INFORMATICA, 2011, Vol. 22, No. 1, 11–26

From Multiblock Partial Least Squares to Multiblock Redundancy Analysis. A Continuum Approach

Stéphanie BOUGEARD[1], El Mostafa QANNARI[2], Coralie LUPO[3], Mohamed HANAFI[2]

[1]*French Agency for Food, Environmental, and Occupational Health Safety*
Department of Epidemiology
22440 Ploufragan, France
[2]*Nantes-Atlantic National College of Veterinary Medicine, Food Science and Engineering*
Department of Chemometrics and Sensometrics
Rue de la Géraudière, 44322 Nantes, France
[3]*French Research Institute for Exploitation of the Sea*
Avenue de Mus de loup, 17390 La Tremblade, France
e-mail: stephanie.bougeard@anses.fr, elmostafa.qannari@oniris-nantes.fr, coralie.lupo@ifremer.fr, mohamed.hanafi@oniris-nantes.fr

Received: October 2009; accepted: October 2010

Abstract. For the purpose of exploring and modelling the relationships between a dataset and several datasets, multiblock Partial Least Squares is a widely-used regression technique. It is designed as an extension of PLS which aims at linking two datasets. In the same vein, we propose an extension of Redundancy Analysis to the multiblock setting. We show that PLS and multiblock Redundancy Analysis aim at maximizing the same criterion but the constraints are different. From the solutions of both these approaches, it turns out that they are the two end points of a continuum approach that we propose to investigate.

Keywords: multiblock PLS, multiblock redundancy analysis, continuum approach, Ridge-type regularization, multicolinearity.

1. Introduction

This paper deals with the description and the prediction of multiblock data organized in $(K+1)$ blocks consisting of K explanatory blocks $(X_1, \dots, X_K)$ and a Y dataset to be explained. The first issue is to describe the multiblock tables and sum up the relationships between the variables and between the datasets. For this purpose, we seek overall variables (latent variables) which highlight the relationships between the various datasets. The second issue is to predict Y from the K tables $(X_1, \dots, X_K)$, determine which X_k blocks are best related to the Y variables and within these blocks which variables have an impact on Y.

Multiblock Partial Least Squares (Wold, 1984) is a regression technique that is widely used in the field of chemometrics, sensometrics and process monitoring for the purpose of exploring and modelling the relationships between several datasets to be predicted from several other datasets. Thus, not only a multiblock approach makes it possible to combine several sources of information, but it also highlights the importance of each block in the prediction of the response variables. In the case where only one block of variables is explained by several blocks of explanatory variables, (Westerhuis *et al.*, 1998; Qin *et al.*, 2001; Vivien, 2002) show that the solution obtained from the iterative algorithm of multiblock PLS ($mbPLS$) is equivalent to the solution obtained from a PLS regression of Y and X, where X is the merged dataset, namely $X = [X_1|, \dots, |X_K]$. Redundancy Analysis, RA (Rao, 1964; Van Den Wollenberg, 1977), is yet another popular method for linking two datasets. In a previous paper, we compared the merits of RA and PLS regression (Bougeard *et al.*, 2008). We propose to extend Redundancy Analysis to the multiblock setting and compare this approach to $mbPLS$. Redundancy Analysis, also called Principal Component Analysis with respect to Instrumental Variables, was introduced by Rao (1964) and further investigated by Van der Wollenberg (1997) and Sabatier (1984) among others. These authors gave several formulations of RA which clearly show how this method of analysis can be seen as a regression of Y upon linear combinations of the variables $(x_1, \dots, x_P)$ or as a principal component analysis of the Y variables where components are constrained to be linear combinations of $(x_1, \dots, x_P)$. Similarly to PLS regression, the components thus obtained may be used for an exploratory purpose to investigate the relationships between $(x_1, \dots, x_P)$ and Y or to set up prediction models. This latter approach is called *reduced-rank regression* (Muller, 1981; Davies and Tso, 1982). For the purpose of exploring and modelling the relationships between a dataset Y and several datasets $(X_1, \dots, X_K)$, we propose, in a first stage, a new method called multiblock Redundancy Analysis ($mbRA$), based on the same maximization criterion as $mbPLS$ with different constraints on the components to be determined. In a second stage, we highlight the connection between multiblock Redundancy Analysis and multiblock PLS. It turns out that $mbPLS$ and $mbRA$ are the two end points of a continuum approach that we propose to investigate. As this continuum approach establishes a bridge between $mbPLS$ and $mbRA$, we shall refer to it as “multiblock Continuum Redundancy PLS regression" ($mbCR$). We discuss how the proposed methods are related to other statistical techniques. Finally, the interest of the multiblock methods and the properties of the continuum are illustrated on the basis of a simulation study and on a real dataset in the field of veterinary epidemiology.

2. Methods

2.1. *Notations*

Consider the multiblock setting where we have $(K + 1)$ datasets: a dataset Y to be predicted from K datasets X_k ($k = 1, \dots, K$). The Y table contains Q variables and each table X_k contains p_k variables. The merged dataset X is defined as $[X_1|, \ldots, |X_K]$ and contains $P = \sum_k p_k$ explanatory variables. All these quantitative variables are measured on the same N individuals and supposed to be column centred.

2.2. *Multiblock PLS Regression*

Wold introduced multiblock Partial Least Squares as an alternative procedure based on the Non-linear Iterative PArtial Least Squares (NIPALS) algorithm (Wold, 1984; Struc and Pavesic, 2009). This algorithm was further investigated by Wangen and Kowalski (1988). The initial method aims at linking several tables X_k ($k = 1, \ldots, K$) to one (or more) data table(s) Y. Westerhuis *et al.* (1998), Qin *et al.* (2001), Vivien (2002) showed that the solution obtained from the iterative algorithm of $mbPLS$ is equivalent to the solution obtained from a PLS regression of Y and X where X is the merged dataset. More precisely, Vivien (2002) proved that $mbPLS$ seeks, in a first step, a component $t^{(1)} = Xw^{(1)}$ which is highly related to a component $u^{(1)} = Yv^{(1)}$ and which sums up partial components $t_k^{(1)}$ respectively associated with the blocks X_k. More formally, $mbPLS$ consists in maximizing criterion (1):

$$\begin{aligned} &\operatorname{cov}^2\big(u^{(1)}, t^{(1)}\big) \quad \text{with } t^{(1)} = \sum_{k=1}^{K} a_k^{(1)} t_k^{(1)},\ u^{(1)} = Yv^{(1)},\ t_k^{(1)} = X_k w_k^{(1)}, \\ &\sum_{k=1}^{K} {a_k^{(1)}}^2 = 1, \quad \|w_k^{(1)}\| = \|v^{(1)}\| = 1. \end{aligned} \tag{1}$$

The optimal vector of loadings $w^{(1)}$ is given by the eigenvector of the matrix $(X'YY'X)$ associated with the largest eigenvalue $\lambda_{mbPLS}^{(1)}$ (Westerhuis *et al.*, 1998). The vector $v^{(1)}$ is given by the eigenvector of $M_{mbPLS} = (Y'XX'Y)$ associated with the same eigenvalue. Thereafter the partial vectors of loadings $w_k^{(1)}$ are given by $w_k^{(1)} = w_k^{(1)*}/\|w_k^{(1)*}\|$ where $w_k^{(1)*}$ are the block sub-vectors of $w^{(1)}$, namely $w^{(1)} = [w_1^{(1)*}|, \ldots, |w_K^{(1)*}]'$ (Qin *et al.*, 2001). It is clear that $a_k^{(1)} = \|w_k^{(1)*}\|$ which indeed fulfills the constraint $\sum_k {a_k^{(1)}}^2 = 1$.

Thereafter, the same analysis is performed by replacing $(X_1, \ldots, X_K)$ by their residual in the orthogonal projection onto the subspace spanned by the first global component $t^{(1)}$ (Westerhuis and Smilde, 2001). This process is reiterated in order to determine subsequent components. It is worth noting that Wangen and Kowalski (1988) use a block score deflation, i.e., deflation of each block X_k with respect to its associated partial component t_k. This leads to a slightly different $mbPLS$ strategy of analysis.

2.3. *Proposition of a Multiblock Redundancy Analysis*

For the purpose of exploring and modelling the relationships between two data tables $X = (x_1, \ldots, x_P)$ and Y, it has been shown that Redundancy Analysis and PLS regression are based on the same criterion to maximize, namely $\operatorname{cov}^2(t, u)$ with $t = Xw$

and $u = Yv$, associated with different norm constraints imposed on the components to be determined (Burnham *et al.*, 1996; Bougeard *et al.*, 2008). More precisely, PLS regression imposes the constraints $\|w\| = \|v\| = 1$ whereas Redundancy Analysis imposes the constraints $\|t\| = \|v\| = 1$. In the same vein as $mbPLS$, we propose an extension of Redundancy Analysis to the multiblock setting. This leads us to consider the following maximization problem (2).

$$\begin{aligned}&\operatorname{cov}^2(u^{(1)}, t^{(1)}) \quad \text{with } t^{(1)} = \sum_{k=1}^{K} a_k^{(1)} t_k^{(1)},\ u^{(1)} = Yv^{(1)},\ t_k^{(1)} = X_k w_k^{(1)},\\ &\sum_{k=1}^{K} a_k^{(1)^2} = 1, \quad \|t_k^{(1)}\| = \|v^{(1)}\| = 1.\end{aligned} \tag{2}$$

As previously, the method derives a global component $t^{(1)} = Xw^{(1)}$ oriented towards the explanation of Y, that sums up partial components $t_k^{(1)}$ for $k = (1, \dots, K)$ respectively associated with the blocks X_k. In the case where there is only one block of explanatory variables ($K = 1$), it is clear that multiblock Redundancy Analysis ($mbRA$) amounts to RA.

Replacing the global component $t^{(1)}$ by its expression as a linear combination of the partial components $t_k^{(1)}$, we are led to maximizing the criterion $\operatorname{cov}^2(u^{(1)}, t^{(1)}) = [\sum_k a_k^{(1)} \operatorname{cov}(u^{(1)}, t_k^{(1)})]^2$ under the constraints stated above. The optimal solutions are given by:

$$a_k^{(1)} = \frac{\operatorname{cov}(u^{(1)}, t_k^{(1)})}{\sqrt{\sum_{l=1}^{K} \operatorname{cov}^2(u^{(1)}, t_l^{(1)})}}.$$

Therefore the criterion to be maximized amounts to $\sum_{k=1}^{K} \operatorname{cov}^2(u^{(1)}, t_k^{(1)})$. The maximization problem becomes (3):

$$\begin{aligned}&\sum_{k=1}^{K} \operatorname{cov}^2\big(u^{(1)}, t_k^{(1)}\big) \quad \text{with } t_k^{(1)} = X_k w_k^{(1)},\\ &\quad u^{(1)} = Yv^{(1)},\ \|t_k^{(1)}\| = \|v^{(1)}\| = 1.\end{aligned} \tag{3}$$

The criterion (3) highlights the optimal link between datasets Y and $(X_1, \dots, X_K)$. It follows:

$$\sum_{k=1}^{K} \operatorname{cov}^2\big(u^{(1)}, t_k^{(1)}\big) = \sum_k \big[w_k^{(1)'} X_k' u^{(1)}\big]^2 = \sum_k \big[b_k^{(1)'} (X_k' X_k)^{-1/2} X_k' u^{(1)}\big]^2, \tag{4}$$

where $b_k^{(1)}$ is defined as $b_k^{(1)} = (X_k' X_k)^{1/2} w_k^{(1)}$. In the previous equations, we have dropped the term $1/N$ from the expression of the covariance, for simplicity sake. The constraint $\|t_k^{(1)}\| = 1$ can be expressed as $\|b_k^{(1)}\| = 1$. The maximization of the criterion

(4) leads to $b_k^{(1)} = (X_k'X_k)^{-1/2}X_k'u^{(1)}/\|(X_k'X_k)^{-1/2}X_k'u^{(1)}\|$. Including this expression in criterion (4) and replacing $u^{(1)}$ by $Yv^{(1)}$, we are led to:

$$\sum_{k=1}^{K} \mathrm{cov}^2\big(u^{(1)}, t_k^{(1)}\big) = \sum_k v^{(1)\prime} Y' X_k \big(X_k'X_k\big)^{-1} X_k' Y v^{(1)}. \tag{5}$$

It directly follows that the solution is given by $v^{(1)}$ the normalized eigenvector of the matrix $M_{mbRA} = \sum_k Y'X_k(X_k'X_k)^{-1}X_k'Y$ associated with the largest eigenvalue $\lambda_{mbRA}^{(1)}$. The partial components $(t_1^{(1)}, \dots, t_K^{(1)})$ are therefore given by $t_k^{(1)} = X_k w_k^{(1)} = X_k(X_k'X_k)^{-1/2}b_k^{(1)} = P_{X_k}u^{(1)}/\|P_{X_k}u^{(1)}\|$, where $P_{X_k} = X_k(X_k'X_k)^{-1}X_k'$ is the projector onto the subspace spanned by the X_k variables. The partial components $(t_1^{(1)}, \dots, t_K^{(1)})$ are given by the projection of $u^{(1)}$ on each subspace respectively spanned by $X_1, \dots$ and X_K. The coefficients $a_k^{(1)}$ can also be given by $a_k^{(1)} = \mathrm{cov}(u^{(1)}, t_k^{(1)})/\sqrt{\sum_l \mathrm{cov}^2(u^{(1)}, t_l^{(1)})} = \|P_{X_k}u^{(1)}\|/\sqrt{\sum_l \|P_{X_l}u^{(1)}\|^2}$. These coefficients reflect the link between the Y and the X_k datasets for $k = (1, \dots, K)$. This implies that the global component $t^{(1)} = \sum_k a_k^{(1)} t_k^{(1)} = \sum_k P_{X_k}u^{(1)}/\sqrt{\sum_l \|P_{X_l}u^{(1)}\|^2}$.

We recall that the optimal solution of the maximization of the criterion (2) is based on the eigenvector of the matrix $M_{mbRA} = \sum_k Y'X_k(X_k'X_k)^{-1}X_k'Y$. Because projectors are symmetric and idempotent ($P_{X_k}^2 = P_{X_k}$), it follows that $M_{mbRA} = \sum_k (P_{X_k}Y)'(P_{X_k}Y)$. From this standpoint, $mbRA$ appears as a principal component analysis of the table obtained by the vertical concatenation of the projection of Y onto each subspace spanned by the X_k blocks. Moreover, criterion (5) can also be written as:

$$(5) = v^{(1)\prime} Y' \sum_{k=1}^{K} P_{X_k} Y v^{(1)} = u^{(1)\prime} \sum_k P_{X_k} u^{(1)} = \sum_k \mathrm{var}\big(P_{X_k} u^{(1)}\big). \tag{6}$$

It follows that $mbRA$ consists in maximizing the sum of the variance of the projections of $u^{(1)} = Yv^{(1)}$ onto the subspace spanned by the X_k variables.

As a summing up, the various components in $mbRA$ can be determined as follows:

1. Compute $P_{X_k} = X_k(X_k'X_k)^{-1}X_k'$ and $M_{mbRA} = \sum_k (P_{X_k}Y)'(P_{X_k}Y)$.
2. Compute $v^{(1)}$, the normalized eigenvector of M_{mbRA} associated with the largest eigenvalue and set $u^{(1)} = Yv^{(1)}$.
3. Set $t_k^{(1)} = P_{X_k}u^{(1)}/\|P_{X_k}u^{(1)}\|$.
4. Set $t^{(1)} = \sum_k P_{X_k}u^{(1)}/\sqrt{\sum_k \|P_{X_l}u^{(1)}\|^2}$.

In order to obtain second order solutions, i.e., a global component $t^{(2)}$ and partial components $(t_1^{(2)}, \dots, t_K^{(2)})$ and $u^{(2)}$, we propose to follow the same strategy as $mbPLS$. This consists in deflating the $(X_1, \dots, X_K)$ datasets by projection onto $t^{(1)}$ and considering the residuals. Subsequent components can be found by reiterating this process.

2.4. *Continuum Between Multiblock PLS and Multiblock Redundancy Analysis*

It turns out that $mbRA$ and $mbPLS$ regression are respectively based on the eigenstructure of matrices $M_{mbRA} = \sum_k Y'X_k(X_k'X_k)^{-1}X_k'Y$ and $M_{mbPLS} = Y'XX'Y = \sum_k Y'X_k'X_kY$. Thus, it appears that $mbPLS$ corresponds to a shrinkage of matrices $(X_k'X_k)^{-1}$ towards the identity matrices I_{p_k} for $k = (1, \ldots, K)$. From this standpoint, we can adopt a gradual shrinkage of the matrices $(X_k'X_k)^{-1}$ towards I_{p_k} by considering a convex combination of these matrices (Saudargiene, 1999). More precisely for a scalar γ comprised between 0 and 1, the various components of the continuum approach can be determined as follows:

1. Compute $P_{X_k,\gamma} = X_k'[(1-\gamma)(X_k'X_k) + \gamma I_{p_k}]^{-1}X_k$ and $M_\gamma = \sum_k (P_{X_k,\gamma}Y)'(P_{X_k,\gamma}Y)$.
2. Compute $v_\gamma^{(1)}$ the normalized eigenvector of M_γ associated with the largest eigenvalue $\lambda_\gamma^{(1)}$ and set $u_\gamma^{(1)} = Yv_\gamma^{(1)}$.
3. Set $w_{k,\gamma}^{(1)} = [(1-\gamma)(X_k'X_k)+\gamma I_{p_k}]^{-1}X_k'u_\gamma^{(1)}/\|X_k[(1-\gamma)(X_k'X_k)+\gamma I_{p_k}]^{-1/2}\cdot X_k'u_\gamma^{(1)}\|$ and then the partial components $t_{k,\gamma}^{(1)} = X_kw_{k,\gamma}^{(1)}$.
4. Set the coefficients $a_{k,\gamma}^{(1)} = \mathrm{cov}(u_\gamma^{(1)}, t_{k,\gamma}^{(1)})/\sqrt{\sum_l \mathrm{cov}^2(u_\gamma^{(1)}, t_{l,\gamma}^{(1)})}$ and then set the global component $t_\gamma^{(1)} = \sum_k a_{k,\gamma}^{(1)}t_{k,\gamma}^{(1)}$ or set directly $t_\gamma^{(1)} = \sum_k P_{X_k,\gamma}u_\gamma^{(1)}/\sqrt{\sum_k \|P_{X_l,\gamma}u_\gamma^{(1)}\|^2}$.

It is clear that the case ($\gamma = 0$) corresponds to $mbRA$ applied to the datasets $(Y, X_1, \ldots, X_K)$ whereas the case ($\gamma = 1$) corresponds to $mbPLS$. We shall refer to this strategy of analysis as multiblock Continuum Redundancy PLS regression ($mbCR$). As previously, subsequent components can be obtained by a stagewise procedure by deflating the X_k datasets with respect to the global components obtained in earlier stages.

The introduction of parameter γ is intended to prevent the instability of the prediction models in case of multicolinearity among the variables in X_k. Indeed, the sensitivity to multicolinearity can be reflected by the condition index (Belsley *et al.*, 1980). The condition index η_k of matrix $(X_k'X_k)$ is the ratio of its largest eigenvalue $\lambda_k^{(1)}$ to its smallest eigenvalue $\lambda_k^{(p_k)}$ of matrix $(X_k'X_k)$. A large value of η_k flags the presence of multicolinearity among X_k which is likely to lead to an unstable model. The condition index of each matrix $[(1-\gamma)X_k'X_k + \gamma I_{p_k}]$ is given by:

$$\eta_{k,\gamma} = \frac{[(1-\gamma)\lambda_k^{(1)} + \gamma]}{[(1-\gamma)\lambda_k^{(p_k)} + \gamma]} \quad \text{for } k = (1, \ldots, K).$$

It is easy to prove, by considering its derivative with respect to γ, that each $\eta_{k,\gamma}$ decreases when γ increases. Within $mbCR$, $mbPLS$ corresponds to the smallest values of $\eta_{k,\gamma}$ whereas $mbRA$ corresponds to the largest ones. Thus, parameter γ stands as a regularization parameter as it improves the conditioning of each matrix $(X_k'X_k)$.

2.5. *Prediction of Y from ($X_1, \ldots, X_K$)*

For all the methods previously described, e.g., $mbPLS$, $mbRA$ and $mbCR$, the prediction of the Y variables can be obtained by regressing the Y variables onto the global components $(t^{(1)}, \ldots, t^{(h)})$. These components being orthogonal by construction, the Y table is split up into: $Y = t^{(1)}c^{(1)'} + \cdots + t^{(h)}c^{(h)'} + Y^{(h)}$, $Y^{(h)}$ being the matrix of residuals. Moreover, the global components can be expressed as linear combinations of X: $t^{(1)} = Xw^{*(1)}, \ldots, t^{(h)} = Xw^{*(h)}$. The vectors of loadings w^* and c are defined as in PLS regression. This leads to the model (7):

$$Y = X\big[w^{*(1)}c^{(1)'} + \cdots + w^{*(h)}c^{(h)'}\big] + Y^{(h)}. \tag{7}$$

From a practical point of view, the final model may be obtained by selecting the optimal number h of components to be introduced in the model and the γ parameter, by a validation technique such as cross-validation (Stone, 1974). This consists in splitting the whole dataset into two sets, namely a calibration set and a validation set. The calibration set is used to select the parameters of the model and the root mean square error of calibration ($RMSE_C$) which reflects the fitting ability of the model. The validation set is used to compute the root mean square error of validation ($RMSE_V$) which reflects the prediction ability of the model under consideration.

$$RMSE^{(h)} = \|Y - \hat{Y}^{(h)}\|/\sqrt{Q}, \tag{8}$$

where $\hat{Y}^{(h)}$ is the matrix of predicted values from a model with h components. Thereafter, this procedure is repeated several times. For each number h of components to be introduced in the model, the optimal value of γ is determined by minimizing $RMSE_V$. Among all these models corresponding to the various values of h, a compromised model with a correct fitting ability and a good prediction ability is retained.

2.6. *Alternative Methods*

It is worth mentioning that several methods are proposed in the literature in order to investigate the relationships among datasets. Among these methods, we can distinguish strategies of analysis which fit into the framework of generalized canonical analysis (Horst, 1961; Carroll, 1968). We refer to Kissita (2003) for a review of such methods. Another family of methods pertains to PLS regression and its extensions. We refer to Vivien (2002) for a detailed discussion of these methods. PLS path modelling, PLS–PM (Wold, 1982; Markauskaite, 2001) and more generally structural equation modelling are also worth mentioning in this context. The Generalized Structured Component Analysis, $GSCA$ (Hwang and Takane, 2004), as an alternative method to PLS–PM may also be mentioned. However, this method pertaining to the field of structural equation modelling follows a specific pattern of analysis based on conceptual models which should be set up by the user beforehand. Among all these techniques of analysis, we single out those methods which are based on the same maximization criterion as $mbPLS$ and $mbRA$. Generalized canonical analysis with a reference table, $GCART$ (Kissita, 2003) fits within the framework of generalized canonical analysis, whereas generalized concordance analysis, $CONCORg$ (Lafosse and Ten Berge, 2006) and Orthogonal Multiple Co-Inertia Analysis, $OMCIA$ (Vivien *et al.*, 2005) fit within the framework of PLS regression. A main difference of these methods with $mbPLS$ on the one hand and $mbRA$ on the other hand, lies in the fact that these methods focus on the partial components rather than the global components. This is in particular reflected by the adopted deflation procedure which consists in deflating with respect to the vectors of loadings within each dataset ($CONCORg$) or the partial components ($OMCIA$ and $ACGTR$). Therefore, within each dataset, the vector of loadings or the partial components are orthogonal, but not the global components. We believe that the global components give more insight into the problem under study and give valuable tools both for the prediction and the investigation of the relationships among datasets as will be illustrated in the next section (Westerhuis and Smilde, 2001) .

3. Applications

For the purpose of comparing the performances of the methods, we apply multiblock PLS regression, multiblock Redundancy Analysis and the continuum approach to a simulation study and to a real dataset pertaining to the field of veterinary epidemiology.

3.1. *Simulation Study*

A simulation study is conducted in order to investigate the performance of the three methods under study, e.g., $mbPLS$, $mbRA$ and $mbCR$. A simplified model is specified which involves three datasets X_1, X_2 and Y with two variables per dataset. The conditions considered in this simulation study are the size of multicolinearity among the variables in X_k and the sample size. The multicolinearity within X_1 and X_2 is set to be identical and varied at three levels (low, $\mathrm{Cor} = 0.1$; medium, $\mathrm{Cor} = 0.5$; high, $\mathrm{Cor} = 0.9$). The average correlation between variables in X_k and Y is set to 0.3. Furthermore, five different sample size are considered ($N = 15,\ 25,\ 50,\ 100,\ 200$). At each level of the experimental conditions, i.e., the three levels of multicolinearity times the five levels of the sample sizes, one hundred samples are randomly generated. The methods $mbPLS$, $mbRA$ and $mbCR$ are applied to each sample. The performance of these three methods is evaluated on the basis of a cross-validation procedure, described in Section 2.5, repeated one hundred times. Fitting ability ($RMSE_C$) and prediction ability ($RMSE_V$) are computed using respectively the calibration set and the validation set. As these measures express a lack of fit, the smaller they are, the better the method of analysis is. Moreover, for $mbCR$, the optimal value of the tuning parameter γ is automatically selected in each sample in accordance with the procedure described in Section 2.5. The average value of γ is given for each level of the experimental conditions. Results for all methods under the different conditions, for a model based on the first global component, are given in Table 1.

Table 1

Fitting ability ($R_c = RMSE_C$) and prediction ability ($R_v = RMSE_V$) obtained from multiblock PLS ($mbPLS$), multiblock Redundancy Analysis ($mbRA$) and multiblock Continuum Redundancy PLS regression ($mbCR$) under different simulation conditions. For $mbCR$, the average optimal tuning parameter (γ_{opt}) is also given

	$mbRA$ ($\gamma = 0$)			$mbPLS$ ($\gamma = 1$)			Continuum $mbCR$ (γ_{opt})		
	Cor $= 0.1$	Cor $= 0.5$	Cor $= 0.9$	Cor $= 0.1$	Cor $= 0.5$	Cor $= 0.9$	Cor $= 0.1$	Cor $= 0.5$	Cor $= 0.9$
$N = 15$	$R_c = 0.66$ $R_v = 0.79$	$R_c = 0.72$ $R_v = 0.84$	$R_c = 0.73$ $R_v = 0.88$	$R_c = 0.67$ $R_v = 0.77$	$R_c = 0.71$ $R_v = 0.81$	$R_c = 0.75$ $R_v = 0.83$	$R_c = 0.67$ $R_v = 0.75$ $\gamma_{\mathrm{opt}} = 0.65$	$R_c = 0.71$ $R_v = 0.79$ $\gamma_{\mathrm{opt}} = 0.71$	$R_c = 0.74$ $R_v = 0.81$ $\gamma_{\mathrm{opt}} = 0.61$
$N = 25$	$R_c = 0.71$ $R_v = 0.80$	$R_c = 0.76$ $R_v = 0.86$	$R_c = 0.78$ $R_v = 0.88$	$R_c = 0.71$ $R_v = 0.79$	$R_c = 0.76$ $R_v = 0.83$	$R_c = 0.79$ $R_v = 0.85$	$R_c = 0.71$ $R_v = 0.78$ $\gamma_{\mathrm{opt}} = 0.65$	$R_c = 0.76$ $R_v = 0.83$ $\gamma_{\mathrm{opt}} = 0.74$	$R_c = 0.79$ $R_v = 0.84$ $\gamma_{\mathrm{opt}} = 0.64$
$N = 50$	$R_c = 0.75$ $R_v = 0.81$	$R_c = 0.81$ $R_v = 0.87$	$R_c = 0.83$ $R_v = 0.89$	$R_c = 0.76$ $R_v = 0.81$	$R_c = 0.81$ $R_v = 0.86$	$R_c = 0.83$ $R_v = 0.88$	$R_c = 0.75$ $R_v = 0.80$ $\gamma_{\mathrm{opt}} = 0.67$	$R_c = 0.81$ $R_v = 0.85$ $\gamma_{\mathrm{opt}} = 0.75$	$R_c = 0.83$ $R_v = 0.87$ $\gamma_{\mathrm{opt}} = 0.70$
$N = 100$	$R_c = 0.77$ $R_v = 0.81$	$R_c = 0.83$ $R_v = 0.86$	$R_c = 0.86$ $R_v = 0.90$	$R_c = 0.78$ $R_v = 0.81$	$R_c = 0.83$ $R_v = 0.86$	$R_c = 0.86$ $R_v = 0.89$	$R_c = 0.78$ $R_v = 0.81$ $\gamma_{\mathrm{opt}} = 0.69$	$R_c = 0.83$ $R_v = 0.86$ $\gamma_{\mathrm{opt}} = 0.80$	$R_c = 0.86$ $R_v = 0.89$ $\gamma_{\mathrm{opt}} = 0.71$
$N = 200$	$R_c = 0.79$ $R_v = 0.81$	$R_c = 0.84$ $R_v = 0.87$	$R_c = 0.87$ $R_v = 0.90$	$R_c = 0.79$ $R_v = 0.81$	$R_c = 0.84$ $R_v = 0.86$	$R_c = 0.87$ $R_v = 0.89$	$R_c = 0.79$ $R_v = 0.81$ $\gamma_{\mathrm{opt}} = 0.69$	$R_c = 0.84$ $R_v = 0.86$ $\gamma_{\mathrm{opt}} = 0.76$	$R_c = 0.87$ $R_v = 0.89$ $\gamma_{\mathrm{opt}} = 0.75$

Table 2

Comparison of methods with respect to their fitting and prediction ability under different simulation conditions: proportion of times that the methods, i.e., multiblock Redundancy Analysis ($mbRA$), multiblock PLS ($mbPLS$) or multiblock Continuum Redundancy PLS regression ($mbCR$), outperform each other

Compared methods	N	Cor $= 0.1$		Cor $= 0.5$		Cor $= 0.9$	
		Fitting ab.	Prediction ab.	Fitting ab.	Prediction ab.	Fitting ab.	Prediction ab.
$mbRA > mbPLS$	15	83%	29%	81%	8%	90%	11%
	25	78%	27%	75%	4%	91%	13%
	50	84%	34%	71%	5%	84%	11%
	100	83%	31%	75%	15%	89%	3%
	200	73%	42%	73%	24%	87%	7%
$mbCR > mbPLS$	15	80%	100%	67%	100%	84%	100%
	25	77%	100%	66%	100%	84%	100%
	50	79%	100%	54%	100%	73%	100%
	100	68%	100%	70%	100%	76%	100%
	200	60%	100%	63%	100%	75%	100%
$mbCR > mbRA$	15	22%	100%	20%	100%	9%	100%
	25	27%	100%	20%	100%	8%	100%
	50	20%	100%	29%	100%	15%	100%
	100	16%	100%	27%	100%	9%	100%
	200	34%	100%	26%	100%	13%	100%

When the level of multicolinearity increases, regardless of the sample size and the method, $RMSE_C$ and $RMSE_V$ become also larger. In presence of high multicolinearity within the X_k datasets, the performance of $mbRA$, $mbPLS$ and $mbCR$ decreases. For $mbCR$, the average γ value is higher for a medium level of multicolinearity within X_k (Cor $= 0.5$) than for a high or a low level (Cor $= 0.1$ or Cor $= 0.9$). As expected, on the one hand, multiblock Redundancy Analysis has a better fitting ability than multiblock PLS especially when the level of multicolinearity is low. On the other hand, multiblock PLS has a better prediction ability for medium and high level of multicolinearity. The multiblock Continuum Redundancy PLS regression has a good and comparable fitting ability to $mbRA$. Moreover, whatever the size of the sample size and the level of multicolinearity, $mbCR$ has a better fitting ability than $mbPLS$ and $mbRA$. Results can also be viewed from a more general perspective by computing the number of times that the methods outperform each other (Table 2). It can be seen that, overall, $mbRA$ has a better fitting ability than $mbPLS$ and $mbCR$ and, contrariwise, it is less effective insofar as the prediction is concerned. $mbCR$ outperforms the other two methods in terms of prediction ability.

3.2. *Case Study*

The dataset consists in the measurements of several variables on 404 chicken flocks that were studied during rearing, transport and at slaughterhouse (Lupo *et al.*, 2008).

The Y table to be explained contains two quantitative variables which reflect the official reasons for condemnation at slaughterhouse, i.e., rate of infectious ($INFECT$) or traumatic ($TRAUMA$) origin. The explanatory table is organized in three blocks. Table X_1 contains 26 variables pertaining to the rearing features. Table X_2 contains 11 variables which refer to the transport conditions between farm and slaughterhouse. Table X_3 contains 4 variables pertaining to the slaughtering conditions. The condition index computed for each explanatory table flags the presence of multicolinearity in X_1. Indicator (dummy) variables are considered for the categorical variables. Variables are column centred and scaled to unit variance.

The relationships between (X_1, X_2, X_3) and Y can be investigated using the global components $(t^{(1)}, \ldots, t^{(h)})$. The graphical displays in Fig. 1 depict the loadings associated with the first two components $t^{(1)}$ and $t^{(2)}$ for $mbRA$ ($\gamma = 0$) and $mbPLS$ regression ($\gamma = 1$). It highlights the relationships among the explanatory variables from X_1, X_2 and X_3, and makes it possible to identify some risk factors associated with the condemnation reasons (Y table). The graphical displays associated with $mbRA$ and $mbPLS$ show that the Y variables are strongly related to the first two components. For simplicity sake, we will only interpret the graphical display associated with $mbRA$. The condemnation rate at slaughterhouse for infectious reason ($INFECT$) is in particular associated with the age of the poultry house ($ebanage$ in X_1), the frequency of visits of the farmer to the poultry house during the starting period ($EPassage$ in X_1) and the standard chicken type ($eilotyp5$ in X_1), among others. The condemnation rate at slaughterhouse for traumatic reason ($TRAUMA$) is in particular linked to the presence of an operator at the evisceration line ($ievinbr$ in X_3) and the average lairage time at the slaughterhouse ($dattentemoy$ in X_2), among others. This means that particular care with respect to these variables should be taken in order to reduce the number of carcasses which are condemned at slaughterhouse.

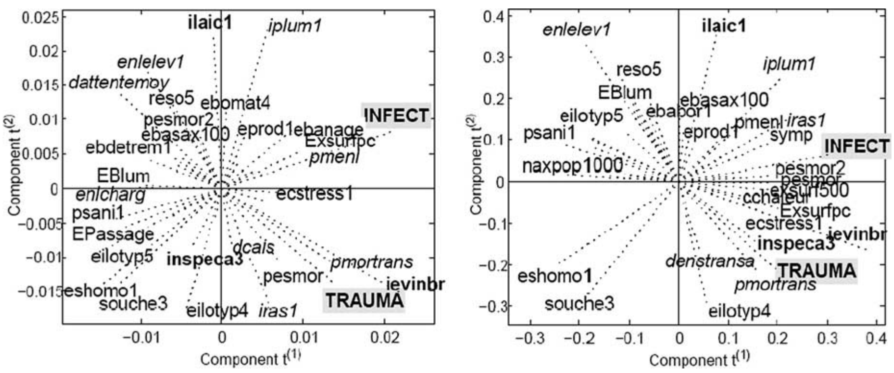

Fig. 1. Plots of the variable loadings associated with the first two components, for $mbRA$ and $mbPLS$. 13 (resp. 14) variables that were not deemed important for the interpretation of the $mbRA$ (resp. $mbPLS$) graphical display were discarded from the plot (although these variables were included in the analysis). Y variables are bold with a grey background, X_1 variables are normal, X_2 variables are slanted and X_3 variables are bold.

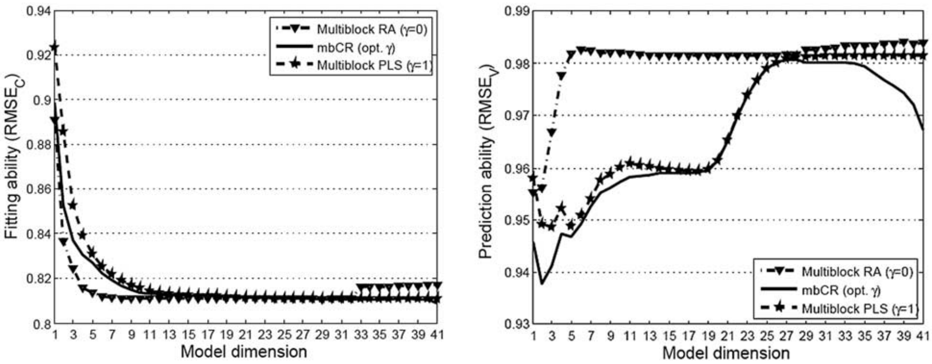

Fig. 2. Fitting and prediction ability as functions of the number of components introduced in the model. Comparison of $mbCR$ (optimal parameter), $mbRA$ ($\gamma = 0$), $mbPLS$ ($\gamma = 1$).

The choice of the optimal model (i.e., optimal number of components and γ parameter) is a compromise achieved by both minimizing the root mean square error of calibration ($RMSE_C$) and validation ($RMSE_V$), which respectively reflect the fitting ability and the prediction ability of the model under consideration. The cross-validation procedure is repeated $(m = 200)$ times by setting one third of the individuals out and by varying the γ value from 0 to 1 with an increment of 0.01. We undertake a comparison of $mbCR$, $mbRA$ and $mbPLS$ on the basis of $RMSE_C$ and $RMSE_V$ criteria. Figure 2 shows $RMSE_C$ and $RMSE_V$ criteria as functions of the number h of components $(t^{(1)}, \ldots, t^{(h)})$ introduced in the model. It can be seen in Fig. 2 that $mbCR$, $mbRA$ and $mbPLS$ have comparable fitting abilities, although $mbRA$ slightly outperforms the other methods. Insofar as the prediction is concerned, the continuum approach $mbCR$ outperforms $mbPLS$ and $mbRA$ especially for the first four dimensions. We can notice that the performances of the methods under study, especially the fitting ability, depend on the number of components introduced in the model. It leads to the usual dilemma between performance (i.e., keep a large number of components) and parsimony (i.e., reduce the number of components). A compromise between a correct fitting and a good prediction ability makes the choice easier and leads to a model with two components for all the methods considered herein. The best method to predict Y from (X_1, X_2, X_3) is obtained by $mbCR$ with two components. For a two-dimensional model, the median value from $(m = 200)$ cross-validations of the optimal γ parameter is 0.98. The median value is given because the optimal parameter distribution is bimodal: 17% of the γ value are comprised between 0 and 0.19 and 83% between 0.77 and 1. This kind of optimal parameter value, close to one of the bound of the continuum, is also found in other continuum methods (Gonzalez *et al.*, 2008; Hwang, 2009).

The importance of each block X_k for explaining Y is reflected by the coefficients $a_k^{(h)^2}$ for a given dimension h. For h components introduced in the model, the importance of the block X_k is given by the average value $\overline{a_k}^{(1-h)^2}$ of the coefficients $(a_k^{(1)^2}, \ldots, a_k^{(h)^2})$ associated with dimensions $(1, \ldots, h)$. Table 3 gives the importance of the rearing features (X_1), the transport conditions (X_2) and the slaughtering conditions (X_3) in the explanation of the official reasons for condemnation at slaughterhouse (Y).

Table 3

Importance of (X_1, X_2, X_3) in the Y explanation for the optimal model with $(h = 2)$ components. Comparison of the block weight of multiblock Redundancy Analysis ($\gamma = 0$), continuum $mbCR$ ($\gamma_{\text{opt.}} = 0.98$) and multiblock PLS ($\gamma = 1$)

	$mbRA$ ($\gamma = 0$)	$mbCR$ ($\gamma_{\text{opt.}} = 0.98$)	$mbPLS$ ($\gamma = 1$)
$\%\overline{a_1}^{(1-2)2}$	51%	51%	48%
$\%\overline{a_2}^{(1-2)2}$	39%	39.5%	44%
$\%\overline{a_3}^{(1-2)2}$	10%	9.5%	8%
Total	100%	100%	100%

As discussed above, a prediction model can be set up by regressing the Y variables on the basis of the first two global components. Table 4 gives a comparison of the regression coefficients obtained by $mbRA$, $mbCR$ and $mbPLS$. We use the results of the $(m = 200)$ cross-validated regression coefficients in order to compute the standard deviations for the various coefficients. Each variable from X is considered to be significantly linked with each variable from Y when the 95% confidence interval associated with the regression coefficient does not contain zero. It turns out that 19 (46%) explanatory variables are interpreted in a same way whatever the method used. For example, the variable $eprod1$ (i.e., presence of other animal productions on the farm, in X_1) is a risk factor both for the infectious and the traumatic reasons for all the methods under study. We can notice that 15 (37%) explanatory variables have a different interpretation when using $mbPLS$ instead of $mbRA$ or $mbCR$. For example, the variable $denstransa$ (i.e., chicken density in crates, in X_2) is highlighted as a risk factor for infectious reason only by $mbPLS$ (positive regression coefficients). The X variables which most influence Y, especially the infectious reason ($INFECT$), are in particular the frequency of visits of the farmer to the poultry house during the starting period ($EPassage$, in X_1), the area of the poultry house ($exsurf500$, in X_1) and factors related to whether the production is made with standard chicken ($eilotyp5$, in X_1) or the presence of the farmer during bird crating ($enlelev1$, in X_2). This means that particular care with respect to these variables should be taken in order to reduce the number of carcasses which are condemned at slaughterhouse.

4. Concluding Remarks

For the purpose of exploring and modelling the relationships between one block of variables Y and several blocks of explanatory variables $(X_1, \ldots, X_K)$, we propose an extension of Redundancy Analysis in order to improve the fitting ability of multiblock PLS regression. As $mbPLS$ and $mbRA$ are based on the same criterion to be maximized associated with different norm constraints, we also investigate a continuum approach, called multiblock Continuum Redundancy PLS regression ($mbCR$). The key feature of this approach is the shrinkage of the variance-covariance matrices $(X_k'X_k)$ towards the identity

Table 4

Comparison of the regression coefficients of $X = [X_1|X_2|X_3]$ on Y using two global components. X variables which have a regression coefficient with an asterisk have a significant link with Y

Block	Variable	$mbRA$ ($\gamma = 0$)		$mbCR$ ($\gamma = 0.98$)		$mbPLS$ ($\gamma = 1$)	
		$INFECT$	$TRAUMA$	$INFECT$	$TRAUMA$	$INFECT$	$TRAUMA$
X_1	ebasax100	0.02	−0.05	0.03	−0.05	0.01	**−0.04***
	cchaleur	0.04	0.01	0.04	0.01	0.04	0.04
	ebabor1	−0.04	**−0.07***	−0.03	**−0.08***	−0.01	**−0.06***
	ebanage	**0.10***	0.01	**0.09***	0.01	0.01	−0.01
	ebchauf1	−0.04	−0.01	−0.05	−0.01	−0.02	−0.02
	ebdetrem1	**−0.07***	**−0.08***	−0.05	**−0.08***	−0.02	**−0.06***
	EBlum	**−0.07***	**−0.05***	**−0.06***	−0.05	−0.02	**−0.06***
	ebomat4	0.03	**−0.07***	0.03	**−0.07***	−0.01	**−0.06 ***
	ecstress1	0.03	0.04	0.04	0.04	**0.05***	**0.05***
	edesins2	**−0.08***	−0.01	**−0.08***	−0.01	−0.03	−0.03
	eilotyp4	−0.08	0.06	−0.03	0.06	0.04	0.04
	eilotyp5	**−0.19***	**0.10***	**−0.14***	**0.10***	0.00	**0.09***
	EPassage	**−0.26***	0.00	**−0.19***	−0.02	**−0.05***	**−0.07***
	eprod1	**−0.13***	**−0.05***	**−0.12***	**−0.06***	**−0.05***	**−0.07***
	ESatpres1	0.07	0.01	0.05	0.01	−0.02	0.01
	eshomo1	0.07	−0.03	**0.07***	−0.04	0.01	−0.02
	estri1	−0.06	−0.02	−0.05	−0.02	0.02	−0.01
	exsurf500	**−0.25***	−0.01	**−0.24***	−0.01	**−0.09***	−0.01
	exsurfpc	0.06	0.04	0.06	0.04	**0.03***	0.02
	frac	−0.02	−0.01	−0.01	0.00	0.02	−0.01
	pesmor2	0.09	0.02	**0.09***	0.02	**0.06***	**0.04***
	pesmor	0.11	0.04	**0.09***	0.04	**0.06***	**0.06***
	psani1	**0.06***	0.02	**0.06***	0.02	**0.03***	0.01
	reso5	−0.08	−0.04	−0.05	−0.03	0.03	0.00
	souche3	0.18	0.00	**0.08***	0.03	**0.06***	0.03
	symp	−0.03	0.07	0.04	0.04	**0.06***	0.03
X_2	enlcais2	**−0.11***	−0.05	**−0.11***	−0.06	**−0.07***	**−0.08***
	enlcharg	0.00	**−0.13***	0.00	**−0.13***	−0.03	**−0.09***
	enlelev1	**−0.18***	−0.02	**−0.17***	0.00	**−0.07***	0.02
	iplum1	**0.07***	0.05	**0.07***	0.05	**0.06***	0.02
	iras1	−0.05	−0.03	−0.05	−0.03	**−0.04***	−0.03
	isolei1	**−0.15***	**−0.08***	**−0.13***	**−0.06***	−0.02	**−0.06***
	pmenl	−0.04	**−0.16***	−0.04	**−0.16***	−0.04	**−0.13***
	pmortrans	**0.14***	−0.12	**0.11***	−0.07	**0.08***	−0.01
	denstransa	−0.04	0.14	0.00	0.08	**0.07***	0.02
	dattentemoy	−0.04	−0.05	−0.04	−0.05	−0.03	**−0.06***
	dcais	**0.09***	0.01	**0.09***	0.00	**0.04***	0.00
X_3	ievinbr	**0.08***	**0.14***	**0.07***	**0.13***	0.04	**0.09***
	ilaic1	−0.02	**0.06***	−0.02	**0.06***	0.01	**0.05***
	inspeca3	−0.08	**−0.13***	**−0.07***	**−0.10***	**−0.05***	**−0.06***
	naxpop1000	−0.03	0.03	−0.04	0.00	**−0.05***	**−0.03***

matrices. This continuum is easy to grasp and implement because the solutions are derived from an eigenanalysis of a matrix. The practical advantage of the $mbCR$ approach lies in the fact that the tuning parameter makes it possible to explore a wide range of methods in order to find an optimal set of coefficients. This approach gives a unified framework so as to deal with potential multicolinearity problems.

The tuning parameter stands as a regularization parameter as it improves the conditioning of each matrix $(X_k' X_k)$. The optimal value of this parameter may be determined through a cross-validation procedure. From the simulation study, we show that multiblock Redundancy Analysis has a better fitting ability than multiblock PLS but has a lower prediction ability for medium and high level of multicolinearity. The continuum approach can be viewed as a ridge-type regularization of multiblock Redundancy Analysis. We show that whatever the size of the sample size and the level of multicolinearity, $mbCR$ has similar or better fitting and prediction ability than $mbPLS$ and $mbRA$. From the case study, we found that $mbCR$ slightly outperforms $mbPLS$ and $mbRA$. To summarize, we can advice to use $mbCR$ in lieu of $mbRA$ or $mbPLS$ when the level of multicolinearity is high. When no multicolinearity occurs, the proposed multiblock Redundancy Analysis could be recommended.

Moreover, further research is needed in order to investigate more deeply the benefits gained from introducing the regularization procedure considering that it entails the cost of introducing a new parameter. Another topic for future research is to investigate the connection between the tuning parameter and the number of components to be introduced in the model. Different tuning parameters $(\gamma_1, \ldots, \gamma_K)$ could also be included in the model, depending of the level of multicolinearity within each datasets $(X_1, \ldots, X_K)$.

References

Belsley, D.A., Kuh, E., Welsch, R.E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. Wiley, New York.

Björkström, A., Sundberg, R. (1996). Continuum regression is not always continuous. *Journal of Royal Statatistical Society B*, 58(4), 703–710.

Bougeard, S., Hanafi, M., Qannari, E.M. (2008). Continuum redundancy PLS regression: a simple continuum approach. Application to epidemiological data. *Computational Statistics and Data Analysis*, 52, 3686–3696.

Burnham, A.J., Viveros, R., Macgregor, J.F. (1996). Framework for latent variable multivariate regression. *Journal of Chemometrics*, 10, 31–45.

Carroll, J.D. (1968). A generalization of canonical correlation analysis to three or more sets of variables. *76th Annual Convention of the American Psychological Association*, 227–228.

Davies, P.T., Tso, M.K.S. (1982). Procedures for reduced-rank regression. *Appl. Stat.*, 31, 244–255.

Gonzalez I., Dejean S., Martin P.G.P., Baccini A. (2008). CCA: An R package to extend canonical correlation analysis. *Jornal of Statistical Software*, 23.

Horst, P. (1961). Relations among m sets of measures. *Psychometrika*, 26, 129–149.

Hwang, H. (2009). Regularized generalized structured component analysis. *Psychometrika*, 74, 517–530.

Hwang, H., Takane, Y. (2004). Generalized structured component analysis. *Psychometrika*, 69, 81–99.

Kissita, G. (2003). *Les analyses canoniques généralisées avec tableau de référence généralisé : éléments théoriques et appliqués*. PhD Thesis, University of Paris Dauphine.

Lafosse, R., Ten Berge, J.M.F. (2006). A simultaneous CONCOR algorithm for the analysis of two partitioned matrices. *Computational Statistics and Data Analysis*, 50, 2529–2535.

Lupo, C., Chauvin, C., Balaine, L., Petetin, I., Péraste, J., Colin, P., Le Bouquin, S. (2008). Post mortem condemnation of processed broiler chickens in Western France. *The Veterinary Record*, 162, 709–713.

Markauskaite, L. (2001). PLSpath modelling of various factors' influence on students' knowledge of informatics. *Informatica*, 12(3), 413–430.

Muller, K.E. (1981). Relationships between redundancy analysis, canonical correlation and multivariate regression. *Psychometrika*, 46, 139–142.

Qin, S.J., Valle, S. and Piovoso, M.J. (2001). On unifying multiblock analysis with application to decentralized process monitoring. *Journal of Chemometrics*, 15, 715–742.

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

Sabatier, R. (1984). Quelques généralisations de l'analyse en composantes principales de variables instrumentales. *Statistique et Analyse de Données*, 9, 75–103.
Saudargiene, A. (1999). Structurization of the covariance matrix by process type and block-diagonal models in the classifier design. *Informatica*, 10, 245–269.
Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of Royal Statistical Society*, 36, 111–147.
Struc, V., Pavesic, N. (2009). Gabor-based kernel partial-least-squares discrimination features for face recognition. *Informatica*, 20, 115–138.
Van Den Wollenberg, A. (1977). Redundancy analysis: an alternative for canonical correlation analysis. *Psychometrika*, 42, 207–219.
Vivien, M. (2002). *Approches PLS Linéaires et Non-linéaires pour la Modélisation de Multi-tableaux: Théorie et Applications*. PhD Thesis, University of Montpellier 1.
Vivien, M., Verron, T., Sabatier, R. (2005). Comparing and predicting sensory profiles by *NIRS*: use of the *GOMCIA* and *GOMCIA-PLS* multi-block methods. *Journal of Chemometrics*, 19, 162–170.
Wangen, L.E. and Kowalski, B.R. (1988). A multiblock partial least squares algorithm for investigating complex chemical systems. *Journal of Chemometrics*, 3, 3–20.
Westerhuis, J.A., Kourti, T. and MacGregor, J.F. (1998). Analysis of multiblock and hierarchical *PCA* and *PLS* model. *Journal of Chemometrics*, 12, 301–321.
Westerhuis J.A. and Smilde A.K. (2001). Deflation in multiblock *PLS* (short communication). *Joyrnal of Chemometrics*, 15, 485–493.
Wold, H. (1982). *Soft Momdelling: The Basic Design and Some Extensions. System Under Indirect Observation*, Part 2. H.K.G Joreskog and Wold (Ed.). North-Holland, Amsterdam.
Wold, S. (1984). Three *PLS* algorithms according to SW. In: *Symposium MULDAST (Multivariate Analysis in Science and Technology)*, Umea University, Sweden. pp. 26–30.

S. Bougeard is researcher in the veterinary epidemiological team of the French agency for food, environmental and occupational health safety (Anses). Her interests are multiblock methods oriented toward a dependent block explanation.

E.M. Qannari is professor at Nantes-Atlantic National College of Veterinary Medicine, Food Science and Engineering (Oniris), Laboratory of Chemometrics and Sensometric. His interests are primarily in multivariate analysis with applications in sensory analysis and chemometrics.

C. Lupo is a veterinary specialized in epidemiology. She participated in collaboration with the French Ministry of Agriculture to the modernization of poultry meat inspection, based on risk analysis.

M. Hanafi is researcher at Nantes-Atlantic National College of Veterinary Medicine, Food Science and Engineering (Oniris), Department of Chemometrics and Sensometric. His interests are primarily in multivariate analysis with applications in sensory analysis and chemometrics.

Nuo daugiablokių mažiausių kvadratų iki daugiablokės pertekliškumo analizės: kontinumo metodas

Bougeard STÉPHANIE, Qannari El MOSTAFA, Lupo CORALIE, Hanafi MOHAMED

Daugiablokis dalinis mažiausių kvadratų (DMK) metodas yra dažnai taikomas regresiniuose uždaviniuose, tiriant ir modeliuojant sąryšius tarp duomenų bazės ir kelių duomenų bazių. Šis metodas yra DMK, susiejančio dvi duomenų bazes, apibendrinimas. Darbe yra pasiūlytas DMK plėtinys daugiablokėje formuluotėje. Parodyta, kad daugiablokis DMK ir daugiablokė pertekliškumo analizė maksimizuoja tą patį kriterijų skirtingais ribojimais. Pasirodo, abu sprendiniai priklauso tai pačiai kontinumo aibei, tiriamai darbe.