Food Quality and Preference 24 (2012) 8–16

Contents lists available at SciVerse ScienceDirect

Food Quality and Preference

journal homepage: www.elsevier.com/locate/foodqual

# Preference mapping by PO-PLS: Separating common and unique information in several data blocks

Ingrid Måge [a,*], Elena Menichelli [a,b], Tormod Næs [a,c]

[a] Nofima AS, Olsoveien 1, NO-1430 Ås, Norway
[b] University of Life Sciences, NO-1430 Ås, Norway
[c] University of Copenhagen, Faculty of Life Sciences, Dept. Food Science, Rolighedsvej 30, 1958 Fredriksberg Copenhagen, Denmark

ARTICLE INFO

*Article history:*
Received 7 January 2011
Received in revised form 16 August 2011
Accepted 16 August 2011
Available online 25 August 2011

*Keywords:*
PO-PLS
PLS regression
Multi-block analysis
Preference mapping

ABSTRACT

In food development, preference mapping is an important tool for relating product sensory attributes to consumer preferences. The sensory attributes are often divided into several categories, such as visual appearance, smell, taste and texture. This forms a so-called multi-block data set, where each block is a collection of related attributes. The current paper presents a new method for analysing such multi-block data: Parallel Orthogonalised Partial Least Squares regression (PO-PLS). The main objective of PO-PLS is to find common and unique components among several data blocks, and thereby improve interpretation of models. In addition to that, PO-PLS overcomes some challenges from the standard multi-block PLS regression when it comes to scaling and dimensionality of blocks.

The method is illustrated by two case studies. One of them is based on a collection of flavoured waters that are characterised by both odour and flavour attributes, forming two blocks of sensory descriptors. A consumer test has also been performed, and PO-PLS is used to create a preference map relating the sensory blocks to consumer liking. The new method is also compared to a preference map created by standard PLS regression. The same is done for the other data set where instrumental data are applied together with sensory data when predicting consumer liking. Here the sensory variables are divided into two blocks: one related to appearance and mouth feel attributes and the other one describing odour and taste properties. In both cases the results clearly illustrate that PO-PLS and PLS regression are equivalent in terms of model fit, but PO-PLS offer some interpretative advantages.



## 1. Introduction

In modern food development, actual prototypes are often assessed by several measurement principles such as chemical analysis, descriptive sensory analysis and various types of consumer liking or choice tests. Typically, one will relate these different measurements to each other in order to obtain improved information about what are the main "drivers of liking" and how the values of these "drivers" can be optimised (Helgesen, Solheim, & Næs, 1997; Moskowitz & Silcher, 2006; Næs, Lengard, Johansen, & Hersleth, 2010). The focus of the present paper will be relations between sensory attributes and instrumental measurements on one side and consumer liking of products on the other.

Descriptive sensory analysis data often consists of different groups or types of attributes. For food products the most important groups are attributes related to visual appearance, smell, taste and texture. In many cases all these attributes are considered together (Helgesen & Næs, 1995; McEwan, 1996; Wold, Veberg, & Nilsen, 2006), while in other cases one will also be interested in how the different groups of attributes, here called data blocks, relate to each other (Martens, Tenenhaus, Vinzi, & Martens, 2007). Likewise, within consumer testing one may be interested in the relation among different measurements taken, for instance among expectation, blind and informed liking or between consumer attributes such as attitudes and habits. The challenge is then not only to find the relation between the main categories of data as listed in the first paragraph, but also relations within each of the categories.

Multivariate data analysis tools such as Principal Component Analysis (PCA) and Partial Least Squares (PLS) regression (Martens & Næs, 1989), are essential for interpreting relationships between many variables. When several data blocks are present, a straightforward solution is to put all variables together into one large data matrix and analyse it with conventional PLS and PCA, depending on whether a predictive direction is present or not. This is often referred to as multi-block PCA and PLS, respectively (Westerhuis, Kourti, & MacGregor, 1998). The drawback of this approach is that variables from different blocks are mixed together, which might obscure interpretation (Jørgensen, Segtnan, Thyholt, & Næs, 2004). The solution will also depend heavily on how the different

* Corresponding author. Tel.: +47 64970100; fax: +47 64970333.
*E-mail address:* ingrid.mage@nofima.no (I. Måge).

0950-3293/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.foodqual.2011.08.003

variable blocks are scaled relative to each other and there may be problems for situations with different dimensionality within each of the blocks. A variant of this approach which proposes a special type of weighting is Multiple Factor Analysis (MFA, Escofier & Pagès, 1994) based on PCA of a concatenated matrix after weighting of each block separately. An approach which solves the problem of different scale is Canonical Correlation Analysis (CCA), introduced by Hotelling (1936). In CCA, linear combinations of two blocks of variables are obtained in such a way that the squared correlation between the linear combinations is maximised. A generalization of the method, called GCA (Carroll, 1968), allows for more than two data blocks. Even though GCA is invariant to scale, it has other problems related to over-fitting and instability when the number of variables is large. Other related approaches can be found in Kettenring (1971); Hanafi and Kiers (2006); Dahl and Næs (2006) and Kohler, Bertrand, Møretrø, and Qannari (2009).

In the area of chemometrics, a couple of methods have recently been developed for solving these problems based on sequential use of PLS regression on matrices that are orthogonalised with respect to each other. These methods are invariant with respect to the relative scale of the data blocks, they allow for different dimensionality of the blocks, allow for high collinearity within and between blocks, and enhance interpretation (Jørgensen, Mevik, & Næs, 2007; Jørgensen et al., 2004; Måge, Mevik, & Næs, 2008). So far the methods have mainly been tested for predictive modelling of production processes with a recent exception of Næs, Tomic, Mevik, and Martens (2010) where one version of it is used within the context of path modelling. Two variants of this type of modelling exist, namely the sequential procedure (SO-PLS, Jørgensen et al., 2004, 2007; Næs et al., 2010) and the so-called parallel method (PO-PLS, Måge, Mevik, & Næs, 2008). The two variants are useful for different purposes, and the difference lies in the way the data blocks are incorporated and which type of information is extracted. In SO-PLS, the focus is on incorporating blocks of data one at a time and assessing and interpreting the incremental or additional contribution of the different blocks added. For the PO-PLS method the focus is on first identifying the information that is common between the blocks and then on identifying the information in each block that is unique.

The present paper is a study of the use of PO-PLS in the area of preference mapping. The method is a combination of PLS regression and GCA. In the approached situation one is interested in the relation between sensory attributes and consumer liking with a special focus on how different blocks of sensory data relate to each other and to the consumer preference data. In one of the examples used for illustration, instrumental data will be applied together with sensory data when predicting consumer liking. In this way the paper is also an illustration of how one can incorporate instrumental data together with sensory data in preference mapping using one single analysis. It will be shown that this type of modelling can be used for obtaining more information than standard preference mapping which will also be tested on the same data. The examples chosen are particularly useful for showing what the new methods does in comparison with standard approaches. An additional scope of the paper is to make the methodology known to the sensory and consumer science community. The method will be illustrated by analysing two data sets.

## 2. Materials and methods

### 2.1. Data sets

#### 2.1.1. Flavoured waters

The main objective of this study was to develop a new type of flavoured water. Sensory and consumer trials were performed in order to optimise the recipe and gain knowledge about which sensory attributes the consumers respond positively (or negatively) to. The data set is collected in such a way that it is suitable for investigating how different groups of attributes are affected by the recipe, and how consumers relate to these groups.

Eighteen water samples were prepared according to a full factorial design with three design factors: Flavour type (A or B), flavour dose (0.2%, 0.6% or 0.8%) and sugar content (L (low), M (medium) or H (high)). A trained sensory panel consisting of 11 assessors evaluated the samples first by smelling (9 descriptors) and then by tasting (14 descriptors). The test was done according to a standard descriptive analysis protocol using a scale between 1 and 9 for each of the attributes. Two data blocks were then obtained for the odour and taste attributes separately by averaging both data sets over the assessors. The sensory descriptors are listed in Table 1.

In addition, 180 consumers tested 10 of the waters each, and rated their overall liking on a scale from 1 (“Dislikes very much”) to 9 (“Likes very much”). The ten waters per consumer were selected according to an incomplete block design, and were presented in two sessions with five waters in each session. The consumers were selected according to relevant market figures: 50% males/females aged between 20 and 49 years. The missing observations, due to the incomplete design structure, were here estimated by PCA (The Unscrambler X, version 10.0.1, CAMO Software AS, Oslo, Norway). The NIPALS algorithm was used for estimation, and the consumers were mean centered but not scaled. The percentage of missing values is high (44%), but the estimates are regarded as adequate since the number of consumers is relatively high and the structure of the data is strong (Hedderley & Wakeling, 1995).

The data thus consists of four data blocks: design matrix, two sensory data sets and one consumer liking data set (see Fig. 1). In this case, the design matrix is only used for interpretation purposes. Further information about the data set can be found in a series of application notes from CAMO Software AS (Måge, 2008a, 2008b, 2008c).

#### 2.1.2. Jams

This data set stems from the Norwegian food research institute (now called Nofima). It consists of 12 raspberry jams selected according to a factorial design based on four production places (C1-C4) and three harvesting times (H1-H3). It is used as a tutorial data set in *The Unscrambler* (CAMO Software AS, Oslo, Norway), and is also thoroughly described and analysed by Esbensen (2002). The jams were evaluated by a trained sensory panel, rating 12 attributes on a 9 point scale, and overall liking was scored by 114 representative consumers. For illustration purposes we will here

**Table 1**
Sensory descriptors in the flavoured waters data set. To distinguish the two groups “odour” and “flavour”, odour descriptors are always given in upper-case letters, and flavour descriptors in lower-case letters.

| Odour | Flavour |
|---|---|
| RIPE | Ripe |
| TROPICAL | Tropical |
| CANDY | Candy |
| SYNTHETIC | Synthetic |
| LACTONIC | Lactonic |
| SULFURIC | Sulfurous |
| SKIN | Skin |
| GREEN | Green |
| FLORAL | Floral |
| | Sweet |
| | Sour |
| | Bitter |
| | Dry |
| | Sticky |