Chemometrics and Intelligent Laboratory Systems 129 (2013) 98–107

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

On the added value of multiset methods for three-way data analysis ☆

Kim De Roover [a,*], Marieke E. Timmerman [b], Iven Van Mechelen [a], Eva Ceulemans [a]

[a] KU Leuven, Belgium
[b] University of Groningen, The Netherlands

ARTICLE INFO

Article history:
Received 30 November 2012
Received in revised form 24 April 2013
Accepted 8 May 2013
Available online 18 May 2013

Keywords:
Three-way component analysis
Simultaneous component analysis
Clusterwise simultaneous component analysis

ABSTRACT

Three-way three-mode data are collected regularly in scientific research and yield information on the relation between three sets of entities. To summarize the information in such data, three-way component methods like CANDECOMP/PARAFAC (CP) and Tucker3 are often used. When applying CP and Tucker3 in empirical practice, one should be cautious, however, because they rely on very strict assumptions. We argue that imposing these assumptions may obscure interesting structural information included in the data and may lead to substantive conclusions that are appropriate for some part of the data only. As a way out, this paper demonstrates that this structural information may be elegantly captured by means of component methods for multiset data, that is to say, simultaneous component analysis (SCA) and its clusterwise extension (clusterwise SCA).

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Three-way three-mode data yield information on the relation between three distinct sets of entities (for an introduction, see [1]). For example, in sensory profiling research (see, e.g., [2]), one often asks a number of panelists to rate the same set of food samples (e.g., different cream cheeses or breads) on a variety of attributes (e.g., "tastefulness", and "crispness").

To summarize the information in such data, one often uses generalizations of standard principal component analysis. The most popular three-way component models are CANDECOMP/PARAFAC (CP) [3–5] and Tucker3 [6]. CP models the data in terms of a few underlying dimensions or components, on which the elements of the three modes receive a score. In case of sensory profiling data, these components can be interpreted as underlying dimensions of experience. As an example, one component could represent a dimension of overall experienced appreciation. The sample component scores reflect the positions of the food samples on the dimensions of experience (e.g., how well a particular bread sample is appreciated). The attribute component scores reflect which attributes determine the different experience dimensions (e.g., overall appreciation might be determined by "tastefulness" and "crispness"). Furthermore, the panelist component scores represent how salient these characteristics are for the different panelists, implying that the ratings of some panelists may be strongly influenced by a particular dimension whereas this is much less the case for the ratings of other panelists. Tucker3, which encompasses CP as a special case, reduces each of the three modes to a separate set of components and describes their interrelations by means of a core array. In comparison to CP, the interpretation of a Tucker3 model is much more complicated.

When applying CP and Tucker3 in empirical practice, one should be cautious, however, because they rely on very strict assumptions about the data. For example, in case of sensory profiling data, using CP implies that one is willing to endorse, among others, the following two major assumptions: (a) the underlying dimensions of experience are the same for all panelists, and (b) the positions of the food samples on the dimensions of experience (i.e., the sample component scores) are identical for all panelists. Reflecting on these two assumptions, we argue that imposing them may obscure two important types of individual differences among the panelists. First, it could be that different panelists rely on different dimensions of experience when rating the samples. For instance, for one subgroup of panelists overall experienced appreciation may be determined by "tastefulness", and "crispness", whereas for a second subgroup it is related to "tastefulness" and "softness". Second, even if (some of the) panelists rely on the same underlying dimensions of experience, they may disagree about the extent to which each dimension applies to particular samples: experience of food will always remain subjective, no matter how well-trained the panelists are. In the remainder of this paper, we will refer to the first type of differences as qualitative differences and to the second type as quantitative. Obviously, ignoring these individual differences may obscure interesting structural information included in the data, and may lead to substantive conclusions that are appropriate for some part of the data only.

Although CP and Tucker3 can in principle model qualitative differences by using a large number of components, this will hardly ever

☆ The research reported in this paper was partially supported by the fund for Scientific Research-Flanders (Belgium), project no. G.0477.09 awarded to Eva Ceulemans, Marieke Timmerman and Patrick Onghena and by the Research Council of KU Leuven (GOA/2010/02).

* Corresponding author at: Methodology of Educational Sciences Research Unit, Andreas Vesaliusstraat 2, B-3000 Leuven, Belgium.
E-mail address: Kim.DeRoover@ppw.kuleuven.be (K. De Roover).

0169-7439/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.chemolab.2013.05.002

work in practice. Specifically, in CP such differences would have to show up in the panelist-specific saliencies, in that a panelist component score of zero indicates that a panelist does not use the corresponding dimension. In Tucker3, the situation is somewhat more complex, as individual differences in the used dimensions may be represented by the component scores of the panelists as well as by the core array, with a consistent zero pattern for a specific sample or attribute component implying that the corresponding dimension of experience is irrelevant for the panelist in question. However, in practice, CP and Tucker3 models with large numbers of components will be very prone to error fitting, which interferes with the modeling of qualitative differences, as the overall fit of the model will in most cases be higher when each panelist has a non-zero score on all dimensions. Furthermore, CP and Tucker3 cannot represent quantitative individual differences.

This paper aims to demonstrate that both types of individual differences may be elegantly captured by means of component methods for multiset data [7]: simultaneous component analysis (SCA) [8] and its clusterwise extension [9,10]. Clusterwise SCA simultaneously performs a clustering of the panelists and fits a separate SCA model to the data within each cluster, implying that clusterwise SCA encompasses SCA as a special case. This implies that clusterwise SCA may yield insight into qualitative differences by means of a clustering of the panelists, with panelists that use different dimensions being assigned to different clusters. Furthermore, SCA models may reveal possible quantitative differences, because it allows the sample component scores to differ across panelists, but does not reveal qualitative differences because it restricts the attribute component scores to be the same.

The remainder of this paper is organized as follows. Section 2 describes the data structure and preprocessing, and introduces the sensory profiling data set that will be used in this paper. Section 3 concisely recapitulates the methods under consideration. For simplicity's sake, we will not discuss Tucker3 further, because the point that we want to make is essentially the same for CP and Tucker3. Therefore, we will focus on CP, in order to not needlessly complicate our line of reasoning. Section 4 discusses model selection, which is a challenging issue. Section 5 presents a simulation study that demonstrates the added value of (clusterwise) SCA over CP. In Section 6, each of these methods is applied to the sensory profiling data set, illustrating the benefits of multiset methods. Section 7 concludes the paper with a few points of discussion and directions for future research.

2. Data structure and preprocessing

2.1. Data structure

In this paper, we will re-analyze the bread data [11] that can be downloaded from http://www.models.kvl.dk/Sensory_Bread. These data pertain to 10 bread samples, rated on 11 attributes by 8 panelists. The bread samples correspond to five types of bread, with two samples included for each type. These data can be presented in terms of an $I \times J \times K$ data array $\underline{\mathbf{X}}$, where $i = 1, \ldots, I$ refers to the samples, $j = 1, \ldots, J$ refers to the attributes, and $k = 1, \ldots, K$ corresponds to the panelists.

2.2. Preprocessing

With respect to preprocessing, different alternatives can be considered (e.g., [12–14]). In this paper, we standardize the attributes for each panelist, removing some types of individual differences from the data. To give more insight in how the data are transformed by this type of preprocessing and which information is retained in the analysis, we decompose the raw (i.e., unpreprocessed) data matrices $\mathbf{X}_k^r$ as follows:

$$\mathbf{X}_k^r = \mathbf{1}_I \bar{\mathbf{x}}_k^r + \mathbf{X}_k \mathbf{S}_k, \quad \text{with} \quad \mathbf{X}_k = (\mathbf{X}_k^r - \mathbf{1}_I \bar{\mathbf{x}}_k^r)\,\mathbf{S}_k^{-1} \tag{1}$$

where $\mathbf{1}_I$ is a $I \times 1$ vector of ones, $\bar{\mathbf{x}}_k^r$ is a $1 \times J$ vector containing the attribute means of panelist k, $\mathbf{S}_k$ is a diagonal matrix containing the standard deviations of the attributes for panelist k, and $\mathbf{X}_k$ denotes the columnwise standardized data of panelist k (i.e., mean of zero and standard deviation of one for each attribute). Thus, standardizing the attributes per panelist implies that we analyze the $\mathbf{X}_k$ matrices from Eq. (1).

Eq. (1) clarifies that we first discard between-panelist differences in attribute means $\bar{\mathbf{x}}_k^r$ (i.e., the so-called between-structure), in order to focus on the within-panelist variability in $\mathbf{X}_k\mathbf{S}_k$ (i.e., the within-structure). Indeed, our stance is that these means should never be included in three-way or multiset component analysis without assuring that the between-structure is the same as the within-structure, to avoid confounding differences in within- and between-structures. Of course, individual differences in means can be interesting in empirical practice. Therefore, one may opt to model the between-structure by a separate principal component analysis, as is done in multilevel component analysis [15].

Moreover, we remove between-panelist differences in the variances of the attributes (i.e., differences in $\mathbf{S}_k$ matrices), as they may have a strong impact on the results with individuals with larger variances dominating the analysis and the thus obtained dimensions. Indeed, analyzing the $\mathbf{X}_k$ matrices implies that we give each panelist an equal weight in the analysis and that we focus on individual differences in the attribute correlations rather than their covariances. Of course, in general, the answer to the question whether or not variance differences should be removed from the data depends on the possible sources of these differences (e.g., measurement error or real variability) and their reliability or stability and on the research question that one wants to answer.

3. Modeling individual differences in three-way data

3.1. CP: capturing individual differences in sensitivity to the dimensions

CP [3–5] decomposes $\underline{\mathbf{X}}$ into three component matrices that contain the scores of the elements of the three modes on the Q components: $\mathbf{A}$ ($I \times Q$), $\mathbf{B}$ ($J \times Q$) and $\mathbf{C}$ ($K \times Q$). Specifically, the data matrix $\mathbf{X}_k$ of panelist k is modeled as:

$$\mathbf{X}_k = \mathbf{A}\mathbf{D}_k\mathbf{B}' + \mathbf{E}_k. \tag{2}$$

As $\mathbf{A}$ and $\mathbf{B}$ are the same for all panelists, it becomes clear that the underlying experience dimensions, which can be labeled on the basis of $\mathbf{B}$, as well as the scores of the bread samples on these dimensions ($\mathbf{A}$) are assumed to be identical across panelists. $\mathbf{D}_k$ is a diagonal matrix with the elements of the k-th row of $\mathbf{C}$ on its diagonal, which reflect the panelist-specific 'saliencies'. Thus, the inclusion of $\mathbf{D}_k$ leaves room for individual differences in sensitivity to the dimensions. Under some mild conditions (e.g., [16]), a CP solution is essentially unique, i.e., can be identified up to permutation and reflection of the columns of $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$.

3.2. SCA: capturing quantitative individual differences

The model equation of SCA can be written in a form that closely resembles that of Eq. (2):

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

under the constraint that the diagonal elements of $\mathbf{A}_k'\mathbf{A}_k$ equal I. $\mathbf{B}$ holds the attribute component scores that are the same for all panelists, implying that the panelists are assumed to use the same dimensions. The diagonal matrix $\mathbf{D}_k$ again leaves room for individual differences in sensitivity to these dimensions. Comparing Eqs. (2) and (3), one can conclude that SCA differs from CP in that the sample component scores in

$\mathbf{A}_k$ are allowed to differ across panelists. Thus, $\mathbf{A}_k$ models the quantitative individual differences under consideration.

Note that different variants of SCA can be distinguished based on their restrictions on the correlations and variances of $\mathbf{A}_k\mathbf{D}_k$ [8], and thus on the off-diagonal and diagonal elements of $\mathbf{D}_k{}'\mathbf{A}_k{}'\mathbf{A}_k\mathbf{D}_k$ (note that $\mathbf{A}_k$ is centered, because $\mathbf{X}_k$ is centered). Because of these restrictions, some SCA variants are essentially unique under mild conditions [8].

3.3. Clusterwise SCA: capturing qualitative and quantitative individual differences

Clusterwise SCA assigns the panelists to G clusters according to the underlying component structure and performs an SCA variant within each cluster. Hence, clusterwise SCA decomposes the data of panelist k as follows

$$\mathbf{X}_k = \sum_{g=1}^{G} p_{kg}\left(\mathbf{A}_k\mathbf{D}_k\mathbf{B}_g{}'\right) + \mathbf{E}_k, \tag{4}$$

where p_{kg} is an element of the $K \times Q$ partition matrix $\mathbf{P}$ – which equals 1 when panelist k belongs to cluster g and 0 otherwise – and where the index of $\mathbf{B}_g$ indicates that the attribute component scores, which are used to interpret the components, are cluster-specific. Thus, the qualitative differences under consideration are modeled through $\mathbf{P}$ and $\mathbf{B}_g$. Moreover, as $\mathbf{A}_k$ and $\mathbf{D}_k$ are panelist-specific, panelists can still differ in how they score each bread on the cluster-specific dimensions (quantitative differences) and in their sensitivity to the dimensions. The number of clusters G can range from one (equivalent to SCA) to K (equivalent to separate principal component analysis per panelist), making it a very flexible method.

4. Model selection

Given the multitude of models and variants, selecting the correct model, variant, complexity (how many dimensions and, if appropriate, clusters) and data slicing (see below), becomes a big issue. In this section we present and discuss a decision tree (see Fig. 1), that may help users through this model selection process. Note that this decision tree deals with the different aspects of the model selection process sequentially. Of course, one may wonder why we don't propose a simultaneous model selection strategy, implying that one fits all possible solutions (i.e., including all models, variants, complexities and slicing options) and selects the one which balances description of the data and complexity best. An important impediment for such a simultaneous strategy is that it is difficult to come up with a proper complexity measure that can be easily computed and sensibly compared for the different solutions. Indeed, it has been suggested to simply use the number of free parameters, but simulation studies have shown that this measure sometimes leads to an inferior model selection performance [17,18], probably because the impact of the different parameters is not always the same (e.g., the elements of the partition versus component score matrices; [19]). Moreover, a simultaneous model selection strategy would imply performing a very large number of analyses, which would be rather time-consuming.

The first step of the decision tree in Fig. 1 consists of deciding whether one wants to allow for quantitative differences. If not, CP or Tucker3 is the appropriate model. Next, if one wants to allow for quantitative differences and, thus, opts to use multiset methods, a second step pertains to evaluating whether qualitative differences are or interest. If so, clusterwise SCA is the best choice; if not, SCA is most appropriate. Two steps remain, which we will elaborate below: choosing the appropriate slicing of the three-way data array into data matrices (in case of a multiset analysis) on the one hand, and selecting the model variant and number of components and clusters on the other hand.

Regarding slicing, taking into account that rows and columns have a different status in multiset methods, the slicing can be done in six different ways, with each way of slicing allowing to represent other types of quantitative and qualitative differences. For example, in the present paper we slice the bread data array into a collection of bread by attribute matrices, with each matrix pertaining to one panelist, to look for individual differences. Yet, we could also transform

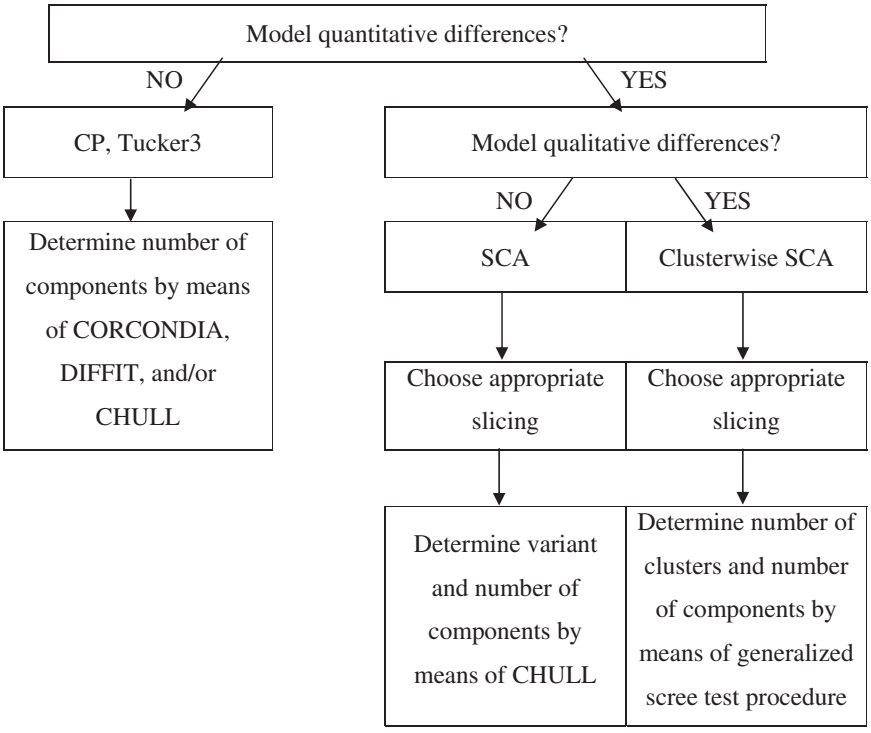


Fig. 1. Decision tree for selection of most appropriate three-way or multiset component method for a given three-way data array.

the original data array into a collection of panelist by attribute matrices, with each matrix pertaining to one particular bread sample. In that case, clusterwise SCA could provide insight into qualitative differences among the breads. Such differences could, for instance, include that the overall experienced appreciation for white bread is based on how soft and moist the bread is, while for gray bread a more grainy and tough structure could be preferred. As the examples indicate, substantive arguments may guide in selecting the appropriate slicing.

Regarding the final step, different variants of CP, Tucker3, SCA and clusterwise SCA are available and all of these methods can be applied with different numbers of components (and, in case of clusterwise SCA, also with different numbers of clusters). Therefore, given a particular data set and model, one has to decide which model variant and which model complexity yield the most adequate description of the data, in that the model describes the data well, without becoming overly complex. For this model selection aspect, a number of tools are available that can help the user in making a decision. Specifically, the CHULL procedure can be used for selecting among three-way component models or simultaneous component models of different variants and complexities [17,20–22]. For selecting the most appropriate number of components for CP one can also make use of CORCONDIA [23], whereas for Tucker3 the DIFFIT procedure [24] can be applied. For clusterwise SCA, the generalized scree test procedure described by De Roover, Ceulemans and Timmerman [7] can be used to select the number of clusters and number of components. Of course, additional information on the data or theoretical knowledge may be used as well. Moreover, some form of cross-validation may be necessary to ensure that the obtained solution is stable.

5. Simulation study

5.1. Problem

In this section, we present a simulation study in which we assess the fit and recovery performance of CP, SCA-IND,[1] and clusterwise SCA-IND when applied to data that contain quantitative and/or qualitative differences (note that we only evaluate the (clusterwise) SCA variant that is used in the application section, in order to not needlessly complicate the study). As we expect that CP will perform weakly in the presence of quantitative or qualitative differences and that SCA-IND will break down when data contain qualitative differences, we generate data according to CP, SCA-IND, or clusterwise SCA-IND and apply all three methods to these data, with the correct number of components, to verify these hypotheses. Moreover, we expect a data array that is generated according to a more restricted model to be adequately modeled by a more flexible method, although some overfitting can occur.

5.2. Design and procedure

In this simulation study, the size of the data array **X** is fixed at 50 (samples) × 12 (attributes) × 30 (panelists) and the number of components at two. The following three factors were systematically varied in a complete factorial design:

1. the *underlying model* at 3 levels: CP, SCA-IND, clusterwise SCA-IND (with two clusters);
2. the *cluster sizes* at 2 levels: in case of clusterwise SCA-IND, the first cluster contains either 60% or 80% of the panelists; note that in case of CP or SCA-IND, this factor has no effect;
3. the *error level e*, which is the expected proportion of error variance in the data matrices $\mathbf{X}_k$, at 2 levels: .20, .40;

The data matrices $\mathbf{X}_k$ are constructed on the basis of Eq. (4). The partition matrix **P** was generated taking the number of clusters (factor 1: $G = 1$ for CP and SCA-IND, and $G = 2$ for clusterwise SCA-IND) and the cluster size (factor 2) into account. The sample component score matrices $\mathbf{A}_k$ were randomly sampled from a multivariate normal distribution with zero means and identity covariance matrix; in case of CP data, all $\mathbf{A}_k$ matrices were identical (i.e. $\mathbf{A}_k = \mathbf{A}$ for $k = 1, \ldots, K$). The saliencies in the $\mathbf{D}_k$ matrices were uniformly sampled between $\sqrt{0.25}$ and $\sqrt{1.75}$, implying that the panelist-specific component variances fluctuate between 0.25 and 1.75. The attribute component scores (i.e., the $\mathbf{B}_g$ matrices) were sampled uniformly between −1 and 1; in case of CP and SCA-IND, a single $\mathbf{B}_g$ matrix (i.e., $\mathbf{B}_g = \mathbf{B}$) was drawn. The entries from the error matrices $\mathbf{E}_k$ were randomly sampled from a standard normal distribution. The resulting $\mathbf{E}_k$ and $\mathbf{B}_g$ matrices were rescaled to obtain data that contain the desired proportion of error variance (factor 3). Finally, the thus obtained data matrices $\mathbf{X}_k$ were columnwise standardized.

For each cell of the design, 20 data arrays were generated, yielding 240 data sets. Each of them was analyzed with CP, SCA-IND, and clusterwise SCA-IND, using two components. In case of CP or SCA-IND data, clusterwise SCA-IND was applied with one cluster, boiling down to SCA-IND; in case of clusterwise SCA-IND data two clusters were used. In the CP analysis, we used a rational start based on the direct trilinear decomposition method [25]. The SCA-IND analysis was run with one rational start based on a singular value decomposition (SVD) of the complete data [8]. In the clusterwise SCA-IND analysis, we applied a multistart procedure using 25 different random initializations of the clustering [9,10]; the SCA-IND analyses within the clusters are started rationally with SVD.

5.3. Results

5.3.1. Goodness of fit

To evaluate the goodness of fit of the obtained solutions, we computed the corresponding percentages of variance accounted for as follows:

$$\mathrm{VAF}(\%) = \frac{\sum_{k=1}^{K} \left\| \hat{\mathbf{X}}_k \right\|^2}{\sum_{k=1}^{K} \left\| \mathbf{X}_k \right\|^2} \times 100 \tag{4}$$

where $\hat{\mathbf{X}}_k$ equals $\mathbf{A}\mathbf{D}_k\mathbf{B}'$, $\mathbf{A}_k\mathbf{D}_k\mathbf{B}'$ and $\sum_{g=1}^{G} p_{kg}\left(\mathbf{A}_k\mathbf{D}_k\mathbf{B}_g{}'\right)$ for CP, SCA-IND and clusterwise SCA-ECP solutions, respectively.

In Table 1, the mean VAF(%) is tabulated for each cell of the design. In this table (and in Tables 2 and 3), the values that are obtained using a method that matches the data generating model or that is less restrictive, are printed in boldface. It can be concluded that fit extremely deteriorates when quantitative differences are present in the data, but not modeled, and that not modeling qualitative differences also has a clear, but less dramatic effect. Moreover, CP slightly underfits the data (i.e., VAF(%) < (1 − *e*) × 100%), whereas SCA-IND and its clusterwise extension somewhat overfit the data (i.e., VAF(%) > (1 − *e*) × 100%), especially when the data contain more error. The effect of cluster size is negligible.

5.3.2. Goodness of recovery

5.3.2.1. Recovery of the attribute component scores. To evaluate the recovery of the attribute component scores within true cluster *g*, we obtained a goodness-of-attribute-score-recovery statistic ($GOASR_g$) by computing Tucker congruence coefficients φ [26] between the

[1] SCA-IND restricts $\mathbf{D}_k{}'\mathbf{A}_k{}'\mathbf{A}_k\mathbf{D}_k$ to be a diagonal matrix for all panelists $k = 1, \ldots, K$, containing the variances of $\mathbf{A}_k\mathbf{D}_k$. Under some mild uniqueness assumptions, SCA-IND estimates are unique up to permutation, reflection and rescaling of the components [8].

Table 1
The mean VAF(%) of the estimated models for the data arrays in each cell of the simulation study design. Values that are obtained using a method that matches the data generating model or that is less restrictive, are printed in boldface.

Data characteristics			Analysis applied		
Cluster sizes	Error level	Generating model	CP	SCA-IND	Clusterwise SCA-IND
60% of panelists in first cluster (in case of clusterwise SCA-IND data)	20%	CP	**79%**	**82%**	**82%**
		SCA-IND	9%	**81%**	**81%**
		Clusterwise SCA-IND	8%	59%	**81%**
	40%	CP	**60%**	**66%**	**66%**
		SCA-IND	8%	**65%**	**65%**
		Clusterwise SCA-IND	6%	48%	**64%**
80% of panelists in first cluster (in case of clusterwise SCA-IND data)	20%	CP	**78%**	**81%**	**81%**
		SCA-IND	10%	**81%**	**81%**
		Clusterwise SCA-IND	9%	68%	**81%**
	40%	CP	**59%**	**66%**	**66%**
		SCA-IND	7%	**65%**	**65%**
		Clusterwise SCA-IND	7%	55%	**65%**

true and estimated scores and averaging these coefficients across components:

$$GOASR_g = \frac{\sum_{q=1}^{Q} \varphi\left(\mathbf{b}_{gq}^{T}, \mathbf{b}_{gq}^{M}\right)}{Q} \quad (4)$$

with φ being the Tucker phi coefficient and $\mathbf{b}_{gq}^{T}$ and $\mathbf{b}_{gq}^{M}$ indicating the true and estimated attribute scores on the q-th component of cluster g. The estimated components are permuted and reflected such that $GOASR_g$ is maximized. In case of two clusters, the permutational freedom of the clusters of a clusterwise SCA-IND model (i.e., the columns of **P** can be permuted without altering the fit of the solution) is dealt with by matching the largest estimated cluster with the largest true one (see factor 2). Moreover, for clusterwise SCA-IND data, the results of CP and SCA-IND analyses are compared to each true cluster to investigate which true cluster is captured by the estimated model, and to which extent.

Table 2
Mean $GOASR_g$ of the estimated models for the data arrays in each cell of the simulation study design. Values that are obtained using a method that matches the data generating model or that is less restrictive, are printed in boldface.

Data characteristics				Analysis applied			
				CP	SCA-IND	Clusterwise SCA-IND	
Cluster sizes	Error level	Generating model				Cluster 1	Cluster 2
60% of panelists in first cluster (in case of clusterwise SCA-IND data)	20%	CP		**1.00**	**1.00**	**1.00**	
		SCA-IND		.77	**.99**	**.99**	
		Clusterwise SCA-IND	Cluster 1	.74	.82	**.99**	
			Cluster 2	.34	.56		**.99**
	40%	CP		**1.00**	**.98**	**.98**	
		SCA-IND		.74	**1.00**	**1.00**	
		Clusterwise SCA-IND	Cluster 1	.81	.84	**1.00**	
			Cluster 2	.30	.57		**.98**
80% of panelists in first cluster (in case of clusterwise SCA-IND data)	20%	CP		**.99**	**.98**	**.98**	
		SCA-IND		.75	**.98**	**.98**	
		Clusterwise SCA-IND	Cluster 1	.80	.95	**1.00**	
			Cluster 2	.26	.40		**.97**
	40%	CP		**1.00**	**1.00**	**1.00**	
		SCA-IND		.77	**1.00**	**1.00**	
		Clusterwise SCA-IND	Cluster 1	.82	.96	**1.00**	
			Cluster 2	.29	.39		**.98**

Table 3
Mean $GOSSR_g$ of the estimated models for the data arrays in each cell of the simulation study design. Values that are obtained using a method that matches the data generating model or that is less restrictive, are printed in boldface.

Data characteristics				Analysis applied			
				CP	SCA-IND	Clusterwise SCA-IND	
Cluster sizes	Error level	Generating model				Cluster 1	Cluster 2
60% of panelists in first cluster (in case of clusterwise SCA-IND data)	20%	CP		**.99**	**.96**	**.96**	
		SCA-IND		.01	**.95**	**.95**	
		Clusterwise SCA-IND	Cluster 1	.03	.88	**.95**	
			Cluster 2	.01	.73		**.95**
	40%	CP		**.99**	**.91**	**.91**	
		SCA-IND		.00	**.92**	**.92**	
		Clusterwise SCA-IND	Cluster 1	.01	.84	**.92**	
			Cluster 2	.00	.71		**.91**
80% of panelists in first cluster (in case of clusterwise SCA-IND data)	20%	CP		**.98**	**.94**	**.94**	
		SCA-IND		.02	**.94**	**.94**	
		Clusterwise SCA-IND	Cluster 1	.03	.92	**.95**	
			Cluster 2	.00	.70		**.93**
	40%	CP		**.99**	**.92**	**.92**	
		SCA-IND		.02	**.91**	**.91**	
		Clusterwise SCA-IND	Cluster 1	.00	.89	**.92**	
			Cluster 2	.00	.59		**.91**

In Table 2, the mean $GOASR_g$ values are tabulated for each cell of the design. Using the guidelines by Lorenzo-Seva and ten Berge [27] (i.e., $\varphi > .85$ indicates good recovery and $\varphi > .95$ excellent recovery), we conclude that when the analysis method matches the data generating model or is less restrictive, results are excellent. In contrast, but as hypothesized, results are bad when the analysis method is more restrictive than the data generating method. There is one exception though: in case the first cluster is large (i.e., containing 80% of the panelists), SCA-IND recovers the corresponding attribute scores very well, because this cluster dominates the analysis. The error level hardly influences the results.

5.3.2.2. Recovery of the sample component scores. For quantifying the recovery of the sample component scores within true cluster g, we obtained a goodness-of-sample-score-recovery statistic ($GOSSR_g$) by computing congruence coefficients between the true and estimated sample component scores for each panelist and averaging these coefficients across components and across the panelists that are assigned[2] to this true cluster:

$$GOSSR_g = \frac{\sum_{k=1}^{K} p_{kg} \sum_{q=1}^{Q} \varphi\left(\mathbf{a}_{kq}^{T}, \mathbf{a}_{kq}^{M}\right)}{Q \sum_{k=1}^{K} p_{kg}} \quad (4)$$

with $\mathbf{a}_{kq}^{T}$ and $\mathbf{a}_{kq}^{M}$ indicating the true and estimated scores of panelist k on the q-th component of cluster g. The permutational freedom of the components and clusters is handled in the same way as for $GOASR_g$.

Table 3 summarizes the $GOSSR_g$ results for each cell of the design. The same conclusions can be drawn as in the previous subsection: recovery is bad when a too restrictive analysis method is used and good to excellent otherwise; also, SCA-IND succeeds quite well in recovering the sample scores within the largest clusterwise SCA-IND cluster. The $GOSSR_g$ values being somewhat smaller than the corresponding $GOASR_g$ ones, is probably due to the fact that $GOSSR_g$ is calculated over multiple 50×2 score matrices while $GOASR_g$ pertains to only a single 12×2 matrix.

[2] Clusterwise SCA-IND always yielded the correct clustering of the panelists.

Table 4
Component matrices of the CP model with two components for the bread data (attribute component scores with an absolute value > .45 are printed in boldface).

	A			B			C	
	Salt	Yeast		Salt	Yeast		Salt	Yeast
Bread 1	−1.82	−.46	Bread-od	.24	−.05	Panelist 1	.76	.92
Bread 2	−1.85	−.40	Yeast-od	.28	**.63**	Panelist 2	.94	1.35
Bread 3	.01	1.57	Off-flav	**−.54**	−.40	Panelist 3	.98	.75
Bread 4	−.05	1.35	Color	.32	−.01	Panelist 4	1.16	1.01
Bread 5	.19	.59	Moisture	**.75**	.15	Panelist 5	1.05	.93
Bread 6	.09	.79	Tough	**.74**	−.21	Panelist 6	.92	.21
Bread 7	1.17	−1.24	Salt-t	**.71**	−.35	Panelist 7	.85	1.33
Bread 8	1.06	−1.62	Sweet-t	**−.83**	−.15	Panelist 8	1.25	1.04
Bread 9	.55	−.29	Yeast-t	.13	**.67**			
Bread 10	.65	−.29	Other-t	**−.74**	−.48			
			Total	.29	.17			

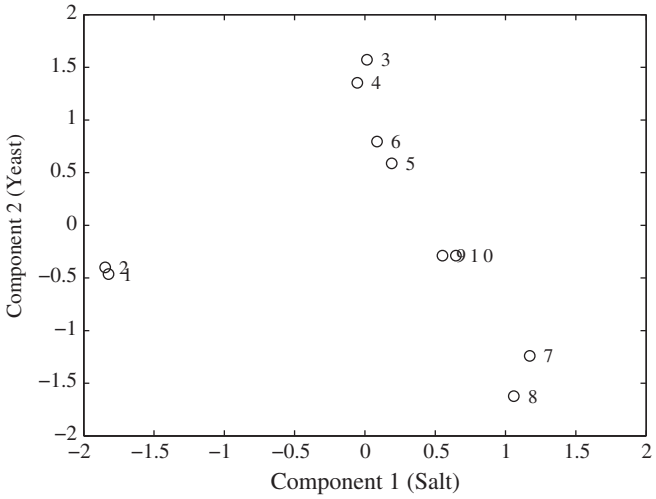


Fig. 2. Sample component score (**A**) plot for the CP model with two components for the bread data.

6. Application

In this section, we apply the three methods under consideration to the bread data.

6.1. CP: capturing individual differences in sensitivity to the dimensions

We start by fitting a CP model to the data. Like Bro [11], we use two components.[3] The component matrices of the model (which explain 44% of the variance in the data) are given in Table 4. To interpret the components, we inspect the scores of the attributes in matrix **B**. The scores for component 1 indicate that lower ratings on "salt taste", "tough" and "moisture" are associated with higher ratings on "off-flavor", "sweet taste" and "other taste", and vice versa. This component can be interpreted as a "salt" dimension, because salt tightens gluten (gives a tougher structure), balances flavor (masks the off-flavors and makes the bread less sweet) and helps bread retain moisture. The scores of the attributes on component 2 suggest that "yeast-odor" co-occurs with "yeast-taste", implying that this component may be interpreted as a "yeast" dimension.

The scores of the bread samples on these dimensions can be found in **A** (see Table 4 and Fig. 2). As could be expected, samples of the same bread type, which are indicated by adjacent numbers, have similar component scores. Furthermore, the first type of bread (samples 1 and 2) differs strongly from the other bread types with respect to the first dimension, for which we have no explanation as we have no information on the ingredients of the different bread types. Moreover, bread samples 3 and 4 have the highest score on the second component, implying that they are judged as having the strongest yeast odor and yeast taste. Matrix **C** contains the individual differences in sensitivity to the components; e.g., judge 6 has a very low sensitivity to the second component, implying that his/her ratings are not strongly influenced by the amount of yeast in the bread.

6.2. SCA: capturing quantitative individual differences

To explore whether there are quantitative differences present in the data, in that the scores of the bread samples differ across the panelists, we estimate an SCA-IND[4] model with two components. This model explains 52% of the variance in the data and the attribute component scores (**B**) are printed in Table 5. The first component appears to be a salt dimension that is very similar to the first CP component. The second component points to a negative correlation between "yeast odor" and "yeast taste" on the one hand and "salt taste" and "tough" on the other hand. Thus, this component can be read as a "yeast" dimension, but with a slightly different structure as in the CP solution (i.e., a stronger negative score of "salt taste" and "tough"). The negative correlation between the yeast flavor attributes and "salt taste" may be explained by the fact that yeast growth is tempered by salt.

To quantify the similarity of the SCA-IND and CP components, Tucker congruence coefficients [26] were calculated between both sets of components (see Table 9). Using the guidelines of Lorenzo-Seva and ten Berge [27], we can state that the first SCA-IND component is identical (i.e., congruence > .95) to the first CP component and that the second SCA-IND component is very similar (i.e., congruence > .85) to the second CP component.

The advantage of SCA-IND over CP is that the component scores of the bread samples ($\mathbf{A}_k$) can differ among panelists. The panelist-specific score plots are given in Fig. 3. A few examples of individual differences that could not be captured by CP: for panelist 6, bread samples 1 and 2 lie farther apart from each other with respect to the "yeast" dimension than in the other score patterns; for panelist 2, bread samples 7 and 8 clearly get the lowest score on the "yeast" dimension, while they lie in the middle of the plot for judge 3; etcetera.

How different the sample component score patterns are among the panelists, is quantified by the mean congruence coefficient (i.e., the mean over the two components) for the $\mathbf{A}_k$-matrices for each pair of panelists (see Table 6): As according to Lorenzo-Seva & ten Berge [27] a congruence coefficient lower than .85 indicates a low similarity, SCA-IND captures some considerable quantitative differences among panelists in the score pattern of the bread samples; especially panelist 6 appears to have a very different score pattern. Additionally, we can inspect the individual differences in sensitivity to the components by looking at the saliencies in Table 5: as was the case in the CP model, panelist 6 is less sensitive to the "yeast" dimension.

6.3. Clusterwise SCA: capturing qualitative and quantitative individual differences

Next to quantitative differences, we are also interested in possible qualitative differences in the dimensions. To explore these differences, clusterwise SCA-IND is applied with two clusters and two components per cluster.[5] This model explains 57% of the variance in the data and its first cluster contains panelists 1, 2 and 7, while the five

[3] The core consistency diagnostic [23] was 100% for the reported PARAFAC model, suggesting that a solution with two components is adequate.

[4] The four SCA variants were fitted with two components and SCA-IND was selected as the best one (in terms of balance between fit and complexity) by means of the CHULL model selection procedure [17,20].

[5] Selected by means of the model selection procedure described by De Roover and colleagues [7,9].

Table 5
Attribute component scores (**B**) and saliencies ($\mathbf{D}_k$) of the SCA-IND model with two components for the bread data (attribute component scores with an absolute value > .45 are printed in boldface).

	Attribute component scores			Saliencies	
Bread-od	.34	−.08	Panelist 1	.68	1.09
Yeast-od	.35	**.69**	Panelist 2	.95	1.25
Off-flav	**−.63**	−.30	Panelist 3	.96	.88
Color	.44	−.05	Panelist 4	1.20	.88
Moisture	**.82**	−.06	Panelist 5	1.10	.90
Tough	**.69**	**−.49**	Panelist 6	.89	.66
Salt-t	**.62**	**−.63**	Panelist 7	.87	1.20
Sweet-t	**−.81**	.14	Panelist 8	1.24	1.02
Yeast-t	.24	**.72**			
Other-t	**−.76**	−.23			
Total	.32	.04			

other panelists are assigned to cluster 2. The differences in dimensions can be investigated by comparing the two attribute score matrices $\mathbf{B}_g$ in Table 7.

With respect to cluster 1, we can state that the first component is a salt dimension (like in the CP and SCA-IND solution) but with an extra high (negative) score of the "total" rating, indicating a negative association between the "salt taste" rating and the total appreciation of the bread sample. Consequently, "salt & disfavor" seems an appropriate label for this component. The second component of cluster 1 is again a "yeast" dimension that is essentially identical to the second component of the SCA-IND solution (i.e., congruence > .95; see Table 9).

The high attribute scores on the first component of cluster 2 indicate that a stronger salt taste is associated with a higher appreciation of the bread sample as well as with a higher moisture level, tougher structure, and less yeast and sweet taste. This component is thus labeled "salt & favor". The second component has high scores of almost all attributes, and it implies that "off-flavor", "sweet taste" and "other taste" are negatively related to most other attributes (including the total appreciation of the bread sample), with the exception of "salt taste". Therefore, this component is labeled "other flavor" (i.e., other than salt flavor).

When comparing the first components of both clusters, we encounter a very important difference between the two clusters, i.e., the panelists in cluster 1 do not appreciate bread with a strong salty taste whereas the panelists in cluster 2 seem to prefer it. Moreover, the overall component structure is very different between the two clusters, as indicated by the relatively low congruences between the components of both clusters (i.e., <.85; see Table 9).

On top of these qualitative differences in dimensions, clusterwise SCA-IND also allows for differences in bread sample component scores and saliencies within the clusters. The component scores $\mathbf{A}_k$ are plotted in Fig. 4 for the panelists in cluster 1, and in Fig. 5 for the panelists in cluster 2. Examples of individual differences in the score pattern are that, within cluster 1, panelist 7 gives bread sample 8 the lowest rating on the "salt & disfavor" component, while panelists 1 and 2 give bread sample 7 the lowest rating, and, within cluster 2, panelist 6 rates bread sample 2 much lower on the "salt & favor" component than the other panelist. To express how much the panelists within a cluster differ from one another in their component scores for the bread samples, the mean congruence coefficient (i.e., the mean over the two components) for the $\mathbf{A}_k$-matrices is computed for each pair of panelists in the same cluster (see Table 8). These congruence coefficients point to considerable differences within each cluster (i.e., most congruences < .85). The largest differences appear to exist between panelist 6 and the other panelists in cluster 2. In Table 7, we see that panelist 6 also stands out with a much lower saliency for the "other flavor" component of cluster 2, next to other, more subtle differences in sensitivity.

6.4. Conclusion

When comparing the different models for the bread data, we conclude that the multiset methods have contributed to a better understanding of the underlying structure of the data. In particular, in

Table 6
Congruence coefficients among sample component scores $\mathbf{A}_k$ (mean over the two components) for the SCA-IND model for the bread data (congruences > .85 are printed in boldface).

	Panelist 1	Panelist 2	Panelist 3	Panelist 4	Panelist 5	Panelist 6	Panelist 7	Panelist 8
Panelist 1	–	–	–	–	–	–	–	–
Panelist 2	**.87**	–	–	–	–	–	–	–
Panelist 3	.70	.75	–	–	–	–	–	–
Panelist 4	.84	**.91**	.84	–	–	–	–	–
Panelist 5	.73	.76	.77	.78	–	–	–	–
Panelist 6	.51	.51	.61	.50	.54	–	–	–
Panelist 7	.77	.79	.77	.77	.81	.53	–	–
Panelist 8	.78	.82	**.89**	**.87**	.71	.50	**.87**	–

Table 7
Attribute component scores ($\mathbf{B}_g$) and saliencies ($\mathbf{D}_k$) of the clusterwise SCA-IND model with two clusters and two components for the bread data (attribute component scores with an absolute value > .45 are printed in boldface).

	Attribute component scores					Saliencies			
	Cluster 1		Cluster 2			Cluster 1		Cluster 2	
	Salt & disfavor	Yeast	Salt & favor	Other flavor		Salt & disfavor	Yeast	Salt & favor	Other flavor
Bread-od	−.26	−.07	.28	**.46**	Panelist 1	.98	.86	–	**–**
Yeast-od	.38	**.78**	−.26	**.66**	Panelist 2	1.02	1.07	–	**–**
Off-flav	**−.49**	−.22	−.12	**−.74**	Panelist 3	**–**	–	.96	.91
Color	.24	−.36	.05	**.56**	Panelist 4	–	–	.87	1.14
Moisture	**.81**	−.19	**.47**	**.64**	Panelist 5	**–**	–	.88	1.10
Tough	**.49**	**−.71**	**.72**	**.46**	Panelist 6	**–**	**–**	1.10	.64
Salt-t	**.45**	**−.74**	**.82**	.23	Panelist 7	1.00	1.05	**–**	–
Sweet-t	**−.74**	.32	**−.56**	**−.63**	Panelist 8	**–**	–	1.15	1.12
Yeast-t	.43	**.71**	**−.48**	**.57**					
Other-t	**−.86**	−.28	−.39	**−.62**					
Total	**−.57**	.39	**.58**	**.47**					

Table 8
Congruence coefficients among sample component scores $\mathbf{A}_k$ (mean over two components) for the clusterwise SCA-IND model with two clusters and two components for the bread data (congruences > .85 are printed in boldface).

	Cluster 1				Cluster 2				
	Panelist 1	Panelist 2	Panelist 7		Panelist 3	Panelist 4	Panelist 5	Panelist 6	Panelist 8
Panelist 1	–	–	–	Panelist 3	–	–	–	–	–
Panelist 2	**.85**	–	–	Panelist 4	**.85**	–	–	–	–
Panelist 7	.80	.80	–	Panelist 5	.78	.75	–	–	–
				Panelist 6	.58	.46	.53	–	–
				Panelist 8	**.87**	.82	.64	.53	–

Table 9
Congruence coefficients among the attribute component scores (per component) of the different models for the bread data (congruences > .85 are printed in boldface).

		CP		SCA-IND		Clusterwise SCA-IND			
		S	Y	S	Y	S & D (cl. 1)	Y (cl. 1)	S & F (cl. 2)	OF (cl. 2)
CP	S	–	–	–	–	–	–	–	–
	Y	.16	–	–	–	–	–	–	–
SCA-IND	S	**.98**	.32	–	–	–	–	–	–
	Y	−.24	**.92**	−.06	–	–	–	–	–
Clusterwise	S & D (cl. 1)	.80	.42	.82	.09	–	–	–	–
SCA-IND	Y (cl. 1)	−.35	.82	−.19	**.95**	−.13	–	–	–
	S & F (cl. 2)	.84	−.30	74	−.64	.41	−.60	–	–
	OF (cl. 2)	.83	.59	**.92**	.27	.71	.13	.49	–

S = Salt; Y = Yeast; S&D = Salt & Disfavor; S&F = Salt & favor; OF = Other flavor.

addition to the between-panelist differences in component scores of the bread samples (that were also captured by SCA-IND), clusterwise SCA-IND reveals remarkable qualitative differences in the dimensions underlying the ratings, and these differences may partly be explained by the fact that the clustering separated the panelists that dislike salty bread from panelists that prefer it.

7. Discussion

The goal of this paper was to show the added value of multiset methods for three-way data analysis. Specifically, SCA can capture quantitative differences in the component score patterns of different data slices. On top of that, clusterwise SCA can reveal qualitative differences in the component structure that underlies these data slices. For illustrative purposes, we re-analyzed sensory profiling data, but the argument of course holds for other types of three-way data as well.

A first point of discussion pertains to the fact that the violations of the CP-assumptions that we encountered in the sensory profiling case are only one instance of a broader problem in three-way three-mode data, namely the problem that in quite a few cases the substantive meaning of an element of some data mode does not remain constant across different $\mathbf{X}_k$ data slices. This problem, which implies that data entries of different data slices can no longer be meaningfully compared, has been analyzed extensively by Van Mechelen and Smilde [28]. For example, in the sensory profiling case we argued that the

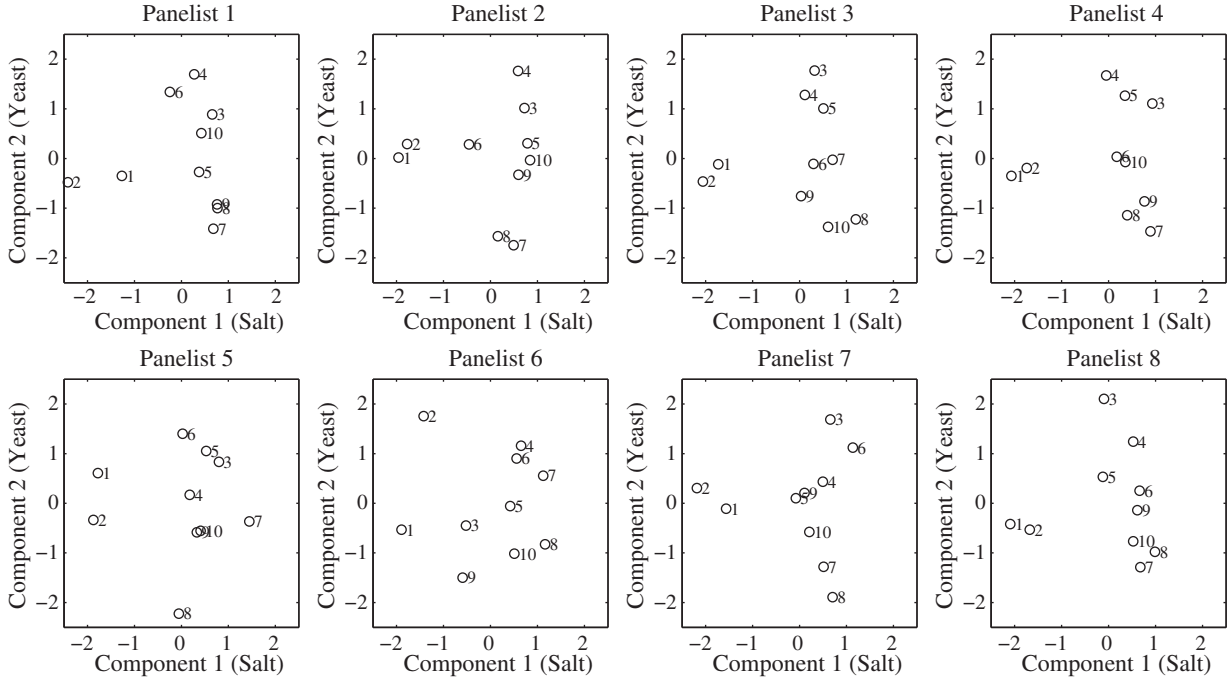


Fig. 3. Sample component score ($\mathbf{A}_k$) plots for the SCA-IND model with two components for the bread data.

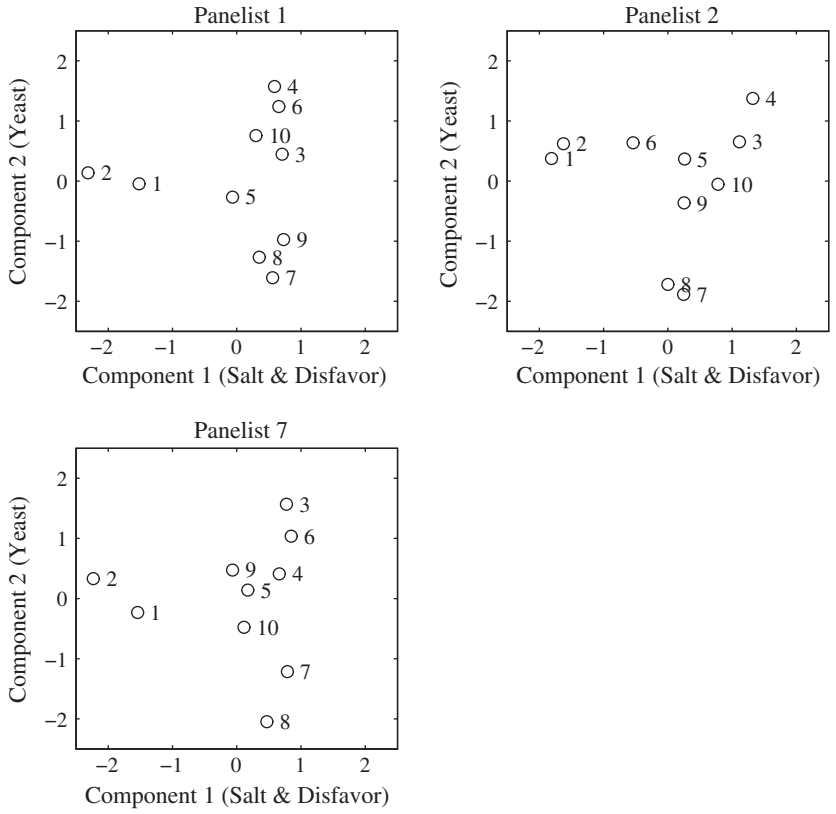


Fig. 4. Sample component score ($\mathbf{A}_k$) plots for the panelists in cluster 1 of the clusterwise SCA-IND model with two clusters and two components for the bread data.

meaning of a particular bread sample may not remain constant across the $\mathbf{X}_k$ slices because that sample may be experienced differently by different panelists, since, as was mentioned before, the experience of food is a highly subjective matter. As a second example, one can think of longitudinal (time × metabolite × batch) metabolomics data, with some batches pertaining to blood samples and others to urine samples. In this case, the biological meaning of a particular physical time point does not remain constant over batches because, after food ingestion, the metabolites will appear first in the blood and somewhat later in the urine. This lack of constancy is often

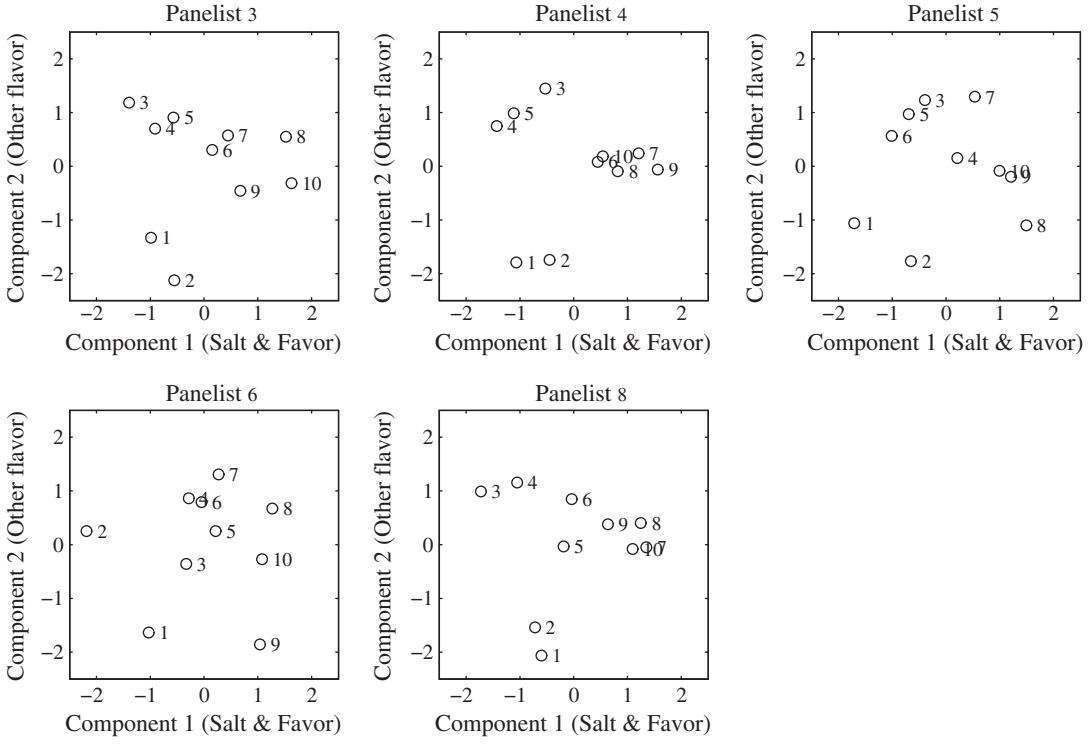


Fig. 5. Sample component score ($\mathbf{A}_k$) plots for the panelists in cluster 2 of the clusterwise SCA-IND model with two clusters and two components for the bread data.

referred to as an alignment problem, and may in some cases be dealt with by PARAFAC2 (see, e.g., [29]). Therefore, it is no surprise that PARAFAC2 is equivalent to one of the SCA variants (i.e., SCA-PF2; see [8]). Hence, one may safely conclude that multiset methods indeed appear to be an interesting option to analyze three-way data that suffer from lack of constancy in meaning. On top of the PARAFAC2 solution, the present paper offers new and interesting possibilities, i.e., the use of the other SCA variants on the one hand and clusterwise SCA on the other hand.

Secondly, in this paper we did not consider a method for three-way data that allows for qualitative differences only, simply because such a method does not yet exist. Such a method would cluster the elements of one of the three modes (e.g., panelists) and model the data within each cluster with, for example, a CP model (allowing for differences in saliencies only). This method may prove to be extremely useful to provide parsimonious accounts of qualitative differences.

References

[1] P.M. Kroonenberg, Applied Multiway Data Analysis, Wiley, Hoboken, NJ, 2008.
[2] R. Bro, E.M. Qannari, H.A.L. Kiers, T. Næs, M.B. Frøst, Multi-way models for sensory profiling data, Journal of Chemometrics 22 (2008) 36–45, http://dx.doi.org/10.1002/cem.1097.
[3] J.D. Carroll, J.J. Chang, Analysis of individual differences in multidimensional scaling via an N-way generalization of 'Eckart–Young' decomposition, Psychometrika 35 (1970) 283–319, http://dx.doi.org/10.1007/BF02310791.
[4] R.A. Harshman, Foundations of the PARAFAC procedure: models and conditions for an "explanatory" multi-modal factor analysis, UCLA Working Papers in Phonetics 16 (1970) 1–84.
[5] F.L. Hitchcock, The expression of a tensor or a polyadic as a sum of products, Journal of Mathematics and Physics 6 (1927) 164–189.
[6] L.R. Tucker, Some mathematical notes on three-mode factor analysis, Psychometrika 31 (1966) 279–311.
[7] K. De Roover, E. Ceulemans, M.E. Timmerman, How to perform multiblock component analysis in practice, Behavior Research Methods 44 (2012) 41–56, http://dx.doi.org/10.3758/s13428-011-0129-1.
[8] M.E. Timmerman, H.A.L. Kiers, Four simultaneous component models of multivariate time series from more than one subject to model intraindividual and interindividual differences, Psychometrika 86 (1) (2003) 105–122.
[9] K. De Roover, E. Ceulemans, M.E. Timmerman, P. Onghena, A clusterwise simultaneous component method for capturing within-cluster differences in component variances and correlations, British Journal of Mathematical and Statistical Psychology 86 (2013) 81–102, http://dx.doi.org/10.1111/j.2044-8317.2012.02040.x.
[10] K. De Roover, E. Ceulemans, M.E. Timmerman, K. Vansteelandt, J. Stouten, P. Onghena, Clusterwise simultaneous component analysis for analyzing structural differences in multivariate multiblock data, Psychological Methods 17 (1) (2012) 100–119.
[11] R. Bro, Multi-Way Analysis in the Food Industry: Models, Algorithms, and Applications. (Ph.D. thesis) University of Amsterdam, 1998.
[12] R. Bro, A.K. Smilde, Centering and scaling in component analysis, Journal of Chemometrics 17 (2003) 16–33, http://dx.doi.org/10.1002/cem.773.
[13] R.A. Harshman, M.E. Lundy, Data preprocessing and the extended PARAFAC model, in: H.G. Law, C.W. Snyder Jr., J.A. Hattie, R.P. McDonald (Eds.), Research Methods for Multimode Data Analysis, Praeger, 1984, pp. 122–215.
[14] R.A. van den Berg, H.C.J. Hoefsloot, J.A. Westerhuis, A.K. Smilde, M. van der Werf, Centering, scaling, and transformations: improving the biological information content of metabolomics data, BMC Genomics 7 (2006) 142, http://dx.doi.org/10.1186/1471-2164-7-142.
[15] M.E. Timmerman, Multilevel component analysis, British Journal of Mathematical and Statistical Psychology 59 (2006) 301–320, http://dx.doi.org/10.1348/000711005X67599.
[16] A. Stegeman, On uniqueness conditions for Candecomp/Parafac and Indscal with full column rank in one mode, Linear Algebra and its Applications 431 (2009) 211–227, http://dx.doi.org/10.1016/j.laa.2009.02.025.
[17] E. Ceulemans, M.E. Timmerman, H.A.L. Kiers, The CHULL procedure for selecting among multilevel component solutions, Chemometrics and Intelligent Laboratory Systems 106 (1) (2011) 12–20, http://dx.doi.org/10.1016/j.chemolab.2010.08.001.
[18] K. De Roover, E. Ceulemans, M.E. Timmerman, J.B. Nezlek, P. Onghena, Modeling differences in the dimensionality of multiblock data by means of clusterwise simultaneous component analysis, Psychometrika (2013), http://dx.doi.org/10.1007/s11336-013-9318-4, (advance online publication).
[19] K. Bulteel, T.F. Wilderjans, F. Tuerlinckx, E. Ceulemans, CHull as an alternative to AIC and BIC in the context of mixtures of factor analyzers, Behavior Research Methods (2013), http://dx.doi.org/10.3758/s13428-012-0293-y, (advance online publication).
[20] T.F. Wilderjans, E. Ceulemans, K. Meers, CHull: a generic convex-hull-based model selection method, Behavior Research Methods 45 (2013) 1–15, http://dx.doi.org/10.3758/s13428-012-0238-5.
[21] E. Ceulemans, H.A.L. Kiers, Selecting among three-mode principal component models of different types and complexities: a numerical convex hull based method, British Journal of Mathematical and Statistical Psychology 59 (2006) 133–150, http://dx.doi.org/10.1348/000711005X64817.
[22] E. Ceulemans, H.A.L. Kiers, Discriminating between strong and weak structures in three-mode principal component analysis, British Journal of Mathematical and Statistical Psychology 62 (2009) 601–620, http://dx.doi.org/10.1348/000711008X369474.
[23] R. Bro, H.A.L. Kiers, A new efficient method for determining the number of components in PARAFAC models, Journal of Chemometrics 17 (2003) 274–286, http://dx.doi.org/10.1002/cem.801.
[24] M.E. Timmerman, H.A.L. Kiers, Three-mode principal component analysis: choosing the numbers of components and sensitivity to local optima, British Journal of Mathematical and Statistical Psychology 53 (2000) 1–16, http://dx.doi.org/10.1348/000711000159132.
[25] C.A. Andersson, R. Bro, The N-way toolbox for MATLAB, Chemometrics and Intelligent Laboratory Systems 52 (2000) 1–4, http://dx.doi.org/10.1016/S0169-7439(00)00071-X.
[26] L.R. Tucker, A Method for Synthesis of Factor Analysis Studies (Personnel Research section Rep. No. 984), Washington, DC, Department of the Army, 1951.
[27] U. Lorenzo-Seva, J.M.F. ten Berge, Tucker's congruence coefficient as a meaningful index of factor similarity, Methodology 2 (2) (2006) 57–64, http://dx.doi.org/10.1027/1614-1881.2.2.57.
[28] I. Van Mechelen, A.K. Smilde, Comparability problems in the analysis of multiway data, Chemometrics and Intelligent Laboratory Systems 106 (2011) 2–11, http://dx.doi.org/10.1016/j.chemolab.2010.08.006.
[29] R. Bro, C.A. Anderson, H.A.L. Kiers, Parafac 2 — part II. Modeling chromatographic data with retention time shifts, Journal of Chemometrics 13 (1999) 295–309, http://dx.doi.org/10.1002/(SICI)1099-128X(199905/08)13:3/4<295::AID-CEM547>3.0.CO;2-Y.