Chapter 23

Partial Least Squares Methods: Partial Least Squares Correlation and Partial Least Square Regression

Hervé Abdi and Lynne J. Williams

Abstract

Partial least square (PLS) methods (also sometimes called *projection to latent structures*) relate the information present in two data tables that collect measurements on the same set of observations. PLS methods proceed by deriving latent variables which are (optimal) linear combinations of the variables of a data table. When the goal is to find the shared information between two tables, the approach is equivalent to a correlation problem and the technique is then called *partial least square correlation* (PLSC) (also sometimes called PLS-SVD). In this case there are two sets of latent variables (one set per table), and these latent variables are required to have maximal covariance. When the goal is to *predict* one data table the other one, the technique is then called *partial least square regression*. In this case there is one set of latent variables (derived from the predictor table) and these latent variables are required to give the best possible prediction. In this paper we present and illustrate PLSC and PLSR and show how these descriptive multivariate analysis techniques can be extended to deal with inferential questions by using cross-validation techniques such as the bootstrap and permutation tests.

Key words: Partial least square, Projection to latent structure, PLS correlation, PLS-SVD, PLS-regression, Latent variable, Singular value decomposition, NIPALS method, Tucker inter-battery analysis

1. Introduction

Partial least square (PLS) methods (also sometimes called *projection to latent structures*) relate the information present in two data tables that collect measurements on the same set of observations. These methods were first developed in the late 1960s to the 1980s by the economist Herman Wold (55, 56, 57) but their main early area of development were chemometrics (initiated by Herman's son Svante, (59)) and sensory evaluation (34, 35). The original approach of Herman Wold was to develop a least square algorithm (called NIPALS (56)) for estimating parameters in path analysis

Brad Reisfeld and Arthur N. Mayeno (eds.), *Computational Toxicology: Volume II*, Methods in Molecular Biology, vol. 930,
DOI 10.1007/978-1-62703-059-5_23,

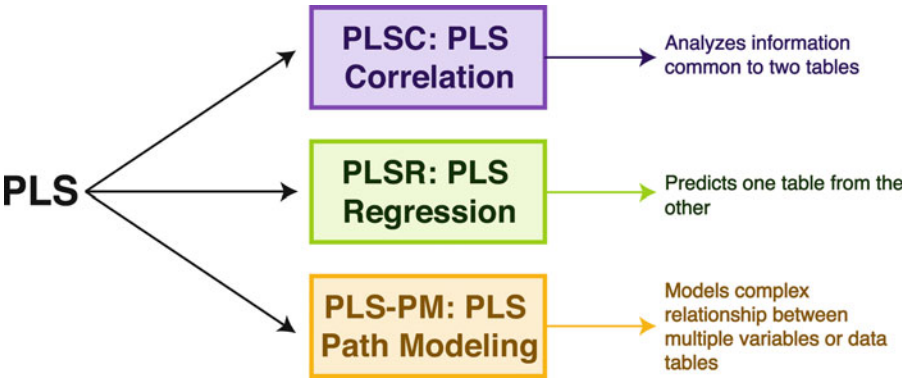

Fig. 1. The PLS family.

models (instead of the maximum likelihood approach used for structural equation modeling such as, e.g., LISREL). This first approach gave rise to partial least square path modeling (PLS-PM) which is still active today (see, e.g., (26, 48)) and can be seen as a least square alternative for structural equation modeling (which uses, in general, a maximum likelihood estimation approach). From a multivariate descriptive analysis point of view, however, most of the early developments of PLS were concerned with defining a latent variable approach to the analysis of two data tables describing one set of observations. Latent variables are new variables obtained as linear combinations of the original variables. When the goal is to find the shared information between these two tables, the approach is equivalent to a correlation problem and the technique is then called partial least square correlation (PLSC) (also sometimes called PLS-SVD (31)). In this case there are two sets of latent variables (one set per table), and these latent variables are required to have maximal covariance. When is goal is to *predict* one data table the other one, the technique is then called partial least square regression (PLSR, see (4, 16, 20, 42)). In this case there is one set of latent variables (derived from the predictor table) and these latent variables are computed to give the best possible prediction. The latent variables and associated parameters are often called *dimension*. So, for example, for PLSC the first set of latent variables is called the first dimension of the analysis.

In this chapter we will present PLSC and PLSR and illustrate them with an example. PLS-methods and their main goals are described in Fig. 1.

2. Notations

Data are stored in matrices which are denoted by upper case bold letters (e.g., **X**). The *identity* matrix is denoted **I**. Column vectors

are denoted by lower case bold letters (e.g., $\mathbf{x}$). Matrix or vector transposition is denoted by an uppercase superscript $\mathbf{T}$ (e.g., $\mathbf{X}^{\mathrm{T}}$). Two bold letters placed next to each other imply matrix or vector multiplication unless otherwise mentioned. The number of rows, columns, or sub-matricesis denoted by an uppercase italic letter (e.g., I) and a given row, column, or sub-matrixis denoted by a lowercase italic letter (e.g., i).

PLS methods analyze the information common to two matrices. The first matrix is an I by J matrix denoted $\mathbf{X}$ whose generic element is $x_{i,j}$ and where the rows are observations and the columns are variables. For PLSR the $\mathbf{X}$ matrix contains the predictor variables (i.e., independent variables). The second matrix is an I by K matrix, denoted $\mathbf{Y}$, whose generic element is $y_{i,k}$. For PLSR, the $\mathbf{Y}$ matrix contains the variables to be predicted (i.e., dependent variables). In general, matrices $\mathbf{X}$ and $\mathbf{Y}$ are statistically preprocessed in order to make the variables comparable. Most of the time, the columns of $\mathbf{X}$ and $\mathbf{Y}$ will be rescaled such that the mean of each column is zero and its norm (i.e., the square root of the sum of its squared elements) is one. When we need to mark the difference between the original data and the preprocessed data, the original data matrices will be denoted $\mathbf{X}$ and $\mathbf{Y}$ and the rescaled data matrices will be denoted $\mathbf{Z}_{\mathbf{X}}$ and $\mathbf{Z}_{\mathbf{Y}}$.

3. The Main Tool: The Singular Value Decomposition

The main analytical tool for PLS is the singular value decomposition (SVD) of a matrix (see (3, 21, 30, 47), for details and tutorials). Recall that the SVD of a given $J \times K$ matrix $\mathbf{Z}$ decomposes it into three matrices as:

$$\mathbf{Z} = \mathbf{U}\Delta\mathbf{V}^{\mathbf{T}} = \sum_{\ell}^{L} \delta_{\ell}\mathbf{u}_{\ell}\mathbf{v}_{\ell}^{\mathbf{T}} \tag{1}$$

where $\mathbf{U}$ is the J by L matrix of the normalized left singular vectors (with L being the rank of $\mathbf{Z}$), $\mathbf{V}$ the K by L matrix of the normalized right singular vectors, $\mathbf{\Delta}$ the L by L diagonal matrix of the L singular values. Also, δ_{ℓ}, $\mathbf{u}_{\ell}$, and $\mathbf{v}_{\ell}$ are,respectively, the ℓth singular value, left, and right singular vectors. Matrices $\mathbf{U}$ and $\mathbf{V}$ are orthonormal matrices (i.e., $\mathbf{U}^{\mathbf{T}}\mathbf{U} = \mathbf{V}^{\mathbf{T}}\mathbf{V} = \mathbf{I}$).

The SVD is closely related to and generalizes the well-known *eigen-decomposition* because $\mathbf{U}$ is also the matrix of the normalized eigenvectors of $\mathbf{Z}\mathbf{Z}^{\mathrm{T}}$, $\mathbf{V}$ is the matrix of the normalized eigenvectors of $\mathbf{Z}^{\mathrm{T}}\mathbf{Z}$, and the singular values are the square root of the eigenvalues of $\mathbf{Z}\mathbf{Z}^{\mathrm{T}}$ and $\mathbf{Z}^{\mathrm{T}}\mathbf{Z}$ (these two matrices have the same eigenvalues). *Key property:* the SVD provides the best reconstitution (in a least squares sense) of the original matrix by a matrix with a lower rank (for more details, see, e.g., (1–3, 47)).

4. Partial Least Squares Correlation

PLSC generalizes the idea of correlation between two variables to two tables. It was originally developed by Tucker (51), and refined by Bookstein (14, 15, 46). This technique is particularly popular in brain imaging because it can handle the very large data sets generated by these techniques and can easily be adapted to handle sophisticated experimental designs (31, 38–41). For PLSC, both tables play a similar role (i.e., both are dependent variables) and the goal is to analyze the information *common* to these two tables. This is obtained by deriving two new sets of variables (one for each table) called latent variables that are obtained as linear combinations of the original variables. These latent variables, which describe the observations, are required to "explain" the largest portion of the *covariance* between the two tables. The original variables are described by their *saliences*.

For each latent variable, the **X** or **Y** variable saliences have a large magnitude, and have large weights for the computation of the latent variable. Therefore, they have contributed a large amount to creating the latent variable and should be used to interpret that latent variable (i.e., the latent variable is mostly "made" from these high contributing variables). By analogy with principal component analysis (see, e.g., (13)), the latent variables are akin to factor scores and the saliences are akin to loadings.

4.1. Correlation Between the Two Tables

Formally, the pattern of relationships between the columns of **X** and **Y** is stored in a $K \times J$ cross-product matrix, denoted **R** (that is usually a correlation matrix in that we compute it with $\mathbf{Z_X}$ and $\mathbf{Z_Y}$ instead of **X** and **Y**). **R** is computed as:

$$\mathbf{R} = \mathbf{Z_Y}^T \mathbf{Z_X}. \tag{2}$$

The SVD (see Eq. 1) of **R** decomposes it into three matrices:

$$\mathbf{R} = \mathbf{U}\Delta\mathbf{V}^{\mathbf{T}}. \tag{3}$$

In the PLSC vocabulary, the singular vectors are called saliences: so **U** is the matrix of **Y**-saliences and **V** is the matrix of **X**-saliences. Because they are singular vectors, the norm of the saliences for a given dimension is equal to one. Some authors (e.g., (31)) prefer to normalize the salience to their singular values (i.e., the delta-normed **Y** saliences will be equal to **U** Δ instead of **U**) because the plots of the salience will be interpretable in the same way as factor scores plots for PCA. We will follow this approach here because it makes the interpretation of the saliences easier.

4.1.1. Common Inertia

The quantity of common information between the two tables can be directly quantified as the *inertia* common to the two tables. This quantity, denoted $\mathcal{I}_{\text{Total}}$, is defined as

$$\mathcal{I}_{Total} = \sum_{\ell}^{L} \delta_{\ell}, \tag{4}$$

where δ_{ℓ} denotes the singular values from Eq. 3 (i.e., δ_l is the ℓth diagonal element of Δ) and L is the number of nonzero singular values of $\mathbf{R}$.

4.2. Latent Variables

The latent variables are obtained by projecting the original matrices onto their respective saliences. So, a latent variable is a linear combination of the original variables and the weights of this linear combination are the saliences. Specifically, we obtain the latent variables for $\mathbf{X}$ as:

$$\mathbf{L}_{\mathbf{X}} = \mathbf{Z}_{\mathbf{X}}\mathbf{V}, \tag{5}$$

and for $\mathbf{Y}$ as:

$$\mathbf{L}_{\mathbf{Y}} = \mathbf{Z}_{\mathbf{Y}}\mathbf{U}. \tag{6}$$

(NB: some authors compute the latent variables with $\mathbf{Y}$ and $\mathbf{X}$ rather than $\mathbf{Z}_{\mathbf{Y}}$ and $\mathbf{Z}_{\mathbf{X}}$; this difference is only a matter of normalization, but using $\mathbf{Z}_{\mathbf{Y}}$ and $\mathbf{Z}_{\mathbf{X}}$ has the advantage of directly relating the latent variables to the maximization criterion used). The latent variables combine the measurements from one table in order to find the common information between the two tables.

4.3. What Does PLSC Optimize?

The goal of PLSC is to find pairs of latent vectors $\mathbf{l}_{\mathbf{X},\,\ell}$ and $\mathbf{l}_{\mathbf{Y},\,\ell}$ with maximal covariance and with the additional constraints that (1) the pairs of latent vectors made from two different indices are uncorrelated and (2) the coefficients used to compute the latent variables are normalized (see (48, 51), for proofs).

Formally, we want to find

$$\mathbf{l}_{\mathbf{X},\ell} = \mathbf{Z}\mathbf{X}\mathbf{v}_{\ell} \quad \textbf{and} \quad \mathbf{l}_{\mathbf{Y},\ell} = \mathbf{Z}_{\mathbf{Y}}\mathbf{u}_{\ell}$$

such that

$$\operatorname{cov}\left(\mathbf{l}_{\mathbf{X},\ell}, \mathbf{l}_{\mathbf{Y},\ell}\right) \propto \mathbf{l}_{\mathbf{X},\ell}^{\mathrm{T}}\mathbf{l}_{\mathbf{Y},\ell} = \max \tag{7}$$

[where $\operatorname{cov}\left(\mathbf{l}_{\mathbf{X},\ell}, \mathbf{l}_{\mathbf{Y},\ell}\right)$ denotes the covariance between $\mathbf{l}_{\mathbf{X},\,\ell}$ and $\mathbf{l}_{\mathbf{Y},\,\ell}$] under the constraints that

$$\mathbf{l}_{\mathbf{X},\ell}^{\mathrm{T}}\mathbf{l}_{\mathbf{Y},\ell'} = \mathbf{0} \textbf{ when } \ell \neq \ell' \tag{8}$$

(note that $\mathbf{l}_{\mathbf{X},\ell}^{T}\mathbf{l}_{\mathbf{X},\ell'}$ and $\mathbf{l}_{\mathbf{Y},\ell}^{T}\mathbf{l}_{\mathbf{Y},\ell'}$ are *not* required to be null) and

$$\mathbf{u}_{\ell}^{\mathrm{T}}\mathbf{u}_{\ell} = \mathbf{v}_{\ell}^{\mathrm{T}}\mathbf{v}_{\ell} = \mathbf{1}. \tag{9}$$

It follows from the properties of the SVD (see, e.g., (13, 21, 30, 47)) that $\mathbf{u}_{\ell}$ and $\mathbf{v}_{\ell}$ are singular vectors of $\mathbf{R}$. In addition, from Eqs. 3, 5,

and 6, the covariance of a pair of latent variables $\mathbf{l}_{\mathbf{X},\ell}$ and $\mathbf{l}_{\mathbf{Y},\ell}$ is equal to the corresponding singular value:

$$\mathbf{l}_{\mathbf{X},\ell}^{\mathrm{T}}\mathbf{l}_{\mathbf{Y},\ell} = \delta_{\ell}. \tag{10}$$

So, when $\ell = 1$, we have the largest possible covariance between the pair of latent variables. When $\ell = 2$ we have the largest possible *covariance* for the latent variables under the constraints that the latent variables are uncorrelated with the first pair of latent variables (as stated in Eq. 8, e.g., $\mathbf{l}_{\mathbf{X},1}$ and $\mathbf{l}_{\mathbf{Y},2}$ are uncorrelated), and so on for larger values of ℓ.

So in brief, for each dimension, PLSC provides two sets of saliences (one for $\mathbf{X}$ one for $\mathbf{Y}$) and two sets of latent variables. The saliences are the weights of the linear combination used to compute the latent variables which are ordered by the amount of covariance they explain. By analogy with principal component analysis, saliences are akin to loadings and latent variables are akin to factor scores (see, e.g., (13)).

4.4. Significance

PLSC is originally a descriptive multivariate technique. As with all these techniques, an additional inferential step is often needed to assess if the results can be considered reliable or "significant." Tucker (51) suggested some possible analytical inferential approaches which were too complex and made too many assumptions to be routinely used. Currently, statistical significance is assessed by computational cross-validation methods. Specifically, the significance of the global model and of the dimensions can be assessed with permutation tests (29); whereas the significance of specific saliences or latent variables can be assessed via the Bootstrap (23).

4.4.1. Permutation Test for Omnibus Tests and Dimensions

The permutation test—originally developed by Student and Fisher (37)—provides a nonparametric estimation of the sampling distribution of the indices computed and allows for null hypothesis testing. For a permutation test, the rows of $\mathbf{X}$ and $\mathbf{Y}$ are randomly permuted (in practice only one of the matrices need to be permuted) so that any relationship between the two matrices is now replaced by a random configuration. The matrix $\mathbf{R}_{\mathrm{perm}}$ is computed from the permuted matrices (this matrix reflects only random associations of the original data because of the permutations) and the analysis of $\mathbf{R}_{\mathrm{perm}}$ is performed: The singular value decomposition of $\mathbf{R}_{\mathrm{perm}}$ is computed. This gives a set of singular values, from which the overall index of effect $\mathcal{I}_{\mathrm{Total}}$ (i.e., the common inertia) is computed. The process is repeated a large number of times (e.g., 10,000 times). Then, the distribution of the overall index and the distribution of the singular values are used to estimate the probability distribution of $\mathcal{I}_{\mathrm{Total}}$ and of the singular values, respectively. If the common inertia computed for the sample is rare enough (e.g., less than 5%) then this index is considered statistically

significant. This test corresponds to an omnibus test (i.e., it tests an *overall* effect) but does not indicate which dimensions are significant. The significant dimensions are obtained from the sampling distribution of the singular values of the same order. Dimensions with a rare singular value (e.g., less than 5%) are considered significant (e.g., the first singular values are considered significant if they are rarer than 5% of the first singular values obtained form the $\mathbf{R}_{\text{perm}}$ matrices). Recall that the singular values are ordered from the largest to the smallest. In general, when a singular value is considered significant all the smaller singular values are considered to be nonsignificant.

4.4.2. What are the Important Variables for a Dimension

The Bootstrap (23, 24) can be used to derive confidence intervals and bootstrap ratios (5, 6, 9 , 40) which are also sometimes "test-values" (32). Confidence intervals give lower and higher values, which together comprise a given proportion (e.g., often 95%) of the values of the saliences. If the zero value is not in the confidence interval of the saliences of a variable, this variable is considered relevant (i.e., "significant"). Bootstrap ratios are computed by dividing the mean of the bootstrapped distribution of a variable by its standard deviation. The bootstrap ratio is akin to a Student t criterion and so if a ratio is large enough (say 2.00 because it roughly corresponds to an $\alpha = .05$ critical value for a t-test) then the variable is considered important for the dimension. The bootstrap estimates a sampling distribution of a statistic by computing multiple instances of this statistic from bootstrapped samples obtained by sampling with replacement from the original sample. For example, in order to evaluate the saliences of $\mathbf{Y}$, the first step is to select with replacement a sample of the rows. This sample is then used to create $\mathbf{Y}_{\text{boot}}$ and $\mathbf{X}_{\text{boot}}$ that are transformed into $\mathbf{Z}_{\mathbf{Y}\text{boot}}$ and $\mathbf{Z}_{\mathbf{X}\text{boot}}$, which are in turn used to compute $\mathbf{R}_{\text{boot}}$ as:

$$\mathbf{R}_{boot} = \mathbf{Z}\mathbf{Y}_{boot}^{T}\mathbf{Z}_{X\,boot}. \tag{11}$$

The Bootstrap values for $\mathbf{Y}$, denoted $\mathbf{U}_{\text{boot}}$, are then computed as

$$\mathbf{U}_{\mathbf{boot}} = \mathbf{R}_{\mathbf{boot}}\mathbf{V}\Delta^{-1}. \tag{12}$$

The values of a large set (e.g., 10,000) are then used to compute confidence intervals and bootstrap ratios.

4.5. PLSC: Example

We will illustrate PLSC with an example in which $I = 36$ wines are described by a matrix $\mathbf{X}$ which contains $J = 5$ objective measurements (price, total acidity, alcohol, sugar, and tannin) and by a matrix $\mathbf{Y}$ which contains $K = 9$ sensory measurements (fruity, floral, vegetal, spicy, woody, sweet, astringent, acidic, hedonic) provided (on a 9 point rating scale) by a panel of trained wine assessors (the ratings given were the median rating for the group of assessors). Table 1 gives the raw data (note that columns two to four, which

Table 1
Physical and chemical descriptions (matrix X) and assessor sensory evaluations (matrix Y) of 36 wines

	Wine descriptors			X: Physical/Chemical description					Y: Assessors' evaluation								
Wine	Varietal	Origin	Color	Price	Total acidity	Alcohol	Sugar	Tannin	Fruity	Floral	Vegetal	Spicy	Woody	Sweet	Astringent	Acidic	Hedonic
1	Merlot	Chile	Red	13	5. 33	13. 8	2. 75	559	6	2	1	4	5	3	5	4	2
2	Cabernet	Chile	Red	9	5. 14	13. 9	2. 41	672	5	3	2	3	4	2	6	3	2
3	Shiraz	Chile	Red	11	5. 16	14. 3	2. 20	455	7	1	2	6	5	3	4	2	2
4	Pinot	Chile	Red	17	4. 37	13. 5	3. 00	348	5	3	2	2	4	1	3	4	4
5	Chardonnay	Chile	White	15	4. 34	13. 3	2. 61	46	5	4	1	3	4	2	1	4	6
6	Sauvignon	Chile	White	11	6. 60	13. 3	3. 17	54	7	5	6	1	1	4	1	5	8
7	Riesling	Chile	White	12	7. 70	12. 3	2. 15	42	6	7	2	2	2	3	1	6	9
8	Gewurztraminer	Chile	White	13	6. 70	12. 5	2. 51	51	5	8	2	1	1	4	1	4	9
9	Malbec	Chile	Rose	9	6. 50	13. 0	7. 24	84	8	4	3	2	2	6	2	3	8
10	Cabernet	Chile	Rose	8	4. 39	12. 0	4. 50	90	6	3	2	1	1	5	2	3	8
11	Pinot	Chile	Rose	10	4. 89	12. 0	6. 37	76	7	2	1	1	1	4	1	4	9
12	Syrah	Chile	Rose	9	5. 90	13. 5	4. 20	80	8	4	1	3	2	5	2	3	7
13	Merlot	Canada	Red	20	7. 42	14. 9	2. 10	483	5	3	2	3	4	3	4	4	3
14	Cabernet	Canada	Red	16	7. 35	14. 5	1. 90	698	6	3	2	2	5	2	5	4	2
15	Shiraz	Canada	Red	20	7. 50	14. 5	1. 50	413	6	2	3	4	3	3	5	1	2
16	Pinot	Canada	Red	23	5. 70	13. 3	1. 70	320	4	2	3	1	3	2	4	4	4
17	Chardonnay	Canada	White	20	6. 00	13. 5	3. 00	35	4	3	2	1	3	2	2	3	5
18	Sauvignon	Canada	White	16	7. 50	12. 0	3. 50	40	8	4	3	2	1	3	1	4	8

19	Riesling	Canada	White	16	7. 00	11. 9	3. 40	48	7	5	1	1	3	3	1	7	8
20	Gewurztraminer	Canada	White	18	6. 30	13. 9	2. 80	39	6	5	2	2	2	3	2	5	6
21	Malbec	Canada	Rose	11	5. 90	12. 0	5. 50	90	6	3	3	3	2	4	2	4	8
22	Cabernet	Canada	Rose	10	5. 60	1. 25	4. 00	85	5	4	1	3	2	4	2	4	7
23	Pinot	Canada	Rose	12	6. 20	13. 0	6. 00	75	5	3	2	1	2	3	2	3	7
24	Syrah	Canada	Rose	12	5. 80	13. 0	3. 50	83	7	3	2	3	3	4	1	4	7
25	Merlot	USA	Red	23	6. 00	13. 6	3. 50	578	7	2	2	5	6	3	4	3	2
26	Cabernet	USA	Red	16	6. 50	14. 6	3. 50	710	8	3	1	4	5	3	5	3	2
27	Shiraz	USA	Red	23	5. 30	13. 9	1. 99	610	8	2	3	7	6	4	5	3	1
28	Pinot	USA	Red	25	6. 10	14. 0	0.00	340	6	3	2	2	5	2	4	4	2
29	Chardonnay	USA	White	16	7. 20	13. 3	1. 10	41	6	4	2	3	6	3	2	4	5
30	Sauvignon	USA	White	11	7. 20	13. 5	1. 00	50	6	5	5	1	2	4	2	4	7
31	Riesling	USA	White	13	8. 60	12. 0	1. 65	47	5	5	3	2	2	4	2	5	8
32	Gewurztraminer	USA	White	20	9. 60	12. 0	0.00	45	6	6	3	2	2	4	2	3	8
33	Malbec	USA	Rose	8	6. 20	12. 5	4. 00	84	8	2	1	4	3	5	2	4	7
34	Cabernet	USA	Rose	9	5. 71	12. 5	4. 30	93	8	3	3	3	2	6	2	3	8
35	Pinot	USA	Rose	11	5. 40	13. 0	3. 10	79	6	1	1	2	3	4	1	3	6
36	Syrah	USA	Rose	10	6. 50	13. 5	3. 00	89	9	3	2	5	4	3	2	3	5

describe the varietal, origin, and color of the wine, are not used in the analysis but can help interpret the results).

4.5.1. Centering and Normalization

Because $\mathbf{X}$ and $\mathbf{Y}$ measure variables with very different scales, each column of these matrices is centered (i.e., its mean is zero) and rescaled so that its norm (i.e., square root of the sum of squares) is equal to one. This gives two new matrices called $\mathbf{Z_X}$ and $\mathbf{Z_Y}$ which are given in Table 2.

The $K = 5$ by $J = 9$ matrix of correlations $\mathbf{R}$ is then computed from $\mathbf{Z_X}$ and $\mathbf{Z_Y}$ as

$$\mathbf{R} = \mathbf{Z}\mathbf{Y}^T\mathbf{Z}\mathbf{X} = \begin{bmatrix} -0.278 & -0.083 & 0.068 & 0.115 & 0.481 & -0.560 & 0.407 & -0.020 & -0.540 \\ 0.029 & 0.531 & 0.348 & -0.168 & -0.162 & 0.084 & -0.098 & 0.202 & 0.202 \\ -0.044 & -0.387 & -0.016 & 0.431 & 0.661 & -0.445 & 0.730 & -0.399 & -0.850 \\ 0.305 & -0.187 & -0.198 & -0.118 & -0.400 & 0.469 & -0.326 & -0.054 & 0.418 \\ 0.008 & -0.479 & -0.132 & 0.525 & 0.713 & -0.408 & 0.936 & -0.336 & -0.884 \end{bmatrix} \tag{13}$$

The $\mathbf{R}$ matrix contains the correlation between each of variable in $\mathbf{X}$ with each of variable in $\mathbf{Y}$.

4.5.2. SVD of R

The SVD (cf., Eqs. 1 and 3) of $\mathbf{R}$ is computed as

$$\mathbf{R} = \mathbf{U}\Delta\mathbf{V}^{\mathbf{T}} = \begin{bmatrix} 0.366 & -0.423 & -0.498 & 0.078 & 0.658 \\ -0.180 & -0.564 & 0.746 & -0.021 & 0.304 \\ 0.584 & 0.112 & 0.206 & -0.777 & -0.005 \\ -0.272 & 0.652 & 0.145 & -0.077 & 0.689 \\ 0.647 & 0.255 & 0.364 & 0.620 & 0.006 \end{bmatrix} \begin{bmatrix} 2.629 \\ 0.881 \\ 0.390 \\ 0.141 \\ 0.077 \end{bmatrix} \times \begin{bmatrix} -0.080 & 0.338 & 0.508 & -0.044 & 0.472 \\ -0.232 & -0.627 & 0.401 & 0.005 & -0.291 \\ -0.030 & -0.442 & 0.373 & -0.399 & 0.173 \\ 0.265 & 0.171 & 0.206 & 0.089 & -0.719 \\ 0.442 & -0.133 & -0.057 & 0.004 & -0.092 \\ -0.332 & 0.388 & 0.435 & 0.084 & -0.265 \\ 0.490 & -0.011 & 0.433 & 0.508 & 0.198 \\ -0.183 & -0.307 & -0.134 & 0.712 & 0.139 \\ -0.539 & 0.076 & -0.043 & 0.243 & -0.088 \end{bmatrix}^T . \tag{14}$$

4.5.3. From Salience to Factor Score

The saliences can be plotted as a PCA-like map (one per table), but here we preferred to plot the delta-normed saliences $\mathbf{F_X}$ and $\mathbf{F_Y}$, which are also called factor scores. These graphs give the same information as the salience plots, but their normalization makes

Table 2
The matrices Z_X and Z_Y (corresponding to X and Y)

	Wine descriptors			Z_X: Centered and normalized version of X: Physical/Chemical description					Z_Y: Centered and normalized version of Y: Assessors' evaluation								
Wine Name	Varietal	Origin	Color	Price	Total acidity	Alcohol	Sugar	Tannin	Fruity	Floral	Vegetal	Spicy	Woody	Sweet	Astringent	Acidic	Hedonic
1	Merlot	Chile	Red	− 0.046	− 0.137	0.120	− 0.030	0.252	− 0.041	− 0.162	− 0.185	0.154	0.211	− 0.062	0.272	0.044	− 0.235
2	Cabernet	Chile	Red	− 0.185	− 0.165	0.140	− 0.066	0.335	− 0.175	− 0.052	− 0.030	0.041	0.101	− 0.212	0.385	− 0.115	− 0.235
3	Shiraz	Chile	Red	− 0.116	− 0.162	0.219	− 0.088	0.176	0.093	− 0.271	− 0.030	0.380	0.211	− 0.062	0.160	− 0.275	− 0.235
4	Pinot	Chile	Red	0.093	− 0.278	0.061	− 0.003	0.098	− 0.175	− 0.052	− 0.030	− 0.072	0.101	− 0.361	0.047	0.044	− 0.105
5	Chardonnay	Chile	White	0.023	− 0.283	0.022	− 0.045	− 0.124	− 0.175	0.058	− 0.185	0.041	0.101	− 0.212	− 0.178	0.044	0.025
6	Sauvignon	Chile	White	− 0.116	0.049	0.022	0.015	− 0.118	0.093	0.168	0.590	− 0.185	− 0.229	0.087	− 0.178	0.204	0.155
7	Riesling	Chile	White	− 0.081	0.210	− 0.175	− 0.093	− 0.127	− 0.041	0.387	− 0.030	− 0.072	− 0.119	− 0.062	− 0.178	0.364	0.220
8	Gewurztraminer	Chile	White	− 0.046	0.064	− 0.136	− 0.055	− 0.120	− 0.175	0.497	− 0.030	− 0.185	− 0.229	0.087	− 0.178	0.044	0.220
9	Malbec	Chile	Rose	− 0.185	0.034	− 0.037	0.444	− 0.096	0.227	0.058	0.125	− 0.072	− 0.119	0.386	− 0.066	− 0.115	0.155
10	Cabernet	Chile	Rose	− 0.220	− 0.275	− 0.234	0.155	− 0.091	− 0.041	− 0.052	− 0.030	− 0.185	− 0.229	0.237	− 0.066	− 0.115	0.155
11	Pinot	Chile	Rose	− 0.150	− 0.202	− 0.234	0.352	− 0.102	0.093	− 0.162	− 0.185	− 0.185	− 0.229	0.087	− 0.178	0.044	0.220
12	Syrah	Chile	Rose	− 0.185	− 0.054	0.061	0.123	− 0.099	0.227	0.058	− 0.185	0.041	− 0.119	0.237	− 0.066	− 0.115	0.090
13	Merlot	Canada	Red	0.197	0.169	0.337	− 0.098	0.197	− 0.175	− 0.052	− 0.030	0.041	0.101	− 0.062	0.160	0.044	− 0.170
14	Cabernet	Canada	Red	0.058	0.159	0.258	− 0.119	0.354	− 0.041	− 0.052	− 0.030	− 0.072	0.211	− 0.212	0.272	0.044	− 0.235
15	Shiraz	Canada	Red	0.197	0.181	0.258	− 0.162	0.145	− 0.041	− 0.162	0.125	0.154	− 0.009	− 0.062	0.272	− 0.435	− 0.235
16	Pinot	Canada	Red	0.301	− 0.083	0.022	− 0.141	0.077	− 0.309	− 0.162	0.125	− 0.185	− 0.009	− 0.212	0.160	0.044	− 0.105
17	Chardonnay	Canada	White	0.197	− 0.039	0.061	− 0.003	− 0.132	− 0.309	− 0.052	− 0.030	− 0.185	− 0.009	− 0.212	− 0.066	− 0.115	− 0.040
18	Sauvignon	Canada	White	0.058	0.181	− 0.234	0.049	− 0.128	0.227	0.058	0.125	− 0.072	− 0.229	− 0.062	− 0.178	0.044	0.155
19	Riesling	Canada	White	0.058	0.108	− 0.254	0.039	− 0.122	0.093	0.168	− 0.185	− 0.185	− 0.009	− 0.062	− 0.178	0.523	0.155
20	Gewurztraminer	Canada	White	0.127	0.005	0.140	− 0.024	− 0.129	− 0.041	0.168	− 0.030	− 0.072	− 0.119	− 0.062	− 0.066	0.204	0.025

(continued)

Table 2 (continued)

	Wine descriptors			Z_X: Centered and normalized version of X: Physical/Chemical description					Z_Y: Centered and normalized version of Y: Assessors' evaluation								
Wine Name	Varietal	Origin	Color	Price	Total acidity	Alcohol	Sugar	Tannin	Fruity	Floral	Vegetal	Spicy	Woody	Sweet	Astringent	Acidic	Hedonic
21	Malbec	Canada	Rose	− 0.116	− 0.054	− 0.234	0.261	− 0.091	− 0.041	− 0.052	0.125	0.041	− 0.119	0.087	− 0.066	0.044	0.155
22	Cabernet	Canada	Rose	− 0.150	− 0.098	− 0.136	0.102	− 0.095	− 0.175	0.058	− 0.185	0.041	− 0.119	0.087	− 0.066	0.044	0.090
23	Pinot	Canada	Rose	− 0.081	− 0.010	− 0.037	0.313	− 0.102	− 0.175	− 0.052	− 0.030	− 0.185	− 0.119	− 0.062	− 0.066	− 0.115	0.090
24	Syrah	Canada	Rose	− 0.081	− 0.068	− 0.037	0.049	− 0.097	0.093	− 0.052	− 0.030	0.041	− 0.009	0.087	− 0.178	0.044	0.090
25	Merlot	USA	Red	0.301	− 0.039	0.081	0.049	0.266	0.093	− 0.162	− 0.030	0.267	0.321	− 0.062	0.160	− 0.115	− 0.235
26	Cabernet	USA	Red	0.058	0.034	0.278	0.049	0.363	0.227	− 0.052	− 0.185	0.154	0.211	− 0.062	0.272	− 0.115	− 0.235
27	Shiraz	USA	Red	0.301	− 0.142	0.140	− 0.110	0.290	0.227	− 0.162	0.125	0.493	0.321	0.087	0.272	− 0.115	− 0.300
28	Pinot	USA	Red	0.370	− 0.024	0.160	− 0.320	0.092	− 0.041	− 0.052	− 0.030	− 0.072	0.211	− 0.212	0.160	0.044	− 0.235
29	Chardonnay	USA	White	0.058	0.137	0.022	− 0.204	− 0.127	− 0.041	0.058	− 0.030	0.041	0.321	− 0.062	− 0.066	0.044	− 0.040
30	Sauvignon	USA	White	− 0.116	0.137	0.061	− 0.214	− 0.121	− 0.041	0.168	0.435	− 0.185	− 0.119	0.087	− 0.066	0.044	0.090
31	Riesling	USA	White	− 0.046	0.342	− 0.234	− 0.146	− 0.123	− 0.175	0.168	0.125	− 0.072	− 0.119	0.087	− 0.066	0.204	0.155
32	Gewurztraminer	USA	White	0.197	0.489	− 0.234	− 0.320	− 0.124	− 0.041	0.278	0.125	− 0.072	− 0.119	0.087	− 0.066	− 0.115	0.155
33	Malbec	USA	Rose	− 0.220	− 0.010	− 0.136	0.102	− 0.096	0.227	− 0.162	− 0.185	0.154	− 0.009	0.237	− 0.066	0.044	0.090
34	Cabernet	USA	Rose	− 0.185	− 0.082	− 0.136	0.134	− 0.089	0.227	− 0.052	0.125	0.041	− 0.119	0.386	− 0.066	− 0.115	0.155
35	Pinot	USA	Rose	− 0.116	− 0.127	− 0.037	0.007	− 0.100	− 0.041	− 0.271	− 0.185	− 0.072	− 0.009	0.087	− 0.178	− 0.115	0.025
36	Syrah	USA	Rose	− 0.150	0.034	0.061	− 0.003	− 0.092	0.361	− 0.052	− 0.030	0.267	0.101	− 0.062	− 0.066	− 0.115	− 0.040

Each column has a mean of zero and a sum of squares of one

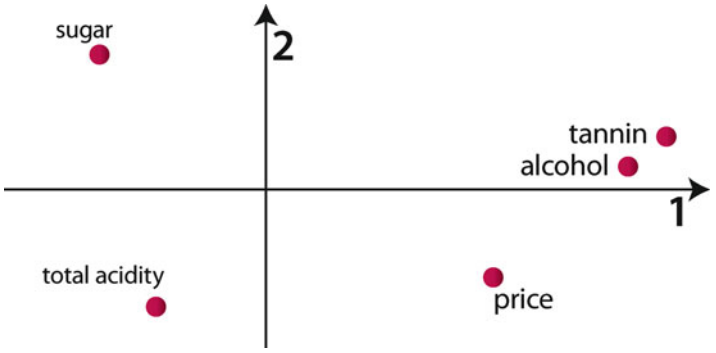

Fig. 2. The Saliences (normalized to their eigenvalues) for the physical attributes of the wines.

the interpretation of a plot of several saliences easier. Specifically, each salience is multiplied by its singular value, then, when a plot is made with the saliences corresponding to two different dimensions, the distances on the graph will directly reflect the amount of explained covariance of $\mathbf{R}$. The matrices $\mathbf{F_X}$ and $\mathbf{F_Y}$ are computed as

$$\mathbf{FX} = \mathbf{U}\Delta = \begin{bmatrix} 0.962 & -0.373 & -0.194 & 0.011 & 0.051 \\ -0.473 & -0.497 & 0.291 & -0.003 & 0.024 \\ 1.536 & 0.098 & 0.080 & -0.109 & 0.000 \\ -0.714 & 0.574 & 0.057 & -0.011 & 0.053 \\ 1.700 & 0.225 & 0.142 & 0.087 & 0.000 \end{bmatrix} \tag{15}$$

$$\mathbf{FY} = \mathbf{V}\Delta = \begin{bmatrix} -0.210 & 0.297 & 0.198 & -0.006 & 0.037 \\ -0.611 & -0.552 & 0.156 & 0.001 & -0.023 \\ -0.079 & -0.389 & 0.145 & -0.056 & 0.013 \\ 0.696 & 0.151 & 0.080 & 0.013 & -0.056 \\ 1.161 & -0.117 & -0.022 & 0.001 & -0.007 \\ -0.871 & 0.342 & 0.169 & 0.012 & -0.021 \\ 1.287 & -0.009 & 0.169 & 0.072 & 0.015 \\ -0.480 & -0.271 & -0.052 & 0.100 & 0.011 \\ -1.417 & 0.067 & -0.017 & 0.034 & -0.007 \end{bmatrix} \tag{16}$$

Figures 2 and 3 show the $\mathbf{X}$ and $\mathbf{Y}$ plot of the saliences for Dimensions 1 and 2.

4.5.4. Latent Variables

The latent variables for $\mathbf{X}$ and $\mathbf{Y}$ are computed according to Eqs. 5 and 6. These latent variables are shown in Tables 3 and 4. The corresponding plots for Dimensions 1 and 2 are given in Figures 4

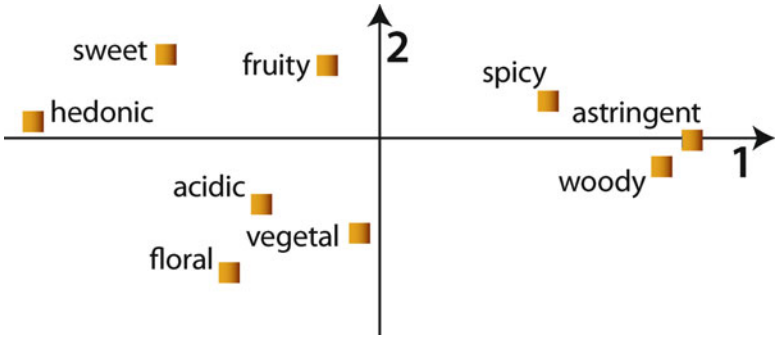

Fig. 3. The Saliences (normalized to their eigenvalues) for the sensory evaluation of the attributes of the wines.

Table 3
PLSC. The X latent variables. $L_X = Z_X V$

Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
0.249	0.156	0.033	0.065	− 0.092
0.278	0.230	0.110	0.093	− 0.216
0.252	0.153	0.033	− 0.060	− 0.186
0.184	0.147	− 0.206	0.026	− 0.026
0.004	0.092	− 0.269	− 0.083	− 0.102
− 0.119	0.003	0.058	− 0.101	− 0.052
− 0.226	− 0.197	0.102	0.054	− 0.053
− 0.170	− 0.098	− 0.009	0.030	− 0.049
− 0.278	0.320	0.140	− 0.080	0.194
− 0.269	0.300	− 0.155	0.102	− 0.121
− 0.317	0.355	− 0.110	0.084	0.083
− 0.120	0.171	0.047	− 0.132	− 0.054
0.392	− 0.155	0.155	− 0.120	0.113
0.405	− 0.073	0.255	0.030	0.005
0.328	− 0.225	0.120	− 0.086	0.073
0.226	− 0.150	− 0.200	0.067	0.076
0.030	− 0.090	− 0.163	− 0.113	0.114
− 0.244	− 0.153	0.019	0.099	0.128
− 0.236	− 0.119	− 0.040	0.121	0.098
0.051	− 0.090	− 0.081	− 0.177	0.067

(continued)

Table 3
(continued)

Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
− 0.299	0.200	− 0.026	0.097	0.088
− 0.206	0.146	− 0.046	0.029	− 0.058
− 0.201	0.214	0.034	− 0.065	0.159
− 0.115	0.076	− 0.046	− 0.040	− 0.040
0.323	0.004	− 0.058	0.123	0.221
0.399	0.112	0.193	0.009	0.083
0.435	− 0.029	− 0.137	0.106	0.080
0.379	− 0.310	− 0.183	− 0.013	0.016
− 0.018	− 0.265	0.002	− 0.079	− 0.062
− 0.051	− 0.192	0.097	− 0.118	− 0.183
− 0.255	− 0.326	0.164	0.106	− 0.026
− 0.146	− 0.626	0.127	0.134	0.058
− 0.248	0.126	0.054	0.021	− 0.077
− 0.226	0.174	− 0.010	0.027	− 0.054
− 0.108	0.096	− 0.080	− 0.040	− 0.110
− 0.084	0.025	0.079	− 0.117	− 0.092

Table 4
PLSC. The Y-latent variables. $L_Y = Z_X U$

Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
0.453	0.109	− 0.040	0.197	− 0.037
0.489	− 0.088	− 0.018	0.062	0.025
0.526	0.293	0.083	− 0.135	− 0.145
0.243	− 0.201	− 0.280	0.013	0.090
0.022	− 0.112	− 0.308	0.015	− 0.145
− 0.452	− 0.351	0.236	− 0.157	0.208
− 0.409	− 0.357	− 0.047	0.225	− 0.062
− 0.494	− 0.320	0.019	0.006	− 0.150
− 0.330	0.186	0.325	− 0.112	0.030

(continued)

Table 4 (continued)

Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
− 0.307	0.170	0.005	− 0.062	0.040
− 0.358	0.252	− 0.167	0.053	0.142
− 0.206	0.280	0.171	− 0.006	− 0.060
0.264	− 0.072	− 0.075	0.090	− 0.042
0.412	− 0.125	− 0.050	0.103	0.160
0.434	0.149	0.152	− 0.268	− 0.030
0.202	− 0.194	− 0.237	0.016	0.160
0.065	− 0.138	− 0.330	− 0.134	0.029
− 0.314	− 0.021	0.066	− 0.094	0.159
− 0.340	− 0.194	− 0.173	0.368	0.138
− 0.169	− 0.186	− 0.057	0.120	0.019
− 0.183	0.019	0.017	− 0.002	− 0.045
− 0.154	0.037	− 0.120	0.112	− 0.188
− 0.114	− 0.010	− 0.196	− 0.096	0.051
− 0.161	0.114	− 0.025	− 0.019	− 0.035
0.490	0.141	0.076	− 0.031	− 0.083
0.435	0.180	0.162	0.072	0.035
0.575	0.208	0.365	− 0.024	− 0.167
0.357	− 0.124	− 0.098	0.046	0.137
0.145	− 0.113	− 0.078	0.002	− 0.087
− 0.268	− 0.299	0.177	− 0.161	0.114
− 0.283	− 0.232	− 0.008	0.109	− 0.068
− 0.260	− 0.158	0.147	− 0.124	− 0.081
− 0.106	0.373	0.078	0.117	− 0.065
− 0.275	0.275	0.305	− 0.102	− 0.019
− 0.060	0.300	− 0.238	− 0.091	0.004
0.130	0.209	0.162	− 0.110	− 0.030

and 5. These plots show clearly that wine color is a major determinant of the wines both for the physical and the sensory points of view.

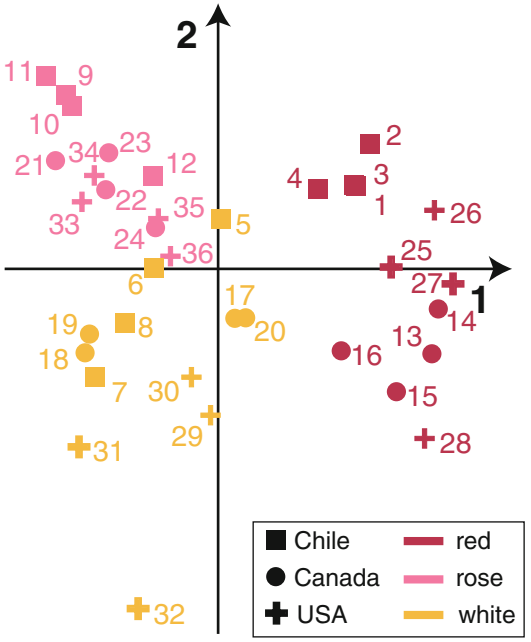

Fig. 4. Plot of the wines: The **X**-latent variables for Dimensions 1 and 2.

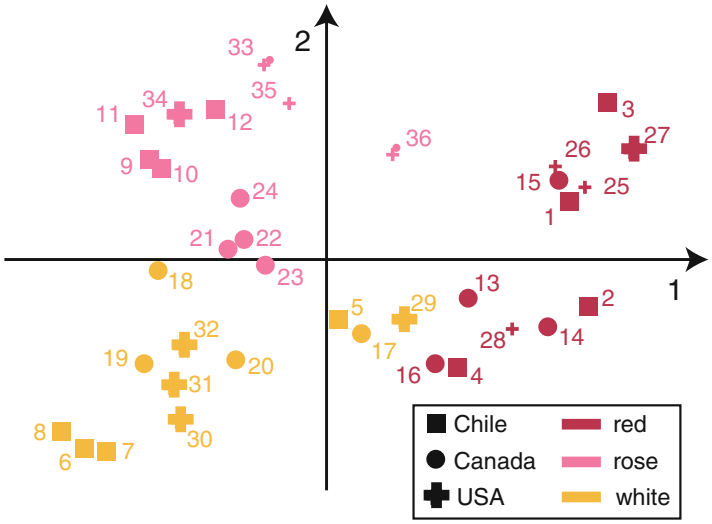

Fig. 5. Plot of the wines: The **Y**-latent variables for Dimensions 1 and 2.

4.5.5. Permutation Test

In order to evaluate if the overall analysis extracts relevant information, we computed the total inertia extracted by the PLSC. Using Eq. 4, we found that the inertia common to the two tables was equal to $\mathcal{I}_{\text{Total}} = 7.8626$. To evaluate its significance, we generated 10,000 **R** matrices by permuting the rows of **X**. The distribution of the values of the inertia is given in Fig. 6, which shows that the

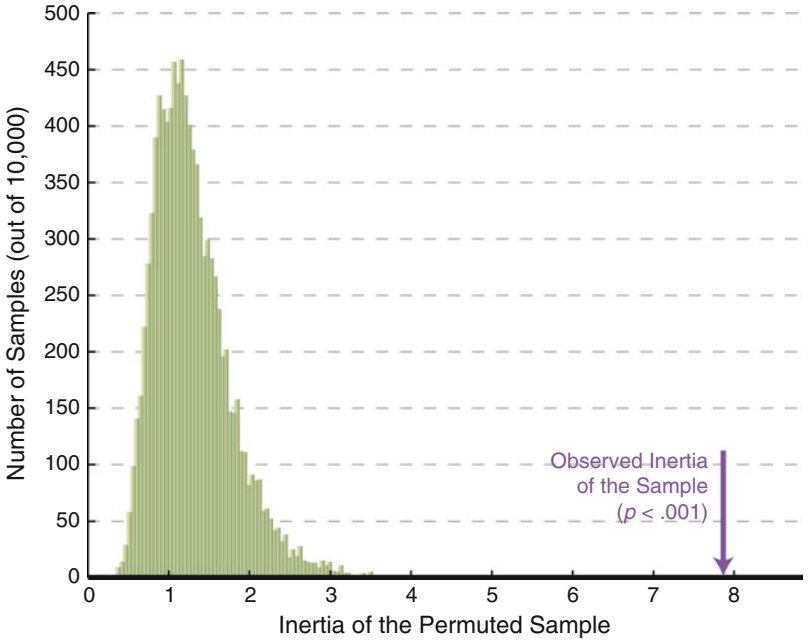

Fig. 6. Permutation test for the inertia explained by the PLSC of the wine. The observed value was never obtained in the 10,000 permutation. Therefore we conclude that PLSC extracted a significant amount of common variance between these two tables $P < 0.0001$).

value of $\mathcal{I}_{\text{Total}} = 7.8626$ was never obtained in this sample. Therefore we conclude that the probability of finding such a value by chance alone is smaller than $\frac{1}{10{,}000}$ (i.e., we can say that $p < .0001$).

The same approach can used to evaluate the significance of the dimensions extracted by PLSC. The permutation test found that only the first two dimensions could be considered significant at the $\alpha = .05$ level: For Dimension 1, $p < .0001$ and for Dimension 2 $p = .0043$. Therefore, we decided to keep only these first two dimensions for further analysis.

4.5.6. Bootstrap

Bootstrap ratios and 95% confidence intervals for **X** and **Y** are given for Dimensions 1 and 2 in Table 5. As it is often the case, bootstrap ratios and confidence intervals concur in indicating the relevant variables for a dimension. For example, for Dimension 1, the important variables (i.e., variables with a Bootstrap ratio > 2 or whose confidence interval excludes zero) for **X** are Tannin, Alcohol, Price, and Sugar; whereas for **Y** they are Hedonic, Astringent, Woody, Sweet, Floral, Spicy, and Acidic.

Table 5
PLSC. Bootstrap Ratios and Confidence Intervals for X and Y.

	Dimension 1			Dimension 2		
	Bootstrap ratio	Lower 95 % CI	Upper 95 % CI	Bootstrap ratio	Lower 95 % CI	Upper 95 % CI
X						
Price	3.6879	0.1937	0.5126	− 2.172	− 0.7845	− 0.1111
Acidity	− 1.6344	− 0.3441	0.0172	−3.334	− 0.8325	− 0.2985
Alcohol	13.7384	0.507	0.642	0.5328	− 0.2373	0.3845
Sugar	− 2.9555	− 0.4063	− 0.1158	4.7251	0.4302	0.8901
Tannin	16.8438	0.5809	0.7036	1.4694	− 0.0303	0.5066
Y						
Fruity	− 0.9502	− 0.2188	0.0648	2.0144	0.0516	0.5817
Floral	−3.9264	− 0.3233	− 0.1314	−3.4383	− 0.9287	− 0.3229
Vegetal	− 0.3944	− 0.139	0.0971	−2.6552	− 0.7603	− 0.195
Spicy	3.2506	0.1153	0.3709	1.0825	− 0.0922	0.4711
Woody	9.1335	0.3525	0.5118	− 0.6104	− 0.4609	0.2165
Sweet	−6.9786	− 0.408	− 0.2498	1.9499	0.043	0.6993
Astringent	16.6911	0.439	0.5316	− 0.0688	− 0.3099	0.291
Acidic	−2.5518	− 0.2778	− 0.0529	−1.443	− 0.6968	0.05
Hedonic	−22.7344	− 0.5741	− 0.4968	0.3581	− 0.285	0.4341

5. Partial Least Square Regression

Partial least square *Regression* (PLSR) is used when the goal of the analysis is to *predict* a set of variables (denoted $\mathbf{Y}$) from a set of predictors (called $\mathbf{X}$). As a regression technique, PLSR is used to predict a whole table of data (by contrast with standard regression which predicts one variable only), and it can also handle the case of multicolinear predictors (i.e., when the predictors are not linearly independent). These features make PLSR a very versatile tool because it can be used with very large data sets for which standard regression methods fail.

In order to predict a table of variables, PLSR finds latent variables, denoted $\mathbf{T}$ (in matrix notation), that model $\mathbf{X}$ and simultaneously predict $\mathbf{Y}$. Formally this is expressed as a double decomposition of $\mathbf{X}$ and the predicted $\widehat{\mathbf{Y}}$:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}} \quad \text{and} \quad \widehat{\mathbf{Y}} = \mathbf{T}\mathbf{B}\mathbf{C}^{\mathrm{T}}, \tag{17}$$

where **P** and **C** are called (respectively) **X** and **Y** loadings (or weights) and **B** is a diagonal matrix. These latent variables are ordered according to the amount of variance of $\widehat{\mathbf{Y}}$ that they explain. Rewriting Eq. 17 shows that $\widehat{\mathbf{Y}}$ can also be expressed as a regression model as

$$\widehat{\mathbf{Y}} = \mathbf{TBC}^{\mathbf{T}} = \mathbf{XB}_{\mathbf{PLS}} \tag{18}$$

with

$$\mathbf{B}_{PLS} = \mathbf{P}^{\mathbf{T}+}\mathbf{BC}^{\mathbf{T}} \tag{19}$$

(where $\mathbf{P}^{\mathrm{T}+}$ is the Moore–Penrose pseudoinverse of $\mathbf{P}^{\mathrm{T}}$, see, e.g., (12), for definitions). The matrix $\mathbf{B}_{\mathrm{PLS}}$ has J rows and K columns and is equivalent to the regression weights of multiple regression (Note that matrix **B** is diagonal, but that matrix $\mathbf{B}_{\mathrm{PLS}}$ is, in general *not* diagonal).

5.1. Iterative Computation of the Latent Variables in PLSR

In PLSR, the latent variables are computed by iterative applications of the SVD. Each run of the SVD produces orthogonal latent variables for **X** and **Y** and corresponding regression weights (see, e.g., (4) for more details and alternative algorithms).

5.1.1. Step One

To simplify the notation we will assume that **X** and **Y** are mean-centered and normalized such that the mean of each column is zero and its sum of squares is one. At step one, **X** and **Y** are stored (respectively) in matrices $\mathbf{X}_0$ and $\mathbf{Y}_0$. The matrix of correlations (or covariance) between $\mathbf{X}_0$ and $\mathbf{Y}_0$ is computed as

$$\mathbf{R}_1 = \mathbf{X}_0^T\mathbf{Y}_0. \tag{20}$$

The SVD is then performed on $\mathbf{R}_1$ and produces two sets of orthogonal singular vectors $\mathbf{W}_1$ and $\mathbf{C}_1$, and the corresponding singular values Δ_1 (compare with Eq. 1):

$$\mathbf{R}_1 = \mathbf{W}_1\Delta_1\mathbf{C}_1^T. \tag{21}$$

The first pair of singular vectors (i.e., the first columns of $\mathbf{W}_1$ and $\mathbf{C}_1$) are denoted $\mathbf{w}_1$ and $\mathbf{c}_1$ and the first singular value (i.e., the first diagonal entry of Δ_1) is denoted δ_1. The singular value represents the maximum covariance between the singular vectors. The first latent variable of **X** is given by (compare with Eq. 5 defining $\mathbf{L}_{\mathbf{X}}$):

$$\mathbf{t}_1 = \mathbf{X}_0\mathbf{w}_1 \tag{22}$$

where $\mathbf{t}_1$ is normalized such that $\mathbf{t}_1^T\mathbf{t}_1$. The loadings of $\mathbf{X}_0$ on $\mathbf{t}_1$ (i.e., the projection of $\mathbf{X}_0$ on the space of $\mathbf{t}_1$) are given by

$$\mathbf{p}_1 = \mathbf{X}_0^{\mathbf{T}}\mathbf{t}_1. \tag{23}$$

The least square estimate of $\mathbf{X}$ from the first latent variable is given by

$$\widehat{\mathbf{X}}_1 = \mathbf{t}_1^T \mathbf{p}_1. \tag{24}$$

As an intermediate step we derive a first pseudo latent variable for $\mathbf{Y}$ denoted $\mathbf{u}_1$ and obtained as

$$\mathbf{u}_1 = \mathbf{Y}_0 \mathbf{c}_1. \tag{25}$$

Reconstituting $\mathbf{Y}$ from its pseudo latent variable as

$$\widehat{\mathbf{Y}}_1 = \mathbf{u}_1 \mathbf{c}_1^{\mathbf{T}}, \tag{26}$$

and then rewriting Eq. 26 we obtained the prediction of $\mathbf{Y}$ from the $\mathbf{X}$ latent variable as

$$\widehat{\mathbf{Y}}_1 = \mathbf{t}_1 b_1 \mathbf{c}_1^{\mathrm{T}} \tag{27}$$

with

$$b_1 = \mathbf{t}_1^T \mathbf{u}_1. \tag{28}$$

The scalar b_1 is the slope of the regression of $\widehat{\mathbf{Y}}_1$ on $\mathbf{t}_1$.

Matrices $\widehat{\mathbf{X}}_1$ and $\widehat{\mathbf{Y}}_1$ are then subtracted from the original $\mathbf{X}_0$ and original $\mathbf{Y}_0$ respectively, to give *deflated* $\mathbf{X}_1$ and $\mathbf{Y}_1$:

$$\mathbf{X}_1 = \mathbf{X}_0 - \widehat{\mathbf{X}}_1 \quad \text{and} \quad \mathbf{Y}_1 = \mathbf{Y}_0 - \widehat{\mathbf{Y}}_1. \tag{29}$$

5.1.2. Last Step

The iterative process continues until $\mathbf{X}$ is completely decomposed into L components (where L is the rank of $\mathbf{X}$). When this is done, the weights (i.e., all the $\mathbf{w}_\ell$'s) for $\mathbf{X}$ are stored in the J by L matrix $\mathbf{W}$ (whose ℓth column is $\mathbf{w}_\ell$). The latent variables of $\mathbf{X}$ are stored in the I by L matrix $\mathbf{T}$. The weights for $\mathbf{Y}$ are stored in the K by L matrix $\mathbf{C}$. The pseudo latent variables of $\mathbf{Y}$ are stored in the I by L matrix $\mathbf{U}$. The loadings for $\mathbf{X}$ are stored in the J by L matrix $\mathbf{P}$. The regression weights are stored in a diagonal matrix $\mathbf{B}$. These regression weights are used to predict $\mathbf{Y}$ from $\mathbf{X}$; therefore, there is one b_ℓ for every pair of $\mathbf{t}_\ell$ and $\mathbf{u}_\ell$, and so $\mathbf{B}$ is an $L \times L$ diagonal matrix.

The predicted $\mathbf{Y}$ scores are now given by

$$\widehat{\mathbf{Y}} = \mathbf{TBC}^{\mathrm{T}} = \mathbf{XB}_{\mathbf{PLS}}, \tag{30}$$

where, $\mathbf{B}_{\mathrm{PLS}} = \mathbf{P}^{\mathrm{T}+}\,\mathbf{BC}^{\mathrm{T}}$, (where $\mathbf{P}^{\mathrm{T}+}$ is the Moore-Penrose pseudoinverse of $\mathbf{P}^{\mathrm{T}}$). $\mathbf{B}_{\mathrm{PLS}}$ has J rows and K columns.

5.2. What Does PLSR Optimize?

PLSR finds a series of L latent variables $\mathbf{t}_\ell$ such that the covariance between $\mathbf{t}_1$ and $\mathbf{Y}$ is maximal and such that $\mathbf{t}_1$ is uncorrelated with $\mathbf{t}_2$ which has maximal covariance with $\mathbf{Y}$ and so on for all L latent variables (see, e.g., (4, 17, 19, 26, 48, 49), for proofs and developments). Formally, we seek a set of L linear transformations of $\mathbf{X}$ that satisfies (compare with Eq. 7):

$$\mathbf{t}_\ell = \mathbf{X}\mathbf{w}_\ell \quad \text{such that} \quad \mathrm{cov}(\mathbf{t}_\ell, \mathbf{Y}) = \max \tag{31}$$

(where $\mathbf{w}_\ell$ is the vector of the coefficients of the ℓth linear transformation and cov is the covariance computed between $\mathbf{t}$ and each column of $\mathbf{Y}$) under the constraints that

$$\mathbf{t}_\ell^T \mathbf{t}_{\ell'} = 0 \quad \text{when} \quad \ell \neq \ell' \tag{32}$$

and

$$\mathbf{t}_\ell^T \mathbf{t}_\ell = 1. \tag{33}$$

5.3. How Good is the Prediction?

5.3.1. Fixed Effect Model

A common measure of the quality of prediction of observations within the sample is the Residual Estimated Sum of Squares (RESS), which is given by (4)

$$\text{RESS} = \| \mathbf{Y} - \widehat{\mathbf{Y}} \|^2, \tag{34}$$

where $\| \|^2$ is the square of the norm of a matrix (i.e., the sum of squares of all the elements of this matrix). The smaller the value of RESS, the better the quality of prediction (4, 13).

5.3.2. Random Effect Model

The quality of prediction generalized to observations outside of the sample is measured in a way similar to RESS and is called the Predicted Residual Estimated Sum of Squares (PRESS). Formally PRESS is obtained as (4):

$$\text{PRESS} = \| \mathbf{Y} - \widetilde{\mathbf{Y}} \|^2 . \tag{35}$$

The smaller PRESS is, the better the prediction.

5.3.3. How Many Latent Variables?

By contrast with the fixed effect model, the quality of prediction for a random model does not always increase with the number of latent variables used in the model. Typically, the quality first increases and then decreases. If the quality of the prediction decreases when the number of latent variables increases this indicates that the model is *overfitting* the data (i.e., the information useful to fit the observations from the learning set is not useful to fit *new* observations). Therefore, for a random model, it is critical to determine the optimal number of latent variables to keep for building the model. A straightforward approach is to stop adding latent variables as soon as the PRESS decreases. A more elaborated approach (see, e.g., (48)) starts by computing the ratio Q_ℓ^2 for the ℓth latent variable, which is defined as

$$Q_\ell^2 = 1 - \frac{\text{PRESS}_\ell}{\text{RESS}_\ell - 1,} \tag{36}$$

with PRESS_ℓ (resp. $\text{RESS}_{\ell-1}$) being the value of PRESS (resp. RESS) for the ℓth (resp. $\ell-1$) latent variable [where $\text{RESS}_0 = \text{K} \times (\text{I} - 1)$]. A latent variable is kept if its value of Q_ℓ^2 is larger than some arbitrary value generally set equal to $(1 - .95^2) = .0975$ (an

alternative set of values sets the threshold to .05 when $I \leq 100$ and to 0 when $I > 100$, see, e.g., (48, 58), for more details). Obviously, the choice of the threshold is important from a theoretical point of view, but, from a practical point of view, the values indicated above seem satisfactory.

5.3.4. Bootstrap Confidence Intervals for the Dependent Variables

When the number of latent variables of the model has been decided, confidence intervals for the predicted values can be derived using the Bootstrap. Here, each bootstrapped sample provides a value of $\mathbf{B}_{\text{PLS}}$ which is used to estimate the values of the observations in the testing set. The distribution of the values of these observations is then used to estimate the sampling distribution and to derive Bootstrap ratios and confidence intervals.

5.4. PLSR: Example

We will use the same example as for PLSC (see data in Tables 1 and 2). Here we used the physical measurements stored in matrix $\mathbf{X}$ to predict the sensory evaluation data stored in matrix $\mathbf{Y}$. In order to facilitate the comparison between PLSC and PLSR, we have decided to keep two latent variables for the analysis. However if we had used the Q^2 criterion of Eq. 36, with values of 1. 3027 for Dimension 1 and $-$ 0.2870 for Dimension 2, we should have kept only one latent variable for further analysis.

Table 6 gives the values of the latent variables ($\mathbf{T}$), the reconstituted values of $\mathbf{X}$ ($\widehat{\mathbf{X}}$) and the predicted values of $\mathbf{Y}$ ($\widehat{\mathbf{Y}}$). The value of $\mathbf{B}_{\text{PLS}}$ computed with two latent variables is equal to

$$
\begin{aligned}
&\mathbf{B}_{PLS} \\
&= \begin{bmatrix}
-0.0981 & 0.0558 & 0.0859 & 0.0533 & 0.1785 & -0.1951 & 0.1692 & 0.0025 & -0.2000 \\
-0.0877 & 0.3127 & 0.1713 & -0.1615 & -0.1204 & -0.0114 & -0.1813 & 0.1770 & 0.1766 \\
-0.0276 & -0.2337 & -0.0655 & 0.2135 & 0.3160 & -0.20977 & 0.3633 & -0.1650 & -0.3936 \\
0.1253 & -0.1728 & -0.1463 & 0.0127 & -0.1199 & 0.1863 & -0.0877 & -0.0707 & 0.1182 \\
0.0009 & -0.3373 & -0.1219 & 0.2675 & 0.3573 & -0.2072 & 0.4247 & -0.2239 & -0.4536
\end{bmatrix}.
\end{aligned}
\tag{37}
$$

The values of $\mathbf{W}$ which play the role of loadings for $\mathbf{X}$ are equal to

$$
\mathbf{W} = \begin{bmatrix}
\mathbf{-0.3660} & \mathbf{-0.4267} \\
\mathbf{0.1801} & \mathbf{-0.5896} \\
\mathbf{-0.5844} & \mathbf{0.0771} \\
\mathbf{0.2715} & \mathbf{0.6256} \\
\mathbf{-0.6468} & \mathbf{0.2703}
\end{bmatrix}. \tag{38}
$$

A plot of the first two dimensions of $\mathbf{W}$ given in Fig. 7 shows that $\mathbf{X}$ is structured around two main dimensions. The first dimension opposes the wines rich in alcohol and tannin (which are the red wines) are opposed to wines that are sweet or acidic. The second dimension opposes sweet wines to acidic wines (which are also more expensive) (Figs. 8 and 9).

Table 6
PLSR: Prediction of the sensory data (matrix Y) from the physical measurements (matrix X). Matrices T, U, $\widehat{X}$, $\widehat{Y}$

	T		U		$\widehat{X}$					$\widehat{Y}$								
Wine	Dim 1	Dim 2	Dim 1	Dim 2	Price	Total acidity	Alcohol	Sugar	Tannin	Fruity	Floral	Vegetal	Spicy	Woody	Sweet	Astringent	Acidic	Hedonic
1	− 0.16837	0.16041	−2.6776	0.97544	15.113	5.239	14.048	3.3113	471.17	6.3784	2.0725	1.7955	3.6348	4.3573	2.9321	4.0971	3.1042	2.8373
2	− 0.18798	0.22655	−2.8907	− 0.089524	14.509	4.8526	14.178	3.6701	517.12	6.4612	1.6826	1.6481	3.8384	4.5273	2.9283	4.337	2.9505	2.4263
3	− 0.17043	0.15673	−3.1102	2.1179	15.205	5.2581	14.055	3.2759	472.39	6.3705	2.0824	1.8026	3.6365	4.3703	2.9199	4.108	3.1063	2.8139
4	− 0.12413	0.14737	−1.4404	−1.0106	14.482	5.3454	13.841	3.438	413.67	6.4048	2.3011	1.8384	3.4268	4.0358	3.0917	3.7384	3.2164	3.5122
5	− 0.0028577	0.07931	− 0.13304	− 0.5399	13.226	5.8188	13.252	3.5632	245.11	6.4267	3.0822	2.0248	2.7972	3.14	3.4934	2.7119	3.5798	5.422
6	0.080038	− 0.015175	2.6712	−2.671	13.069	6.4119	12.82	3.319	113.62	6.3665	3.8455	2.2542	2.2783	2.5057	3.7148	1.9458	3.9108	6.8164
7	0.15284	− 0.18654	2.4224	−2.2504	14.224	7.4296	12.383	2.4971	−31.847	6.1754	4.9385	2.6436	1.6593	1.9082	3.8112	1.1543	4.3538	8.205
8	0.11498	− 0.09827	2.9223	−2.4331	13.636	6.9051	12.61	2.9187	43.514	6.2735	4.3742	2.4429	1.9796	2.2185	3.7601	1.5647	4.1249	7.4844
9	0.18784	0.21492	1.952	0.98895	7.6991	5.1995	12.482	5.4279	62.365	6.8409	3.1503	1.8016	2.2553	1.8423	4.3943	1.4234	3.7333	7.975
10	0.18149	0.21809	1.8177	0.95013	7.7708	5.1769	12.513	5.4187	71.068	6.8391	3.1112	1.7927	2.2876	1.8891	4.3728	1.4767	3.7149	7.8756
11	0.21392	0.25088	2.1158	1.4184	6.6886	5.017	12.388	5.8026	43.283	6.9247	3.0763	1.7341	2.2134	1.6728	4.5368	1.2706	3.7243	8.2918
12	0.080776	0.11954	1.2197	1.3413	11.084	5.6554	12.902	4.2487	158.44	6.5782	3.2041	1.9678	2.524	2.5621	3.8671	2.121	3.6803	6.5773
13	− 0.26477	− 0.085879	−1.5647	− 0.88629	20.508	6.5509	14.323	1.1469	503.24	5.8908	2.8881	2.2864	3.5804	4.9319	2.2795	4.5097	3.3327	1.8765
14	− 0.27335	− 0.012467	−2.4386	− 0.84706	19.593	6.1319	14.409	1.6096	538.44	5.9966	2.5048	2.1273	3.7516	5.0267	2.3272	4.6744	3.1889	1.6141
15	− 0.22148	− 0.14773	−2.5658	0.88267	20.609	6.931	14.089	0.93334	430.34	5.8398	3.3465	2.4328	3.2863	4.595	2.3812	4.0928	3.5271	2.6278
16	− 0.15251	− 0.089213	−1.1964	−1.2017	18.471	6.6538	13.817	1.6729	367.45	6.0044	3.3258	2.332	3.1078	4.1299	2.7176	3.6395	3.5663	3.5337
17	− 0.020577	− 0.072286	− 0.3852	− 0.65881	15.773	6.6575	13.235	2.4344	214.94	6.1706	3.7406	2.3412	2.5909	3.197	3.2555	2.6449	3.805	5.4426
18	0.16503	− 0.15453	1.8587	− 0.17362	13.53	7.2588	12.349	2.7767	−35.61	6.2384	4.8313	2.5797	1.6678	1.8359	3.8946	1.1032	4.3234	8.3249
19	0.15938	− 0.12373	2.0114	−1.163	13.184	7.0815	12.394	2.9608	−18.379	6.2806	4.6627	2.5123	1.7481	1.8903	3.9066	1.1882	4.2588	8.1846
20	− 0.034285	− 0.071934	0.99958	−1.5624	16.023	6.6453	13.297	2.3698	231.5	6.1567	3.6874	2.3357	2.6485	3.2949	3.202	2.7511	3.7765	5.2403
21	0.20205	0.12592	1.0834	0.53399	8.7377	5.7103	12.362	4.8856	15.13	6.7166	3.6292	1.9958	2.0319	1.7003	4.3514	1.1944	3.9154	8.3491

22	0.13903	0.095646	0.90872	0.44113	10.351	5.8333	12.626	4.3693	80.458	6.6025	3.5372	2.0386	2.2379	2.1358	4.0699	1.6397	3.8397	7.4783
23	0.13566	0.14176	0.67329	0.26414	9.7392	5.5716	12.67	4.6698	100.15	6.6711	3.3041	1.9394	2.3371	2.1809	4.1077	1.7276	3.7534	7.3432
24	0.077587	0.048002	0.95125	0.55919	12.19	6.055	12.871	3.7413	137.99	6.4628	3.5342	2.1189	2.4052	2.5521	3.7752	2.0495	3.797	6.6631
25	−0.21821	0.043304	−2.897	1.0065	17.752	5.8598	14.197	2.2626	491.22	6.1423	2.4453	2.0275	3.6255	4.659	2.6061	4.3241	3.2047	2.3216
26	−0.26916	0.13515	−2.5723	0.85536	17.355	5.3054	14.484	2.6448	583.5	6.2322	1.8146	1.8147	4.0069	5.0644	2.5074	4.8404	2.9431	1.4018
27	−0.29345	0.034272	−3.4006	1.2348	19.282	5.8542	14.529	1.8326	578.41	6.0485	2.2058	2.021	3.9215	5.1914	2.3	4.8922	3.0675	1.2318
28	−0.25617	−0.20133	−2.1121	−0.9005	22.038	7.2062	14.211	0.39528	453.76	5.7192	3.4724	2.5349	3.3314	4.8178	2.1853	4.2883	3.549	2.2172
29	0.011979	−0.21759	−0.85732	−0.18988	17.295	7.4986	12.996	1.5947	126.59	5.9776	4.5577	2.6614	2.1872	2.8984	3.2224	2.1988	4.1213	6.1909
30	0.034508	−0.16317	1.5868	−2.1363	16.08	7.2096	12.93	2.079	118.03	6.0867	4.3821	2.5534	2.1941	2.7626	3.3714	2.0981	4.0733	6.4213
31	0.17235	−0.29489	1.6713	−1.187	15.448	8.0531	12.226	1.8476	−92	6.0264	5.5299	2.8808	1.3781	1.7195	3.7677	0.85837	4.58	8.6928
32	0.098879	−0.52412	1.5407	−1.0685	20.167	9.2864	12.41	−0.087465	−81.643	5.5898	6.35	3.3433	1.26	2.1385	3.2243	1.1171	4.8257	8.0376
33	0.1672	0.072228	0.62606	2.4774	10.171	5.986	12.484	4.3461	38.717	6.5956	3.7551	2.0981	2.0776	1.9242	4.1547	1.391	3.9372	7.9361
34	0.15281	0.11474	1.6241	1.4483	9.816	5.7363	12.576	4.5679	70.404	6.6469	3.4977	2.0027	2.2159	2.0463	4.1453	1.5591	3.8348	7.6456
35	0.072566	0.066931	0.35548	1.7924	12.006	5.9449	12.906	3.8469	150.44	6.4872	3.4248	2.0769	2.461	2.5965	3.7765	2.1136	3.7542	6.5542
36	0.056807	0.0071035	−0.76977	1.6816	13.174	6.2693	12.938	3.3586	149.05	6.3768	3.6517	2.1988	2.416	2.6815	3.6481	2.1548	3.8253	6.4334

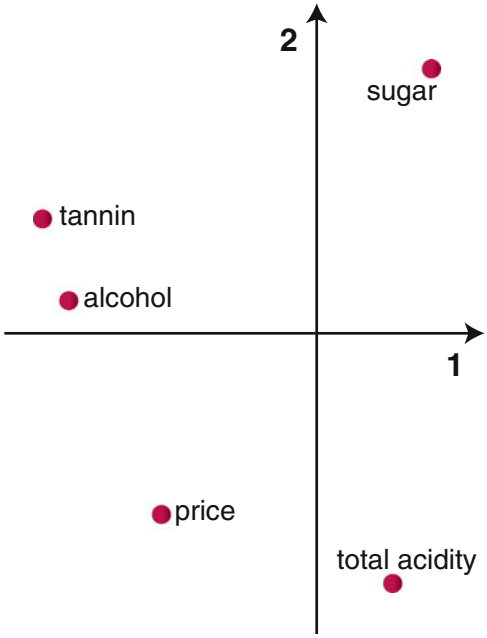

Fig. 7. The **X**-loadings for Dimensions 1 and 2.

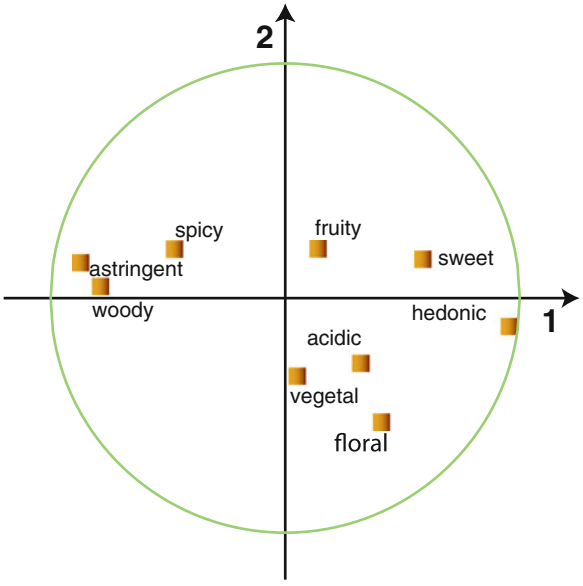

Fig. 8. The circle of correlation between the **Y** variables and the latent variables for Dimensions 1 and 2.

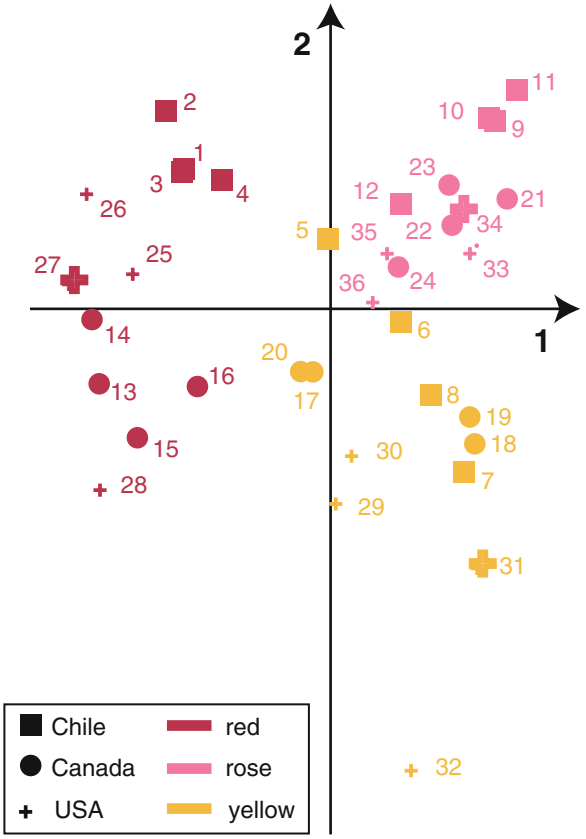

Fig. 9. PLSR. Plot of the latent variables (wines) for Dimensions 1 and 2.

6. Software

PLS methods necessitate sophisticated computations and therefore they critically depends on the availability of software.

PLSC is used intensively is neuroimaging, and most of the analyses in this domain are performed with a special MATLAB toolboox (written by McIntosh, Chau, Lobaugh, and Chen). The programs and a tutorial are freely available from www.rotman-baycrest.on.ca:8080. These programs (which are the standard for neuroimaging) can be adapted for other types of data than neuroimaging (as long as the data are formatted in a compatible format). The computations reported in this paper were performed with MATLAB and can be downloaded from the home page of the first author (www.utdallas.edu/~herve).

For PLSR there are several available choices. The computations reported in this paper are performed with MATLAB and can be downloaded from the home page of the first author (www.utdallas.

edu/~herve). A public domain set of MATLAB programs is also available from the home page of the *N*-Way project (www.models.kvl.dk/source/nwaytoolbox/) along with tutorials and examples. The statistic toolbox from MATLAB includes a function to perform PLSR. The public domain program R implements PLSR through the package PLS (43). The general purpose statistical packages SAS, SPSS, and XLSTAT (which has, by far the most extensive implementation of PLS methods) can be also used to perform PLSR. In chemistry and sensory evaluation, two main programs are used: the first one called SIMCA-P was developed originally by Wold (who also pioneered PLSR), the second one called the UNSCRAMBLER was first developed by Martens who was another pioneer in the field. And finally, a commercial MATLAB toolbox has also been developed by EIGENRESEARCH.

7. Related Methods

A complete review of the connections between PLS and the other statistical methods is, clearly, out of the scope of an introductory paper (see, however, (17, 48, 49, 26), for an overview), but some directions are worth mentioning. PLSC uses the SVD in order to analyze the information common to two or more tables, and this makes it closely related to several other SVD (or eigen-decomposition) techniques with similar goals. The closest technique is obviously inter-battery analysis (51) which uses the same SVD as PLSC, but on non structured matrices. Canonical correlation analysis (also called simply canonical analysis, or canonical variate analysis, see (28, 33), for reviews) is also a related technique that seeks latent variables with largest *correlation* instead of PLSC's criterion of largest *covariance*. Under the assumptions of normality, analytical statistical tests are available for canonical correlation analysis but cross-validation procedures analogous to PLSC could also be used.

In addition, several multi-way techniques encompass as a particular case data sets with two tables. The oldest and most well-known technique is multiple factor analysis which integrates different tables into a common PCA by normalizing each table with its first singular value (7, 25). A more recent set of techniques is the STATIS family which uses a more sophisticated normalizing scheme whose goal is to extract the common part of the data (see (1, 8–11), for an introduction). Closely related techniques comprise common component analysis (36) which seeks a set of factors common to a set of data tables, and co-inertia analysis which could be seen as a generalization of Tucker's (1958) (51) inter-battery analysis (see, e.g., (18, 22, 50, 50, 54), for recent developments).

PLSR is strongly related to regression-like techniques which have been developed to cope with the multi-colinearity problem. These include principal component regression, ridge regression, redundancy analysis (also known as PCA on instrumental variables (44, 52, 53), and continuum regression (45), which provides a general framework for these techniques.

8. Conclusion

Partial Least Squares (PLS) methods analyze data from multiple modalities collected on the same observations. We have reviewed two particular PLS methods: Partial Least Squares *Correlation* or PLSC and Partial Least Squares *Regression* or PLSR. PLSC analyzes the shared information between two or more sets of variables. In contrast, PLSR is *directional* and predicts a set of dependent variables from a set of independent variables or predictors. The relationship between PLSC and PLSR are also explored in (17) and, recently (27) proposed to integrate these two approaches into a new predictive approach called BRIDGE-PLS. In practice, the two techniques are likely to give similar conclusions because the criteria they optimize are quite similar.

References

1. Abdi H (2001) Linear algebra for neural networks. In: Smelser N, Baltes P (eds) International encyclopedia of the social and behavioral sciences. Elsevier, Oxford UK
2. Abdi H (2007a) Eigen-decomposition: eigenvalues and eigenvectors. In: Salkind N (ed) Encyclopedia of measurement and statistics. Sage, Thousand Oaks, CA
3. Abdi H (2007) Singular value decomposition (SVD) and generalized singular value decomposition (GSVD). In: Salkind N (ed) Encyclopedia of measurement and statistics. Sage, Thousand Oaks, CA
4. Abdi H (2010) Partial least square regression, projection on latent structure regression, PLS-regression. Wiley Interdiscipl Rev Comput Stat 2:97–106
5. Abdi H, Dunlop JP, Williams LJ (2009) How to compute reliability estimates and display confidence and tolerance intervals for pattern classifiers using the Bootstrap and 3-way multidimensional scaling (DISTATIS). NeuroImage 45:89–95
6. Abdi H, Edelman B, Valentin D, Dowling WJ (2009b) Experimental design and analysis for psychology. Oxford University Press, Oxford
7. Abdi H, Valentin D (2007a) Multiple factor analysis (MFA). In: Salkind N (ed) Encyclopedia of measurement and statistics. Sage, Thousand Oaks, CA
8. Abdi H, Valentin D (2007b) STATIS. In: Salkind N (ed) Encyclopedia of measurement and statistics. Sage, Thousand Oaks, CA
9. Abdi H, Valentin D, O'Toole AJ, Edelman B (2005) DISTATIS: the analysis of multiple distance matrices. In: Proceedings of the IEEE computer society: international conference on computer vision and pattern recognition pp 42–47
10. Abdi H, Williams LJ (2010a) Barycentric discriminant analysis. In: Salkind N (ed) Encyclopedia of research design. Sage, Thousand Oaks, CA
11. Abdi H, Williams LJ (2010b) The jackknife. In: Salkind N (ed) Encyclopedia of research design. Sage, Thousand Oaks, CA

12. Abdi H, Williams LJ (2010c) Matrix algebra. In: Salkind N (ed) Encyclopedia of research design. Sage, Thousand Oaks, CA
13. Abdi H, Williams LJ (2010d) Principal components analysis. Wiley Interdiscipl Rev Comput Stat 2:433–459
14. Bookstein F (1982) The geometric meaning of soft modeling with some generalizations. In: Jöreskog K, Wold H (eds) System under indirect observation, vol 2. North-Holland, Amsterdam.
15. Bookstein FL (1994) Partial least squares: a dose-response model for measurement in the behavioral and brain sciences. Psycoloquy 5
16. Boulesteix AL, Strimmer K (2006) Partial least squares: a versatile tool for the analysis of high-dimensional genomic data. Briefing in Bioinformatics 8:32–44
17. Burnham A, Viveros R, MacGregor J (1996) Frameworks for latent variable multivariate regression. J Chemometr 10:31–45
18. Chessel D, Hanafi M (1996) Analyse de la co-inertie de *k* nuages de points. Revue de Statistique Appliquée 44:35–60
19. de Jong S (1993) SIMPLS: an alternative approach to partial least squares regression. Chemometr Intell Lab Syst 18:251–263
20. de Jong S, Phatak A (1997) Partial least squares regression. In: Proceedings of the second international workshop on recent advances in total least squares techniques and error-in-variables modeling. Society for Industrial and Applied Mathematics
21. de Leeuw J (2007) Derivatives of generalized eigen-systems with applications. Department of Statistics Papers, 1–28
22. Dray S, Chessel D, Thioulouse J (2003) Co-inertia analysis and the linking of ecological data tables. Ecology 84:3078–3089
23. Efron B, Tibshirani RJ (1986) Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat Sci 1:54–77
24. Efron B, Tibshirani RJ (1993) An introduction to the bootstrap. Chapman & Hall, New York
25. Escofier B, Pagès J (1990) Multiple factor analysis. Comput Stat Data Anal 18:120–140
26. Esposito-Vinzi V, Chin WW, Henseler J, Wang H (eds) (2010) Handbook of partial least squares: concepts, methods and applications. Springer, New York.
27. Gidskehaug L, Stødkilde-Jørgensen H, Martens M, Martens H (2004) Bridge-PLS regression: two-block bilinear regression without deflation. J Chemometr 18:208–215
28. Gittins R (1985) Canonical analysis. Springer, New York
29. Good P (2005) Permutation, parametric and bootstrap tests of hypotheses. Springer, New York
30. Greenacre M (1984) Theory and applications of correspondence analysis. Academic, London
31. Krishnan A, Williams LJ, McIntosh AR, Abdi H (2011) Partial least squares (PLS) methods for neuroimaging: a tutorial and review. NeuroImage 56:455–475
32. Lebart L, Piron M, Morineau A (2007) Statistiques exploratoires multidimensionelle. Dunod, Paris
33. Mardia KV, Kent JT, Bibby JM (1979) Multivariate analysis. Academic, London
34. Martens H, Martens M (2001) Multivariate analysis of quality: an introduction. Wiley, London
35. Martens H, Naes T (1989) Multivariate calibration. Wiley, London
36. Mazerolles G, Hanafi M, Dufour E, Bertrand D, Qannari ME (2006) Common components and specific weights analysis: a chemometric method for dealing with complexity of food products. Chemometr Intell Lab Syst 81:41–49
37. McCloskey DN, Ziliak J (2008) The cult of statistical significance: how the standard error costs us jobs, justice, and lives. University of Michigan Press, Michigan
38. McIntosh AR, Gonzalez-Lima F (1991) Structural modeling of functional neural pathways mapped with 2-deoxyglucose: effects of acoustic startle habituation on the auditory system. Brain Res 547:295–302
39. McIntosh AR, Lobaugh NJ (2004) Partial least squares analysis of neuroimaging data: applications and advances. NeuroImage 23: S250–S263
40. McIntosh AR, Chau W, Protzner A (2004) Spatiotemporal analysis of event-related fMRI data using partial least squares. NeuroImage 23:764–775
41. McIntosh AR, Bookstein F, Haxby J, Grady C (1996) Spatial pattern analysis of functional brain images using partial least squares. NeuroImage 3:143–157
42. McIntosh AR, Nyberg L, Bookstein FL, Tulving E (1997) Differential functional connectivity of prefrontal and medial temporal cortices during episodic memory retrieval. Hum Brain Mapp 5:323–327
43. Mevik B-H, Wehrens R (2007) The PLS package: principal component and partial least

squares regression in R. J Stat Software 18:1–24

44. Rao C (1964) The use and interpretation of principal component analysis in applied research. Sankhya 26:329–359
45. Stone M, Brooks RJ (1990) Continuum regression: cross-validated sequentially constructed prediction embracing ordinary least squares, partial least squares and principal components regression. J Roy Stat Soc B 52:237–269
46. Streissguth A, Bookstein F, Sampson P, Barr H (1993) Methods of latent variable modeling by partial least squares. In: The enduring effects of prenatal alcohol exposure on child development. University of Michigan Press
47. Takane Y (2002) Relationships among various kinds of eigenvalue and singular value decompositions. In: Yanai H, Okada A, Shigemasu K, Kano Y, Meulman J (eds) New developments in psychometrics. Springer, Tokyo
48. Tenenhaus M (1998) La regression PLS. Technip, Paris
49. Tenenhaus M, Tenenhaus A (in press) Regularized generalized canonical correlation analysis. Psychometrika
50. Thioulouse J, Simier M, Chessel D (2003) Simultaneous analysis of a sequence of paired ecological tables. Ecology 20:2197–2208
51. Tucker L (1958) An inter-battery method of factor analysis. Psychometrika 23:111–136
52. Tyler DE (1982) On the optimality of the simultaneous redundancy transformations. Psychometrika 47:77–86
53. van den Wollenberg A (1977) Redundancy analysis: an alternative to canonical correlation. Psychometrika 42:207–219
54. Williams LJ, Abdi H, French R, Orange JB (2010) A tutorial on Multi-Block Discriminant Correspondence Analysis (MUDICA): a new method for analyzing discourse data from clinical populations. J Speech Lang Hear Res 53:1372–1393
55. Wold H (1966) Estimation of principal component and related methods by iterative least squares. In: Krishnaiah PR (ed) Multivariate analysis. Academic Press, New York
56. Wold H (1973) Nonlinear Iterative Partial Least Squares (NIPALS) modeling: some current developments. In: Krishnaiah PR (ed) Multivariate analysis. Academic Press, New York
57. Wold H (1982) Soft modelling, the basic design and some extensions. In: Wold H, Jöreskog K-G (eds) Systems under indirect observation: causality-structure-prediction, Part II. North-Holland, Amsterdam
58. Wold S (1995) PLS for multivariate linear modelling. In: van de Waterbeenl H (ed) QSAR: chemometric methods in molecular design, methods and principles in medicinal chemistry, vol 2. Verla Chemie, Weinheim Germany
59. Wold S, Sjöström M, Eriksson L (2001) PLS-regression: a basic tool of chemometrics. Chemometr Intell Lab Syst 58:109–130