ELSEVIER Computational Statistics & Data Analysis 40 (2002) 447–470

COMPUTATIONAL STATISTICS & DATA ANALYSIS

www.elsevier.com/locate/csda

Three-way component analysis with smoothness constraints

Marieke E. Timmerman*, Henk A.L. Kiers

Heymans Institute of Psychology, DPMG, University of Groningen, Grote Kruisstraat 211, 9712 TS Groningen, The Netherlands

Abstract

Tucker3 Analysis and CANDECOMP/PARAFAC (CP) are closely related methods for three-way component analysis. Imposing constraints on the Tucker3 or CP solutions can be useful to improve estimation of the model parameters. In the present paper, a method is proposed for applying smoothness constraints on Tucker3 or CP solutions, which is particularly useful in analysing functional three-way data. The usefulness of smoothness constraints on Tucker3 and CP solutions is examined by means of a simulation experiment. Generally, the results of the experiments indicate better estimations of the model parameters. An empirical example illustrates the use of smoothness constraints. The constrained model is more stable and easier to interpret than the unconstrained model. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Three-way data; Longitudinal data; Splines; Smoother

1. Introduction

This paper deals with methods for analysing three-way functional data. A univariate data series is functional if the (true) scores can be described as a function of a certain predictor, for example time or distance (Ramsay and Silverman, 1997). An example is a series of scores on a variable indicating reading ability, that is repeatedly collected from a pupil receiving reading education. In practice, one often collects multivariate functional data, for example repeatedly obtained reading ability scores from a number of pupils. A commonly used approach to revealing the structure in two-way functional data is principal component analysis (PCA) (Tucker, 1958, 1966; Rao, 1958). Functional PCA, as it is called by Ramsay and Silverman (1997), aims at describing the

* Corresponding author. Tel.: +31-50-363-6255; fax: +31-50-363-6304.
E-mail address: m.e.timmerman@ppsw.rug.nl (M.E. Timmerman).

PII: S0167-9473(02)00059-2

dominant modes of variation of a functional data set and at obtaining a comprehensible representation of the structure of variability. An example of functional three-way data is repeatedly measured scores of a number of pupils on a number of variables, that are indicators of reading ability. As Tucker3 analysis (Tucker, 1966; Kroonenberg and De Leeuw, 1980) and CANDECOMP/PARAFAC (CP; Carroll and Chang, 1970; Harshman, 1970) are three-way generalisations of principal component analysis, they are natural analysis approaches for three-way functional data.

Although Tucker3 and CP analyses are often used in an exploratory way, previous knowledge of data generating processes can be used to constrain the model (e.g., see Bro, 1998). Possible advantages of constraining a model are more stable parameter estimates, and reduction of numerical problems and computation time. In the present paper, the use of *smoothness* constraints, which is particularly useful in the case of functional data analysis, in Tucker3 and CP models is elaborated. Smoothness constraints could be combined with monotonicity constraints, which is possibly useful in the case of growth data. It will be shown that the use of a particular smoothness constraint on the components leads to equivalent parameter estimates as smoothing the observed measurements before analysis. This is important because it obviates the difficult question whether one should analyse the original data by a method where the results are *constrained* to be smooth, or smooth the data, and then analyse the smoothed data. The usefulness of smoothing in the Tucker3 model and CP model will be examined under different conditions in a simulation experiment. An empirical example illustrates the use of smoothness constraints in the Tucker3 model.

2. Tucker3 analysis and CANDECOMP/PARAFAC

Tucker3 analysis (Tucker, 1966; Kroonenberg and De Leeuw, 1980; Kroonenberg, 1983) is a generalisation of two-way PCA. In a two-way PCA, the data are decomposed into two matrices. In a Tucker3 analysis, the three-way data are decomposed into three component matrices and a so-called core array, which denotes the importances of the relationships between the different components. In this paper, we start from functional three-way data that are collected in an $I \times J \times K$ three-way array $\underline{\mathbf{X}}$, where $i = 1,\dots,I$ refers to subject i, $j = 1,\dots,J$ to variable j and $k = 1,\dots,K$ to measurement k. Such a data set could result, for example, from a research into reading ability development of I pupils during the first year of reading education. During this first year we could weekly gather (for K weeks) scores on J variables that are indicative of aspects of reading ability.

The Tucker3 model is defined as

$$\mathbf{X_c} = \mathbf{C}\mathbf{G_c}(\mathbf{B}' \otimes \mathbf{A}') + \mathbf{E_c}, \tag{1}$$

where $\mathbf{X_c}$ $(K \times IJ)$, $\mathbf{G_c}$ $(R \times PQ)$, and $\mathbf{E_c}$ $(K \times IJ)$ denote the 'matricised' versions of the three-way data array $\underline{\mathbf{X}}$ $(I \times J \times K)$, the core array $\underline{\mathbf{G}}$ $(P \times Q \times R)$, and residual array $\underline{\mathbf{E}}$ $(I \times J \times K)$, respectively (e.g., see Kiers, 2000), $\mathbf{A}$ $(I \times P)$, $\mathbf{B}$ $(J \times Q)$, and $\mathbf{C}$ $(K \times R)$ denote the subject, variable and occasion component matrices, respectively, and $\otimes$ denotes the Kronecker product.

The Tucker3 model is usually fitted to data in the least squares sense, hence by minimising the sum of squared residuals, with **A**, **B**, and **C** usually restricted to orthonormality without loss of fit. Kroonenberg and De Leeuw (1980) offered an alternating least squares (ALS) algorithm to fit the Tucker3 model to data. Various variants and improvements of that algorithm have been proposed. The algorithms are essentially based on alternately updating the matrices **A**, **B**, **C**, and $\mathbf{G_c}$, keeping the other matrices fixed. Each iteration decreases the sum of squared residuals, implying that the function value decreases monotonically. Because of the boundedness of the function it converges to a stable function value, although it does not necessarily converge to the global minimum. By using several differently (e.g., randomly) started runs, one can reduce the chance to miss the global minimum. The Tucker3 model is not uniquely defined. All matrices $\mathbf{A}, \mathbf{B}$, and **C** can be rotated orthogonally or obliquely, provided that such rotations are compensated in the core.

The CANDECOMP/PARAFAC (CP) model (Carroll and Chang, 1970; Harshman, 1970) is a constrained version of the Tucker3 model. The CP model is defined as

$$\mathbf{X_c} = \mathbf{CH}(\mathbf{B}' \otimes \mathbf{A}') + \mathbf{E_c}, \tag{2}$$

where the matrix **H** is the $(R \times R^2)$ two-way version of the 'superidentity' three-way array $\underline{\mathbf{H}}$, that is, an array with $h_{pqr} = 1$ if $p = q = r$, and $h_{pqr} = 0$ otherwise. The model is uniquely defined under certain (weak) conditions, which are usually met in practice. That is, the estimations of **A**, **B**, and **C** are unique up to an arbitrary scaling of the columns in two of the component matrices and an arbitrary simultaneous permutation of the columns in the component matrices (Harshman, 1972).

The CP model is fitted to data in the least squares sense, and the parameters of CP are estimated using an ALS algorithm (Harshman, 1970; Carroll and Chang, 1970). Just as in fitting the Tucker3 model, the CP algorithm can end up in a local minimum. The use of several differently started runs can increase the chance to attain the global minimum.

A Tucker3 or CP analysis can be applied to a raw data array $\underline{\mathbf{X}}$, but usually the data are centred and/or scaled before analysis (Harshman and Lundy, 1984; Bro and Smilde, 2001).

3. Smoothing in the Tucker3 model and CP model

3.1. The use of smoothness constraints in the Tucker3 model and CP model

Generally, fitting a model to observed data aims at obtaining an *interpretable* model with a small degree of overall error, which refers to the lack of fit of the current model fitted to the current data set to the population data (Browne and Cudeck, 1992). The use of prior knowledge to imposing constraints on the Tucker3 and CP may lead to a better interpretable model, that has a smaller degree of overall error than its unconstrained counterpart. Additionally, a constrained solution may be less likely to end up in a local minimum and/or it can converge in fewer iterations, because the solution space is limited. We expect this kind of positive effects in CP analyses rather than the Tucker3 analyses, because CP analyses, particularly of large data sets and/or data with mild to severe multicollinearity, are known for its computational problems.

The type of constraint we will discuss here is useful to apply in the Tucker3 or CP analysis of functional three-way data, hence if the three-way data consist of various samples from what can be expected to be continuous curves with measurement error. Suppose we are interested in reading ability development of pupils during the first year of reading education. During this year we could weekly gather scores on a number of variables that are indicative of aspects of reading ability. The various univariate series, that consists of scores of a pupil on a variable gathered on the successive weeks, can be viewed as evaluations of the underlying reading ability curve at the particular measurement times. If the scores on a certain variable could be expressed as any *known* function of time with known parameters, then any deviations in observed scores are measurement error. In social sciences, specific functions generating the data are usually unknown, and even an appropriate functional form for describing the data is difficult to choose. Hence, a technique that does not impose rigid parametric assumptions about the dependence of scores and time is warranted. An approach is the use of a so-called smoother. Smoothing in the Tucker3 or CP model can be performed by smoothing the raw data, hence the various univariate series per subject and variable, before analysis by the unconstrained Tucker3 or CP model. By doing this, one aims at (partly) eliminating measurement error from the data. Alternatively, one could constrain the component scores in the Tucker3 or CP model to be smooth and thus fit a constrained model to the raw data. One hopes that the thus fitted model covers less measurement error than its unconstrained counterpart. Eventually, in both approaches, one aims at fitting models that has a smaller degree of overall error than the unconstrained counterparts. In the case of growth data, it might be useful to combine smoothness constraints with monotonicity constraints. Fortunately, the question which of the two approaches to take is trivialised in an important class of cases. As will be shown in the following, fitting the smooth descriptions of the observed data by the unconstrained Tucker3 or CP model is equivalent to fitting the Tucker3 or CP model with smoothness constraints in a particular class of cases.

To facilitate the explanation, it is assumed that the variable scores are gathered at the same time points for all subjects. The latter is not strictly necessary in a smoothness constrained Tucker3 or CP, as will be explained in the Discussion section.

3.2. The choice of a smoother

A smoother is used to describe a response measurement as a smooth function of one or more predictor measurements (Hastie and Tibshirani, 1990), usually by so-called local averaging. In our applications, the predictor is the time at which a measurement is taken. Local averaging aims at averaging the observed measurements associated with predictor values close to each other (i.e., in each others neighbourhood). The different types of smoothers mainly differ in their method of averaging. The size of the neighbourhood influences the smoothness, and the accuracy: a large neighbourhood leads to an estimate with low variance (i.e., high smoothness) but high potential bias (i.e., low accuracy), whereas the opposite holds for small neighbourhoods (Hastie and Tibshirani, 1990).

Hastie and Tibshirani (1990) and Ramsay and Silverman (1997) offer overviews of different smoothers and their properties. Polynomial regression splines form a class of smoothers that is computationally convenient, and they will be used here. Polynomial regression splines are constructed from different polynomial pieces, which are joined at certain predictor values, the knots. A popular type of polynomial splines are B-splines (De Boor, 1978), which can be used easily for smoothing the data before analysis (see Alsberg and Kvalheim (1993) for an example involving three-way data). Monotone smoothness constraints, which can be useful in longitudinal applications, can be imposed by using I-splines (Ramsay, 1988).

B-splines (basis splines) are non-negative basis functions. The degree (d) of a B-spline is the degree of the polynomial pieces on which it is based. Each B-spline is determined by its degree and by its knot sequence. The knots are positioned in the domain between the minimal value of the predictor ($t_{\min}$) and the maximal value of the predictor ($t_{\max}$). If they are of equal degree and they are positioned equidistantly, the basis functions are equal in size and shape. The polynomial pieces join at d inner knots, and at these joining points, the derivative up to order $d-1$ is continuous. The number of non-zero B-splines (N) on the domain $t_{\min}$ to $t_{\max}$ is equal to the total number of knots plus the degree of the polynomial minus 1. A B-spline is positive on a domain spanned by $d+2$ knots, and everywhere else it is zero. Any degree of the polynomial can be chosen. Given the degree and the location of the knots, B-splines can be defined by a recursive formula (De Boor, 1978).

Usually, a set of response measurements collected in $\mathbf{y}$ $(K \times 1)$ is to be approximated by linear combinations of the B-splines, that are evaluated in the values of the predictor $\mathbf{t}$. Let $\mathbf{B}^{\mathrm{s}}$ denote a $(K \times N)$ B-spline matrix, in which the nth column contains the values of the nth B-spline that is evaluated in all values of the predictor $\mathbf{t}$ $(K \times 1)$; let $\mathbf{w}$ $(N \times 1)$ denote the vector with weights for the N B-splines, and $\hat{\mathbf{y}}$ the vector with estimated response measurements, which is called the *smooth* in the sequel, then

$$\hat{\mathbf{y}} = \mathbf{B}^{\mathrm{S}}\mathbf{w}. \tag{3}$$

I-splines (integrated splines; Ramsay, 1988) are monotonically increasing basis functions. They are based on integrated M-splines, which are proportional to B-splines. Because M-splines are non-negative everywhere, the integrated M-splines are a natural basis for monotone splines. Since bases for I-splines are monotonically increasing, a non-negativity constraint on the set of coefficients of the I-splines leads to monotonically non-decreasing estimated response variables.

The use of B-splines and I-splines requires the choice of the 'smoothing parameters', that is the degree of the splines and the number and the position of the knots. Commonly, the degree of the splines is fixed. For B-splines, a popular choice is a third degree B-spline (Hastie and Tibshirani, 1990; p. 22); smoothers based on higher degree splines tend to oscillate wildly (Van Rijckevorsel, 1988). Ramsay (1988) claims that the use of low (e.g., second) degree I-splines generally suffices. While fixing the degree of the splines, the number and location of knots are used to influence the smooth. More knots in a region lead to a greater flexibility of estimation in that region. The smoothing parameters can be selected by subjective comparison of the observed and several estimated response variables. Automatic selection methods for the smoothing parameters are also available (see Hastie and Tibshirani, 1990, pp. 42–52). Although the usefulness of these methods is debatable, these methods can be helpful in deciding about the number of knots. A commonly used procedure is cross-validation by means of the leave-one-out approach. Hastie and Tibshirani (1990, pp. 46–48) showed that the cross-validation sum of squares for linear smoothers can be computed by

$$\mathrm{CV}(\lambda)=\frac{1}{K}\sum_{k=1}^{K}\left(\frac{y_k-\hat{y}_k}{1-S(\lambda)_{kk}}\right)^2, \tag{4}$$

where λ denotes the smoothing parameters (i.e., the degree, and the number and the positions of the knots), and $\mathbf{S}(\lambda)_{kk}$ are the diagonal elements of the projection-matrix $\mathbf{S}(\lambda)$, which relates $\hat{\mathbf{y}}$ to $\mathbf{y}$. To select one's smoothing parameters, one may search those that minimise $\mathrm{CV}(\lambda)$.

As to choosing the *position* of the knots, a simple approach is to position them uniformly over the predictor domain. Alternatively, one could place them at appropriate quantiles of the predictor variable. Knot optimisation techniques exist (De Boor, 1978), but they are only useful if the function to be estimated has distinct and known discontinuities (Van Rijckevorsel, 1988).

3.3. How to smooth in the Tucker3 and CP?

Smoothing in a Tucker3 or CP analysis can be performed by smoothing the raw data before an unconstrained Tucker3 or CP analysis, or by constraining the component scores to be smooth in the (constrained) Tucker3 or CP model. Ramsay and Silverman (1997, Chap. 7) used the latter approach in a functional PCA by applying a roughness penalty to the estimated principal components. We choose a different approach, which has the advantage that smoothing the raw data and smoothing the components lead to the same estimated model parameters. We propose to impose a smoothness constraint on the occasion component matrix $\mathbf{C}$ by constraining $\mathbf{C}$ $(K \times R)$ such that it can be written as $\mathbf{B}^{\mathrm{s}}\mathbf{U}$, for a B-spline matrix $\mathbf{B}^{\mathrm{s}}$ $(K \times N)$ and a particular weight matrix $\mathbf{U}$ $(N \times R)$, and where $N \geqslant R$. As a result the Tucker3 and CP model with smoothness constraints on the occasion component matrix can be written as

$$\mathbf{X_c}=\mathbf{B^SUG_c}(\mathbf{B}'\otimes\mathbf{A}')+\mathbf{E_c}, \tag{5}$$

where $\mathbf{X_c}$ denote the $K \times IJ$ matricised data array $\underline{\mathbf{X}}$, $\mathbf{B}^{\mathrm{s}}$ $(K \times N)$ a B-spline matrix, $\mathbf{U}$ $(N \times R)$ $(N \geqslant R)$ a weight matrix, $\mathbf{A}$ $(I \times P)$ and $\mathbf{B}$ $(J \times Q)$ component matrices, $\mathbf{G_c}$ $(R \times PQ)$ the supermatrix containing the lateral slices of the core array $\underline{\mathbf{G}}$ $(P \times Q \times R)$, and $\mathbf{E_c}$ $(K \times IJ)$ the matricised error array $\underline{\mathbf{E}}$; in the case of the CP model with smoothness constraints the core array $\underline{\mathbf{G}}$ is fixed at superidentity. The B-spline bases are computed using 'time' as a predictor. Note that the same basis is used for all components. In fact, formula (3) is used repeatedly for $r = 1,\ldots,R$ as $\mathbf{c}_r = \mathbf{B^S u}_r$. If monotonicity restrictions are required on the component matrix $\mathbf{C}$, it is proposed to replace the B-splines basis matrix by an I-splines basis matrix, and to impose non-negativity constraints on the weights.

The Tucker3 and CP models with smoothness constraints are fitted to data by minimising the sum of squared residuals, just as their unconstrained counterpart. Now, it will be shown that restricting the component matrix **C** to be in the column space of the B-spline matrix $\mathbf{B}^{\mathbf{s}}$ in the Tucker3 or CP is equivalent to analysing the projection of the data matrix $\mathbf{X}_{\mathbf{c}}$ on $\mathbf{B}^{\mathbf{s}}$ by the unrestricted Tucker3 or CP, which in turn comes down to Tucker3 or CP applied to the B-spline smoothed data. To show this, we replace $\mathbf{B}^{\mathbf{s}}$ by the QR-factorisation $\mathbf{B}^{\mathbf{s}} = \mathbf{QR}$ (see Golub and Van Loan, 1989, p. 211), with $\mathbf{Q}$ $(K \times N)$ columnwise orthonormal, and $\mathbf{R}$ $(N \times N)$ a square upper triangular matrix. Note that since $\mathbf{B}^{\mathbf{s}}$ is of full rank, **R** is non-singular. Then, the function to be minimised is

$$f_1(\mathbf{U}, \mathbf{A}, \mathbf{B}, \mathbf{G_c}) = \|\mathbf{X_c} - \mathbf{QRUG_c}(\mathbf{B}' \otimes \mathbf{A}')\|^2, \tag{6}$$

with $\mathbf{X_c}$ $(K \times IJ)$ the matricised data array $\underline{\mathbf{X}}$, and $\mathbf{G_c}$ $(R \times PQ)$ the matricised core in Tucker3, or the matricised superidentity array in the case of CP. As already noted by Carroll et al. (1980, p. 7), minimisation of (6) is equivalent to minimising

$$f_2(\tilde{\mathbf{U}}, \mathbf{A}, \mathbf{B}, \mathbf{G_c}) = \|\mathbf{Q}'\mathbf{X_c} - \tilde{\mathbf{U}}\mathbf{G_c}(\mathbf{B}' \otimes \mathbf{A}')\|^2, \tag{7}$$

with $\tilde{\mathbf{U}}$ $(N \times R)$ written for **RU**.

It will now be shown that minimising (7) is equivalent to analysing the smoothed version of **X** (using the B-spline matrix $\mathbf{B}^{\mathbf{s}}$) by unrestricted Tucker3 or CP. Smoothing the data matrix $\mathbf{X_c}$ by means of B-splines before Tucker3 or CP analysis is achieved by minimising

$$f_3(\mathbf{W}) = \|\mathbf{X_c} - \mathbf{B}^{\mathbf{s}}\mathbf{W}\|^2. \tag{8}$$

The optimal weights **W** are given by $(\mathbf{B}^{\mathbf{S}\prime}\mathbf{B}^{\mathbf{S}})^{-1}\mathbf{B}^{\mathbf{S}\prime}\mathbf{X_c}$, hence the smooth of $\mathbf{X_c}$ is $\hat{\mathbf{X}}_{\mathbf{c}} = \mathbf{B}^{\mathbf{S}}(\mathbf{B}^{\mathbf{S}\prime}\mathbf{B}^{\mathbf{S}})^{-1}\mathbf{B}^{\mathbf{S}\prime}\mathbf{X_c}$, the projection of $\mathbf{X_c}$ on $\mathbf{B}^{\mathbf{s}}$. Analysing this projection by Tucker3 or CP comes down to minimising

$$\begin{aligned} f_4(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{G_c}) &= |\hat{\mathbf{X}}_{\mathbf{c}} - \mathbf{CG_c}(\mathbf{B}' \otimes \mathbf{A}')\|^2 \\ &= \|\mathbf{B}^{\mathbf{s}}(\mathbf{B}^{\mathbf{s}\prime}\mathbf{B}^{\mathbf{s}})^{-1}\mathbf{B}^{\mathbf{s}\prime}\mathbf{X_c} - \mathbf{CG_c}(\mathbf{B}' \otimes \mathbf{A}')\|^2. \end{aligned} \tag{9}$$

Let $\mathbf{B}^{\mathbf{s}}$ be replaced by the QR-factorisation as $\mathbf{B}^{\mathbf{s}} = \mathbf{QR}$. Note that **R** is non-singular. Minimisation of (9) comes down to minimising

$$f_4(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{G_c}) = \|\mathbf{QQ}'\mathbf{X_c} - \mathbf{CG_c}(\mathbf{B}' \otimes \mathbf{A}')\|^2. \tag{10}$$

The optimal **C** will be in the column space of **Q**, hence **C** can be written as $\mathbf{Q}\tilde{\mathbf{C}}$, and minimising (10) is equivalent to minimising

$$\begin{aligned} f_5(\mathbf{A}, \mathbf{B}, \tilde{\mathbf{C}}, \mathbf{G}) &= \|\mathbf{QQ}'\mathbf{X_c} - \mathbf{Q}\tilde{\mathbf{C}}\mathbf{G_c}(\mathbf{B}' \otimes \mathbf{A}')\|^2 \\ &= \|\mathbf{Q}(\mathbf{Q}'\mathbf{X_c} - \tilde{\mathbf{C}}\mathbf{G_c}(\mathbf{B}' \otimes \mathbf{A}'))\|^2 \\ &= \|\mathbf{Q}'\mathbf{X_c} - \tilde{\mathbf{C}}\mathbf{G_c}(\mathbf{B}' \otimes \mathbf{A}')\|^2, \end{aligned} \tag{11}$$

(Kiers and Harshman, 1997). Clearly, minimising (11) is equivalent to minimising (7); the solutions for $\mathbf{A},\mathbf{B}$, and $\mathbf{G_c}$ of (7), and of (11) are equivalent; the solution for $\tilde{\mathbf{U}}$ in (7) is equivalent to that for $\tilde{\mathbf{C}}$ in (11). Because $\tilde{\mathbf{U}}$ leads to $\mathbf{C}$ by $\mathbf{C}=\mathbf{B^sU}=\mathbf{QRU}=\mathbf{Q}\tilde{\mathbf{U}}$, and $\tilde{\mathbf{C}}$ leads to $\mathbf{C}$ by $\mathbf{C}=\mathbf{Q}\tilde{\mathbf{C}}$, we see that both methods give the same solution for $\mathbf{C}$ as well. It has thus been shown that analysing the original data by means of a smooth constrained Tucker3 or CP is equivalent to analysing smoothed data by the unconstrained Tucker3 or CP, as long as smoothness is defined in terms of unrestricted linear combinations of B-splines.

The matrix $\mathbf{Q'X_c}$ in (11) is a compressed version of the smooth matrix $\mathbf{B^S}(\mathbf{B^{S\prime}B^S})^{-1}\mathbf{B^{S\prime}X_c}$ (see Kiers and Harshman, 1997). Since the matrix $\mathbf{Q'X_c}$ $(N \times IJ)$ is (much) smaller than the matrix $\hat{\mathbf{X}}_\mathbf{c}$, minimisation of (11) over $\mathbf{A}$, $\mathbf{B}$, $\tilde{\mathbf{C}}$, and $\mathbf{G_c}$ can be considerably faster than minimisation of (9) over $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{G_c}$. Moreover, the use of (unrestricted) splines on the data rather than on the components may be easier to handle, since standard software can be used to obtain the smooths and to analyse the smooths subsequently by Tucker3 or CP. Incidentally, if $I > JN$ or $J > IN$, minimisation of (11) can be speeded up further by applying a procedure discussed in Kiers and Harshman (1997, p. 37).

Constraining the B-spline or I-spline weights imposes constraints on the smooth. If I-spline weights are restricted to non-negativity, the smooth is non-negative and positive monotone increasing. The smooth can be restricted to be non-negative by requiring the B-splines weights to be non-negative. (Optimal) non-negative weights for the splines can be found by treating the problem as a non-negative least squares problem, which is solved by Lawson and Hanson (1974, pp. 158–164). Note that if spline weights are constrained, imposing a spline basis on a component matrix will have a different effect from imposing a spline basis on the data matrix. If a spline basis is imposed on a *component matrix* with constrained weights, we have to minimise (6) over $\mathbf{U},\mathbf{A},\mathbf{B}$, and $\mathbf{G_c}$, subject to appropriate constraints. If a spline basis is imposed on the *data matrix* with constrained weights $\mathbf{W}$, the constrained smooths will generally not be given by $\mathbf{B^S}(\mathbf{B^{S\prime}B^S})^1\mathbf{B^{S\prime}X_c}$, and the equivalence of minimising (11) and $f_6(\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{G_c})=\|\hat{\mathbf{X}}_\mathbf{c}-\mathbf{CG_c}(\mathbf{B}'\otimes\mathbf{A}')\|^2$ no longer holds.

In the case of modelling functional three-way data, it is expected that a smoothness constrained Tucker3 or CP model is better interpretable and has a smaller degree of overall error than its unconstrained counterparts, and that the smoothness constrained Tucker3 and CP models take less computing time per iteration. Furthermore, it is expected that the algorithm to fit a smoothness constrained CP will land in a local minimum less frequently, especially in the case of high multicollinearity of the component matrices. The usefulness of smoothness constraints are examined in a simulation experiment and an empirical example.

4. Comparisons of the restricted with unrestricted Tucker3 and CP

To test the usefulness of smoothing in Tucker3 and CP, we performed a simulation study on the basis of 960 data sets for the Tucker3 model, and 480 for the CP model. It is examined to what extent an underlying structure is recovered by unconstrained and smoothness constrained CP or Tucker3 analyses from data with smooth structure. Additionally, the computational properties of the constrained and unconstrained models are examined. The algorithms were programmed in MATLAB (2001), and the analyses were carried out on a Pentium 333 Mhz 256 Mb RAM personal computer in a Windows 98 environment.

4.1. Construction of the data for the simulation study

4.1.1. CP data for the simulation study

For the CP simulation study, 480 data sets were constructed with known CP structure with smooth components in one mode, and various data sizes, numbers of components, degrees of multicollinearity in $\mathbf{A}, \mathbf{B}$, and $\mathbf{C}$, and error levels. The data matrices $\mathbf{X_c}$ $(K \times IJ)$ were constructed according to

$$\mathbf{X_c} = \mathbf{C}_o\mathbf{H}(\mathbf{B}_o{}' \otimes \mathbf{A}_o{}') + \varepsilon\mathbf{N_c}, \tag{12}$$

where $\mathbf{A}_o$ $(I \times Q)$, $\mathbf{B}_o$ $(J \times Q)$, and $\mathbf{C}_o$ $(K \times Q)$ are 'true' component matrices for the respective three-modes, $\mathbf{H}$ $(Q \times Q^2)$ is the matrix version of the superdiagonal three-way array $\underline{\mathbf{H}}$, ε is a multiplication coefficient, and $\mathbf{N_c}$ $(K \times IJ)$ denotes the matrix expression of the three-way error array $\underline{\mathbf{N}}$ $(I \times J \times K)$.

The data sizes I, J, K were 10,10,20; 10,10,50; 10,50,20 and 10,50,50. The numbers of components were two and four. The elements of the matrices $\mathbf{A}_o$ and $\mathbf{B}_o$ were drawn randomly from the uniform [0,1] distribution (mild multicollinearity condition), and from the uniform [0.5,1.5] distribution (severe multicollinearity condition). Every component of $\mathbf{C}_o$ followed a smooth function evaluated at K equidistant points (to be denoted by $t_1, \ldots, t_K$). Half of the components of $\mathbf{C}_o$ followed the first order derivative to the growth parameter of an exponential function plus a constant, and half of the components followed a logistic function plus a constant. The constant was added to manipulate the condition number of $\mathbf{C}_o$. The parameters were varied so that in the mild multicollinearity condition of $\mathbf{C}_o$ the condition numbers for two and four components of $\mathbf{C}_o$ were 2 and 6, respectively, whereas in the severe multicollinearity condition the condition numbers were 6 and 42. The values of $\mathbf{N_c}$ were drawn randomly from the standard normal distribution and multiplied by a scalar ε, which influences the variance of the distribution of the error part. The scalar ε was chosen so that the expected percentages of error sums of squares in $\underline{\mathbf{X}}$ were 2%, 26%, or 50%. The number of replications was five. The design was fully crossed, leading to a total of four (data sizes) × two (numbers of components) × two (degrees of multicollinearity of $\mathbf{A}_o$ and $\mathbf{B}_o$)× two (degrees of multicollinearity of $\mathbf{C}_o$)× three (error levels) × five (replications) $=480$ matrices.

4.1.2. Tucker3 data for the simulation study

For the Tucker3 simulation study, 960 data sets were constructed with known Tucker3 model structure with smooth components in one mode, and various data sizes, numbers of components, degrees of multicollinearity in the core, and error levels. The data matrices were constructed as

$$\mathbf{X_c} = \mathbf{C}_o\mathbf{G}_o(\mathbf{B}_o{}' \otimes \mathbf{A}_o{}') + \varepsilon\mathbf{N_c}, \tag{13}$$

where $\mathbf{A}_o$ $(I \times P)$, $\mathbf{B}_o$ $(J \times Q)$, and $\mathbf{C}_o$ $(K \times R)$ are 'true' component matrices for the respective three modes, $\mathbf{G}_o(R \times PQ)$ is the matricised version of the three-way core array $\underline{\mathbf{G}}_o$ (the subscript '**c**' is omitted for notational simplicity), ε is a multiplication coefficient, and $\mathbf{N_c}$ $(K \times IJ)$ denotes the matrix expression of the three-way error array $\underline{\mathbf{N}}$.

The data sizes of the data array $\underline{\mathbf{X}}$ I,J,K were 10,10,20; 10,20,20; 10,10,50; 10,50,20; 10,20,50; 30,20,20; 10,50,50 and 30,20,50. The numbers of components P,Q,R for the three modes were 2,2,2; 2,4,2; 2,2,4 and 4,4,4. The component matrices $\mathbf{A}_o, \mathbf{B}_o$, and $\mathbf{C}_o$ were chosen column-wise orthonormal. The components of the smooth $\mathbf{C}_o$ followed the same functions as in the CP simulation study, except for the fact that orthonormal bases of the matrices concerned were used. The matrices $\mathbf{A}_o$ and $\mathbf{B}_o$ were obtained by taking the orthonormal bases of a matrix with equal size as $\mathbf{A}_o$ and $\mathbf{B}_o$ with elements drawn randomly from the uniform [0,1] distribution. These choices do not place severe limitations on the simulation study, since the component matrices in the Tucker3 solution, and hence any set of 'true' component matrices of a Tucker3 model in a simulation study, can be transformed to orthonormality, provided that this transformation is compensated in the core. However, transformation of a multicollinear true component matrix to orthonormality and compensation for this in the core array would lead to a multicollinear core. For example, suppose we have a matrix $\mathbf{C}$ and $\mathbf{G}$, where $\text{cond}(\mathbf{C}) = 100$, and $\mathbf{G}$ is row-wise orthonormal so that $\text{cond}(\mathbf{G})$ is 1, where cond( ) means the condition number. Orthonormalization of $\mathbf{C}$ into $\tilde{\mathbf{C}}$, and compensation for the orthonormalization in $\mathbf{G}$ by transforming $\mathbf{G}$ into $\tilde{\mathbf{G}}$ results in $\text{cond}(\tilde{\mathbf{C}}) = 1$. This can be achieved by taking the QR-decomposition of $\mathbf{C} = \mathbf{QR}$, defining $\tilde{\mathbf{C}} = \mathbf{Q} = \mathbf{CR}^{-1}$, and $\tilde{\mathbf{G}} = \mathbf{RG}$, and, as a result, $\text{cond}(\tilde{\mathbf{G}}) = \text{cond}(\mathbf{RG}) = 100$. Therefore, to represent a reasonable range of possible data matrices, the degree of multicollinearity of the core is varied in this study. The elements of $\mathbf{G}_o$ were drawn randomly from the uniform [0,1] distribution in the low multicollinearity condition, and from the uniform [0.5,1.5] distribution in the high multicollinearity condition. The error level was varied in the same way as in the CP simulation study, that is the expected percentages of error sum of squares of $\underline{\mathbf{X}}$ were 2%, 26%, and 50%. The number of replications in each condition was five. The design was fully crossed, leading to a total of eight (data sizes) × four (numbers of components) × two (degrees of multicollinearity of $\mathbf{G}_o$)× three (error levels) × five (replications) =960 matrices.

4.2. Analyses of simulation data

The simulated data sets $\mathbf{X_c}$ were all analysed by one unconstrained CP or Tucker3 analysis, and by two CP or Tucker3 analyses with smoothness constraints. Specifically, in the analyses with smoothness constraints, the estimated component matrix $\mathbf{C}$ was restricted to be in the column space of a set of B-splines $\mathbf{B^s}$ of degree three. The knots were equidistantly placed on the time interval $t_1,\dots,t_K$, with a knot at t_1 and one at t_K. The CP or Tucker3 analyses with smoothness constraints were performed on the compressed data array (see (11)) instead of the full data array to reduce computation time. In one of the analyses with smoothness constraints, the number of knots was chosen so that the sum of the cross-validation sum of squares, CV(λ), see (4), over columns of $\mathbf{X_c}$ was minimised. That is, for a fixed number of knots, the CV(λ) was computed for each column of $\mathbf{X_c}$, and then the sum of the CV(λ)'s obtained in this way was computed. The sum of the CV(λ)'s was computed successively for solutions based on $2,3,\ldots,t_K$ knots, and the number of knots that goes with the minimal sum of CV(λ)'s was chosen. The CP and Tucker3 analyses with these restrictions are referred to as CP-Bs(CV) and Tucker3-Bs(CV), respectively. In the other analysis with smoothness constraints, the number B-spline knots was fixed at three. This number is somewhat arbitrary, although we chose deliberately a small number of knots to prevent overfitting. The CP and Tucker3 analysis with this restriction are referred to as CP-Bs(3) and Tucker3-Bs(3), respectively. The CP algorithm of Harshman (1970), and Carroll and Chang (1970) was used to fit each CP model to data. Each Tucker3 analysis was performed using the efficient algorithm by Andersson and Bro (1998). The CP and Tucker3 algorithms were run from five different starts, one started rationally and four randomly, to reduce the chance of missing the global minimum for each analysis. The rationally started runs were started with the parameters resulting from Tucker's Method I (Tucker, 1966). The convergence criterion was set at 10^{-6}.

4.3. Criteria of interest

The main interest in this study was how well the original component matrices (and core array in the case of the Tucker3 model) were recovered by each of the methods. Different comparison criteria are used for the CP and the Tucker3 analyses, due to the disparity of transformational freedom.

4.3.1. CP analyses: criteria of interest

In the CP analyses, comparing the estimated component matrices and the original component matrices has to take into account possible permutations, rescalings, and sign reversions of the estimated component matrices. Following Kiers (1998), and Mitchell and Burdick (1994), we compared the CP solutions by computing the cosines between the tensor products $\mathbf{a}_r^{\mathrm{o}} \otimes \mathbf{b}_r^{\mathrm{o}} \otimes \mathbf{c}_r^{\mathrm{o}},\ r=1,\ldots,R$, for the original component matrices and, $\hat{\mathbf{a}}_r \otimes \hat{\mathbf{b}}_r \otimes \hat{\mathbf{c}}_r,\ r=1,\ldots,R$, for the estimated component matrices, where the subscript r denotes the rth column of the matrix at hand. Given a data array $\underline{\mathbf{X}}$ that is represented by a set of R tensor products of components, which are collected in component matrices $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$, other sets of component matrices that yield the same representation of $\underline{\mathbf{X}}$ are constituted of the same such tensor products, although possibly in a different order. Therefore, a useful comparison measure of the original and the estimated component matrices is the mean of the R cosines between the tensor products of the original components and the tensor products of the estimated components, with the latter tensor products ordered such that they lead to the highest mean of cosines. The cosines are computed as Tucker's coefficient of congruence (Tucker, 1951), and the highest mean of cosines is denoted by φ.

One rationally started and four randomly started runs of the CP analysis were carried out. The runs which led to a sub-optimal solution (defined here as a solution with a function value higher than 1.001 times the fit of the optimal solution, out of the five runs) were counted to get an impression of the sensitivity to local minima of the constrained and unconstrained analyses. The number of iterations and the computing time per analysis were recorded to get an idea of the computational complexity.

4.3.2. *Tucker3 analyses: criteria of interest*

To investigate how well the original matrices of the Tucker3 model are recovered, the recovery of both the column spaces of the component matrices, and the interaction weights of the components are of importance. The column spaces of the component matrices are compared as follows: a comparison of the estimated to the underlying component matrices has to take into account the fact that the component matrices can be transformed without loss of fit, provided that such transformations are compensated in the core. Therefore, the estimated component matrices $\hat{\mathbf{A}}$, $\hat{\mathbf{B}}$, and $\hat{\mathbf{C}}$ are transformed towards the original component matrices $\mathbf{A}_\mathrm{o}$, $\mathbf{B}_\mathrm{o}$, and $\mathbf{C}_\mathrm{o}$ by postmultiplying $\hat{\mathbf{A}}, \hat{\mathbf{B}}$, and $\hat{\mathbf{C}}$ by the matrices $\mathbf{S}$, $\mathbf{T}$, and $\mathbf{V}$, espectively. The transformation matrices $\mathbf{S}$, $\mathbf{T}$, and $\mathbf{V}$ are found by minimising the Euclidean distance between the original component matrices $\mathbf{A}_\mathrm{o}, \mathbf{B}_\mathrm{o}$, and $\mathbf{C}_\mathrm{o}$ and the transformed component matrices $\hat{\mathbf{A}}\mathbf{S}$, $\hat{\mathbf{B}}\mathbf{T}$, and $\hat{\mathbf{C}}\mathbf{V}$, respectively. The transformations are compensated in the estimated core matrix $\hat{\mathbf{G}}$ by computing the transformed core array $\tilde{\mathbf{G}} = \mathbf{V}^{-1}\hat{\mathbf{G}}((\mathbf{T}')^{-1} \otimes (\mathbf{S}')^{-1})$. The component matrices $\hat{\mathbf{A}}\mathbf{S}, \hat{\mathbf{B}}\mathbf{T}$, and $\hat{\mathbf{C}}\mathbf{V}$ are compared to the original component matrices $\mathbf{A}_\mathrm{o}, \mathbf{B}_\mathrm{o}$, and $\mathbf{C}_\mathrm{o}$ by computing the proportion of agreement ($\mathrm{PA}_\mathbf{A}$, $\mathrm{PA}_\mathbf{B}$, and $\mathrm{PA}_\mathbf{C}$, respectively) as

$$\mathrm{PA}_\mathbf{A} = 1 - \frac{\|\mathbf{A}_\mathrm{o} - \hat{\mathbf{A}}\mathbf{S}\|^2}{\|\mathbf{A}_\mathrm{o}\|^2}, \qquad \mathrm{PA}_\mathbf{B} = 1 - \frac{\|\mathbf{B}_\mathrm{o} - \hat{\mathbf{B}}\mathbf{T}\|^2}{\|\mathbf{B}_\mathrm{o}\|^2},$$

$$\mathrm{PA}_\mathbf{C} = 1 - \frac{\|\mathbf{C}_\mathrm{o} - \hat{\mathbf{C}}\mathbf{V}\|^2}{\|\mathbf{C}_\mathrm{o}\|^2}. \tag{14}$$

The average of $\mathrm{PA}_\mathbf{A}, \mathrm{PA}_\mathbf{B}$, and $\mathrm{PA}_\mathbf{C}$, denoted as $\mathrm{PA}_\mathbf{ABC}$, is used as the measure of agreement between the original and the estimated component matrices.

The recovery of the interaction weigths of the components is examined by comparing the transformed core matrix $\tilde{\mathbf{G}}$ to the original core matrix $\mathbf{G}_\mathrm{o}$ via the proportion of agreement ($\mathrm{PA}_\mathbf{G}$):

$$\mathrm{PA}_\mathbf{G} = 1 - \frac{\|\mathbf{G}_\mathrm{o} - \tilde{\mathbf{G}}_\mathbf{c}\|^2}{\|\mathbf{G}_\mathrm{o}\|^2}. \tag{15}$$

Note that the transformed component matrices are optimally transformed toward the original component matrices, whereas the associated core matrix is not optimally transformed towards the original core matrix. Hence, it can be expected that the $\mathrm{PA}_\mathbf{G}$ is smaller than the $\mathrm{PA}_\mathbf{ABC}$ in the case of a Tucker3 solution deviating from the original matrices. The number of iterations and the computing time per analysis were recorded as a measure of computational complexity.

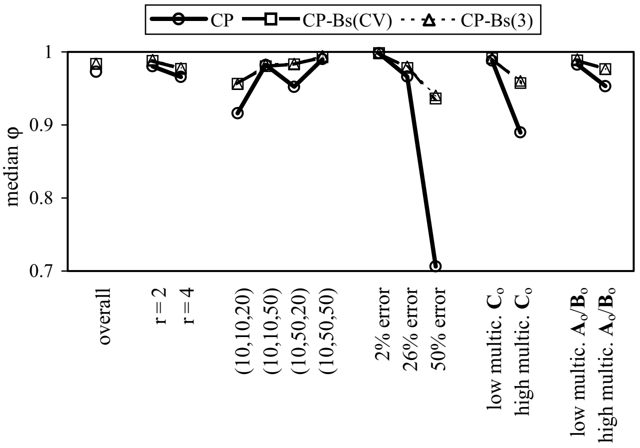

Fig. 1. Median φ-values of CP, CP-Bs(CV) and CP-Bs(3) per condition. ‘Multic.’ denotes ‘multicollinearity’.

4.4. Results of the simulation experiments

4.4.1. Results of the CP simulation experiment

The original component matrices and the estimated component matrices, as obtained by unconstrained CP analysis (CP) and CP with smoothness constraints (CP-Bs(CV) and CP-Bs(3)), are compared by inspecting the φ-values. Recall that the φ-value is the highest mean of cosines between the tensor products of the original components and the tensor products of the estimated components. The φ-values have a negatively skewed distribution over the replications within each condition. The median φ-values of the three analysis methods are plotted overall as well as per main condition in Fig. 1.

The following observations can be made in Fig. 1. The median φ-value of the constrained CP solutions is larger than the median φ-value of the unconstrained CP solutions, whereas virtually no difference was found between the median φ-value of CP-Bs(CV) and CP-Bs(3). Furthermore, the difference between the unconstrained and the constrained CP solutions gets clearly larger with increasing condition numbers of $\mathbf{C}_o$, and with increasing error level, and varies in a more complicated manner with data size (see Fig. 1).

A repeated measurement ANOVA was performed to test whether the observed effects of type of analysis and of the interactions of analysis method with the various manipulated factors could be distinguished from random fluctuations. To correct for the deviation from normality for the repeated measurement ANOVA, the φ-values were transformed into $\tilde{\varphi} = \log(\varphi/(1-\varphi))$ before analysis, where the two observed negative φ-values were excluded from the analysis. The transformation of negatively skewed φ-values on the interval [0,1] results in approximately normally distributed $\tilde{\varphi}$-values on the interval $[-\infty, \infty]$. The effects which were described in the previous paragraph, were all found to be significant at $\alpha = 0.001$ in the repeated measurement ANOVA of the $\tilde{\varphi}$-values.

Table 1
Frequencies of good ($\varphi \geqslant 0.75$) and bad ($\varphi < 0.75$) solution per analysis method (CP with CP-Bs(CV) and CP with CP-Bs(3))

		CP-Bs(CV)		CP-Bs(3)	
		Good	Bad	Good	Bad
CP	Good	348	3	348	3
	Bad	43	86	40	89

In addition to the φ-values of the three analysis methods in the different conditions, the number of cases in which the unconstrained CP leads to a ‘good’ solution, and the constrained CP to a ‘bad’ solution is of interest. On the basis of inspection of a number of plots of original and estimated components and the accompanying φ-value, solutions with a φ-value smaller than 0.75 were considered to be bad. The resulting frequencies according to this criterion are presented in Table 1.

In a large number of cases, both the constrained and the unconstrained CP model lead to a good solution. The most important finding is that if the unconstrained CP leads to a bad solution, the constrained CP model leads to a good solution in about 33% of the cases. Furthermore, it is rarely found that the unconstrained CP model leads to a good solution and the constrained CP model to a bad solution. The proportion of bad solutions, as well as the differences between the constrained and unconstrained CP model increases with error level, and condition number of $\mathbf{C}_0$. Thus, on the basis of these results we can conclude that, if there is a smooth underlying structure, B-spline constrained CP is helpful in a fair number of cases, and that there is very little risk in replacing unconstrained CP by CP with smoothness constraints. Moreover, the choice for the number of knots does not seem crucial.

Differences between the constrained and unconstrained CP analyses in sensitivity to local minima were also studied. The constrained CP analyses led to a sub-optimal solution a little less frequently (both 0.10 out of five starts on average) than the unconstrained analyses (0.16 on average). No difference in average numbers of local minima has been found between the rationally and the randomly started runs. The number of local minima increased with increasing error level, whereas no substantial interaction between any other of the manipulated factors and type of analysis was found.

The number of iterations and computing time of the optimal solution of the constrained and unconstrained CP analyses were inspected to get an idea of the computational complexity. The average computing time per analyses of the unconstrained CP analysis (2.11 s) was much longer than of the constrained CP analyses (0.65 s for CP-Bs(CV) and 0.48 s CP-Bs(3)). The most relevant independent variable related to computing time is the data size, where larger data sizes lead to longer computing time. The observed interaction between type of analysis and data size suggests that, not surprisingly, smoothness constrained CP analyses are particularly useful to reduce computing time in the case of a large number of elements in the occasion mode. For example, the mean computing time for data size 10,50,50 is 5.45 s. for CP analysis

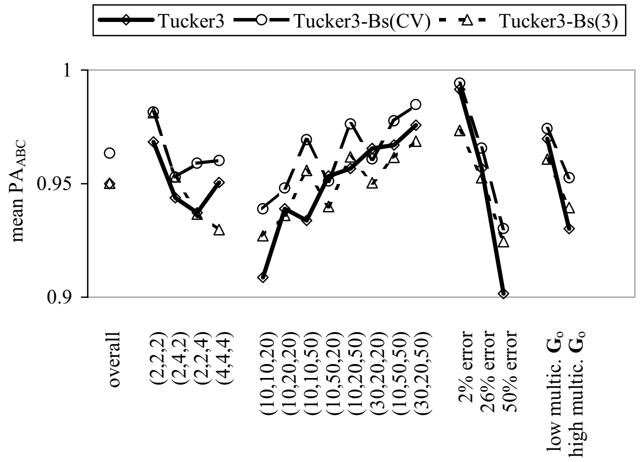

Fig. 2. Mean $\mathrm{PA}_{\mathbf{ABC}}$ of Tucker3, Tucker3-Bs(CV) and Tucker3-Bs(3) per condition. 'Multic.' denotes 'multicollinearity'.

and 0.87 for CP-Bs(CV) and 0.78 s CP-Bs(3). The average number of iterations of the CP analyses (130) was much lower than of the CP-Bs(CV) (200) and CP-Bs(3) (219). The differences in mean number of iterations between the unconstrained CP analysis and the smoothness constrained CP analysis gets larger with increasing condition number (both of $\mathbf{C}_0$ and of $\mathbf{A}_0/\mathbf{B}_0$), and increasing number of components, but is not influenced by error level or data size. All effects reported in this section appeared to be significant ($p < 0.01$) in the repeated measurement ANOVA of the number of iterations and of the computing time.

4.5. Results of the Tucker3 simulation experiment

The original component matrices and the original core matrix were compared to the estimated component matrices and the estimated core matrix by means of the proportion of agreement of the component matrices and the core matrix, the $\mathrm{PA}_{\mathbf{ABC}}$, which is based on the average of the expressions in (14), and the $\mathrm{PA}_{\mathbf{G}}$, (15), respectively. The average $\mathrm{PA}_{\mathbf{ABC}}$ values per analysis method give a good impression of the condition effects, and they are plotted per condition in Fig. 2.

As can be seen in Fig. 2, the $\mathrm{PA}_{\mathbf{ABC}}$ of Tucker3-Bs(CV) is generally higher than that of Tucker3 and Tucker3-Bs(3), whereas almost no difference was found between the $\mathrm{PA}_{\mathbf{ABC}}$ of Tucker3 and Tucker3-Bs(3), over all conditions. The difference in $\mathrm{PA}_{\mathbf{ABC}}$ between the three methods of analysis increases with increasing core size, error percentage, and degree of multicollinearity of the core, and varies with data size. The gain of the smooth Tucker3 over the unconstrained Tucker3 is largest in the case of a relatively large size of the smooth mode, and relatively small sizes of the non-smooth modes, for example, data size 10,10,50. If the size of the smooth mode is smaller than the size of one of the non-smooth modes, the performance of Tucker3 is better than

Table 2
Frequencies of good ($\mathrm{PA}_{\mathbf{ABC}} > 0.9$ and $\mathrm{PA}_{\mathbf{G}} > 0.9$) and bad solution per analysis method (Tucker3 with Tucker3-BS(CV) and Tucker3 with Tucker3-BS(3))

		Tucker3-Bs(CV)		Tucker3-Bs(3)	
		Good	Bad	Good	Bad
Tucker3	Good	811	7	808	10
	Bad	63	79	62	80

that of the smooth Tucker3 (e.g., data sizes 30,20,20 and 10,50,20). Tucker3 clearly outperforms Tucker3-Bs(3) in the case of low error level (2%) and low multicollinearity of the core, whereas Tucker3-Bs(CV) performs best of the three. In high error level and high multicollinearity conditions, Tucker3-Bs(3) performs better than Tucker3, but Tucker3-Bs(CV) gives best recovery of the component matrices. This finding suggests that the smoothness restricted Tucker3 is sensitive to the choice of number of knots, and that in ‘easy conditions’ an unconstrained Tucker3 model performs even better than a smoothness constrained Tucker3 model with a non-optimal number of knots.

A repeated measurement ANOVA was performed to test whether the observed effects of type of analysis and of the interactions of analysis method with the various manipulated factors could be distinguished from random fluctuations. For the repeated measurement ANOVA, the $\mathrm{PA}_{\mathbf{ABC}}$-values were transformed to correct for the observed heterogeneity of variances for the groups by computing $\tilde{\mathrm{P}}\tilde{\mathrm{A}}_{\mathbf{ABC}} = \arcsin(\mathrm{PA}_{\mathbf{ABC}})^{1/2}$ (Stevens, 1992). The effects that were explicitly described in the previous paragraph, were found to be significant at $\alpha = 0.001$ in the repeated measurement ANOVA of the $\tilde{\mathrm{P}}\tilde{\mathrm{A}}_{\mathbf{ABC}}$-values.

The estimated component matrices $\hat{\mathbf{A}}$, $\hat{\mathbf{B}}$, and $\hat{\mathbf{C}}$ are optimally transformed to the original component matrices $\mathbf{A}_{\mathrm{o}}, \mathbf{B}_{\mathrm{o}}$, and $\mathbf{C}_{\mathrm{o}}$, whereas the transformation of the estimated core matrix $\hat{\mathbf{G}}$ is so that the transformations of the original component matrices are compensated. Therefore, a non-optimal recovery will be expressed in a low $\mathrm{PA}_{\mathbf{G}}$ value, and possibly in a low $\mathrm{PA}_{\mathbf{ABC}}$ value. The $\mathrm{PA}_{\mathbf{G}}$ values appear highly negatively skewed, with some extremely low values, hence the median $\mathrm{PA}_{\mathbf{G}}$ values give a better insight into the condition effects than the mean $\mathrm{PA}_{\mathbf{G}}$ values. The median $\mathrm{PA}_{\mathbf{G}}$ values per condition appeared to be high (> 0.985), and they hardly differ from each other, neither between type of analysis nor between conditions. The extremely low values all occurred in the ‘more difficult’ conditions, namely large core size, small data size, high condition number of the core $\mathbf{G}_{\mathrm{o}}$, and high error level. The Tucker3 analysis showed more extremely low $\mathrm{PA}_{\mathbf{G}}$ values than the Tucker3-Bs(CV) and Tucker3-Bs(3) analyses, as is indicated by, for example, the percentages of the cases with $\mathrm{PA}_{\mathbf{G}}$ values lower than 0.5 of 4.7%, 1.1% and 1.3%, respectively.

A second way of comparing the achievement of the three methods of analysis is to inspect the number of cases that are recovered well by the different methods. On the basis of inspection of a number of original and estimated components, cores and associated $\mathrm{PA}_{\mathbf{ABC}}$ and $\mathrm{PA}_{\mathbf{G}}$ solutions with a $\mathrm{PA}_{\mathbf{ABC}}$ or a $\mathrm{PA}_{\mathbf{G}}$ smaller than 0.9 were considered to be bad. The resulting frequencies of good and bad solutions are presented in Table 2. It can be seen in this table that if Tucker3 leads to a bad solution, Tucker3-Bs(CV) leads to a good solution in 44% of the cases. In only 1% of the cases, the Tucker3-Bs(CV) is bad, while the Tucker3 solution is good. According to the frequencies in Table 2, Tucker3-Bs(3) performs almost as well as Tucker3-Bs(CV).

Although the number of solutions that were reasonably recovered by Tucker3-Bs(3) does not deviate much from the number of reasonable recoveries using Tucker3-Bs(CV), the smoothing technique is sensitive to the choice for the number of knots, as indicated by the better recovery of the underlying component structure by Tucker3-Bs(CV) than of Tucker3-Bs(3). Thus, we can conclude on the basis of these results, that if there is a smooth underlying structure, a smoothness constrained Tucker3 model is helpful in a reasonable number of cases, and that, conversely, there is very little risk in using smoothness constrained instead of unconstrained Tucker3. Tucker3-Bs(CV), the method with optimal knot selection, performed best, and is therefore preferable to Tucker3-Bs(3).

The computational complexity of the three Tucker3 analysis types was evaluated via the number of iterations and computing time of the optimal solution. The average computing time per analyses of the unconstrained Tucker3 analysis (0.05 s) was longer than of the Tucker3-Bs(CV) and Tucker3-Bs(3) analyses (both 0.01 s). The average number of iterations of the unconstrained Tucker3 analysis (8) is somewhat higher than of the smoothness constrained Tucker3 analyses (both 5). Both the average computing time and the number of iterations increases with increasing error level, core size, condition number of the core and the core size. The effects of the independent variables on the average computing time and the number of iterations are larger for the unconstrained Tucker3 analyses than for the constrained ones, and hence an interaction is observed between type of analysis and each of the independent variables. All effects reported in this section appeared to be significant ($p < 0.01$) in the repeated measurement ANOVA of the number of iterations and of the computing time.

5. Example: learning to read study

In this section, an empirical example is presented to illustrate the use of a smoothness constrained Tucker3 model, and monotonicity constrainted data in an unconstrained Tucker3 model. The degree of overall error of the models is investigated and compared by using cross-validation.

The learning to read study (Bus and Kroonenberg, 1982) investigates the learning process of reading. Seven pupils were tested weekly (except for holidays) on 37 occasions by means of five different tests, which intended to measure different aspects of reading ability. The primary research questions were focused on whether the development of the pupils per test and over tests was equal over time.

Before analysis, the raw scores were rescaled so that the scores of the five tests ranged from zero to one. As a result, the scores are comparable between tests, while all the differences in variation were maintained in the data. The rescaled scores were collected in the data array $\underline{\mathbf{Y}}$ ($7 \times 5 \times 37$) and analysed by the unconstrained Tucker3 model.

Table 3
Subject component scores of the unconstrained Tucker3 solution

A (subjects)	First component	Second component
1	1.06	−0.42
2	0.96	−0.30
3	0.99	−0.38
4	1.28	1.00
5	1.16	0.19
6	1.09	−0.01
7	0.89	−0.42

The scores are viewed as evaluations of growth curves, which are assumed to follow some smooth curves in the course of time. Therefore, in the second analysis, the scores on the components of the occasion mode (**C**) are constrained to follow smooth curves. A smoothness constrained Tucker3 model (T3-Bs) is fitted to $\mathbf{Y_c}$ by minimising (11), which is equivalent to minimising (6). The degree of the B-spline was fixed at three. The knots were placed equidistantly, and their number was chosen such that the sum of cross-validation sum of squares (CV(λ), see (4)) of the columns of $\mathbf{Y_c}$ was minimised, by computing the sum of CV(λ)'s related to B-splines with $2,3,\ldots,10$ knots, and choosing the number of knots that goes with the minimal sum of CV(λ)'s.

Because the data pertain to learning data, it might be reasonable to assume that the true scores per variable per subject are non-decreasing on subsequent occasions, assuming that the reading ability of the child never decreases in the course of time. To model non-decreasing true scores, a smoothed data matrix $\tilde{\mathbf{Y}}_\mathbf{c} = \mathbf{B^iW}$ is obtained by minimising $\|\mathbf{Y_c} - \mathbf{B^iW}\|^2$, where $\mathbf{B^i}$ is an I-spline matrix and $\mathbf{W}$ the weights for the I-splines that are restricted to non-negativity, and as a result $\tilde{\mathbf{Y}}_\mathbf{c}$ is restricted in the sense that $\tilde{y}_{ijk} \leqslant \tilde{y}_{ij(k+1)}$, for all $i = 1,\ldots,7$; $j = 1,\ldots,5$, and $k = 1,\ldots,36$. An unconstrained Tucker3 analysis is applied to these smoothed data. This analysis will be referred to as T3-Bi. The degree of the I-spline matrix was fixed at two. The number of knots was selected by subjective comparison of the observed and several estimated response variables.

For all three analyses, the numbers of components were chosen to be 2,1, and 2 for the subject, variable and occasion modes, respectively. As will be explained, the latter model is stable, parsimonious, fits the data well, and the solution is well interpretable.

The fit of the unconstrained Tucker3 model is 96.26%. The estimated core matrix $\mathbf{G}_\mathrm{a}$ of the model positioned in principal axes orientation was diagonal. The core matrix was transformed to identity, and this rescaling was compensated in the subject component matrix. The columns of the component matrices were rescaled such that the solution was easy to interpret (for example, the maximum component value of the variables was rescaled to 1). The component matrices for the subjects (**A**) and the variables (**B**) of the unconstrained Tucker3 of $\underline{\mathbf{Y}}$ are presented in Tables 3 and 4. The occasions component scores are plotted in Fig. 3.

To interpret the component scores, we start with the component scores for the occasion mode. In Fig. 3, it can be seen that the component scores of the first component

Table 4
Variable component scores of the unconstrained Tucker3 solution

B (variables)	
Letter knowledge	0.91
Regular orthographic short words	1.00
Regular orthographic long words	0.87
Regular orthographic long and short words within context	0.99
Irregular orthographic long and short words	0.58

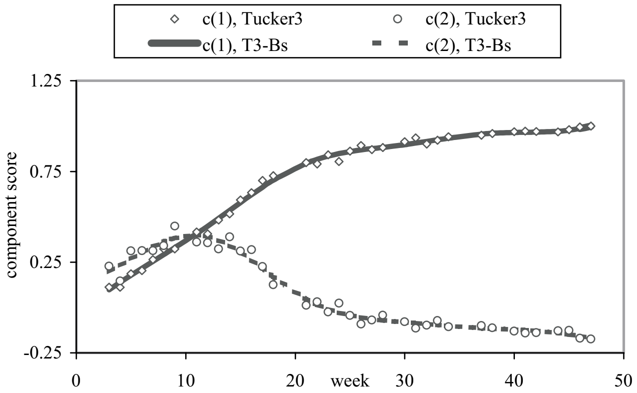

Fig. 3. Component scores for occasions resulting from the unconstrained Tucker3 model (denoted by ‘Tucker3’) and the T3-Bs model; c(1) denotes the scores on the first component, c(2) the scores on the second component.

(indicated by diamonds) increase gradually from week 3 to 20, and then levels off to an asymptote of one. The scores on the second component (indicated by circles) show a steady increase to week 10, a steady decrease from week 10 to 20, and then levels off to slightly below zero. We would interpret the first component as indicating general performance level, and the second component as approximately reflecting learning rate. The latter component is not entirely interpretable as learning rate, because of the negative component scores, which are due to the estimated model parameters, and does not indicate that the performance decreases in the end.

The core matrix is identity, implying that the first component of the subject component matrix (**A**) is only related to the first component of the occasion component matrix (**C**), and the same holds for the second component of **A** and **C**; moreover the combinations of the first and second components are equally weighted.

Now, the subject component matrices can be interpreted. Recall that the general performance level of a subject (hence apart from specific variable effects) is a weighted sum of the two occasion components, which reflect general performance level and learning rate. A relatively high score on the first component means that the subject concerned performs above the general performance level. A relatively high score on the second component implies that the subject shows a relatively fast learning rate.

Thus, for example, Subject 4 performs by far best, as (s)he has a high performance level and a high learning rate. Subject 4 is followed at some distance by Subject 5, as (s)he has a second position for performance level, and learning rate. Subjects 1 and 6 show approximately the same weighting for general performance level, but Subject 1 has a lower weight for the second component. Hence, their asymptote scores are more or less equal, but Subject 1 develops much more slowly than Subject 6. The performance order between Subjects 1 and 3 is somewhat difficult to see at once, as the performance level of a subject is a weighted sum of the two occasion components, and the weights are close to each other. One could plot the weighted occasion component scores for the three subjects, and this would reveal that Subject 1 performs best of the three, and Subject 2 worst.

As there is only one variable component, the relative sizes of the variable component scores denote the difficulties of the items. The variable 'Irregular Orthographic Long and Short words' is by far the most difficult variable, as indicated by the lowest variable component score. Hence, the scores on 'Irregular Orthographic Long and Short Words' develop slowly in the course of time, compared to the other variables. The variable component scores of the 'Regular Orthographic Short Words' and the 'Regular Orthographic Long and Short Words within Context' are the highest variable component scores, showing that these scores develop fastest in the course of time. The variable component score of the 'Letter Knowledge' is slightly larger, and thus develops slightly faster, than the 'Regular Orthographic Long Words'.

The stability of the model just discussed was investigated via a split-half analysis, following the guidelines by Kiers and Van Mechelen (2001). That is, the data were split into two halves over the subject mode, resulting in one data set of three subjects, and one of four subjects, to be denoted as $\underline{\mathbf{Y}}_1$ and $\underline{\mathbf{Y}}_2$. A Tucker3 analysis was performed for each of the two data sets. The solutions for **B** and **C** for each of the data sets were optimally transformed (in the least squares sense) to the solutions for **B** and **C** of the full data set (as presented in Table 4 and Fig. 3, respectively). The two transformed occasion component matrices obtained in this way were compared by computing the coefficients of congruence (Tucker, 1951; see also Section 4.3) between the columns of the matrices. The two variable component matrices were compared analogously. The subject component matrices were compared as follows: the two solutions for **A** for each of the splits were collected in one matrix $\mathbf{A}_{12}$, in which the rows pertain to the same subjects as in **A** of the full data set (as presented in Table 3). The matrix $\mathbf{A}_{12}$ was regressed on **A**, and the resulting transformed component matrix was compared to **A** by computing the coefficients of congruence between the two columns of the matrices. For each of the splits, the transformations of $\mathbf{A}, \mathbf{B}$ and **C** were compensated in the core array. The separate split-half core arrays and the full data set core array were compared by computing the mean absolute difference between the split-half core array and the full data set core array.

The split-half procedure was repeated for every possible combination of the seven subjects split into two groups of three and four subjects, resulting in 35 split-half analyses. The mean coefficients of congruence for the subject component matrices over the 35 analyses was 1.000 and 0.997. The mean coefficient of congruence for the variable components was 0.998. The mean coefficient of congruence for the occasion components was 0.996 and 0.713 for the first and second occasion component, respectively. This implies that the stability of the subject components, the variable component and the first occasion component is high, whereas the stability of the second occasion component is moderate. The mean absolute difference between the split-half core array and the full data set core array averaged over the 35 analyses was 0.000. On the basis of these results, we conclude that the current Tucker3 model is sufficiently stable, given the small sample size at hand. We now turn to the results of the smoothness constrained analyses (T3-Bs and the T3-Bi), which will be discussed successively.

In the smooth Tucker3 analysis with B-splines, the T3-Bs, an unconstrained Tucker3 analysis was performed on the smoothed data array. The number of knots as indicated by the (minimal) cross-validation sum of squares was five. (The CV values for two through seven knots were 0.0096, 0.0089, 0.0089, 0.0087, 0.0089, 0.0097, respectively.) The fit of this constrained model to the data array $\underline{\mathbf{Y}}$ was 96.18%, which is only 0.08% less than the fit of the unconstrained Tucker3 model. The estimated core matrix of the solution in principal axes orientation was diagonal, and was rescaled to identity. The component score matrices of the T3-Bs model were rescaled in the same way as was done with the unconstrained Tucker3 model. The estimated component matrices **A** and **B** of T3-Bs are compared to the solutions of the unconstrained Tucker3 model by computing the coefficient of congruence between the pairs of components concerned. This coefficient was large (> 0.999) for all pairs, and therefore the solutions of **A** and **B** for T3-Bs can be interpreted in the same way as the corresponding solutions for the unconstrained Tucker3. The component scores for the occasions for T3-Bs are plotted in Fig. 3 by lines. Not surprisingly, the component scores of the T3-Bs solution follow more or less the same curve as the ones of the unconstrained Tucker3 solution. However, the wiggles have disappeared, and the overall trend in the component scores of the occasions is more clear.

It is interesting to investigate whether the stability of the occasion component matrices of the T3-Bs has been improved on the unconstrained Tucker3 model. Additionally, it is important to check whether the subject and variable component matrices, and the core array of the T3-Bs model have a high stability, just as their counterparts in the Tucker3 model. The stability of the T3-Bs model was investigated using the split-half procedure, as discussed before for the unconstrained Tucker3 model. The mean coefficients of congruence of the occasion component matrices were 0.998 and 0.792, which is higher than the coefficients of congruence found in the unconstrained analyses of 0.996 and 0.713, respectively. The mean coefficients of congruence for the subject component matrices and the variable component matrices were equal to the ones found for the unconstrained Tucker3 model (1.000 and 0.997 for the subject component matrices, and 0.998 for the variable component matrices). Also, the mean absolute difference between the split-half core array and the full data set core array averaged over the 35 analyses was 0.000. On the basis of this results, one can conclude that the stability of the second occasion component of the T3-Bs model has indeed been improved somewhat compared to the unconstrained counterpart. The subject and variable component matrices, the first occasion component and the core array of the T3-Bs model are highly stable, just as their unconstrained counterparts.

In the T3-Bi analysis, the subsequent scores per variable and per subject were restricted to be non-decreasing in the course of time before analysis. The number of knots for the second degree I-spline matrix was chosen to be seven, on the basis of subjective comparison of the observed variables and several estimated response variables. The I-splines were defined on the interval from week 0 to 50. The fit of the resulting estimates of the Tucker3 model to the unconstrained data array $\underline{\mathbf{Y}}$ was 96.17%. The core matrix of the solution in principal axis direction was diagonal. The estimated core and component matrices were rescaled in the same way as in the Tucker3 and Tucker3-Bs models. The estimated solutions for T3-Bi of **A** and **B** were compared to the associated solutions for the unconstrained T3 by the coefficient of congruence. The coefficients were high (> 0.999) for all pairs concerned, and **A** and **B** are interpreted in the same way as **A** and **B** of the unconstrained Tucker3. The occasion component scores for T3-Bi resemble the occasion component scores for T3-Bs closely, as indicated by the coefficients of congruence (1.000 and 0.999, respectively). Therefore, our interpretation of this model is identical to the interpretation of the T3-Bs model.

6. Discussion and conclusion

The results from the Tucker3 and CP simulation experiments demonstrate that, if smooth underlying components are present, applying smoothness constraints in Tucker3 model and CP is generally useful to estimate the (underlying) components of the Tucker3 and CP model (and the core of the Tucker3 model) better. The gain in estimation accuracy of constrained estimation is more salient in the case of larger numbers of components, high condition numbers of the component matrices and high error levels.

In the simulation experiment, the smoothness constraints were imposed by requiring that the smooth component matrix lies in the column space of a B-spline matrix. The performance of the constrained Tucker3 model is considerably better if the number of knots of the B-splines was optimised according to the cross-validation criterion compared to the fixed knots choice (of 3 knots). Contrarily, the performance of CP does not appear to be influenced by the method for choosing the number of knots. This finding suggests that the performance of the smoothness constrained Tucker3 model is more sensitive to the choice of the number of knots than the smoothness constrained CP model. This might be due to the more constrained character of the CP model.

The smoothness constrained Tucker3 and CP models are estimated faster than their unconstrained counterparts. This is not surprising as the data array to be analysed is much smaller in the case of a smoothness constrained model.

The empirical example demonstrates an application of a smoothness constrained Tucker3 model, and such a model combined with monotonicity constraints. The subject and variable component matrices and the core of the constrained Tucker3 models are equally interpreted as the unconstrained Tucker3 model. The T3-Bi model, in which the analysed data are constrained to be non-decreasing in the course of time, appears to be reasonable for the data at hand. However, the monotonicity constraint additional to the smoothness constraint did not alter the interpretation of the solution at all, and therefore the simpler T3-Bs model can be preferred here. The interpretation of the time component scores of the smooth constrained T3-Bs solution is more clear than the unconstrained Tucker3 solution, as it is hard to judge whether certain wiggles in the plot of the time component scores of the unconstrained Tucker3 model should be considered important. Additionally, the stability of the T3-Bs solution is higher than of the unconstrained Tucker3 solution. Therefore, in the case of (presumed) smooth components, it appears to be useful to use smoothness constraints on the Tucker3 and CP model.

The commonly used procedures to estimate the Tucker3 model or CP require all elements of the three-way data box to be observed. In the case of data with a smooth mode, the use of the proposed procedures for smoothing the data can be helpful in estimating missing data elements. In longitudinal data, this procedure can be particularly useful if all measurements take place in the same time span, but at different sets of time points for different variables and/or occasions, where the missing data can be assumed to be missing completely at random (Little and Rubin, 1987). Note that the B-spline matrix $\mathbf{B}^{\mathrm{s}}$ $(K \times N)$ is a matrix with N B-splines which are evaluated in all values K of the predictor. In Section 3.3, the K measurements of the predictor, which simply represents the measurement times in the case of longitudinal data, were assumed to be equal for all i $(i = 1,\ldots,I)$, and j $(j = 1,\ldots,J)$, thus the B-spline matrix $\mathbf{B}^{\mathrm{s}}$ is equal for all i and all j, and (6) can be used. However, if there are different measurement occasions for different subjects and variables, hence for different (i,j) combinations, a B-spline matrix $\mathbf{B}^{\mathrm{s}}_{ij}$ must be defined for every combination of i and j. Now, provided that $\mathbf{B}^{\mathrm{s}}_{ij}$ is of full column rank, $\mathbf{x}_{ij}$ $(K_{ij} \times 1)$, the datavector containing the measurements of subject i on variable j at K_{ij} occasions, can be projected on $\mathbf{B}^{\mathrm{s}}_{ij}$ by minimising

$$f_6(\mathbf{w}_{ij})\|\mathbf{x}_{ij} - \mathbf{B}^{\mathrm{s}}_{ij}\mathbf{w}_{ij}\|^2. \tag{16}$$

The weights $\mathbf{w}_{ij}$ can be used to estimate $\hat{\mathbf{x}}_{ij}$ on the same time points for all i and j, namely by defining $\hat{\mathbf{x}}_{ij} = \mathbf{B}^{\mathrm{s}}\mathbf{w}_{ij}$. If the vectors $\hat{\mathbf{x}}_{ij}$ are collected in $\hat{\mathbf{X}}$ $(K \times IJ)$, $\hat{\mathbf{X}}$ can be analysed by unrestricted Tucker3 or CP procedures. Hence, the use of smoothness constraints in the Tucker3 model and CP is not only useful in enlarging the estimation accuracy, but also in dealing with data measured at unequal sets of time points.

Acknowledgements

This research has been made possible by funding from the Netherlands Organization of Scientific Research (NWO) to the first author. The authors are obliged to J.M.F. ten Berge, and F.J. Oort for useful comments on an earlier version of this paper, to P.H.C. Eilers for providing us some Matlab code and to A.G. Bus for putting her data set at our disposal.

References

Alsberg, B.K., Kvalheim, O.M., 1993. Compression of *n*th-order data arrays by B-splines. Part 1: theory. J. Chemometrics 7, 61–73.

Andersson, C.A., Bro, R., 1998. Improving the speed of multi-way algorithms: Part I. Tucker3. Chemometrics and Intelligent Laboratory Systems 42, 93–103.

Bro, R., 1998. Multi-way analysis in the food industry. Models, algorithms and applications. Unpublished Doctoral Thesis, University of Amsterdam, Amsterdam.

Bro, R., Smilde, A.K., 2001. Centering and scaling in component analysis. Submitted for publication.

Browne, M.W., Cudeck, R., 1992. Alternative ways of assessing model fit. Sociological Meth. Res. 21 (2), 230–258.

Bus, A.G., Kroonenberg, P.M., 1982. Reading instruction and learning to read: a longitudinal study. Internat. Report, SOL/82-08. Dept. of Education, University of Groningen, Groningen.

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

Carroll, J.D., Pruzansky, S., Kruskal, J.B., 1980. CANDELINC: a general approach to multidimensional analysis of many-way arrays with linear constraints on parameters. Psychometrika 1, 3–24.

De Boor, C., 1978. A Practical Guide to Splines. Springer, Berlin.

Golub, G.H., Van Loan, C.F., 1989. Matrix Computations. Johns Hopkins University Press, Baltimore, MD.

Harshman, R.A., 1970. Foundations of the PARAFAC procedure: models and conditions for an 'exploratory' multi mode factor analysis. UCLA Working Papers in Phonetics 16, 1–84.

Harshman, R.A., 1972. Determination and proof of minimum uniqueness conditions for PARAFAC1. UCLA Working Papers in Phonetics 22, 111–117.

Harshman, R.A., Lundy, M.E., 1984. Data preprocessing and the extended PARAFAC model. In: Law, H.G., Snyder, C.W., Hattie, J.A., McDonald, R.P. (Eds.), Research Methods for Multimode Data Analysis. Praeger Publishers, New York, pp. 216–284.

Hastie, T.J., Tibshirani, R.J., 1990. Generalized Additive Models. Chapman & Hall, London.

Kiers, H.A.L., 1998. A three-step algorithm for CANDECOMP/PARAFAC analysis of large data sets with multicollinearity. J. Chemometrics 12, 155–171.

Kiers, H.A.L., 2000. Towards a Standardized notation and terminology in multiway analysis. J. Chemometrics 14, 105–122.

Kiers, H.A.L., Harshman, R.A., 1997. Relating two proposed methods for speedup of algorithms for fitting two- and three-way principal component and related multilinear models. Chemometrics Intelligent Laboratory Systems 36, 31–40.

Kiers, H.A.L., Van Mechelen, I., 2001. Three-way component analysis: principles and illustrative application. Psychological Meth. 6, 84–110.

Kroonenberg, P.M., 1983. Three-mode principal component analysis. Theory and Applications. DSWO Press, Leiden.

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.

Lawson, C.L., Hanson, R.J., 1974. Solving Least Squares Problems. Prentice-Hall, Englewood Cliffs, NJ.

Little, R.J.A., Rubin, D.B., 1987. Statistical Analysis with Missing Data. Wiley, New York.

MATLAB, 2001. MATLAB. The Language of Technical Computing. The Mathworks, Inc., Natick, MA.

Mitchell, B.C., Burdick, D.S., 1994. Slowly converging PARAFAC sequences: swamps and two-factor degeneracies. J. Chemometrics 8, 155–168.

Ramsay, J.O., 1988. Monotone regression splines in action. Statist. Sci. 3 (4), 425–461.

Ramsay, J.O., Silverman, B.W., 1997. Functional Data Analysis. Springer, New York.

Rao, C.R., 1958. Some statistical methods for comparison of growth curves. Biometrics 14, 1–17.

Stevens, J., 1992. Applied Multivariate Statistics for the Social Sciences. Lawrence Erlbaum Ass, Hillsdale.

Tucker, L.R., 1951. A method for synthesis of factor analysis studies. Personnel Research Section Report No. 984. Dept. of the Army, Washington D.C.

Tucker, L.R., 1958. Determination of parameters of a functional relationship by factor analysis. Psychometrika 23, 19–23.

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

Van Rijckevorsel, J.L.A., 1988. Fuzzy coding and B-splines. In: Van Rijckevorsel, J.L.A., De Leeuw, J. (Eds.), Component and Correspondence Analysis. Dimension Reduction by Functional Approximation. Wiley, Chichester.