

# Supervised strategies of multiblock data analysis: a unified approach and extensions

Essomanda **Tchandao Mangamana**<sup>(a, b)\*</sup>, Romain **Glèlè Kakai**<sup>b</sup> and  
El Mostafa **Qannari**<sup>a</sup>

<sup>a</sup>*StatSC, ONIRIS, INRA, 44322, Nantes, France*

<sup>b</sup>*LABEF, University of Abomey-Calavi, 04 BP 1525, Cotonou, Bénin*

\*Corresponding author: tchanesso@yahoo.fr

## Abstract

Within the framework of multiblock data analysis, a unified approach of supervised methods is discussed. It encompasses multiblock redundancy analysis (MB-RA) and multiblock partial least squares (MB-PLS) regression. Moreover, we develop new supervised strategies of multiblock data analysis, which can be seen as variants of one or the other of these two methods. They are respectively referred to as multiblock weighted redundancy analysis (MB-WRA) and multiblock weighted covariate analysis (MB-WCov). The four methods are based on the determination of latent variables associated with the various blocks of variables. They are derived from clear optimization criteria whose aim is to maximize either the sum of the covariances or the sum of squared covariances between the latent variable associated with the response block of variables and the block latent variables associated with the various explanatory blocks of variables. We also propose indices to help better interpreting the outcomes of the analyses. The methods are illustrated and compared based on simulated and real datasets.

Keywords: Multiblock data analysis, redundancy analysis, PLS regression, supervised methods, multicollinearity.

## 1 Introduction

With the advent of technology, several blocks of variables are very often collected in different areas to study some phenomena of interest. As example, for the characterization and optimization of food products, several measurements can be made (sensory, instrumental, physico-chemical) and linked to consumers preferences. In ecology, we may be interested in exploring the relationships between the abundance of some species in different sites, on the one hand, and, the variables describing those sites (environment, biodiversity, spatial situation, etc.), on the other hand.

We consider the setting where all the blocks of variables are measured on the same individuals, but the variables can be different from one block to another. For the purpose of predicting a block of variables from other blocks of variables, supervised methods are often used. Examples of such methods are Multiblock Redundancy Analysis (MB-RA) [1–4], Multiblock PLS (MB-PLS) regression [5–7], Hierarchical PLS (H-PLS) [6] and P-ComDim [8].

The purpose of this paper is to: (i) set up a unified approach for MB-RA and MB-PLS regression, (ii) propose new supervised strategies for the analysis of multiblock data, (iii) highlight the similarities and differences among all these methods, (iv) propose new indices for a better interpretation of the results of these supervised strategies.

Although self containing, this paper can be seen as a follow up of the unified framework for unsupervised strategies of multiblock data analysis [9].

We start by introducing Redundancy Analysis [10,11] and PLS2 regression analysis [12] using an original presentation. Then, from two different viewpoints, we extend these two strategies to the framework of multiblock data analysis. The first extension of RA leads to MB-RA, whereas the second extension gives a new strategy of analysis that can be seen as a variant of MB-RA. This new strategy of analysis is referred to as Multiblock Weighted Redundancy Analysis (MB-WRA). In the same vein, we consider a first extension of PLS2 regression, which leads to MB-PLS regression and a second extension which leads to an interesting variant of MB-PLS, to which we refer to as Multiblock Weighted Covariate Analysis (MB-WCov). We also exhibit indices that reflect the importance of the latent

variables determined at the successive stages and indices that highlight the contribution of the various explanatory blocks of variables to the determination of the global latent variables. All the methods of analysis are compared and illustrated using simulated and real datasets.

The paper is organized as follows. In section 2, we introduce the various methods of analysis. Then, in section 3, we illustrate and compare these methods on the basis of simulated and real datasets. We end the paper by a discussion and some concluding remarks.

## 2 Methods

### 2.1 Relationships between two datasets

#### 2.1.1 Redundancy analysis

Let us denote by  $\mathbf{X}$  ( $n \times p$ ) and  $\mathbf{Y}$  ( $n \times q$ ), two datasets measured on the same individuals and supposed to be column-centered. The aim is to predict  $\mathbf{Y}$  from  $\mathbf{X}$ . Let us denote by  $\mathbf{P}_X = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top$ , the orthogonal projector upon the space generated by the variables of  $\mathbf{X}$ . We consider a latent variable associated with  $\mathbf{Y}$ ,  $\mathbf{u} = \mathbf{Y}\boldsymbol{\nu}$ , where  $\boldsymbol{\nu}$  is the vector of loading weights, which is constrained to be of length equal to 1 ( $||\boldsymbol{\nu}|| = 1$ ). The projection of  $\mathbf{u}$  upon the space generated by the variables of  $\mathbf{X}$  is given by the latent variable  $\mathbf{t} = \mathbf{P}_X \mathbf{u}$ . We seek  $\mathbf{u}$ , and, therefore,  $\mathbf{t}$  so as to maximize the criterion:

$$\text{cov}(\mathbf{u}, \mathbf{t}) = \frac{1}{n} \mathbf{u}^\top \mathbf{t} = \frac{1}{n} \boldsymbol{\nu}^\top \mathbf{Y}^\top \mathbf{P}_X \mathbf{Y} \boldsymbol{\nu} \quad (1)$$

The rationale behind this maximization problem is clear enough: we seek a latent variable  $\mathbf{u}$  in the  $\mathbf{Y}$ -space and a latent variable  $\mathbf{t}$  in the  $\mathbf{X}$ -space that are as close as possible to each other.

It is clear that the vector,  $\boldsymbol{\nu}$ , that maximizes such a criterion is given by the eigenvector of  $\mathbf{Y}^\top \mathbf{P}_X \mathbf{Y}$  associated with the largest eigenvalue. Thus, we are led to the same solution as redundancy analysis of  $\mathbf{X}$  and  $\mathbf{Y}$ . A Non Iterative PArtial Least Squares (NIPALS)-like

algorithm to run this analysis is as follows:

0. Choose randomly  $\nu$  and set  $\nu = \nu / ||\nu||$ ;
1. Compute the  $\mathbf{Y}$ -latent variable:  $\mathbf{u} = \mathbf{Y}\nu$ ;
2. Compute the  $\mathbf{X}$ -latent variable:  $\mathbf{t} = \mathbf{P}_X\mathbf{u}$ ;
3. Update  $\nu$ :  $\nu = \mathbf{Y}^\top \mathbf{t} / ||\mathbf{Y}^\top \mathbf{t}||$ ;
4. Iterate starting from step 1, until convergence. That is, the criterion to be maximized ceases to increase by more than a pre-specified threshold (e.g.,  $\epsilon = 10^{-8}$ ).

We can notice that, since the projector  $\mathbf{P}_X$  is idempotent ( $\mathbf{P}_X^2 = \mathbf{P}_X$ ) and symmetric ( $\mathbf{P}_X^\top = \mathbf{P}_X$ ), we have:

$$\mathbf{Y}^\top \mathbf{P}_X \mathbf{Y} = \mathbf{Y}^\top \mathbf{P}_X \mathbf{P}_X \mathbf{Y} = \mathbf{Y}^\top \mathbf{P}_X^\top \mathbf{P}_X \mathbf{Y} = (\mathbf{P}_X \mathbf{Y})^\top (\mathbf{P}_X \mathbf{Y}) \quad (2)$$

It follows that  $\text{trace}(\mathbf{Y}^\top \mathbf{P}_X \mathbf{Y}) = n \sum_{j=1}^q \text{var}(\hat{y}_j)$ , where  $\hat{y}_j = \mathbf{P}_X \mathbf{y}_j$  is the projection of the  $j^{\text{th}}$  variable,  $\mathbf{y}_j$  of  $\mathbf{Y}$  on the space spanned by the variables in  $\mathbf{X}$ . In other words,  $\text{trace}(\mathbf{Y}^\top \mathbf{P}_X \mathbf{Y})$  reflects the total variance in  $\mathbf{Y}$  explained by  $\mathbf{X}$ . Similarly,

$$\nu^\top \mathbf{Y}^\top \mathbf{P}_X \mathbf{Y} \nu = \mathbf{u}^\top \mathbf{P}_X \mathbf{u} = (\mathbf{P}_X \mathbf{u})^\top (\mathbf{P}_X \mathbf{u}) = n \times \text{var}(\mathbf{P}_X \mathbf{u}) \quad (3)$$

Therefore, an index of particular interest is:

$$I = \frac{\nu^\top \mathbf{Y}^\top \mathbf{P}_X \mathbf{Y} \nu}{\text{trace}(\mathbf{Y}^\top \mathbf{P}_X \mathbf{Y})} = \frac{\text{var}(\mathbf{P}_X \mathbf{u})}{\sum_{j=1}^q \text{var}(\mathbf{P}_X \mathbf{y}_j)} = \frac{\text{var}(\mathbf{t})}{\sum_{j=1}^q \text{var}(\mathbf{P}_X \mathbf{y}_j)} \quad (4)$$

This index reflects the variation in  $\mathbf{P}_X \mathbf{Y}$  recovered by the latent variable  $\mathbf{t}$ . It ranges between 0 and 1. Obviously, the analysis seeks to maximize this index  $I$ .

It is worth noting that for the particular case where  $\mathbf{Y} = [\mathbf{y}]$  (a single variable), we are led to multiple linear regression.

Subsequent latent variables could be retrieved after a deflation of  $\mathbf{X}$  and  $\mathbf{Y}$ , with respect to the  $\mathbf{X}$ -latent variable  $\mathbf{t} = \mathbf{P}_X \mathbf{u}$ . More precisely, this deflation procedure

consists of subtracting from the datasets  $\mathbf{X}$  and  $\mathbf{Y}$ , the information that has already been explained by the previous latent variables in  $\mathbf{X}$  [13, 14]. This information is computed by projecting the datasets  $\mathbf{X}$  and  $\mathbf{Y}$  onto these latent variables. This leads us to compute the  $\mathbf{Y}$ -latent variables  $\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \dots$  and correspondingly, the  $\mathbf{X}$ -latent variables  $\mathbf{t}^{(1)}, \mathbf{t}^{(2)}, \dots$  etc. At each stage, we can compute the index:

$$I_h = \frac{\text{var}(\mathbf{t}^{(h)})}{\sum_{j=1}^q \text{var}(\mathbf{P}_X \mathbf{y}_j)} \quad (5)$$

These indices could be plotted as a function of the latent variables at each stage,  $h$ , and could be interpreted similarly to the so-called scree diagram in Principal Components Analysis (PCA) [15] to select an appropriate number of latent variables to be retained from the analysis.

A prediction model of the variables in  $\mathbf{Y}$  from  $\mathbf{X}$  can be set by regressing  $\mathbf{Y}$  upon the latent variables  $\mathbf{t}^{(1)}, \mathbf{t}^{(2)}, \dots, \mathbf{t}^{(A)}$ . The number of latent variables,  $A$ , to be introduced in the prediction model can be selected by a cross-validation procedure [16].

The projector,  $\mathbf{P}_X$ , involves the inversion of the matrix,  $\mathbf{X}^\top \mathbf{X}$ . The advantage of this inversion is that the correlations as well as the variances of the variables in  $\mathbf{X}$  are shaded off. Therefore, the analysis is targeted at recovering the variation in  $\mathbf{Y}$  by means of the variables of  $\mathbf{X}$ . However, a major drawback of the inversion of such a matrix is that, in case of high redundancy among the variables of  $\mathbf{X}$ , the analysis is likely to lead to unstable results because of the well known problem of quasi-collinearity [17].

### 2.1.2 PLS2 regression

We consider the same setting as in the previous section. By way of circumventing the problem of quasi-collinearity, we consider the operator  $\mathbf{W}_X = \mathbf{X} \mathbf{X}^\top$  instead of the projector  $\mathbf{P}_X = \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$ . This means that the matrix  $(\mathbf{X}^\top \mathbf{X})^{-1}$  is replaced by the identity matrix.

To a latent variable  $\mathbf{u} = \mathbf{Y} \boldsymbol{\nu}$  ( $\|\boldsymbol{\nu}\| = 1$ ), we associate the latent variable  $\mathbf{t} = \mathbf{W}_X \mathbf{u}$ .

Thereafter, we seek  $\mathbf{u}$  (and, consequently,  $\mathbf{t}$ ) so as to maximize the criterion:

$$\text{cov}(\mathbf{u}, \mathbf{t}) = \frac{1}{n} \mathbf{u}^\top \mathbf{t} = \frac{1}{n} \boldsymbol{\nu}^\top \mathbf{Y}^\top \mathbf{X} \mathbf{X}^\top \mathbf{Y} \boldsymbol{\nu}. \quad (6)$$

It follows that the vector,  $\boldsymbol{\nu}$ , that maximizes this quadratic form is given by the eigenvector of  $\mathbf{Y}^\top \mathbf{X} \mathbf{X}^\top \mathbf{Y}$  associated with the largest eigenvalue. Obviously, we are led to the same solution as PLS regression [12, 18]. A NIPALS-like algorithm to run this method is the following:

0. Choose randomly the vector of loadings,  $\boldsymbol{\nu}$ , associated with  $\mathbf{Y}$  and set  $\boldsymbol{\nu} = \boldsymbol{\nu} / \|\boldsymbol{\nu}\|$ ;
1. Compute the latent variable,  $\mathbf{u}$ , associated with  $\mathbf{Y}$ :  $\mathbf{u} = \mathbf{Y} \boldsymbol{\nu}$ ;
2. Compute the latent variable,  $\mathbf{t}$ , associated with  $\mathbf{X}$ :  $\mathbf{t} = \mathbf{W}_X \mathbf{u}$ ;
3. Update the vector  $\boldsymbol{\nu}$ :  $\boldsymbol{\nu} = \mathbf{Y}^\top \mathbf{t} / \|\mathbf{Y}^\top \mathbf{t}\|$ ;
4. Iterate starting from step 1, until convergence. That is the criterion to be maximized ceases to increase by more than a pre-specified threshold (e.g.,  $\epsilon = 10^{-8}$ ).

The convergence of this algorithm is granted by the fact that at each step, the criterion to be maximized increases. Since this criterion is upper-bounded, it results that the series of values of this criterion generated in the course of the iterations converges.

Since PLS2 regression revolves around the eigenanalysis of the matrix  $\mathbf{Y}^\top \mathbf{X} \mathbf{X}^\top \mathbf{Y}$ , we can remark that  $\text{trace}(\mathbf{Y}^\top \mathbf{X} \mathbf{X}^\top \mathbf{Y}) = \text{trace}(\mathbf{X} \mathbf{X}^\top \mathbf{Y} \mathbf{Y}^\top)$ . By using the relationships  $\mathbf{X} \mathbf{X}^\top = \sum_{i=1}^p \mathbf{x}_i \mathbf{x}_i^\top$  and  $\mathbf{Y} \mathbf{Y}^\top = \sum_{j=1}^q \mathbf{y}_j \mathbf{y}_j^\top$ , where  $\mathbf{x}_i$  (respectively,  $\mathbf{y}_j$ ) is the  $i^{\text{th}}$  (respectively,  $j^{\text{th}}$ ) variable of  $\mathbf{X}$  (respectively,  $\mathbf{Y}$ ), it readily follows that  $\text{trace}(\mathbf{Y}^\top \mathbf{X} \mathbf{X}^\top \mathbf{Y}) = n^2 \sum_{i=1}^p \sum_{j=1}^q \text{cov}^2(\mathbf{x}_i, \mathbf{y}_j)$ . Therefore, the quantity:

$\text{trace}(\mathbf{Y}^\top \mathbf{X} \mathbf{X}^\top \mathbf{Y})$  reflects the strength of the link between the variables in  $\mathbf{X}$  and those in  $\mathbf{Y}$ . This index was introduced by Robert and Escouffier [19, 20] and it is tightly linked to the so-called RV coefficient, which is widely used in sensometrics and chemometrics [21–25]. Similarly, we have:  $\boldsymbol{\nu}^\top \mathbf{Y}^\top \mathbf{X} \mathbf{X}^\top \mathbf{Y} \boldsymbol{\nu} = n^2 \sum_{i=1}^p \text{cov}^2(\mathbf{x}_i, \mathbf{u})$ . Conse-

quently, the index:

$$\begin{aligned} I &= \frac{\nu^\top \mathbf{Y}^\top \mathbf{X} \mathbf{X}^\top \mathbf{Y} \nu}{\text{trace}(\mathbf{Y}^\top \mathbf{X} \mathbf{X}^\top \mathbf{Y})} = \frac{n \times \text{cov}(t, u)}{\text{trace}(\mathbf{Y}^\top \mathbf{X} \mathbf{X}^\top \mathbf{Y})} = \frac{\sum_{i=1}^p \text{cov}^2(x_i, u)}{\sum_{i=1}^p \sum_{j=1}^q \text{cov}^2(x_i, y_j)} \\ &= \frac{\lambda_1}{\text{trace}(\mathbf{Y}^\top \mathbf{X} \mathbf{X}^\top \mathbf{Y})} \end{aligned} \quad (7)$$

where  $\lambda_1$  is the largest eigenvalue of  $\mathbf{Y}^\top \mathbf{X} \mathbf{X}^\top \mathbf{Y}$ . This index reflects the proportion of covariation (i.e.,  $\text{trace}(\mathbf{Y}^\top \mathbf{X} \mathbf{X}^\top \mathbf{Y})$ ) recovered by  $\mathbf{t}$  (and  $\mathbf{u}$ ).

Subsequent latent variables could be determined after a deflation of  $\mathbf{X}$  and  $\mathbf{Y}$  with respect to the latent variables associated with  $\mathbf{X}$ , determined at previous stages. Therefore, we obtain the  $\mathbf{Y}$ -latent variables  $\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \dots$  and the corresponding  $\mathbf{X}$ -latent variables  $\mathbf{t}^{(1)}, \mathbf{t}^{(2)}, \dots$  etc. At each stage, the index:

$$I_h = \frac{\sum_{i=1}^p \text{cov}^2(x_i, \mathbf{u}^{(h)})}{\sum_{i=1}^p \sum_{j=1}^q \text{cov}^2(x_i, y_j)} \quad (8)$$

could be computed. These indices could be plotted as a scree diagram. This yields a tool to help choosing the number of latent variables to be retained.

## 2.2 Multiblock data analysis

We consider the multiblock setting where we have a dataset  $\mathbf{Y}$  ( $n \times q$ ) to be predicted by  $K$  datasets  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K$ ; the dimension of  $\mathbf{X}_k$  ( $k = 1, 2, \dots, K$ ) is  $n \times p_k$ . All these datasets are measured on the same individuals and assumed to be column-centered. Moreover, in order to set all the  $\mathbf{X}$ -datasets on the same footing, they can be pre-scaled so as to have their norms equal to 1. This is achieved by dividing each dataset  $\mathbf{X}_k$  by its norm  $\|\mathbf{X}_k\| = \sqrt{\text{trace}(\mathbf{X}_k^\top \mathbf{X}_k)}$ .

### 2.2.1 Multiblock Redundancy Analysis

Let us denote by  $\mathbf{P}_k = \mathbf{X}_k (\mathbf{X}_k^\top \mathbf{X}_k)^{-1} \mathbf{X}_k^\top$  the orthogonal projector upon the space generated by the variables of dataset  $\mathbf{X}_k$ .

Starting with a  $\mathbf{Y}$ -latent variable  $\mathbf{u} = \mathbf{Y} \nu$  ( $\|\nu\| = 1$ ), we consider its orthogonal

projection upon the space generated by the variables of  $\mathbf{X}_k$ . Thus, we obtain  $\mathbf{t}_k = \mathbf{P}_k \mathbf{u}$ , which define the block latent variables associated with the datasets  $\mathbf{X}_k$  ( $k = 1, 2, \dots, K$ ). We seek  $\mathbf{u}$ , and, consequently  $\mathbf{t}_k$ , so as to maximize:

$$\sum_{k=1}^K \text{cov}(\mathbf{u}, \mathbf{t}_k) = \frac{1}{n} \mathbf{u}^\top \sum_{k=1}^K \mathbf{t}_k = \frac{1}{n} \boldsymbol{\nu}^\top \mathbf{Y}^\top \left( \sum_{k=1}^K \mathbf{P}_k \right) \mathbf{Y} \boldsymbol{\nu}. \quad (9)$$

The rationale behind this problem is clear: we seek a latent variable  $\mathbf{u}$  in the  $\mathbf{Y}$ -space that is as close as possible to the  $\mathbf{X}_k$ -spaces. The optimal vector,  $\boldsymbol{\nu}$ , of this quadratic form is given by the eigenvector of  $\mathbf{Y}^\top \left( \sum_{k=1}^K \mathbf{P}_k \right) \mathbf{Y}$  associated with the largest eigenvalue. Clearly, we are led to MB-RA [2–4].

We can remark that the criterion to be maximized can also be written as:

$$\sum_{k=1}^K \text{cov}(\mathbf{u}, \mathbf{t}_k) = \text{cov}(\mathbf{u}, \sum_{k=1}^K \mathbf{t}_k) = \text{cov}(\mathbf{u}, \mathbf{t}) \quad (10)$$

where  $\mathbf{t} = \sum_{k=1}^K \mathbf{t}_k$  stands as the global latent variable.

A NIPALS-like algorithm to solve the above maximization problem is as follows:

0. Choose randomly a vector  $\boldsymbol{\nu}$  and set  $\boldsymbol{\nu} = \boldsymbol{\nu} / \|\boldsymbol{\nu}\|$ ;
1. Determine the latent variable associated with  $\mathbf{Y}$ :  $\mathbf{u} = \mathbf{Y} \boldsymbol{\nu}$  and the block latent variable associated with  $\mathbf{X}_k$ :  $\mathbf{t}_k = \mathbf{P}_k \mathbf{u}$ ;
2. Compute the global latent variable  $\mathbf{t} = \sum_{k=1}^K \mathbf{t}_k$ ;
3. Update the vector  $\boldsymbol{\nu}$ :  $\boldsymbol{\nu} = \mathbf{Y}^\top \mathbf{t} / \|\mathbf{Y}^\top \mathbf{t}\|$ ;
4. Iterate starting from step 1, until convergence.

The block latent variable  $\mathbf{t}_k = \mathbf{X}_k (\mathbf{X}_k^\top \mathbf{X}_k)^{-1} \mathbf{X}_k^\top \mathbf{u}$  is a linear combination of the variables in  $\mathbf{X}_k$ :  $\mathbf{t}_k = \mathbf{X}_k \mathbf{w}_k$ , with  $\mathbf{w}_k = (\mathbf{X}_k^\top \mathbf{X}_k)^{-1} \mathbf{X}_k^\top \mathbf{u}$ . For interpretational purpose, this latter vector could be standardized to unit length. Similarly, the global latent variable  $\mathbf{t}$  can be written as  $\mathbf{t} = \mathbf{X} \mathbf{w}$ , where  $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K]$  and  $\mathbf{w} = (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_K)^\top$ . Again for interpretational purpose, this latter vector could be standardized to unit length.

Latent variables of higher order than 1 could be obtained by following the same strategy of analysis after deflating all the datasets with respect to the global latent variables

associated with  $\mathbf{X}$ , determined at previous stages. Therefore, we are led to computing the  $\mathbf{Y}$ -latent variables  $\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \dots$ , the  $\mathbf{X}_k$ -block latent variables  $\mathbf{t}_k^{(1)}, \mathbf{t}_k^{(2)}, \dots$  ( $k = 1, 2, \dots, K$ ) and the global latent variables  $\mathbf{t}^{(1)}, \mathbf{t}^{(2)}$ , etc.

As for the case of RA, the index:

$$I_h = \frac{\sum_{k=1}^K \mathbf{u}^{(h)\top} \mathbf{P}_k \mathbf{u}^{(h)}}{\sum_{k=1}^K \text{trace}(\mathbf{Y}^\top \mathbf{P}_k \mathbf{Y})} = \frac{\sum_{k=1}^K \text{var}(\mathbf{P}_k \mathbf{u}^{(h)})}{\sum_{k=1}^K \sum_{j=1}^q \text{var}(\mathbf{P}_k \mathbf{y}_j)} = \frac{\sum_{k=1}^K \text{var}(\mathbf{t}_k^{(h)})}{\sum_{k=1}^K \sum_{j=1}^q \text{var}(\mathbf{P}_k \mathbf{y}_j)} \quad (11)$$

can be computed at each stage to reflect the variation in datasets  $\mathbf{P}_k \mathbf{Y}$  ( $k = 1, 2, \dots, K$ ) recovered by  $\mathbf{t}^{(h)}$ . These indices could be plotted as a function of the number of latent variables,  $h$ , and could be interpreted as a scree diagram to help choosing the number of latent variables to be retained.

At stage  $h$ , the quantity:

$$\text{cont}_k^{(h)} = \frac{\text{var}(\mathbf{P}_k \mathbf{u}^{(h)})}{\sum_{l=1}^K \text{var}(\mathbf{P}_l \mathbf{u}^{(h)})} = \frac{\text{var}(\mathbf{t}_k^{(h)})}{\sum_{l=1}^K \text{var}(\mathbf{t}_l^{(h)})} \quad (12)$$

reflects the contribution of the block  $\mathbf{X}_k$  in the determination of the components  $\mathbf{u}^{(h)}$  and  $\mathbf{t}^{(h)}$ .

### 2.2.2 Multiblock weighted redundancy analysis (MB-WRA)

Let us consider the same setting as in the previous section. We operate the same centering and pre-scaling of the datasets.

We consider a variant to the maximization problem. Instead of maximizing the quantity  $\sum_{k=1}^K \text{cov}(\mathbf{u}, \mathbf{P}_k \mathbf{u})$ , we propose to maximize the quantity:

$$\sum_{k=1}^K \text{cov}^2(\mathbf{u}, \mathbf{P}_k \mathbf{u}) = \frac{1}{n^2} \sum_{k=1}^K (\boldsymbol{\nu}^\top \mathbf{Y}^\top \mathbf{P}_k \mathbf{Y} \boldsymbol{\nu})^2 \quad (13)$$

under the constraint that  $\|\boldsymbol{\nu}\| = 1$ .

Obviously, the rationale behind this problem is exactly the same as previously, that is seeking a direction in the  $\mathbf{Y}$ -space that is as close as possible to the  $\mathbf{X}_k$ -spaces. To solve this problem, let us use the Lagrangian method. The Lagrange expression associated with

this maximization criterion is the following:

$$\sum_{k=1}^K (\nu^\top Y^\top P_k Y \nu)^2 - 2\mu(\nu^\top \nu - 1) \quad (14)$$

where  $2\mu$  is the Lagrange multiplier associated with the constraint  $\|\nu\| = 1$  or equivalently  $\|\nu\|^2 = \nu^\top \nu = 1$ . By deriving this Lagrange expression with respect to  $\nu$  and setting this derivative to 0, we obtain:

$$4 \sum_{k=1}^K (\nu^\top Y^\top P_k Y \nu) Y^\top P_k Y \nu - 4\mu\nu = 0. \quad (15)$$

If we denote by  $\lambda_k$ , the quantity  $\lambda_k = \nu^\top Y^\top P_k Y \nu$ , we can write:

$$\sum_{k=1}^K \lambda_k Y^\top P_k Y \nu = \mu\nu \quad (16)$$

Multiplying the two members of this equality by  $\nu^\top$  and setting  $\nu^\top \nu = 1$ , we obtain  $\mu = \sum_{k=1}^K \lambda_k^2$ . From this, we can derive the stationnary point as:  $\nu = \frac{\sum_{k=1}^K \lambda_k Y^\top P_k Y \nu}{\sum_{k=1}^K \lambda_k^2}$ .

This suggests the following iterative algorithm:

0. Choose randomly the vector  $\nu$  and set  $\nu = \nu / \|\nu\|$ ;
1.  $\lambda_k = \nu^\top Y^\top P_k Y \nu$ ;
2.  $\nu = \sum_{k=1}^K \lambda_k Y^\top P_k Y \nu / \sum_{k=1}^K \lambda_k^2$ ;
3.  $\nu = \nu / \|\nu\|$ ;
4. Iterate starting from step 1, until convergence.

We show that this algorithm converges (see appendix). More precisely, we show that at each iteration, the criterion to be maximized increases and since this criterion is upper-bounded, the series of the values of the criterion that we seek to maximize generated in the course of the algorithm converges.

Let us consider again the criterion that we sought to maximize. We have:

$$\sum_{k=1}^K cov^2(u, t_k) = \frac{1}{n} \sum_{k=1}^K \lambda_k cov(u, t_k) = \frac{1}{n} cov(u, \sum_{k=1}^K \lambda_k t_k) = \frac{1}{n} cov(u, t) \quad (17)$$

where  $t = \sum_{k=1}^K \lambda_k t_k$  stands as the global latent variable and appears as a linear combination of the  $X_k$ -block latent variables  $t_k = P_k u$ . More precisely, since  $\lambda_k = u^\top t_k = n \times cov(u, t_k)$ , it follows that  $t$  is proportional to the first PLS component of  $u$  upon  $t_1, t_2, \dots, t_K$ . These remarks suggest an alternative algorithm to solve the maximization problem above. Indeed, this criterion can also be expressed as:

$$\sum_{k=1}^K cov^2(u, t_k) = n \sum_{k=1}^K \lambda_k cov(u, t_k) = \nu^\top \left( \sum_{k=1}^K \lambda_k Y^\top P_k Y \right) \nu \quad (18)$$

Therefore, it follows that for fixed values of  $\lambda_k$ , the optimal vector,  $\nu$ , is given by the eigenvector of  $\sum_{k=1}^K \lambda_k Y^\top P_k Y$  associated with the largest eigenvalue. Conversely, for a fixed value of  $\nu$ ,  $\lambda_k$  is given by:  $\lambda_k = \nu^\top Y^\top P_k Y \nu$ . The algorithm associated with this solution is the following:

0. Set  $\lambda_k = 1$  for  $k = 1, 2, \dots, K$ ;
1. Set  $\nu$  to the eigenvector of  $\sum_{k=1}^K \lambda_k Y^\top P_k Y$  associated with the largest eigenvalue;
2. Update  $\lambda_k = \nu^\top Y^\top P_k Y \nu$ ;
3. Iterate from step 1 until convergence, that is until the criterion to be maximized ceases to increase by more than a pre-specified threshold (e.g.,  $\epsilon = 10^{-8}$ ).

Subsequent latent variables of higher order can be determined by following the same strategy of analysis after a deflation of all the blocks of variables with respect to the global latent variable  $t$ .

Let us denote by  $t^{(h)}$  the global latent variable and by  $t_1^{(h)}, t_2^{(h)}, \dots, t_K^{(h)}$ , its associated block latent variables. Similarly, we denote by  $u^{(1)}, u^{(2)}, \dots, u^{(h)}$  the successive latent variables in the  $Y$ -space. Finally, we denote by  $\lambda_k^{(h)} = ncov(u^{(h)}, t_k^{(h)})$ . The following

indices can be very useful to better interpret the results.

$$I_h = \frac{\sum_{k=1}^K \lambda_k^{(h)}}{\sum_{k=1}^K \text{trace}(\mathbf{Y}^\top \mathbf{P}_k \mathbf{Y})} \quad (19)$$

highlights the importance of the global component  $\mathbf{t}^{(h)}$  in explaining the covariation of  $\mathbf{Y}$  and  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K$ .

$$\text{cont}_k^{(h)} = \frac{\lambda_k^{(h)}}{\sum_{k=1}^K \lambda_k^{(h)}} \quad (20)$$

reflects the contribution of the block  $\mathbf{X}_k$  in the determination of the latent variables  $\mathbf{t}^{(h)}$  and  $\mathbf{u}^{(h)}$ .

### 2.2.3 Multiblock PLS regression

In order to counteract the problems of collinearity that arise from the inversion of the matrices  $\mathbf{X}_k^\top \mathbf{X}_k$ , we propose, as for the case of PLS regression, to replace the matrices  $(\mathbf{X}_k^\top \mathbf{X}_k)^{-1}$  by the identity matrix and, therefore, consider the operators  $\mathbf{W}_k = \mathbf{X}_k \mathbf{X}_k^\top$ , instead of the projectors  $\mathbf{P}_k = \mathbf{X}_k (\mathbf{X}_k^\top \mathbf{X}_k)^{-1} \mathbf{X}_k^\top$ .

Starting from the  $\mathbf{Y}$ -latent variable,  $\mathbf{u} = \mathbf{Y}\boldsymbol{\nu}$  ( $\|\boldsymbol{\nu}\| = 1$ ), we define the  $\mathbf{X}_k$ -block latent variable as  $\mathbf{t}_k = \mathbf{W}_k \mathbf{u}$ . Thereafter, we seek  $\mathbf{u}$  and, therefore,  $\mathbf{t}_k$ , so as to maximize the criterion:

$$\begin{aligned} \sum_{k=1}^K \text{cov}(\mathbf{u}, \mathbf{t}_k) &= \frac{1}{n} \mathbf{u}^\top \sum_{k=1}^K \mathbf{t}_k = \frac{1}{n} \boldsymbol{\nu}^\top \mathbf{Y}^\top \left( \sum_{k=1}^K \mathbf{W}_k \right) \mathbf{Y} \boldsymbol{\nu} \\ &= \frac{1}{n} \boldsymbol{\nu}^\top \sum_{k=1}^K \mathbf{Y}^\top \mathbf{X}_k \mathbf{X}_k^\top \mathbf{Y} \boldsymbol{\nu} = \mathbf{n} \boldsymbol{\nu}^\top \left( \sum_{k=1}^K \mathbf{V}_{Yk} \mathbf{V}_{kY} \right) \boldsymbol{\nu} \end{aligned} \quad (21)$$

where  $\mathbf{V}_{kY} = \frac{1}{n} \mathbf{X}_k^\top \mathbf{Y}$  is the covariance matrix between  $\mathbf{X}_k$  and  $\mathbf{Y}$  and  $\mathbf{V}_{Yk} = \mathbf{V}_{kY}^\top$ . Since  $\sum_{k=1}^K \text{cov}(\mathbf{u}, \mathbf{t}_k) = \text{cov}(\mathbf{u}, \sum_{k=1}^K \mathbf{t}_k)$ , the latent variable  $\mathbf{t} = \sum_{k=1}^K \mathbf{t}_k$  stands as the global latent variable.

All in all, we are led to the same solution as MB-PLS regression. A straightforward algorithm to run this method is the following:

0. Randomly choose  $\boldsymbol{\nu}$ , with  $\|\boldsymbol{\nu}\| = 1$ ;

1.  $\mathbf{u} = \mathbf{Y}\boldsymbol{\nu}$ :  $\mathbf{Y}$ -latent variable;
2.  $\mathbf{t}_k = \mathbf{X}_k \mathbf{X}_k^\top \mathbf{u}$ :  $\mathbf{X}_k$  block latent variables;
3.  $\mathbf{t} = \sum_{k=1}^K \mathbf{t}_k$ : global latent variable;
4.  $\boldsymbol{\nu} = \mathbf{Y}^\top \mathbf{t} / \|\mathbf{Y}^\top \mathbf{t}\|$ ;
5. Iterate starting from step 1, until convergence.

It is clear that at each step of this algorithm, the criterion to be maximized increases and since this criterion is upper-bounded, the series of values of this criterion computed at the successive step of this algorithm converges.

We can also note that:

$$\sum_{k=1}^K \text{cov}(\mathbf{u}, \mathbf{t}_k) = \frac{1}{n} \boldsymbol{\nu}^\top \mathbf{Y}^\top \sum_{k=1}^K \mathbf{X}_k \mathbf{X}_k^\top \mathbf{Y} \boldsymbol{\nu} = \frac{1}{n} \boldsymbol{\nu}^\top \mathbf{Y}^\top \mathbf{X} \mathbf{X}^\top \mathbf{Y} \boldsymbol{\nu} \quad (22)$$

where  $\mathbf{X} = [\mathbf{X}_1 | \mathbf{X}_2 | \dots | \mathbf{X}_K]$  is the dataset obtained by horizontally merging the datasets  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K$ . This entails that the optimal vector  $\boldsymbol{\nu}$  is the eigenvector of matrix  $\mathbf{Y}^\top \mathbf{X} \mathbf{X}^\top \mathbf{Y}$  associated with the largest eigenvalue. In other words, the solution to MB-PLS regression amounts to performing PLS regression of  $\mathbf{Y}$  upon  $\mathbf{X}$  [6].

We can compute the index:

$$\mathbf{I} = \frac{\boldsymbol{\lambda}}{\sum_{k=1}^K \text{trace}(\mathbf{Y}^\top \mathbf{X}_k \mathbf{X}_k^\top \mathbf{Y})} = \frac{\boldsymbol{\lambda}}{\text{trace}(\mathbf{Y}^\top \mathbf{X} \mathbf{X}^\top \mathbf{Y})} \quad (23)$$

where  $\boldsymbol{\lambda} = \mathbf{n} \times \text{cov}(\mathbf{u}, \mathbf{t})$ . This index reflects the proportion of covariation between  $\mathbf{Y}$ , on the one hand, and  $\mathbf{X}_k$  ( $k = 1, 2, \dots, K$ ), on the other hand, that is explained by the global latent variable. Moreover, since we have  $\boldsymbol{\lambda} = \sum_{k=1}^K \boldsymbol{\lambda}_k$ , where  $\boldsymbol{\lambda}_k = \mathbf{u}^\top \mathbf{t}_k = \mathbf{n} \times \text{cov}(\mathbf{u}, \mathbf{t}_k)$ , we can compute the indices  $\text{cont}_k = \frac{\boldsymbol{\lambda}_k}{\boldsymbol{\lambda}}$ , which reflects the contribution of the blocks of variables  $\mathbf{X}_k$  in the determination of the global latent variable. In other words, these indices reflect the importance that each block of variables,  $\mathbf{X}_k$ , attaches to the global latent variable.

Subsequent latent variables can be computed following the same strategy of analysis after deflation with respect to the global latent variables associated with  $\mathbf{X}$ .

#### 2.2.4 Multiblock Weighted Covariate Analysis (MB-WCov)

We consider the same setting and the same notations as in the previous section. As a variant to the maximization criterion that led us to introduce MB-PLS regression, namely  $\sum_{k=1}^K \text{cov}(\mathbf{u}, \mathbf{t}_k)$ , we consider the criterion  $\sum_{k=1}^K \text{cov}^2(\mathbf{u}, \mathbf{t}_k)$ . The rationale behind these two criteria is the same, that is exploring the covariation between, on the one hand,  $\mathbf{Y}$  and, on the other hand, the blocks of variables  $\mathbf{X}_k$ . The latter maximization criterion will lead us to a strategy of analysis that we shall refer to as Multiblock Weighted Covariate Analysis (MB-WCov). Indeed, as with MB-WRA, this method of analysis will explicitly exhibit weights that reflect the contribution of the  $\mathbf{X}_k$ -blocks of variables to the determination of the latent variables at each stage.

To the  $\mathbf{Y}$ -latent variable  $\mathbf{u} = \mathbf{Y}\boldsymbol{\nu}$  ( $\|\boldsymbol{\nu}\| = 1$ ), we associate the variables  $\mathbf{t}_k = \mathbf{W}_k \mathbf{u}$ , which are the block latent variables associated with the datasets  $\mathbf{X}_k$ . Thereafter, we seek  $\mathbf{u}$  so as to maximize:

$$\sum_{k=1}^K \text{cov}^2(\mathbf{u}, \mathbf{t}_k) = \frac{1}{n^2} \sum_{k=1}^K (\boldsymbol{\nu}^\top \mathbf{Y}^\top \mathbf{W}_k \mathbf{Y} \boldsymbol{\nu})^2 \quad (24)$$

The Lagrangian expression associated to this problem is:

$$\sum_{k=1}^K (\boldsymbol{\nu}^\top \mathbf{Y}^\top \mathbf{W}_k \mathbf{Y} \boldsymbol{\nu})^2 - 2\mu(\boldsymbol{\nu}^\top \boldsymbol{\nu} - 1) \quad (25)$$

where  $2\mu$  is the Lagrange multiplier associated with the constraint  $\|\boldsymbol{\nu}\| = 1$  or equivalently  $\|\boldsymbol{\nu}\|^2 = \boldsymbol{\nu}^\top \boldsymbol{\nu} = 1$ . By deriving this Lagrange expression with respect to  $\boldsymbol{\nu}$  and setting this derivative to 0, we obtain  $4 \sum_{k=1}^K (\boldsymbol{\nu}^\top \mathbf{Y}^\top \mathbf{W}_k \mathbf{Y} \boldsymbol{\nu}) \mathbf{Y}^\top \mathbf{W}_k \mathbf{Y} \boldsymbol{\nu} - 4\mu\boldsymbol{\nu} = 0$ . By denoting  $\boldsymbol{\lambda}_k = \boldsymbol{\nu}^\top \mathbf{Y}^\top \mathbf{W}_k \mathbf{Y} \boldsymbol{\nu}$ , we have  $\sum_{k=1}^K \boldsymbol{\lambda}_k \mathbf{Y}^\top \mathbf{W}_k \mathbf{Y} \boldsymbol{\nu} = \mu\boldsymbol{\nu}$ . Multiplying the two members of this equality by  $\boldsymbol{\nu}^\top$  and setting  $\boldsymbol{\nu}^\top \boldsymbol{\nu} = 1$ , we obtain  $\mu = \sum_{k=1}^K \boldsymbol{\lambda}_k^2$ . The stationnary point is therefore given by  $\boldsymbol{\nu} = \frac{\sum_{k=1}^K \boldsymbol{\lambda}_k \mathbf{Y}^\top \mathbf{W}_k \mathbf{Y} \boldsymbol{\nu}}{\sum_{k=1}^K \boldsymbol{\lambda}_k^2}$ . This suggests the following iterative algorithm:

0. Choose randomly the vector  $\boldsymbol{\nu}$  and set  $\boldsymbol{\nu} = \boldsymbol{\nu} / \|\boldsymbol{\nu}\|$ ;
1.  $\boldsymbol{\lambda}_k = \boldsymbol{\nu}^\top \mathbf{Y}^\top \mathbf{W}_k \mathbf{Y} \boldsymbol{\nu}$ ;

$$2. \nu = \sum_{k=1}^K \lambda_k \mathbf{Y}^\top \mathbf{W}_k \mathbf{Y} \nu / \sum_{k=1}^K \lambda_k^2;$$

$$3. \nu = \nu / ||\nu||;$$

4. Iterate starting from step 1, until convergence.

The convergence of this algorithm can be shown using very similar developments as for the case of MB-WRA (appendix).

The criterion that we sought to maximize can be written as  $\sum_{k=1}^K \lambda_k \text{cov}(\mathbf{u}, \mathbf{t}_k) = \text{cov}(\mathbf{u}, \sum_{k=1}^K \lambda_k \mathbf{t}_k)$ . This yields the global latent variable  $\mathbf{t} = \sum_{k=1}^K \lambda_k \mathbf{t}_k$ , with  $\lambda_k = n \times \text{cov}(\mathbf{u}, \mathbf{t}_k)$ . It follows that  $\mathbf{t}$  is proportional to the first PLS component of  $\mathbf{u}$  upon  $\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_K$ . Thus, for fixed values of  $\lambda_k$ , the optimal vector,  $\nu$ , is given by the eigenvector of  $\sum_{k=1}^K \lambda_k \mathbf{Y}^\top \mathbf{W}_k \mathbf{Y}$  associated with the largest eigenvalue. Conversely, for a fixed value of  $\nu$ ,  $\lambda_k$  is given by:  $\lambda_k = \nu^\top \mathbf{Y}^\top \mathbf{W}_k \mathbf{Y} \nu$ . From these developments, we can propose an alternative algorithm for the resolution of the MB-WCov maximization criterion as follows.

0. Set  $\lambda_k = 1$  for  $k = 1, 2, \dots, K$ ;

1. Set  $\nu$  to the eigenvector of  $\sum_{k=1}^K \lambda_k \mathbf{Y}^\top \mathbf{W}_k \mathbf{Y}$  associated with the largest eigenvalue;

2. Update  $\lambda_k = \nu^\top \mathbf{Y}^\top \mathbf{W}_k \mathbf{Y} \nu$ ;

3. Iterate starting from step 1 until convergence, that is until the criterion to be maximized ceases to increase by more than a pre-specified threshold (e.g.,  $\epsilon = 10^{-8}$ ).

Subsequent latent variables of higher order can be determined after deflating all the blocks of variables with respect to the global latent variable.

Similar indices as for the case of MB-PLS regression can be computed to enhance the interpretation of the results.

### 2.3 Comparison of methods

Table 1 sums up the four methods of multiblock data analysis discussed in this paper.

Table 1: Overview of the supervised multiblock methods.

$$(\mathbf{u} = \mathbf{Y}\boldsymbol{\nu} (||\boldsymbol{\nu}|| = 1); \mathbf{P}_k = \mathbf{X}_k(\mathbf{X}_k^\top \mathbf{X}_k)^{-1}\mathbf{X}_k^\top; \mathbf{W}_k = \mathbf{X}_k\mathbf{X}_k^\top)$$

Method	$\mathbf{X}_k$ -block latent variable	Maximization criterion	Global latent variable	First algorithm	NIPALS-like Algorithm
MB-RA	$t_k = \mathbf{P}_k \mathbf{u}$	$\sum_{k=1}^K \text{cov}(\mathbf{u}, t_k)$	$t = \sum_{k=1}^K t_k$	$\boldsymbol{\nu}$ is the first eigenvector of $\mathbf{Y}^\top \left( \sum_{k=1}^K \mathbf{P}_k \right) \mathbf{Y}$ .	0. Initial $\boldsymbol{\nu} (  \boldsymbol{\nu}   = 1)$ 1. $\mathbf{u} = \mathbf{Y}\boldsymbol{\nu}; t_k = \mathbf{P}_k \mathbf{u}$ 2. $t = \sum_{k=1}^K t_k$ 3. $\boldsymbol{\nu} = \mathbf{Y}^\top t /   \mathbf{Y}^\top t  $ 4. Iterate from 1.
MB-WRA	$t_k = \mathbf{P}_k \mathbf{u}$	$\sum_{k=1}^K \text{cov}^2(\mathbf{u}, t_k)$	$t = \sum_{k=1}^K \lambda_k t_k$ with $\lambda_k = \mathbf{u}^\top t_k$	0. Set $\boldsymbol{\lambda}_k = \mathbf{1}$ 1. $\boldsymbol{\nu}$ is the first eigenvector of $\mathbf{Y}^\top \left( \sum_{k=1}^K \lambda_k \mathbf{P}_k \right) \mathbf{Y}$ . 2. $\lambda_k = \boldsymbol{\nu}^\top \mathbf{Y}^\top \mathbf{P}_k \mathbf{Y} \boldsymbol{\nu}$ 3. Iterate from 1.	0. Initial $\boldsymbol{\nu} (  \boldsymbol{\nu}   = 1)$ 1. $\mathbf{u} = \mathbf{Y}\boldsymbol{\nu}; t_k = \mathbf{P}_k \mathbf{u}$ 2. $\boldsymbol{\lambda}_k = \mathbf{u}^\top t_k$ 3. $t = \sum_{k=1}^K \lambda_k t_k$ 4. $\boldsymbol{\nu} = \mathbf{Y}^\top t /   \mathbf{Y}^\top t  $ 5. Iterate from 1.
MB-PLS	$t_k = \mathbf{W}_k \mathbf{u}$	$\sum_{k=1}^K \text{cov}(\mathbf{u}, t_k)$	$t = \sum_{k=1}^K t_k$	$\boldsymbol{\nu}$ is the first eigenvector of $\mathbf{Y}^\top \left( \sum_{k=1}^K \mathbf{W}_k \right) \mathbf{Y}$ .	0. Initial $\boldsymbol{\nu} (  \boldsymbol{\nu}   = 1)$ 1. $\mathbf{u} = \mathbf{Y}\boldsymbol{\nu}; t_k = \mathbf{W}_k \mathbf{u}$ 2. $t = \sum_{k=1}^K t_k$ 3. $\boldsymbol{\nu} = \mathbf{Y}^\top t /   \mathbf{Y}^\top t  $ 4. Iterate from 1.
MB-WCov	$t_k = \mathbf{W}_k \mathbf{u}$	$\sum_{k=1}^K \text{cov}^2(\mathbf{u}, t_k)$	$t = \sum_{k=1}^K \lambda_k t_k$ with $\lambda_k = \mathbf{u}^\top t_k$	0. Set $\boldsymbol{\lambda}_k = \mathbf{1}$ 1. $\boldsymbol{\nu}$ is first eigenvector of $\mathbf{Y}^\top \left( \sum_{k=1}^K \lambda_k \mathbf{W}_k \right) \mathbf{Y}$ . 2. $\lambda_k = \boldsymbol{\nu}^\top \mathbf{Y}^\top \mathbf{W}_k \mathbf{Y} \boldsymbol{\nu}$ 3. Iterate from 1.	0. Initial $\boldsymbol{\nu} (  \boldsymbol{\nu}   = 1)$ 1. $\mathbf{u} = \mathbf{Y}\boldsymbol{\nu}; t_k = \mathbf{W}_k \mathbf{u}$ 2. $\boldsymbol{\lambda}_k = \mathbf{u}^\top t_k$ 3. $t = \sum_{k=1}^K \lambda_k t_k$ 4. $\boldsymbol{\nu} = \mathbf{Y}^\top t /   \mathbf{Y}^\top t  $ 5. Iterate from 1.

It is clear that the four methods of multiblock data analysis discussed herein can be differentiated by two main features. The first key of differentiation is how the  $\mathbf{X}_k$ -block latent variables,  $t_k$ , are computed from the  $\mathbf{Y}$ -latent variable,  $\mathbf{u}$ . Two options are offered to us, namely, whether we consider  $t_k = \mathbf{P}_k \mathbf{u}$  or  $t_k = \mathbf{W}_k \mathbf{u}$ . The first option leads to methods of analysis pertaining to redundancy analysis (i.e., MB-RA and MB-WRA). The second option leads to methods of analysis akin to PLS regression (i.e., MB-PLS and MB-WCov). The second key of differentiation between the methods is the relationship between the  $\mathbf{X}_k$ -block latent variable,  $t_k$ , and the global latent variable,  $t$ . This relationship directly stems from the optimization criterion to determine the latent

variable. More precisely, the criterion based on the covariance between  $\mathbf{t}_k$  and  $\mathbf{u}$  leads to  $\mathbf{t} = \sum_{k=1}^K \mathbf{t}_k$  and the criterion based on the squared covariance between  $\mathbf{t}_k$  and  $\mathbf{u}$  leads to  $\mathbf{t} = \sum_{k=1}^K \lambda_k \mathbf{t}_k$ , where  $\lambda_k = \mathbf{u}^\top \mathbf{t}_k$ . In other words, the global latent variable sums up the block latent variables by stating that  $\mathbf{t}$  is proportional to the average of the block latent variables  $\mathbf{t}_k$  ( $k = 1, 2, \dots, K$ ), in the former case and, in the latter case, stating that  $\mathbf{t}$  is proportional to the first PLS component of  $\mathbf{u}$  upon  $\mathbf{t}_k$  ( $k = 1, 2, \dots, K$ ).

### 3 Illustrations

The supervised strategies of multiblock data analysis are illustrated and compared based on a simulation study and a real case study.

#### 3.1 Simulation study

This simulation study is, to a large extent, similar to that of Westerhuis et al. [6]. It consists of considering two orthogonal variables  $\mathbf{d}_1$  and  $\mathbf{d}_2$ , four explanatory datasets  $\mathbf{X}_1$  to  $\mathbf{X}_4$  and a response dataset,  $\mathbf{Y}$ . These datasets are defined as follows:  $\mathbf{X}_1 = [\mathbf{d}_1, \mathbf{d}_1, \mathbf{d}_1, \mathbf{d}_1, \mathbf{d}_1]$ ,  $\mathbf{X}_2 = \mathbf{X}_3 = \mathbf{X}_4 = [\mathbf{d}_2, \text{randn}(4)]$  and  $\mathbf{Y} = [\mathbf{d}_1, \mathbf{d}_2]$ , where  $\text{randn}(4)$  stands for four columns of normally distributed random variables. In each dataset, twenty percent of random noise was added to the variables  $\mathbf{d}_1$  and  $\mathbf{d}_2$ . MB-RA, MB-WRA, MB-PLS regression and MB-WCov were performed with the aim of predicting  $\mathbf{Y}$  from  $\mathbf{X}_1$ ,  $\mathbf{X}_2$ ,  $\mathbf{X}_3$  and  $\mathbf{X}_4$ . The simulation study was done based on fifty observations.

Figure 1a depicts the proportion of covariation between the predictive blocks of variables  $\mathbf{X}_1$  to  $\mathbf{X}_4$ , and the response block of variables  $\mathbf{Y}$  explained by the first six latent variables derived from MB-RA, MB-WRA, MB-PLS and MB-WCov. As mentioned above, this figure can be interpreted as a scree diagram to help determining the number of latent variables to be retained. We can see that the curves for all the methods sharply drop between the second and the third components and form a plateau starting from the third component. This suggests to retain the first three components.

Figure 1b shows the proportions of total variance in  $\mathbf{Y}$  recovered by the six successive

components obtained by the various methods of analysis. From this figure, we can draw exactly the same conclusions as those from Figure 1a.

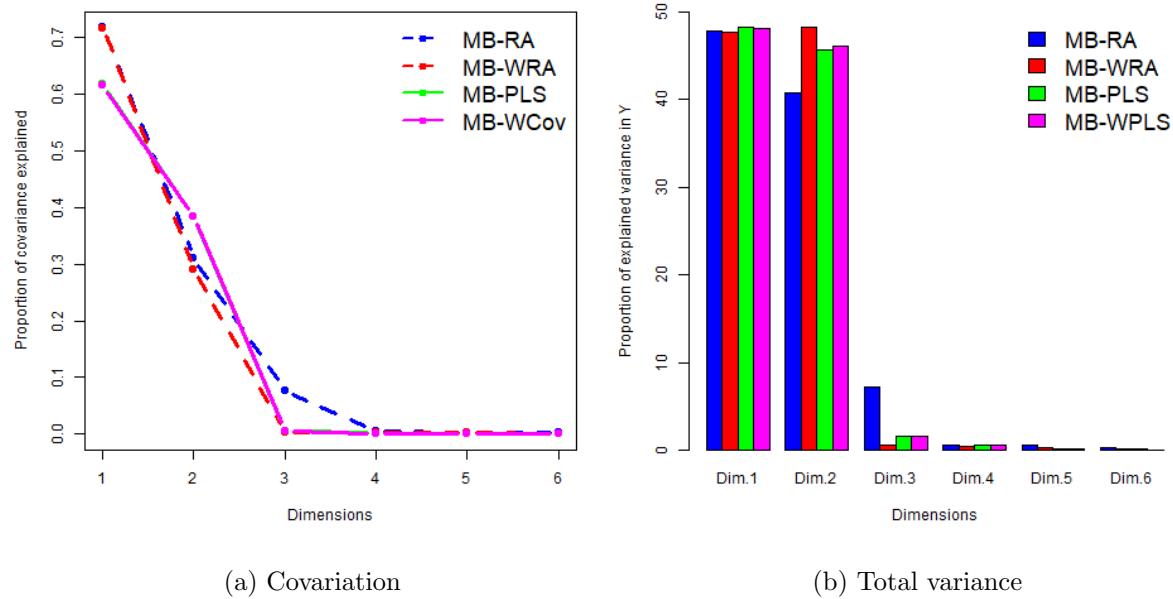


Figure 1: Simulated data: (a) Proportion of covariation between the predictive and response blocks of variables explained by the first six latent variables derived from MB-RA, MB-WRA, MB-PLS and MB-WCov and (b) Proportion of total variance of the response block of variables recovered by the first six latent variables computed by means of MB-RA, MB-WRA, MB-PLS and MB-WCov.

Table 2 gives the correlations between the global latent variables  $\mathbf{t}^{(1)}$  and  $\mathbf{t}^{(2)}$  with the building variables  $\mathbf{d}_1$  and  $\mathbf{d}_2$ . It also gives the contributions of the blocks of variables  $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$  and  $\mathbf{X}_4$  to the determination of the components  $\mathbf{t}^{(1)}$  and  $\mathbf{t}^{(2)}$ .

The first component,  $\mathbf{t}^{(1)}$ , obtained by means of MB-RA and MB-WRA is highly correlated to  $\mathbf{d}_2$  and the second component,  $\mathbf{t}^{(2)}$  is highly correlated with  $\mathbf{d}_1$ . MB-PLS and MB-WCov show an opposite pattern in that sense that their respective first components are highly correlated with  $\mathbf{d}_1$  and their respective second components are highly correlated with  $\mathbf{d}_2$ . This can be explained by the fact that since the two methods pertaining to MB-RA do not take account of the variation (i.e., variances and correlations) within each predictive dataset, it is the variable  $\mathbf{d}_2$  that takes the lead because it appears in  $\mathbf{Y}$ , on the one hand, and  $\mathbf{X}_2, \mathbf{X}_3$  and  $\mathbf{X}_4$ , on the other hand. Therefore, it appears as a common pattern to all the datasets but  $\mathbf{X}_1$ . This is a configuration favored by MB-RA

and MB-WRA. By contrast, the variable  $\mathbf{d}_1$  appears in the dataset  $\mathbf{X}_1$  five times but since MB-RA and MB-WRA do not take account of the within variation, it will be counted as one variable. Contrariwise, this variable will be the leading variable for MB-PLS and MB-WPLS because these methods of analysis take account of the variation within the predictive datasets in addition to taking account of their relationships with  $\mathbf{Y}$ .

As regards the contributions of the datasets  $\mathbf{X}_1$ ,  $\mathbf{X}_2$ ,  $\mathbf{X}_3$  and  $\mathbf{X}_4$  to the determination of the component  $\mathbf{t}^{(1)}$  (Table 2), we can see that, not surprisingly, this component is almost equally determined by  $\mathbf{X}_2$ ,  $\mathbf{X}_3$  and  $\mathbf{X}_4$  insofar as MB-RA and MB-WRA are concerned. With MB-PLS and MB-WCov, the first component is almost determined by the block of variables  $\mathbf{X}_1$ . Similar conclusions regarding the second component  $\mathbf{t}^{(2)}$  can easily be drawn. All these conclusions are in line with the rationale governing the various methods, namely that the two methods pertaining to MB-RA do not take account of the within variation in the blocks of variables, whereas the two methods pertaining to MB-PLS do.

Table 2: Simulated data: Correlations between the global latent variables  $\mathbf{t}^{(1)}$ ,  $\mathbf{t}^{(2)}$  with the building variables  $\mathbf{d}_1$  and  $\mathbf{d}_2$  and contributions of the various blocks of variables to the determination of the global latent variables  $\mathbf{t}^{(1)}$  and  $\mathbf{t}^{(2)}$ .

		Correlations		Contributions			
		$\mathbf{d}_1$	$\mathbf{d}_2$	$\mathbf{X}_1$	$\mathbf{X}_2$	$\mathbf{X}_3$	$\mathbf{X}_4$
MB-RA	$\mathbf{t}^{(1)}$	0.04	<b>-0.99</b>	0.02	<b>0.32</b>	<b>0.33</b>	<b>0.33</b>
	$\mathbf{t}^{(2)}$	<b>-0.90</b>	-0.06	<b>0.78</b>	0.10	0.04	0.08
MB-WRA	$\mathbf{t}^{(1)}$	-0.01	<b>0.99</b>	0.02	<b>0.32</b>	<b>0.33</b>	<b>0.33</b>
	$\mathbf{t}^{(2)}$	<b>0.99</b>	0.01	<b>0.84</b>	0.06	0.02	0.08
MB-PLS	$\mathbf{t}^{(1)}$	<b>-0.99</b>	0.09	<b>0.96</b>	0.01	0.01	0.02
	$\mathbf{t}^{(2)}$	0.09	<b>0.98</b>	0.00	<b>0.34</b>	<b>0.33</b>	<b>0.33</b>
MB-WCov	$\mathbf{t}^{(1)}$	<b>-1.00</b>	0.01	<b>0.97</b>	0.01	0	0.02
	$\mathbf{t}^{(2)}$	-0.01	<b>-0.98</b>	0	<b>0.34</b>	<b>0.33</b>	<b>0.33</b>

In order to assess the prediction ability of the four methods of analysis, we divided the datasets into calibration sets with thirty observations and test sets with twenty observations. The former datasets were used to set up prediction models. These models

were applied to the predictive blocks of variables from the test data. The predictions thus obtained were compared to the actual values from  $\mathbf{Y}$  by means of the root mean squared errors of prediction (RMSEP, [16]). It can be seen in Figure 2, which shows the evolution of the RMSEP values as a function of the number of latent variables introduced in the models, that the four methods show more or less the same pattern. More precisely, the RMSEP sharply decreases when the second latent variable is introduced, then it very slightly decreases with the introduction of the third component. Thereafter, the various curves form a plateau or tend to slightly increase. It is worth noting that the smallest RMSEP value is obtained by means of MB-WRA with three components.

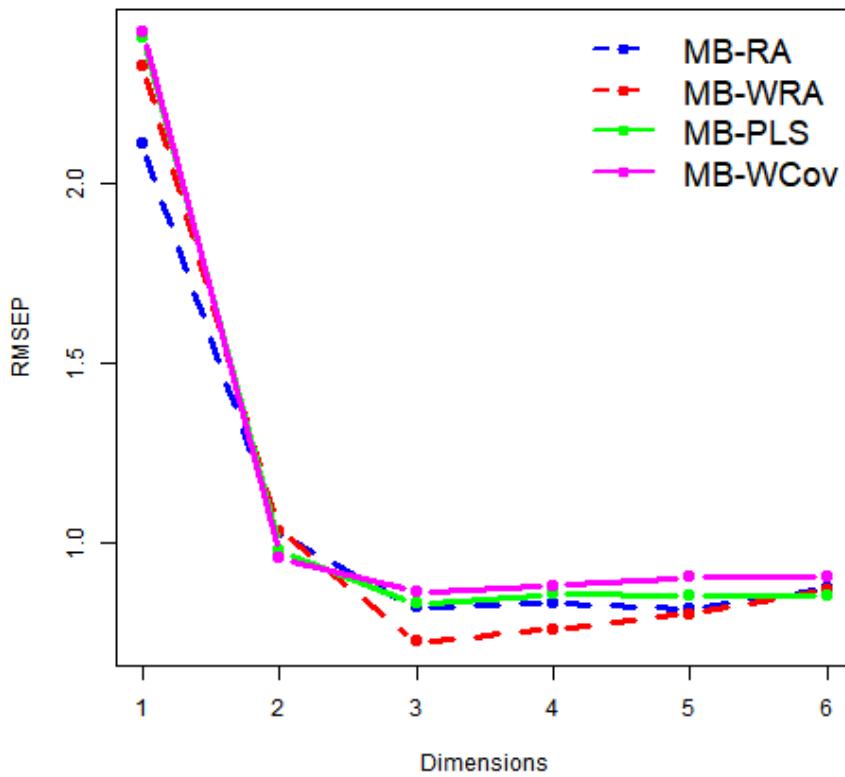


Figure 2: Simulated data: Root mean squared errors of prediction (RMSEP) for MB-RA, MB-WRA, MB-PLS and MB-WCov as a function of the number of latent variables introduced in the models.

### 3.2 Case study: Potatoes data

The multiblock data used to compare the methods of analysis are described in more details in [26]. The aim is to predict sensory attributes from measurement data. Twenty potato samples were analyzed after one month of storage and six additional samples, after eight months of storage. A panel of assessors profiled the texture of cooked potatoes with respect to nine texture attributes. The sensory data were averaged accross assessors, yielding a dataset  $\mathbf{Y}$  (26 potatoes samples  $\times$  9 sensory attributes). The block  $\mathbf{X}_1$  is given by the chemical analysis of the potatoes samples (14 variables) and a second block of variables,  $\mathbf{X}_2$ , concerns the uniaxial compression at six deformation rates (6 variables). Each predictive dataset was column-centered and pre-scaled so as to have its norm equal to 1. Thereafter, they were submitted to MB-RA, MB-WRA, MB-PLS and MB-WCov analyses.

Figure 3a shows the proportion of covariation between the predictive and response blocks of variables explained by the first six latent variables derived from MB-RA, MB-WRA, MB-PLS and MB-WCov. It turns out that, for all these methods, there is no more improvement in the explained covariation when we move from the third latent variable to the fourth latent variable. Therefore, the first three latent variables can be retained.

The proportions of total variance in the response block of variables,  $\mathbf{Y}$ , recovered by the first six latent variables derived from MB-RA, MB-WRA, MB-PLS and MB-WCov are depicted in Figure 3b. These values decrease as we move from the first latent variable to the fourth latent variable. For the first latent variable, MB-RA and its variant MB-WRA are the methods that recover the highest proportion of variance. MB-PLS and MB-WCov show more or less the same behaviour.

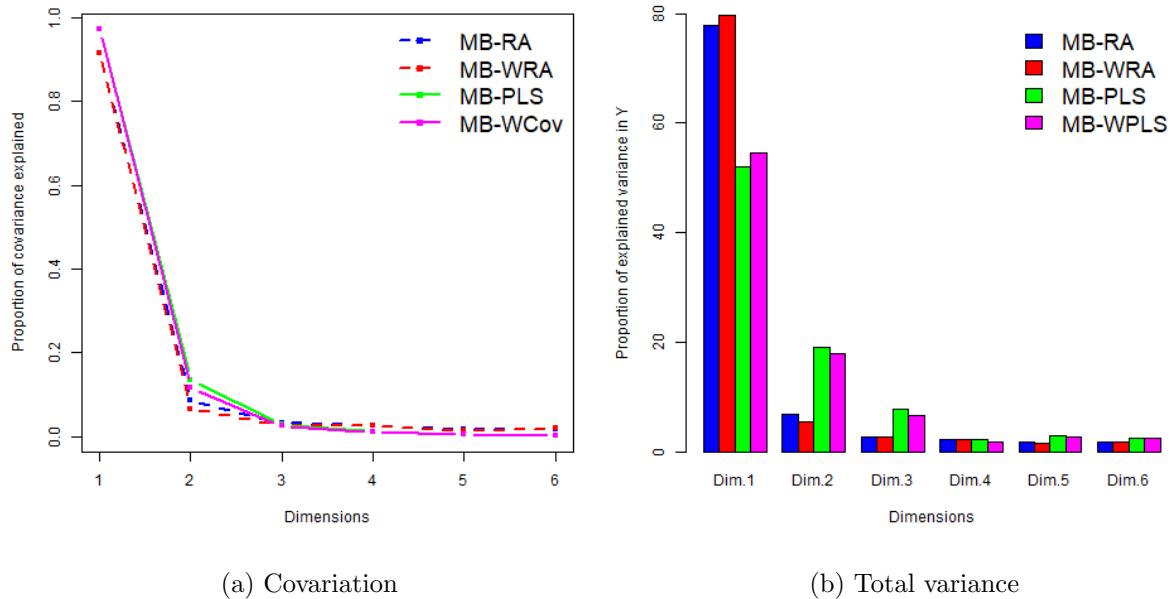


Figure 3: Potatoes data: (a) Proportion of covariation between the predictive and response blocks of variables explained by the first six latent variables derived from MB-RA, MB-WRA, MB-PLS and MB-WCov and (b) Proportion of total variance of the response block of variables recovered by the first six latent variables computed by means of MB-RA, MB-WRA, MB-PLS and MB-WCov.

The contributions of each dataset in determining the global latent variables of MB-RA, MB-WRA, MB-PLS and MB-WCov are presented in Table 3. Globally, it appears that, for all the methods, the chemical dataset ( $\mathbf{X}_1$ ) is the one that contributes most to the determination of the first two latent variables.

Table 3: Potatoes data: Contributions of the blocks of variables  $\mathbf{X}_1$  and  $\mathbf{X}_2$  to the determination of the global latent variables  $\mathbf{t}^{(1)}$  and  $\mathbf{t}^{(2)}$ .

	MB-RA		MB-WRA		MB-PLS		MB-WCov	
	$\mathbf{X}_1$	$\mathbf{X}_2$	$\mathbf{X}_1$	$\mathbf{X}_2$	$\mathbf{X}_1$	$\mathbf{X}_2$	$\mathbf{X}_1$	$\mathbf{X}_2$
$\mathbf{t}^{(1)}$	0.59	0.41	0.59	0.41	0.55	0.45	0.56	0.44
$\mathbf{t}^{(2)}$	0.55	0.45	0.56	0.44	0.67	0.33	0.67	0.33

Figure 4 depicts the root mean squared errors (RMSECV) obtained by a leave one out (LOO) cross validation procedure as a function of the first six global latent variables obtained by the various methods. We can see that the RMSECV values associated with MB-PLS and MB-WCov decrease with the number of global latent variables introduced

in the model and reach a plateau starting from the third component. Moreover, these two methods of analysis seem to have the same prediction performance. The conclusions regarding RMSECV obtained by means of MB-RA and MB-WRA are very different. With a model based on only the first global component from MB-RA or MB-WRA, the RMSECV is smaller than those obtained by means of MB-PLS and MB-WCov with one component. However, as new latent variables are introduced in the models, the performance dramatically deteriorates. These findings are typical of those methods that are vulnerable to the problem of multicollinearity.

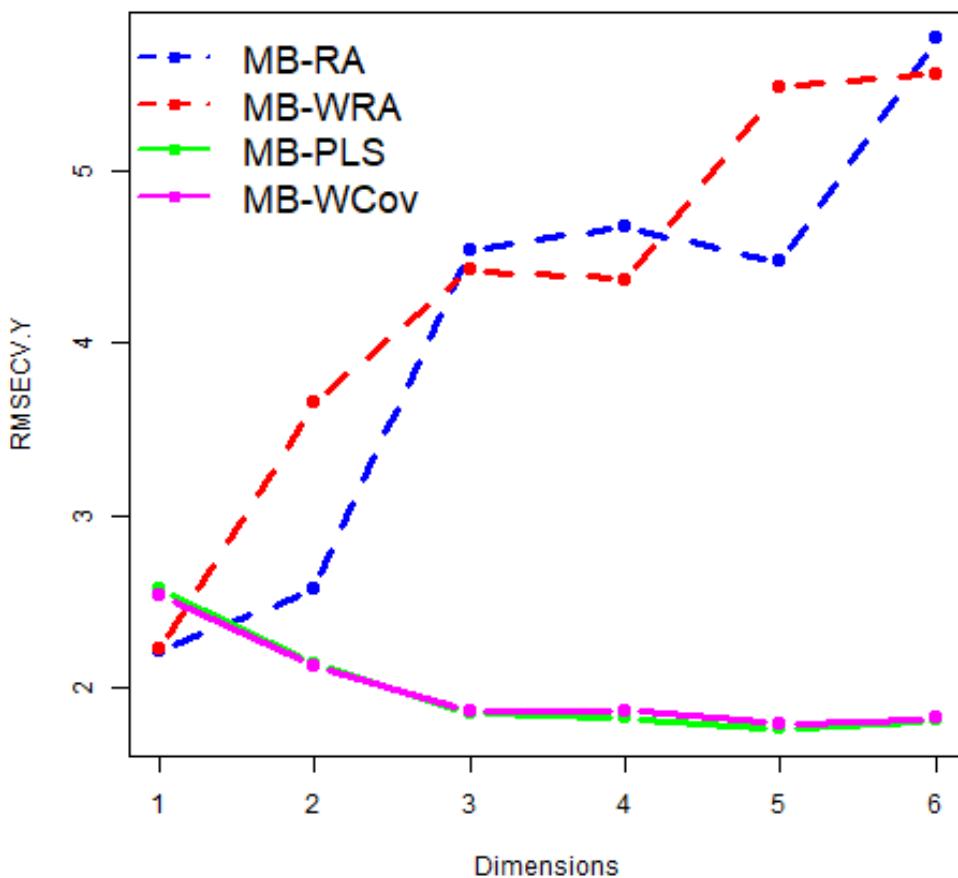


Figure 4: Potatoes data: Root mean squared errors obtained by a LOO cross validation procedure for the first six latent variables derived from MB-RA, MB-WRA, MB-PLS and MB-WCov.

## 4 Discussion and concluding remarks

We have set up a unified approach for two supervised methods, namely, MB-RA and MB-PLS regression. We have also proposed two new strategies for predicting a block of variables from other blocks of variables. These four methods are based on clear optimization criteria and can be differentiated according to two traits.

The first trait concerns the relationship between the  $\mathbf{Y}$ -latent variable  $\mathbf{u}$  and the  $\mathbf{X}_k$ -block latent variables  $\mathbf{t}_k$ . We have considered herein two types of relationships: (i)  $\mathbf{t}_k = \mathbf{X}_k(\mathbf{X}_k^\top \mathbf{X}_k)^{-1}\mathbf{X}_k^\top \mathbf{u}$  and (ii)  $\mathbf{t}_k = \mathbf{X}_k \mathbf{X}_k^\top \mathbf{u}$ . The first relationship leads to the methods pertaining to redundancy analysis, namely, MB-RA and MB-WRA. Since, this strategy of analysis involves the inversion of the matrices  $\mathbf{X}_k^\top \mathbf{X}_k$ , the variances and the correlations of the datasets  $\mathbf{X}_k$  are shaded off. Therefore, the methods of analysis will focus on recovering the variation of  $\mathbf{Y}$  by means of the latent variables of the  $\mathbf{X}_k$  datasets. However, in case of multicollinearity among the variables of  $\mathbf{X}_k$ , the inversion of such matrices is likely to lead to unstable models. By considering the relationship (ii), we circumvent this problem and we are led to methods pertaining to PLS regression, namely, MB-PLS regression and MB-WCov.

The second distinguishing trait is the relationship between the global latent variable  $\mathbf{t}$  and its associated block latent variables  $\mathbf{t}_k$ . Again, we have considered two kinds of relationships: (i) the global latent variable  $\mathbf{t}$  is equal to the sum of its block latent variables  $\mathbf{t}_k$  ( $\mathbf{t} = \sum_{k=1}^K \mathbf{t}_k$ ) and (ii) the global latent variable  $\mathbf{t}$  is equal to the linear combination of its block latent variables  $\mathbf{t}_k$  ( $\mathbf{t} = \sum_{k=1}^K \boldsymbol{\lambda}_k \mathbf{t}_k$ , with  $\boldsymbol{\lambda}_k = \mathbf{u}^\top \mathbf{t}_k$ ). This latter relationships means that  $\mathbf{t}$  is the first PLS component of  $\mathbf{u}$  upon  $\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_K$ . For the relationship (i), we are led to MB-RA and MB-PLS regression. By using the relationship (ii), a specific weight,  $\boldsymbol{\lambda}_k$ , is attached to each dataset  $\mathbf{X}_k$ . These weights reflect the importance of each dataset in the computation of the latent variables  $\mathbf{t}$  and  $\mathbf{u}$ . With the relationship (ii), we obtain the methods MB-WRA and MB-WCov.

We have already noted the relationship  $\mathbf{t} = \sum_{k=1}^K \boldsymbol{\lambda}_k \mathbf{t}_k$ , with  $\mathbf{t}_k = \mathbf{u}^\top \mathbf{t}_k$  means that the latent variable  $\mathbf{t}$  is proportional to the first PLS component of  $\mathbf{u}$  upon the block

components  $\mathbf{t}_k$ . From this standpoint, it appears that MB-WCov bears some similarities to Multiblock Hierarchical PLS (MB-HPLS) [6] which basically enjoys the same property. However, since MB-HPLS is not grounded on a clear optimization criterion, it suffers from convergence problems [6, 27].

The fact that, with MB-WRA and MB-WCov, the global latent variable  $\mathbf{t}$  is the first PLS component of the  $\mathbf{Y}$ -latent variable  $\mathbf{u}$  upon the block latent variables  $\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_K$  suggests new interesting developments. Indeed, instead of the usual PLS regression, we could use a sparse PLS regression. This means that, at each stage, the blocks of variables that do not have a significant contribution to the determination of the latent components computed at the current stage will be discarded. As a consequence, we are led to parsimonious models that are easier to interpret without affecting the prediction ability.

Moving from the projectors  $\mathbf{P}_k = \mathbf{X}_k(\mathbf{X}_k^\top \mathbf{X}_k)^{-1}\mathbf{X}_k^\top$  to the operators  $\mathbf{W}_k = \mathbf{X}_k\mathbf{X}_k^\top$  was dictated by a requirement to circumvent the tricky problem of quasi-collinearity. In effect, this corresponds to a drastic shrinkage of the matrices  $\mathbf{X}_k^\top \mathbf{X}_k$  towards the identity matrix. A softer shrinkage may consist in considering the operators  $\mathbf{P}_{k\gamma} = \mathbf{X}_k [\gamma \mathbf{I} + (1 - \gamma) \mathbf{X}_k^\top \mathbf{X}_k]^{-1} \mathbf{X}_k^\top$  where  $\gamma$  is a tuning parameter comprised between 0 and 1. This yields a continuous strategy of analysis whose two extreme points (i.e.,  $\gamma = 0$  and  $\gamma = 1$ ) are the methods of analysis discussed herein. In practice, the tuning parameter could be determined together with the number of latent variables to be included in the model by a technique of cross-validation [28].

We have also proposed indices for the interpretation of the results of the strategies of analysis. Among these indices, we have an index that indicates the proportion of covariation between the blocks of variables. This index could give a hint regarding the number of latent variables to be retained. We have also proposed an index that highlights the contribution of the block of variables  $\mathbf{X}_k$  in the determination of the global latent variable.

## Acknowledgements

The first author is very grateful to the cooperation and cultural action service of the Embassy of France in Togo and the Center of Excellence in Mathematical Sciences, Computer Science and Applications of the University of Abomey-Calavi (Bénin) for their financial support.

## Appendix

This appendix aims at proving the convergence of the algorithm associated with the maximization of the criterion that defines MB-WRA.

The maximization with respect to  $\nu$  ( $||\nu|| = 1$ ) of the criterion:

$C(\nu) = \sum_{k=1}^K (\nu^\top Y^\top P_k Y \nu)^2$  led us to the following iterative algorithm:

0. Initial  $\nu$ , with  $||\nu|| = 1$ .
1.  $\nu = \frac{\sum_{k=1}^K \lambda_k A_k}{\sum_l \lambda_l^2} \nu$ , where  $A_k = Y^\top P_k Y$  and  $\lambda_k = \nu^\top A_k \nu$ .
2.  $\nu = \nu / ||\nu||$ .
3. Iterate until convergence.

The algorithm generates a series of (positive) real values  $C(n) = \sum_{k=1}^K (\nu_n^\top A_k \nu_n)^2$  where  $\nu_n$  is the vector,  $\nu$ , determined at the  $n^{th}$  iteration.

We prove that this series increases with  $n$ . Since it is upper bounded by  $\sum_{k=1}^K ||Y^\top P_k Y||^2$ , the convergence will be granted.

Let us denote by  $G(n) = \frac{\sum_{k=1}^K \lambda_k(n) A_k}{\sum_{k=1}^K (\lambda_k(n))^2}$  where  $\lambda_k(n) = \nu_n^\top A_k \nu_n$ .

We have  $\nu_{n+1} = \frac{G(n)\nu_n}{||G(n)\nu_n||}$ .

We aim at proving that  $C(n) \leq C(n+1)$ .

We have the following property:

$$\nu_n^\top G(n) \nu_n \leq \nu_{n+1}^\top G(n) \nu_n \quad (26)$$

This property is readily proven by remarking that by virtue of Cauchy-Schwarz inequality, the maximum (with respect to  $\mathbf{x}$ ,  $\|\mathbf{x}\| = 1$ ) of the function  $\mathbf{x}^\top \mathbf{G}(\mathbf{n}) \mathbf{\nu}_n$  is achieved for  $\mathbf{x} = \frac{\mathbf{G}(\mathbf{n}) \mathbf{\nu}_n}{\|\mathbf{G}(\mathbf{n}) \mathbf{\nu}_n\|} = \mathbf{\nu}_{n+1}$ .

By expanding the left term of inequality (26), it is easy to check that it is equal to 1. The right term of inequality (26) is:  $\mathbf{\nu}_{n+1}^\top \mathbf{G}(\mathbf{n}) \mathbf{\nu}_n = \frac{\sum_{k=1}^K \lambda_k(n) \mathbf{\nu}_{n+1}^\top \mathbf{A}_k \mathbf{\nu}_n}{\sum_{l=1}^K (\lambda_l(n))^2}$ . Since the matrices  $\mathbf{A}_k = \mathbf{Y}^\top \mathbf{P}_k \mathbf{Y}$  are semi-definite positive, we have by virtue of Cauchy-Schwarz theorem  $\mathbf{\nu}_{n+1}^\top \mathbf{A}_k \mathbf{\nu}_n \leq \sqrt{\mathbf{\nu}_{n+1}^\top \mathbf{A}_k \mathbf{\nu}_{n+1}} \sqrt{\mathbf{\nu}_n^\top \mathbf{A}_k \mathbf{\nu}_n}$ .

Thus we have:

$$1 \leq \mathbf{\nu}_{n+1}^\top \mathbf{G}(\mathbf{n}) \mathbf{\nu}_n \leq \frac{\sum_{k=1}^K \lambda_k(n) \sqrt{\mathbf{\nu}_{n+1}^\top \mathbf{A}_k \mathbf{\nu}_{n+1}} \sqrt{\mathbf{\nu}_n^\top \mathbf{A}_k \mathbf{\nu}_n}}{\sum_{l=1}^K (\lambda_l(n))^2} \quad (27)$$

Again, by using Cauchy-Schwarz inequality it follows that the last term of the inequality (27) is smaller than:

$$\frac{\sqrt{\sum_{k=1}^K (\lambda_k(n))^2} \sqrt{\sum_{k=1}^K (\mathbf{\nu}_{n+1}^\top \mathbf{A}_k \mathbf{\nu}_{n+1}) (\mathbf{\nu}_n^\top \mathbf{A}_k \mathbf{\nu}_n)}}{\sum_{l=1}^K (\lambda_l(n))^2} = \frac{\sqrt{\sum_{k=1}^K \lambda_k(n) \mathbf{\nu}_{n+1}^\top \mathbf{A}_k \mathbf{\nu}_{n+1}}}{\sqrt{\sum_{l=1}^K (\lambda_l(n))^2}} \quad (28)$$

Using one last time Cauchy-Schwarz inequality, it follows that:

$$\sum_{k=1}^K \lambda_k(n) \mathbf{\nu}_{n+1}^\top \mathbf{A}_k \mathbf{\nu}_{n+1} \leq \sqrt{\sum_{k=1}^K (\lambda_k(n))^2} \sqrt{\sum_{k=1}^K (\mathbf{\nu}_{n+1}^\top \mathbf{A}_k \mathbf{\nu}_{n+1})^2} \quad (29)$$

Combining the inequalities (27), (28) and (29), we have:

$$1 \leq \frac{\left(\sum_{k=1}^K (\lambda_k(n))^2\right)^{1/4} \left(\sum_{k=1}^K (\mathbf{\nu}_{n+1}^\top \mathbf{A}_k \mathbf{\nu}_{n+1})^2\right)^{1/4}}{\sqrt{\sum_{l=1}^K (\lambda_l(n))^2}} = \frac{\left(\sum_{k=1}^K (\mathbf{\nu}_{n+1}^\top \mathbf{A}_k \mathbf{\nu}_{n+1})^2\right)^{1/4}}{\left(\sum_{l=1}^K (\lambda_l(n))^2\right)^{1/4}} \quad (30)$$

By remarking that  $\lambda_k(n) = \mathbf{\nu}_n^\top \mathbf{A}_k \mathbf{\nu}_n$ , it readily follows that:

$$\left(\sum_{l=1}^K (\mathbf{\nu}_n^\top \mathbf{A}_k \mathbf{\nu}_n)^2\right)^{1/4} \leq \left(\sum_{k=1}^K (\mathbf{\nu}_{n+1}^\top \mathbf{A}_k \mathbf{\nu}_{n+1})^2\right)^{1/4} \quad (31)$$

or equivalently:

$$\sum_{k=1}^K (\nu_n^\top A_k \nu_n)^2 \leq \sum_{k=1}^K (\nu_{n+1}^\top A_k \nu_{n+1})^2. \quad (32)$$

This is precisely the property that we aim to prove.

## References

- [1] S. Bougeard, M. Hanafi, E.M. Qannari, ACPVI multibloc, Application à des données d'épidémiologie animale, *Journal de la Société Française de Statistique* 148 (2007) 77-94.
- [2] S. Bougeard, E.M. Qannari, C. Lupo, M. Hanafi, From Multiblock Partial Least Squares to Multiblock Redundancy Analysis. A continuum approach, *Informatica* 22 (2011) 11-26.
- [3] S. Bougeard, E.M. Qannari, C. Lupo, C. Chauvin, Multiblock redundancy analysis from a user's perspective. Application in epidemiology, *Electron. J. App. Stat. Anal.* 4 (2) (2011) 203-214.
- [4] S. Bougeard, E.M. Qannari, N. Rose, Multiblock redundancy analysis: interpretation tools and application in epidemiology, *J. Chemom.* 25 (2011) 467-475.
- [5] L.E. Wangan, B.R. Kowalski, A multiblock partial least squares algorithm for investigating complex chemical systems, *J. Chemom.* 3 (1) (1989) 3-20.
- [6] J.A. Westerhuis, T. Kourti, J.F. MacGregor, Analysis of multiblock and hierarchical PCA and PLS models, *J. Chemom.* 12 (5) (1998) 301-321.
- [7] M. Vivien, Approches PLS linéaires et non linéaires pour la modélisation de multi-tableaux. Théorie et applications, Phd dissertation, Université de Montpellier, France, (2002).

[8] A. El Ghaziri, E.M. Qannari, Measures of association between two datasets; Application to sensory data, *Food Qual. Prefer.* 40 (2015) 116–124. <http://dx.doi.org/10.1016/j.foodqual.2014.09.010>.

[9] E. Tchandao Mangamana, V. Cariou, E. Vigneau, R.L. Glèlè Kakaï, *Journal of Chemometrics and Intelligent Laboratory Systems* 194 (2019) 103856. <https://doi.org/10.1016/j.chemolab.2019.103856>

[10] C.R. Rao, The use and interpretation of principal component analysis in applied research, *Sankhya A*. 26 (1964) 329–358.

[11] R. Sabatier, Analyse factorielle de données structurées et métriques, *Statistique et Analyse des Données* 12 (1987) 75-96.

[12] H. Wold, Estimation of principal components and related models by iterative least squares. In: Krishnaiah, (Ed.), *Multivariate Analysis*, Academic press, New York (1966).

[13] S. Wold, L. Eriksson, J. Trygg, N. Kettaneh, The PLS method-partial least squares projections to latent structures-and its applications in industrial RDP (research, development, and production), Umea University, 2004.

[14] B.S. Dayal, J.F. MacGregor, Improved PLS algorithms, *J. Chemom.* 11 (1) (1997) 73–85.

[15] I. Jolliffe, *Principal Component Analysis*, Springer-Verlag, New York, 2nd edition, 2002.

[16] M. Stone, Cross-validatory choice and assessment of statistical predictions, *J. R. Statist. Soc.* 36 (1974) 111–147.

[17] N.R. Draper, S. Harry, *Applied Regression Analysis*, 3rd ed., Replika Press, 2005.

[18] A. Höskuldsson, PLS regression methods, *J. Chemom.* 2 (1988) 211–228.

[19] Y. Escoufier, Le traitement des variables vectorielles, *Biometrics* 29 (1973) 751–761.

[20] P. Robert, Y. Escoufier, A unifying tool for linear multivariate statistical methods: The RV-coefficient, *Applied Statistics* 25 (1976) 257–265.

[21] P. Schlich. (1996), Defining and validating assessor compromises about product distances and attribute correlations. In T. Ns & E. Risvik (Eds.), *Data Handling in Science and Technology* (vol. 16, pp. 259–306). Elsevier. Sibson, R. (1978). *Studies in the robustness of multidimension*.

[22] P. Faye, D. Brémaud, M. Durand Daubin, P. Courcoux, A. Giboreau, H. Nicod, Perceptive free sorting and verbalization tasks with naive subjects: An alternative to descriptive mappings, *Food Qual. Prefer.* 15 (2004) 781–791.

[23] P. Faye, D. Brémaud, E. Teillet, P. Courcoux, A. Giboreau, H. Nicod, An alternative to external preference mapping based on consumer perceptive mapping, *Food Qual. Prefer.* 17 (2006) 604–614.

[24] T. Worch, Prefmfa, a solution taking the best of both internal and external preference mapping techniques, *Food Qual. Prefer.* 30 (2013) 180–191.

[25] H. Abdi, D. Valentin, S. Chollet, C. Chrea, Analyzing assessors and products in sorting tasks: Distatis, theory and applications, *Food Qual. Prefer.* 18 (2007) 627–640.

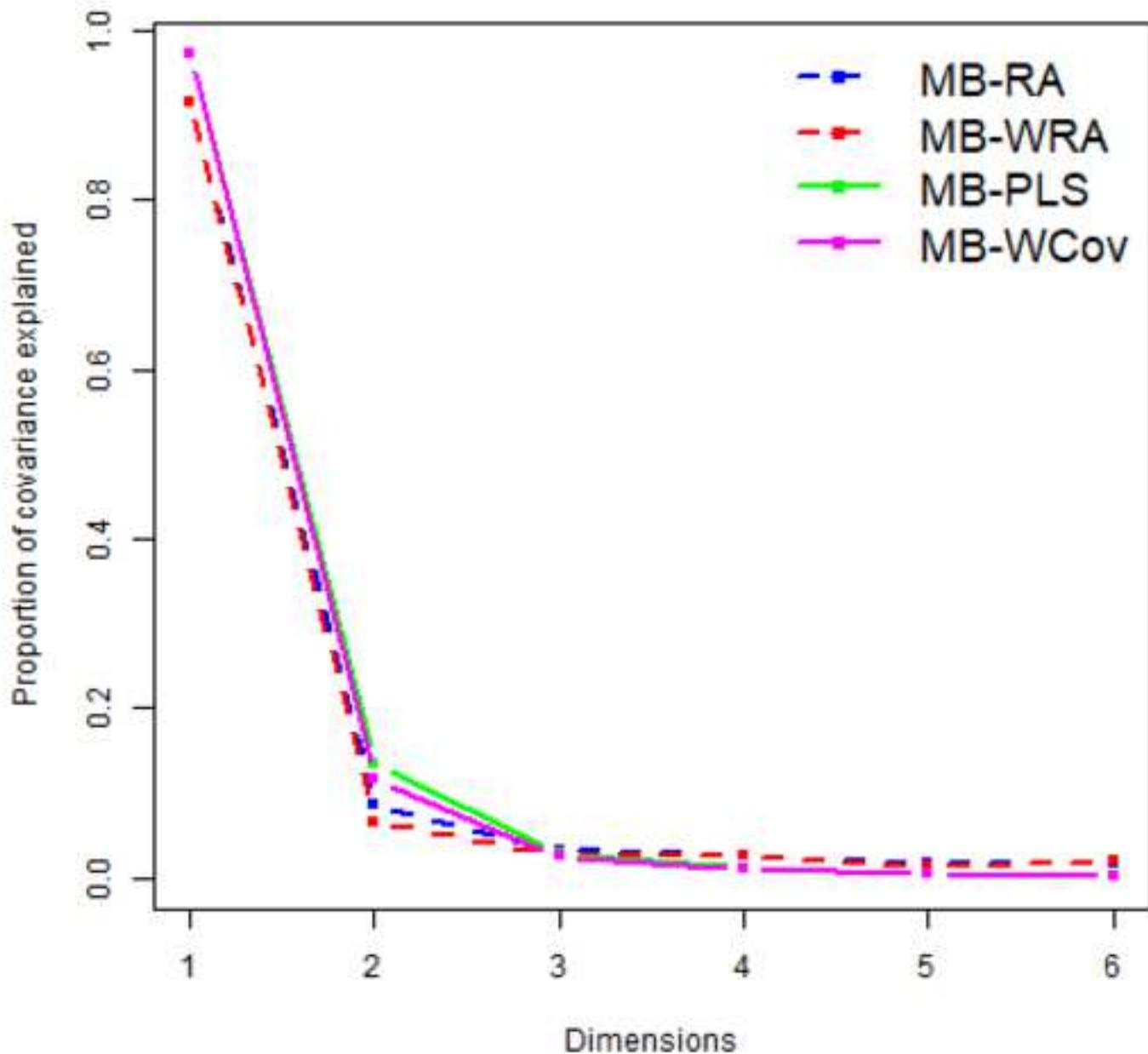
[26] A.K. Thybo, I.E. Bechmann, M. Martens, S.B. Engelsen, Prediction of sensory texture of cooked potatoes using uniaxial compression near infrared spectroscopy and low field  $^1\text{H}$  NMR spectroscopy, *LWT Food Science and Technology* 23 (2) (2000) 103-111.

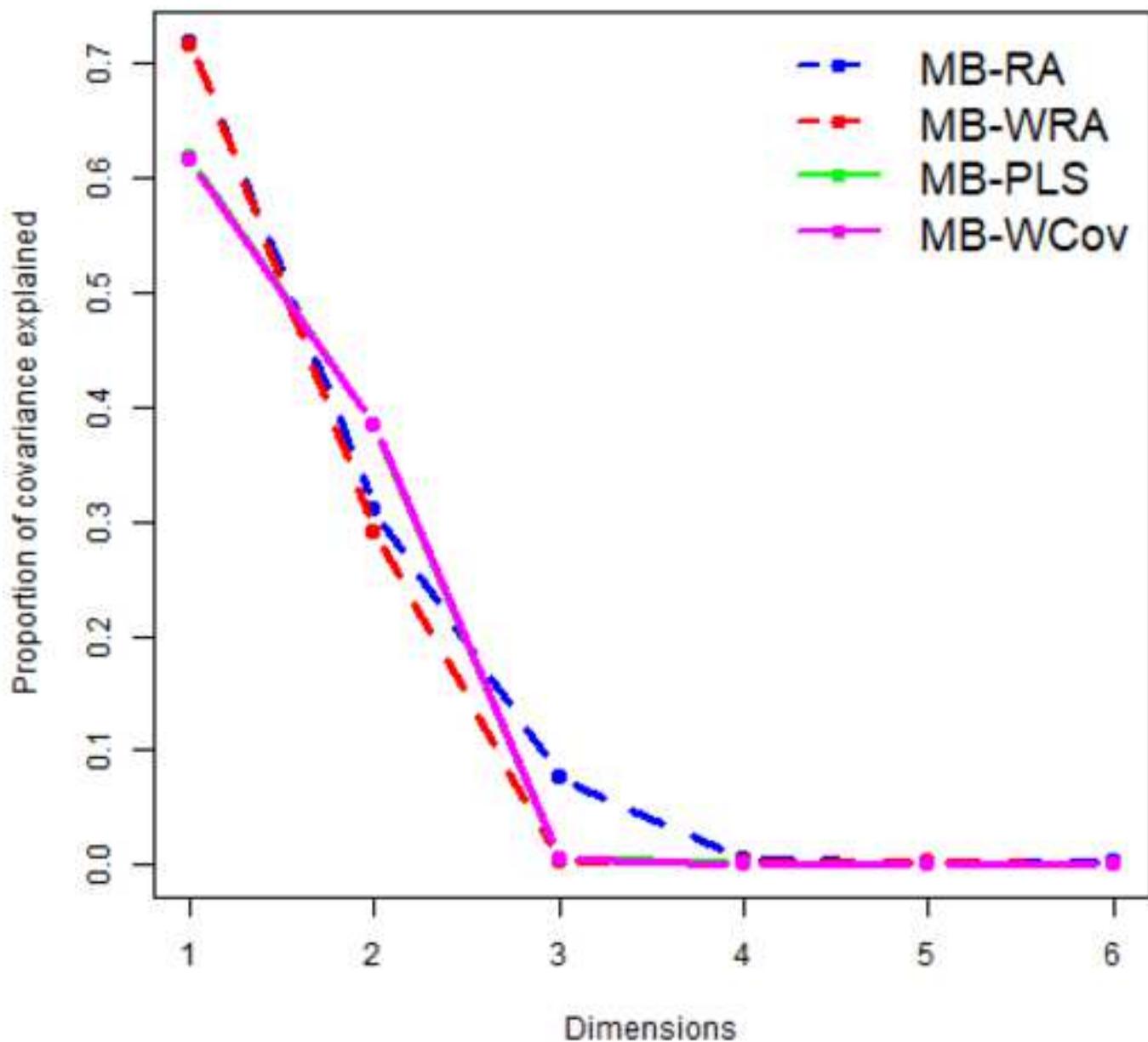
[27] S. Wold, N. Kettaneh, K. Tjesssem, Hierarchical multiblock PLS and PC models for easier interpretation and as an alternative to variable selection, *J. Chemom.* 10 (5-6) (1996) 463-482.

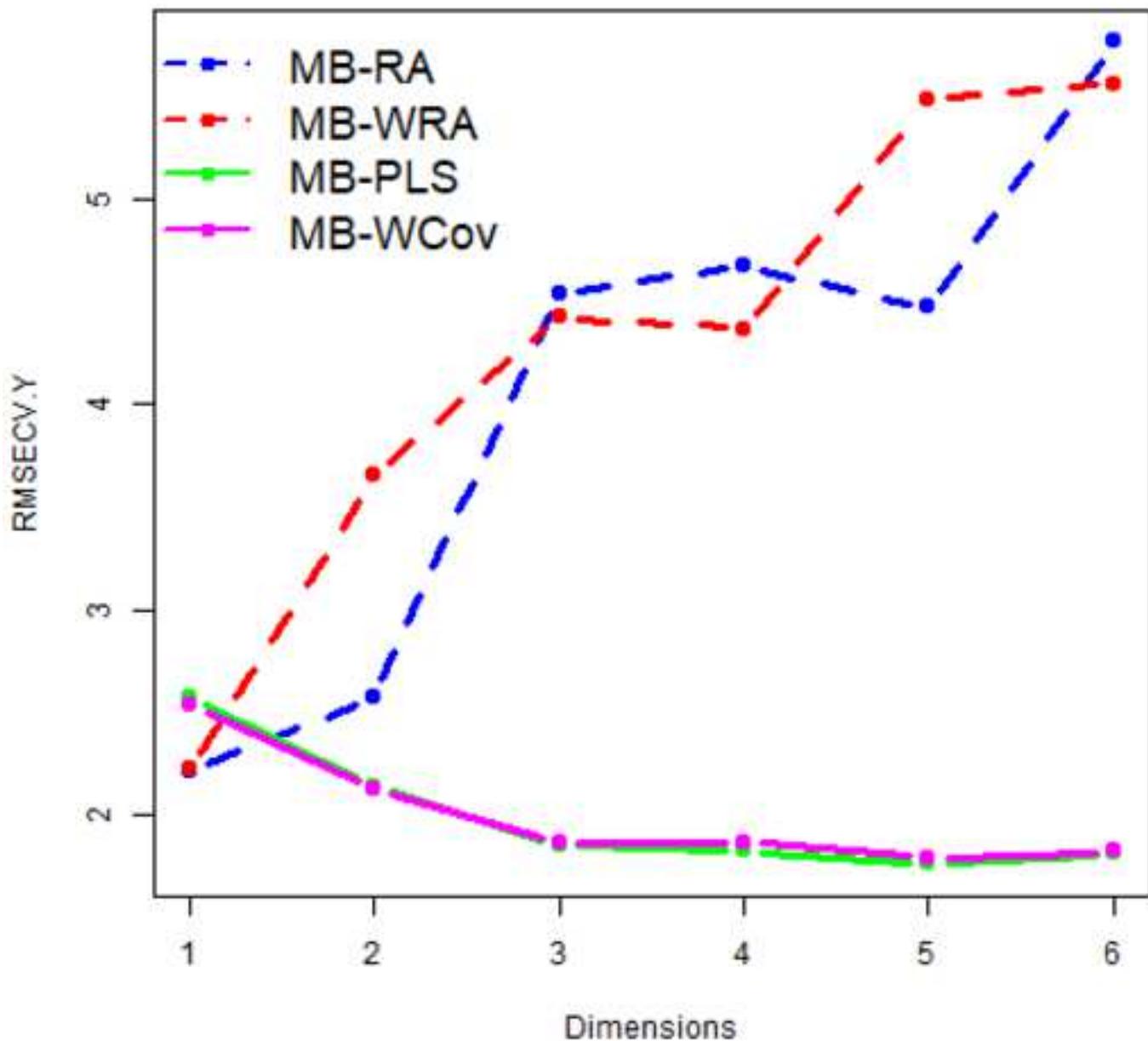
[28] M. Stone and R. J. Brooks, Continuum Regression: Cross-validated Sequentially Constructed Prediction Embracing Ordinary Least Squares, Partial Least Squares and Principal Components Regression, *J. R. Statist. Soc. B* 52 (2) (1990) 237-269.

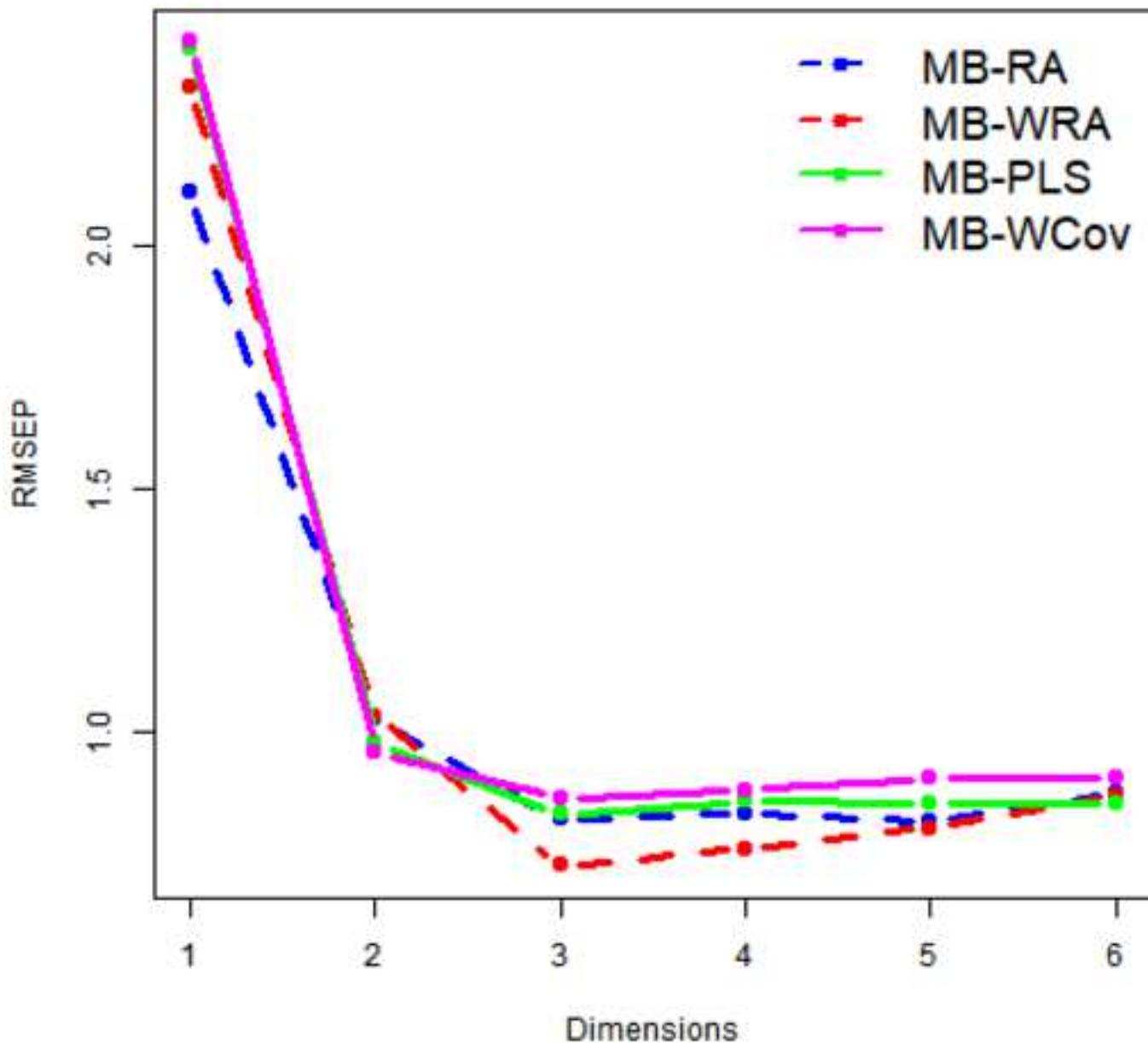


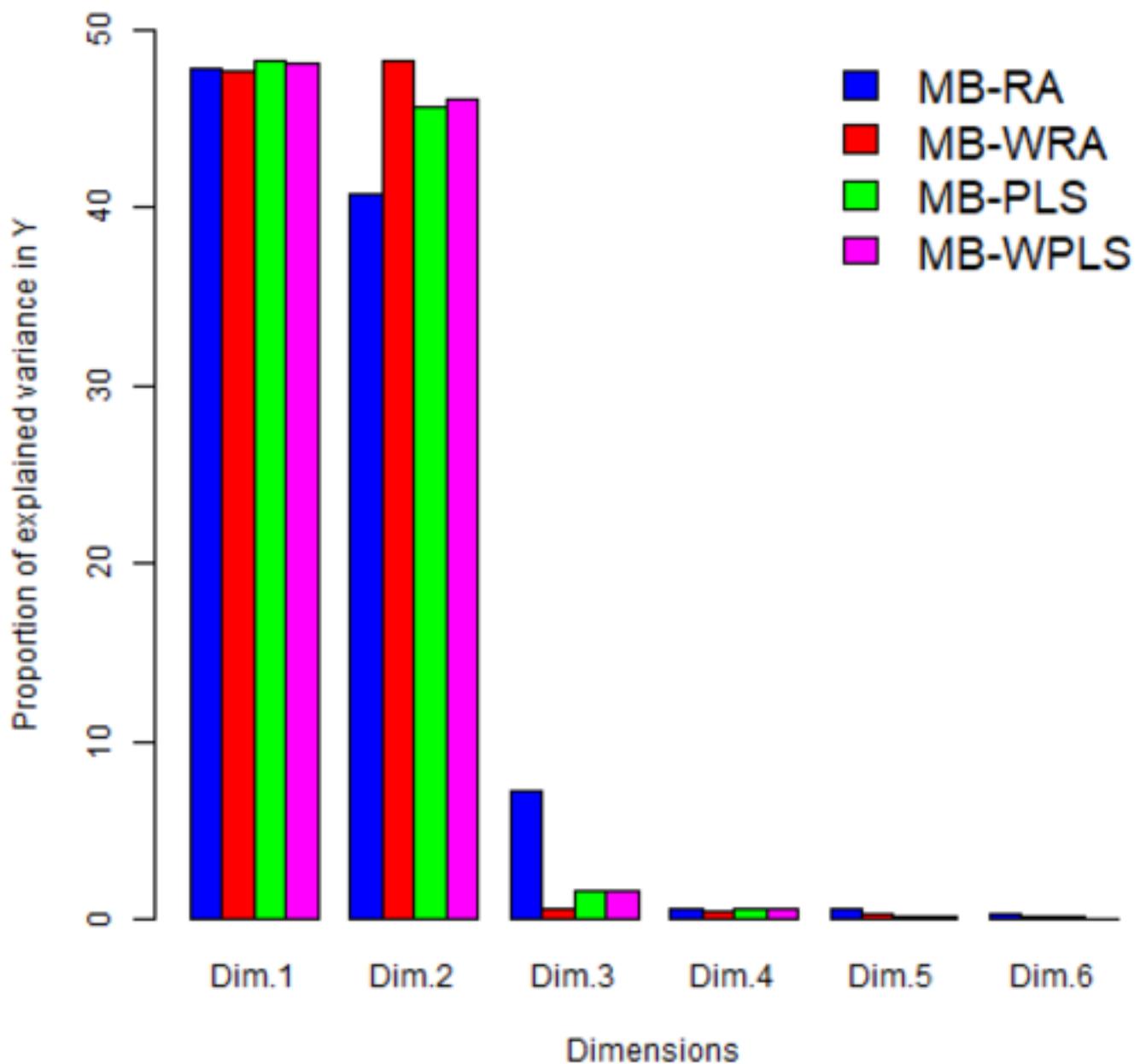
Click here to access/download  
**LaTeX Source File**  
**WPLS\_V0.tex**

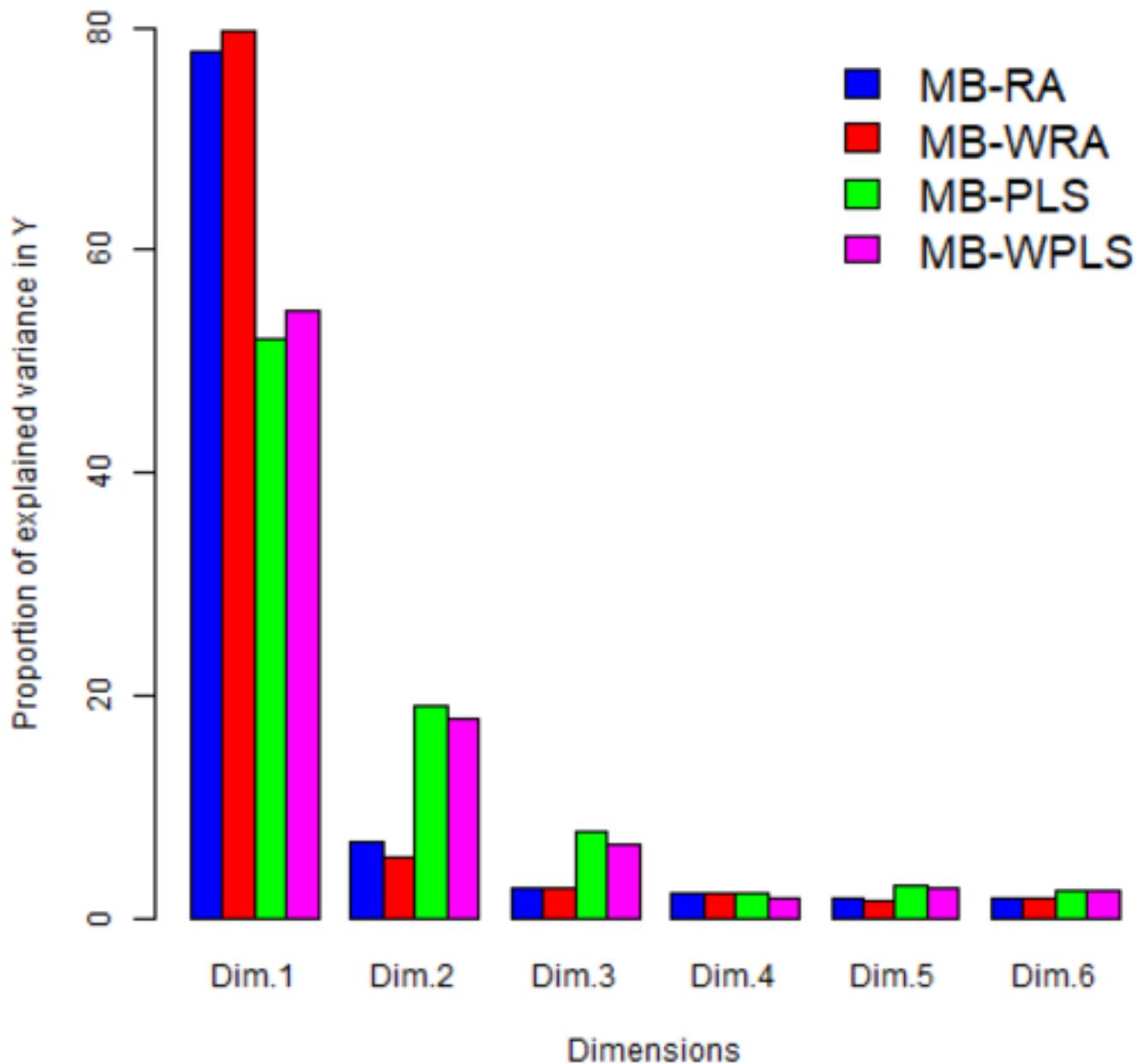








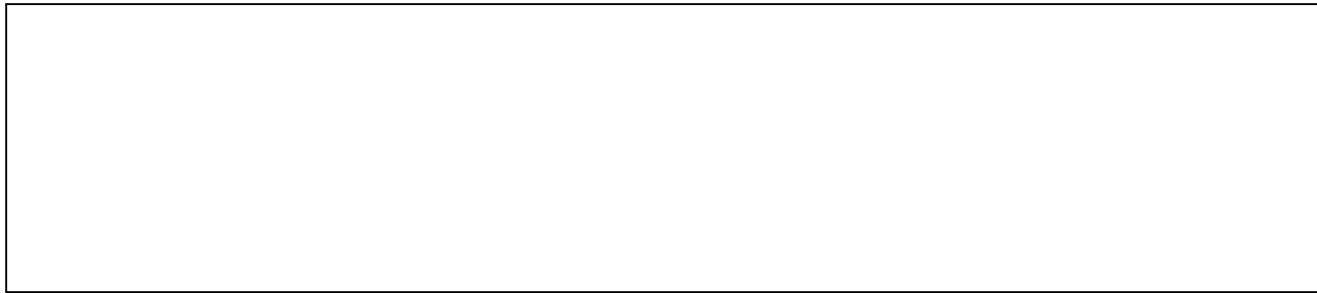




**Declaration of interests**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

A large, empty rectangular box with a thin black border, occupying the lower half of the page. It is intended for authors to declare any potential competing interests.