British Journal of Mathematical and Statistical Psychology (2013), 66, 81–102
© 2012 The British Psychological Society



The British Psychological Society

www.wileyonlinelibrary.com

A clusterwise simultaneous component method for capturing within-cluster differences in component variances and correlations

Kim De Roover[1]*, Eva Ceulemans[1], Marieke E. Timmerman[2] and Patrick Onghena[1]

[1]Katholieke Universiteit Leuven, Belgium
[2]University of Groningen, The Netherlands

This paper presents a clusterwise simultaneous component analysis for tracing structural differences and similarities between data of different groups of subjects. This model partitions the groups into a number of clusters according to the covariance structure of the data of each group and performs a simultaneous component analysis with invariant pattern restrictions (SCA-P) for each cluster. These restrictions imply that the model allows for between-group differences in the variances and the correlations of the cluster-specific components. As such, clusterwise SCA-P is more flexible than the earlier proposed clusterwise SCA-ECP model, which imposed equal average cross-products constraints on the component scores of the groups that belong to the same cluster. Using clusterwise SCA-P, a finer-grained, yet parsimonious picture of the group differences and similarities can be obtained. An algorithm for fitting clusterwise SCA-P solutions is presented and its performance is evaluated by means of a simulation study. The value of the model for empirical research is illustrated with data from psychiatric diagnosis research.

1. Introduction

Behavioural researchers often examine whether the underlying structure of a set of variables differs between known groups of subjects. To this end one may, firstly, perform a separate principal component analysis (PCA: Jolliffe, 1986; Pearson, 1901) for each group (e.g., McCrae & Costa, 1997). This implies that, for each group, the variables are reduced to a smaller number of components (see Table 1) which explain as much of the variance in the data as possible. The resulting group-specific loading matrices represent the relations between the variables and the components and yield insight into the structure of the variables within the different groups. This approach leaves plenty

*Correspondence should be addressed to Kim De Roover, Department of Educational Sciences, Katholieke Universiteit Leuven, Andreas Vesaliusstraat 2, B-3000 Leuven, Belgium (e-mail: Kim.DeRoover@ppw.kuleuven.be).

DOI:10.1111/j.2044-8317.2012.02040.x

Table 1. Restrictions imposed by the different component methods for modelling the within-group structure of multivariate data from different groups.

Method	Component loadings	Component variances	Component correlations
PCA by group (Jolliffe, 1986)	Free	Free	Free
Clusterwise SCA-P (this paper)	Equal for all groups in the same cluster	Free	Free
Clusterwise SCA-ECP (De Roover *et al.*, 2011)	Equal for all groups in the same cluster	Equal for all groups in the same cluster	Equal for all groups in the same cluster
SCA-P (Timmerman & Kiers, 2003)	Equal for all groups	Free	Free
SCA-ECP (Timmerman & Kiers, 2003)	Equal for all groups	Equal for all groups	Equal for all groups

of freedom to trace differences between the groups, but it may be hard to gain insight into the structural similarities. Besides, when the number of groups is large, comparing all the loading matrices is practically infeasible.

Secondly, one may perform simultaneous component analysis (SCA: Kiers, 1990; Kiers & ten Berge, 1994a; Timmerman & Kiers, 2003). In SCA, the data of all groups are modelled simultaneously, assuming that the same components underlie the data of the different groups and thus that a common loading matrix can be used to summarize the data. As such, SCA is much more parsimonious than the separate PCA strategy and sheds light on the structural similarities of the groups. On the downside, having only one loading matrix for all groups makes it hard to trace structural differences between the groups. Specifically, the only differences that can be detected are differences between groups in the variances (across subjects within a group) of and the correlations between the components. Which of these differences can be uncovered depends on the SCA variant used (Timmerman & Kiers, 2003). In the most constrained variant, called SCA-ECP (i.e., with equal average cross-products constraints), component correlations and variances must be equal across the groups, which implies that there is no room for structural differences between the groups (see Table 1). Using the most general variant, SCA-P (i.e., with invariant pattern constraints), one can trace differences in component correlations as well as variances (see Table 1).

Recently, a generic modelling strategy that encompasses both SCA and separate PCA as special cases was proposed that deals with the disadvantages of these approaches: clusterwise SCA (De Roover *et al.*, 2011). In clusterwise SCA, the different groups of subjects are assigned to a limited number of mutually exclusive clusters and the data within each cluster are modelled with SCA. Thus, groups that are classified into to the same cluster share a loading matrix, whereas groups that are assigned to different clusters have different loading matrices. Note that, although factor-analytic alternatives exist for PCA and SCA (e.g., Dolan, Oort, Stoel, & Wicherts, 2009; Lawley & Maxwell, 1962), no factor-analytic counterpart exists for clusterwise SCA, that is, no model is available that provides a clustering of the groups of subjects based on the differences and similarities in factor loading structure.

Within the clusterwise SCA framework, one specific model had already been developed: clusterwise SCA-ECP, which uses the most constrained SCA variant, SCA-ECP, within each cluster. Hence, clusterwise SCA-ECP imposes a very strict concept

of structural similarity (see Table 1). First, within each cluster, the correlations among the component scores are constrained to be equal for all groups. This is less ideal if some groups have the same component structure, but differ strongly with respect to component correlations. In such cases, clusterwise SCA-ECP would require additional clusters to adequately summarize the data.

Second, in clusterwise SCA-ECP the variances of the component scores are constrained to be one for each group. This is too restrictive if one is interested in modelling between-group differences in variability across subjects. For example, when a personality questionnaire is administered to several groups of subjects, the 'neuroticism' personality trait may underlie the data of all groups, but the variance of this component can be different for groups of healthy persons and clinical groups. In this case, thoughtless application of clusterwise SCA-ECP could even result in inappropriate model estimates. To avoid such problems, the model could be fitted to autoscaled data (i.e., data in which each variable is standardized by group). However, this type of preprocessing has the clear disadvantage that the between-group differences in variability are lost.

To meet the need for a clusterwise SCA model that allows for within-cluster differences in component variances and correlations, we introduce clusterwise SCA-P which models the data within a cluster with SCA-P. Thus, compared to clusterwise SCA-ECP, clusterwise SCA-P is based on a less strict concept of structural similarity which only concerns the component loadings (see Table 1).

The remainder of this paper is organized as follows. In Section 2 the clusterwise SCA-ECP model is recapitulated and the new clusterwise SCA-P model is introduced. Section 3 describes the loss function and an algorithm for clusterwise SCA-P analysis, followed by a model selection heuristic. In Section 4 an extensive simulation study is presented to evaluate the performance of this algorithm and model selection heuristic. In Section 5 clusterwise SCA-P is applied to data from psychiatric diagnosis research. In Section 6 we conclude with a few points of discussion, including directions for future research.

2. Model

2.1. Data and preprocessing

In this paper we assume that for each of the K groups under study, an I_k (subjects) $\times$ J (variables) data matrix $\mathbf{X}_k$ ($k = 1, \ldots, K$) is available.[1] As the focus is on between-group differences in within-group structure, it is essential that the data of each group are centred by variable, implying that between-group differences in variable means are removed from the data. Moreover, to eliminate arbitrary scale differences between variables, the variables may be standardized across the groups, thus retaining the information on between-group differences in within-group variability. Because the latter standardization eases the interpretation of the loadings of clusterwise SCA-P (i.e., they can be scaled such that they are correlations between components and variables in the case of orthogonal components; see Appendix), it will be assumed in what follows that data are standardized across groups.

[1]Note that fully crossed or three-way, three-mode data (for an introduction, see Kroonenberg, 2008) are a special case of the hierarchical data structure described above, in which all the groups consist of the same subjects – for example, the same subjects measured under different conditions.

2.2. Clusterwise SCA-ECP

Clusterwise SCA-ECP (De Roover *et al.*, 2011; De Roover, Ceulemans, & Timmerman, 2011) captures between-group differences in underlying structure by partitioning the K groups into C clusters and modelling the data of the groups within each cluster with SCA-ECP (Timmerman & Kiers, 2003). The number of components Q of the cluster-specific SCA-ECP models is assumed to be the same across the clusters, which means that clusterwise SCA-ECP aims to find differences in the nature of the underlying dimensions rather than differences in the number of dimensions.

Formally, the data matrix $\mathbf{X}$, which is obtained by vertically concatenating the K matrices $\mathbf{X}_k$, is decomposed into a binary partition matrix $\mathbf{P}$ of dimension $K \times C$, K component score matrices $\mathbf{F}_k$ of dimension $I_k \times Q$, and C cluster loading matrices $\mathbf{B}^c$ of dimension $J \times Q$. Specifically, the decomposition rule reads as

$$\mathbf{X}_k = \sum_{c=1}^{C} p_{kc}\mathbf{F}_k\left(\mathbf{B}^c\right)' + \mathbf{E}_k = \mathbf{F}_k\left(\mathbf{B}^c_{k\in c}\right)' + \mathbf{E}_k, \tag{1}$$

where p_{kc} denotes the entries of the binary partition matrix $\mathbf{P}$ ($K \times C$), which equal one when group k is assigned to cluster c ($c = 1, \ldots, C$) and zero otherwise, and $\mathbf{E}_k$ ($I_k \times J$) denotes the matrix of residuals. The columns of each component score matrix $\mathbf{F}_k$ are restricted to have a variance of one; furthermore, the correlations between the columns of $\mathbf{F}_k$ (i.e., the cluster-specific components) must be equal for the groups that are assigned to the same cluster. These restrictions imply that clusterwise SCA-ECP leaves no room for between-group differences in component variances and correlations within a cluster. If such differences were present in the data, additional clusters would be required to adequately model these differences. To facilitate the interpretation of the components, the cluster-specific SCA-ECP solutions can be freely rotated using an orthogonal (e.g., varimax: Kaiser, 1958), or oblique (e.g., Harris-Kaiser independent cluster or HKIC rotation: Harris & Kaiser, 1964; Kiers & ten Berge, 1994b) rotation criterion.

To illustrate the characteristics and interpretation of the clusterwise SCA-ECP model, we make use of the hypothetical data matrix $\mathbf{X}$ in Table 2. These data pertain to the amount of overt aggression (e.g., pushing another person), relational aggression (e.g., spreading gossip about someone) and prosocial behaviour (e.g., helping another person) that children of six different ages (7–12 years old) display at school and at home. The data are columnwise centred by age group and standardized over all groups. As a consequence, we observe between-group differences in variability: for instance, the younger children vary less on the six variables than the older children.

The clusterwise SCA-ECP solution with three clusters and two components explains 99.7% of the overall variance of $\mathbf{X}$. Note that, because of the considerable differences between the age groups in variability, $\mathbf{X}$ could only be fitted perfectly with clusterwise SCA-ECP if as many clusters as age groups are formed (i.e., $C = K$). The partition matrix $\mathbf{P}$ of the solution with three clusters and two components is displayed in Table 3 and the cluster loading matrices in Table 4. From Table 3 it can be derived that each of the three clusters consists of two consecutive age groups. From the varimax rotated cluster loading matrix $\mathbf{B}^1$ in Table 4 it can be read that for the 7- and 8-year-olds the behaviour at home has high positive or negative loadings on the first component, whereas the behaviour in school loads strongly on the second component. Hence, the components can be labelled 'home behaviour' and 'school behaviour'. For cluster 2, containing ages

Table 2. Hypothetical data matrix **X** with the (rounded off) scores of school children of six different ages on six variables concerning aggressive and prosocial behaviour, after standardization over groups. 'O' denotes overt aggression, 'R' relational aggression and 'P' prosocial behaviour, while 'h' and 's' refers to home and school, respectively.

Age	Subj.	Oh	Os	Rh	Rs	Ph	Ps	Age	Subj.	Oh	Os	Rh	Rs	Ph	Ps
7	1	0.0	0.1	0.0	0.1	0.0	−0.1	10	1	−0.9	−0.2	−0.9	−0.2	0.9	0.2
	2	−0.6	−0.4	−0.6	−0.4	0.6	0.4		2	2.3	2.0	2.3	2.0	−2.2	−2.0
	3	0.3	−1.4	0.3	−1.4	−0.3	1.4		3	−0.9	−1.8	−0.9	−1.8	0.9	1.8
	4	−0.8	0.8	−0.8	0.8	0.8	−0.8		4	−0.3	−0.6	−0.3	−0.6	0.3	0.6
	5	0.5	0.1	0.5	0.1	−0.5	−0.1		5	0.2	0.4	0.2	0.4	−0.2	−0.4
	6	1.3	0.8	1.3	0.8	−1.3	−0.8		6	0.6	−0.1	0.6	−0.1	−0.6	0.1
	7	−0.7	0.1	−0.7	0.1	0.7	−0.1		7	−0.9	0.3	−0.9	0.3	0.9	−0.3
8	1	−1.0	−0.6	−1.0	−0.6	1.0	0.6	11	1	0.3	0.3	0.3	0.3	0.5	0.5
	2	0.2	−0.2	0.2	−0.2	−0.2	0.2		2	0.1	0.1	0.1	0.1	0.4	0.4
	3	−0.4	−1.3	−0.4	−1.3	0.4	1.3		3	−0.4	−0.4	−0.4	−0.4	1.7	1.7
	4	1.4	0.0	1.4	0.0	−1.4	0.0		4	1.7	1.7	1.7	1.7	−0.5	−0.5
	5	−0.3	−0.1	−0.3	−0.1	0.3	0.1		5	−1.7	−1.7	−1.7	−1.7	−1.5	−1.5
	6	−0.6	1.8	−0.6	1.8	0.6	−1.8		6	−1.8	−1.8	−1.8	−1.8	−0.1	−0.1
	7	−0.4	0.6	−0.4	0.6	0.4	−0.6		7	1.2	1.2	1.2	1.2	−1.9	−1.9
	8	1.0	−0.2	1.0	−0.2	−1.0	0.2		8	0.5	0.5	0.5	0.5	1.3	1.3
9	1	0.0	0.0	0.0	0.0	0.0	0.0	12	1	−0.8	−0.8	−0.8	−0.8	1.2	1.2
	2	0.3	−0.4	0.3	−0.4	−0.3	0.4		2	1.0	1.0	1.0	1.0	−0.2	−0.2
	3	−1.7	−1.4	−1.7	−1.4	1.7	1.4		3	0.9	0.9	0.9	0.9	−2.1	−2.1
	4	−0.3	−0.1	−0.3	−0.1	0.3	0.1		4	1.8	1.8	1.8	1.8	0.7	0.7
	5	−1.2	−1.1	−1.2	−1.1	1.2	1.1		5	−1.7	−1.7	−1.7	−1.7	−1.0	−1.0
	6	1.9	1.4	1.9	1.4	−1.9	−1.4		6	0.1	0.1	0.1	0.1	1.7	1.7
	7	0.2	−0.2	0.2	−0.2	−0.2	0.2		7	−1.3	−1.3	−1.3	−1.3	−0.2	−0.2
	8	0.6	0.3	0.6	0.3	−0.6	−0.3								
	9	0.2	1.5	0.2	1.5	−0.2	−1.5								

Table 3. Partition matrix **P** of the clusterwise SCA-ECP decomposition with three clusters and two components of **X** in Table 2 and of the clusterwise SCA-P decomposition with two clusters and two components.

	Clusterwise SCA-ECP			Clusterwise SCA-P	
Groups	Cluster 1	Cluster 2	Cluster 3	Cluster 1	Cluster 2
7 years	1	0	0	1	0
8 years	1	0	0	1	0
9 years	0	1	0	1	0
10 years	0	1	0	1	0
11 years	0	0	1	0	1
12 years	0	0	1	0	1

Table 4. Cluster loading matrices of the clusterwise SCA-ECP and clusterwise SCA-P decompositions of **X** in Table 2. 'OA' denotes overt aggression, 'RA' relational aggression and 'PB' prosocial behaviour.

	Clusterwise SCA-ECP					
	Cluster 1		Cluster 2		Cluster 3	
	Home behaviour	School behaviour	Home behaviour	School behaviour	Aggression	Prosocial behaviour
OA home	0.75	0.00	1.01	0.00	1.19	0.00
OA school	0.00	0.78	0.00	0.99	1.18	0.00
RA home	0.75	0.00	1.01	0.00	1.19	0.00
RA school	0.00	0.78	0.00	0.99	1.18	0.00
PB home	−0.74	0.00	−1.01	0.00	0.00	1.19
PB school	0.00	−0.77	0.00	−0.99	.00	1.19

	Clusterwise SCA-P			
	Cluster 1		Cluster 2	
	Home behaviour	School behaviour	Aggression	Prosocial behaviour
OA home	0.90	0.00	1.19	0.00
OA school	0.00	0.90	1.18	0.00
RA home	0.90	0.00	1.19	0.00
RA school	0.00	0.90	1.18	0.00
PB home	−0.89	0.00	0.00	1.20
PB school	0.00	−0.89	0.00	1.19

9 and 10, the HKIC rotated loadings[2] in Table 4 display the same structure (home behaviour versus school behaviour), but the component scores are strongly correlated (i.e., correlation of .80). The varimax rotated loadings of cluster 3, which consists of the 11- and 12-year-olds, reveal a different pattern: the components refer to the type of

[2]In the case of obliquely rotated components, the term 'pattern matrix' (rather than 'loading matrix') is often used to refer to the weight matrix for the components. For the sake of simplicity, we will continue using the term 'loadings'.

behaviour instead of the context, with overt and relational aggression constituting the first component (labelled 'aggression') and prosocial behaviour the second component (labelled 'prosocial behaviour').

2.3. Clusterwise SCA-P: A more general clusterwise SCA model

We propose clusterwise SCA-P to model the between-group differences in the component variances and correlations in a more comprehensive and/or parsimonious way than clusterwise SCA-ECP, where parsimony refers to the number of clusters and thus the number of loading matrices that are to be inspected and compared after the analysis. Clusterwise SCA-P is built on the same principle as clusterwise SCA-ECP: a clustering of the groups – which is represented in a partition matrix $\mathbf{P}$ – and a separate SCA with Q components on the data of each cluster, yielding a different loading matrix $\mathbf{B}^c$ for each cluster c. In clusterwise SCA-P, the component model within each cluster is an SCA-P model, however, which implies that the variances and correlations of the component scores may differ across the groups belonging to the same cluster. Thus, both models share the same decomposition rule (equation (1)), but clusterwise SCA-P imposes no active constraints on the component scores (collected in $\mathbf{F}_k$ ($k = 1, \ldots, K$)); to partly identify the solution the variance of each cluster-specific component is scaled to unity across all groups within a cluster.

The cluster-specific SCA-P models can be orthogonally or obliquely rotated within each cluster to make them easier to interpret. Also, the loadings and component scores of a clusterwise SCA-P model can be rescaled such that the loadings can be read as correlations between components and variables across all clusters, in the case of orthogonal components. Given this rescaling, the sizes of the component scores are no longer comparable over clusters, however. The pros and cons of the different scaling options are discussed in the Appendix.

The hypothetical data in Table 2 are also used to illustrate the properties of the clusterwise SCA-P model. $\mathbf{X}$ can be perfectly reconstructed by a clusterwise SCA-P model with two clusters and two components. The partition matrix $\mathbf{P}$ in Table 3 reveals that ages 7–10 are now combined into one cluster, while ages 11 and 12 form the second cluster. The cluster loading matrices $\mathbf{B}^c$ in Table 4 – rotated obliquely using the HKIC criterion for the first cluster and orthogonally according to the varimax criterion for the second – show that the components for the cluster of younger children can again be interpreted as 'home behaviour' versus 'school behaviour', whereas the components for the cluster of older children can be labelled 'aggression' and 'prosocial behaviour'.

The variances and correlations of the component scores for each age group are presented in Table 5. These variances and correlations give additional insight into the data. For instance, one can derive that in cluster 1, the variability on home and school behaviour seems to increase with age. Furthermore, the component correlations in Table 5 indicate that the home and school behaviour components are uncorrelated for the two youngest age groups but highly correlated for the 9- and 10-year-olds.

We conclude that the clusterwise SCA-P solution fits the hypothetical data slightly better (100% variance explained versus 99.7%) than the clusterwise SCA-ECP solution and is more parsimonious in that only two clusters are needed. Indeed, ages 9 and 10 have the same loading structure as ages 7 and 8, but differ with respect to the correlation between these components. Because clusterwise SCA-P can handle such differences in correlations, these four groups are assigned to the same cluster in the clusterwise SCA-P solution, while in clusterwise SCA-ECP two separate clusters had to be formed. In

Table 5. Variances and correlations of component score matrices $\mathbf{F}_k$ of the clusterwise SCA-P decomposition with two clusters and two components of $\mathbf{X}$ in Table 2.

Cluster	Group	Components	Variances	Correlations
1	7 years	home behaviour	0.6	.05
		school behaviour	0.6	
	8 years	home behaviour	0.8	−.05
		school behaviour	0.9	
	9 years	home behaviour	1.2	.82
		school behaviour	1.1	
	10 years	home behaviour	1.4	.78
		school behaviour	1.4	
2	11 years	aggression	1.0	−.03
		prosocial behaviour	1.0	
	12 years	aggression	1.0	.03
		prosocial behaviour	1.1	

addition, clusterwise SCA-P sheds light on the between-group differences in variability within a cluster.

3. Data analysis

3.1. Loss function

For given numbers of clusters C and components Q and data matrices $\mathbf{X}_k$, the aim of a clusterwise SCA-P analysis is to find the partition matrix $\mathbf{P}$, the component score matrices $\mathbf{F}_k$ and the cluster loading matrices $\mathbf{B}^c$ that minimize the loss function:

$$L = \sum_{c=1}^{C} \sum_{k=1}^{K} p_{kc} \left\| \mathbf{X}_k - \mathbf{F}_k \left(\mathbf{B}^c\right)' \right\|. \tag{2}$$

Note that on the basis of the loss function value L, one can compute the percentage of variance in the data that is accounted for by the clusterwise SCA-P solution:

$$VAF(\%) = \frac{\|\mathbf{X}\|^2 - L}{\|\mathbf{X}\|^2} \times 100 \tag{3}$$

3.2. Algorithm

In clusterwise SCA-P analysis, we follow a deterministic perspective in that no distributional assumptions are made about the component scores, loadings, cluster memberships, and residuals, as is done in stochastic approaches (e.g., mixture modelling: McLachlan & Peel, 2000). As such, no likelihood function can be specified and, as is common for deterministic models, an alternating least squares (ALS) algorithm is used to fit a clusterwise SCA-P solution with C clusters and Q components to a data matrix $\mathbf{X}$. This algorithm was implemented in Matlab R2010a and the m-files can be obtained freely from the first author.

The ALS procedure alternately updates each row of the partition matrix – that is, the cluster membership of one group – conditional upon the other rows of $\mathbf{P}$ and thus

upon the cluster memberships of the other groups. Specifically, the clusterwise SCA-P algorithm consists of five steps:

1. *Randomly initialize the partition matrix* **P**. Initialize the partition matrix **P** by randomly assigning the K groups to one of the C clusters, where the probability of assigning a group to a certain cluster is equal for all clusters. If one of the clusters is empty, repeat this procedure until all clusters contain at least one group.
2. *Estimate the component score matrices* $\mathbf{F}_k$ *and the cluster loading matrices* $\mathbf{B}^c$. For each cluster c, estimate $\mathbf{B}^c$ and the corresponding $\mathbf{F}^c$ matrix by performing SCA-P on the data matrix $\mathbf{X}^c$, where $\mathbf{F}^c$ and $\mathbf{X}^c$ consist of the component score matrices $\mathbf{F}_k$ and the data matrices $\mathbf{X}_k$ of all the groups that belong to cluster c, respectively. Specifically, given the singular value decomposition of $\mathbf{X}^c$ into $\mathbf{U}^c$, $\mathbf{S}^c$ and $\mathbf{V}^c$ with $\mathbf{X}^c = \mathbf{U}^c\mathbf{S}^c(\mathbf{V}^c)'$, least squares estimates of $\mathbf{F}^c$ and $\mathbf{B}^c$ are obtained by $\mathbf{F}^c = \sqrt{I^c}\mathbf{U}^c_Q$ and $\mathbf{B}^c = \sqrt{\frac{1}{I^c}}\mathbf{V}^c_Q\mathbf{S}^c_Q$. $\mathbf{U}^c_Q$ and $\mathbf{V}^c_Q$ are the first Q columns of $\mathbf{U}^c$ and $\mathbf{V}^c$ respectively, $\mathbf{S}^c_Q$ consists of the first Q columns and the first Q rows of $\mathbf{S}^c$. I^c denotes the total number of subjects in cluster c.
3. *For each group* k*, re-estimate row* k *of the partition matrix* **P** *conditionally on the other rows of* **P** *and update each* $\mathbf{B}^c$ *and* $\mathbf{F}_k$ *accordingly.* Reassign group k to each of the C clusters and compute the $\mathbf{B}^c$ and $\mathbf{F}_k$ matrices for each of the C resulting clusterings, as described in step 2, together with the corresponding loss function values. Subsequently, group k is placed in the cluster for which L is minimal and the corresponding estimates of the $\mathbf{B}^c$ and $\mathbf{F}_k$ matrices are retained.
4. *When one of the C clusters is empty, move the group that fits its current cluster least to the empty cluster*. Re-estimate each $\mathbf{B}^c$ and $\mathbf{F}_k$ as described in step 2.
5. *Repeat steps 3 and 4* until the decrease in the loss function value L for the current iteration is smaller than the convergence criterion of 1×10^{-6}.

To reduce the probability of ending up in a local minimum, it is advisable to use a multistart procedure with different random initializations of the partition matrix **P**.

3.3. Model selection

When performing clusterwise SCA analysis, two model selection questions have to be answered. First, which model is most appropriate for the substantive question at hand: clusterwise SCA-ECP or clusterwise SCA-P? Second, given one of these models, how many clusters and components should be used?

3.3.1. Applying clusterwise SCA-ECP or clusterwise SCA-P

To choose whether clusterwise SCA-ECP or clusterwise SCA-P is the most appropriate approach for a specific data analysis problem, one may consider the following three questions:

1. Are you interested in between-group differences in the variability of the observed variables and the resulting components?
2. Should any differences in component variability within groups be captured in different clusters, or should those differences be captured within clusters (i.e., do you want groups with the same loading structure but with different component variances to be assigned to the same cluster)?

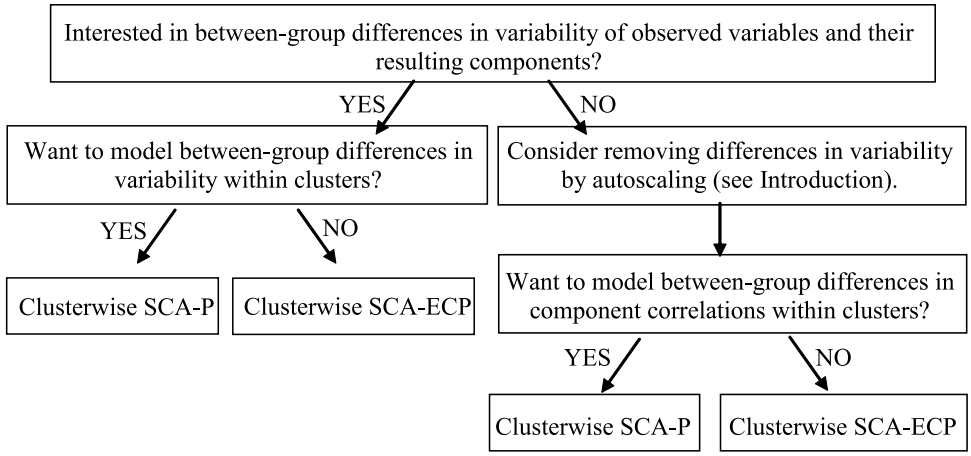


Figure 1. Decision tree for making the choice between applying clusterwise SCA-ECP and clusterwise SCA-P for a specific data analysis problem.

3. Should any differences in component correlations across groups be captured in different clusters, or should those differences be captured within clusters (i.e., do you want groups with the same loading structure but with different component correlations to be assigned to the same cluster)?

These three questions make up a decision tree, depicted in Figure 1, that guides the user to the most appropriate approach.

3.3.2. Selecting the number of clusters and components

When performing clusterwise SCA-(EC)P analysis, the number of underlying clusters C and components Q is usually unknown. To determine appropriate C- and Q-values, one may apply the following model selection procedure (see De Roover, Ceulemans, & Timmerman, 2011, for more details). First, solutions are estimated using several values for C and Q. Next, to select the most appropriate number of clusters, called C^{best}, one computes – given the different Q-values – the following scree ratio $sr_{(C|Q)}$ for all C-values for which $C_{\min} < C < C_{\max}$, with $C_{\min}$ and $C_{\max}$ being the smallest and largest number of clusters considered, respectively:

$$sr_{(C|Q)} = \frac{\text{VAF}_{C|Q} - \text{VAF}_{C-1|Q}}{\text{VAF}_{C+1|Q} - \text{VAF}_{C|Q}} \tag{4}$$

where $\text{VAF}_{C|Q}$ denotes the variance-accounted-for percentage of the solution with C clusters and Q components (for a general description of the scree ratio, see Ceulemans & Kiers, 2006). The C-value which has the highest average scree ratio across the different Q-values is retained as C^{best}. Finally, to assess the best number of components Q^{best}, similar scree ratios are calculated, with the number of clusters equal to C^{best}:

$$sr_{(Q|C^{\text{best}})} = \frac{\text{VAF}_{Q|C^{\text{best}}} - \text{VAF}_{Q-1|C^{\text{best}}}}{\text{VAF}_{Q+1|C^{\text{best}}} - \text{VAF}_{Q|C^{\text{best}}}}. \tag{5}$$

The Q-value for which equation (5) is maximal is retained as Q^{best}.

4. Simulation studies

In this section, we first present an extensive simulation study in which the clusterwise SCA-P algorithm is evaluated with respect to sensitivity for local minima and goodness of recovery. In a second simulation study, we examine whether the model selection procedure presented succeeds in selecting C and Q correctly.

4.1. Simulation study 1

4.1.1. Design and procedure

In this simulation study, seven factors were systematically varied in a complete factorial design, keeping the number of variables J fixed at 12:

(a) the *number of groups* K at 2 levels: 20, 40;
(b) the *number of subjects per group* I_k at 2 levels: $I_k \sim$U[30; 70], $I_k \sim$U[80; 120], with U denoting a uniform distribution;
(c) the *number of clusters* C at 2 levels: 2, 4;
(d) the *cluster size*, at 3 levels (see Brusco & Cradit, 2001; Steinley, 2003): equal (equal number of groups in each cluster); unequal with minority (10% of the groups in one cluster and the remaining groups distributed equally across the other clusters); unequal with majority (60% of the groups in one cluster and the remaining groups distributed equally across the other clusters);
(e) the *number of components* Q at 2 levels: 2, 4;
(f) the *error level* e, which is the expected proportion of error variance in the data matrices $\mathbf{X}_k$, at 3 levels: .00, .20, .40;
(g) the *amount of congruence between the cluster loading matrices* $\mathbf{B}^c$ at 3 levels: low, medium, and high, which respectively imply that the Tucker congruence coefficients (Tucker, 1951) between the corresponding components of the cluster loading matrices amount to .41, .72 and .93 on average, when these matrices are orthogonally Procrustes rotated to each other. The clustering of the groups is less distinct when the congruence between the cluster loading matrices is high.

These seven factors will be considered as random effects.

For each cell of the simulation design, 50 data matrices $\mathbf{X}$ were generated using the following procedure. Each component score matrix $\mathbf{F}_k$ was randomly sampled from a multivariate normal distribution, of which the mean vector consists of zeros, and of which the variance–covariance matrix was obtained by uniformly sampling the component correlations and variances between $-.5$ and .5 and between 0.25 and 1.75 respectively. To construct the partition matrix $\mathbf{P}$, the groups were randomly assigned to the clusters, making sure that each cluster had the correct size. The cluster loading matrices $\mathbf{B}^c$ were generated according to the procedure described by De Roover *et al.* (2011), where all loadings had values between -1 and 1. Subsequently, the proportion of variance accounted for by each cluster was manipulated by multiplying the cluster loading matrix of the cth cluster by $\sqrt{s^c \frac{I}{I^c}}$ where $s^c \sim$ U[.10, .90], subject to the restriction that all s^c-values sum to one. For each group k an error matrix $\mathbf{E}_k$ was randomly sampled from the standard normal distribution, and subsequently the cluster loading matrices $\mathbf{B}^c$ and the error matrices $\mathbf{E}_k$ were rescaled by multiplying these matrices by $\sqrt{e}$ and $\sqrt{1-e}$ respectively, such that the data contain the correct amount of error. Finally, $\mathbf{X}$ was obtained by computing the $\mathbf{X}_k$ matrices of the K groups as $\mathbf{F}_k(\mathbf{B}^c)' + \mathbf{E}_k$.

All 21,600 data matrices **X** were centred by group and columnwise standardized across all groups. Subsequently, the data matrices were analysed with the clusterwise SCA-P algorithm, using the correct *C*- and *Q*-values. The algorithm was run 25 times, each time using a different random start, and the best solution out of the 25 runs was retained. Additionally, the data matrices were also analysed with the clusterwise SCA-ECP algorithm, again using the correct *C* and *Q* as well as 25 random starts.

4.1.2. Results

4.1.2.1. Goodness of fit and sensitivity to local minima. To evaluate the sensitivity of the clusterwise SCA-P algorithm to local minima, the loss function value of the retained solution should be compared to that of the global minimum. This global minimum is unknown, however, for instance because the simulated data are perturbed with error. As a way out, we use the solution that results from seeding the algorithm with the true $\mathbf{F}_k$, $\mathbf{B}^c$ and **P** matrices as a proxy of the global minimum.

First, we evaluated whether the best-fitting solution out of the 25 randomly started runs from the multistart procedure had a higher loss function value than the proxy, which would imply that the retained solution is a local minimum for sure. The results indicate that this is the case for only 1 out of the 21,600 simulated data matrices (0.005%).

Furthermore, we determined what proportion of the 25 solutions resulting from the multistart procedure had a loss function value that was equal to that of the retained solution or to that of the proxy of the global minimum, whichever was the lowest. This proportion is referred to as the 'global minimum proportion'. On average, the global minimum proportion equals .96 with a standard deviation of 0.09, which implies that most of the runs ended in the retained solution.

To assess the effects of the different factors, we performed an analysis of variance with the global minimum proportion – the values of which were logit-transformed to improve normality – as the dependent variable. In this analysis the seven main effects and all possible two-way and higher-order interactions were included. Thus, 128 effects were tested, which implies that reporting the full ANOVA table would not be very insightful. As advocated by Skrondal (2000), we examined the 'practical significance' of the obtained ANOVA effects, by computing intraclass correlations $\hat{\rho}_I$ (Haggard, 1958; Kirk, 1995) as a measure of effect size. We only discuss the effects that account for more than 10% of the variance of the dependent variable (i.e., $\hat{\rho}_I > 10$). The results reveal a main effect of the number of clusters *C* ($\hat{\rho}_I = 42$): the higher the number of clusters, the lower the global minimum proportion. The number of clusters *C* further interacts with the amount of error ($\hat{\rho}_I = 22$): the effect of the number of clusters is more pronounced when error is present in the data (Figure 2).

Finally, we compared the percentage of VAF (equation (5)) of the clusterwise SCA-P and clusterwise SCA-ECP solution that was obtained for each of the simulated data sets. On average, the clusterwise SCA-P solution explains about 7% ($SD = 2.58\%$) more variance in the data than the clusterwise SCA-ECP solution.

4.1.2.2. Goodness of recovery. The goodness of recovery will be evaluated with respect to (1) the clustering of the groups and (2) the cluster loading matrices.

4.1.2.2.1. Recovery of the clustering of the groups. To examine the recovery of the clustering of the groups, the *Adjusted Rand Index* (ARI: Hubert & Arabie, 1985) is

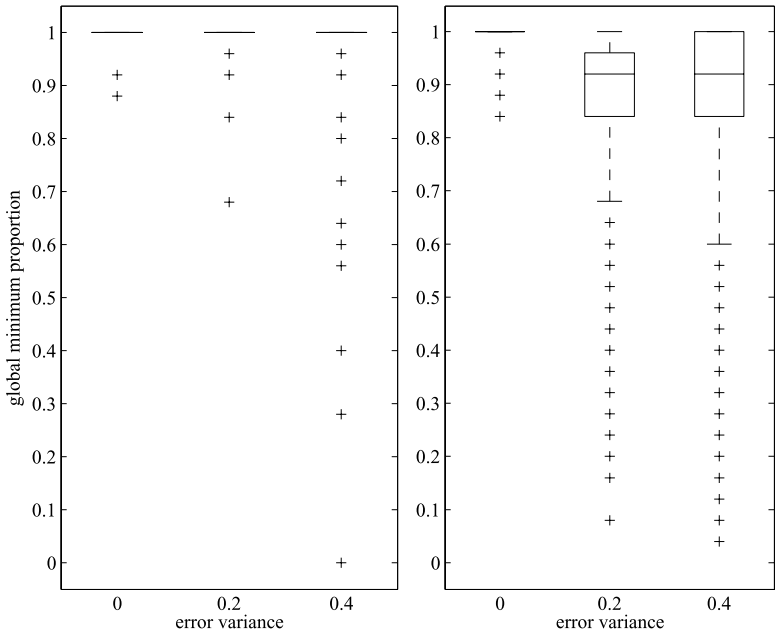


Figure 2. The proportion of random runs with a loss function value equal to that of the proxy of the global minimum ('global minimum proportion') as a function of the amount of error *e* when the number of clusters *C* is two (left panel) and when *C* is four (right panel).

calculated between the true partition matrix and the estimated partition matrix. The ARI equals one if the two partitions are identical, and equals zero when the agreement between the true and estimated partitions is at chance level.

On average, the ARI amounts to .99 ($SD = 0.04$), which indicates that the clustering of the groups is recovered very well. No analysis of variance was performed since only 2.94% (636) of the data sets resulted in an ARI smaller than one. The majority of these 636 data sets (531) are situated in the conditions with highly congruent loading matrices and 40% of error variance.

4.1.2.2.1. Recovery of the cluster loading matrices. To evaluate how well the cluster loading matrices are recovered, we calculated a goodness-of-cluster-loading-recovery statistic (GOCL) by computing congruence coefficients φ (Tucker, 1951) between the components of the true and estimated loading matrices and averaging these coefficients across components and clusters as follows:

$$\text{GOCL} = \frac{\sum_{c=1}^{C} \sum_{q=1}^{Q} \varphi\left(\mathbf{B}_q^{cT}, \mathbf{B}_q^{cM}\right)}{C\,Q}, \tag{6}$$

with $\mathbf{B}_q^{cT}$ and $\mathbf{B}_q^{cM}$ denoting the qth component of the true and estimated cluster loading matrices, respectively. The rotational freedom of the clusterwise SCA-P model was dealt with by rotating the estimated loading matrices towards the true loading matrices using an orthogonal Procrustes rotation. Moreover, the permutational freedom of the clusters (i.e., the columns of **P** can be permuted without altering the fit of the solution) was taken into account by selecting the column permutation of **P** that maximizes the GOCL

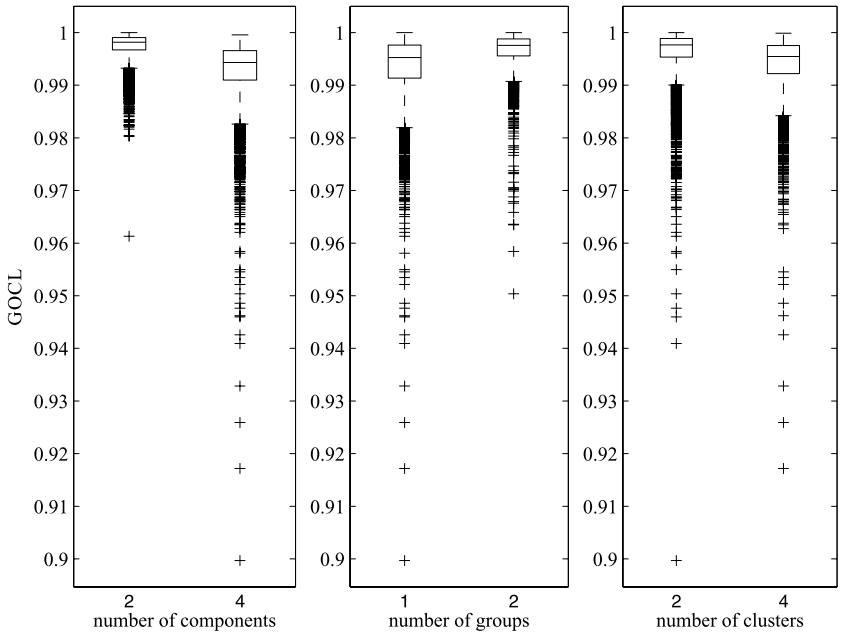


Figure 3. Box plots of the goodness-of-cluster-loading-recovery statistic (GOCL) as a function of the number of components (left), the number of groups (middle), and the number of clusters (right).

value. The GOCL statistic takes values between zero (no recovery at all) and one (perfect recovery).

On average, the GOCL statistic has a value of .99, with a standard deviation of 0.005, showing that the $\mathbf{B}^c$ matrices are recovered very well by the clusterwise SCA-P algorithm. An analysis of variance with the logit-transformed GOCL as the dependent variable and the seven factors as independent variables reveals a main effect of the number of components ($\hat{\rho}_I = 41$): this main effect implies that the recovery of the cluster loading matrices deteriorates when the number of components increases (see Figure 3). Moreover, a main effect is found of the number of groups ($\hat{\rho}_I = 14$) and of the number of clusters ($\hat{\rho}_I = 10$): the cluster loading matrices are recovered slightly better when the clusters contain more groups, that is, when the number of groups is higher or when the number of clusters is lower (Figure 3).

4.2. Simulation study 2

To investigate whether the model selection procedure presented succeeds in selecting the correct *C*- and *Q*-values, we used the first five replicates in each design cell of Simulation study 1, discarding the errorless data sets. We analysed each of these 1,440 data matrices with the clusterwise SCA-P algorithm, with *C* and *Q* varying from 1 to 6 and using 25 random starts per analysis, and applied the model selection procedure.

The procedure selects the correct *C*- and *Q*-value for 1,289 out of the 1,440 data sets (89.5%). When examining the results for the remaining data sets, we find that for respectively 7.1%, 2.8%, and 0.6% of the cases, only *C*, only *Q*, and both *C* and *Q* were selected incorrectly. The majority of the model selection mistakes (150 out of the 151

mistakes) were made in the conditions with four underlying clusters, 40% error variance and/or highly congruent cluster loading matrices.

4.3. Conclusion

From the simulation studies above, we can conclude (1) that clusterwise SCA-P rarely ends in a local minimum when 25 random starts are used,[3] (2) that clusterwise SCA-P explains more variance of the data than clusterwise SCA-ECP, (3) that the true underlying clustering as well as the within-cluster component models are recovered very well by the clusterwise SCA-P analysis,[3] and (4) that the model selection procedure retains the correct clusterwise SCA-P model in the majority of the simulated cases.

A limitation of the performed study might be that we use completely synthetic data, sampling the parameters from specific distributions. However, an advantage of this approach, in comparison with more realistic simulation studies in which some of the parameters are taken from the analysis of an empirical data set, is that we were able to evaluate the performance of our algorithm in a wide variety of well-defined conditions.

5. Application

In this section, we illustrate clusterwise SCA-P by applying it to data from psychiatric diagnosis research. In this field, the structure of diagnostic categories is extensively investigated, given the heavy criticism of standard diagnostic systems such as the different versions of the DSM (Kendel & Jablensky, 2003; Kendler, 1990; Zachar & Kendler, 2007). Specifically, as these systems define a diagnostic category by indicating which pattern of symptoms is typical for patients that belong to this category, a number of questions can be raised: one can wonder (1) whether clinicians agree about the extent to which different symptoms apply, (2) whether some structure can be discerned in the opinions of clinicians who disagree (do they disagree on the presence of single symptoms that seem randomly selected or on the presence of meaningful types of symptoms?), and (3) whether for some categories clinicians agree more than for others.

To shed light on these questions, we applied clusterwise SCA-P to data that were collected by Mezzich and Solomon (1980). These authors asked 22 clinicians to imagine a typical patient for four diagnostic categories: manic-depressive depressed (MDD), manic-depressive manic (MDM), simple schizophrenic (SS) and paranoid schizophrenic (PS). These categories are part of the nomenclature of mental disorders (DSM-II) issued in 1968 by the American Psychiatric Association. Subsequently, the 22 clinicians rated each archetypal patient on 17 psychopathological symptoms, on a 0 (absent) to 6 (extremely severe) Likert scale. As such an 88 patients by 17 symptoms data set was obtained (see Mezzich & Solomon, 1980)), where each patient belonged to one of the four diagnostic categories. Considering the diagnostic categories as the groups and the patients as the subjects, nested within the groups, we centred the data for each diagnostic category

[3]We also evaluated the performance in case of eight clusters, using the same design as in Section 4.1. The medium congruence level of the cluster loading matrices was omitted, however, since the data generation procedure for this level could not be readily generalized to eight clusters. The overall results are as follows: a mean ARI of .95 ($SD = 0.16$), a mean GOCL of .99 ($SD = 0.01$) and a mean 'global minimum proportion' of .77 ($SD = 0.24$), with the algorithm yielding for sure a local minimum for 0.50% of the simulated data sets.

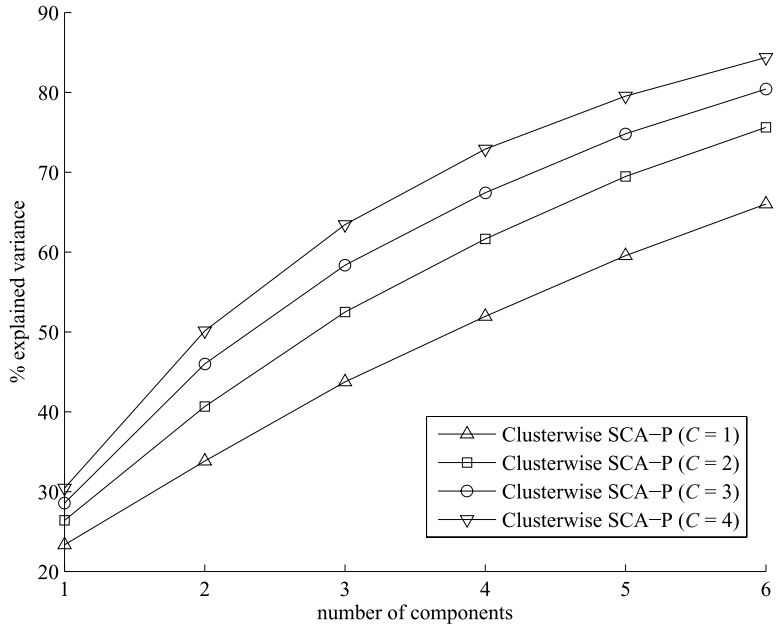


Figure 4. Percentage of explained variance for clusterwise SCA-P solutions with the number of components varying from 1 to 6, and the number of clusters for clusterwise SCA-ECP varying from 1 to 4, for the archetypal patients data.

separately and standardized the symptoms across categories (see Section 2.1). In this way the mean symptom profiles of the four diagnostic categories are removed from the data, but the information on the amount of disagreement for each category is retained.

To these data we fitted clusterwise SCA-P models with Q varying from 1 to 6 and C varying from 1 to 4 (i.e., the number of diagnostic categories). In Figure 4, the VAF percentage of the obtained solutions is plotted. The model selection procedure presented (see Section 3.3) suggests retaining two clusters, since the average scree ratio is maximal for the solutions with two clusters (Table 6, above). With two as the number of clusters, the solution with three components has the highest scree ratio (Table 6, below). Therefore, we decided to retain the solution with two clusters and three components.

Table 6. Scree ratios for the number of clusters C given the number of components Q (top), and for the number of components Q given two clusters (bottom), for the archetypal patients data. The maximal scree ratio in each column is highlighted in bold face.

	No. of components						
No. of clusters	1	2	3	4	5	6	average
C given Q							
2	**1.41**	1.28	**1.49**	**1.68**	**1.85**	**2.00**	**1.62**
3	1.16	**1.29**	1.16	1.06	1.13	1.21	1.17
Q given $C = 2$							
3		1.21	**1.29**	1.17	1.27		

Table 7. Varimax rotated loadings for the clusterwise SCA-P solution for the archetypal patients data with two clusters and three components. Loadings which are larger than ± .50 are highlighted in bold face.

	Cluster 1: Schizophrenia			Cluster 2: Manic depression		
	Grandiosity	Affective symptoms	Behavioural symptoms	Blunted affect	Anxiety	Cognitive symptoms
Depressive mood	0.17	**0.87**	0.24	0.00	0.14	−0.08
Excitement	−0.24	**0.59**	0.32	−0.03	−0.08	0.05
Guilt feelings	0.02	**0.79**	0.13	−0.06	0.47	−0.21
Anxiety	0.14	**0.63**	−0.14	−0.02	**0.91**	0.05
Tension	0.05	**0.81**	0.01	0.45	−0.18	0.43
Somatic concern	0.31	**0.62**	0.12	−0.13	**0.84**	0.12
Conceptual disorganization	−0.05	**0.65**	0.44	0.36	0.01	**0.65**
Unusual thought content	0.39	0.43	0.33	0.27	0.09	**0.92**
Hallucinatory behaviour	0.32	0.33	**0.61**	−0.38	0.03	**0.69**
Mannerisms and posturing	0.05	0.20	**1.00**	0.09	0.30	0.16
Motor retardation	−0.01	0.04	**1.09**	0.17	0.17	0.09
Grandiosity	**0.88**	0.16	0.01	0.30	0.01	0.28
Uncooperativeness	**0.53**	−0.13	0.36	0.29	**0.59**	**0.51**
Suspiciousness	0.45	0.17	0.07	−0.28	0.06	**1.01**
Hostility	0.37	0.00	−0.08	**0.60**	−0.30	0.07
Blunted affect	−0.40	−0.04	0.39	**0.95**	0.01	−0.12
Emotional withdrawal	−0.21	0.17	0.33	0.48	0.41	−0.01

In the solution selected, the partition matrix **P** (not shown) reveals that the PS and SS categories are assigned to the first cluster and the MDD and MDM categories to the second cluster. Therefore, these clusters can be called 'schizophrenia' and 'manic depression', respectively.

The varimax rotated component loadings of these two clusters are displayed in Table 7. In the schizophrenia cluster, the first component can be labelled 'grandiosity' since this is the only symptom with a very strong loading on the component. Given the high loadings for 'tension', 'depressive mood', and 'guilt feelings', the second component of this cluster is named 'affective symptoms'. On the third component motor and behavioural symptoms such as 'mannerisms and posturing', 'hallucinatory behaviour' and 'motor retardation' load high; therefore, it is labelled 'behavioural symptoms'.

In the manic depression cluster, the first component is called 'blunted affect', because of the high loading of this symptom. The 'somatic concern' and 'anxiety' symptoms have high loadings on the second component, which is thus labelled 'anxiety'. On the third component cognitive symptoms such as 'conceptual disorganization', 'suspiciousness' and 'unusual thought content' load high; therefore it is named 'cognitive symptoms'.

The variances and correlations of the component scores are presented in Table 8. From this table, it can be concluded that the variances of the component scores differ substantially between the diagnostic categories that belong to the same cluster. Specifically, in the schizophrenia cluster, the variance on the 'behavioural symptoms' component is larger for the simple schizophrenic patients than for the paranoid schizophrenic patients. This indicates a relatively large disagreement among psychiatrists about the severity of behavioural symptoms in simple schizophrenic patients. For the manic-depressive patients with depression, there appears to be strong disagreement

Table 8. Variances and correlations of the component scores by diagnostic category for the clusterwise SCA-P solution for the archetypal patients data with two clusters and three components.

Cluster			Variances	Correlations	
Cluster 1: Schizophrenia				Affective symptoms	Behavioural symptoms
	Simple schizophrenia	Grandiosity	0.96	.22	.21
		Affective symptoms	1.15		−.23
		Behavioural symptoms	1.58		
	Paranoid schizophrenia	Grandiosity	1.04	−.24	−.40
		Affective symptoms	0.96		.51
		Behavioural symptoms	0.42		
Cluster 2: Manic depression				Anxiety	Cognitive symptoms
	Manic depression, depressive	Blunted affect	1.67	.04	−.08
		Anxiety	1.14		.09
		Cognitive symptoms	0.85		
	Manic depression, manic	Blunted affect	0.33	−.12	.16
		Anxiety	0.86		−.09
		Cognitive symptoms	1.15		

about the extent to which they are characterized by 'blunted affect'. These differences in the amount of disagreement about the symptoms of PS and SS, on the one hand, and MDM and MDD, on the other hand, may be explained by the fact that the symptoms of simple schizophrenia and manic depression depressive are mostly 'negative' (i.e., normal aspects of a person's behaviour disappear), such as mental and motor retardation, reduction of interest, apathy and impoverishment of interpersonal relations. In contrast, paranoid schizophrenia and manic-depressive illness manic are psychiatric disorders with very salient 'positive' symptoms (i.e., abnormal symptoms that are added to the behaviour), such as hallucinations, aggression, talkativeness, accelerated speech and motor activity. Therefore, it is not surprising that there is less disagreement about the symptoms of these disorders than about the symptoms of simple schizophrenia and manic-depressive illness depressive.

Table 8 also shows the correlations between the component scores for each of the four diagnostic categories. In general, these component correlations are rather low. This indicates that the opinion of clinicians on symptoms of one type is quite independent of their opinion on symptoms of another type.

We conclude that clusterwise SCA-P allows us to formulate fine-grained yet parsimonious answers to the three research questions outlined above. (1) The psychiatrists indeed disagree on the symptoms of the four disorders. (2) The specific symptoms for which disagreement exists can be grouped into meaningful types, which differ between the schizophrenia and the manic-depressive disorders. (3) The amount of disagreement about the types of symptoms differs between the categories within a cluster. More specifically, the clinicians disagree more about the disorders with negative symptoms (MDD and SS) than about the disorders with positive symptoms (MDM and PS).

6. Discussion

In this paper, the clusterwise SCA-P model was proposed for detecting and modelling structural differences and similarities between data of several groups. Clusterwise SCA-P is more flexible than clusterwise SCA-ECP, as clusterwise SCA-P allows component variances and correlations to vary freely within each cluster. Therefore, clusterwise SCA-P may result in more comprehensive and/or more parsimonious solutions (in terms of the number of clusters) than clusterwise SCA-ECP. For the sake of clarity, we focused on data from different groups of subjects in this paper. However, clusterwise SCA is also applicable to multivariate time series data from multiple subjects (for illustrative applications, see De Roover *et al.*, 2011; De Roover, Ceulemans, & Timmerman, 2011).

We see at least three possible directions for further research. First, in this paper, the number of components was fixed across the clusters. Due to this restriction, differences in the nature of the underlying dimensions are captured rather than differences in number of underlying dimensions. This is often not ideal. For example, in personality psychology, personality trait structure is often defined by five dimensions (Goldberg, 1990). However, some authors claim that in some cultures extra dimensions might be needed to adequately describe the structure of personality (Diaz-Loving, 1998). Therefore, in future research it would be useful to allow the number of components to vary between clusters. This generalization is not as straightforward as it may seem, as it would result in non-arbitrary problems with respect to the model estimation. Meanwhile, researchers can use the following strategy: inspect the within-cluster component models of the clusterwise SCA solution obtained and look for signs of overextraction (e.g., one of the components is determined by only one variable, or has low loadings for all variables) and, when indicated, fit an SCA solution with a lower number of components to the data of the groups that belong to the cluster at hand.

Second, clusterwise SCA clusters the groups on the basis of the within-group structures, ignoring between-group differences in variable means. However, these differences in means could reveal interesting additional information. Therefore, one may consider developing an extension of clusterwise SCA in which the group means are modelled as well. Such an extension has already been described for SCA (Timmerman, 2006), which implies a PCA of the group means next to an SCA of the within-group structure. Alternatively, one could model the group means by reduced K-means (Bock, 1987; de Soete & Carroll, 1994; Timmerman, Ceulemans, Kiers, & Vichi, 2010), which would entail a clustering of the groups as well as a dimension reduction of the variables.

Third, it may be useful to introduce group-specific weights to correct for the unwanted dominance of some groups (for an overview of possible weighting strategies, see Van Deun, Smilde, van der Werf, Kiers, & Van Mechelen, 2009). For instance, one may want to give more weight to the data of smaller groups, to avoid the analysis results being primarily influenced by the larger groups.

Acknowledgements

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 Katholieke Universiteit Leuven (GOA/2010/02).

References

Bock, H. H. (1987). On the interface between cluster analysis, principal component analysis, and multidimensional scaling. In H. Bozdogan & A. K. Gupta (Eds.), *Multivariate statistical modeling and data analysis* (pp. 17–34). Dordrecht: Reidel.

Brusco, M. J., & Cradit, J. D. (2001). A variable selection heuristic for *K*-means clustering. *Psychometrika*, *66*, 249–270.

Ceulemans, E., & Kiers, H. A. L. (2006). 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*, 133–150.

De Roover, K., Ceulemans, E., Timmerman, M. E. (2011). How to perform multiblock component analysis in practice. *Behavior Research Methods*. doi:10.3758/s13428-011-0129-1

De Roover, K., Ceulemans, E., Timmerman, M. E., Vansteelandt, K., Stouten, J., & Onghena, P. (2011). Clusterwise SCA-ECP for analyzing structural differences in multivariate multiblock data. *Psychological Methods*. doi:10.1037/a0025385

de Soete, G., & Carrol, J. D. (1994). *K*-means clustering in a low-dimensional Euclidean space. In E. Diday, Y. Léchevallier, M. Schader, P. Bertrand, & B. Burtschy (Eds.) *New approaches in classification and data analysis* (pp. 212–219). Berlin: Springer.

Diaz-Loving, R. (1998). Contributions of Mexican ethnopsychology to the resolution of the etic-emic dilemma in personality. *Journal of Cross-Cultural Psychology*, *29*, 104–118.

Dolan, C. V., Oort, F. J., Stoel, R. D., & Wicherts, J. M. (2009). Testing measurement invariance in the target rotated multigroup exploratory factor model. *Structural Equation Modeling*, *16*, 295–314.

Goldberg, L. R. (1990). An alternative 'description of personality': The Big-Five factor structure. *Journal of Personality and Social Psychology*, *59*, 1216–1229.

Haggard, E. A. (1958). *Intraclass correlation and the analysis of variance*. New York: Dryden.

Harris, C. W., & Kaiser, H. F. (1964). Oblique factor analytic solutions by orthogonal transformations. *Psychometrika*, *29*, 347–362.

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, *2*, 193–218.

Jolliffe, I. T. (1986). *Principal component analysis.* New York: Springer.

Kaiser, H. F. (1958). The Varimax criterion for analytic rotation in factor analysis. *Psychometrika*, *23*, 187–200.

Kendell, R., & Jablensky, A. (2003). Distinguishing between the validity and utility of psychiatric diagnosis. *American Journal of Psychiatry*, *160*, 4–12.

Kendler, K. S. (1990). Toward a scientific psychiatric nosology: Strengths and limitations. *Archives of General Psychiatry*, *47*, 969–973.

Kiers, H. A. L. (1990). *SCA. A program for simultaneous components analysis of variables measured in two or more populations*. Groningen: iec ProGAMMA.

Kiers, H. A. L., & ten Berge, J. M. F. (1994a). Hierarchical relations between methods for simultaneous components analysis and a technique for rotation to a simple simultaneous structure. *British Journal of Mathematical and Statistical Psychology*, *47*, 109–126.

Kiers, H. A. L., & ten Berge, J. M. F. (1994b). The Harris-Kaiser independent cluster rotation as a method for rotation to simple component weights. *Psychometrika*, *59*, 81–90.

Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Pacific Grove, CA: Brooks/Cole.

Kroonenberg, P. M. (2008). *Applied multiway data analysis*. Hoboken, NJ: Wiley.

Lawley, D. N., & Maxwell, A. E. (1962). Factor analysis as a statistical method. *The Statistician*, *12*, 209–229.

McCrae, R. R., & Costa, P. T., Jr. (1997). Personality trait structure as a human universal. *American Psychologist*, *52*, 509–516.

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. New York: Wiley.

Mezzich, J. E., & Solomon, H. (1980). *Taxonomy and behavioral science: Comparative performance of grouping methods*. London: Academic Press.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, *2*, 559–572.

Skrondal, A. (2000). Design and analysis of Monte Carlo experiments: Attacking the conventional wisdom. *Multivariate Behavioral Research, 35*, 137–167.

Steinley, D. (2003). Local optima in *K*-means clustering: What you don't know may hurt you. *Psychological Methods*, *8*, 294–304.

Timmerman, M. E. (2006). Multilevel component analysis. *British Journal of Mathematical and Statistical Psychology*, *59*, 301–320.

Timmerman, M. E., Ceulemans, E., Kiers, H. A. L., & Vichi, M. (2010). Factorial and reduced *K*-means reconsidered. *Computational Statistics & Data Analysis*, *54*, 1858–1871.

Timmerman, M. E., & Kiers, H. A. L. (2003). Four simultaneous component models of multivariate time series from more than one subject to model intraindividual and interindividual differences. *Psychometrika*, *86*, 105–122.

Tucker, L. R. (1951). *A method for synthesis of factor analysis studies* (Personnel Research Section Rep. No. 984). Washington, DC: Department of the Army.

Van Deun, K., Smilde, A. K., van der Werf, M. J., Kiers, H. A. L., & Van Mechelen, I. (2009). A structured overview of simultaneous component based data integration. *BMC Bioinformatics*, *10*, 246.

Zachar, P., & Kendler, K. S. (2007). Psychiatric disorders: A conceptual taxonomy. *American Journal of Psychiatry*, *164*, 557–565.

Received 3 February 2011; revised version received 28 October 2011

Appendix: Two different scalings of clusterwise SCA-P and SCA-ECP solutions

As mentioned in Section 2.3, the variance of the component scores is fixed at one across all groups belonging to the same cluster, to partly identify the clusterwise SCA-P solution. This type of scaling will be referred to as 'scaling by cluster'. An alternative way of scaling the component scores can be considered, however, which will be referred to as 'scaling across clusters'. Both types of scaling, which are also applicable to clusterwise SCA-ECP, will be discussed below.

For ease of explanation, we rewrite the decomposition rule (equation (1)) of the clusterwise SCA-ECP and clusterwise SCA-P models as follows:

$$\mathbf{X} = \mathbf{F}\mathbf{B}' + \mathbf{E}, \tag{7}$$

where $\mathbf{F}$ is an $I \times CQ$ matrix, of which the cth set of Q columns consists of the $\mathbf{F}_k$ matrices for the groups that belong to cluster c and zeros for the groups that belong to another cluster, $\mathbf{B} = [\mathbf{B}^1\ \mathbf{B}^2\ \ldots\ \mathbf{B}^c]$ is a $J \times CQ$ matrix that concatenates the C cluster loading matrices, and $\mathbf{E}$ ($I \times J$) denotes the matrix of residuals. For example, given the partition matrix $\mathbf{P}$ (see Table 3) of the clusterwise SCA-P decomposition of the hypothetical data in Table 2, equation (7) would read as follows:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \mathbf{X}_3 \\ \mathbf{X}_4 \\ \mathbf{X}_5 \\ \mathbf{X}_6 \end{bmatrix} = \begin{bmatrix} \mathbf{F}_1 & [\mathbf{0}] \\ \mathbf{F}_2 & [\mathbf{0}] \\ \mathbf{F}_3 & [\mathbf{0}] \\ \mathbf{F}_4 & [\mathbf{0}] \\ [\mathbf{0}] & \mathbf{F}_5 \\ [\mathbf{0}] & \mathbf{F}_6 \end{bmatrix} \left[\mathbf{B}^1\mathbf{B}^2\right]' + \begin{bmatrix} \mathbf{E}_1 \\ \mathbf{E}_2 \\ \mathbf{E}_3 \\ \mathbf{E}_4 \\ \mathbf{E}_5 \\ \mathbf{E}_6 \end{bmatrix}.$$

Note that this example shows that the components of the different clusters are orthogonal to each other.

When scaling by cluster is applied, the variance of the non-zero component scores is set to one for each column of **F** in equation (7). This implies that the relative sizes of the component scores are independent of the cluster size (i.e., the number of individuals belonging to a cluster) and thus can be compared across clusters. For each loading b^c_{jq}, $(b^c_{jq})^2/(s^c_j)^2$ is the proportion of the cluster-specific variance of the jth variable that is explained by component q, where $(s^c_j)^2$ is the variance of variable j across all groups that make up cluster c. If the data are standardized across all groups, these cluster-specific variances will not necessarily equal one. This implies that the loadings cannot be interpreted as correlations. Only if the variables are autoscaled rather than standardized across all groups do the squared loadings $(b^c_{jq})^2$ equal the proportion of cluster-specific variance of variable j that is explained by component q. Then the loadings are also correlations between components and variables, in the case of orthogonal components.

Scaling across clusters implies that the variance of the complete columns of **F** in equation (7), thus including the zero entries, is set to one. The cluster loading matrices $\tilde{\mathbf{B}}^c$ and corresponding component score matrices $\tilde{\mathbf{F}}_k$ of a solution that is scaled across all clusters can be obtained directly from the solution that is scaled by cluster, namely as $\tilde{\mathbf{B}}^c = \sqrt{\frac{I^c}{I}}\mathbf{B}^c$ and $\tilde{\mathbf{F}}_k = \sqrt{\frac{I}{I^c}}\mathbf{F}_k$, where I^c is the number of subjects within cluster c. When the component scores are scaled across clusters the sizes of the component scores can only be compared within a cluster, because the size of the component scores is affected by the cluster size. Specifically, in clusters that contain a relatively low number of subjects, the absolute values of the scores will be higher than in clusters that contain more subjects. The squared loadings $(\tilde{b}^c_{jq})^2$ equal the proportion of total variance of the jth variable (i.e., across all clusters) that is explained by component q. Furthermore, if the components are orthogonal within each cluster, the loadings are correlations between the variables and components (across all clusters).

Summarizing, which scaling is to be preferred, scaling by cluster or scaling across clusters, depends on which aspect of the solution should be comparable across clusters. If the size of the component scores should be comparable irrespective of cluster size, one should use scaling by cluster. If one is interested in loadings that are independent of the cluster-specific variances of the variables and that can be read as correlations between variables and components across all clusters, then scaling across clusters is to be preferred.