The Construction of a Partial Least Squares Biplot

Opeoluwa Oyedele[1, 2] and Sugnet Gardner-Lubbe[1]
[1] University of Cape Town, Cape Town, SOUTH AFRICA
[2] Corresponding author: Opeoluwa Oyedele, e-mail: opeoluwaoyedele@gmail.com

Abstract

A graphical display of PLS regression of a data set is presented. Biplots in regression analysis has many advantages, including demonstrating the association between samples and variables graphically. The PLS biplot provides a single graphical representation of the samples alongside the predictor and response variables, as well as their interrelationships.

Keywords: Biplots, Partial Least Squares Regression, Principal Component Analysis biplot.

1. INTRODUCTION

One of the most often asked questions in data analysis is the question of how to model one or more response variables using a set of predictor variables. Usually the modelling of the responses is done using the Multiple Linear Regression (MLR) technique. However, with modern day measuring instruments, data can be very large, strongly correlated and sometimes incomplete. For this reason, Partial Least Squares Regression (PLSR) can be a useful tool for modelling. PLSR is a technique that combines features from Principal Component Analysis (PCA) and MLR, with the goal of predicting the responses using a set of orthogonal latent variables extracted from the predictors.

Results found by most recognised statistical methods of analysis can be visualized graphically using some form of graphical display such as biplots - a joint graphical display of all rows and of all columns of a data matrix. Biplots are often referred to as the multivariate version of scatter plots since they allow for the display of samples as points and each variable by an axis on the plot.

In this paper, a brief overview of Partial Least Squares (PLS) is provided before the construction of the PLS biplot is derived in detail. This biplot allows for the simultaneous representation of the sample points and variables. Both the predictor variables and response variables are represented, as well as the matrix of PLSR coefficients.

2. PARTIAL LEAST SQUARES

PLS searches for a set of orthogonal factors, components or latent variables, that performs a simultaneous decomposition of a set of predictor variables $\mathbf{X}: \mathrm{N} \times \mathrm{P}$ and response variables $\mathbf{Y}: \mathrm{N} \times \mathrm{M}$, with the restriction that these components explain as much as possible of the covariance between $\mathbf{X}$ and $\mathbf{Y}$. The PLS model aims to find a few ($\mathrm{A} < \mathrm{P}$) latent variables called X-scores. They are denoted by $\mathbf{T}$ and since the latent variables are orthogonal $\mathbf{T}^{\mathrm{T}}\mathbf{T} = \mathbf{I}_{\mathrm{A}}$.

Different iterative algorithms have been proposed for computing these latent variables (Martins *et al*., 2010). The most popular algorithms are the NIPALS (Nonlinear Iterative PArtial Least Squares), Kernel and SIMPLS (Statistical Inspired Modification to Partial Least Squares) algorithms. Let $\mathrm{a} = 1, \ldots, \mathrm{A}$ represent the successive orthogonal latent variables, then the columns of the following matrices are calculated successively: $\mathbf{R}: \mathrm{P} \times \mathrm{A}$ a matrix of (transformed) predictor weights, $\mathbf{C}: \mathrm{M} \times \mathrm{A}$ a matrix of response weights, $\mathbf{T}: \mathrm{N} \times \mathrm{A}$ a matrix of X-scores, $\mathbf{P}: \mathrm{P} \times \mathrm{A}$ a matrix of X-loadings and $\mathbf{Q}: \mathrm{M} \times \mathrm{A}$ a matrix of Y-loadings. The first step is to column centre both the $\mathbf{X}$ and $\mathbf{Y}$ matrices by subtracting the mean of each variable

respectively. Denote the centred data matrices by $\mathbf{X}_0$ and $\mathbf{Y}_0$. The different algorithms can result in different PLS decompositions, but in all cases $\mathbf{T} = \mathbf{X}_0\mathbf{R}$ where $\mathbf{X}_0 = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \mathbf{E}$ and $\mathbf{Y}_0 = \mathbf{T}\mathbf{Q}^{\mathrm{T}} + \mathbf{F}$. If A = P there is no dimension reduction in the matrix of X-scores and the error component $\mathbf{E}$ vanishes. The $\mathbf{T}$-matrix of X-scores represent the simultaneous decomposition of both $\mathbf{X}_0 = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \mathbf{E}$ and $\mathbf{Y}_0 = \mathbf{T}\mathbf{Q}^{\mathrm{T}} + \mathbf{F}$.

2.1 Partial Least Squares Regression

In MLR, M Y-variables are modelled by solving the equation $\mathbf{Y} = \mathbf{X}\mathbf{B}$. The unknown coefficients matrix $\mathbf{B}$: P × M is obtained by $\widehat{\mathbf{B}} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y}$. Thus, $\mathbf{Y}$ is estimated as $\widehat{\mathbf{Y}} = \mathbf{X}\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y} = \mathbf{X}\widehat{\mathbf{B}}$.

With the aim of predicting $\mathbf{Y}$ using the few components extracted from $\mathbf{X}$, PLSR uses the X-scores as predictors in the modelling of $\mathbf{Y}$. As a result, $\mathbf{Y}$ is estimated as $\widehat{\mathbf{Y}}_0 = \mathbf{T}\left(\mathbf{T}^{\mathrm{T}}\mathbf{T}\right)^{-1}\mathbf{T}^{\mathrm{T}}\mathbf{Y}_0 = \mathbf{T}\mathbf{T}^{\mathrm{T}}\mathbf{Y}_0 \cong \mathbf{T}\mathbf{Q}^{\mathrm{T}} = \mathbf{X}_0\mathbf{R}\mathbf{Q}^{\mathrm{T}} = \mathbf{X}_0\widehat{\mathbf{B}}_{\mathrm{PLSR}}$, where $\left(\mathbf{T}^{\mathrm{T}}\mathbf{T}\right) = \mathbf{I}_{\mathrm{A}}$, A is the number of PLS components and $\widehat{\mathbf{B}}_{\mathrm{PLSR}} = \mathbf{R}\mathbf{Q}^{\mathrm{T}}$ is the estimated coefficients matrix for the PLS model, with $\mathbf{Q} \cong \mathbf{Y}_0^{\mathrm{T}}\mathbf{T}$.

3. BIPLOTS

According to Barnett (1981), the biplot is a joint graphical display of rows and columns of a data matrix $\mathbf{D}$: G × H by means of markers $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_{\mathrm{G}}$ for its rows and markers $\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_{\mathrm{H}}$ for its columns. Each marker is chosen in such a way that the inner product $\mathbf{a}_{\mathrm{i}}^{\mathrm{T}}\mathbf{b}_{\mathrm{j}}$ represents d_{ij}, the $(\mathrm{i},\mathrm{j})^{\mathrm{th}}$ element of the data matrix $\mathbf{D}$. That is, $\widehat{\mathbf{D}} = \mathbf{A}\mathbf{B}^{\mathrm{T}}$. Usually the biplot is constructed in two dimensions, so that both $\mathbf{A}$ and $\mathbf{B}$ have two columns. In PLS, the data has the form of N sample points measured on P predictor variables and M response variables. Since the roles of the columns and rows cannot be interchanged, the resulting biplot will be an asymmetric biplot. In asymmetric biplots, rows of a data matrix are represented as points while columns are represented as axes. Therefore, biplot points will be defined by the row markers of the data matrix and biplot axes by the column markers. That is, G rows of $\mathbf{A}$ will serve as the biplot points for the biplot of a data matrix $\mathbf{D}$ while H rows of $\mathbf{B}$ will give the directions of the biplot axes.

3.1 PCA biplot

The simplest form of an asymmetric biplot is the Principal Component Analysis (PCA) biplot. In PCA, only a single set of variables $\mathbf{X}$: N × P is decomposed into a set of principal components (latent variables) $\mathbf{T} = \mathbf{X}_0\mathbf{V}_{[\mathrm{A}]}$ where $\mathbf{V}_{[\mathrm{A}]}$ contains the first A columns of the right singular vectors of $\mathbf{X}_0 = \mathbf{U}\mathbf{D}\mathbf{V}^{\mathrm{T}}$. It follows that $\mathbf{T} = \mathbf{X}_0\mathbf{V}_{[\mathrm{A}]} = \mathbf{U}\mathbf{D}\mathbf{V}^{\mathrm{T}}\mathbf{V}_{[\mathrm{A}]} = \mathbf{U}\mathbf{D}\mathbf{J} = (\mathbf{U}\mathbf{D})_{[\mathrm{A}]}$ where $\mathbf{J}$: P × P $= \begin{bmatrix}\mathbf{I}_{\mathrm{A}} \\ \mathbf{0}\end{bmatrix}$. Furthermore the best A-dimensional approximation of $\mathbf{X}$ is given by $\mathbf{X}_0 = \mathbf{X}_0\mathbf{V}\mathbf{V}^{\mathrm{T}} \cong \mathbf{X}_0\mathbf{V}\mathbf{J}\mathbf{J}^{\mathrm{T}}\mathbf{V}^{\mathrm{T}} = \mathbf{X}_0\mathbf{V}_{[\mathrm{A}]}\mathbf{V}_{[\mathrm{A}]}^{\mathrm{T}} = \mathbf{T}\mathbf{V}_{[\mathrm{A}]}^{\mathrm{T}}$ (Eckart & Young, 1936). In the PCA biplot the samples are represented by the row markers defined by the rows of $\mathbf{T}$ and the directions of the biplot axes are given by the rows of the matrix $\mathbf{V}_{[\mathrm{A}]}$.

Gower & Hand (1996) defines the processes of interpolation and prediction as finding the position of a P-dimensional sample in the A-dimensional biplot space and inferring the values of the P original variables for any point in the A-dimensional biplot space respectively. These authors show that given a sample $\mathbf{x}$, it is interpolated into the biplot space by the equation $\mathbf{t}^{\mathrm{T}} = (\mathbf{x} - \bar{\mathbf{x}})^{\mathrm{T}}\mathbf{V}_{[\mathrm{A}]}$ where $\bar{\mathbf{x}}$ is the column means of $\mathbf{X}$.

To trace the prediction biplot axis for the k^{th} original variable, values of $\mu_0 \in (-\infty, \infty)$ is substituted into the expression $\frac{\mu_0}{\mathbf{e}_{\mathrm{k}}^{\mathrm{T}}\mathbf{V}_{[\mathrm{A}]}\mathbf{V}_{[\mathrm{A}]}^{\mathrm{T}}\mathbf{e}_{\mathrm{k}}}\mathbf{V}_{[\mathrm{A}]}^{\mathrm{T}}\mathbf{e}_{\mathrm{k}}$ where $\mathbf{e}_{\mathrm{k}}$ is the unit vector with zeros except a one in the k^{th} position. The values μ_0 are in terms of the

centred observations and calibration markers are fitted where $\mu = \mu_0 + \bar{x}_k$ are sensible scale marker values.

3.2 PLS biplot

PLS can be viewed as an approximation technique since it approximates a data matrix using only a few components. As a result, $\mathbf{X}$ is approximated using a set of latent variables ($\mathbf{T}$) and a set of loadings ($\mathbf{P}$) such that $\mathbf{X}_0 \cong \mathbf{T}\mathbf{P}^T$ and similarly $\mathbf{Y}$ is approximated using the same set of latent variables ($\mathbf{T}$) and a set of loadings ($\mathbf{Q}$) such that $\mathbf{Y}_0 \cong \mathbf{T}\mathbf{Q}^T$. Analogous to the PCA biplot, the samples are represented by the rows of the matrix $\mathbf{T}$ while the directions of the biplot axes are given by the rows of $\mathbf{P}$ and $\mathbf{Q}$ for the predictor variables and response variables respectively. Although this is true for any number of components A, it is only practical to plot two or three dimensions. For two-dimensional biplots, $A = 2$ components will be extracted and plotted in a new A-dimensional orthogonal Cartesian axes system referred to as the scaffolding axes. These scaffolding axes are not seen but since $\mathbf{T} = \mathbf{X}_0\mathbf{R}$, they are defined by the first A columns of $\mathbf{R}$, the PLS X-weights matrix. Consider $\mathbf{x}$ and $\mathbf{y}$ a pair of predictor and response samples. Both $\mathbf{x}$ and $\mathbf{y}$ are interpolated into the biplot space using the equation $\mathbf{t}^T = (\mathbf{x} - \bar{\mathbf{x}})^T\mathbf{R}$. Given that $\mathbf{X}_0 = \mathbf{T}\mathbf{P}^T$, for any point $\mathbf{t}$ in the biplot space $\mathbf{x}_0^T = \mathbf{t}^T\mathbf{P}^T$. The k^{th} centred original variable value is given by $\mathbf{x}_0^T\mathbf{e}_k = \mathbf{t}^T\mathbf{P}^T\mathbf{e}_k$. If this value is μ_{0X} then

$$\mu_{0X} = \mathbf{t}^T\mathbf{P}^T\mathbf{e}_k. \quad (1)$$

This defines a line in the two dimensional biplot space and for different values $\mu_{0X} \in (-\infty, \infty)$ parallel lines are obtained. To facilitate orthogonal projection onto the biplot axes, similar to scatter plots, the line through the origin orthogonal to (1) is selected as the biplot axis for variable k. Any point on this biplot axis will have the form $\phi\mathbf{P}^T\mathbf{e}_k$. As a result, the point on the biplot axis predicting the value μ_{0X} for the k^{th} variable will have

$$\mathbf{t}_{\mu_{0X}} = \phi\mathbf{P}^T\mathbf{e}_k. \quad (2)$$

Substituting (2) in (1) yields $\mu_{0X} = \phi\mathbf{e}_k^T\mathbf{P}\mathbf{P}^T\mathbf{e}_k$. Solving for ϕ yields

$$\phi = \frac{\mu_{0X}}{\mathbf{e}_k^T\mathbf{P}\mathbf{P}^T\mathbf{e}_k}. \quad (3)$$

Furthermore, replacing ϕ in (2) by (3) gives the marker μ_{0X} on the k^{th} prediction biplot axis as $\frac{\mu_{0X}}{\mathbf{e}_k^T\mathbf{P}\mathbf{P}^T\mathbf{e}_k}\mathbf{P}^T\mathbf{e}_k$. Similarly, since $\mathbf{Y}_0 = \mathbf{T}\mathbf{Q}^T$, any point predicting μ_{0Y} for the k^{th} response variable will have $\mu_{0Y} = \mathbf{t}^T\mathbf{Q}^T\mathbf{e}_k$ with the biplot axis of the form $\phi\mathbf{Q}^T\mathbf{e}_k$. If the value μ_{0Y} is predicted, $\mathbf{t}_{\mu_{0Y}} = \phi\mathbf{Q}^T\mathbf{e}_k$ and $\mu_{0Y} = \phi\mathbf{e}_k^T\mathbf{Q}\mathbf{Q}^T\mathbf{e}_k$ so that the marker μ_{0Y} on the k^{th} response prediction biplot axis is given by $\frac{\mu_{0Y}}{\mathbf{e}_k^T\mathbf{Q}\mathbf{Q}^T\mathbf{e}_k}\mathbf{Q}^T\mathbf{e}_k$. The values μ_{0X} and μ_{0Y} are in terms of the centred samples and calibration markers are fitted where $\mu_X = \mu_{0X} + \bar{x}_k$ and $\mu_Y = \mu_{0Y} + \bar{y}_k$ are sensible scale marker values.

With the PLSR coefficients matrix defined as $\widehat{\mathbf{B}}_{PLSR} = \mathbf{R}\mathbf{Q}^T$, the i^{th} row of $\widehat{\mathbf{B}}_{PLSR}$ can be rewritten as $\mathbf{b}_{(i)}^T = \mathbf{r}_{(i)}^T\mathbf{Q}^T$, akin to $\mathbf{y}_0^T = \mathbf{t}^T\mathbf{Q}^T$, so that the regression coefficients are predicted by the response variable prediction biplot axes. Instead of predicting a sample point $\mathbf{t}^T = (\mathbf{x} - \bar{\mathbf{x}})_i^T\mathbf{R}$, $\mathbf{e}_i^T\mathbf{R}$ are projected onto these axes. That is, projecting each of the rows of $\mathbf{R}$ onto the prediction axes defined by $\mathbf{Q}$ yields the estimated PLSR coefficients matrix $\widehat{\mathbf{B}}_{PLSR}$. Note that this is true without adjusting for centring, so here $\mu_b = \mu_{0Y}$ and not $\mu_Y = \mu_{0Y} + \bar{y}_k$. Therefore two sets of prediction marker calibrations are needed, a set corrected for the mean for $\mathbf{Y}$ and a set in terms of the centred values for $\widehat{\mathbf{B}}_{PLSR}$.

3.3 Measure of fit for a PLS biplot

The quality of the representation provided by the PLS biplot is required so as to determine the adequacy of the representation of the original data matrix. With the

original data matrix $\mathbf{D}$ defined as $\mathbf{D} = [\mathbf{X} \ \ \mathbf{Y}]$, the approximated data matrix can be written as $\widehat{\mathbf{D}} = [\widehat{\mathbf{X}} \ \ \widehat{\mathbf{Y}}] = \mathbf{T}[\mathbf{P} \ \ \mathbf{Q}]^{\mathrm{T}}$, for $\widehat{\mathbf{X}} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \mathbf{1}\bar{\mathbf{x}}$ and $\widehat{\mathbf{Y}} = \mathbf{T}\mathbf{Q}^{\mathrm{T}} + \mathbf{1}\bar{\mathbf{y}}$, where $\mathbf{1}: \mathrm{N} \times 1$ a vector of ones, $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$ are the column means of $\mathbf{X}$ and $\mathbf{Y}$ respectively. Overall, the quality of representation can be measured in terms of the percentage of variation in $\mathbf{D}$ explained by the extracted latent variables $\mathbf{T}$. That is,

$$\text{Overall quality} = \text{sum}\left(\text{diag}\left(\widehat{\mathbf{D}}^{\mathrm{T}}\widehat{\mathbf{D}}\right)\right)\left[\text{sum}\left(\text{diag}\left(\mathbf{D}^{\mathrm{T}}\mathbf{D}\right)\right)\right]^{-1}.$$

4. A SMALL EXAMPLE

The following example is an illustration of a PLS biplot using the Olive Oil data from Mevik & Wehrens (2007). This data shows the sensory and chemical quality evaluation of sixteen olive oil samples. There were five chemical quality measurements (Acidity, Peroxide, K232, K270 and DK) taken and six sensory panel characteristics (Yellow, Green, Brown, Glossy, Transp and Syrup) used in this evaluation. The sixteen olive oils are assigned as the samples. The chemical quality measurements and the sensory panel characteristics are the predictor and response variables respectively. As a result, the olive oil data can be viewed as a data matrix $\mathbf{X}: 16 \times 5$ of predictors and a matrix $\mathbf{Y}: 16 \times 6$ of responses.

A 2-component PLS was performed and the asymmetric PLS biplot is shown in Figure **1**. With an overall quality of 0.973, various relationships can be deduced from this biplot such as a relation between the response Syrup and predictors K232 and Peroxide. Acidity can be seen to have no clear relation with the others. Each sample point in the PLS biplot is orthogonally projected onto the axes and the respective values read off to give the approximated values of the olive oil data. For example, sample point G5 projected onto the Acidity and Syrup axes yields the values 0.51 and 46.7 respectively as shown in Figure **2**. Similarly, the PLSR coefficient values bi, for $\mathrm{i} = 1, 2, \cdots, 5$, are projected onto the axes representing the sensory panel characteristics. However, the purple markers are used to read off these values. For example, point b4 projected onto the Syrup axis gives a value of 1.2 as shown in Figure **3**. The predicted regression coefficient values as well as the predicted values of the olive oil data are shown in Tables **1** and **2** respectively. The orthogonal latent variables $\mathbf{T}$ are shown in Table **3**.

5. CONCLUSION

When there is the need to predict a set of response variables from a (very) large set of predictor variables and there is no practical need to limit the number of predictor variables, PLSR can be a useful tool. Results found by the PLSR analysis of a data set can be visualized graphically using biplots. A PLS biplot provides a single graphical representation of the samples together with the predictor and response variables, as well as their interrelationships.

SOFTWARE

A collection of functions has been developed in the R language (R Development Core Team, 2012) to produce the PLS biplot. These functions are available electronically upon request.

ACKNOWLEDGEMENTS

Associate Professor S. Gardner-Lubbe is thanked for the valuable suggestions that have helped in the improvement of this paper. Financial support from the University of Cape Town Postgraduate Centre and Funding Office is gratefully acknowledged.

Table 1 The predicted PLSR coefficient values.

	Yellow	**Green**	**Brown**	**Glossy**	**Transp**	**Syrup**
b1: **Acidity**	-1.175	1.343	-1.112	-0.357	-0.604	-0.119
b2: **Peroxide**	-0.921	0.675	2.379	-1.638	-1.364	1.862
b3: **K232**	-1.169	0.977	1.984	-1.646	-1.448	1.741
b4: **K270**	-1.365	1.301	0.942	-1.348	-1.314	1.206
b5: **DK**	-1.042	1.071	0.046	-0.748	-0.818	0.515

Table 2 The predicted olive oil values.

	Acidity	**Peroxide**	**K232**	**K270**	**DK**	**Yellow**	**Green**	**Brown**	**Glossy**	**Transp**	**Syrup**
G1	0.77	12.7	1.820	0.153	0.001	23.0	68.9	9.4	77.1	71.8	48.5
G2	0.22	12.3	1.620	0.107	-0.003	58.7	24.3	11.6	82.8	81.0	47.1
G3	0.32	10.3	1.520	0.108	-0.003	57.2	27.5	8.8	84.3	82.4	46.0
G4	0.61	13.7	1.830	0.143	0.000	31.2	57.9	11.4	77.5	72.9	48.9
G5	0.47	11.2	1.630	0.123	-0.001	45.2	41.9	9.0	81.8	78.6	46.9
I1	0.29	18.7	2.050	0.137	0.000	39.7	44.0	19.0	74.4	70.6	51.7
I2	0.27	15.3	1.830	0.123	-0.001	48.8	34.8	15.0	78.7	75.9	49.3
I3	0.24	18.5	2.020	0.132	-0.001	43.6	39.2	19.0	75.1	71.8	51.4
I4	0.33	15.6	1.860	0.128	-0.001	44.7	39.8	15.1	77.8	74.6	49.6
I5	0.26	19.4	2.090	0.137	0.000	40.2	43.0	20.0	73.8	70.1	52.1
S1	0.17	10.5	1.490	0.097	-0.004	66.0	16.3	9.7	85.5	84.4	45.8
S2	0.25	8.2	1.360	0.094	-0.004	66.4	17.3	6.5	87.6	86.6	44.3
S3	0.23	12.5	1.630	0.109	-0.003	57.9	25.1	11.8	82.5	80.6	47.3
S4	0.19	11.0	1.520	0.100	-0.003	64.0	18.5	10.3	84.8	83.4	46.2
S5	0.15	10.8	1.500	0.096	-0.004	66.6	15.3	10.2	85.3	84.3	45.9
S6	0.22	11.4	1.560	0.104	-0.003	60.7	22.4	10.5	83.9	82.3	46.5

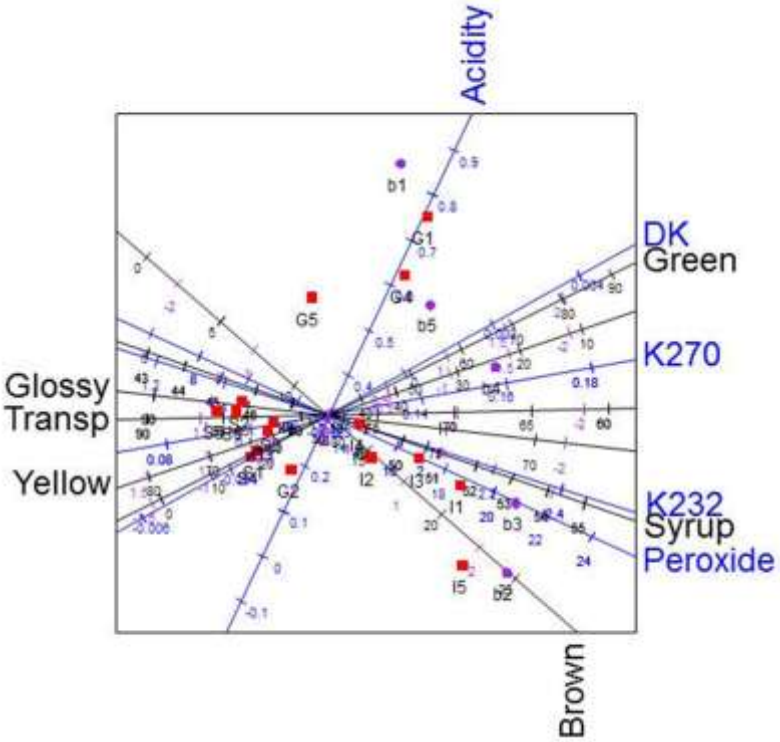

Figure 1 The PLS biplot of the olive oil data.

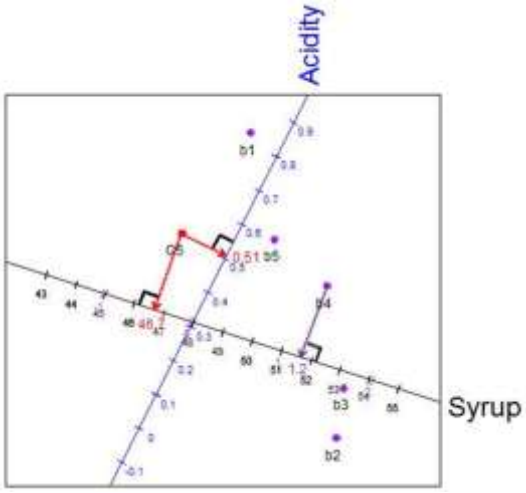

Figure 2 Examples of orthogonal projections.

Table 3 The orthogonal latent variables **T**.

	Component 1	**Component 2**
G1	-0.039	0.709
G2	-0.074	-0.135
G3	-0.227	0.020
G4	0.036	0.452
G5	-0.158	0.258
I1	0.421	-0.050
I2	0.158	-0.070
I3	0.404	-0.132
I4	0.181	0.013
I5	0.475	-0.102
S1	-0.213	-0.206
S2	-0.394	-0.078
S3	-0.059	-0.127
S4	-0.174	-0.186
S5	-0.191	-0.237
S6	-0.144	-0.129

REFERENCES

Barnett, V. (1981) *Interpreting Multivariate Data*, Wiley Series in Probability and Mathematical Statistics.

Eckart, C. and Young, G. (1936) "The Approximation of One Matrix by Another of Lower Rank," *Psychometrika*, 1, 211-218.

Gower, J.C. and Hand, D.J. (1996) *Biplots*, *Monographs on Statistics and Applied Probability 54*, Chapman & Hall, UK.

Martins, J.P.A., Teófilo, R.F. and Ferreira, M.M.C. (2010) "Computational Performance and Cross-Validation Error Precision of Five PLS Algorithms using Designed and Real Data Sets," *Journal of Chemometrics*, 24, 320-332.

Mevik, B.H. and Wehrens, R. (2007) "The pls Package: Principal Component and Partial Least Squared Regression in R," *Journal of Statistical Software*, 2(18), 1-24.

R Development Core Team (2012) *R: A Language and Environment for Statistical Computing*, the R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.