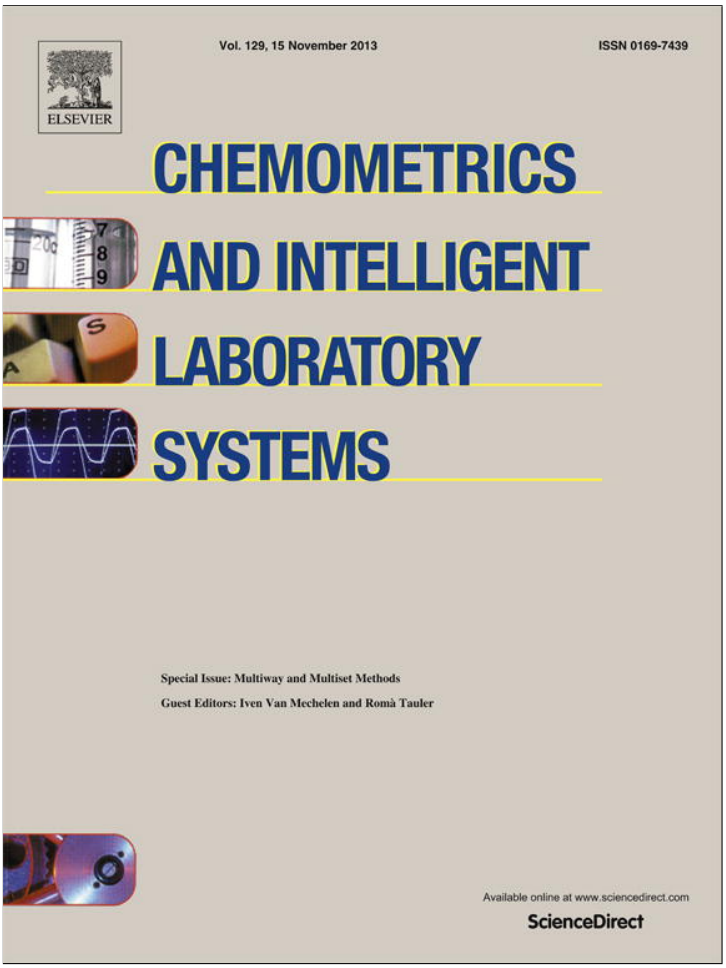
Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/authorsrights

Author's personal copy

Chemometrics and Intelligent Laboratory Systems 129 (2013) 33–39

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Robust multilevel simultaneous component analysis

Eva Ceulemans [a], Mia Hubert [b,*], Peter Rousseeuw [b]

[a] KU Leuven, Methodology of Educational Sciences Research Group, Andreas Vesaliusstraat 2, BE-3000 Leuven, Belgium
[b] KU Leuven, Department of Mathematics, Celestijnenlaan 200B, BE-3001 Leuven, Belgium

ARTICLE INFO

Article history:
Received 15 January 2013
Received in revised form 20 June 2013
Accepted 29 June 2013
Available online 8 July 2013

Keywords:
Multilevel
Multivariate
Outliers
Principal components
Visualization

ABSTRACT

Multilevel simultaneous component analysis (MSCA) is a class of techniques for analyzing multivariate data with more than one level. However, as MSCA uses classical statistics such as the arithmetic mean, the resulting estimates are highly sensitive to outlying measurements. In this paper we propose a robust version of MSCA-P (the least restrictive MSCA variant) which can withstand the effects of outliers, while its associated outlier maps reveal the outliers visually. The technique is applied to data from an eating disorder study as well as to chemical data. We show by simulation that robust MSCA-P succeeds well in detecting outliers. The latter conclusion also holds when the data are generated according to the other MSCA variants.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Multilevel multivariate data are encountered in many fields of science. A first example is sensory profiling, where panelists are asked to rate a set of food samples on a number of attributes (e.g., [1]). The resulting food samples by attributes data matrix has a multilevel structure, as the food samples (level 1) are nested within the panelists (level 2). As another example, in clinical psychology one often asks patients to report on their momentary cognitions, feelings, and actions at multiple randomly selected time points in daily life (e.g., [2]). In this case the measurement occasions constitute the level 1 units which are nested within the patients (level 2). In this paper we will limit ourselves to data with two levels, and throughout we will refer to the level 2 units as subjects and the level 1 units as measurement occasions.

Given two-level multivariate data, one often wants to summarize the correlation structure among the variables at the different levels. It is important to realize that these level 1 and level 2 correlations may be very different. For instance, in psychology [3] found that general levels of positive and negative affect were independent across subjects (level 2) but negatively related within subjects (level 1).

To study the level 1 and level 2 correlations, multilevel simultaneous component analysis (MSCA, [4]) can be used. MSCA specifies a separate component model for each level in the data. The level 2 or between-subjects model captures the correlation structure of the subject means by reducing the variables to a few between-components. Similarly, the correlation structure in the deviations from the subject means is summarized by a level 1 or within-subjects model, making use of within-components. The within-subjects model is a simultaneous component model, of which different variants have been proposed [5]. MSCA has been successfully applied in different fields, such as cross-cultural psychology [6], metabolomics [7], social psychology [8], and process data analysis [9].

However, as MSCA uses classical statistics such as the arithmetic mean, the resulting estimates are highly sensitive to outlying measurements. In this setting, outliers can occur at level 1 and at level 2. Outliers at level 1 correspond to outlying measurement occasions, whereas outliers at level 2 correspond to outlying subjects. To protect the MSCA analysis against both types of outliers we propose to use robust estimators at each stage of the analysis. Specifically, we present a robust version of MSCA-P, the least restrictive MSCA variant. Like all robust statistical methods (see e.g., [10] and the references therein), robust MSCA-P looks for a model that describes the majority of the data, and as a by-product yields diagnostic tools to detect and visualize the outliers.

Moreover, as MSCA-P subsumes the three other MSCA variants as special cases, robust MSCA-P can also be useful if one wants to apply one of the more restrictive MSCA variants. In such cases, a sequential strategy can be adopted in which one first applies robust MSCA-P to detect possibly outlying measurement occasions and subjects. Next, these outliers are removed and the restrictive MSCA model is fitted to the remaining data.

In the next section we recapitulate MSCA. Next, we propose robust MSCA-P. In Section 4 we illustrate the usefulness of robust MSCA-P by applying it to a study on eating disorders and to chemical data. Section 5 discusses a simulation study in which we examine whether robust MSCA-P succeeds in detecting outlying subjects and measurement occasions. The

* Corresponding author. Tel.: + 3216322023
E-mail address: Mia.Hubert@wis.kuleuven.be (M. Hubert).

0169-7439/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.chemolab.2013.06.016

Author's personal copy

simulated data are generated according to the four MSCA variants introduced by [4].

2. Multilevel simultaneous component analysis

MSCA has been developed for modeling the between-subjects and within-subjects correlation structure of a two-level data matrix $\mathbf{X}$ with N observations by J variables. This matrix $\mathbf{X}$ is obtained by stacking I data matrices $\mathbf{X}_i(N_i \times J)$, each of which holds the data of subject i ($i = 1, \ldots, I$) on the associated N_i measurement occasions. The columns of $\mathbf{X}$ will be denoted by $X_1, \ldots, X_J$.

The MSCA model decomposes each of these I data matrices $\mathbf{X}_i$ as follows:

$$\boldsymbol{X}_i = \mathbf{1}_{Ni}\boldsymbol{m}^T + \mathbf{1}_{Ni}\boldsymbol{f}_i^b\left(\boldsymbol{B}^b\right)^T + \boldsymbol{F}_i^w\left(\boldsymbol{B}^w\right)^T + \boldsymbol{E}_i \tag{1}$$

where $\mathbf{1}_{Ni}$ is a $N_i \times 1$ vector of ones, $\mathbf{m}$ $(J \times 1)$ contains the offsets of the J variables across all measurement occasions of all subjects, $\boldsymbol{f}_i^b$ $(1 \times Q_b)$ is the i-th row of the $I \times Q_b$ between-subjects component scores matrix $\boldsymbol{F}^b$, $\boldsymbol{B}^b$ $(J \times Q_b)$ denotes the between-subjects loading matrix, $\boldsymbol{F}_i^w$ $(N_i \times Q_w)$ denotes the within-subjects component scores matrix of subject i, $\boldsymbol{B}^w$ $(J \times Q_w)$ denotes the within-subjects loading matrix, which does not depend on i, and $\boldsymbol{E}_i$ $(N_i \times J)$ is the matrix of residuals. Here, Q_b is the number of between-subjects components and Q_w the number of within-subjects components. To make the model identifiable it is imposed that the mean between-subjects component scores over all N measurement occasions are zero:

$$\sum_{i=1}^{I} N_i \boldsymbol{f}_i^b = 0,$$

and that the mean within-subjects component scores across the N_i measurement occasions of subject i are zero, i.e.,

$$\left(\mathbf{1}_{N_i}\right)^T \boldsymbol{F}_i^w = 0 \qquad \text{for } i = 1,\ldots,I \ .$$

Timmerman [4] proposed four MSCA variants, that only differ with respect to the within-subjects model. MSCA-P, the least restrictive variant, imposes that $(\boldsymbol{F}^w)^T\boldsymbol{F}^w$ (with $\boldsymbol{F}^w$ denoting the vertical concatenation of all within-subjects component scores matrices) has $\sqrt{N}$ on the diagonal, implying that the variance of the component scores amounts to one across subjects. MSCA-PF2 adds the restriction that the correlations among the subject-specific component scores in $\boldsymbol{F}_i^w$ are the same for all subjects. MSCA-IND further constrains these correlations to zero. Finally, MSCA-ECP, the most restrictive variant, imposes that the variance of the component scores amounts to one for each separate subject and that the correlations of these scores are the same across subjects.

In this paper we will propose a robust version of MSCA-P. In this variant, both the between-subjects loading matrix $\boldsymbol{B}^b$ as well as the within-subjects loading matrix $\boldsymbol{B}^w$ have rotational freedom. Specifically, $\boldsymbol{B}^b$ and $\boldsymbol{B}^w$ can be rotated without altering the fit of the model, provided that the associated between- and within-subjects component scores matrices $\boldsymbol{F}^b$ and $\boldsymbol{F}_i^w$ are counterrotated. We will make use of this rotational freedom and reparametrize the model without loss of generality by replacing the MSCA-P restriction discussed above by the constraint that $\boldsymbol{B}^w$ is columnwise orthonormal, implying that $(\boldsymbol{B}^w)^T\boldsymbol{B}^w$ is an identity matrix. The same restriction is imposed on $\boldsymbol{B}^B$.

To obtain a classical MSCA-P solution with Q_b between-subjects components and Q_w within-subjects components for a given data matrix $\boldsymbol{X}$, one first estimates the overall mean

$$\hat{\boldsymbol{m}} = \frac{\boldsymbol{X}^T\mathbf{1}_N}{N}. \tag{2}$$

Then $\boldsymbol{X}$ is mean-centered:

$$\tilde{\boldsymbol{X}} = \boldsymbol{X} - \mathbf{1}_N\hat{\boldsymbol{m}}^T. \tag{3}$$

Next, each submatrix $\tilde{\boldsymbol{X}}_i$ is split into a between-subjects part $\tilde{\boldsymbol{X}}_i^b$ given by

$$\tilde{\boldsymbol{X}}_i^b = \mathbf{1}_{N_i}\frac{\left(\mathbf{1}_{N_i}\right)^T\tilde{\boldsymbol{X}}_i}{N_i} \tag{4}$$

where each row corresponds to the empirical average of $\tilde{\boldsymbol{X}}_i$, and a within-subjects part $\tilde{\boldsymbol{X}}_i^w$ defined as

$$\tilde{\boldsymbol{X}}_i^w = \tilde{\boldsymbol{X}}_i - \mathbf{1}_{N_i}\frac{\left(\mathbf{1}_{N_i}\right)^T\tilde{\boldsymbol{X}}_i}{N_i}. \tag{5}$$

Subsequently, the between-subjects and within-subjects components can be estimated by applying singular value decompositions to the vertical concatenation of all between-subjects and within-subjects parts, respectively.

Often differences in variability between the J variables are not of interest and should therefore be removed beforehand. To this end, one may first standardize the data by dividing the variables X_j by their standard deviation across both levels. Note that this commonly used standardization does not imply that the variances of the between-subjects (c.q. within-subjects) parts of the variables are identical.

3. Robust MSCA-P

To perform a robust MSCA-P, we proceed as follows. First, instead of using the overall mean (2), we robustly center the observed data $\boldsymbol{X}$ (of which we denote the rows as $\boldsymbol{x}_1^T, \ldots, \boldsymbol{x}_N^T$). As a robust measure of centrality we use the spatial median, also known as the L^1-median. It is defined as the point $\hat{\boldsymbol{m}}_s$ which minimizes the sum of the distances to all observations, that is,

$$\hat{\boldsymbol{m}}_s = \underset{\hat{\boldsymbol{m}}}{\operatorname{argmin}} \sum_{k=1}^{N} \|\boldsymbol{x}_k - \hat{\boldsymbol{m}}\|.$$

The spatial median is highly robust to outliers as it can withstand up to 50% of outlying observations $\boldsymbol{x}_k$. More formally, it has a 50% breakdown value ([11]). Moreover, it is *orthogonally equivariant*, which means that it behaves properly under orthogonal transformations (such as rotations and reflections) of the data. That is, if the $\boldsymbol{x}_k$ yield the spatial median $\hat{\boldsymbol{m}}_s$, then the spatial median of the $\boldsymbol{A}\boldsymbol{x}_k$ equals $\boldsymbol{A}\hat{\boldsymbol{m}}_s$ for all orthogonal matrices $\boldsymbol{A}$. Therefore, the spatial median preserves the rotational invariance of MSCA-P. The spatial median can be computed by different algorithms [12]. (Note that one could also center the data by the coordinatewise median $(med(X_1), med(X_2), \ldots, med(X_J))^T$ which is equally robust, but it is not orthogonally equivariant.)

Secondly, we decompose $\tilde{\boldsymbol{X}}_i$ into a between-subjects part and a within-subjects part as in Eqs. (4) and (5), now using the spatial median of the N_i measurement occasions instead of their mean.

In the third step of MSCA-P we compute robust loadings using the robust principal component analysis method ROBPCA [13]. In general ROBPCA decomposes an $N \times J$ data matrix $\boldsymbol{Y}$ into a robust center $\hat{\boldsymbol{\mu}}^T$, a $J \times Q$ orthogonal loading matrix $\boldsymbol{B}$ and an $N \times Q$ scores matrix $\boldsymbol{F}$ following the PCA model:

$$\boldsymbol{Y} = \mathbf{1}_N\hat{\boldsymbol{\mu}}^T + \boldsymbol{F}\boldsymbol{B}^T + \boldsymbol{E}.$$

In robust statistics it is commonly assumed that at least half of the data points are uncontaminated, otherwise no method can distinguish between the regular observations and the outliers. In many applications one assumes that $h = [\alpha N]$ observations are uncontaminated, with

Author's personal copy

$0.5 < \alpha < 1$. One can run ROBPCA for a specific rate α (or h). The most robust procedure is attained for $\alpha = 0.5$ as then half of the data points could be contaminated. As this reduces the efficiency at data sets with lower amounts of contamination, one often puts $\alpha = 0.75$.

ROBPCA combines ideas of projection pursuit and robust covariance estimation. A rough outline of the method is as follows. The first step of the algorithm computes a measure of outlyingness for each data point $\boldsymbol{y}_i$ (the i-th row of $\boldsymbol{Y}$). This is done by projecting the data on many straight lines through two data points. In each projection one computes robust univariate estimators of its center and scale, and for every data point its standardized distance to that center is measured. For each observation the largest distance over all directions is stored, which is called its outlyingness. The $h = [\alpha N]$ points with the smallest outlyingness then form an h-subset H_0 and their empirical covariance matrix $\hat{\boldsymbol{\Sigma}}_0$ is computed. From the eigenvalues of $\hat{\boldsymbol{\Sigma}}_0$ the number of principal components Q is selected (as in classical PCA). As initial robust Q-dimensional subspace estimate, ROBPCA considers the subspace V_0 spanned by the Q dominant eigenvectors of $\hat{\boldsymbol{\Sigma}}_0$.

To increase the efficiency a first reweighting is then carried out. All data points which are close to V_0 receive weight 1, whereas the observations far away from it receive zero weight. More precisely, for each observation $\boldsymbol{y}_i$ its orthogonal distance is computed as

$$\mathrm{OD}(\boldsymbol{y}_i) = ||\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i|| \tag{6}$$

with $\hat{\boldsymbol{y}}_i$ the projection of $\boldsymbol{y}_i$ on the subspace V_0. In [13] it is argued that the orthogonal distances to the power 2/3 are approximately normally distributed $N(\mu_d, \sigma_d^2)$ where μ_d and σ_d are unknown parameters which we estimate by univariate robust estimators $\hat{\mu}_d$ and $\hat{\sigma}_d$. Consequently, observations with OD smaller than $\left(\hat{\mu}_d + \hat{\sigma}_d \Phi^{-1}(0.975)\right)^{3/2}$ are retained and their covariance matrix $\hat{\boldsymbol{\Sigma}}_1$ is computed. An improved robust subspace estimate is now obtained as the subspace V_1 spanned by the Q dominant eigenvectors of $\hat{\boldsymbol{\Sigma}}_1$.

The second stage of ROBPCA projects the data points onto V_1. The projection of $\boldsymbol{y}_i$ is denoted as $\boldsymbol{y}_i^*$. Their center $\hat{\boldsymbol{\mu}}$ and scatter matrix $\hat{\boldsymbol{\Sigma}}_R$ are estimated using the Minimum Covariance Determinant (MCD) estimator [14,15]. The MCD is a highly robust estimator of location and scatter for low-dimensional data. First the method looks for an optimal h-subset H_1 that contains h observations for which the determinant of their empirical covariance matrix is minimal. The center $\hat{\boldsymbol{\mu}}_{raw}$ and scatter matrix $\hat{\boldsymbol{\Sigma}}_{raw}$ of these h observations are computed. Next, a second reweighting procedure is performed to include more than h points in the computations. Weights are based on the robust distance of every point with respect to $\hat{\boldsymbol{\mu}}_{raw}$ and $\hat{\boldsymbol{\Sigma}}_{raw}$. More precisely, observations obtain a zero weight if their robust distance

$$d\left(\boldsymbol{y}_i^*, \hat{\boldsymbol{\mu}}_{raw}, \hat{\boldsymbol{\Sigma}}_{raw}\right) = \sqrt{(\boldsymbol{y}_i^* - \hat{\boldsymbol{\mu}}_{raw})^T \hat{\boldsymbol{\Sigma}}_{raw}^{-1} (\boldsymbol{y}_i^* - \hat{\boldsymbol{\mu}}_{raw})} \tag{7}$$

is larger than $\sqrt{\chi^2_{Q,0.975}}$, the square root of the 0.975-quantile of the χ^2 distribution with Q degrees of freedom. The other observations receive weight 1. Finally, the reweighted MCD center $\hat{\boldsymbol{\mu}}$ and covariance matrix $\hat{\boldsymbol{\Sigma}}_R$ are determined as the classical mean and covariance matrix of the observations with weight 1.

The last stage of ROBPCA consists of representing $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}_R$ in the original J-dimensional data space. The spectral decomposition of this scatter matrix $\hat{\boldsymbol{\Sigma}}_R = \boldsymbol{BLB}^T$ yields the robust loadings $\boldsymbol{B}$ and the robust eigenvalues $\boldsymbol{L}$. Finally the scores are computed as $\boldsymbol{F} = \left(\boldsymbol{Y} - \boldsymbol{1}_N \hat{\boldsymbol{\mu}}^T\right)\boldsymbol{B}$.

Note that the reweighting steps typically increase the size of the subsets on which the computations are performed. Therefore, it cannot be said that the final estimates are solely based on h well chosen observations. This is especially important in smaller data sets, as the reweighting steps decrease the variability of the estimates.

Here we apply ROBPCA to the stacked $\tilde{\boldsymbol{X}}_i^b$ which yields

$$\tilde{\boldsymbol{X}}_i^b = \boldsymbol{1}_{Ni}\left(\hat{\boldsymbol{\mu}}^b\right) + \boldsymbol{f}_{R,i}^b \left(\boldsymbol{B}_R^b\right)^T + \boldsymbol{E}_i^b.$$

Analogously we decompose

$$\tilde{\boldsymbol{X}}_i^w = \boldsymbol{1}_{Ni}(\hat{\boldsymbol{\mu}}^w)^T + \boldsymbol{F}_{R,i}^w (\boldsymbol{B}_R^w)^T + \boldsymbol{E}_i^w.$$

The overall robust offset is then given by

$$\hat{\boldsymbol{m}}_R = \hat{\boldsymbol{m}}_s + \hat{\boldsymbol{\mu}}^b + \hat{\boldsymbol{\mu}}^w.$$

Note that the sums of squares of the robust between- and within-parts do not add up to the total sum of squares, a property which does hold for classical MSCA-P. This is however to be expected as such ANOVA decompositions only hold for least-squares estimators, which are known not to be robust. (In robust statistics we do not want to add all the squares when some of them are due to outliers.) But when robust MSCA-P is used in a sequential procedure in which first the outliers are removed and next an MSCA model is fitted, the decomposition will hold on the reduced data set.

The robust analysis allows the detection of level 1 and level 2 outliers by means of several diagnostics. The level 1 outliers (outlying measurement occasions) are flagged by the within-subjects analysis, whereas the level 2 outliers (outlying subjects) are detected by the between-subjects analysis. Both robust PCA analyses yield an *outlier map*, which for each observation plots its orthogonal distance to the PCA subspace versus its robust distance within the PCA subspace. The orthogonal distance is defined as in Eq. (6) with $\hat{\boldsymbol{y}}_i$ the projection of $\boldsymbol{y}_i$ on the ROBPCA subspace. The score distance is the robust distance as in Eq. (7) but now with respect to the reweighted MCD estimates $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}_R$. A cutoff value for the orthogonal distances can again be derived based on their values to the power of 2/3. For the scores distances, we use $\sqrt{\chi^2_{Q,0.975}}$. These cutoff values allow the separation of the regular data points from the outliers. The outlier maps and their interpretation will be discussed for the real data examples in Section 4.

In the examples below we will first standardize the data and set $\alpha = 0.5$. We analyze the data in MATLAB® (The MathWorks, Inc.) with the robust functions from the LIBRA toolbox [16]. The l1median function implements the algorithm developed in [17]. Users of *R* can employ the function l1median in the *R* package pcaPP and the function PcaHubert in the *R* package rrcov [18].

4. Examples

4.1. Eating disorder data

As a first example we analyze data from a study on eating disorders, conducted by [2]. These authors studied 32 female patients of the specialized inpatient eating disorder unit of the University Psychiatric Centre in Leuven, Belgium. Of these patients, 19 suffered from anorexia nervosa and 13 from bulimia nervosa. The study focused on the relations between the drive for thinness, affect, and physical activity of these patients in daily life (for more details, see [19]). To this end, the patients were asked to fill out a questionnaire at nine randomly selected times a day, during one week. The questionnaire consisted of 22 items that aimed at measuring the momentary drive for thinness, positive and negative emotional states, urge to be physically active, and actual physical activity. An electronic device was used to signal the patients that they had to fill out the questionnaire. On average, the patients reacted to 45.60 of the 63 signals ($SD = 10.78$), yielding a 1459 observations by 22 variables data matrix $\boldsymbol{X}$, consisting of 32 matrices $\boldsymbol{X}_i$, one for each patient.

Author's personal copy

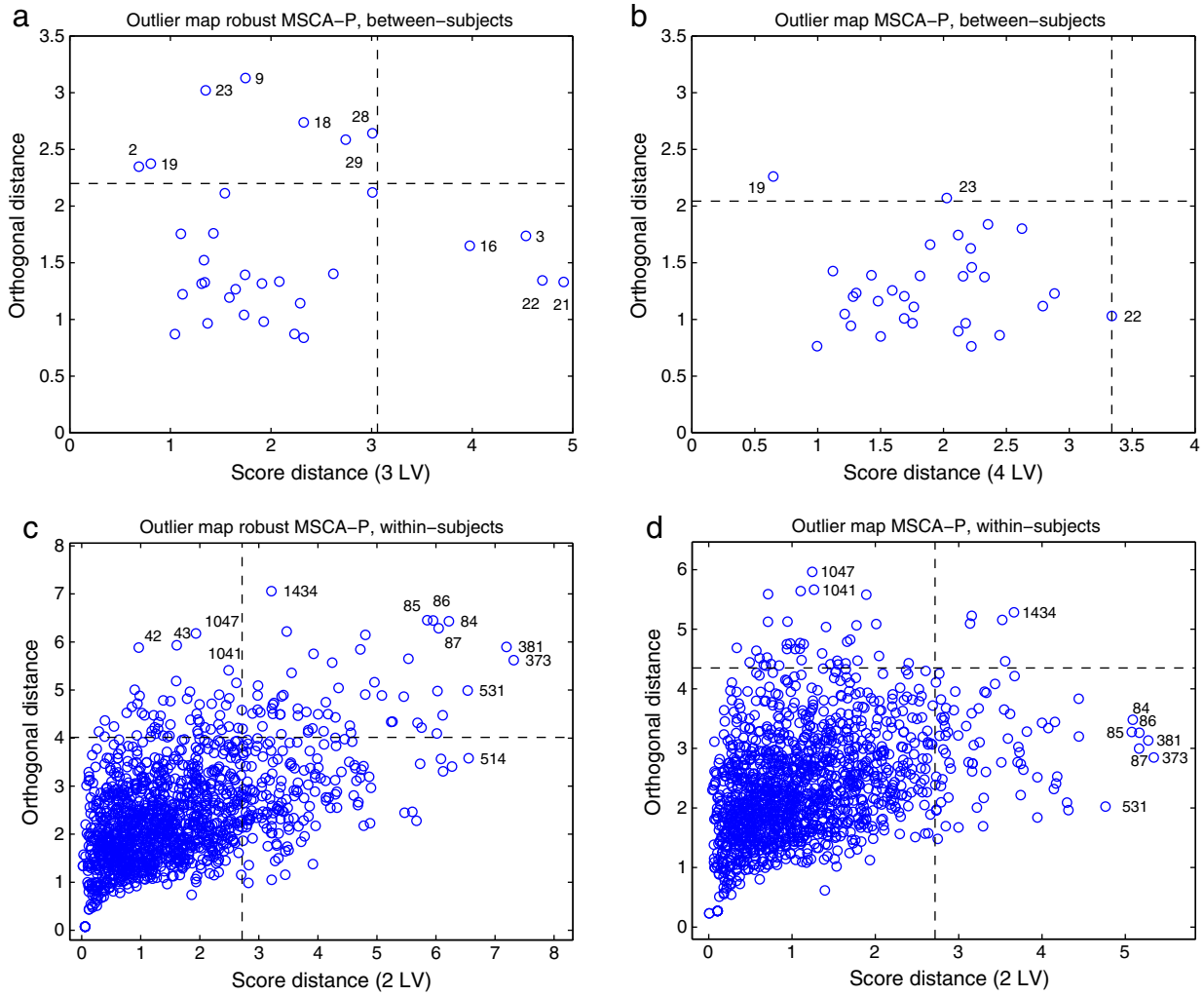


Fig. 1. Eating disorder data: between-subjects (top) and within-subjects (bottom) outlier maps for robust MSCA-P (left) and classical MSCA-P (right).

We first report the results of the between-subjects analysis. For the robust analysis, we select $Q_b = 3$ based on a scree plot of the robust eigenvalues. The outlier map from the robust MSCA-P is shown in Fig. 1(a). The labels on this plot correspond to the patients' number. The dotted lines are drawn at the cutoff values for the score and orthogonal distances. They separate the regular data points from the outliers. We see that several patients have a deviating center. Looking at the raw data, we noted for example that patient 21 had large values for the last variables. Patient 22 is atypical as she had large values for the first four variables, and mostly zero entries for the remaining ones. The classical MSCA-P used the non-robust mean values of each patient. Here, we select four components following the analyses reported by [20]. Its outlier map in Fig. 1(b) has three patients outside the cutoff lines, but their outlyingness is not pronounced. Comparing the robust and classical unrotated loadings (not shown), we see that mostly the loadings of the second and the third component are different. Indeed, the congruence between the first robust loading vector and the first classical loading vector (i.e.. the cosine of the angle between them) is 0.909, whereas it is only 0.289 and 0.242 for the second and third loading vectors.

Based on [20] and on the robust eigenvalues, the within-subjects analyses were conducted with $Q_w = 2$ for both MSCA-P and the robust MSCA-P. These analyses yield the outlier maps in Fig. 1(c–d). Here, the labels correspond to the 1459 measurement occasions. We note important differences between the robust and the classical analysis. Robust MSCA-P finds several points (such as 1047) whose orthogonal distance sticks out whereas their score distance doesn't. Such points are called 'orthogonal outliers'. It also detects many points (such as 86) with outlying score distance and at the same time outlying orthogonal distance. Such points are called 'bad leverage points' because they lie far from the subspace formed by the majority, and their projections on that subspace also lie far from the projection of the majority. As the mechanical term 'leverage' suggests, bad leverage points tend to distort the classical PCA result. Finally there are some 'good leverage points' (such as 514) whose score distance stands out but their orthogonal distance doesn't. A closer analysis of the distances revealed that the outlying measurement occasions occurred among many patients. Patients 5, 7 and 24 had proportionally the most outlying measurement occasions (54%, 63% and 45% respectively). This could indicate that their within-subjects correlation structure is different.

By contrast, the outlier map from classical MSCA-P looks quite different. The classical analysis clearly sees the abovementioned bad leverage points as good leverage points, which implies that classical MSCA-P was highly influenced (rotated) by them. The unrotated within-subjects loadings are quite similar for the first component (congruence 0.996) but somewhat different for the second component (congruence 0.864). From Table 1 we see that mainly the last 6 variables are weighted differently. This is interesting because these variables were the main focus of the study of [2], who wanted to unravel the processes beneath the excessive exercise patterns of patients with eating disorders. In the robust analysis the variables that refer to the urge to be active get a higher weight than the items that indicate

Author's personal copy

Table 1
Robust and classical unrotated within loadings of the eating disorder data set.

Variable	Within subjects			
	Robust		Classical	
Pleased	−.40	−.07	−.38	−.00
Happy	−.37	−.07	−.35	.01
Appreciated	−.23	−.06	−.22	−.01
Love	−.23	−.12	−.23	−.03
Sad	.38	.01	.36	−.04
Angry	.35	−.04	.35	−.09
Lonely	.26	.01	.25	−.03
Anxious	.20	−.09	.20	−.07
Tense	.25	−.12	.27	−.08
Guilty	.19	−.06	.19	−.04
Irritated	.24	−.10	.26	−.07
Ashamed	.16	−.07	.16	−.04
Feel fat	.10	−.13	.12	−.07
Feel ugly	.11	−.10	.12	−.06
Want to burn calories	.03	−.26	.04	−.17
Want to lose weight	.07	−.14	.09	−.08
Want to be active	−.01	−.42	−.00	−.25
Want to move	−.01	−.44	.00	−.26
Want to sport	.00	−.43	.01	−.28
Am active	−.09	−.35	−.12	−.49
Am moving	−.05	−.32	−.10	−.51
Am sporting	−.03	−.18	−.09	−.47

actual activity, whereas the opposite is the case for the classical analysis. This makes sense as actual physical activity was partly restricted by the therapeutic program, implying that the urge items may more reliably measure the symptoms under study.

4.2. Herring data

As a second example we apply robust MSCA-P to the herring data available from http://www.models.kvl.dk/Ripening_of_Herring. The data come from ripening experiments with salted herring, collected at three laboratories in Denmark, Iceland and Norway [21]. The data contain $J = 10$ physicochemical measurements (such as protein content and pH) which reflect the properties of the ripening process at different time points. As different countries, stocks and preprocessing methods are used, the full data matrix can be considered as a multilevel data set in which the different countries, stocks and treatment conditions constitute the level 2 measurements.

Since the data set contains many missing values, we first removed those level 2 data for which at least one variable was fully unobserved. Next, we removed 4 observations with all or all but one missing value, yielding a data set with $N = 237$ observations pertaining to $I = 14$ conditions. First we imputed the remaining missing values. Specifically, each missing score was replaced by the median score of the other observations within the same condition on the variable under study. This is our final data set $\boldsymbol{X}$ to which we apply MSCA-P and robust MSCA-P.

Here we focus on the within-subjects analysis. (A between-subjects analysis would only have 14 points in 10 dimensions, making it difficult to draw meaningful conclusions.) Based on the scree plots we retained $Q_w = 4$ components for robust and classical MSCA-P. The resulting outlier map for the robust analysis is displayed in Fig. 2(a). We immediately spot many observations with unusual large score and/or orthogonal distances. Most prominent are the red-colored observations with largest orthogonal distance. They correspond to all the time-zero measurements in the data set. In [21] it was mentioned that these samples behave significantly differently from the others, though similarly across countries. Our robust MSCA-P finds these outlying measurement occasions by a single analysis and through one diagnostic plot.

On the other hand, classical MSCA-P does not find them automatically as the outlier map in Fig. 2(b) does not show very deviating measurements and clearly cannot distinguish between the time-zero samples and the others. Note that similar results were obtained for other choices of Q_w.

5. Simulation study

We perform a simulation study to assess how well classical and robust MSCA-P succeed in detecting outlying measurement occasions and subjects, and how much the classical estimates of the different MSCA variants are deformed by these outliers. Specifically, based on the experiment in [22], we generated data according to the four MSCA variants. To these data we added outliers at the between-subjects level and/or at the within-subjects level.

5.1. Design

In this study the population offsets ($\boldsymbol{m}$) were set to zero, the number of subjects to 40, the number of variables to 18, and the number of between-subjects and within-subjects components to two. For each subject, the number of measurement occasions was sampled uniformly between 30 and 70. The measurement occasions and the

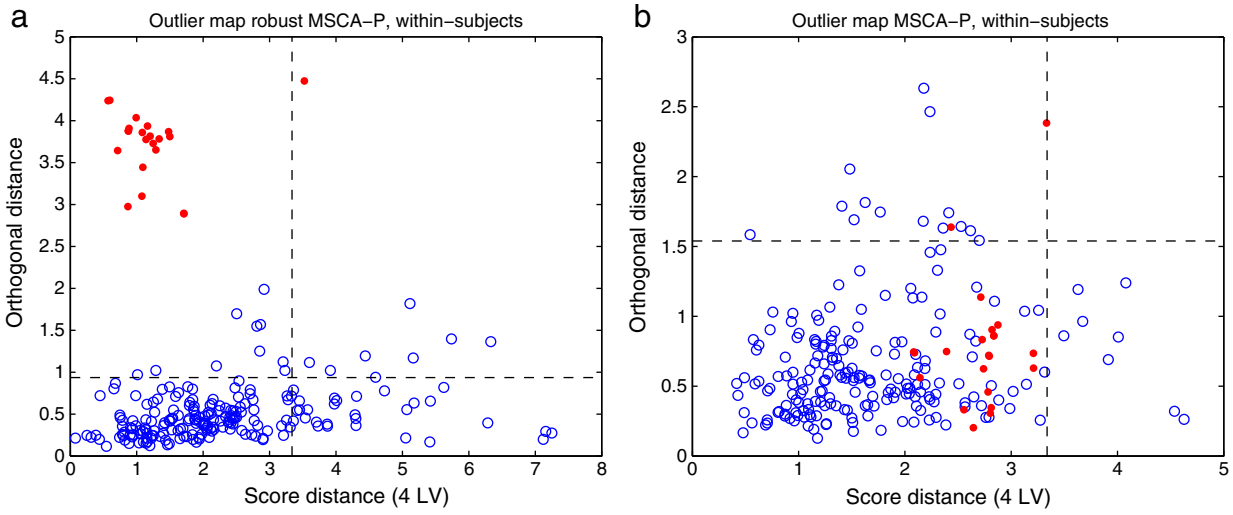


Fig. 2. Herring data within-subjects: outlier map for robust MSCA-P (left) and classical MSCA-P (right). The red dots correspond to the time-zero samples.

Author's personal copy

subjects were considered random samples from a population of measurement occasions and subjects, respectively (for more details, see [23]).

For each subject a simulated data matrix $\mathbf{X}\text{sim}_i$ was created as follows:

$$\mathbf{X}\text{sim}_i = \mathbf{1}_N \boldsymbol{f}_i^b \left(\boldsymbol{B}^b\right)^T + \boldsymbol{F}_i^w \left(\boldsymbol{B}^w\right)^T + \boldsymbol{E}_i$$

where the entries of $\boldsymbol{f}_i^b$ and $\boldsymbol{E}_i$ were sampled from a standard normal distribution and those of $\boldsymbol{F}_i^w$ from a multivariate normal distribution, the covariance matrix of which depended on the MSCA variant used (see below). $\boldsymbol{B}^b$ and $\boldsymbol{B}^w$ were constructed in such a way that the first between- and within-subjects component comprises the first six variables only, and the second one the seventh to twelfth variables. The last six variables constitute a third between- and within-subjects component that is used to add outliers (see below), implying that regular subjects and measurement occasions have a zero score on this component. Note that the non-zero between-subjects loadings equaled $\sqrt{.2}$ and the non-zero within-subject loadings $\sqrt{.8}$, implying that the proportion of structural variance that is contributed by the between-subjects and within-subjects parts of the simulated data was set to .20 and .80.

The following three factors were fully crossed:

- level at which outliers occur (3 levels): were the outliers added at the between-subjects level (B), at the within-subjects level (W), or at both levels (BW). In the B and BW case, the data of the first eight subjects were altered by sampling a score on the third between-subjects component from a standard normal distribution and adding 4 to their three between-subjects component scores. In the W and BW case, the data of 350 randomly selected measurement occasions were altered in the same way.
- MSCA variant used (4 levels): were the data generated according to an MSCA-ECP, MSCA-IND, MSCA-PF2, or MSCA-P model. In the MSCA-IND, MSCA-PF2 and MSCA-P cases, the variances of the within-subjects component scores were uniformly sampled between .25 and 1.75. Moreover, in the MSCA-PF2 case the correlations of the within-subjects components were set to .25, whereas in the MSCA-P case these correlations were uniformly sampled between −.50 and .50.
- amount of error: 10% or 30% (2 levels).

In each of the 24 cells of the design we had 25 replicates, leading to 600 generated data matrices. Each simulated data matrix was first analyzed with robust and classical MSCA-P, using two between-subjects and two within-subjects components. Subsequently, the detected outlying subjects and measurement occasions (taking both score distance and orthogonal distance into account) were removed from the data. Both types of remaining data (without 'classical' outliers and without 'robust' outliers) were analyzed with the appropriate classical MSCA approach, again with two between-subjects and two within-subjects components and using five random starts and one start based on the singular value decomposition of the data.

5.2. Results

First, we examined how many of the outlying subjects and measurement occasions were classified as outliers by robust and classical MSCA-P. From the proportion of undetected outliers in Table 2, it is clear that robust MSCA-P finds most outliers, whereas classical MSCA-P only detects 20 to 30%.

Next, we inspected how well the appropriate classical MSCA analysis on the data without robust or classical outliers recovers the between-subjects and within-subjects component loadings. To this end we computed the Tucker congruence coefficient between the simulated loadings and the estimated ones, where values larger than .95 indicate that the components can be considered identical. Note that the rotational freedom of the between-subjects components and the MSCA-ECP and MSCA-P within-subjects components was taken into account by Procrustes rotating the estimated loadings towards the simulated ones. Fig. 3 shows that the recovery of the loadings deteriorates when outliers were present and a classical approach to detecting them was used, whereas the robust analysis correctly removed influential outliers. The effects of the other factors (error level and MSCA variant used) were quite small.

Table 2
Proportion of undetected outliers.

Level	Variant	Error	Between		Within	
			Robust	Classical	Robust	Classical
B	ECP	.1	0	.7343	–	–
B	ECP	.3	0	.7160	–	–
B	IND	.1	0	.7488	–	–
B	IND	.3	0	.7824	–	–
B	PF2	.1	0	.7631	–	–
B	PF2	.3	0	.8049	–	–
B	P	.1	0	.7558	–	–
B	P	.3	0	.7364	–	–
W	ECP	.1	–	–	0	.7360
W	ECP	.3	–	–	.0029	.8086
W	IND	.1	–	–	0	.7440
W	IND	.3	–	–	.0040	.8073
W	PF2	.1	–	–	0	.7211
W	PF1	.3	–	–	.0110	.7779
W	P	.1	–	–	.0001	.7381
W	P	.3	–	–	.0032	.8046
BW	ECP	.1	0	.7498	.0001	.7419
BW	ECP	.3	0	.7814	.0048	.8119
BW	IND	.1	0	.7895	0	.7515
BW	IND	.3	0	.7352	.0040	.8240
BW	PF2	.1	0	.7639	.0001	.7465
BW	PF2	.3	0	.7841	.0176	.8088
BW	P	.1	0	.7803	0	.7515
BW	P	.3	0	.7091	.0054	.8127

6. Conclusion

In this paper we presented robust MSCA-P. As this technique uses robust estimators in all phases of the analysis, the obtained between- and within-subjects results are less influenced by outlying subjects and outlying measurement occasions. Users may interpret the robust results directly, as we did in the examples. Alternatively, as illustrated in the simulation study, they may use the robust technique as a screening method for detecting outliers, remove these outliers, and perform the classical analysis on the remaining data. Or, if they prefer a more gradual approach, they can conduct a weighted classical analysis, in which weights are assigned to the data entries in function of the outlyingness of the corresponding subject or measurement occasion. An advantage of such a weighted classical analysis (possibly with zero–one weights) is that model selection heuristics are available for determining which variant and which number of between- and within-subjects components yield a good but parsimonious description of the data [22], and that bootstrap techniques have been proposed for obtaining confidence intervals for the different model parameters [23].

As we saw in the examples, the within-subjects ROBPCA reveals outlying measurement occasions. Since these occasions are nested in the subjects, this analysis also yields information on the subjects. Specifically, it often happens that the within-subjects correlation structure is not shared by all subjects. Such deviant subjects can be detected by inspecting the proportion of outlying occasions per subject. If this proportion is high (say above 50%) for multiple subjects, it may be worthwhile to replace the within-subjects PCA by a clusterwise simultaneous component analysis, to examine whether

Author's personal copy

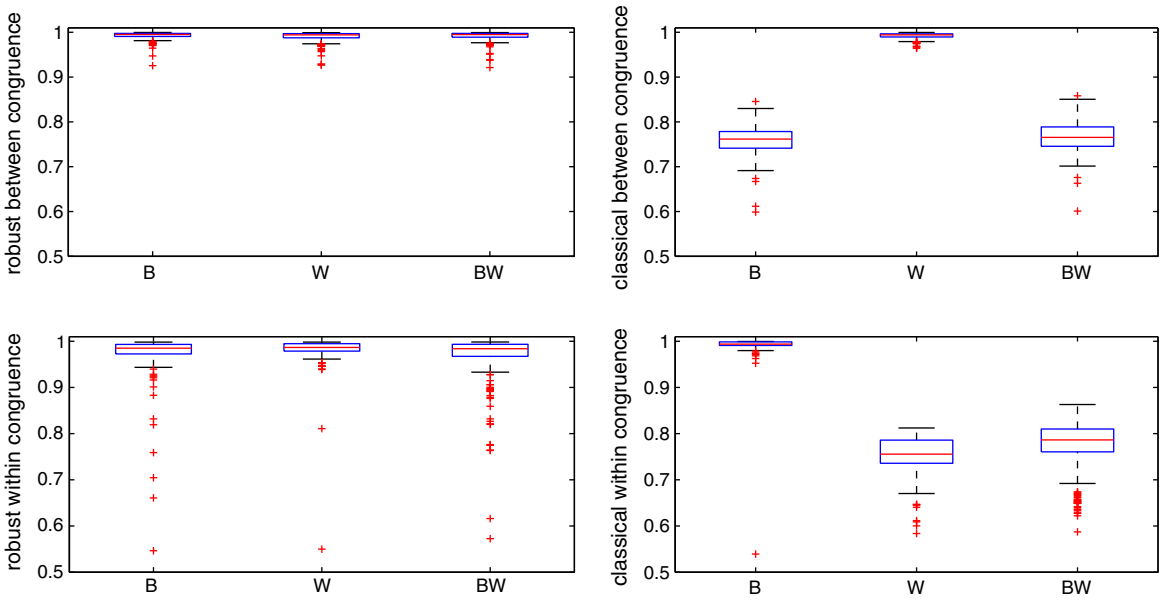


Fig. 3. Mean between-subjects (top) and within-subjects (bottom) congruence for the appropriate classical MSCA analysis on the data without the robust (left) or classical (right) outliers, as a function of the outlier level (B, W, or BW).

the data contain subgroups of subjects which are characterized by a different correlation structure and thus by different latent processes [19,24]. MATLAB-based software for conducting this analysis has been provided by [25].

Acknowledgment

The research reported in this paper was supported by the IAP research network grant nr. P7/06 of the Belgian Government (Belgian Science Policy) and by the KU Leuven grant GOA/12/14. We also thank the reviewers for their useful comments and suggestions which have improved our paper.

References

[1] R. Bro, E. Qannari, H. Kiers, T. Naes, M. Frost, Multi-way models for sensory profiling data, Journal of Chemometrics 22 (2008) 36–45.
[2] K. Vansteelandt, F. Rijmen, G. Pieters, M. Probst, J. Vanderlinden, Drive for thinness, affect regulation and physical activity in eating disorders: a daily life study, Behaviour Research and Therapy 45 (2007) 1717–1734.
[3] K. Vansteelandt, I. Van Mechelen, J. Nezlek, The co-occurrence of emotions in daily life: a multilevel approach, Journal of Research in Personality 39 (2005) 325–335.
[4] M. Timmerman, Multilevel component analysis, British Journal of Mathematical and Statistical Psychology 59 (2006) 301–320.
[5] M. Timmerman, H. Kiers, Four simultaneous component models of multivariate time series from more than one subject to model intraindividual and interindividual differences, Psychometrika 86 (2003) 105–122.
[6] P. Kuppens, E. Ceulemans, M. Timmerman, E. Diener, C. Kim-Prieto, Universal intracultural, and intercultural dimensions of the recalled frequency of emotional experience, Journal of Cross Cultural Psychology 37 (2006) 491–515.
[7] J. Jansen, H. Hoefsloot, J. van der Greef, M. Timmerman, A. Smilde, Multilevel component analysis of time-resolved metabolic fingerprinting, Analytica Chimica Acta 530 (2005) 173–183.
[8] J. Stouten, E. Ceulemans, M. Timmerman, A. Van Hiel, Tolerance of justice violations: the effects of need on emotional reactions after violating equality in social dilemmas, Journal of Applied Social Psychology 41 (2011) 357–380.
[9] O. De Noord, E. Theobald, Multilevel component analysis and multilevel PLS of chemical process data, Journal of Chemometrics 19 (2005) 301–307.
[10] P. Rousseeuw, M. Debruyne, S. Engelen, M. Hubert, Robustness and outlier detection in chemometrics, Critical Reviews in Analytical Chemistry 36 (2006) 221–242.
[11] P. Rousseeuw, A. Leroy, Robust Regression and Outlier Detection, Wiley-Interscience, New York, 1987.
[12] H. Fritz, P. Filzmoser, C. Croux, A comparison of algorithms for the multivariate L1-median, Computational Statistics 27 (2012) 393–410.
[13] M. Hubert, P. Rousseeuw, K. Vanden Branden, ROBPCA: a new approach to robust principal components analysis, Technometrics 47 (2005) 64–79.
[14] P. Rousseeuw, Least median of squares regression, Journal of the American Statistical Association 79 (1984) 871–880.
[15] P. Rousseeuw, K. Van Driessen, A fast algorithm for the minimum covariance determinant estimator, Technometrics 41 (1999) 212–223.
[16] S. Verboven, M. Hubert, LIBRA: a MATLAB library for robust analysis, Chemometrics and Intelligent Laboratory Systems 75 (2005) 127–136.
[17] O. Hössjer, C. Croux, Generalizing univariate signed rank statistics for testing and estimating a multivariate location parameter, Journal of Nonparametric Statistics 4 (1995) 293–308.
[18] V. Todorov, P. Filzmoser, An object-oriented framework for robust multivariate analysis, Journal of Statistical Software 32 (2009) 1–47.
[19] K. De Roover, E. Ceulemans, M. Timmerman, K. Vansteelandt, J. Stouten, P. Onghena, Clusterwise simultaneous component analysis for analyzing structural differences in multivariate multiblock data, Psychological Methods 17 (2012) 100–119.
[20] M. Timmerman, E. Ceulemans, A. Lichtwarck-Aschoff, K. Vansteelandt, Multilevel component analysis for studying intra-individual variability and inter-individual differences, in: J. Valsiner, P. Molenaar, M. Lyra, N. Chaudhary (Eds.), Dynamic Process Methodology in the Social and Developmental Sciences, Springer, New York, 2009, pp. 291–318.
[21] R. Bro, H. Nielsen, G. Stefansson, T. Skara, A phenomenological study of ripening of salted herring. Assessing homogeneity of data from different countries and laboratories, Journal of Chemometrics 16 (2002) 81–88.
[22] E. Ceulemans, M. Timmerman, H. Kiers, The Chull procedure for selecting among multilevel component solutions, Chemometrics and Intelligent Laboratory Systems 106 (2011) 12–20.
[23] M. Timmerman, H. Kiers, A. Smilde, E. Ceulemans, J. Stouten, Bootstrap confidence intervals in multi-level simultaneous component analysis, British Journal of Mathematical and Statistical Psychology 62 (2009) 299–318.
[24] K. De Roover, E. Ceulemans, M. 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.
[25] K. De Roover, E. Ceulemans, M. Timmerman, How to perform multiblock component analysis in practice, Behavior Research Methods 44 (2012) 41–56.