*, Law HG, Snyder CW, Hattie JA, McDonald RP (eds). Praeger: New York, 1984; 122–215.
11. Kiers HAL. Towards a standardized notation and terminology in multiway analysis. *J. Chemometrics* 2000; **14**: 105–122.
12. Kiers HAL. Hierarchical relations among three-way methods. *Psychometrika* 1991; **56**: 449–470.
13. Harshman RA, De Sarbo WS. An application of PARAFAC to a small sample problem, demonstrating preprocessing, orthogonality constraints, and split-half diagnostic technique. In *Research Methods for Multimode Data Analysis*, Law HG, Snyder CW, Hattie JA, McDonald RP (eds). Praeger: New York, 1984; 602–642.
14. Harshman RA. Determination and proof of minimum uniqueness conditions for PARAFAC1. *UCLA Working Papers Phonet*. 1972; **22**: 111–117.
15. Harshman RA, Lundy ME. PARAFAC: parallel factor analysis. *Comput. Statist. Data Anal.* 1994; **18**: 39–72.
16. Lundy ME, Harshman RA, Kruskal JB. A two-stage procedure incorporating good features of both trilinear and quadrilinear models. In *Multiway Data Analysis*, Coppi R, Bocasco S (eds). Elsevier: Amsterdam, 1989; 123–130.
17. Meyer JP. Causal attribution for success and failure: a multivariate investigation of dimensionality, formation, and consequences. *J. Personality Social Psychol.* 1980; **38**: 704–718.
18. Harshman RA, Ladefoged P, Goldstein L. Factor analysis of tongue shapes. *J. Acoust. Soc. Am.* 1977; **62**: 693–707.
19. Kroonenberg PM. Multivariate and longitudinal data on growing children. Solutions using a three-mode principal component analysis and some comparison results with other approaches. In *Data Analysis. The Ins and Outs of Solving Real Problems*, Janssen J, Marcotorchini F, Proth JM (eds). Plenum: New York, 1987; 89–112.
20. Kroonenberg PM, De Leeuw J. TUCKALS2: a principal component analysis of three mode data. Res. Bull. RB 001–77, University of Leiden, 1977.
21. Kaiser HF. The varimax criterion for analytic rotation in factor analysis. *Psychometrika* 1958; **23**: 187–200.
22. Doledec S, Chessel D. Rythmes saisonniers et composantes stationelles en milieu aquatique. I. Description d'un plan d'observation complet par projection de variables. *Acta Oecol./Oecol. Gen.* 1987; **8**: 403–406.
23. Sabatier R. Méthodes factorielles en analyse des données. Doctoral *Dissertation*, University of Montpellier, 1987.
24. Timmerman ME, Kiers HAL. Three-mode principal components analysis: choosing the numbers of components and sensitivity to local optima. *Br. J. Math. Statist. Psychol.* in press.
25. Bove G, Di Ciaccio A. Comparisons among three factorial methods for analysing three-mode data. In *Multiway Data Analysis*, Coppi R, Bolasco S (eds). Elsevier: Amsterdam, 1989; 103–113.
26. Nomikos P, MacGregor JF. Monitoring batch processes using multiway principal component analysis. *AIChE J* 1994; **40**: 1361–1375.
27. Kiers HAL. Joint orthomax rotation of the core and component matrices resulting from three-mode principal components analysis. *J. Classif.* 1998; **15**: 245–263.

