Available online at www.sciencedirect.com
SCIENCE DIRECT®

Computational Statistics & Data Analysis 48 (2005) 103 – 123

COMPUTATIONAL STATISTICS & DATA ANALYSIS

www.elsevier.com/locate/csda

Regression of a data matrix on descriptors of both its rows and of its columns via latent variables: L-PLSR

Harald Martens[a,b,c,*], Endre Anderssen[b], Arnar Flatberg[b], Lars Halvor Gidskehaug[b], Martin Høy[b], Frank Westad[c], Anette Thybo[d], Magni Martens[a]

[a] *The Royal Veterinary and Agricultural University, DK-1958 Frederiksberg, Denmark*
[b] *The Norwegian University of Science and Technology, N-7034 Trondheim, Norway*
[c] *Matforsk, The Norwegian Food Research Institute, Oslov. 1, N-1430 Ås, Norway*
[d] *Danish Institute of Agricultural Sciences, Department of Horticulture, DK-5792 Aarslev, Denmark*

Received 9 October 2003; received in revised form 9 October 2003; accepted 10 October 2003

Abstract

A new approach is described, for extracting and visualising structures in a data matrix $\mathbf{Y}$ in light of additional information BOTH about the ROWS in $\mathbf{Y}$, given in matrix $\mathbf{X}$, AND about the COLUMNS in $\mathbf{Y}$, given in matrix $\mathbf{Z}$. The three matrices $\mathbf{Z}$–$\mathbf{Y}$–$\mathbf{X}$ may be envisioned as an "L-shape"; $\mathbf{X}(I \times K)$ and $\mathbf{Z}(J \times L)$ share no matrix size dimension, but are connected via $\mathbf{Y}(I \times J)$. A few linear combinations (components) are extracted from $\mathbf{X}$ and from $\mathbf{Z}$, and their interactions are used for bi-linear modelling of $\mathbf{Y}$, as well as for bi-linear modelling of $\mathbf{X}$ and $\mathbf{Z}$ themselves. The components are defined by singular value decomposition (SVD) of $\mathbf{X}'\mathbf{YZ}$. Two versions of the L-PLSR are described—using one single SVD for all components, or component-wise SVDs after deflation.

The method is applied to the analysis of consumer liking data $\mathbf{Y}$ of six products assessed by 125 persons, in light of 10 other product descriptors $\mathbf{X}$ and 15 other person descriptors $\mathbf{Z}$. Its performance is also checked on artificial data.

* Corresponding author. Matforsk, The Norwegian Food Research Institute, Oslov. 1, N-1432 Aas, Norway. Tel.: +47-64970100; fax: +47-6470333.

E-mail address: harald.martens@matforsk.no (H. Martens).

0167-9473/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2003.10.004

Keywords: Multivariate; Regression; Partial least squares; PLS; PLSR; L-PLSR; Bookstein; Bi-linear model; Eigenvector; Singular value decomposition; SVD; Three-block; Latent variables

1. Introduction

1.1. A two-way table with two extra tables

Traditionally, good science demanded a one-to-one relationship between a cause and a measurement. That tradition can now be a hindrance to the study of more complex systems, which are characterised by many-to-many relationships, often hidden in big data tables. To find these requires humble, but aggressive explorative data analysis with powerful structure extraction- and display-tools, not just traditional testing of hypotheses.

It is common practice to organise empirical data in a two-way data table, $\mathbf{Y}$. These data may be interpreted in light of other descriptors of its *rows*, organised in data table $\mathbf{X}$ with the same number of rows as $\mathbf{Y}$. Good multivariate methods have been developed for relating *two* such tables to each other, for prediction *and* interpretation (see e.g. Martens and Naes, 1989; Martens and Martens, 2001). For instance, if $\mathbf{X}$ and/or $\mathbf{Y}$ have strongly inter-correlated columns, a method like the PLS Regression (Wold et al., 1983) utilises this multi-collinearity as a stabilising advantage in a linear/bi-linear model. In this journal the method was recently assessed (Elden, 2004), and applied, e.g. to biotechnology data (Nguyen and Rocke, 2004).

However, the data table $\mathbf{Y}$ may also have descriptors of its *columns*, organised in a *third* data table $\mathbf{Z}'$ with the same number of columns as $\mathbf{Y}$. The question is how to utilise the information in all three matrices $\mathbf{X}, \mathbf{Y}$ and $\mathbf{Z}$ efficiently, in a way that is interpretable and statistically stabile.

1.2. Example of L-shaped input data tables

Consumer studies represent an application field where such "L-shaped" data matrix structures $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ are common: A set of I products has been assessed by a set of J consumers, e.g. with respect to liking, with results collected in "liking" data table $\mathbf{Y}(I \times J)$. In addition, each of the I products has been "measured" by K product descriptors ("$\mathbf{X}$-variables"), reflecting chemical or physical measurements, sensory descriptions, production facts etc., in data table $\mathbf{X}(I \times K)$. Moreover, each of the J consumers have been characterised by L consumer descriptors ("$\mathbf{Z}$-variables"), comprising sociological background variables like gender, age, income, etc., as well as the individual's general attitude and consumption patterns; these are collected in data table $\mathbf{Z}(J \times L)$. Relevant questions could then be: Is it possible to find reliable patterns of variation in the liking data $\mathbf{Y}$, which can be explained from both product descriptors $\mathbf{X}$ and from consumer descriptors $\mathbf{Z}$? Is it possible to predict how a new product will be liked by these consumers, by measuring its $\mathbf{X}$-variables? Is it possible to predict how a new consumer group will like these products, from their background $\mathbf{Z}$-variables?

Consumer response data are of very low precision. With so noisy data it is important to have a modelling method that only extracts the most dominant, relevant and reliable structures from the data: the number of model parameters to be estimated independently from the response data $\mathbf{Y}$ must be low. The level of statistical validity of the obtained model parameters can be checked, e.g. by various cross-validation /jack-knifing schemes, keeping different consumers and/or products "secret", in turn, for model testing. This can also reveal outliers and gross errors in the data. For simplicity, statistical validation will not be pursued here.

Moreover, consumer studies are wrought with methodological problems that may give systematic errors of various kinds. Therefore, it is important to have efficient graphical displays to inspect and interpret the structures obtained from the data.

In the present case, Danish children's liking of apples is being studied. Their response to various apple types is termed $\mathbf{Y}$. Chemical, physical and sensory descriptors of these apple types are called $\mathbf{X}$, and sociological and attitude descriptors on these children is called $\mathbf{Z}$. The purpose of the analysis is to find patterns in these $\mathbf{X}/\mathbf{Y}/\mathbf{Z}$ data that are causally interpretable and have predictive reliability.

In the following, matrices will be written in upper-case (e.g. $\mathbf{X}$) letters, vectors in lower-case (e.g. $x_k, k=1,2,\ldots,K$) and scalar elements in italics (e.g. $x_{ik}, i=1,2,\ldots,I; k=1,2,\ldots,K$); all vectors are column vectors unless otherwise specified.

1.3. Modelling alternatives

Since $\mathbf{X}(I\times K), \mathbf{Y}(I\times J)$ and $\mathbf{Z}(J\times L)$ have different size dimensions, the $\mathbf{Z}$-variables cannot be modelled with $\mathbf{X}$ and $\mathbf{Y}$ by regression over I objects in today's two-block bi-linear model framework.

A two-step approach for using the $\mathbf{Z}$-information is to fold the information from $\mathbf{Z}$ with $\mathbf{Y}'$ in order to give the $\mathbf{Z}$-information a dimension I, common with $\mathbf{Y}$ and $\mathbf{X}$. Kubberød et al. (2002) first estimated the reduced-rank regression coefficient matrix $\mathbf{B}'_{Z,Y}(I \times L)$ from the mean-centred model $\mathbf{Y}' \approx \mathbf{Z} \cdot \mathbf{B}_{Z,Y}$ by PLSR, and secondly regressed both $\mathbf{Y}$ and $\mathbf{B}'_{Z,Y}$ on $\mathbf{X}$ based on the linear model $[\mathbf{Y}, \mathbf{B}'_{Z,Y}] \approx \mathbf{X} \cdot \mathbf{B}_X$, in another reduced-rank PLSR step. Thybo et al. (2003) used a similar two-step approach, but in order to simplify the analysis, they replaced the regression coefficient matrix $\mathbf{B}'_{Z,Y}$ by the matrix of correlation coefficients $\mathbf{R}_{Z,Y}(I \times L)$ between the I rows in $\mathbf{Y}$ and the L columns in $\mathbf{Z}$, correlated over J elements. In either case, the $\mathbf{X}/\mathbf{Y}/\mathbf{Z}$ interpretations were meaningful. But the two-step procedures are cumbersome and have complicated mean-centring properties.

Another approach could be to unfold the two-way $\mathbf{Y}(I \times J)$ into a vector $\mathbf{y}$ with $N = I \cdot J$ elements, and submit this as regressand to a regression model with $\mathbf{X}$ and $\mathbf{Z}$ as regressors, using the additive model $y_{ij} = \mu + \sum_{k=1}^{K} x_{ik}\beta_\kappa + \sum_{l=1}^{L} z_{jl}\gamma_l + f_{ij}$. This assumes independent $\mathbf{X}$ and $\mathbf{Z}$ contributions to $\mathbf{y}$. In the present application that would preclude consumer/product segmentation: A product property x_{ik} would then be assumed to have the same impact for all consumers ("when an apple i is very red, then it is very much liked by everybody"), and a consumer consumption descriptor z_{jl} would mean the same for all products ("when a person j says he particularly likes red apples, then that person likes all the apples"). Moreover, collinearity between the

X-variables and between the **Z**-variables would give estimation problems, if a full-rank regression method had been used.

A multiplicative model like $y_{ij} = \mu + (\sum_{k=1}^{K} x_{ik}\beta_{\kappa})(\sum_{l=1}^{L} z_{jl}\gamma_l) + f_{ij}$ could be an alternative. Here **y** is modelled as the bi-linear product between a linear combination of the **X**-variables and a linear combination of the **Z**-variables. This model would only allow *one* pattern of **Y**-relevant variation among the **X**-variables, defined by $\beta_{\kappa}, k = 1, 2, \ldots, K$, and one pattern of **Y**-relevant variation among the **Z**-variables, defined by $\gamma_l,\ l = 1, 2, \ldots, L$. That is an unnatural restriction, because consumers may differ in more than one way. A generalisation of this approach will be presented here: **Y** is approximated by a bi-linear model, which is defined as the sum of the interactions between *several* orthogonal linear combinations of **X** and of **Z**. These linear combinations represent **Y**-relevant latent variables in **X** and in **Z**, defined according to the PLS principle of maximising explained covariance. The present approach may be seen as an extension of an earlier, iterative NIPALS-based method presented by Wold et al. (1987). Höskuldsson (2001) outlines other ways to combine three or more matrices within the PLS framework. The first author's account of the development of the two-block PLS regression is given in Martens (2001).

2. Materials and methods

2.1. Experimental input data

2.1.1. $I = 6$ products

The data are taken from Thybo et al. (2003). $I = 6$ products were the apple cultivars "*Jonagold*", "*Mutsu*", "*Gala*", "*Gloster*", "*Elstar*" and "*GrannySmith*". All cultivars were selected due to commercial relevance for the Danish market and due to the fact that the cultivars were known to span a large variation in sensory quality (Kühn and Thybo, 2001). *Gloster* was chosen as a wine-red cultivar with particularly high glossiness, *Gala* and *Jonagold* as red cultivars with 80–90% red flushed surface, *Mutsu* as a yellow–green cultivar and *GrannySmith* as a green and particularly round shaped cultivar. *GrannySmith* was known to be a rather popular cultivar for some children, due to its texture and moistness characteristics. Only apples with shape and colour deemed representative for their cultivar were used.

2.1.2. $K = 10$ *product descriptors* **X**

Sensory profile descriptors: A panel of ten assessors was trained in quantitative descriptive analysis of apple types as described in Kühn and Thybo (2001). Conventional statistical design w.r.t. replication and serving order was applied. The panel average of a subset of the appearance, texture, taste and flavour descriptors determined will be used here: *RED, SWEET, SOUR, GLOSSY, HARD* and *ROUND*.

Chemical and instrumental product descriptors: Texture firmness was evaluated instrumentally by penetration (*FIRM*, *INSTR.*). Content of acid (*ACIDS*) and sugar (*SUGARS*) were determined as malic acid and soluble solids, respectively. Based on prior theory on human sensation of sourness, the ratio $ACIDS/SUGARS$ was included as a separate variable (Kühn and Thybo, 2001).

Together, the sensory, chemical and instrumental variables constituted $K=10$ product descriptors, which will here be referred to as $\mathbf{X}(I \times K)$ for the $I=6$ products.

2.1.3. $J=125$ consumers

The consumers were children aged 6–10 years (51% boys, 49% girls), recruited from a local elementary school. A total of 146 children were tested and included in the original publication (Thybo et al., 2003). For simplicity, only the $J=125$ children that had no missing values in their liking and background data are included in the present study.

2.1.4. $L=15$ consumer descriptors $\mathbf{Z}$

First, each child was asked to look at a table with five different fruits (a red and a green apple, a banana, a pear and an orange (mandarin)), and answer the questions: "If you were asked to eat a fruit, which fruit would you then choose, and which fruit would be your last choice?" The resulting responses will here be named "$\langle fruit \rangle First$" and "$\langle fruit \rangle$ *Last*", where $\langle fruit \rangle$ is one of [*Red Apple*, *Green Apple*, *Pear*, *Banana*, *Orange* or *Apple*]. (Summaries were later computed for apple liking: $AppleFirst = RedAppleFirst + GreenAppleFirst$ and $AppleLast = RedAppleLast + GreenAppleLast$.) The child was also questioned about how often he/she ate apples, by having the following opportunities: "every day" (here coded as value 4), "a couple of times weekly" (3), "a couple of times monthly" (2), "very seldom" (1); this descriptor is here named "*EatAOften*". (A few of the children responded "do not know" to how often he/she ate apples. To reduce the number of missing values, this was for simplicity taken as indicating very low apple consumption, and coded as 0.) In addition, the child's *gender* and *age* were noted. These two sociological descriptors were used, together with the attitude variables $\langle fruit \rangle$ *First* and $\langle fruit \rangle$ *Last* and eating habit-variable *EatAOften*, as $L=15$ consumer background descriptors $\mathbf{Z}(J \times L)$ for the $J=125$ children.

2.1.5. Product liking $\mathbf{Y}$

Then each child was asked to express the liking of the appearance of the six apple cultivars, using a five-point hedonical facial scale: the scale expressed: 1 ="not at all like to eat it", 2 ="not like to eat it", 3 ="it is okay", 4 ="like to eat it", 5 ="very much like to eat it". One apple at a time was shown to the child to avoid that the child concentrated on comparing the appearances. All samples were presented in randomised order. The resulting *liking* data for the $I=6$ products $\times$ $J=125$ consumers will here be termed $\mathbf{Y}(I \times J)$.

2.1.6. The three-block input data

Fig. 1 depicts the actual data in this example, with the three input data tables or "blocks", product descriptors $\mathbf{X}(I \times K)$, liking data $\mathbf{Y}(I \times J)$ and consumer descriptors $\mathbf{Z}(J \times L)$. Since $\mathbf{X}$ and $\mathbf{Z}'$ each share one of the dimensions with $\mathbf{Y}$, and none with each other, the data tables form an "L". The new PLSR method to be presented

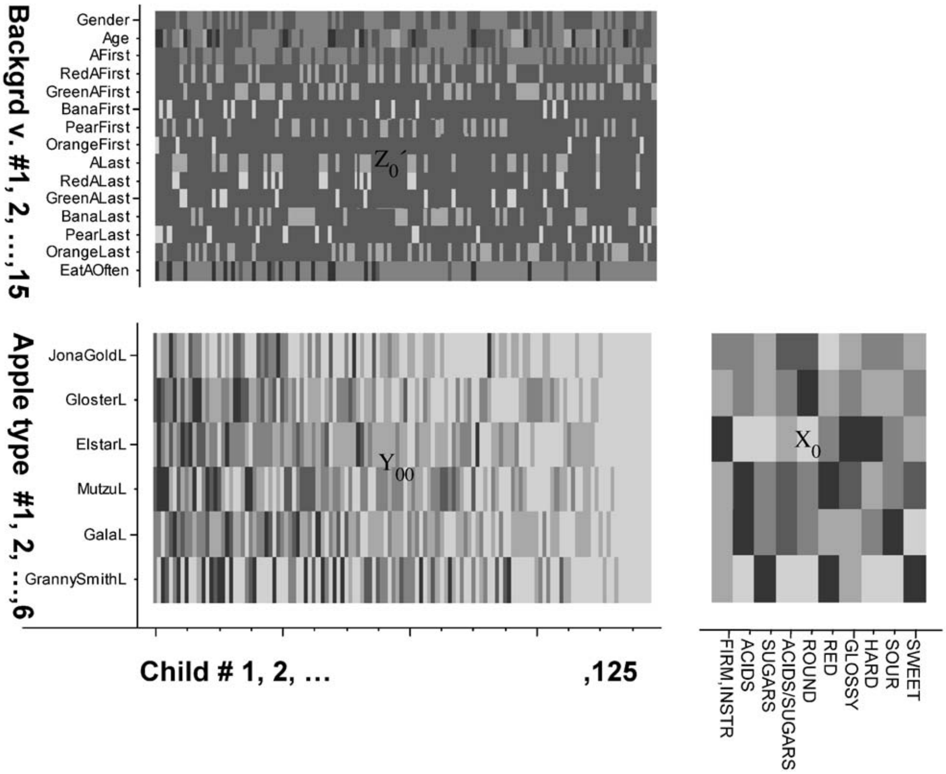

Fig. 1. Overview of input data tables **X**, **Y** and **Z**. **X**: 10 sensory and chemical/instrumental descriptors (standardised) of six apple types; dark/light = −1.8/1.6. **Y**: 125 children's liking (1–5) of six apple types; dark/light = 1/5. **Z**′ : 15 background descriptors (standardised) for 125 children; dark/light = −2.5/3.6.

is therefore named "L-PLSR". In this particular data sets the input dimensions are $\mathbf{X}(6 \times 10), \mathbf{Y}(6 \times 125)$ and $\mathbf{Z}'(15 \times 125)$.

2.1.7. Software

The software for L-PLSR and correlation loadings was written in Matlab™ (www.mathworks.com). The remaining plots were drawn using The Unscrambler™ (www.camo.com).

2.2. Theory

2.2.1. Pre-processing: mean centring and re-scaling

Since the K input variables in $\mathbf{X}(I \times K)$ are given in units with different offsets, they are brought to a common origin by mean centring over the I rows. Likewise, the L input variables in $\mathbf{Z}(J \times L)$ are brought to a common origin by mean centring over the J rows:

$$\mathbf{X}_0 = \mathbf{X} - \mathbf{1}\bar{\mathbf{x}}', \quad \mathbf{Z}_0 = \mathbf{Z} - \mathbf{1}\bar{\mathbf{z}}'. \tag{1a}$$

$\mathbf{Y}(I \times J)$ is then to be modelled from interactions of the mean centred X- and **Z**-data. The column- and row-means in **Y** will automatically be ignored, since these mean vectors will be orthogonal to $\mathbf{X}_0$ and $\mathbf{Z}_0$, respectively. Therefore **Y** is subjected to conventional double mean-centring:

$$\mathbf{Y}_{00} = \mathbf{Y} - \mathbf{1}_I \bar{\mathbf{y}}'_{.J} - \bar{\mathbf{y}}_{I.} \mathbf{1}'_J + \mathbf{1}_I \bar{\bar{y}} \mathbf{1}'_J, \tag{1b}$$

where $\bar{\mathbf{y}}_{I.}(I \times 1)$ contains the mean for each of the I rows (the "average consumer", in plots named "*RowsMean*"), $\bar{\mathbf{y}}_{.J}(J \times 1)$ contains the mean for each of the J columns (the "average product", "*ColsMean*"), and $\bar{\bar{y}}$ is the grand mean of **Y**.

The column- and row-mean vectors $\bar{\mathbf{y}}_{.J}$ and $\bar{\mathbf{y}}_{I.}$ in this data set represent the average liking score for each of the I products and for each of the J consumers, respectively. They will be studied graphically as if they represented an extra "**Y**-column" (the "average consumer") and "**Y**-row" (the "average product") in the final model overview.

Since the different **X**-variables are given in units with very different ranges, they will be rescaled, e.g. by conventional standardisation to a common total initial standard deviation of 1. The **Z**-variables will likewise be standardised, while the **Y**-data are left unscaled.

2.2.2. Model overview

Fig. 2 gives an overview of the matrices involved in the three-block bi-linear model. The three input matrices are shown as filled rectangles, the main and additional parameter matrices as dense and dotted rectangle outlines, respectively. The L-PLSR estimation algorithm is outlined by arrows.

The L-PLSR is primarily intended to reveal patterns in **Y** that correspond to patterns in both **X** and **Z**, after mean centring. This restrictive condition is intended to act as an efficient filter against random noise in the **Y**— on the condition that both **X** and **Z** span the interesting variation types. If such **X**/**Y**/**Z** patterns are found, they may be used for prediction of future **Y** from more easily available information **X** and **Z**, as well as for graphical interpretation of the likely **X**/**Y**/**Z** causality.

Note how **X** and **Z** share no size dimension; **Y** acts as an "instrumental matrix" to connect them. In the present data example, the goal is to find related **Y**-relevant patterns among the product descriptors **X** and among the consumer descriptors **Z**, and to approximate **Y** from these. The random noise in consumer liking data **Y** is expected to be high, but it is hoped that the restrictive L-PLSR model will stabilise the modelling. It is assumed that both the K variables in **X** and the L variables in **Z** have been chosen so that they have a good chance of spanning the interesting types of variations in **Y**. If such **X**/**Y**/**Z** relationships in fact do exist, then the method should be able to reveal consumer segments with different product preference patterns, and indicate which product properties and which consumer descriptors point to these systematic differences.

2.2.3. Three-block bi-linear structure model

Full-rank interaction model of **Y**: A rather restrictive approximation model may be written as

$$\mathbf{Y}_{00} = \mathbf{X}_0 \mathbf{C} \mathbf{Z}'_0 + \mathbf{F}, \tag{2}$$

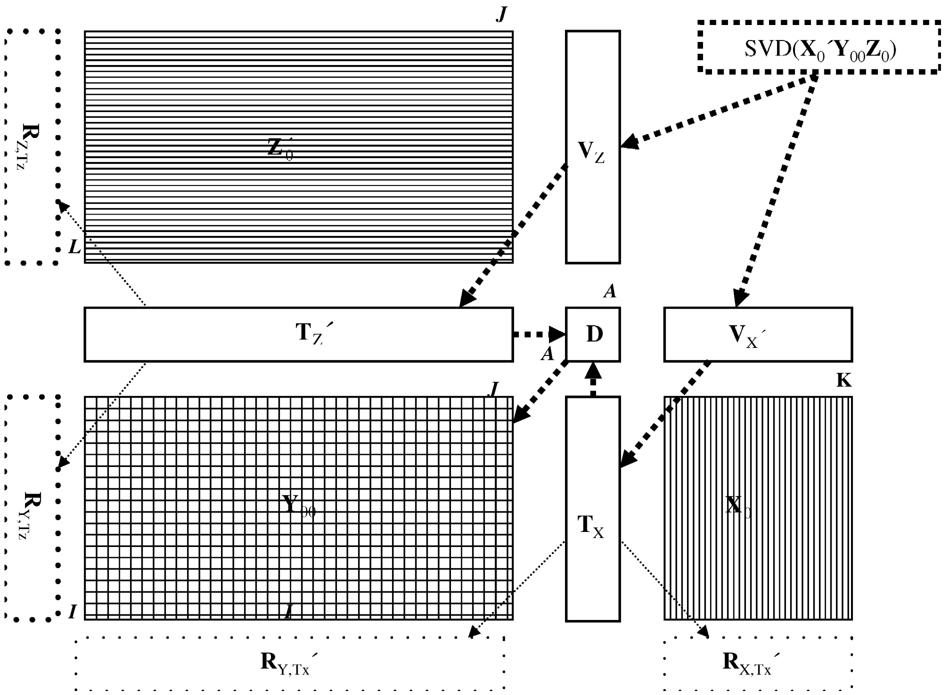

Fig. 2. L-PLS regression method. Mean-centred input data $\mathbf{Y}_{00}$, $\mathbf{X}_0$ and $\mathbf{Z}_{0,}$ model structure (weights $\mathbf{V}_X$ and $\mathbf{V}_Z$, scores $\mathbf{T}_X$ and $\mathbf{T}_Z$) and estimation algorithm: singular value decomposition(s) of $\mathbf{X}_0'\mathbf{Y}_{00}\mathbf{Z}_0$ (with or without deflation) yield $\mathbf{V}_X$ and $\mathbf{V}_Z$ which define scores as $\mathbf{T}_X = \mathbf{X}_0\mathbf{V}_X$ and $\mathbf{T}_Z = \mathbf{Z}_0\mathbf{V}_Z$. For graphical display, correlation loadings $[\mathbf{R}_{X,T_X},\ \mathbf{R}_{Y,T_X}]$ are obtained by simple correlations between $[\mathbf{X},\mathbf{Y}]$ and $\mathbf{T}_X$, and $[\mathbf{R}_{Z,T_Z},\ \mathbf{R}_{Y,T_Z}]$ by simple correlation between $[\mathbf{Z},\mathbf{Y}']$ and $\mathbf{T}_Z$.

where matrix $\mathbf{C}(K \times L)$ represents the $\mathbf{Y}$-relevant $\mathbf{X}/\mathbf{Z}$ interactions and $\mathbf{F}(I \times J)$ the $\mathbf{Y}$-residuals. If the K $\mathbf{X}$-variables had been linearly independent and the L $\mathbf{Z}$-variables had been linearly independent, the interactions parameters might have been estimated by ordinary least-squares regression. However, usually neither $\mathbf{X}$ nor $\mathbf{Z}$ have full column rank, and full-rank least squares solution of Eq. (2) then leads to numerical problems or statistical variance inflation.

Reduced-rank bi-linear model of $\mathbf{Y}$: Assume instead a reduced-rank version of Eq. (2)

$$\mathbf{Y}_{00} = \mathbf{X}_0\mathbf{C}_A\mathbf{Z}_0' + \mathbf{F}_A, \tag{3a}$$

where, for simplicity, only the A most important underlying variation types are being modelled, by $\mathbf{C}_A(K \times L)$, leaving the rest as unmodelled residuals in $\mathbf{F}_A(I \times J)$. The elements in $\mathbf{C}_A$ are to be estimated, to rank A, by making the elements in $\mathbf{F}_A$ "small" in some way; the details will depend on the methodology chosen. Here it will be implemented as an extension of the well-established two-block rank-reduced bi-linear regression modelling (see e.g. Martens and Martens, 2001), in terms of A latent variables (linear combinations or components) from $\mathbf{X}$ and from $\mathbf{Z}$:

$$\mathbf{T}_X = \mathbf{X}_0\mathbf{V}_{X,A}, \quad \mathbf{T}_X = \mathbf{Z}_0\mathbf{V}_{Z,A}, \tag{3b}$$

where $\mathbf{V}_{X,A}(K \times A)$ and $\mathbf{V}_{Z,A}(L \times A)$ are the component weight matrices, and $\mathbf{T}_X(I \times A)$ and $\mathbf{T}_Z(J \times A)$ are the resulting component score matrices in $\mathbf{X}$ and $\mathbf{Z}$, respectively. The bi-linear $\mathbf{X}/\mathbf{Z}$ model of $\mathbf{Y}$ in Eq. (3a) may then be rewritten as

$$\mathbf{Y}_{00} = \mathbf{T}_X\mathbf{D}_A\mathbf{T}_Z' + \mathbf{F}_A, \tag{3c}$$

where the elements $\mathbf{D}_A(A \times A)$ describe the $\mathbf{Y}$-relevant interaction structures between $\mathbf{X}$ and $\mathbf{Z}$. The number of components from both $\mathbf{X}$ and $\mathbf{Z}$ is assumed to be the same; this is done for convenience and is not a necessity.

The component-wise estimate of $diag(\mathbf{D}_A)$ at rank A is based on the model $\mathbf{Y}_{00} = \mathbf{t}_{X,a}d_{a,a}\mathbf{t}_{Z,a}' + \mathbf{F}_a, a = 1, 2, \ldots, A$:

$$d_{a,a} = (\mathbf{t}_{X,a}'\mathbf{t}_{X,a})^{-1}\mathbf{t}_{X,a}'\mathbf{Y}_{00}\mathbf{t}_{Z,a}(\mathbf{t}_{Z,a}'\mathbf{t}_{Z,a})^{-1}, \tag{3d}$$

while a full matrix, including the non-diagonal elements, is

$$\mathbf{D}_A = (\mathbf{T}_X'\mathbf{T}_X)^{-1}\mathbf{T}_X'\mathbf{Y}_{00}\mathbf{T}_Z(\mathbf{T}_Z'\mathbf{T}_Z)^{-1}. \tag{3e}$$

If $\mathbf{D}_A$ is found to be highly non-diagonal, it may be further diagonalized by svd, with a corresponding rotation of $\mathbf{T}_X$ and $\mathbf{T}_Z$ but this is not pursued here.

By inserting the definitions of $\mathbf{T}_X$ and $\mathbf{T}_Z$ (Eq. (3b) into Eq. (3c)), the definition of the $\mathbf{X}$–$\mathbf{Z}$ interaction matrix $\mathbf{C}_A(K \times L)$ in Eq. (3) is obtained:

$$\mathbf{C}_A = \mathbf{V}_{X,A}\mathbf{D}_A\mathbf{V}_{Z,A}' \tag{3f}$$

with $\mathbf{Y}$-residuals

$$\mathbf{F}_A = \mathbf{Y}_{00} - \mathbf{T}_X\mathbf{D}_A\mathbf{T}_Z' = \mathbf{Y}_{00} - \mathbf{X}_0\mathbf{C}_A\mathbf{Z}_0'. \tag{3g}$$

Reduced-rank interaction model of $\mathbf{X}$ *and* $\mathbf{Z}$: The A components from $\mathbf{X}$ and $\mathbf{Z}$ may be used to model $\mathbf{X}$ and $\mathbf{Z}$ themselves:

$$\mathbf{X}_0 = \mathbf{T}_X\mathbf{P}_X' + \mathbf{E}_{X,A}, \quad \mathbf{Z}_0 = \mathbf{T}_Z\mathbf{P}_Z' + \mathbf{E}_{Z,A}, \tag{4a}$$

where the $\mathbf{X}$- and $\mathbf{Z}$-loadings $\mathbf{P}_X(K \times A)$ and $\mathbf{P}_Z(L \times A)$ are obtained by projection of $\mathbf{X}$ and $\mathbf{Z}$ on their respective score matrices

$$\mathbf{P}_X = \mathbf{X}_0'\mathbf{T}_X(\mathbf{T}_X'\mathbf{T}_X)^{-1}, \mathbf{P}_Z = \mathbf{Z}_0'\mathbf{T}_Z(\mathbf{T}_Z'\mathbf{T}_Z)^{-1} \tag{4b}$$

and the corresponding residuals $\mathbf{E}_{X,A}(I \times K)$ and $\mathbf{E}_{Z,A}(J \times L)$ obtained as

$$\mathbf{E}_{X,A} = \mathbf{X}_0 - \mathbf{T}_X\mathbf{P}_X', \quad \mathbf{E}_{Z,A} = \mathbf{Z}_0 - \mathbf{T}_Z\mathbf{P}_Z'. \tag{4c}$$

Extending the interaction model of $\mathbf{Y}$ *with additive terms*: Even though the present focus is on extracting $\mathbf{Y}$-structures that are seen both in $\mathbf{X}$ AND $\mathbf{Z}$, it is of course possible that $\mathbf{Y}$ may have some variation that is seen only in $\mathbf{X}$ OR in $\mathbf{Z}$. One way to model that is to assume these additive effects to be picked up by the latent variables already obtained from $\mathbf{X}$ and from $\mathbf{Z}$. Eq. (3c) may then be extended to

$$\mathbf{Y}_{00} = \mathbf{T}_X\mathbf{D}_A\mathbf{T}_Z' + \mathbf{T}_X\mathbf{Q}_X' + \mathbf{Q}_Z\mathbf{T}_Z' + \mathbf{F}_A. \tag{5}$$

The additive loadings, $\mathbf{Q}_X(J \times A)$ and $\mathbf{Q}_Z(I \times A)$ may, e.g. be estimated after removal of the interaction effect, which is the primary aim. Let $\mathbf{G}_A = \mathbf{Y}_{00} - \mathbf{T}_X\mathbf{D}_A\mathbf{T}_Z'$. $\mathbf{Q}_X$

and $\mathbf{Q}_Z$ may be obtained by ordinary least-squares regression of $\mathbf{G}_A$ on $\mathbf{T}_X$ and of $\mathbf{G}'_A$ on $\mathbf{T}_Z$, or by projection of $\mathbf{G}_A$ on $\mathbf{T}_X$ and $\mathbf{T}_Z$ simultaneously, after unfolding $\mathbf{Y}$ into $N = I \cdot J$ "observations". This will not be pursued here. Instead, the relationships between $\mathbf{Y}$ and the latent variables from $\mathbf{X}$ and $\mathbf{Z}$ will be displayed via their correlation loadings (Martens and Martens, 2001), i.e. conventional product–moment correlations between input data and latent variables.

Display of results as correlations between data and model: The $\mathbf{X}$-loadings $\mathbf{P}_X$ in Eq. (4b) may be expressed as a matrix of unit-free correlations $\mathbf{R}_{X,T_X}(K \times A)$ between the $\mathbf{X}$-variables and the $\mathbf{X}$-score vectors $\mathbf{T}_X$, as illustrated in Fig. 2:

$$\mathbf{R}_{X,T_X} : r_{k,a} = \frac{\mathbf{x}'_{0,k}\mathbf{t}_{X,a}}{(\mathbf{x}'_{0,k}\mathbf{x}_{0,k})^{1/2}(\mathbf{t}'_{X,a}\mathbf{t}_{X,a})^{1/2}}. \tag{6a}$$

Due to the analogy to the estimated $\mathbf{X}$-loadings (Eq. (4c)), they are called "correlation loadings for the $\mathbf{X}$-variables". Likewise, the $\mathbf{Z}$-loadings $\mathbf{P}_Z$ in Eq. (4b) may be expressed as matrix of unit-free correlations $\mathbf{R}_{Z,T_Z}(L \times A)$ between the $\mathbf{Z}$-variables and the $\mathbf{Z}$-score vectors $\mathbf{T}_{Z,A}$

$$\mathbf{R}_{Z,T_X} : r_{l,a} = \frac{\mathbf{z}'_{0,l}\mathbf{t}_{Z,a}}{(\mathbf{z}'_{0,l}\mathbf{z}_{0,l})^{1/2}(\mathbf{t}'_{Z,a}\mathbf{t}_{Z,a})^{1/2}}. \tag{6b}$$

These may be used to study how the input variables in $\mathbf{X}$ and $\mathbf{Z}$ are modelled by their latent variables.

Similarly, the modelling of $\mathbf{Y}$ can be studied. The columns in $\mathbf{Y}$ are correlated to the $\mathbf{X}$-scores $\mathbf{T}_X$, by $\mathbf{R}_{Y,T_X}(J \times A)$, defined by

$$\mathbf{R}_{Y,T_X} : r_{j,a} = \frac{\mathbf{y}'_{00,j}\mathbf{t}_{X,a}}{(\mathbf{y}'_{00,j}\mathbf{y}_{00,j})^{1/2}(\mathbf{t}'_{X,a}\mathbf{t}_{X,a})^{1/2}}. \tag{6c}$$

The correlation coefficients in $\mathbf{R}_{Y,T_X}$ reflect $\mathbf{T}_Z\mathbf{D}'_A$ (for model (3c)) or $\mathbf{T}_Z\mathbf{D}'_A + \mathbf{Q}_X$ (for model (5)). Likewise, the rows in $\mathbf{Y}$ are correlated to the $\mathbf{Z}$-scores $\mathbf{T}_Z$, by $\mathbf{R}_{Y,T_Z}(I \times A)$, defined as

$$\mathbf{R}_{Y,T_Z} : r_{i,a} = \frac{\mathbf{y}_{00,i}\mathbf{t}_{Z,a}}{(\mathbf{y}_{00,i}\mathbf{y}'_{00,i})^{1/2}(\mathbf{t}'_{Z,a}\mathbf{t}_{Z,a})^{1/2}}. \tag{6d}$$

$\mathbf{R}_{Y,T_Z}$ reflects $\mathbf{T}_X\mathbf{D}_A$ (for model 3c) or $\mathbf{T}_X\mathbf{D}_A + \mathbf{Q}_Z$ (for model (5)).

The row and column means in $\mathbf{Y}$, here representing the "average consumer's" liking of the I products, $\bar{\mathbf{y}}_{I.}(I \times 1)$, and the J consumers' liking of the "average product", $\bar{\mathbf{y}}_{.J}(J \times 1)$, were removed in Eq. (1b). They may be correlated to the latent variables from $\mathbf{X}$ and $\mathbf{Z}$, respectively. Hence, in this case, the hypothetical "average consumer" ("*RowsMean*") may be related to the latent variables from $\mathbf{T}_{X,A}$

$$\mathbf{r}_{Y,RowsMean} : r_a = \frac{\bar{\mathbf{y}}'_{.I}\mathbf{t}_{X,a}}{(\bar{\mathbf{y}}'_{.I}\bar{\mathbf{y}}_{.I})^{1/2}(\mathbf{t}'_{X,a}\mathbf{t}_{X,a})^{1/2}} \tag{6e}$$

and the "average product" ("*ColsMean*") may be related to $\mathbf{T}_{Z,A}$

$$\mathbf{r}_{Y,ColsMean} : r_a = \frac{\bar{\mathbf{y}}'_{J.}\mathbf{t}_{Z,a}}{(\bar{\mathbf{y}}'_{J.}\bar{\mathbf{y}}_{J.})^{1/2}(\mathbf{t}'_{Z,a}\mathbf{t}_{Z,a})^{1/2}}. \tag{6f}$$

2.2.4. PLS estimation of weights for the latent variables

The arrows in Fig. 2 show that the chosen algorithm starts with the **X**- and **Z**-weights $\mathbf{V}_{X,A}(K\times A)$ and $\mathbf{V}_{Z,A}(L\times A)$. In the literature, two different ways to use Herman Wold's (1982) PLS-principle in two-block **X**/**Y** modelling have been developed, apparently somewhat independently of one another. The same two approaches may be used also for the three-block **X**/**Y**/**Z** modelling.

Simultaneous extraction of all the components: The two-block PLS method of Bookstein et al. (1996) is most recent, but simplest. It extracts all the components in one singular value decomposition (SVD) of the **X**/**Y** covariance matrix: $\mathrm{SVD}(\mathbf{X}_0'\mathbf{Y}_0)$. This means that if there is only one **Y**-variable, it can only extract one latent variable. This stops the method from being useful in some cases, but for the present type of data with many **Y**-variables, it may be good. A three-block extension of this approach is simply to employ $\mathrm{SVD}(\mathbf{X}_0'\mathbf{Y}_{00}\mathbf{Z}_0)$, i.e.

$$\mathbf{U}_X\mathbf{S}\mathbf{U}_Z' = \mathbf{X}_0'\mathbf{Y}_{00}\mathbf{Z}_0, \tag{7a}$$

where, for $M=\min(K,L), \mathbf{U}_X(K\times M)$ and $\mathbf{U}_Z(L\times M)$ are the orthonormal left- and right-hand singular vectors of $(\mathbf{X}_0'\mathbf{Y}_{00}\mathbf{Z}_0)$, and $\mathbf{S}(M\times M)$ is the diagonal matrix of singular values.

To avoid collinearity problems and to simplify the graphical interpretation, only the A first, reliable components are used:

$$\mathbf{V}_X = [\mathbf{u}_{X,a}, a=1,2,\ldots,A], \quad \mathbf{V}_Z = [\mathbf{u}_{Z,a}, a=1,2,\ldots,A]. \tag{7b}$$

Equivalently, the **X**-weights $\mathbf{V}_X$ may be obtained as the A first eigenvectors of the **Y**/**Z**-weighted **X**-covariance $\mathbf{X}_0'(\mathbf{Y}_{00}\mathbf{Z}_0\mathbf{Z}_0'\mathbf{Y}_{00}')\mathbf{X}_0$, and the **Z**-weights $\mathbf{V}_Z$ as the A first eigenvectors of the **Y**/**X**-weighted **Z**-covariance $\mathbf{Z}_0'(\mathbf{Y}_{00}'\mathbf{X}_0\mathbf{X}_0'\mathbf{Y}_{00})\mathbf{Z}_0$.

This solution has the properties that the weights are orthogonal, but not the scores,

$$\mathbf{V}_X'\mathbf{V}_X = \mathbf{I}, \qquad \mathbf{V}_Z'\mathbf{V}_Z = \mathbf{I}, \mathbf{T}_X'\mathbf{T}_X \neq diag, \quad \mathbf{T}_Z'\mathbf{T}_Z \neq diag.$$

Sequential extraction of the components: PLS Regression by Wold et al. (1983) is the two-block PLS method most commonly used in, e.g. chemometrics. It employs a sequence of singular value decompositions of deflated **X**/**Y** covariances, $\mathrm{SVD}(\mathbf{X}_{a-1}'\mathbf{Y}_0)$, $a=1,2,\ldots,A$. Each time only the *first* left-hand singular vector is retained (Höskuldsson, 1988) as basis vectors for the weights $\mathbf{V}_{X,A}$. These orthonormal basis vectors $\mathbf{W}_X = [\mathbf{w}_{X,a}, a=1,2,\ldots,A]$ are called loading weights. The three-block extension of this two-block PLS Regression likewise requires deflation of **X**; it employs $\mathrm{SVD}(\mathbf{X}_{a-1}'\mathbf{Y}_{00}\mathbf{Z}_{a-1})$, $a=1,2,\ldots,A$. The estimation algorithm therefore repeats Eqs. (3b,c) and (4c,d) for each component:

For $a=1,2,\ldots,A$:

$$\mathbf{U}_X\mathbf{S}\mathbf{U}_Z' = \mathbf{X}_{a-1}'\mathbf{Y}_{00}\mathbf{Z}_{a-1} \quad \text{(SVD)} \tag{8a}$$

$$\mathbf{w}_{X,a} = \mathbf{u}_{X,1}, \quad \mathbf{w}_{Z,a} = \mathbf{u}_{Z,1} \quad \text{(use only the first singular vector)}$$

$$\mathbf{t}_{X,a} = \mathbf{X}_{a-1}\mathbf{w}_{X,a}, \quad \mathbf{t}_{Z,a} = \mathbf{Z}_{a-1}\mathbf{w}_{Z,a} \quad [(\approx \text{Eq. (3b)})] \tag{8b}$$

$$\mathbf{p}_{X,A} = \mathbf{X}_{a-1}'\mathbf{t}_{X,a}(\mathbf{t}_{X,a}'\mathbf{t}_{X,a})^{-1}, \quad \mathbf{p}_{Z,A} = \mathbf{Z}_{a-1}'\mathbf{t}_{Z,a}(\mathbf{t}_{Z,a}'\mathbf{t}_{Z,a})^{-1} \quad [(=\text{Eq. (4b)})] \tag{8c}$$

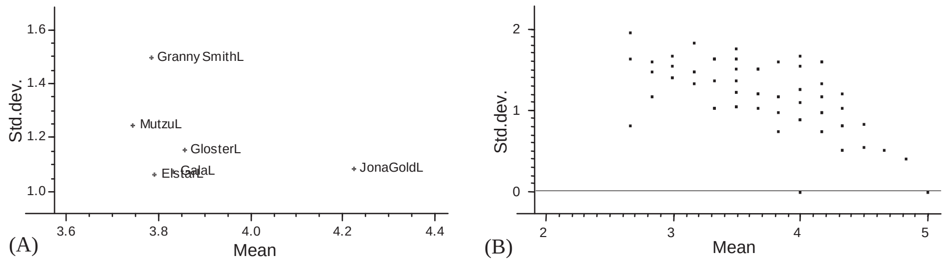

Fig. 3. Summary of input data **Y**. Mean (abscissa) vs. total standard deviation (ordinate). (A) Row mean $\bar{\mathbf{y}}_{I.}$ vs. row std. dev. of six apple types (statistics over 125 children). (B) Col. mean $\bar{\mathbf{y}}_{.J}$ vs. col. std. dev. of 125 children (statistics over six apple types).

$$\mathbf{E}_{X,a} = \mathbf{X}_{a-1} - \mathbf{t}_{X,a}\mathbf{p}'_{X,a}, \quad \mathbf{E}_{Z,a} = \mathbf{Z}_{a-1} - \mathbf{t}_{Z,a}\mathbf{p}'_{Z,a} \quad [(= \text{Eq. (4c)})] \tag{8d}$$

$$\mathbf{X}_a = \mathbf{E}_{X,a}, \quad \mathbf{Z}_a = \mathbf{E}_{Z,a}, \tag{8e}$$

end

$$\mathbf{W}_X = [\mathbf{w}_{X,a}, a = 1, 2, \ldots, A], \quad \mathbf{W}_Z = [\mathbf{w}_{Z,a}, a = 1, 2, \ldots, A],$$

$$\mathbf{T}_X = [\mathbf{t}_{X,a}, a = 1, 2, \ldots, A], \quad \mathbf{T}_Z = [\mathbf{t}_{Z,a}, a = 1, 2, \ldots, A],$$

$$\mathbf{P}_X = [\mathbf{p}_{X,a}, a = 1, 2, \ldots, A], \quad \mathbf{P}_Z = [\mathbf{p}_{Z,a}, a = 1, 2, \ldots, A].$$

In analogy to the PLSR, the weights for the equivalent score expressions in Eqs. (3a,b) and Fig. 2 may then be obtained by

$$\mathbf{V}_X = \mathbf{W}_X(\mathbf{P}'_X\mathbf{W}_X)^{-1}, \quad \mathbf{V}_Z = \mathbf{W}_Z(\mathbf{P}'_Z\mathbf{W}_Z)^{-1}. \tag{8f}$$

This solution has the properties that the scores are orthogonal, but not the weights,

$$\mathbf{T}'_X\mathbf{T}_X = diag, \quad \mathbf{T}'_Z\mathbf{T}_Z = diag, \quad \mathbf{V}'_X\mathbf{V}_X \neq diag, \quad \mathbf{V}'_Z\mathbf{V}_Z \neq diag.$$

3. Results

3.1. Input data

Fig. 1 depicted the input data tables **X**, **Y** and **Z**. The grey-scale of **Y** reflects the original 5-point *liking* scale. **X** has been standardised in order to make its different types of descriptors comparable, i.e. each column has been mean centred and rescaled to a total initial standard deviation of 1. **Z** has likewise been standardised in order to make its different types of descriptors comparable.

Fig. 3 shows some simple statistics of the liking data. Fig. 3(A) shows the arithmetic row mean ($\bar{\mathbf{y}}_{I}.$, abscissa) of each of the products in **Y** (each of the six apple types), plotted against their total initial standard deviations (ordinate). It shows that, on the average, product *Jonagold* is the most liked, while *Mutsu* is the least liked. The childrens' liking differs most for *Granny Smith*.

Fig. 3(B) shows the corresponding column summaries: the abscissa shows the mean $\bar{\mathbf{y}}_{.J}$ for each of the consumers in **Y** (each of the 125 children) while the ordinate shows their corresponding standard deviations. The figure shows that children with low mean liking differentiate more between the different apple types. A few children do not differentiate at all in their reported data (std. dev. = 0).

3.2. L-PLS Regression of **Y** on **X** and **Z**

The standardized **X** and **Z** data were submitted to L-PLSR modelling of the double-centred data in **Y**, according to the structure model outlined in Fig. 2, using the estimation algorithm with orthogonal component scores (Eq. (8a–f)).

Cross-validation with 6×125 segments (not shown here) indicated the first PLS component ("PC") to be clearly valid, while the later ones had low predictive ability. The second PC was found to span primarily one single product, but in an interpretable way, so it was included in the results to be presented.

Fig. 4 shows the results for the first two PCs. Figs. 4(a)–(d) show separate sets of parameters, while Fig. 4(e) plots all of them on top of each other, which is possible, since correlation loadings are unit free.

Product descriptors **X**: Fig. 4(a) shows the main patterns of the sensory, instrumental and chemical product descriptors, in terms of correlations $\mathbf{R}_{X,T_X}$ between the $K = 10$ columns in **X** and the first A=two **X**-score vectors in $\mathbf{T}_X$. The horizontal dimension is seen to span the sensory contrast between *SOUR* and *SWEET*, and the chemical contrast between the *ACIDS/SUGARS* ratio and the *SUGAR* content. Sensory *RED colour* is correlated with *SWEET* apples. The vertical dimension mainly contrasts properties like sensory *HARD* and instrumentally *FIRM* against sensory *ROUND* shape and high content of *ACIDS* and *SUGARS*.

Consumer descriptors **Z**: Fig. 4(b) shows the main patterns of the consumer background descriptors, in terms of correlations $\mathbf{R}_{Z,T_Z}$ between **Z** and the first two **Z**-score vectors $\mathbf{T}_Z$. The horizontal dimension spans a tendency to choose the *green apple first* and the *red apple last* (*GreenAFirst, RedALast*), against the tendency to choose the *red apple first* and *the green apple last*. Vertically, a contrast between choosing *pear first* and *banana last* against choosing *banana first* and *pear last*. The purely sociological variables (*gender*, *age,* how *often* apples are eat) are not particularly evident in the result, although *gender* (coded as being *girl*) is slightly associated with choosing *green apple first*, *pear first* and *banana last*.

Consumer liking of the products **Y**: Fig. 4(c) shows the main, product-related patterns of the consumer w.r.t. liking, in terms of the correlations $\mathbf{R}_{Y,T_X}$ between **Y** and $\mathbf{T}_X$. Most of the 125 children are gathered towards either end of the horizontal dimension. The second, vertical dimensions is much less extensive, and spans fewer children.

The "average consumer" ("*RowsMean*"=$\bar{\mathbf{y}}_{I.}$, the arithmetic average over the columns **Y**, i.e. abscissa in Fig. 2(a)) is seen to fall to the left of the origin; however, few children fall in the vicinity of *RowsMean*. Hence, this average consumer does not seem particularly interesting.

Product liking by the consumers, as seen from **X** *and from* **Z**: Fig. 4(d) shows the main patterns of the six products. The dotted lines point to correlations $\mathbf{R}_{Y,T_Z}$ (o, with

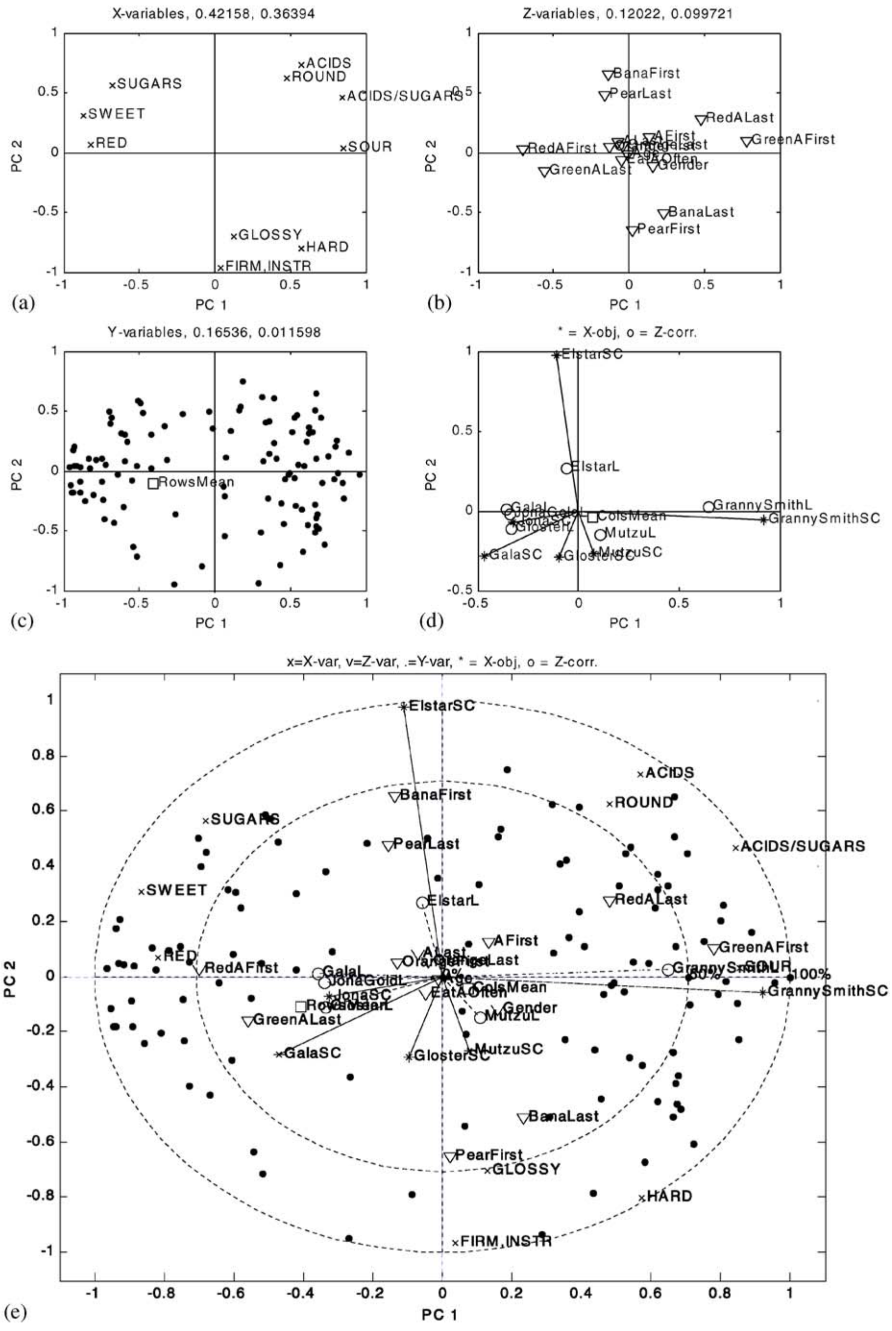

Fig. 4. L-PLS regression; correlation loadings for PCs $a = 1$ (abscissa), $a = 2$ (ordinate). (A) 10 product descriptors $\mathbf{X}$ correlated to $\mathbf{X}$-scores $\mathbf{T}_X(\mathbf{R}_{X,T_X})$. The two components describe 42% and 36% variance in $\mathbf{X}_0$; (B) 15 consumer descriptors $\mathbf{Z}$ correlated to $\mathbf{Z}$-scores scores $\mathbf{T}_Z(\mathbf{R}_{Z,T_Z})$. The components describe 12% and 10% variance in $\mathbf{Z}_0$; (C) 125 $\mathbf{Y}$-variables (children's liking $\mathbf{Y}$) correlated to $\mathbf{X}$-scores scores $\mathbf{T}_X(\mathbf{R}_{Y,Tx})$. The components describe 16% and 1% variance in $\mathbf{Y}_{00}$; (D) six $\mathbf{Y}$-objects (apple types $\mathbf{Y}'$) correlated to $\mathbf{Z}$-scores scores $\mathbf{T}_Z(\mathbf{R}_{Y,T_Z}, 0)$ and six $\mathbf{X}$-objects (apple types $\mathbf{X}'$) correlated to $\mathbf{X}$-indicator matrix $\mathbf{I}_X(\mathbf{R}_{I,Tx},*)$; (E) all correlation loadings plotted together. The dotted ellipses represent 50% and 100% explained variance, respectively.

suffix “L” for “Liking”) between $\mathbf{Y}'$ and the $\mathbf{Z}$-scores $\mathbf{T}_Z$. The solid lines (*, with suffix “SC” for “Sensory-Chemical”), point to an alternative product representation, namely the correlation coefficients between the identity matrix $\mathbf{I}$ $(I \times I)$, and the two first $\mathbf{X}$-score vectors $\mathbf{T}_X$, i.e. the partial leverages in $\mathbf{X}$ (Martens and Martens, 2001). The $\mathbf{X}$- and $\mathbf{Z}$-patterns point in general in the same directions: The horizontal dimension spans the contrast between *Granny Smith* and the other products, mainly *Gala*, *Gloster* and *Jonagold*. The vertical dimension spans the contrast between *Elstar* and the other products. The correlations $\mathbf{R}_{Y,T_Z}$ between $\mathbf{Y}'$ and the $\mathbf{Z}$-scores $\mathbf{T}_Z$ are rather weak, indicating that the variations in the second dimension is weaker than in the first dimension.

The “average liking of apples” (“*ColsMean*” $=\bar{\mathbf{y}}_J$, the arithmetic average over the rows in $\mathbf{Y}$, i.e. abscissa in Fig. 2(b) is seen to fall just to the right of the origin. Hence, the main patterns found in the double-centred liking data do not seem to reflect the children’s general liking of apples.

Overview of the L-PLSR solution: Fig. 4(e) combines Figs. 4(a)–(d). In the horizontal dimension product *GrannySmith* is seen to be particularly *SOUR* and not *SWEET*; it has high ratio *ACID/SUGARS* and low level of *SUGARS*. It is also *HARD* and not *RED*. The products *Gala*, *Gloster* and *Jonagold* appeared to display the opposite tendency.

GrannySmith is seen primarily to be liked by children who were observed to choose *green apple first* and *red apple last*, not by children who were observed to choose *red apple first* and *green apple last*. Again, products *Gala*, *Gloster* and *Jonagold* seem to display the opposite of this tendency.

In the vertical dimension, product *Elstar* is seen to be particularly *round*, with high levels of both *ACIDS* and *SUGARS*, but not instrumentally *FIRM* and sensory *HARD*; nor was it *GLOSSY*. On the contrary, the products *Mutsu*, *Gloster* and *Gala* appeared to be a little more *FIRM* and *GLOSSY*, with less *SUGARS* and *ACIDS* than the others.

Product *Elstar* seems primarily to be liked by children who chose *banana first* and *pear last*, and less liked by children who chose *pear first* and *banana last*. In contrast, e.g. *Mutsu* seemed to be associated with the liking of children who chose *pear first*. Close inspection of the input data (Fig. 1) confirms all of these conclusions, but they were not easy to see directly from the raw data.

The relative size of the components in $\mathbf{X}$ and $\mathbf{Z}$ were

$$\mathbf{T}_X'\mathbf{T}_X = \begin{bmatrix} 21 & 0 \\ 0 & 17 \end{bmatrix}$$

and

$$\mathbf{T}_Z'\mathbf{T}_Z = \begin{bmatrix} 219 & 0 \\ 0 & 170 \end{bmatrix},$$

respectively. This shows that in both $\mathbf{X}$ and $\mathbf{Z}$ the relative importance of the two first components was about the same. The full kernel matrix in the model equation (3c), obtained by Eq. (3f), was

$$D_A = \begin{bmatrix} 0.171 & 0.001 \\ 0.012 & 0.055 \end{bmatrix}.$$

Here, the first component is seen to be clearly the most important one in **Y**. The almost diagonal structure shows that the two phenomena modelled are well separated in both **X** and **Z**, so in this case there was no need for further diagonalization of $\mathbf{D}_A$.

4. Discussion

The present results indicate that PLS modelling with latent variables in two or more dimensions (Wold et al., 1987) is useful and therefore deserves more attention. Fig. 4 illustrated a particular property of the L-PLSR, in that although the **X**- and **Z**-variables share no physical matrix size-dimension, and therefore cannot be correlated directly to each other, they have become connected via **Y**. For instance, *RED* (a column with six numbers from **X**) is seen to be positively related to *RedAFirst* (a column with 125 numbers from **Z**). Likewise *HARD* (from **X**) and *BananaLast* (from **Z**) are positively related. The observed associations seem to make sense in terms of common language and background knowledge.

In general terms PC1 could be interpreted as a general *sour*-and-*green* vs. *sweet*-and-*red* pattern of variation between the apple cultivars. The PC2 may be interpreted as spanning the difference between a hard, pear-like texture and a soft, more banana-like texture.

Children who would choose *red apples first*, tended to choose *green apples last*, and vice versa. They distinguished the green *GrannySmith* and *Mutsu* from the other, more red apples. Children who would choose *banana first* tended to choose *pear last*, and vice versa. As expected, these children distinguished the very soft (“banana-like”) *Elstar* from the others, primarily from the more *hard*, more bland, yellow-green (more “pear-like”) *Mutsu*.

The purely sociological **Z**-descriptors of the consumers (*gender*, *age*) did not give much relevant information about the consumer liking in this case. The strongest correlation between e.g. *gender* and *liking* for any of the apples was only 0.07.

Assessment of conclusions from the L-PLSR modelling. With a relatively complex modelling tool like the L-PLSR, it is important to verify the main aspects of the interpretation by plotting the raw data. Some examples of this are shown in Fig. 5.

Product liking **Y**: Fig. 5(a) shows the input data for comparing the *liking* response for the most extreme products (liking *GrannySmith* vs. liking *Jonagold*). With only five response levels possible, many data points are superimposed and the pattern difficult to see. But their raw liking data are clearly negatively correlated ($r = -0.29$ over the 125 subjects), as expected.

Liking **Y** *vs. consumer background* **Z**. Fig. 5(b) relates liking of the green apple *GrannySmith* to the background response *green apple first*. There is a clear tendency ($r = 0.52$ over 125 subjects) that if children chose green apple first, they reported that they liked *GrannySmith*.

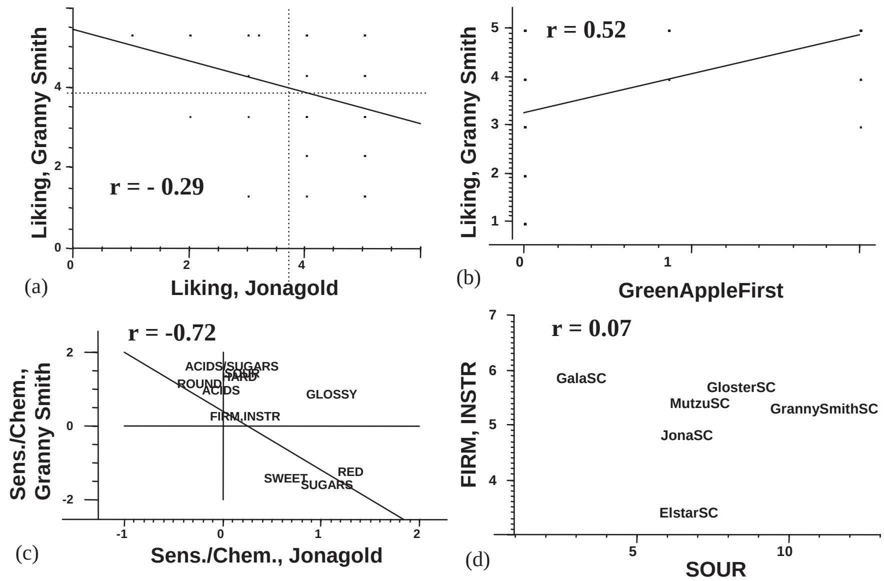

Fig. 5. Checking conclusions in the raw data: (A) A negative relation between rows in **Y**: Liking for the two most extreme products, $\mathbf{y}_{Jonagold}$ (abscissa) vs. $\mathbf{y}_{GrannySmith}$(ordinate). Lines: Regression between ordinate and abscissa, $r =$ correlation coefficient. (B) A positive relation between columns in **Z** and **Y′**: Choosing *green apple first*, $\mathbf{z}_{GreenAFirst}$ (0 or 1, abscissa) vs. the reported liking of apple type *GrannySmith*, $\mathbf{y}_{GrannySmith}$(1–5, ordinate) for the 125 children. (C) A negative relationship between rows in standardized **X**: Row $\mathbf{x}_{Jonagold}$ (abscissa) vs. row $\mathbf{x}_{GrannySmith}$ (ordinate). (D) A weak relations between columns in **X**: Physical descriptor $\mathbf{x}_{FIRM,INSTR}$ (abscissa) vs. the sensory descriptor, $\mathbf{x}_{SOUR}$ (1–5, ordinate) for the six apple types. Suffix "SC" means "sensory-chemical".

Product descriptor rows in **X**: Fig. 5(c) shows the standardized sensory and chemical variables for the two most extreme products, *GrannySmith* and *Jonagold*. Again, these two products are seen to be described quite opposite; *Jonagold* is *SWEET*, *RED* and high in *SUGARS*, while *GrannySmith* has high *ACIDS*/*SUGARS* ratio, is *SOUR*, *HARD* and *ROUND*, and vice versa. The r is -0.72 between these two rows of 10 standardized **X**-variables.

Product descriptor columns in **X**: Fig. 5(d) shows the input data for the sensory descriptor *SOUR* and the instrumental descriptor *FIRM*,*INSTR*. As expected from the L-PLSR model, these two variables are almost orthogonal, with $r = 0.07$ over the six products.

So, the conclusions from the L-PLSR have been confirmed. But which properties of the products were really decisive in *causing* these patterns in the childrens' liking of the appearance of the apples? Such a causal analysis cannot be completed on the basis of this empirical study alone, particularly when only limited range of products and consumers were involved.

But at least the analysis has revealed two interpretable patterns of variation in the double-centred liking data. On this basis one may segment the consumers into groups, distinguishing, e.g. all the children in the left- and right-half of Fig. 5(c) as “likers of red, sweet, soft” and “likers of green, sour, hard” (or “*GrannySmith* likers”).

The present centring of the data removed the children’s differences in general liking of apples, so the background variables *AppleFirst* and *EatAOften* correlated only weakly to the obtained model in Fig. 4, as expected. In general, it is advisable to check the effect of the double centring of the **Y**-data used for the L-PLSR in Fig. 4. In a separate model, the liking data **Y** and $\mathbf{R}_{ZY}$ (the $I \times L$ correlations of consumer descriptors **Z** to $\mathbf{Y}'$) were regressed on the product descriptors **X**. The results are slightly different, as expected with the different mean-centring approach; the results of this were detailed by Thybo et al. (2003).

Simultaneous estimation of all the L-PLSR components: The alternative L-PLSR method, based on a three-block extension of Bookstein’s two-block PLS modelling (Eq. (7a,b), was also tested for these input data. For this simultaneous L-PLSR analysis, the **X**-,**Y**- and **Z**-data were pre-processed just like for the sequential L-PLSR analysis. The results were almost indistinguishable from those in Fig. 4, and therefore they will not be reported here. The only apparent difference was a slight rotation of PC # 2, reflecting the fact that the score vectors $\mathbf{T}_X$ and $\mathbf{T}_Z$ are not completely orthogonal in this method.

Control of the L-PLSR method with artificial data: The Appendix A and Fig. 6 show that the L-PLSR method and software used for obtaining Fig. 4 give perfect representation of artificial, noise-free data constructed according to the model in Eq. (3c).

5. Conclusions

A theory and two estimation algorithm versions have been presented for the “L-PLS regression”, where **Y** is modelled by the bi-linear interactions of latent variables from both **X** and **Z**. It gave an interpretable overview of rather complicated and noisy empirical data, and performed as expected. After mean centring, two patterns of co-variation were found in the present data—a green-and-sour vs. red-and-sweet pattern, and a hard vs. soft pattern.

However, further work is called for. The relative merits of the two L-PLSR versions with and without deflation need further study. The effects of the row- and column-means in **Y** also need further investigation.

Acknowledgements

This work was in part financed by the Danish FØTEK research project “Advanced Quality Monitoring”, and a research project supported by the Danish Dairy Research Foundation (Danish Dairy Board) and the Danish Government. Part of the work was done at The Norwegian Food Research Institute.

Appendix A.

A.1. L-PLSR analysis of artificial, noise-free data

Fig. 6 demonstrates the arithmetic functionality of the L-PLSR method, based on artificial, noise-free $\mathbf{X}, \mathbf{Y}$ and $\mathbf{Z}$ data with rank $A=2$: Bi-linear "liking" data $\mathbf{Y}_{00}$ were constructed according to Eq. (3c) for $I=6$ "products" and $J=125$ "consumers", as a sum of two latent variables with random normally distributed ("randn") $\mathbf{X}$-scores $\mathbf{T}_X = \text{randn}(6,2)$ and random normally distributed $\mathbf{Z}$-scores $\mathbf{T}_Z = \text{randn}(125,2): \mathbf{Y}_{00} = \mathbf{T}_X\mathbf{D}_A\mathbf{T}_Z'$, with $\mathbf{D}_A = \mathbf{I}$. The final, error-free liking $\mathbf{Y}$-data were generated by adding random row offset for the products, $\mathbf{y}_r = \text{randn}(I,1)$, and column offsets for the consumer, $\mathbf{y}_c = \text{randn}(J,1): \mathbf{Y} = \mathbf{Y}_{00} + \mathbf{1}\mathbf{y}_c' + \mathbf{y}_r\mathbf{1}'$.

Product descriptor data $\mathbf{X}$ were then constructed for $K\ =10$ "descriptors" as linear combinations of these two latent variables's $\mathbf{X}$-scores $\mathbf{T}_X$, with random $\mathbf{X}$-loadings $\mathbf{P}_X = \text{randn}(10,2): \mathbf{X} = \mathbf{T}_X\mathbf{P}_X'$. Consumer descriptor data $\mathbf{Z}$ were conversely constructed

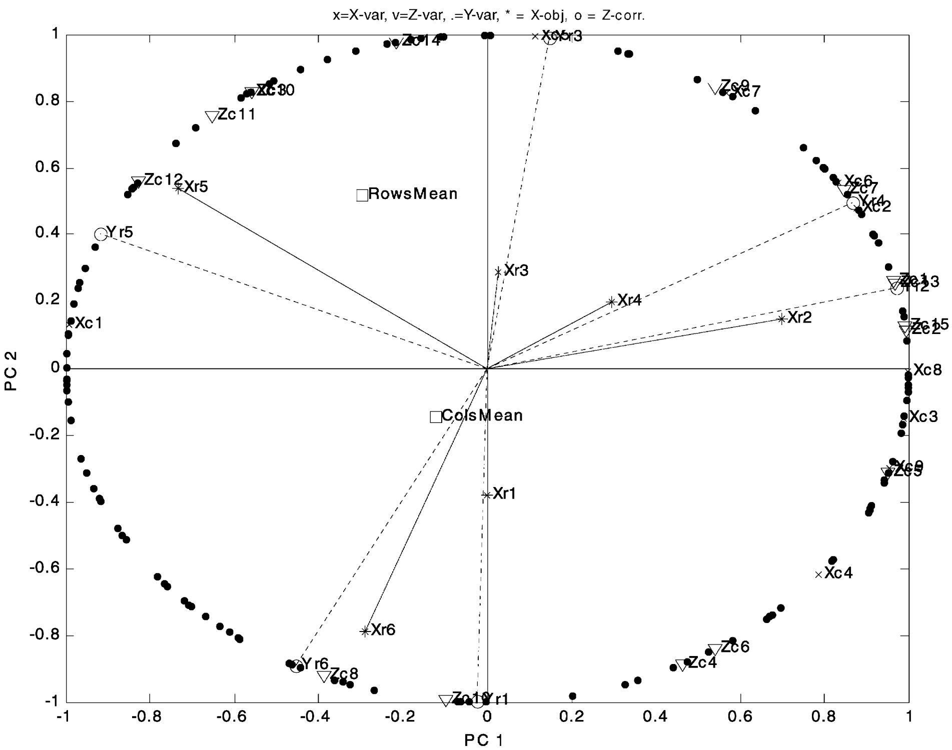

Fig. 6. L-PLSR solution for artificial, noise-free data: $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$ data generated from two latent variables. $\mathbf{X}_c$: columns (product descriptors, $\times$) in $\mathbf{X}$. $\mathbf{Y}_r, \mathbf{Y}_c$: rows (products, o) and columns (consumers,$\bullet$) in $\mathbf{Y}$. $\mathbf{Z}_c$: Columns (consumer descriptors, ∇) in $\mathbf{Z}$. $\mathbf{X}_r$: rows (products, *) in $\mathbf{X}$ (partial leverages). *RowsMean* and *ColsMean* $= \bar{\mathbf{y}}_{I.}$ and $\bar{\mathbf{y}}_{.J}$ from Eq. (1b).

for $L = 15$ descriptors as linear combinations of these two latent variables' **Z**-scores $\mathbf{T}_Z$, with random **Z**-loadings $\mathbf{P}_Z = \mathrm{randn}(15, 2) : \mathbf{Z}_0 = \mathbf{T}_Z\mathbf{P}'_Z$.

These artificial **X**/**Y**/**Z** data were analysed like the real data, with results summarised in Fig. 6, comparable to Fig. 4(e). The 10 "product" descriptors are represented by $\mathbf{r}_{X,T_X,a}$ for a=1 (abscissa) and a=2 (ordinate). They are denoted by $\mathbf{X}_\mathrm{c}\langle$column#1 : 10$\rangle$, the 15 "consumer descriptors" ($\mathbf{r}_{Z,T_Z,a}$) are denoted $\mathbf{Z}_\mathrm{c}\langle$column#1 : 15$\rangle$, the six products liking ($\mathbf{r}_{Y,T_X,a}$) are denoted $\mathbf{Y}_\mathrm{r}\langle$row#1 : 6$\rangle$, while the 125 consumers' liking ($\mathbf{r}_{Y,T_Z,a}$) are simply shown as 125 black dots. The product score configuration in $\mathbf{X}(\mathbf{r}_{I,T_X,a})$ is shown as $\mathbf{X}_\mathrm{r}\langle$row#1 : 6$\rangle$. As before, the mean consumer's liking is called *RowsMean*, and the mean product's liking *ColsMean*.

The double-centred **Y**-data were perfectly fitted after two PCs (even in the element-by-element cross-validation sense). The figure shows, as expected, the 125 consumer columns in **Y**(•) to form a perfect ellipse with multiple $r^2 = 1$ (100% explained variance), together with the column indicators in **X** and **Z**, $\mathbf{X}_\mathrm{c}$ and $\mathbf{Z}_\mathrm{c}$, and the row indicators in $\mathbf{Y}, \mathbf{X}_\mathrm{r}$. This demonstrates that the L-PLS regression formalism works as expected.

The **Y** rows ($\mathbf{Y}_\mathrm{r}$) and the **X** rows ($\mathbf{X}_\mathrm{r}$) point in almost identical directions. However, since the **X** indicators $\mathbf{X}_\mathrm{r}$ are intended to reveal uniqueness among the samples, they do not extend as far as $\mathbf{Y}_\mathrm{r}$ (dotted), which show 100% fit for these artificial data, in contrast to the real data.

The column and row means in **Y**, *RowsMean* and *ColsMean*, are seen to correlate weakly with the latent variables; this fit is incidental.

References

Bookstein, F.L., Sampson, P.D., Streissguth, A.P., Barr, H.M., 1996. Exploiting redundant measurement of dose and developmental outcome: new methods from the behavioral teratology of alcohol. Dev. Psychol. 32 (3), 404–415.

Elden, L., 2004. Partial least-squares vs. Lanczos bidiagonalization I: analysis of a projection method for multiple regression, Comput. Statist. Data Anal., in press; doi: 10.1016/S0167-9473(03)00138-5.

Höskuldsson, A., 1988. PLS regression methods. J. Chemometrics 2, 211–228.

Höskuldsson, A., 2001. Causal and path modelling. Chemometrics Intell. Lab. Systems 58 (2), 287–311.

Kubberød, E., Ueland, Ø., Rødbotten, M., Westad, F., Risvik, E., 2002. Gender specific preferences and attitudes towards meat. Food Qual. Preference 13 (5), 285–294.

Kühn, B.F, Thybo, A.K., 2001. The influence of sensory and physiochemical quality on Danish children's preferences for apples. Food Qual. Preferences 12, 534–550.

Martens, H., 2001. Reliable and relevant modelling of real world data: a personal account of the development of PLS regression. Chemometrics Intell. Lab. Systems 58, 85–95.

Martens, H., Martens, M., 2001. Multivariate Analysis of Quality. An Introduction. Wiley, Chichester, UK.

Martens, H., Naes, T., 1989. Multivariate Calibration. Wiley, Chichester, UK.

Nguyen, D.V., Rocke, D.M., 2004. On partial least squares dimension reduction for microarray-based classification: a simulation study. Comput. Statist. Data Anal., in press; doi: 10.1016/j.csda.2003.08.001.

Thybo, A.K., Kühn, B.F., Martens, H., 2003. Explaining Danish children's preferences for apples using instrumental, sensory and demographic/behavioural data. Food Qual. Preference 15, 53–63.

Wold, H., 1982. Soft modelling. The basic design and some extensions. In: Jöreskog, K.G., Wold, H. (Eds.), Systems Under Indirect Observation. Causality-Structure-Prediction. Vol. I. North-Holland, Amsterdam, pp. 263–271.

Wold, S., Martens, H., Wold, H., 1983. The multivariate calibration problem in chemistry solved by the PLS method. In: Ruhe, A., Kågström,B. (Eds.), Proceedings of the Conference on Matrix Pencils, March 1982, Lecture Notes in Mathematics, Springer, Heidelberg, pp. 286–293.

Wold, S., Hellberg, S., Lundstedt, T., Sjöström, M., Wold, H., 1987. PLS modeling with latent variables in two or more dimensions. In: Proceedings: PLS Model Building: Theory and Applications. Symposium Frankfurt am Main, September 23–25, 1987.