# Clusterwise analysis for multiblock component methods

Stéphanie Bougeard, Hervé Abdi, Gilbert Saporta, Ndèye Niang

# Clusterwise analysis for multiblock component methods

Stéphanie Bougeard, Hervé Abdi, Gilbert Saporta, Ndèye Niang

# Clusterwise analysis for multiblock component methods

**Stéphanie Bougeard**[1] · **Hervé Abdi**[2] · **Gilbert Saporta**[3] · **Ndèye Niang**[3]

Received: 20 October 2016 / Revised: 10 October 2017 / Accepted: 16 October 2017


**Abstract** Multiblock component methods are applied to data sets for which several blocks of variables are measured on a same set of observations with the goal to analyze the relationships between these blocks of variables. In this article, we focus on multiblock component methods that integrate the information found in several blocks of explanatory variables in order to describe and explain one set of dependent variables. In the following, multiblock PLS and multiblock redundancy analysis are chosen, as particular cases of multiblock component methods when one set of variables is explained by a set of predictor variables that is organized into blocks. Because these multiblock techniques assume that the observations come from a homogeneous population they will provide suboptimal results when the observations actually come from different populations. A strategy to palliate this problem—presented in this article—is to use a technique such as clusterwise regression in order to identify homogeneous clusters of observations. This approach creates two new methods that provide clusters that have their own sets of regression coefficients. This combination of clustering and regres-

✉ Stéphanie Bougeard
stephanie.bougeard@anses.fr

Hervé Abdi
herve@utdallas.edu

Gilbert Saporta
gilbert.saporta@cnam.fr

Ndèye Niang
n-deye.niang_keita@cnam.fr

1 Department of Epidemiology, Anses (French agency for food, environmental and occupational health safety), 22440 Ploufragan, France

2 The University of Texas at Dallas, Richardson, TX 75083-0688, USA

3 CEDRIC CNAM, 292 rue St Martin, 75141 Paris Cedex 03, France

Published online: 08 November 2017

sion improves the overall quality of the prediction and facilitates the interpretation. In addition, the minimization of a well-defined criterion—by means of a sequential algorithm—ensures that the algorithm converges monotonously. Finally, the proposed method is distribution-free and can be used when the explanatory variables outnumber the observations within clusters. The proposed clusterwise multiblock methods are illustrated with of a simulation study and a (simulated) example from marketing.

**Keywords** Multiblock component method · Clusterwise regression · Typological regression · Cluster analysis · Dimension reduction

**Mathematics Subject Classification** 62H30 · 62H25 · 91C20

## 1 Introduction

A common problem in the analysis of large data sets is to explore the relationships between several blocks of variables—measured on the same observations—connected according to a relevant (sometimes complex) user-defined arrow diagram. This problem is typically handled with multiblock component methods such as PLS Path Modeling (PLS-PM) (Wold 1985; Lohmoller 1989), Generalized Structured Component Analysis (GSCA) (Hwang and Takane 2004), Regularized Generalized Canonical Correlation Analysis (RGCCA) (Tenenhaus and Tenenhaus 2011) or THEmatic Model Exploration (THEME) (Bry et al. 2012). In this paper, we address a more restrictive problem—called the $(K + 1)$ problem—often encountered in practice, where one set of (dependent) variables is explained by an another set of (explanatory) variables that is organized into $K$ blocks. This $(K + 1)$ problem can be handled (among other techniques, see (Vivien 2002) for a review) with two component-based techniques such as: (1) Multiblock Partial Least Squares (MBPLS) and (2) Multiblock Redundancy Analysis (MBRA). Multiblock Partial Least Squares (Wold 1984)—a covariance based technique—seeks linear combinations of the variables of the $K$ blocks with maximal covariance with the variables in the $(K + 1)$-th block. By contrast, multiblock Redundancy Analysis—a regression based technique—seeks linear combinations of the variables in the $K$ blocks that best predict the variables in the $(K + 1)$-th block (of "dependent variables"). These two methods can be viewed as particular cases of the multiblock component methods previously cited.

These multiblock techniques—that are used in a variety of fields such as chemometrics, sensometrics, or process monitoring among others—assume that the observations come from a homogeneous population. However, it often happens that—unbeknownst to the analyst—the observations originate from different populations, and, in this case, the statistical model computed on the whole set of observations may be of poor quality. A way to palliate this problem is to identify sub-populations (i.e., clusters) of observations and compute one component-based model per cluster of observations. A standard approach to obtain clusters within a regression framework is clusterwise regression (a.k.a. typological regression, see, e.g., (DeSarbo and Cron 1988; Spath 1979)). Clusterwise regression assumes that there is an underlying group structure of the observations and that each cluster can be revealed by the fit of a specific regres-

sion model. In a more formal way, clusterwise regression simultaneously looks for a partition of the observations into clusters and minimizes the sum of squared error computed over all the clusters.

In this article, we combine MBPLS and MBRA with a clusterwise approach to create two new methods that can find the underlying structure of the observations and provide each cluster of observations with its own set of regression coefficients. This combination of clustering and regression improves the fit of the regression, facilitates interpretation, and is particularly well tailored for prediction (i.e., rather than, e.g., for modeling). The remainder of the paper is organized in three sections. In the *method* section, multiblock component methods (i.e., MBPLS and MBRA) are briefly reviewed (and detailed in Appendices 1 and 2), the proposed new methodology is presented and a solution for the prediction of new observations is exposed and discussed. In the *applications* section, clusterwise multiblock methods and their properties are illustrated with both a simulation study and the analysis of a (simulated) example from marketing and consumer satisfaction. Finally, *conclusions* and directions for future work are given.

## 2 Methods

### 2.1 Notations

Matrices are denoted by bold upper-case letters (e.g., $\mathbf{X}$) and column vectors by bold lower-case letters (e.g., $\mathbf{x}$). The symbol $^{\mathsf{T}}$ (e.g., $\mathbf{x}^{\mathsf{T}}$) denotes the matrix transpose operation. The identity matrix is denoted $\mathbf{I}$.

The data describe a sample of $N$ observations and comprise one $N$ by $Q$ matrix of dependent variables denoted $\mathbf{Y} = (y_1, \ldots, y_Q)$ and a set of $K$ matrices (also called blocks) of explanatory variables each of order $N$ by $J^k$ and each denoted $\mathbf{X}^k = (x_1^k, \ldots, x_{J^k}^k)$. The orthogonal projector of $\mathbb{R}^N$ onto the column space of $\mathbf{X}^k$, denoted $\mathbf{P}^k$, is defined as

$$\mathbf{P}^k = \mathbf{X}^k \left( \left( \mathbf{X}^k \right)^{\mathsf{T}} \mathbf{X}^k \right)^{-1} \left( \mathbf{X}^k \right)^{\mathsf{T}} . \tag{1}$$

The $K$ matrices $\mathbf{X}^k$ are stored in the matrix $\mathbf{X} = (\mathbf{X}^1, \ldots, \mathbf{X}^K)$ of order $N$ by $J$ (i.e., $J = \sum_{k=1}^{K} J^k$). For convenience (but without loss of generality), matrices $\mathbf{Y}$ and $\mathbf{X}$ are assumed to be column centered and normalized (i.e., for each variable (column), the sum over all $N$ observations is zero and its norm is one). However, variables (within a given block of variables) could also be weighted (e.g., to handle differences between block sizes) according to the number of variables of their block (Westerhuis and Coenegracht 1997) or to block inertia (Bougeard and Cardinal 2014). Specifically, blocks could be weighted according to their inertia by replacing $\mathbf{X}^k$ and $\mathbf{Y}$ respectively with

$$\tilde{\mathbf{X}}^k = \frac{\mathbf{X}^k}{\sqrt{K \frac{\text{trace}((\mathbf{X}^k)^\top \mathbf{X}^k)}{N-1}}} \quad \text{and} \quad \tilde{\mathbf{Y}} = \frac{\mathbf{Y}}{\sqrt{\frac{\text{trace}(\mathbf{Y}^\top \mathbf{Y})}{N-1}}}.$$

In the following (i.e., simulation study in Sect. 3.1), block weighting is considered as a factor of interest.

In multiblock analysis, the important information from each matrix $\mathbf{Y}$ and $\mathbf{X} = (\mathbf{X}^1, \ldots, \mathbf{X}^K)$ is expressed by a component (or factor) that results as a linear combination of the columns of $\mathbf{Y}$ and $\mathbf{X}$, respectively, and will be denoted by $\mathbf{u}$ and $\mathbf{t}^k$ computed as

$$\mathbf{u} = \mathbf{Y}\mathbf{v} \quad \text{and} \quad \mathbf{t}^k = \mathbf{X}^k \mathbf{w}^k \quad \text{for} \quad k = 1, \ldots, K, \tag{2}$$

where the coefficient vectors $\mathbf{v}$ and $\mathbf{w}^k$ are called the vectors of loadings of, respectively, the $\mathbf{Y}$ and $\mathbf{X}^k$ matrices. The global component that sums up all the explanatory variables, denoted $\mathbf{t}$, is computed as

$$\mathbf{t} = \mathbf{X}\mathbf{w}\,. \tag{3}$$

As relationships between datasets are rarely unidimensional, higher order components are often needed. These higher order components are denoted with the superscript $^{(h)}$; so, for example, the $h$th component for the whole $\mathbf{X}$ matrix is denoted $\mathbf{t}^{(h)}$ with $h = (1, \ldots, H)$, $H$ being the maximum number of dimensions kept for the analysis.

The $N$ vectors of observations are assumed to be clustered into $G$ unknown clusters denoted $\{\mathscr{C}_1, \ldots, \mathscr{C}_G\}$, with $N_g$ denoting the number of observations of the $g$th cluster (and so $N = \sum_{g=1}^{G} N_g$). The matrices $\mathbf{Y}_g$ and $\mathbf{X}_g^k$ denote the data matrices of the $g$th group of (respectively) the dependent variables (i.e., $\mathbf{Y}$) and the $k$th block of explanatory ones (i.e., $\mathbf{X}^k$). More precisely, the matrix $\mathbf{Y}_g$ is of size $(N_g \times Q)$ and the matrix $\mathbf{X}_g$ of size $(N_g \times J^k)$. Figure 1 illustrates the main notations used in this article.

### 2.2 Standard multiblock component methods

As mentioned earlier, we focus on multiblock component methods that integrate the information in several explanatory blocks in order to explain or predict a (single) matrix of dependent variables. Two of the most well-known of these methods (for reviews see (Vivien 2002; Kissita 2003)) are multiblock PLS for the case of a single dependent dataset (MBPLS) and multiblock redundancy analysis (MBRA). Incidentally, both methods can be viewed as special cases of regularized generalized canonical correlation analysis (RGCCA) and, to a lesser extent, of PLS-PM, when choosing specific normalization constraints on explanatory and dependent components. These two methods can also be viewed as particular cases of THEME when relationships between explanatory and dependent components are studied block by block. These methods are component-based analyses for which each dataset (i.e., $\mathbf{Y}$ and $(\mathbf{X}^1, \ldots, \mathbf{X}^K)$), is represented by a (optimum) component denoted (respectively) $\mathbf{u}$ = and $\mathbf{t}^k$ for $k = (1, \ldots, K)$ for the first-order solution. In addition, a global component, denoted

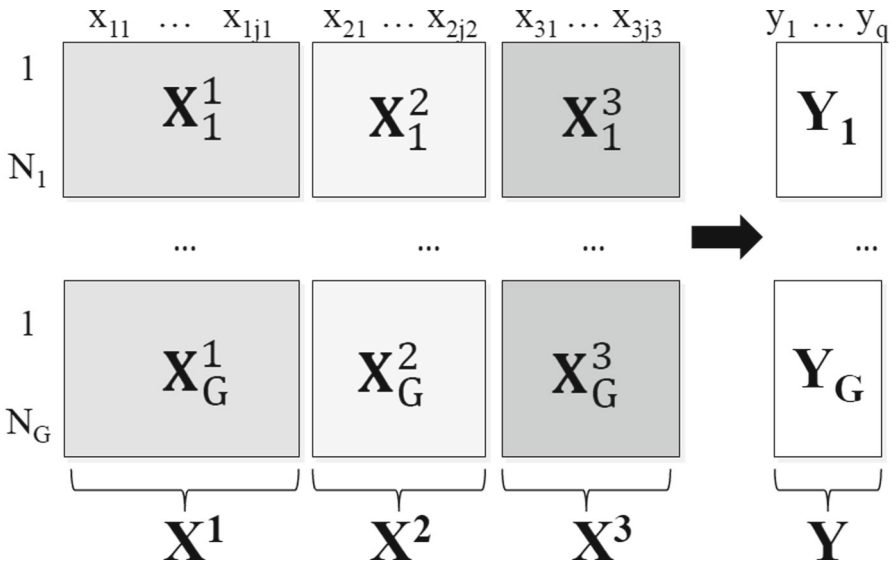


**Fig. 1** Graphical display of clusterwise and multiblock data. In this Figure, the observations are (for convenience) listed clusterwise but this arrangement is not genuine (i.e., does not reflect an unknown ground-truth); it could be obtained after a suitable permutations of the rows of matrix **X** (as obtained, e.g., from one of the algorithms described in this paper). In this Figure, ($K = 3$) blocks are illustrated but it could be extended to a general $K$

**t**, represents the whole set of explanatory variables. Because these methods use components, they can handle multicollinear datasets and therefore can be used when the explanatory variables outnumber the observations. The details of the methods is given in Appendix 1 for MBPLS and Appendix 2 for MBRA.

### 2.3 Existing clusterwise methods

Existing clusterwise methods seek clusters within a regression framework while simultaneously minimizing the sum of squared error computed over all the clusters. These methods can be viewed as extensions of the $K$-means algorithm to the problem of prediction for which the sum of square of prediction errors is minimized instead of the within-inertia. As in standard regression, ordinary least squares or maximum likelihood estimations can be used to assess regression coefficients. Bock (1969), Diday (1976) and Spath (1979) proposed $K$-means like algorithms based on a least square error criterion while DeSarbo and Cron (1988) proposed a method based on conditional mixture likelihood. In the same vein, a multivariate regression for heterogeneous data which takes account of both the between- and the within-cluster variability has also been proposed (Vicari and Vichi 2013). Extensions to principal component regression (PCR) (Charles 1977) and partial least square regression (PLS) (Vinzi et al. 2005; Preda and Saporta 2005) have also been proposed to deal with multicollinearity, small sample size, or large number of variables.

In the framework of component-based path-modeling methods, several clusterwise methods have been applied in the marketing field (for an early review, refer to Sarstedt

2008). Among these methods, some interesting and recent ones can be singled out: (1) the widely-used finite-mixture PLS (FIMIX-PLS)—although it assumes multivariate normally distributed data (Hahn et al. 2002)—(2) fuzzy clusterwise generalized structured component analysis (FCGSCA) (Hwang et al. 2007)—which minimizes the sum of all residuals computed over all the clusters—(3) REBUS-PLS—which is based on the hierarchical clustering based on a *closeness* measure defined from the residuals coming from the same models (Vinzi et al. 2009)—and 4) PLS-IRRS which identifies homogeneous clusters that have similar residual values (Schlittgen et al. 2016).

In the field of multigroup analysis—where the groups of observations are known *a priori*—clusterwise simultaneous component analysis (CW-SCA) seeks clusters among groups of observations rather than among observations (De Roover et al. 2012) (variables are considered to belong to a single block). It is worth noting that methods based on likelihood are relevant for data exploration or modeling but do not focus on prediction as dependent values are needed to compute the likelihood.

### 2.4 Proposed clusterwise multiblock regression models

Clusterwise multiblock regression methods extend the domain of application of standard multiblock regression methods by assuming that the $N$ vectors of observations (i.e., rows of $\mathbf{X}$) originate from $G$ different unknown clusters and that each cluster has a specific regression model. Here, we propose two methods derived, respectively, from MBPLS and MBRA: clusterwise MBPLS (CW.MBPLS) and clusterwise MBRA (CW.MBRA). Both methods look for clusters of observations and compute for each cluster a set of regression coefficients.

#### *2.4.1 Properties*

Table 1 summarizes seven properties of the two proposed clusterwise multiblock regression methods.

It can be seen that they share three common properties but each method has some specific properties (see details in Table 1). Some related methods have already been proposed in the fields of clusterwise principal component regression (PCR), ridge regression, or even plain PLS regression (Charles 1977; Vinzi et al. 2005; Preda and Saporta 2005), but these methods do not fulfil Properties 4 (block data), 2 (decreasing monotonicity), or 3 (prediction).

#### *2.4.2 Optimality criteria*

Just like for standard clusterwise methods, the criterion to minimize is the sum of the squared error computed over all the clusters using local multiblock models applied to each cluster. Let $\mathbf{Y}_g$ and $\mathbf{X}_g = \left(\mathbf{X}_g^1, \ldots, \mathbf{X}_g^k, \ldots, \mathbf{X}_g^K\right)$ be the data matrices corresponding to cluster $\mathscr{C}_g$ for (respectively) the dependent and the explanatory variables. Depending upon the methods, multiblock PLS regression or multiblock redundancy analysis will then be applied to each cluster and so each method minimizes a specific criterion.

**Table 1** Clusterwise multiblock properties

| Property | Description | CW.MBPLS | CW.MBRA |
|---|---|---|---|
| 1 | Distribution-free | x | x |
| 2 | Is based on a goodness of fit criterion that decreases monotonically at each iteration | x | x |
| 3 | Allows the prediction of new observations | x | x |
| 4 | Takes into account the block organization of the explanatory variables | | x |
| 5 | Can be used when the explanatory variables per block outnumber the observations within each cluster ($J^k >> N_g$) | | x |
| 6 | Can be used when the explanatory variables outnumber the observations within each cluster ($J >> N_g$) | x | |
| 7 | Deals with multicollinearity within the explanatory dataset $\mathbf{X}$ | x | |

*Clusterwise multiblock* PLS When the multiblock PLS regression model is applied to each cluster, optimality criteria (10) and (18) of MBPLS (given in Appendix 1) are generalized to the clusterwise context, and so the criterion for clusterwise MBPLS for an optimal number $O$ of components to be included in the model is

$$\underset{(\mathbf{c}_1,\ldots,\mathbf{c}_G)}{\arg\min} \sum_{g=1}^{G} \left\| \mathbf{Y}_g - \sum_{h=1}^{O} \mathbf{t}_g^{(h)} \left(\mathbf{c}_g^{(h)}\right)^{\mathsf{T}} \right\|^2$$

with

$$\mathbf{t}_g^{(h)} = \mathbf{X}_g \mathbf{w}_g^{(h)*} \quad \text{and} \quad \mathbf{w}_g^{(h)*} = \prod_{l=1}^{h-1} \left[ \mathbf{I} - \frac{\mathbf{w}_g^{(l)} \left(\mathbf{t}_g^{(l)}\right)^{\mathsf{T}}}{\|\mathbf{t}_g^{(l)}\|^2} \right] \mathbf{w}_g^{(h)} \tag{4}$$

(cf. Eqs. (18) and (19) for standard MBPLS).

*Clusterwise multiblock redundancy analysis* When the multiblock redundancy analysis model is applied to each cluster, optimality criteria (24) and (28) of standard MBRA (given in Appendix 2) are generalized to the clusterwise context, and so the criterion for clusterwise MBRA, for an optimal number $O$ of components to be included in the model is

$$\underset{(\mathbf{c}_1,\ldots,\mathbf{c}_G)}{\arg\min} \sum_{g=1}^{G} \left\| \mathbf{Y}_g - \sum_{h=1}^{O} \sum_{k=1}^{K} a_g^{k(h)} \mathbf{t}_g^{k(h)} \left(\mathbf{c}_g^{(h)}\right)^{\mathsf{T}} \right\|^2$$

with

$$\mathbf{t}_g^{k(h)} = \mathbf{X}_g^{k\,(h-1)} \mathbf{w}_g^{k(h)} \tag{5}$$

where $\mathbf{X}_g^{k\,(h-1)}$ is the residual of the prediction of $\mathbf{X}_g^k$ from the components $(\mathbf{t}_g^{(1)}, \ldots, \mathbf{t}_g^{(h-1)})$ for $(k = 1, \ldots, K)$ and for $(g = 1, \ldots, G)$ (cf. Eq. (28) for standard MBRA).

### *2.4.3 Algorithm*

The clusterwise multiblock regression algorithm searches an optimal partition of the $N$ vectors of observations into $G$ clusters as well as the corresponding set of regression coefficient matrices $(\mathbf{B}_1, \ldots, \mathbf{B}_G)$ that minimize the overall error respectively described in Eq. (4) for CW.MBPLS and Eq. (5) for CW.MBRA. To ensure that the error decreases monotonically at each iteration of the algorithm, we use a sequential algorithm (Spath 1979) rather than a standard batch algorithm (Charles 1977; Preda and Saporta 2005; Vinzi et al. 2005). So, each vector of observation is assigned to its optimal cluster and the overall error is updated whenever one observation switches cluster. To eliminate a potential effect of the order in which the observations enter the algorithm, the observations are randomly sorted prior to each run. This sequential algorithm iterates the following steps for a given number of clusters $G$ and for factor dimensionality $H$:

Step 1 Choose some initial partition of the $N$ vectors of observations into $G$ clusters $\{\mathscr{C}_1, \ldots, \mathscr{C}_G\}$ and a starting observation denoted $n$.

Step 2 For $n \in \mathscr{C}_g$ examine whether there is a cluster $\mathscr{C}_{g'}$ with $g' \neq g$ such that shifting observation $n$ from $\mathscr{C}_g$ to $\mathscr{C}_{g'}$ reduces the overall error given (respectively) in Eq. (4) for CW.MBPLS or (5) for CW.MBRA.

a. Choose the cluster $\mathscr{C}_{g'}$ that gives the maximal reduction of the overall error. Then, remove observation $n$ from cluster $\mathscr{C}_g$ and assign it to cluster $\mathscr{C}_{g'}$

b. Otherwise, go to Step 3

Step 3 Repeat Step 2 for all the observations $n = (1, \ldots, N)$; doing so guarantees that the overall error decreases monotonically at each change in assignation.

Step 4 Get the final assignation of observations into $G$ clusters $\{\mathscr{C}_1, \ldots, \mathscr{C}_G\}$ and the associated regression coefficient matrices $\left(\mathbf{B}_1^{(H)}, \ldots, \mathbf{B}_G^{(H)}\right)$.

### *2.4.4 Parameter selection*

For both methods, the optimal numbers of clusters and dimensions are unknown and must therefore be determined. Because our methods do not refer to a specific probability model, penalized parametric criteria such as BIC or AIC (DeSarbo and Cron 1988) cannot be used. Other penalized criteria such as those proposed by (Shao and Wu 2005)—while theoretically distribution-free—do not work well in practice when the distributions are not multivariate normal. Following references (Charles 1977; Preda and Saporta 2005) and, because the proposed clusterwise multiblock methods can be used to explicitly predict new vectors of observations, we propose to select the unknown parameters on the basis of a ten-fold cross-validation procedure where the optimal $G$ and $O$ parameters are selected to minimize the average root mean square error (RMSE) of prediction (computed with Eq. (23)). The prediction for a new observation is performed as a two-step process: *First,* predict the cluster of the observation

(based on the $\mathbf{X}^k$ matrices) and *second* predict $\mathbf{Y}$ by applying to the observation the predictive model of its assigned cluster. The first step can be achieved by any appropriate supervised classification method, the categorical variable (to predict) coming from the clusterwise algorithm. For coherence with the properties of the proposed clusterwise multiblock analyses [i.e., Property 1 (distribution-free), Properties 5–6 ($J >> N_g$), and Property 7 (multicollinearity)], we decided to use the non-parametric $K$-nearest neighbors method (as implemented in the `kknn` `R` package).

### *2.4.5 In practice*

*Variable interpretation* When clusterwise multiblock methods are applied to datasets with a large number $J$ or $J^k$ of explanatory variables, the regression coefficients are often uninformative because they often provide information that is too vague for investigating the relationships between variables. In these cases, it is more informative to sort the explanatory variables by order of importance with regard to the whole $\mathbf{Y}$ prediction. To do so, multiblock tools—such as variable importance, denoted *VarImp* (i.e., the explanatory variable importance for the dependent block)—are used (Bougeard et al. 2011b). In the clusterwise framework, the vector (denoted **VarImp**) of the *VarImp* indices is computed for a cluster $\mathscr{C}_g$ and dimension $h$ as

$$\mathbf{VarImp}^{(h)} = \frac{\left(\mathbf{a}^{(h)} \odot \mathbf{a}^{(h)}\right) \odot \left(\mathbf{w}^{(h)*} \odot \mathbf{w}^{(h)*}\right)}{\left\|\left(\mathbf{a}^{(h)} \odot \mathbf{a}^{(h)}\right) \odot \left(\mathbf{w}^{(h)*} \odot \mathbf{w}^{(h)*}\right)\right\|} \quad \text{for} \quad h = 1, \ldots, O \tag{6}$$

where $\odot$ denotes the Hadamard element-wise product and the vector $\mathbf{a}$ of length $J$ in which each scalar $(a^1, \ldots, a^K)$ is respectively repeated $(J^1, \ldots, J^K)$ times. For the case of a model with $O$ dimensions to be taken into account, this index is weighted according to the relative importance of each dimension (i.e., the eigenvalue $\lambda^{(h)}$) as

$$\mathbf{VarImp}^{(O)} = \frac{\sum_{h=1}^{O} \lambda^{(h)} \mathbf{VarImp}^{(h)}}{\sum_{h=1}^{O} \lambda^{(h)}} . \tag{7}$$

*Block interpretation* In the same vein as the Variable Importance, the Block Importance index (denoted *BlockImp*) represents the explanatory block importance for the dependent variables (Bougeard et al. 2011b). In the clusterwise framework, the $BlockImp$ index can be processed within each cluster for a given dimension $h$ such as

$$BlockImp^{k(h)} = (a^{k(h)})^2 \quad \text{for} \quad k = 1, \ldots, K \quad \text{and} \quad h = 1, \ldots, O\ . \tag{8}$$

For the case of a model with several dimensions $O$, this index is weighted according to the relative importance of each dimension (i.e., the eigenvalue $\lambda^{(h)}$ of the matrix $\mathbf{Y}^\mathsf{T}\mathbf{X}\mathbf{X}^\mathsf{T}\mathbf{Y}$ for MBPLS and (25) for MBRA) as

$$BlockImp^{k(O)} = \frac{\sum_{h=1}^{O} \lambda^{(h)} BlockImp^{k(h)}}{\sum_{h=1}^{O} \lambda^{(h)}} \quad \text{for} \quad k = 1, \ldots, K \, . \tag{9}$$

*Algorithm* Because each iteration of the algorithm decreases the criterion that is to be minimized (see Eq. (4) for CW.MBPLS and Eq. (5) for CW.MBRA), the algorithm is guaranteed to converge, but it is, however, not guaranteed to converge to a global minimum. To avoid local optima, several initializations are used and the iteration that minimizes the overall error is kept. According to the literature (Spath 1979; DeSarbo and Cron 1988), and from our experience, using twenty random starts should suffice to give stable results. As detailed in Sect. 2.4.3, the number of steps of the algorithm is equal to the number of observations and from our experience, this number is adequate to achieve convergence. Note that the computational complexity of the algorithm depends on the number of initializations, of observations $N$, of clusters $G$, and of the dimension $H$ because these parameters directly affect the number of computations. Examples of time processing are given in Appendix 3. In practice, there is no restriction imposed on the proposed clusterwise multiblock algorithm in terms of data structure, block structure and cluster structure. The results described in Appendix 3, show that the processing time increases sharply when a large number of explanatory variables is combined with a large number of dimensions.

*Package availability* The clusterwise multiblock methods described in this article are implemented in the `R` (Team 2015) package `mbclusterwise` available from `CRAN`. In its current implementation, `mbclusterwise` uses parallel computing to simultaneously compute the random initializations but processing speed still needs to be improved for large numbers of observations and variables.

## 3 Application

In this section we: (1) assess the performance of clusterwise multiblock methods with a simulation study, and (2) illustrate the proposed clusterwise multiblock methods with a (simulated) example from marketing and consumer satisfaction.

### 3.1 Simulation study

#### *3.1.1 Problem*

A simulation study was conducted to investigate the performance of the two methods under study (CW.MBPLS, CW.MBRA). Thereby we evaluated how the two proposed methods are influenced by ten factors (selected for their theoretical and practical importance): (1) the number of dependent variables, (2) the number of explanatory variables, (3) the number of observations, (4) the number of blocks, (5) the proportion of variables per block, (6) the block weighting scheme when the blocks of variables

have different sizes, (7) the within-block correlation, (8) the number of clusters, (9) the proportion of observations per cluster, and (10) the separation between clusters. The performance of the methods was evaluated with: (1) the Root Mean Square Error of prediction evaluated from a ten-fold cross-validation procedure (compared to the error for the single-cluster case), (2) the adjusted Rand index (Hubert and Arabie 1985), (3) the model complexity (i.e., the number of components included in the model used as a measure of parsimony rather than an estimation of the unknown true dimensionality), (4) the regression coefficients, and (5) the cluster sizes.

### *3.1.2 Design*

The clusterwise method performances were evaluated by 21 case studies described in Table 2.

Explanatory variables were generated from normal mixture models whose parameters (proportions, means and covariances) are $(p_1, \ldots, p_G, \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_G, \boldsymbol{\Sigma_1}, \ldots, \boldsymbol{\Sigma_G})$. The proportion of observations $(p_1, \ldots, p_G)$ in each cluster was chosen as indicated by the vector $Cluster.size$ given in Table 2. The centres of the clusters were either *well-* or *mild*-separated as specified by the $Cluster.sep$ parameter: in case of *well*-separation, each variable in Cluster $\mathscr{C}_1$ comes from a normal distribution with zero mean ($\boldsymbol{\mu}_1 = \mathbf{0}$, a vector of size $J$) and each variable in Cluster $\mathscr{C}_2$ comes from a normal distribution with mean equal to 4 ($\boldsymbol{\mu}_2 = \mathbf{4}$, a vector of size $J$), whereas in case of mild-separation, each variable in Cluster $\mathscr{C}_1$ is simulated as before ($\boldsymbol{\mu}_1 = \mathbf{0}$) and each variable in Cluster $\mathscr{C}_2$ comes from a normal distribution with mean equal to 2 ($\boldsymbol{\mu}_2 = \mathbf{2}$). The covariance matrices of order $J$ are chosen to be identical for all clusters $\mathscr{C}_g$ ($\boldsymbol{\Sigma}_g = \boldsymbol{\Sigma}$). It was defined from the block structure and built in such a way that the $J$ explanatory variables were correlated within their block ($Within.cor$ = .9 or .7) and mildly correlated with the other block variables ($Between.cor$ = .5). Explanatory variables were positively correlated with the dependent variables for the first cluster and negatively correlated with the last cluster (i.e., $\mathbf{B}_1 = 1$ and $\mathbf{B}_2 = -1$ for two clusters; $\mathbf{B}_1 = 1$, $\mathbf{B}_2 = 0$ and $\mathbf{B}_3 = -1$ for three clusters). Dependent variables were computed from the explanatory variables (using the regression coefficients of the model) and weakly perturbed with the addition of a $\mathcal{N}(0, 0.05)$ error component applied to the cluster regression coefficients along with the addition of a random normal error (i.e., $\mathcal{N}(0, 0.1)$) residual.

Clusterwise methods were compared in 21 situations depending on the general structure and the block structure (case studies 1–7) as well as the cluster structure (case studies *a–c*), all details being given in Table 2. Several number of components to be included in the model were selected (following the $Dim$ factor values reported in Table 2); the one that minimized the Root Mean Square Error of prediction was kept (as the $Dim.opt$ values reported in Table 3). For each of the 21 case studies, 20 datasets were simulated.

### *3.1.3 Results*

The results of the simulations are given in Tables 3 and 4.

**Table 2** Description of the 21 case studies evaluated by means of a simulation study with $Q$ denoting the number of dependent variables, $J$ the number of explanatory variables, $N$ the total number of observations, $K$ the number of blocks, $Block.size$ the proportion of variables within blocks, $Block.weight$ the block weighting scheme in case of different block sizes (according to the block inertia; see Sect. 2.1), $Within.cor$ the within-block correlation, $Dim$ the number of components included in the model (the different number of components are computed and the set of values that minimizes the RMSE of prediction is selected), $G$ the number of clusters, $Cluster.size$ the proportion of observations per cluster, and $Cluster.sep$ the cluster separation

| Case | | General structure | | | Block structure | | | | | Cluster structure | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Study | | $Q$ | $J$ | $N$ | $K$ | $Blocksize$ | $Block.weight$ | $Within.cor$ | $Dim$ | $G$ | $Cluster.size$ | $Cluster.sep$ |
| Case 1 | *a* | 1 | 6 | 30 | 2 | [.50; .50] | – | 0.9 | (1, 2, 3) | 2 | [.50; .50] | Well |
| | *b* | 1 | 6 | 30 | 2 | [.50; .50] | – | 0.9 | (1, 2, 3) | 2 | [.70; .30] | Well |
| | *c* | 1 | 6 | 30 | 2 | [.50; .50] | – | 0.9 | (1, 2, 3) | 2 | [.50; .50] | Mild |
| Case 2 | *a* | 2 | 20 | 30 | 2 | [.50; .50] | – | 0.9 | (1, 2, 4) | 2 | [.50; .50] | Well |
| | *b* | 2 | 20 | 30 | 2 | [.50; .50] | – | 0.9 | (1, 2, 4) | 2 | [.70; .30] | Well |
| | *c* | 2 | 20 | 30 | 2 | [.50; .50] | – | 0.9 | (1, 2, 4) | 2 | [.50; .50] | Mild |
| Case 3 | *a* | 2 | 30 | 100 | 2 | [.50; .50] | – | 0.9 | (2, 4, 6) | 3 | [.33; .33; .33] | Well |
| | *b* | 2 | 30 | 100 | 2 | [.50; .50] | – | 0.9 | (2, 4, 6) | 3 | [.70; .15; .15] | Well |
| | *c* | 2 | 30 | 100 | 2 | [.50; .50] | – | 0.9 | (2, 4, 6) | 3 | [.33; .33; .33] | Mild |
| Case 4 | *a* | 2 | 21 | 30 | 3 | [.33; .33; .33] | – | 0.9 | (1, 2, 4) | 2 | [.50; .50] | Well |
| | *b* | 2 | 21 | 30 | 3 | [.33; .33; .33] | – | 0.9 | (1, 2, 4) | 2 | [.70; .30] | Well |
| | *c* | 2 | 21 | 30 | 3 | [.33; .33; .33] | – | 0.9 | (1, 2, 4) | 2 | [.50; .50] | Mild |
| Case 5 | *a* | 2 | 21 | 30 | 3 | [.33; .33; .33] | – | 0.7 | (1, 2, 4) | 2 | [.50; .50] | Well |
| | *b* | 2 | 21 | 30 | 3 | [.33; .33; .33] | – | 0.7 | (1, 2, 4) | 2 | [.70; .30] | Well |
| | *c* | 2 | 21 | 30 | 3 | [.33; .33; .33] | – | 0.7 | (1, 2, 4) | 2 | [.50; .50] | Mild |
| Case 6 | *a* | 2 | 20 | 30 | 3 | [.50; .35; .15] | No | 0.9 | (1, 2, 4) | 2 | [.50; .50] | Well |
| | *b* | 2 | 20 | 30 | 3 | [.50; .35; .15] | No | 0.9 | (1, 2, 4) | 2 | [.70; .30] | Well |
| | *c* | 2 | 20 | 30 | 3 | [.50; .35; .15] | No | 0.9 | (1, 2, 4) | 2 | [.50; .50] | Mild |
| Case 7 | *a* | 2 | 20 | 30 | 3 | [.50; .35; .15] | Yes | 0.9 | (1, 2, 4) | 2 | [.50; .50] | Well |
| | *b* | 2 | 20 | 30 | 3 | [.50; .35; .15] | Yes | 0.9 | (1, 2, 4) | 2 | [.70; .30] | Well |
| | *c* | 2 | 20 | 30 | 3 | [.50; .35; .15] | Yes | 0.9 | (1, 2, 4) | 2 | [.50; .50] | Mild |

**Table 3** Root Mean Square Error of prediction ($RMSE_p1$ for the single-cluster case ($G = 1$) and $RMSE_p$ if $G$ clusters are assumed), adjusted Rand index (AdjRd) and dimensionality of the clusterwise model (*Dim.opt*) associated with the 21 simulation cases (described in Table 2) for CW.MBPLS and CW.MBRA; mean values over the 20 simulated datasets

| Cases | Case *a* (*Well*-sep., equal size) | | | | Case *b* (*Well*-sep., diff. size) | | | | Case *c* (*Mild*-sep., equal size) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $RMSE_p1$ | $RMSE_p$ | AdjRd | *Dim.opt* | $RMSE_p1$ | $RMSE_p$ | AdjRd | *Dim.opt* | $RMSE_p1$ | $RMSE_p$ | AdjRd | *Dim.opt* |
| Case 1 | | | | | | | | | | | | |
| CW.MBRA | 0.83 | 0.19 | 0.82 | 1.1 | 1.17 | 0.30 | 0.75 | 1.3 | 1.32 | 0.81 | 0.74 | 1.4 |
| CW.MBPLS | 0.81 | 0.11 | 0.84 | 1.1 | 1.12 | 0.15 | 0.83 | 1.2 | 1.28 | 0.67 | 0.82 | 1.4 |
| Case 2 | | | | | | | | | | | | |
| CW.MBRA | 0.90 | 0.66 | 0.54 | 1 | 1.35 | 0.56 | 0.58 | 1.6 | 1.56 | 1.28 | 0.43 | 2.35 |
| CW.MBPLS | 0.73 | 0.14 | 0.85 | 1 | 1.11 | 0.10 | 0.89 | 1.3 | 1.29 | 0.60 | 0.91 | 1.60 |
| Case 3 | | | | | | | | | | | | |
| CW.MBRA | 1.87 | 0.34 | 0.77 | 2.3 | 2.15 | 0.65 | 0.68 | 2.9 | 2.25 | 1.54 | 0.48 | 3.6 |
| CW.MBPLS | 1.77 | 0.17 | 0.92 | 2.2 | 2.04 | 0.19 | 0.91 | 2.7 | 2.13 | 1.19 | 0.86 | 3.3 |
| Case 4 | | | | | | | | | | | | |
| CW.MBRA | 0.84 | 0.24 | 0.80 | 1.3 | 1.21 | 0.35 | 0.69 | 1.65 | 1.44 | 0.86 | 0.59 | 1.75 |
| CW.MBPLS | 0.80 | 0.09 | 0.84 | 1.0 | 1.11 | 0.15 | 0.79 | 1.20 | 1.30 | 0.54 | 0.83 | 1.65 |
| Case 5 | | | | | | | | | | | | |
| CW.MBRA | 0.91 | 0.23 | 0.78 | 1.75 | 1.12 | 0.31 | 0.69 | 1.6 | 1.43 | 0.85 | 0.57 | 2.05 |
| CW.MBPLS | 0.81 | 0.11 | 0.87 | 1.15 | 1.02 | 0.13 | 0.87 | 1.6 | 1.32 | 0.61 | 0.90 | 1.40 |
| Case 6 | | | | | | | | | | | | |
| CW.MBRA | 0.27 | 0.13 | 0.97 | 1.15 | 0.38 | 0.16 | 0.91 | 1.3 | 0.56 | 0.26 | 0.93 | 1 |
| CW.MBPLS | 0.21 | 0.12 | 0.60 | 1.00 | 0.32 | 0.15 | 0.57 | 1.0 | 0.48 | 0.24 | 0.55 | 1 |
| Case 7 | | | | | | | | | | | | |
| CW.MBRA | 0.28 | 0.14 | 0.99 | 1 | 0.37 | 0.17 | 0.87 | 1.15 | 0.56 | 0.30 | 0.90 | 1 |
| CW.MBPLS | 0.25 | 0.14 | 0.73 | 1 | 0.34 | 0.17 | 0.68 | 1.00 | 0.51 | 0.29 | 0.73 | 1 |

**Table 4** Regression coefficients and cluster sizes associated with the 21 simulation cases (described in Table 2) for CW.MBPLS and CW.MBRA; mean values over the 20 simulated datasets. Actual values are given in parentheses

| Cases | Case a (*Well*-sep., equal size) | | Case b (*Well*-sep., diff. size) | | Case c (*Mild*-sep., equal size) | |
|---|---|---|---|---|---|---|
| | Regr. coeff. | Cluster size | Regr. coeff. | Cluster size | Regr. coeff. | Cluster size |
| Case 1 | (1 / −1) | (.50 / .50) | (1 / −1) | (.70 / .30) | (1 / −1) | (.50 / .50) |
| CW.MBRA | 0.93 / −0.98 | .46 / .54 | 0.99 / −0.96 | .64 / .36 | 0.90 / −0.96 | .46 / .54 |
| CW.MBPLS | 0.98 / −0.98 | .47 / .53 | 0.98 / −0.99 | .66 / .34 | 0.92 / −1.00 | .47 / .53 |
| Case 2 | (1 / −1) | (.50 / .50) | (1 / −1) | (.70 / .30) | (1 / −1) | (.50 / .50) |
| CW.MBRA | 0.42 / −0.93 | .45 / .55 | 0.89 / −0.89 | .57 / .43 | 0.49 / −0.92 | .47 / .53 |
| CW.MBPLS | 0.95 / −0.98 | .48 / .52 | 0.99 / −0.98 | .66 / .34 | 0.96 / −0.99 | .48 / .52 |
| Case 3 | (1 / −1 / 0) | (.33 / .33 / .33) | (1 / −1 / 0) | (.70 / .15 / .15) | (1 / −1 / 0) | (.33 / .33 / .33) |
| CW.MBRA | 0.92 / −0.92 / −0.05 | .26 / .37 / .36 | 0.91 / −1.10 / −0.10 | .59 / .19 / .22 | 0.41 / −1.06 / −0.27 | .34 / .35 / .31 |
| CW.MBPLS | 0.99/ −1.00 / 0.00 | .31 / .36 / .34 | 0.97 / −0.99 / 0.02 | .67 / .17 / .16 | 0.95 /−1.05 / −0.05 | .32 / .35 / .32 |
| Case 4 | (1 / −1) | (.50 / .50) | (1 / −1) | (.70 / .30) | (1 / −1) | (.50 / .50) |
| CW.MBRA | 0.95 / −0.91 | .44 / .56 | 0.91 / −0.89 | .62 / .38 | 0.66 / −0.86 | .46 / .54 |
| CW.MBPLS | 0.96 / −0.99 | .45 / .55 | 0.97 / −0.96 | .64 / .36 | 0.96 / −0.99 | .46 / .55 |
| Case 5 | (1 / −1) | (.50 / .50) | (1 / −1) | (.70 / .30) | (1 / −1) | (.50 / .50) |
| CW.MBRA | 0.85 / −0.98 | .44 / .56 | 0.92 / −0.93 | .62 / .38 | 0.80 / −0.90 | .44 / .56 |
| CW.MBPLS | 0.95 / −0.97 | .46 / .54 | 0.98 / −0.97 | .67 / .33 | 0.96 / −0.96 | .51 / .49 |
| Case 6 | (1 / −1) | (.50 / .50) | (1 / −1) | (.70 / .30) | (1 / −1) | (.50 / .50) |
| CW.MBRA | 0.48 / −0.57 | .48 / .52 | 0.55 / −0.49 | .66 / .34 | 0.5 / −0.53 | .51 / .49 |
| CW.MBPLS | 0.30/−0.99 | .37 / .63 | 0.43 / −0.95 | .57 / .43 | 0.3 / −0.91 | .42 / .57 |
| Case 7 | (1 / −1) | (.50 / .50) | (1 / −1) | (.70 / .30) | (1 / −1) | (.50 / .50) |
| CW.MBRA | 0.5 / −0.51 | .49 / .51 | 0.55 / −0.55 | .70 / .30 | 0.47 / −0.61 | .49 / .51 |
| CW.MBPLS | 0.3 / −0.83 | .42 / .58 | 0.37 / −0.85 | .64 / .36 | 0.30 / −0.77 | .45 / .55 |

The first performance criterion is the Root Mean Square Error of prediction as evaluated with a ten-fold cross-validation procedure (Table 3). Not surprisingly, clusterwise multiblock analyses always improve the Root Mean Square Error of prediction while taking into account the cluster structure of observations. This effect is particularly clear for the simplest simulation cases (cases 1–5 combined with *a* and *b*) with an average ratio between the two errors (i.e., $\text{RMSE}_p 1$ / $\text{RMSE}_p$) of 5.96. This effect is also clear for the more complex cases (6, 7, and *c*) with an average ratio of 1.92. Then, MBRA and MBPLS have comparable prediction performances for the single-cluster cases with an average ratio (i.e., $\text{RMSE}_{p(\text{MBPLS})}$ / $\text{RMSE}_{p(\text{MBRA})}$) of 0.90 slightly in favor of MBPLS. To a lesser extent, this effect can also be seen for the more complex cases (cases 6, 7, and *c*) with an average ratio of 0.83. Note that CW.MBPLS performs far better than CW.MBRA for the simplest cases (cases 1–5 combined with *a* and *b*) with an average ratio of 0.40. In such cases, CW.MBRA suffers from its strong block constraints whereas it correctly performs for the cases of different block sizes (cases 6 and 7).

The second performance criterion—i.e., the adjusted Rand index—measures how well the actual clusters are recovered (see Table 3). Values of the Rand index close to 1 indicate that the clusterwise multiblock analyses succeeded in recovering the actual clusters, and values close to 0 indicate that the methods failed to recover the actual clusters. Not surprisingly, best performance was achieved for the case *a* (*well*-separated clusters of equal sizes) with an average adjusted Rand of .81 and slightly decreased for the case *b* (*well*-separated clusters of different sizes) with an average adjusted Rand of .77 and for the case *c* (mild-separated clusters of equal sizes) with an average Adjusted Rand of .73. Clusterwise analyses always performed to recover *well*-separated clusters of equal sizes. They also performed to recover *well*-separated clusters of different cluster sizes. But they may encounter difficulties in finding mild-separated clusters. The method CW.MBPLS performed better than CW.MBRA for the cases 1–5 with an average ratio (i.e., $\text{AdjRd}_{(\text{MBPLS})}$ / $\text{AdjRd}_{(\text{MBRA})}$) of 1.36 whereas the opposite was true for the cases 6 and 7 with an average ratio of 0.69. Whatever the cluster structure, CW.MBRA performed better when the blocks have different sizes in recovering the actual clusters although its prediction error was slightly lower than the prediction error of CW.MBPLS— an effect likely due to difficulties in recovering the actual regression coefficients.

The third performance criterion was the model complexity (used here as a measure of parsimony), as measured by the number of components included in the model (Table 3). A higher average number of components was needed for the cases *b* (1.54) and *c* (1.75) than for the simpler case *a* (1.29). The method CW.MBRA always needed a higher number of components than CW.MBPLS with an average ratio (i.e., $\text{Dim}_{(\text{MBPLS})}$ / $\text{Dim}_{(\text{MBRA})}$) of 0.88. The number of components included in the model was, in general, around 1 but higher-order dimensions were needed for higher numbers of clusters (case 3) and to a lesser extent in the case of lower within-block correlations (case 5).

The last two performance criteria evaluated how well the actual regression coefficients and the cluster sizes were recovered (Table 4). The regression coefficients were, most of the time, correctly recovered (cases 1–5) but differed slightly from the actual values when the block sizes differed (cases 6 and 7) and—but to a lesser extent—

when block sizes varied (case *c*); these effect being stronger for CW.MBRA. Cluster sizes were correctly recovered by clusterwise multiblock methods even when sizes were clearly unbalanced (case *b*).

To conclude, clusterwise multiblock analyses always improved the prediction when the observations displayed a cluster structure. The overall performance, except in terms of model complexity, was not really affected by the number of dependent and explanatory variables, the number of observations, the block size, the block weighting scheme, the within-block correlation, the number of clusters, or the proportion of observations per cluster. However, performance was affected by the cluster separation and by the proportion of variables within blocks. For the latter case, CW.MBRA better recovered the clusters but had some difficulties to find the actual regression coefficient values. For the simplest cases, CW.MBPLS performed better than CW.MBRA.

### 3.2 Application to consumer satisfaction data

#### *3.2.1 Data and aim*

We illustrate clusterwise multiblock analysis with a (simulated) example from marketing and consumer satisfaction. This data set is available from the `plspmR` package (`simdata`), and is discussed and described in more details in Trinchera (2007) and Vinzi et al. (2007). These data were generated to mimic the structure of a survey related to consumer satisfaction when consumers are clustered in two clearly different groups of equivalent size. Specifically, the *price fairness* cluster of consumers is characterized by a strong relationship between product price and satisfaction and a weak relationship between product quality and satisfaction whereas the opposite is true for the *quality oriented* cluster of consumers.

In this data set, consumer satisfaction as well as price fairness and product quality are given for ($N = 400$) consumers. The satisfaction construct to be explained ($\mathbf{Y}$) is described by ($Q = 3$) quantitative variables. The ($J = 10$) explanatory variables (stored in $\mathbf{X}$) are organized in ($K = 2$) thematic blocks related to the price fairness ($\mathbf{X}^1$ dataset, 5 variables) and to the product quality ($\mathbf{X}^2$ dataset, 5 variables). The three dependent variables are highly correlated with each other ($\approx .99$ on average); the explanatory variables are correlated within a block ($\approx .70$ on average), but are not correlated between blocks ($\approx .03$ on average); dependent and explanatory variables are mildly correlated ($\approx .30$ on average). Note that, these data simulated highly correlated dependent variables, a case which is sometimes albeit rarely encountered in practice. Such a pattern corresponds to explaining a single variable and affects the clusterwise component-based analysis performance (i.e., lower prediction error than for the case of several non-correlated dependent variables) but does not affect the capacity of the method to explain several variables. Our aim in this example was to identify clusters of consumers that correspond to different predictive models.

### *3.2.2 Parameter selection*

A ten-fold cross-validation procedure was applied to identify two main parameters of interest: (1) the number of clusters (i.e., $G$), and (2) the resulting optimal dimension for the model (i.e., the optimum $O$). These parameters were estimated by minimizing the error of prediction which expresses the predictive power of the model. The Root Mean Square Error (RMSE) values for prediction are shown in Figure 2 for CW.MBRA and CW.MBPLS.

The difference between the values of the RMSE of prediction for ($G = 1$) and ($G > 1$) reflects the improvement of the prediction achieved by clusterwise multiblock models compared to the non-clusterwise models. For example, considering CW.MBRA, the average $\text{RMSE}_p$ moves from 1.18 to .92 when the model moves from one ($G = 1$) to two ($G = 2$) clusters for the observations. The best solution is achieved for a two cluster model (which matches the generative model). The minimum values of the prediction errors from CW.MBRA and CW.MBPLS are comparable and slightly in favour of CW.MBPLS: $\text{RMSE}_p = .901$ for CW.MBRA ($H = 5$ components, $G = 2$ clusters) and $\text{RMSE}_p = .888$ for CW.MBPLS ($H = 1$ component, $G = 2$ clusters). However, as explained in Appendix 1, MBPLS, for the case of a single dependent block $\mathbf{Y}$, is not a true multiblock method because its results do not depend upon the partition of the explanatory variables in blocks. As the explanatory variables have a strong multiblock structure (see 3.2.1), we choose to apply CW.MBRA to analyze the differences between consumers in how satisfaction is related to the variables measuring satisfaction.

As for CW.MBRA, the best solution (with two clusters) is achieved by using ($H = 5$) components ($\text{RMSE}_p = .901$). However using ($H = 5$) components rather than ($H = 1$) slightly improved the prediction error ($\text{RMSE}_p = .901$ for 5 components and $\text{RMSE}_p = .9222$ for 1 component). In practice, according to the parsimony principle and since the regression coefficients are similar, a model with a single component is chosen to be interpreted. This solution assigns 204 consumers to Cluster $\mathscr{C}_1$ and 196 to Cluster $\mathscr{C}_2$. The actual clusters being known, the adjusted Rand index can be computed as well as the percentage of correct classification which are equal respectively to .44 and 83% (confusion matrix in Table 5). Despite the good percentage of correct classification, the adjusted-Rand index may seem disappointing but this index takes into account its expected value (under the generalized hypergeometric distribution assumption for randomness).

### *3.2.3 Cluster interpretation*

As previously stated, one component was needed for best prediction. Not surprisingly in the Redundancy Analysis framework, because the dependent variables are highly correlated, the inertia of the dependent variables (i.e., matrix $\mathbf{Y}$) is mainly explained by the first component for both clusters (99.0% for Cluster $\mathscr{C}_1$ and 99.2% for Cluster $\mathscr{C}_2$). The inertia of the explanatory variables (i.e., matrix $\mathbf{X}$) explained by the first component are comparable: 37.5% for Cluster $\mathscr{C}_1$ and 33.7% for Cluster $\mathscr{C}_2$. But block inertia clearly depends on clusters: the inertia of $\mathbf{X}^1$ (Price block) explained by the first component is equal to 68.9% for Cluster $\mathscr{C}_1$ and 3.1% for Cluster $\mathscr{C}_2$ whereas

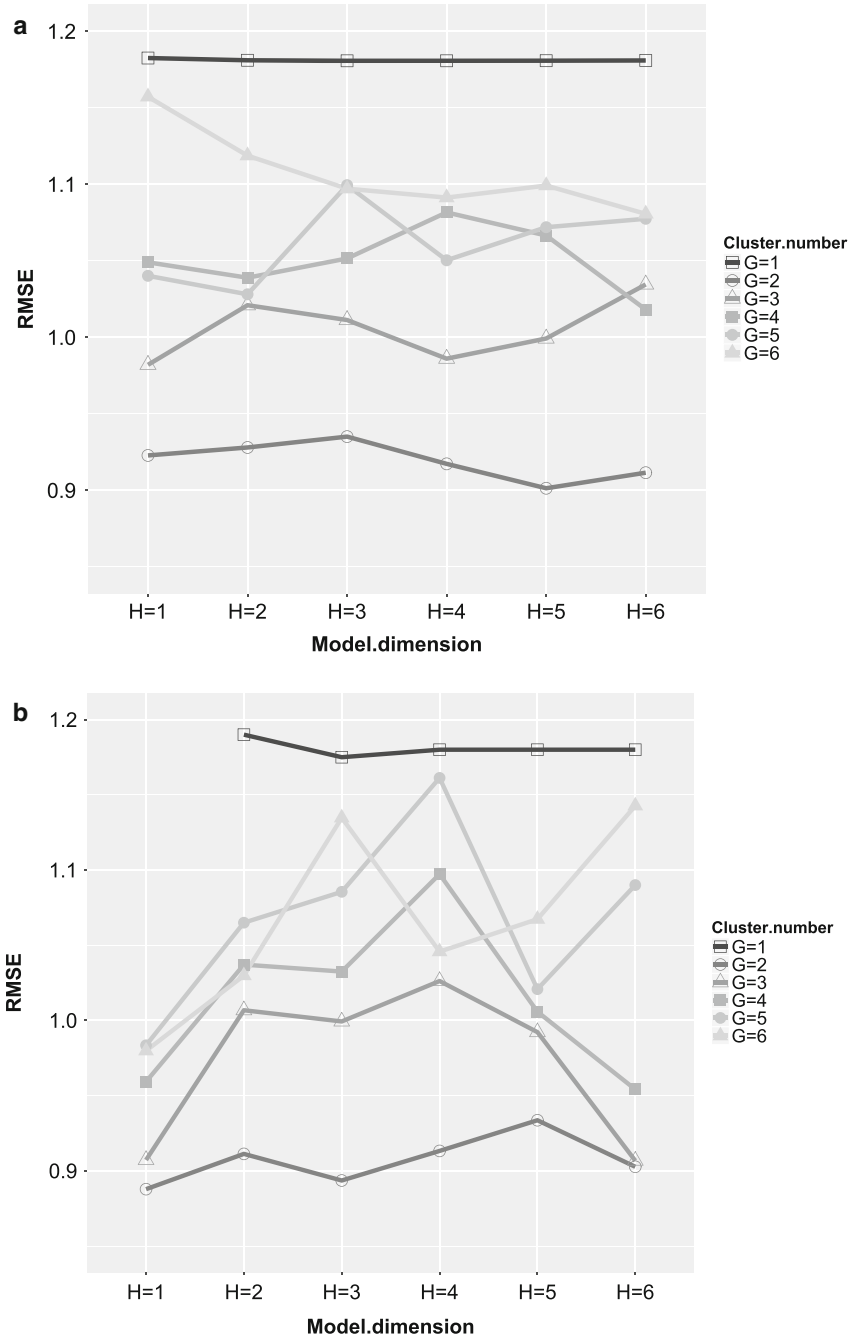


**Fig. 2** Consumer satisfaction data: Mean value of the ten-fold Root Mean Square Error for prediction ($RMSE_p$) for CW.MBRA (Fig. 2a) and CW.MBPLS (Fig. 2b) as a function of the number of clusters ($G = 1, \ldots, 6$) and the number of components in the model ($H = 1, \ldots, 6$). Whatever the method and the model dimension, the solution with two clusters is optimal and therefore the model always finds the correct number of clusters

**Table 5** Confusion matrix of the method CW.MBRA for the model with a single component

| | CW.MBRA Cluster $\mathscr{C}_1$ | CW.MBRA Cluster $\mathscr{C}_2$ | Sum |
|---|---|---|---|
| Actual cluster $\mathscr{C}_1$ | 168 | 32 | 200 |
| Actual cluster $\mathscr{C}_2$ | 36 | 164 | 200 |
| Sum | 204 | 196 | |

the inertia of $\mathbf{X}^2$ (Quality block) is equal to 1.94% for Cluster $\mathscr{C}_1$ and 72.4% for Cluster $\mathscr{C}_2$.

To evaluate the stability of the main parameters, we used the `randboot` function (from the `R` package `ade4`) to generate 500 bootstrap samples that were then used to get standard deviations and confidence intervals for the regression coefficients and the variable importances for each cluster. Results of CW.MBRA are given in Table 6.

Table 6 shows that each cluster has a specific regression model: for Cluster $\mathscr{C}_1$ ($N = 204$) the five measures of consumer satisfaction are associated with price ($Price.1$ to $Price.5$) but also with one element of quality ($Quality.3$), this cluster corresponds to price oriented customers. By contrast, Cluster $\mathscr{C}_2$ ($N = 196$) is associated with quality ($Quality.1$ to $Quality.5$). Note that in this example, the dependent variables are highly correlated and therefore, within a block, the regression coefficients have very similar values.

Finally, the block importance indices also facilitate the interpretation because they quantify the contributions of the explanatory blocks ($\mathbf{X}^1, \mathbf{X}^2$) to the overall consumer satisfaction ($\mathbf{Y}$). Results of CW.MBRA are given in Table 7.

It turns out that the influence of the explanatory blocks on consumer satisfaction differs depending upon the clusters. The 'price' block ($\mathbf{X}^1$) has a much larger importance for Cluster $\mathscr{C}_1$ ($BlockImp_1 = 92.8\%$) than for Cluster $\mathscr{C}_2$. Conversely, the 'quality' block ($\mathbf{X}^2$) mainly influences Cluster $\mathscr{C}_2$ ($BlockImp_2 = 96.2\%$). This means that these blocks and these clusters should be given specific attention in order to increase consumer satisfaction.

## 4 Discussion

### 4.1 Multiblock component methods

Even though MBRA and therefore CW.MBRA have real multiblock constraints in comparison to MBPLS and CW.MBPLS, it is, however, not obvious—as seen from the simulation study—that the prediction performance of CW.MBRA is poorer than CW.MBPLS. From our experience, we recommend to use CW.MBRA rather than CW.MBPLS when the data have a multiblock structure to be taken into account and when the explanatory blocks are not ill-conditioned. As a matter of fact, in our simulation study, CW.MBPLS performed better than CW.MBRA but CW.MBRA better recovered the clusters when the variable proportions differed in their blocks. In our example, the results for CW.MBRA and CW.MBPLS are similar with regard to the number of clusters, their sizes, their assignation (79% of observations are in the same

**Table 6** Consumer satisfaction data: multiblock regression coefficients, their standard deviation (from 500 bootstrap simulations) and their variable importance (**VarImp**) for CW.MBRA for the ($G = 2$) clusters for models involving ($O = 1$) component

| | | Cluster $\mathscr{C}_1$ ($N = 204$) | | | | Cluster $\mathscr{C}_2$ ($N = 196$) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Variable | *Satisf*.1 | *Satisf*.2 | *Satisf*.3 | **VarImp** | *Satisf*.1 | *Satisf*.2 | *Satisf*.3 | **VarImp** |
| | Intercept | 1.14 | 1.17 | 1.17 | | 0.67 | 0.65 | 0.74 () | |
| $\mathbf{X}^1$ | *Price*.1 | 0.11 (0.03)* | 0.11 (0.03)* | 0.11 (0.03)* | 6.8% | 0.10 (0.08) | 0.10 (0.08) | 0.10 (0.08) | 0.2% |
| | *Price*.2 | 0.14 (0.04)* | 0.14 (0.04)* | 0.14 (0.04)* | 12.6% | 0.08 (0.10) | 0.08 (0.10) | 0.07 (0.10) | 0.1% |
| | *Price*.3 | 0.10 (0.04)* | 0.10 (0.04)* | 0.10 (0.04)* | 6.5% | −0.13 (0.12) | −0.13 (0.12) | −0.12 (0.12) | 0.3% |
| | *Price*.4 | 0.17 (0.04)* | 0.17 (0.04)* | 0.17 (0.04)* | 17.7% | 0.06 (0.13) | 0.06 (0.13) | 0.06 (0.13) | 0.1% |
| | *Price*.5 | **0.29** (0.05)* | **0.28** (0.05)* | **0.28** (0.05)* | 49.1% | −0.04 (0.14) | −0.04 (0.14) | −0.03 (0.14) | 0.0% |
| $\mathbf{X}^2$ | *Quality*.1 | −0.03 (0.09) | −0.03 (0.09) | −0.03 (0.09) | 0.0% | 0.19 (0.04)* | 0.19 (0.04)* | 0.19 (0.04)* | 19.7% |
| | *Quality*.2 | −0.07 (0.11) | −0.07 (0.11) | −0.07 (0.11) | 0.2% | **0.28** (0.05)* | **0.28** (0.05)* | **0.28** (0.05)* | 42.1% |
| | *Quality*.3 | **0.37** (0.08)* | **0.36** (0.09)* | **0.36** (0.09)* | 6.3% | −0.15 (0.06)* | −0.15 (0.06)* | −0.15 (0.06)* | 12.7% |
| | *Quality*.4 | −0.13 (0.13) | −0.13 (0.13) | −0.13 (0.13) | 0.8% | 0.17 (0.04)* | 0.17 (0.04)* | 0.17 (0.04)* | 15.6% |
| | *Quality*.5 | 0.01 (0.09) | 0.01 (0.09) | 0.01 (0.09) | 0.0% | 0.13 (0.04)* | 0.13 (0.04)* | 0.13 (0.04)* | 9.1% |

A star indicates that the coefficient has a significant impact on the dependent variable (bold coefficients indicate large coefficients with magnitude $> 0.20$)

**Table 7** Consumer satisfaction data: block importance ($BlockImp$) of CW.MBRA for the ($G = 2$) clusters for models involving ($O = 1$) component

| | Cluster $\mathscr{C}_1$ ($N = 204$) | Cluster $\mathscr{C}_2$ ($N = 196$) |
|---|---|---|
| $\mathbf{X}^1$: price | **92.8%** | 3.8% |
| $\mathbf{X}^2$: quality | 7.3% | **96.2%** |

cluster in terms of interpretation) and their main conclusions (i.e., satisfaction is either related to price or to quality depending on the cluster).

### 4.2 Prediction

In the clusterwise framework, the prediction of new observations has only received scan attention although it is an important feature for clusterwise multiblock methods. In fact, in some cases—such as clusterwise methods based on finite mixture model—it is *de facto* a "non-issue" because likelihood cannot be determined without knowing the true value of $\mathbf{y}$. For techniques amenable to prediction, though, the prediction procedure proposed here also can be used and adapted. We choose the $K$-nearest neighbor technique to assign (new) observations to a given cluster. However, depending on data features, several other supervised classification methods could also be applied.

### 4.3 Cluster numbers

From our experience, the selection of the optimal cluster number based on the prediction error works well when clusters are well-separated. For more subtle cluster differences, a relevant penalized prediction error could also be proposed.

## 5 Conclusion and perspectives

We proposed two new component-based methods that extend clusterwise regression to multiblock data and to model the relationships between one block of dependent variables and several blocks of explanatory variables, when these relations vary across clusters of observations. The proposed methods simultaneously provide local regression models and identify associated clusters with underlying dimensions on which these clusters are based. In addition, the minimization of a well defined criterion by means of a sequential algorithm guarantees convergence. The number of components to be included in the model and the number of clusters are defined so as to minimize the prediction error on the basis of a ten-fold cross-validation. Finally, the proposed methods are distribution-free. The application shows that the proposed clusterwise multiblock regression methods are useful tools that can be used to analyze complex data as found, for example. in marketing, biology, or any field dealing with population mixtures. The clusterwise procedures proposed are oriented both towards modeling and prediction and can be applied to any other multiblock component

methods such as PLS-PM, GSCA, RGCCA, or THEME provided that regression models are given. Specifically, PLS-PM, GSCA, and RGCCA are relevant methods for explanatory issues but present some limitations for prediction. One of their limitation is to fail to take into account more than one component in a model, because modeling is processed dimension by dimension without component selection rules. Then, relationships between block components in PLS-PM and RGCCA are symmetrical as they are based on global correlation or covariance measures and highlight what is common among the blocks (for details, refer to Dolce et al. 2016); by contrast, THEME is based on partial correlation measures which are the basis of regression (Bry et al. 2012). Therefore, THEME being the method most oriented toward prediction may be the most suitable to be extended to the clusterwise purpose.

However, the proposed clusterwise approaches present some limitations and further investigations needs to be undertaken to handle some specific data. The method parameters—namely the number of clusters and the number of components—are assumed to be the same for all blocks of explanatory variables. For future work, one can extend the proposed methods to the case of different numbers of clusters of observations as well as to components varying across blocks. This cannot be achieved through the current multiblock regression in which $\mathbf{Y}$ is explained using all the explanatory blocks at the same time. However, this can be obtained through block regressions where $\mathbf{Y}$ is alternatively explained by each explanatory block, these regressions being then associated with block deflations to ensure block component orthogonality. For the case of a larger number of observations and in a smaller number of variables, the proposed algorithm should be accelerated. In particular, the search of the unknown parameters $G$ and $H$ is based on $(G \times H)$ ten-fold cross-validations and is consequently quite time consuming. For the case of a large number of observations, this time could be reduced by splitting the data into three subsets: the first subset would be used to estimate the regression coefficients with $(1, \ldots, H)$ components, the second subset would be used to select the optimal dimension $H$, and the last subset would be used to compute the prediction error to select the optimal cluster number $G$ (refer to Martella et al. 2015 in another context).

**Acknowledgements** The authors are grateful to two anonymous reviewers for their valuable suggestions that greatly improved the clarity and the relevance of this article.

## Appendix 1: multiblock PLS

In standard multiblock PLS for the case of a single dataset $\mathbf{Y}$ to explain, the relationship between $\mathbf{Y}$ and the $K$ matrices $\mathbf{X}^k$ (stored in $\mathbf{X}$) is first modeled by computing a pair of linear combinations $\mathbf{u}$ and $\mathbf{t}$—called *components*—of the columns of, respectively $\mathbf{Y}$ and $\mathbf{X}$ such that these components have maximal covariance (see, e.g., Qin et al. 2001; Abdi and Williams 2012). After this first step—equivalent to a standard PLS model (Qin et al. 2001)—specific components are computed to relate each $\mathbf{X}^k$ to $\mathbf{Y}$. Formally, MBPLS first implements the following optimization problem

$$\delta = \arg\max_{\mathbf{v},\mathbf{w}} (\text{cov}(\mathbf{u}, \mathbf{t})) \quad \text{with} \quad \mathbf{u} = \mathbf{Y}\mathbf{v}, \quad \text{and} \quad \mathbf{t} = \mathbf{X}\mathbf{w}$$
$$\text{under the constraints that}$$
$$\|\mathbf{w}\|^2 = \|\mathbf{v}\|^2 = 1. \tag{10}$$

The solution of this problem is obtained by taking $\mathbf{v}$ and $\mathbf{w}$ (called, respectively, the $\mathbf{Y}$- and $\mathbf{X}$-loadings) as (respectively) the first left and right singular vectors of matrix $\mathbf{Y}^\mathsf{T}\mathbf{X}$ (and the first singular value $\delta$ gives the thought after maximum of Expression (10)). In a second step, the dependent dataset $\mathbf{Y}$ is predicted (with a standard linear regression) from the component $\mathbf{t}$ as

$$\widehat{\mathbf{Y}} = \frac{\mathbf{t}\mathbf{t}^\mathsf{T}}{\|\mathbf{t}\|^2}\mathbf{Y} = \mathbf{t}\mathbf{c}^\mathsf{T} \quad \text{with} \quad \mathbf{c} = \frac{\mathbf{Y}^\mathsf{T}\mathbf{t}}{\|\mathbf{t}\|^2}\,. \tag{11}$$

The matrix $\widehat{\mathbf{Y}}$ therefore corresponds to the orthogonal projection of $\mathbf{Y}$ onto the component $\mathbf{t}$. In MBPLS, after the components and loadings have been found, block loadings (also called *partial* loadings) are computed (see, e.g., Qin et al. 2001, and Equations (3) and (4) for details) as

$$\mathbf{w}^k = \frac{(\mathbf{X}^k)^\mathsf{T}\mathbf{u}}{\left\|\mathbf{u}^\mathsf{T}\mathbf{X}^k\right\|}\,, \quad \mathbf{t}^k = \mathbf{X}^k\mathbf{w}^k \quad \text{and} \quad a^k = \frac{\mathbf{u}^\mathsf{T}\mathbf{t}^k}{\sqrt{\sum_{k=1}^{K}\left(\mathbf{u}^\mathsf{T}\mathbf{t}^k\right)^2}}. \tag{12}$$

This way of computing the block loading vectors ensures that the global component vector can be obtained as a weighted average of the block vectors, namely

$$\mathbf{t} = \sum_{k=1}^{K} a^k\mathbf{t}^k \quad \text{with} \quad \sum_{k=1}^{K}(a^k)^2 = 1 \quad \text{and} \quad \|\mathbf{w}^k\|^2 = 1\,. \tag{13}$$

The partial loadings $\mathbf{w}^k$ can be seen as normalized sub-vectors of $\mathbf{w}$, and this implies that MBPLS can naturally cope with multicollinearity in $\mathbf{X}_k$ or $\mathbf{Y}$ and will, therefore, provide stable solutions.

As our regression problem is to get a good prediction of $\mathbf{Y}$, this dataset is explained with all the variables in $(\mathbf{X}^1, \ldots, \mathbf{X}^K)$ (Westerhuis and Smilde 2001). As a consequence, the component-based regression is derived from the global component $\mathbf{t}$ rather than on block components $\mathbf{t}^k$. Thereafter, plugging Eq. (13) into Eq. (11) shows that matrix $\mathbf{Y}$ can also be predicted from the partial components as

$$\widehat{\mathbf{Y}} = \mathbf{t}\mathbf{c}^\mathsf{T} = \sum_{k=1}^{K} a^k\mathbf{t}^k\mathbf{c}^\mathsf{T}\,. \tag{14}$$

Note that when—as it is the case here—the matrix $\mathbf{Y}$ does not include blocks, the $\mathbf{X}^k$ block loadings are computed after the global loadings have been estimated, and so the block loadings do not depend upon the partition of the explanatory variables in blocks;

therefore MBPLS, for the case of a single dependent block $\mathbf{Y}$, is not a true multiblock method (Westerhuis et al. 1998; Qin et al. 2001; Vivien 2002).

Because one component rarely completely explain the dependent variables, higher order components are often needed. These higher order components are obtained by first removing from the raw data the previous-order solution (a procedure called "deflation") and then re-iterating the optimization procedure on the deflated data. Because this procedure ensures orthogonality of the components further used in the component-based regression, we choose to deflate the raw data from the global component $\mathbf{t}$ rather than from the block components $\mathbf{t}^k$. Also, as deflating $\mathbf{X}$ or $\mathbf{Y}$ leads to the same prediction (Westerhuis and Smilde 2001), we choose to regress out the effect of the first-order global component from $\mathbf{X}$. Formally, in our deflation step, $\mathbf{X}$ is replaced by $\mathbf{X}^{(2)}$ computed as

$$\mathbf{X}^{(2)} = \left(\mathbf{I} - \frac{\mathbf{t}\mathbf{t}^{\mathsf{T}}}{\|\mathbf{t}\|^2}\right)\mathbf{X}\,. \tag{15}$$

To improve the prediction, $\mathbf{X}$ is replaced in Eq. (10) by its residual defined in Eq. (15). The process can then be re-iterated to obtain subsequent components. We denote by $O$ the optimal number of components to keep in the model (with $O \leq J$)—$O$ is in general estimated by a cross-validation approach. This deflation step ensures that components (i.e., the vectors $\mathbf{t}$) obtained at different steps are orthogonal to each other. Therefore, the predicted dependent dataset can be written according to the global components or according to the block ones

$$\widehat{\mathbf{Y}}^{(O)} = \sum_{h=1}^{O} \mathbf{t}^{(h)}(\mathbf{c}^{(h)})^{\mathsf{T}} = \sum_{h=1}^{O}\sum_{k=1}^{K} a^{k(h)}\mathbf{t}^{k(h)}\left(\mathbf{c}^{(h)}\right)^{\mathsf{T}} \tag{16}$$

with

$$\mathbf{c}^{(h)} = \frac{\mathbf{Y}^{\mathsf{T}}\mathbf{t}^{(h)}}{\|\mathbf{t}^{(h)}\|^2} \tag{17}$$

being the vector of the regression coefficients of $\mathbf{Y}$ on $\mathbf{t}^{(h)}$. This last regression step corresponds to the following optimization problem

$$\arg\min_{\mathbf{c}} \left\|\mathbf{Y} - \sum_{h=1}^{O} \mathbf{t}^{(h)}(\mathbf{c}^{(h)})^{\mathsf{T}}\right\|^2 \quad \text{with} \quad \mathbf{t}^{(h)} = \mathbf{X}^{(h-1)}\mathbf{w}^{(h)} \tag{18}$$

where $\mathbf{X}^{(h-1)}$ is the residual of the prediction of $\mathbf{X}$ from the $h-1$ previous components $(\mathbf{t}^{(1)}, \ldots, \mathbf{t}^{(h-1)})$. Because these components are orthogonal, Expression

$$\mathbf{t}^{(h)} = \mathbf{X}^{(h-1)}\mathbf{w}^{(h)}$$

is equivalent to

$$\mathbf{t}^{(h)} = \mathbf{X}\mathbf{w}^{(h)*}$$

with $\mathbf{w}^{(h)*}$ defined as

$$\mathbf{w}^{(h)*} = \prod_{l=1}^{h-1} \left[ \mathbf{I} - \frac{\mathbf{w}^{(l)} (\mathbf{t}^{(l)})^{\mathsf{T}}}{\|\mathbf{t}^{(l)}\|^2} \right] \mathbf{w}^{(h)} \tag{19}$$

(for proofs see, e.g., Tenenhaus 1998; Wold et al. 1983).

If we define

$$\mathbf{W}^{(O)*} = \left[ \mathbf{w}^{(1)*}, \ldots, \mathbf{w}^{(h)*}, \ldots, \mathbf{w}^{(O)*} \right] \text{ and } \mathbf{C}^{(O)} = \left[ \mathbf{c}^{(1)}, \ldots, \mathbf{c}^{(h)}, \ldots, \mathbf{c}^{(O)} \right], \tag{20}$$

the optimal prediction of $\mathbf{Y}$, denoted $\widehat{\mathbf{Y}}^{(O)}$, can be obtained, in a way analogous to standard multiple linear regression, as

$$\widehat{\mathbf{Y}}^{(O)} = \mathbf{X}\mathbf{B}^{(O)} \quad \text{with} \quad \mathbf{B}^{(O)} = \mathbf{W}^{(O)*} (\mathbf{C}^{(O)})^{\mathsf{T}} . \tag{21}$$

Interestingly, rewriting Eq. (21) shows that it can also be obtained as the solution of the following minimization problem

$$\arg\min_{\mathbf{c}} \left\| \mathbf{Y} - \sum_{h=1}^{O} \mathbf{t}^{(h)} (\mathbf{c}^{(h)})^{\mathsf{T}} \right\|^2 \iff \arg\min_{\widehat{\mathbf{Y}}^{(O)}} \left\| \mathbf{Y} - \widehat{\mathbf{Y}}^{(O)} \right\|^2 . \tag{22}$$

This expression corresponds to a standard least square estimation problem and this indicates, therefore, that the quality of the PLS model can be evaluated like a standard linear regression model using the well-known Root Mean Square Error

$$\text{RMSE} = \frac{1}{\sqrt{Q}} \left\| \mathbf{Y} - \widehat{\mathbf{Y}}^{(O)} \right\| . \tag{23}$$

## Appendix 2: multiblock redundancy analysis

MBRA can be expressed as the solution of the following optimization problem (24)

$$\delta = \arg\max_{\mathbf{v}, \mathbf{t}^k, a^k} \left( \text{cov}^2(\mathbf{u}, \mathbf{t}) \right) \quad \text{with} \quad \mathbf{u} = \mathbf{Y}\mathbf{v}, \quad \mathbf{t} = \sum_{k=1}^{K} a^k \mathbf{t}^k, \quad \mathbf{t}^k = \mathbf{X}^k \mathbf{w}^k$$

$$\text{under the constraints that}$$

$$\sum_{k=1}^{K} (a^k)^2 = 1 \quad \text{and} \quad \|\mathbf{t}^k\|^2 = \|\mathbf{v}\|^2 = 1 . \tag{24}$$

It can be shown that the solution of this problem is obtained by taking $\mathbf{v}$ as the first eigenvector of the matrix

$$\sum_{k=1}^{K} \mathbf{Y}^{\top}\mathbf{P}^{k}\mathbf{Y} \tag{25}$$

(see, e.g., Bougeard et al. 2007, 2011a for proofs and details).

In MBRA, block components come from the normalized projections of $\mathbf{u}$ onto each subspace spanned by the variables of $\mathbf{X}^k$ and are computed as

$$\mathbf{t}^k = \frac{\mathbf{P}^k\mathbf{u}}{\|\mathbf{P}^k\mathbf{u}\|} \quad \text{with} \quad a^k = \frac{\mathbf{u}^{\top}\mathbf{t}^k}{\sqrt{\sum_{k=1}^{K}\left(\mathbf{u}^{\top}\mathbf{t}^k\right)^2}}. \tag{26}$$

In MBRA, the global component is obtained as the weighted sum of the block components, namely

$$\mathbf{t} = \sum_{k=1}^{K} a^k\mathbf{t}^k \quad \text{with} \quad \sum_{k=1}^{K}(a^k)^2 = 1 \quad \text{and} \quad \|\mathbf{t}^k\|^2 = 1\,. \tag{27}$$

It can be noticed that global as well as block components of MBRA take into account the partition of the explanatory variables in blocks. Furthermore—compared to MBPLS—MBRA is more oriented towards the explanation of $\mathbf{Y}$ but will be less stable in case of multicollinearity within explanatory blocks because it requires matrix inversions $\left(\text{i.e., } \left((\mathbf{X}^k)^{\top}\mathbf{X}^k\right)^{-1}\right)$ as indicated in Eqs. (25) and (26) see, for details, Bougeard et al. (2011a).

As for MBPLS, the effect of the component $\mathbf{t}$ is regressed out of $\mathbf{X}$ through the deflation of $\mathbf{X}$ upon this global component following Eq. (15). Subsequent components are then obtained by replacing matrix $\mathbf{X}$ in Eq. (24) by its successive residual matrices.

In a second step, the dependent dataset $\mathbf{Y}$ is predicted using the successive components $(\mathbf{t}^{(1)}, \ldots, \mathbf{t}^{(H)})$ and Eqs. (16) and (17) for $O$—the optimal number of components in the model (in general obtained through a cross-validation procedure).

As for MBPLS, the regression step of MBRA can be interpreted as the solution to the following optimization problem

$$\underset{\mathbf{c}}{\arg\min} \quad \left\| \mathbf{Y} - \sum_{h=1}^{O} a^{k(h)}\mathbf{t}^{k(h)}(\mathbf{c}^{(h)})^{\top} \right\|^2 \quad \text{with} \quad \mathbf{t}^{k(h)} = \mathbf{X}^{k(h-1)}\mathbf{w}^{k(h)} \tag{28}$$

where $\mathbf{X}^{k(h-1)}$ is the residual of the prediction of $\mathbf{X}^k$ from the $h-1$ previous components $(\mathbf{t}^{(1)}, \ldots, \mathbf{t}^{(h-1)})$.

## Appendix 3: computation times for some representative case studies

See Table 8.

**Table 8** Computation times of the clusterwise multiblock algorithm depending on the number of observations $N$, of explanatory variables $J$, of components included in the model $H$, and of clusters $G$

| Parameter values | | | | High-performance computer | Standard computer |
|---|---|---|---|---|---|
| $N$ | $J$ | $O$ | $G$ | (32 processing cores, 192 Go RAM) | (4 processing cores, 8 Go RAM) |
| 100 | 20 | 1 | 2 | 17.0 s | 44.8 s |
| 100 | 20 | 1 | 5 | 28.2 s | 1.72 min |
| 100 | 20 | 10 | 2 | 34.9 s | 2.24 min |
| 100 | 20 | 10 | 5 | 1.43 min | 4.57 min |
| 100 | 100 | 1 | 2 | 32.2 s | 2.04 min |
| 100 | 100 | 1 | 5 | 57.1 s | 4.22 min |
| 100 | 100 | 10 | 2 | 9.48 min | 42.6 min |
| 100 | 100 | 10 | 5 | 23.4 min | 1.79 h |
| 500 | 20 | 1 | 2 | 1.26 min | 5.33 min |
| 500 | 20 | 1 | 5 | 2.12 min | 9.57 min |
| 500 | 20 | 10 | 2 | 5.55 min | 26.1 min |
| 500 | 20 | 10 | 5 | 7.06 min | 32.3 min |
| 500 | 100 | 1 | 2 | 3.43 min | 16.6 min |
| 500 | 100 | 1 | 5 | 6.05 min | 29.6 min |
| 500 | 100 | 10 | 2 | 20.8 min | 1.76 h |
| 500 | 100 | 10 | 5 | 43.3 min | 2.92 h |

Computation times were compared on two computers with a high-level or a standard performance. Parallel computing was used to handle the 20 iterations of the algorithm. Some parameters were fixed because they turned out to have little influence on the computation time—namely CW.MBRA with no block weighting, ($K = 2$) blocks of equal sizes where the within-block correlation was equal to .70 and the between-block correlation was equal to .50, clusters were of equal sizes with a cluster separation of 2 units

## References

Abdi H, Williams L (2012) Partial least squares methods: partial least squares correlation and partial least square regression. In: Reisfeld B, Mayeno A (eds) Methods in molecular biology: computational toxicology. Springer, New York, pp 549–579

Bock H (1969) The equivalence of two extremal problems and its application to the iterative classification of multivariate data. In: Vortragsausarbeitung, Tagung. Mathematisches Forschungsinstitut Oberwolfach

Bougeard S, Cardinal M (2014) Multiblock modeling for complex preference study. Application to European preferences for smoked salmon. Food Qual Prefer 32:56–64

Bougeard S, Hanafi M, Qannari E (2007) ACPVI multibloc. Application à des données d'épidémiologie animale. Journal de la Société Française de Statistique 148:77–94

Bougeard S, Qannari E, Lupo C, Hanafi M (2011a) From multiblock partial least squares to multiblock redundancy analysis. A continuum approach. Informatica 22:11–26

Bougeard S, Qannari E, Rose N (2011b) Multiblock redundancy analysis: interpretation tools and application in epidemiology. J Chemom 25:467–475

Bry X, Verron T, Redont P, Cazes P (2012) THEME-SEER: a multidimensional exploratory technique to analyze a structural model using an extended covariance criterion. J Chemom 26:158–169
Charles C (1977) Régression typologique et reconnaissance des formes. PhD thesis, University of Paris IX, France
De Roover K, Ceulemans C, Timmerman M (2012) Clusterwise simultaneous component analysis for analyzing structural differences in multivariate multiblock data. Psychol Methods 17:100–119
DeSarbo W, Cron W (1988) A maximum likelihood methodology for clusterwise linear regression. J Classif 5:249–282
Diday E (1976) Classification et sélection de paramètres sous contraintes. Technical report, IRIA-LABORIA
Dolce P, Esposito Vinzi V, Lauro C (2016) Path directions incoherence in PLS path modeling: a prediction-oriented solution. In: Abdi H, Esposito Vinzi V, Russolillo G, Saporta G, Trinchera L (eds) The multiple facets of partial least squares and related methods. Springer proceedings in mathematics & statistics. Springer, Berlin, pp 59–59
Hahn C, Johnson M, Hermann AFA (2002) Capturing customer heterogeneity using finite mixture PLS approach. Schmalenbach Bus Rev 54:243–269
Hubert H, Arabie P (1985) Comparing partitions. J Classif 2:193–218
Hwang H, Takane Y (2004) Generalized structured component analysis. Psychometrika 69:81–99
Hwang H, DeSarbo S, Takane Y (2007) Fuzzy clusterwise generalized structured component analysis. Psychometrika 72:181–198
Kissita G (2003) Les analyses canoniques généralisées avec tableau de référence généralisé : éléments théoriques et appliqués. PhD thesis, University of Paris Dauphine, France
Lohmoller J (1989) Latent variables path modeling with partial least squares. Physica-Verlag, Heidelberg
Martella F, Vicari D, Vichi M (2015) Partitioning predictors in multivariate regression models. Stat Comput 25:261–272
Preda C, Saporta G (2005) Clusterwise PLS regression on a stochastic process. Comput Stat Data Anal 49:99–108
Qin S, Valle S, Piovoso M (2001) On unifying multiblock analysis with application to decentralized process monitoring. J Chemom 15:715–742
Sarstedt M (2008) A review of recent approaches for capturing heterogeneity in partial least squares path modelling. J Model Manage 3:140–161
Schlittgen R, Ringle C, Sarstedt M, Becker JM (2016) Segmentation of PLS path models by iterative reweighted regressions. J Bus Res 69:4583–4592
Shao Q, Wu Y (2005) Consistent procedure for determining the number of clusters in regression clustering. J Stat Plan Inference 135:461–476
Spath H (1979) Clusterwise linear regression. Computing 22:367–373
Team R (2015) R: a language and environment of statistical computing. http://cran.r-project.org/
Tenenhaus A, Tenenhaus M (2011) Regularized generalized canonical correlation analysis. Psychometrika 76:257–284
Tenenhaus M (1998) La régression PLS. Technip, Paris
Trinchera L (2007) Unobserved heterogeneity in structural equation models: a new approach to latent class detection in PLS path modeling. PhD thesis, University of Naples Federico II
Vicari D, Vichi M (2013) Multivariate linear regression for heterogeneous data. J Appl Stat 40:1209–1230
Vinzi V, Lauro C, Amato S (2005) PLS typological regression. In: Vichi M, Monari P, Mignani S, Montanari A (eds) New developments in classification and data analysis. Springer, Berlin, pp 133–140
Vinzi V, Ringle C, Squillacciotti S, Trinchera L (2007) Capturing and treating unobserved heterogeneity by response based segmentation in PLS path modeling. a comparison of alternative methods by computational experiments. Technical reports, ESSEC Business School, https://www.academia.edu/168969/Capturing_and_Treating_Unobserved_Heterogeneity_by_Response_Based_Segmentation_in_PLS_Path_Modeling._A_Comparison_of_Alternative_Methods_by_Computational_Experiments
Vinzi V, Trinchera L, Squillacciotti S, Tenenhaus M (2009) REBUS-PLS: a response-based procedure for detecting unit segments in pls path modeling. Appl Stochastic Models Bus Ind 24:439–458
Vivien M (2002) Approches PLS linéaires et non-linéaires pour la modélisation de multi-tableaux : théorie et applications. PhD thesis, University of Montpellier 1, France
Westerhuis J, Coenegracht P (1997) Multivariate modelling of the pharmaceutical two-step process of wet granulation and tableting with multiblock partial least squares. J Chemom 11:379–392
Westerhuis J, Smilde A (2001) Deflation in multiblock PLS. J Chemom 15:485–493

Westerhuis J, Kourti T, MacGregor J (1998) Analysis of multiblock and hierarchical PCA and PLS model. J Chemom 12:301–321

Wold H (1985) Encyclopedia of statistical sciences. In: Kotz S, Johnson N (eds) Partial least squares. Wiley, New York, pp 581–591

Wold S (1984) Three PLS algorithms according to SW. Technical reports, Umea University, Sweden

Wold S, Martens H, Wold H (1983) The multivariate calibration problem in chemistry solved by the PLS method. Matrix Pencils pp 286–293