SO-PLS as an alternative approach for handling multi-dimensionality in modelling different aspects of consumer expectations

Quoc Cuong Nguyen[1,2,3,4*], Kristian Hovde Liland[5], Oliver Tomic[5], Amparo Tarrega[6], Paula Varela[3], Tormod Næs[3,7]

[1]Department of Food technology, Ho Chi Minh City University of Technology (HCMUT), Ho Chi Minh City, Vietnam

[2]Vietnam National University, Ho Chi Minh City, Vietnam

[3]Nofima AS, Osloveien 1, P.O. Box 210, N-1431 Ås, Norway

[4]The Norwegian University of Life Sciences, Faculty of Chemistry, Biotechnology and Food Science (IKBM), Ås, Norway

[5]The Norwegian University of Life Sciences, Faculty of Science and Technology, Ås, Norway

[6]Instituto de Agroquimica y Tecnologia de Alimentos, Valencia, Spain

[7]University of Copenhagen, Department of Food Science, Denmark

* Corresponding Author: Quoc Cuong Nguyen [nqcuong@hcmut.edu.vn]

Abstract

In the development of sensory and consumer science, data are often collected in several blocks responding to different aspects of consumer experience. Sometimes the task of organizing the data and explaining their relation is non-trivial, especially when considering structural (casual) relationship between data sets. In this sense, PLS path modelling (PLS-PM) has been found as a good tool to model such relations, but this approach faces some issues regarding the assumption of uni-dimensionality of consumers' data blocks. Sequential Orthogonalised PLS path modelling (SO-PLS-PM) has been proposed as an alternative approach to handle the multi-dimensionality and to explain the relations between the original data blocks without any preprocessing of the data. This study aims at comparing the efficacy of SO-PLS-PM and PLS-PM (together with splitting blocks into uni-dimensional sub-blocks) for handling multi-dimensionality. Data sets from two satiety perception studies (yoghurt, biscuit) have been used as illustrations.

The main novelty of this paper lies in underlining and solving a major, but little studied problem, related to the assumption of one-dimensional blocks in PLS-PM. The findings from the comparisons indicated that the two approaches (PLS-PM and SO-PLS-PM) highlighted the same main trends for the less complex samples (yoghurt samples): liking was the essential driver of satiation perception and portion size selection; while satiation mainly predicted satiety perception. For the more complex data set - from a sensory perspective - (biscuit samples), the relations between data blocks in PLS-PM model was difficult to interpret, whereas they were well explained by SO-PLS-PM. This underlines the ability of SO-PLS-PM to model multi-dimensional data sets without requiring any preprocessing steps.

Keywords: *consumers; liking; satiety; consumer expectations; path modelling; PLS; SO-PLS; uni-dimensionality*

1. Introduction

In sensory and consumer science one is often interested in analyzing and interpreting the relations between several data sets. In cases with common structure among the sets, like for instance the individual data in projective mapping, one will typically use standard multi-block methods like the MFA (Pagès, 2005; Risvik, McEwan, Colwill, Rogers, & Lyon, 1994). When each data set represents a set of manifest (or observable) variables relating to one latent (unobservable) variable and there are explicit casual relationships between latent variables, some type of path modelling may be useful (Pagès & Tenenhaus, 2001). This is a type of modelling where one can impose a structural (sometimes causal) relationship between the blocks, and then estimate how well and in which way the different blocks are related (Tenenhaus, Vinzi, Chatelin, & Lauro, 2005). Typical examples of this are situations in which several consumer variables like demographics, different types of attitudes and habits are related to each other or to the liking of products (Carrillo, Prado-Gascó, Fiszman, & Varela, 2013; Costa-Font & Gil, 2009; Menichelli, Hersleth, Almøy, & Næs, 2014).

When aspects related to products, as for instance liking, are incorporated in a path model, an additional challenge is apparent; namely how to organize the data (Menichelli, Hersleth, et al., 2014). This situation is typical when interest lies in how different consumer characteristics relate to liking of the different product types (Asioli et al., 2017). Different possibilities exist, as it was demonstrated in Menichelli, Hersleth, et al. (2014). In that paper an organization was recommended where consumers were represented as rows and attributes were organized as columns. Such attributes could consist of both consumer attributes from various questionnaires and/or liking of the different samples (Fig. 1). It was shown in Menichelli, Hersleth, et al. (2014) that with this organization of the data, an ANOVA would be needed to assess the main effects for products. After having eliminated the main effects for products by double centering, the focus is on the 'interactions' between consumer and product.

The assumption of uni-dimensionality in PLS path modelling

Classical path modelling methods like for instance the PLS path modelling require that each block is uni-dimensional (Tenenhaus et al., 2005; Vinzi, Trinchera, & Amato, 2010) or at least that the main variability in each block can be represented by one latent variable only. In particular when product liking values are incorporated, this is in most cases an overoptimistic assumption (Menichelli, Almøy, Tomic, Olsen, & Næs, 2014). One cannot simply assume that the liking of, let us say 5 products, can be decomposed into one principal component. For attitudes and habits, uni-dimensionality is often not a problem since most questionnaires are constructed in such a way that uni-dimensionality is obtained (so-called validated scales) (Karalus & Vickers, 2016; Roininen, Lahteenmaki, & Tuorila, 1999).

A number of different strategies for handling the uni-dimensionality challenge have been proposed (Martens, Tenenhaus, Esposito Vinzi, & Martens, 2007; Menichelli, Hersleth, et al., 2014). Most of these are typically based on splitting blocks up into uni-dimensional sub-blocks and in this way increasing the total number of blocks and then possibly also making interpretation more complex (Nguyen, Næs, Almøy, & Varela, 2020). An alternative approach based on the SO-PLS regression from multi-block analysis has therefore been developed (Menichelli, Almøy, et al., 2014; Næs, Tomic, Mevik, & Martens, 2011; Romano, Tomic, Liland, Smilde, & Næs, 2019). This method does not require uni-dimensionality and can be used for any dimensionality in the original data sets.

The present paper is a comparison of the SO-PLS method for path modelling with PLS-PM accompanied with a strategy for splitting blocks into sub-blocks for handling multi-dimensionality. The particular strategy chosen is simple to apply and is natural to use in this type of studies (Menichelli, Hersleth, et al., 2014). The data sets used here are both based on studies of satiety and specifically related to this challenge, i.e. how to analyze path models when focus is on product related variables.

Satiety perception of products has for several reasons become an important area of research, linked to healthy eating (Brunstrom & Rogers, 2009; Brunstrom & Shakeshaft, 2009). Although consumer expectations (i.e. liking, satiation, satiety, portion size) have been identified as important, very few studies have considered simultaneously all these expectations for understanding consumer perception; therefore, one potential route would be to combine all these blocks of data in an integrated framework and build a predictive model to interpret their relations (Guillocheau et al., 2018). Such an approach results in a composite data set consisting of four blocks of data: liking ($\boldsymbol{X_1}$), satiation ($\boldsymbol{X_2}$), satiety ($\boldsymbol{X_3}$) and portion size ($\boldsymbol{X_4}$) where the data were collected from the same individuals. The path diagram in Fig. 2 describes how the four blocks are linked in this study.

Although both examples presented here are from satiety studies, the methodological issues are general and applicable also to other disciplines in the sensory and consumer area whenever product related variables are involved. The focus here will be on methodological issues such as interpretability and ease of use of the methodologies considered, but some brief discussion will also be given on results relevant for consumer science.

The main novelty of the paper lies in underlining and solving a major, but little studied problem, related to the assumption of one-dimensional blocks in PLS-PM. The problem is particularly important in the cases where the blocks are based on consumer assessments of samples. In such cases one can seldom rely on the one-dimensional assumption. The focus here is on showing how the SO-PLS method is able to directly solve the problem without prior splitting of blocks with subsequent more complex interpretations.

2. Methodological issues

The methodology considered is developed for analyzing relations between J blocks, $\boldsymbol{X_1}, \boldsymbol{X_2}, \ldots, \boldsymbol{X_J}$ of data. We let k_j be the number of columns in block j, and n will be the number of rows. The index i is used to denote consumer i. The special feature of path modelling is that the blocks are linked either according a notion of causality or sequence in time (see e.g., Fig. 2). In the present paper both aspects are implicitly involved in setting up the scheme. All blocks will be mean centered separately for row-wise, that is for each consumer (as for preference mapping) in order to reduce effect of different use of the assessment scale. Since all regression methods used here will center data for each column, this means that the data blocks will essentially be double centered (see e.g. Endrizzi, Menichelli, Johansen, Olsen, & Næs (2011)) in the analysis.

With this organization of data, an ANOVA model is needed to assess the average importance of the products, so-called main effects for products (Menichelli, Almøy, et al., 2014). This will be done using the standard mixed model with fixed main effects for products, random consumer effects plus random error. The main effects for products will be used for assessing the differences in average product effects over the consumer group. The residuals from the model are double centered and therefore identical to the values used as basis for the path modelling (see below). They can be interpreted as the interactions (plus noise) of consumer and product. In other words, they represent how the consumers vary in their assessment of products, which is exactly the relevant information to be used for path modelling.

2.1. PLS path modelling (PLS-PM)

The principle behind PLS-PM is that an iterative algorithm estimates the relationships among blocks of observed variables (indicators or manifest variables MVs), through the construction of non-observed variables (i.e. Latent variables LVs) which describe the main variability in the MVs. The LVs for the different blocks are then linked according to the path model scheme and the MVs related to their respective LV (see Fig. 3).

The PLS-PM algorithm comprises two different stages, the inner and outer estimation (Tenenhaus et al., 2005; Wold, 1980). In the inner estimation stage, LVs are obtained as weighted aggregates of connected LVs. An LV, which never appears as a dependent variable, is called an exogenous variable. Otherwise, it is called an endogenous variable (Tenenhaus et al., 2005). In the outer estimation step, LVs are calculated as weighted aggregates of their corresponding MVs (Latan & Noonan, 2017). The inner weights $\boldsymbol{e_{ij}}$ are estimated using the so-called Centroid, Factor or Path schemes (Vinzi, Trinchera, et al., 2010). There are two ways to estimate the outer weights $\boldsymbol{w_{jk}}$: reflective (mode A) and formative (mode B). In this paper, we will only consider the reflective mode where all manifest variables in block j are considered linear functions of the corresponding latent variables (plus noise), which is usually most natural in consumer science.

The algorithm begins with arbitrary initial outer weights $\boldsymbol{w_{jk}}$ (for simplicity, all weights can be initialized equal to 1), and then iterates between estimating the inner weights and outer weights. Once the algorithm converges, i.e. the sum of absolute changes in weights from one iteration to another falls below a threshold, for instance, 10^{-5} (Henseler, 2010; Wold, 1982), path coefficients are estimated as simple or multiple regression coefficients according to the system of interdependent equations represented by the path diagram (Vinzi, Chin, Henseler, & Wang, 2010). The details of PLS-PM algorithm are provided in (Tenenhaus et al., 2005; Vinzi, Chin, et al., 2010).

These path coefficients represent the most important parts of the results since they are used for interpretation and for calculation of the indirect and direct effects of the different blocks on each other. Usually they are presented together with their standard errors directly in the path diagram (see results section).

Using the path coefficients, the effects (direct, indirect and total) are defined as:

- Direct effects are given by path coefficients, i.e. regression coefficients for the inner relations;
- Indirect effects represent the influence of one block on another block by taking an indirect path calculated as the product of path coefficients;
- Total effects are the sums of both direct and indirect effects.

If there is no relation from one LV to another LV, the effect will be equal to zero. This will apply for both direct and indirect effects.

The bootstrap can be applied to estimate the precision of direct, indirect and total effects. The bootstrap procedure is the following: M samples are created in order to obtain M estimates for each parameter in the PLS model. Each sample is obtained by sampling with replacement from the original data set, with sample size equal to the number of cases in the original data set. The bootstrap estimates are performed with the R package *plspm* (Sanchez, 2013; Sanchez, Trinchera, & Russolillo, 2017).

Alternative approaches for handling the lack of uni-dimensionality

One of the problems with PLS-PM is that it requires uni-dimensionality of the blocks. Various methods exist for solving the problem; for example, removing manifest variables that are far from the model (e.g., manifest variables that are not pointing in the same direction as the other variables in a block), changing the measurement model into a formative model, using a hierarchical model approach or splitting the multidimensional block into uni-dimensional sub-blocks (Becker, Klein, & Wetzels, 2012; Menichelli, Hersleth, et al., 2014; Vinzi, Trinchera, et al., 2010). Although these approaches deal with the uni-dimensionality, they, in general, change the nature of data (removing manifest variables, changing the measurement model) or making the structural model more complicated (using hierarchical model, splitting into uni-dimensional sub-blocks). The approach taken here is one of splitting a block according to the main principal components with a subsequent interpretation of the components as suggested by Menichelli, Hersleth, et al. (2014). However, it is not a straightforward task to decide the number of sub-blocks, especially in cases of complex samples (Nguyen et al., 2020).

2.2. SO-PLS for path modelling (SO-PLS-PM)

Another possibility is to use the newly developed SO-PLS path modelling (SO-PLS-PM) which handles multi-dimensionality directly without any pre-processing (Menichelli, Almøy, et al., 2014; Næs et al., 2011; Romano et al., 2019). As opposed to the methods mentioned above, the SO-PLS-PM method easily handles different underlying dimensionality of the blocks. In addition, it is invariant to the relative scaling of the blocks, meaning that no preprocessing is needed for balancing the influence of the blocks.

The rationale behind SO-PLS-PM is to model each endogenous block separately as a function of all blocks that are input to it (Menichelli, Almøy, et al., 2014; Næs et al., 2011). The separate SO-PLS models (for endogenous blocks) can be interpreted in different ways using the additional explained variance as new blocks are incorporated, the individual PLS models for each block and the principal components of prediction (PCP) method (Langsrud & Næs, 2003).

SO-PLS for multiblock regression

Let us now assume that data consists of three blocks in which $\boldsymbol{X_1}, \boldsymbol{X_2}$ are the explanatory blocks and $\boldsymbol{Y}$ is the response block. Their relations are described as follows:

$$\boldsymbol{Y} = \boldsymbol{X_1 B_1} + \boldsymbol{X_2 B_2} + \boldsymbol{error} \tag{1}$$

where $\boldsymbol{B_1}, \boldsymbol{B_2}$ are regression coefficients.

The SO-PLS method for estimation is based on an iterative use of PLS regression and orthogonalization of blocks with respect to blocks previously fitted, summarized by the following steps: the first step is to fit $\boldsymbol{Y}$ to $\boldsymbol{X_1}$ by PLS regression. The $\boldsymbol{X_2}$ is then orthogonalised with respect to the PLS scores $\boldsymbol{T_{X_1}}$ of $\boldsymbol{X_1}$ to obtain the orthogonalized $\boldsymbol{X_2^{orth}}$; in the second step, the original or deflated $\boldsymbol{Y}$ is fitted to $\boldsymbol{X_2^{orth}}$ using PLS regression, and the PLS scores $\boldsymbol{T_{X_2^{orth}}}$ are estimated; finally, $\boldsymbol{T_{X_1}}$ and $\boldsymbol{T_{X_2^{orth}}}$ are used as independent variables to predict response variables $\boldsymbol{Y}$ in an ordinary least squares (LS) regression. For more blocks, one simply repeats the same procedure. This method provides information of the incremental increase in the explained variance as each new block is incorporated. This is called the additional effect of a block and is important for interpretation.

Determining the number of components

As for regular regression, cross-validation is applied to determine the number of components to use for prediction and assess the quality of the predictor obtained, usually measured by the root mean square error of prediction (RMSEP) (Martens & Næs, 1989). In the SO-PLS regression, the optimal number of components can be selected using global or sequential optimization (Næs et al., 2011). In this paper, we will use the sequential approach since it fits best with the philosophy of using SO-PLS in a path modeling context, i.e. with a focus on additional explained variance.

Direct and indirect effects

Assume that block A imparts block C directly and indirectly through block B (see for instance the (Liking, Satiation, Satiety) part of the model in Fig. 2 with Liking represented by A, Satiation by B and Satiety by C). The effects are defined in the following way:

- The total effect of block A on block C is the explained variance (in %) of C when regressed onto A;
- The direct effect of A on C is defined by how much of C can be explained by A when A is orthogonalized with respect to B;
- The corresponding indirect effects are calculated as the differences between the total effects and the direct effects.

If all information from A to C goes through B, this direct effect will be equal to zero; in all other cases, it will be positive. In order to avoid overoptimistic results, cross-validation is used to estimate the explained variances.

The number of components for the effects are selected as follows: for total effect of A on B and C, the components are selected for A independently for each; for total effect of B on C, the components are selected directly. For direct effect of A on C, the steps are to select first components in B for predicting A, and then components in the residuals of A from this model when predicting C.

For models with more blocks the components are selected in the same way.

With direct, indirect and total effects in SO-PLS-PM a model-based bootstrap is performed where residuals are permuted (see Romano et al. (2019) for details).

Principal components of predictions (PCP)

The PCP aims at providing information about which part of a response block $\boldsymbol{Y}$ can be predicted by which part of a predictor block $\boldsymbol{X}$ (Langsrud & Næs, 2003). The first step is to use PCA on the predicted values $\widehat{\boldsymbol{Y}}$. This gives $\widehat{\boldsymbol{Y}} - \boldsymbol{scores}$ and $\widehat{\boldsymbol{Y}} - \boldsymbol{loadings}$. The $\boldsymbol{X} - \boldsymbol{loadings}$ are obtained by regressing each X-variable onto the $\widehat{\boldsymbol{Y}} - \boldsymbol{scores}$**.** This results in one score plot (for $\widehat{\boldsymbol{Y}}$), and two loading plots (one for $\boldsymbol{X}$, one for $\widehat{\boldsymbol{Y}}$) for each model fitted. Usually, one will concentrate on the first two components of $\widehat{\boldsymbol{Y}}$, but more components are possible (Menichelli, Almøy, et al., 2014; Næs et al., 2011).

3. Case studies

3.1. Yoghurt data

Eight yoghurt samples were prepared from a design of experiment (DOE) based on the same ingredients, but with different texture obtained by using different processing strategies. The samples have the same calories and composition avoiding influence of these parameters on satiety or satiation. The ingredients were commercial natural yoghurt, cereal flakes and a combination of vanilla and high intensity sweetener. The design parameters of the full factorial design were yoghurt viscosity (thin/thick), cereal particle size (flakes/flour) and flavour intensity (low/optimal); see Nguyen, Næs, &

Varela (2018) for details. Table 1 shows the samples with different levels of viscosity, particle size and flavour intensity.

One hundred and one consumers were recruited for the test in the southeast area of Oslo from Nofima's consumer database. Consumers were asked to taste each sample and rate their liking on a Labeled Affective Magnitude (LAM) scale (Schutz & Cardello, 2001), expected satiation on a Satiety Labeled Intensity Magnitude (SLIM) scale (Cardello, Schutz, Lesher, & Merrill, 2005) and expected satiety on a 6-point scale from 1 = "hungry again at once" to 6 = "full for five hours or longer". For their ideal portion size, they chose the amount they would consume as compared to the normal amount of commercial yoghurt product (they were shown a commercial unbranded container). The labeled points on the portion size scale were defined in relation to the provided container as follows: "One-third (of the container)", "A half", "Two-thirds", "One container", "One and a half", "Two", "Three".

3.2. Biscuit data

Eight oat based biscuit samples were used in this study. Samples were prepared following the same idea as for the yoghurt samples, identical composition but different textures. Two parameters of DOE were used: baking powder in two levels (with/without) and four levels of particle sizes (0.5mm, 2.0mm, small commercial flakes, big commercial flakes). The formulations of biscuit samples are shown in Table 2. A consumer test was carried out with 101 consumers at IATA (Valencia, Spain). In this test, consumers tasted the samples and rated the same parameters as in the yoghurt case: liking on LAM scale, expected satiation on SLIM scale and expected satiety on 6-point scale. For portion size selection, they rated how many biscuits they would like to eat on a 6-point scale from "1 biscuit" to "6 or more biscuits".

3.3. Data analyses

The data sets consist in both cases of four blocks $\boldsymbol{X_1}, \boldsymbol{X_2}, \boldsymbol{X_3}, \boldsymbol{X_4}$ corresponding to *liking*, *satiation*, *satiety* and *portion size*. Rows correspond to consumers as discussed above. Before analysis, data are centered for each consumer (as in preference mapping) and block separately (each row) which leads to double-centered data since PCA and PLS regression are always run on column centered data (Endrizzi, Gasperi, Rødbotten, & Næs, 2014; Endrizzi, Menichelli, Johansen, Olsen, & Næs, 2011).

Each uni-dimensional block for PLS-PM (obtained by the splitting step based on principal components) is standardized by dividing by its standard deviation (Tenenhaus et al., 2005). Note that reducing the blocks to two components, means that focus in the path model will be only on the aspects related to these two components (see SO-PLS-PM below for a comparison of this and the results for the full data set).

For the SO-PLS-PM, we here compared solutions based on original data and the principal components (still standardized individually by the standard deviation) used as input for the PLS-PM. The two principal components representing a block (as for PLS-PM) will here, however, not be used separately, only together in a block. For the original data, each original block is double centered as described above and then standardized by dividing by its Frobenius norm (although not needed due to invariance). Also, for the situation with the two principal components (T1 and T2) used together, standardization by the Frobenius norm is applied. Note that comparing results for two components and all the data for blocks implicitly gives a test on whether one loses important information for the path diagram by focusing only on two components.

The R packages *plspm* (Sanchez et al., 2017) and *semPLS* (Monecke & Leisch, 2012) are used for implementing PLS-PM. The computations of SO-PLS are done in Python and SO-PLS-PM in MATLAB with in-house codes.

3.4. Path model considered

For both yoghurt and biscuit data sets, the path diagrams describe the relations between blocks of variables with respect to the sequence of cognitive and physiological processes when people consume a food product (Blundell et al., 2010). This diagram is depicted in Fig. 2 in which liking is incorporated before satiation and satiety expectations, and then these three blocks together impart portion size (Nguyen et al., 2020). This diagram is used directly in the SO-PLS-PM analyses.

For the PLS-PM, the splitting step is done as illustrated in Fig. 4. Instead of the original model (on the upper right side), one applies the PLS-PM on the new one (on the lower right side) which satisfies the assumption of uni-dimensionality. This is essentially the same diagram as in SO-PLS-PM, the only difference is that now each block was replaced by two different blocks with one variable (principal component) in each. The components from the same original block are independent principal components and therefore no relation between them is used in the model.

4. Results

For each data set, two main results were represented; in particular, first the main effect (*product* effect) on consumer expectation (i.e. liking/ satiation/ satiety/ portion) was considered, then the interactions (see beginning of Section 2) between product and consumer effects were investigated in the context of path modelling.

4.1. Yoghurt data

4.1.1. The main effect of product

The average differences in ratings (liking, satiation, satiety, portion) between products were depicted in Fig. 5. The mixed ANOVA model (as described above) showed significant differences between products for liking, satiation, satiety, portion with p-values <0.001. Added to this, the standard errors of the means were added to point at the product separations for each rating. For *liking*, there are four groups of products in the ascending rating: group 1 (TnFkL, TnFrL), group 2 (TnFrH, TnFkH), group 3 (TkFrL, TkFkL), and group 4 (TkFrH, TkFkH). There are two classifiers for this separation: the first one, thickness, distinguishes group 1, 2 (thin products) from group 3, 4 (thick products); the second one, flavor intensity, separates group 1 (low intensity) vs. group 2 (high intensity), and group 3 (low intensity) vs. group 4 (high intensity). For the remaining consumer expectations (*satiation*, *satiety*, *portion*), it is important to see that the difference between products depends on thickness only with products TnFkL, TnFrL, TnFrH, TnFkH in one group, and products TkFrL, TkFkL, TkFrH, TkFkH in another group.

The results do not only highlight how consumers rate their expectations on different products, but also indicate the possible relationships between these expectations due to the similar separations when considering liking, satiation, satiety, portion.

4.1.2. PLS-PM

A PCA was applied to each block (consumers in rows and ratings of products in columns) to split original block into uni-dimensional sub-blocks. With the help of sensory attributes (as supplementary variables) the PCA components were interpreted. For liking, the first component is explained by viscosity with *Thick* and *Liquid* attributes located on opposite sides, whereas the second component is characterized by the particle-size (*Sandy* and *Pieces*). These results are also observed for satiation and portion size, however, for satiety, the components are switched in which the first component became particle-size and the second component was viscosity. The two components explain around 50% of the variation and have clear interpretation for all blocks of data considered. For all blocks the general direction or separation of products in each PCA loading plot is the same for all blocks, i.e. a positive score for particle size for liking corresponds to a positive score for particle size in the other blocks. The loading plots are displayed in the supplementary material in Appendix A (Fig. A1). Component 3 was also discussed briefly in Nguyen et al. (2020), but this did contribute little to the interpretation while also making the model more complicated and was therefore omitted here. The two components were used as separate blocks in the PLS-PM. It is beyond the scope of the present paper to discuss details of product characterizations, but they are available from Nguyen et al. (2020). From now on, the paper will focus on the first two components: the one related to viscosity (V) and the other related to particle-size (P), for example, *LikingV* will be the liking component driven by viscosity, *LikingP* will be the liking component driven by particle size, and so on for the other blocks.

Fig. 6 highlights the relations between the four data blocks using the (V, P) notation. Blue lines indicate positive relationships, red lines negative relationships, dashed lines close to zero relation and the thickness of the lines represent the strengths of the direct relationships between two blocks. It can be noted that all variables were standardized, so that the path coefficients could be compared. The path coefficients are displayed with the corresponding P-values in parentheses.

As can be seen, liking has positive and strong effect on portion size with path coefficients of 0.44 and 0.72 for the component V and P, respectively. In addition, while liking directly influences satiation (*LikingV-SatiationV*: 0.30, *LikingP-SatiationP*: 0.37), it does not contribute directly to satiety for each component separately. On the other hand, satiation strongly (and directly) imparts satiety (*SatiationV-SatietyV*: 0.41, *SatiationP-SatietyP*: 0.48).

The direct, indirect, and total effects and their corresponding P-values are found in Table 3 in which the relations with non-significant values of all direct, indirect, and total effects were eliminated (13 out of 24 relations). It is noted that, in the relation *LikingV-SatietyP*, both indirect and total effects are equal to 0.11 but differ in P-values (0.024 vs. 0.356). A somewhat strange aspect can be noted for the bootstrap-based significance values for *LikingV* vs *SatietyP*; the indirect effect is the same as the total, but the significances are quite different. This is probably due to the fact that in the bootstrap the indirect and total effects are different in each bootstrap replicate, even though the estimate is the same.

In addition to the effects, for each regression in the structural model, the R^2 (the proportion of variance in endogenous LV that is predictable from its independent LVs) is investigated. It is not surprising that *PortionP* is the most explained block with R^2 = 49.8%, followed by *SatietyV* (31.67%) and *SatietyP* (24.82%).

In summary, we can say that liking affects directly both portion size and satiation. Neither satiation nor satiety affect portion size in any significant way. Satiation has a direct effect on satiety. The direct effects dominate completely, only 3 of the indirect effects are significant. The significant effects follow either P or V except the one direct effect from *LikingP* to *SatietyV* (and to a certain extent the indirect of *LikingV* on *SatietyP*). The latter two aspects are somewhat difficult to interpret, in particular the last is difficult given the general structure/size of effects seen in Fig. 6.

4.1.3. SO-PLS-PM for raw data without reduction based on PCA

An essential step here is to determine the number of components for each data block used in the SO-PLS-PM estimation. Based on the path diagram, three SO-PLS models were considered: (1) *Liking* → *Satiation*, (2) *Liking* + *Satiation* → *Satiety*, and

(3) *Liking* + *Satiation* + *Satiety* → *Portion*. For each model, the number of components was selected sequentially by optimizing for the first block and then for the next block while keeping the number of components of previous blocks fixed (sequential optimization). The RMSEP plots (Måge, Mevik, & Næs, 2008), as functions of the total number of components for all three regression methods, show that model 1 was optimized with 5 components of *Liking*; model 2 with 1 component of *Liking* and 5 components of *Satiation*; model 3 with 5 components of *Liking*, 0 component of *Satiation* and 0 component of *Satiety* (Fig. A2 in Appendix A).

The cumulative validated explained variances are displayed in Table 4. For model 1 (*Liking* → *Satiation*), *Liking* predicts 10.5% of the variability of *Satiation*. For model 2 (*Liking* + *Satiation* → *Satiety*), *Satiety* is mostly explained by *Satiation* (14.2%) since *Liking* only explained 0.9% of *Satiety* variance. For model 3 (*Liking* + *Satiation* + *Satiety* → *Portion*), only *Liking* is considered as the regressor of *Portion*, it predicts 20.6% of *Portion* variance. These results clearly indicate a multi-dimensional structure of each data block.

The SO-PLS-PM path diagram (Fig. 7) shows three main/significant relations based on the direct effects: *Liking-Portion*, *Liking-Satiation* and *Satiation-Satiety* with the ‘path coefficients’ (i.e. explained variances) 20.64, 10.45 and 19.23, respectively. These results are consistent with those of PLS-PM which emphasize the relations *Liking-Portion*, *Liking-Satiation* and *Satiation-Satiety*.

The relations *Liking-Portion* and *Satiation-Satiety* are two times higher than the relation *Liking-Satiation*. The relative strengths are slightly different in PLS-PM results where the relations *Liking-Portion* and *Satiation-Satiety* are not twice as high as the relation *Liking-Satiation*, especially regarding the component V. Apart from the relative strengths of relations, the only clear difference is the lack of significant relation between

Liking and *Satiety (*although this effect was quite difficult to interpret for PLS-PM). The indirect and total effects are displayed in Table 5. It can be seen that there are no indirect effects. Total effects were therefore the same as the direct effects.

For further interpretation, PCP plots were obtained for each model. For model 1 (*Liking* → *Satiation*) and 3 (*Liking* + *Satiation* + *Satiety* → *Portion*), it is clear that *Liking* has a positive (i.e. in the same direction) effect on *Satiation* and *Portion* due to the similar configurations between *Liking*, *Satiation* and *Portion* (Fig. A3 in Appendix A). For model 2 (*Liking* + *Satiation* → *Satiety*), the loading plots of the explanatory blocks (i.e. *Liking*, *Satiation*) and response block (i.e. *Satiety*) show that both *Liking* and *Satiation* influence *Satiety* positively. As can be seen in $\hat{Y}$ – loadings (Fig. 8b), the first component separates satiety ratings into two groups: one group (P7, P8, P4) on the left, and another group (P1, P3, P5, P6, P2) on the right side, which is in line with liking or satiation separations (Fig. 8a). On the second component, the classifications of liking, satiation and satiety ratings are roughly consistent with P7, P1, P3 on the top and P4, P2, P5, P6 on the bottom of this component. This shows that an increase in liking and/or satiation results in an increase in satiety.

4.1.4. SO-PLS-PM on preprocessed data

To investigate the effect of the PCA preprocessing step on SO-PLS-PM results, the SO-PLS-PM was also applied on the two components data. Table 6 shows that the direct effects in this model are slightly different as compared with those of SO-PLS-PM on the original data. The main relations are, however, the same: *Liking-Portion* (31.8), *Liking-Satiation* (8.93), and *Satiation-Satiety* (20.18). Consequently, SO-PLS-PM could be used on the original data without changing the main relations between variables.

4.2. Biscuit data

4.2.1. The main effect of product

Like for the yoghurt data, the consumer ratings (liking, satiation, satiety, portion) in different products were also tested for biscuit data (Fig. 9). The mixed ANOVA model (as described above) showed significant differences between products for liking (p-value < 0.001), satiety (p-value 0.012), portion (p-value 0.017), but not for satiation (p-value 0.607). Standard errors of the means were also added to point to the product separations for each rating. There is no clear separation in ratings between products; however, it seems that product s3w is rated high, and product s4wo low in both liking and portion while product s1wo is expected to be the most satiety and, to a certain extent, satiation, indicating the possible relations of liking-portion, and satiation-satiety.

4.2.2. PLS-PM

The same strategy of analyses was applied to the biscuit data set. First, PCA was run on double-centered data; however, the PCA plots did not show the same clear interpretations as for the yoghurt data. For liking and portion size (Fig. 10), there seems to be quite similar classifications along the first component with the product s4w and s4wo (oat flakes in big size, with or without baking powder) on one side and the rest of the products (oat flakes in small size or oat flour, with or without baking powder) on the other side. One can say that the first component can be explained by the differences in particle-size, meaning that samples with big flakes (s4) are separated from the other samples (s1, s2, s3). Component 2 is difficult to explain both for portion and liking, with no clear effect of the baking powder on the perception. Possibly, the component is a combination of two input factors (i.e. particle-size and baking powder), pointing their interaction. Added to this, differences among samples are smaller and the variation in liking/ portion is low, for these reason PCA has no straightforward explanation. This is also observed when considering satiation and satiety with the same tendencies (Fig. B1 in Appendix B).

This implies that the meaning of the first two components is not related to single identifiable properties like viscosity and particle-size in the yoghurt case, and we therefore use the names “1” and “2” as the first and second component in the next analyses. An alternative here could have been to let the different samples represent separate blocks of data as also discussed in Menichelli, Hersleth, et al. (2014), but that would lead to an enormous number of blocks and relations that would be very difficult to interpret. We therefore kept the same procedure as for the yoghurt data and interpret further only the main relations found in the path model below using the PCA plot. Later, it will become evident that component 2 is of less importance in the path diagram than component 1.

The PLS-PM path diagram (Fig. 11) shows the relations between data blocks with the corresponding path coefficients (in the same way as for Fig. 6). The direct, indirect and total effects are given in Table 7 (15 out of 24 relations were eliminated due to non-significant direct, indirect, and total effects with P-values higher than 0.05). In this case, strong positive relations are mostly related to component 1: *Liking1-Satiation1* (0.3), *Satiation1-Satiety1* (0.53), *Satiety1-Portion1* (0.48). There is no significant relation between the two blocks related to component 2, but *Satiation2-Satiety2* estimate (0.29) is close to significance with a P-value 0.09. As can also be seen, *Liking1* is not only related to component 1 but also to component 2; for example, *Liking1-Satiation2* (0.2) in a direct way and *Liking1-Satiety2* (0.11) in an indirect way. In addition, *Satiation1* imparts on both *Portion1* and *Portion2*, but in opposite ways. More specifically, *Satiation1* indirectly imparts *Portion1* with a positive effect (0.20); however, it directly influences *Portion2* with a negative effect (-0.27). These results imply that component 1 dominates the path diagram in the sense that component 1 affects other blocks related to component 1, but also a few related to component 2.

As mentioned previously, the interpretation of the component 2 was difficult, but as can be seen, this component is less important than component 1 in the path diagram. While component 1 displays some main relations: liking-portion size, liking-satiation, satiation-satiety and satiety-portion size, component 2 does not depict any clear relation (at least used as input block). A possible explanation is that consumers relate their expectations (i.e. liking, satiation, satiety and portion size) mostly depending on the particle-size of samples (i.e. component 1 for all blocks of data).

Considering calibrated explained variances (R^2) (note that the explained variances for the SO-PLS-PM are validated with cross-validation and will therefore always be smaller) of data blocks in the structural model, blocks related to component 1 are explained more effectively than those linked to component 2. Among the data blocks, the most explained block is *Portion1* (40.65%), and the least one is *Satiation2* (6.54%).

In summary, the paths related to the blocks driven by component 1 (i.e. particle-size component) are dominating. Generally, liking directly affects portion. Added to this, liking directly influences satiation (both *Satiation1* and *Satiation2*), and then satiation influences satiety. The main difference in this predicted model, as compared to the model for the yoghurt data, is the relation satiety-portion. While this relation (*Satiety1-Portion1* in particular) seems to be significant in the biscuit data, it is not in the yoghurt data. It means that people who expect to feel fullness in longer duration will select a larger amount of food. Care should be taken interpreting this relation because it is only based on component 1.

4.2.3. SO-PLS-PM for raw data without PCA based reduction

Like for the yoghurt data, three SO-PLS models were considered: (1) *Liking* $\rightarrow$ *Satiation*, (2) *Liking* + *Satiation* $\rightarrow$ *Satiety*, and (3) *Liking* + *Satiation* + *Satiety* $\rightarrow$ *Portion*. For model 1, the RMSEP plot shows that *Satiation* is not predicted by *Liking* (0 component of *Liking*). For model 2, 5 components for *Satiation* are selected for predicting *Satiety*. For model 3, *Portion* is explained by 2 components of *Liking*. The corresponding RMSEP plots were shown in Fig. B2 (Appendix B).

Validated explained variances were calculated for each SO-PLS model (Table 8). Model 1 has no predictive power and is not further explained. In model 2, 9.5% of the variability of *Satiety* is explained by *Satiation* and not by *Liking*. Conversely, in model 3, *Portion* is predicted by *Liking* only; in particular, *Liking* explains 7.1% of *Portion* variances.

The relations between blocks were calculated (Table 9) and the path diagram was plotted (Fig. 12). No indirect effects are observed. According to Fig. 12, there are two main relations: *Satiation-Satiety* (15.04) and *Liking-Portion* (7.14). In this path model, the relation *Liking-Satiation* is not found to be significant, whereas it is in the PLS-PM estimation (*Liking1-Satiation1*: 0.3 and *Liking1-Satiation2*: 0.2). Furthermore, the relation *Satiety-Portion* is not significant in SO-PLS-PM estimation, but considerable in the PLS-PM model (*Satiety1-Portion1*: 0.35). In other words, the main difference in terms of significance are the paths between liking and satiation, and satiety and portion size. In fact, the relation *Satiety-Portion* appears and is equal to 1.27, however, the bootstrap-based standard error is high (1.27). Consequently, this relation becomes non-significant.

PCP loading plots were used to interpret the relations between blocks in the path model (Fig. 13). As can be seen in Fig. 13a, the relation *Satiation-Satiety* is positive because their configurations are consistent. In particular, the first component splits the ratings (both satiation and satiety) into two groups: P2, P7, P8 on the left-hand side and P4, P5, P6 on the right-hand side. On the second component, while ratings of P3 and P1 are positioned on the top, ratings of P5 are on the bottom of the loading plot. The plot indicates consensus classifications between satiation and satiety ratings, that is, when satiation ratings increase, satiety ratings also increase, and conversely. This result is consistent with PLS-PM results in which increasing satiation also leads to enhanced satiety perception. Likewise, *Liking-Portion* is considered as a positive relation (Fig. 13b).

It can be noted, for the SO-PLS-PM, that no initial PCA with difficult interpretation is needed.

4.2.4. SO-PLS-PM on preprocessed data

Again, for comparison, SO-PLS-PM was applied to the preprocessed biscuit data. Although the complexity of the data increased (i.e. more complicated in terms of consumer expectations), the effects are still similar as compared with those of SO-PLS-PM on the original data. Particularly, the main relations *Satiation-Satiety* and *Liking-Portion* are 14.53 and 7.27, whereas they are 15.04 and 7.14 in SO-PLS-PM on original data. It is noted that the relation *Satiety-Portion* is 5.58, but its standard error is also high (4.81). Therefore, it was not significant at a 5% level of significance.

5. Discussion

The main focus of this paper been on how to handle multi-dimensionality of blocks in path modelling in consumer science. Special emphasis was given to a method based on principal components proposed in Menichelli, Hersleth, et al. (2014); Nguyen et al. (2020) for obtaining uni-dimensional blocks in PLS-PM. The results from this analysis were compared to results from SO-PLS-PM which handles multi-dimensionality automatically. Interpretation of the results in the context of satiety and satiation were also considered. The focus here was more on the statistical implications rather than the perceptual interpretations of the results. More details on the sensory perception aspects of the yoghurt data can be found in Nguyen et al. (2018).

Uni-dimensional blocks from complex data

To ensure the assumption of uni-dimensionality which is necessary for PLS-PM, PCA was used as a preprocessing step for both data sets (Menichelli, Hersleth, et al., 2014). For the yoghurt data, this strategy works well since the two dominating components are easily interpretable as viscosity and particle-size related. For the biscuit data on the other hand, it is more difficult to interpret the components, which complicates the whole procedure. In other words, the method of splitting based on PCA components was less successful for the biscuit data than for the yoghurt data. The comparison with SO-PLS-PM indicates, however, that in both cases two components capture the most important information for the path modelling.

It must be underlined that other ways of splitting a data block is hard to find in this type of studies with products in focus. Splitting original blocks into uni-dimensional blocks can in general make the interpretation of the path model more complicated since many more blocks have to be taken into account. Some of those relations also seem to be confusing (e.g., *Liking2-Satiety1*, *Satiation1-Portion2*), resulting in difficulty of interpretation of the PLS-PM path model results.

As opposed to PLS-PM, SO-PLS-PM can be applied to the original data without a PCA preprocessing step, and then the interpretations are more straightforward. In addition, PCP loading plots are used to explain how different exploratory blocks are related to the response block.

Resampling – overfitting

As can be seen from the PLS-PM path diagram, the relation *Liking-Satiety* is deemed significant for both yoghurt and biscuit data, but it is not in SO-PLS-PM. A possible explanation for this is that the resampling tests for the effects based on cross-validation are more conservative since they represent a bootstrap on top of a cross-validated estimate. Another possible and related explanation is that the standard PLS-PM is more prone to overfitting. To check this possible overfitting, PLS regression of satiety on liking for both data sets (data not shown) was employed, and the result in fact pointed out that liking explains very low variability of satiety (as opposed to the indication in the PLS-PM results). This points towards simple PLS testing of relations (with cross-validation) if interpretation is found confusing or difficult.

The direct, indirect and total effects

The effects are used to interpret the relations between variables in both PLS-PM and SO-PLS-PM; however, their definitions are different depending on the method used. In PLS-PM, direct effects (also called path coefficients) are the regression coefficients, whereas in SO-PLS-PM, they are the explained variances. This leads to differences in indirect and total effect calculations, but results in Romano et al. (2019) indicated that, in the case of uni-dimensional blocks, they measure the same phenomena. The comparison between PLS-PM and SO-PLS-PM on the path coefficients should generally focus on the main trends instead of the absolute values (see also Romano et al. (2019)). As aforementioned, the values of explained variances in SO-PLS-PM seem to be lower than those of PLS-PM. This is reasonable because these values are *validated*, explained variances calculated by cross-validation instead of just fitted R^2's. In addition, the explained variance results for the SO-PLS-PM are related to the manifest variables while for the PLS-PM they refer to the relation between latent variables.

Other ways of organizing the data

It should also be mentioned, that since both ‘variables’ and ‘samples’ are the same for all blocks, this study could also have been conducted using transposed matrices, but this idea is not pursued here. Note, however, that the same problem of uni-dimensionality would appear also with that approach. In cases, where the variables are different in the different blocks, which is the usual case, such a transposed procedure is not possible.

Further research

There are ongoing discussions on the efficacy of PLS-PM. Some researchers seem to be more inclined to use methods such as common factor models and multi-level modelling (Rönkkö & Evermann, 2013; Rönkkö, McIntosh, & Antonakis, 2015; Rönkkö, McIntosh, Antonakis, & Edwards, 2016). The aim of the present paper, however, is to focus on other aspects, that is, how to deal with the assumption of uni-dimensionality. The SO-PLS-PM presented here is one possibility to solve this issue. Nevertheless, other solutions have been proposed such as summarizing each block by the first principal component (Tenenhaus, 2008) or using multiple dimensions in higher-order constructs (Becker et al., 2012). The SO-PLS-PM should be compared also to these approaches and to other path modelling methods such as Path-ComDim (Cariou, Qannari, Rutledge, & Vigneau, 2018) or RGSCA (Hwang, 2009; Hwang & Takane, 2004).

As a matter of fact, the SO-PLS-PM itself may also face some limitations. One of them is how to establish the dependence order of data blocks if the so-called topological order is not unique (as it is here). In such cases one will need to establish a relation in terms of what is most natural from the researcher's point of view.

6. Conclusion

The main purpose of the path models here was to predict portion size from liking, expected satiation and satiety using PLS-PM and SO-PLS-PM. A procedure based on the use of principal components instead of the original data were tested in order to make the data uni-dimensional, which is a requirement for PLS-PM. For the yoghurt data set, although there were differences in the numerical absolute values, the two approaches showed the same main trends: liking was the essential regressor of expected satiation and portion size; and expected satiation mainly predicted expected satiety. When the complexity of consumer expectations increased, because of higher sensory complexity of a solid product, the uni-dimensionality was not handled well by the PCA preprocessing step as was illustrated using the biscuit data set. The relation between liking and expected satiation became complicated and difficult to interpret in the PLS-PM model. In other words, the splitting procedure tested is not always to be recommended in PLS-PM.

In this study, SO-PLS-PM reveals the ability to model data sets which violate the assumption of uni-dimensionality without requiring any data preprocessing step. This makes the explanation more explicit and avoids the potential problems when applying standard PLS-PM on uni-dimensional blocks obtained by splitting original data blocks.

Acknowledgements

The author Quoc Cuong Nguyen thanks the financial support funded by Ho Chi Minh City University of Technology - VNU-HCM under grand **number T-KTHH-2019-11**. The authors would also like to thank for the financial support received from the Norwegian Foundation for Research Levy on Agricultural Products FFL, through the research program “FoodSMaCK, Spectroscopy, Modelling and Consumer Knowledge” (2017-2020). Special thanks go to Hilde Kraggerud (Tine, Norway) for the support with the sample materials, to Stefan Sahlstrøm (Nofima) for his help with the milling procedure, to Andre Løvas (Nofima) for the help with the baking process, and to Arantxa Rizo, Amparo Gamero for the help with the consumer test in Spain.

References

Asioli, D., Varela, P., Hersleth, M., Almli, V. L., Olsen, N. V., & Næs, T. (2017). A discussion of recent methodologies for combining sensory and extrinsic product properties in consumer studies. *Food Quality and Preference, 56, Part B*, 266-273.

Becker, J.-M., Klein, K., & Wetzels, M. (2012). Hierarchical Latent Variable Models in PLS-SEM: Guidelines for Using Reflective-Formative Type Models. *Long Range Planning, 45* (5), 359-394.

Blundell, J., De Graaf, C., Hulshof, T., Jebb, S., Livingstone, B., Lluch, A., Mela, D., Salah, S., Schuring, E., Van Der Knaap, H., & Westerterp, M. (2010). Appetite control: methodological aspects of the evaluation of foods. *Obesity Reviews, 11* (3), 251-270.

Brunstrom, J. M., & Rogers, P. J. (2009). How Many Calories Are on Our Plate? Expected Fullness, Not Liking, Determines Meal-size Selection. *Obesity, 17* (10), 1884-1890.

Brunstrom, J. M., & Shakeshaft, N. G. (2009). Measuring affective (liking) and non-affective (expected satiety) determinants of portion size and food reward. *Appetite, 52* (1), 108-114.

Cardello, A. V., Schutz, H. G., Lesher, L. L., & Merrill, E. (2005). Development and testing of a labeled magnitude scale of perceived satiety. *Appetite, 44* (1), 1-13.

Cariou, V., Qannari, E. M., Rutledge, D. N., & Vigneau, E. (2018). ComDim: From multiblock data analysis to path modeling. *Food Quality and Preference, 67*, 27-34.

Carrillo, E., Prado-Gascó, V., Fiszman, S., & Varela, P. (2013). Why buying functional foods? Understanding spending behaviour through structural equation modelling. *Food Research International, 50* (1), 361-368.

Costa-Font, M., & Gil, J. M. (2009). Structural equation modelling of consumer acceptance of genetically modified (GM) food in the Mediterranean Europe: A cross country study. *Food Quality and Preference, 20* (6), 399-409.

Endrizzi, I., Gasperi, F., Rødbotten, M., & Næs, T. (2014). Interpretation, validation and segmentation of preference mapping models. *Food Quality and Preference, 32*, 198-209.

Endrizzi, I., Menichelli, E., Johansen, S. B., Olsen, N. V., & Næs, T. (2011). Handling of individual differences in rating-based conjoint analysis. *Food Quality and Preference, 22* (3), 241-254.

Guillocheau, E., Davidenko, O., Marsset-Baglieri, A., Darcel, N., Gaudichon, C., Tomé, D., & Fromentin, G. (2018). Expected satiation alone does not predict actual intake of desserts. *Appetite, 123*, 183-190.

Henseler, J. (2010). On the convergence of the partial least squares path modeling algorithm. *Computational Statistics, 25* (1), 107-120.

Hwang, H. (2009). Regularized Generalized Structured Component Analysis. *Psychometrika, 74* (3), 517-530.

Hwang, H., & Takane, Y. (2004). Generalized structured component analysis. *Psychometrika, 69* (1), 81-99.

Karalus, M., & Vickers, Z. (2016). Satiation and satiety sensations produced by eating oatmeal vs. oranges. a comparison of different scales. *Appetite, 99*, 168-176.

Langsrud, Ø., & Næs, T. (2003). Optimised score plot by principal components of predictions. *Chemometrics and Intelligent Laboratory Systems, 68* (1–2), 61-74.

Latan, H., & Noonan, R. (2017). *Partial Least Squares Path Modeling: Basic Concepts, Methodological Issues and Applications*: Springer International Publishing.

Måge, I., Mevik, B.-H., & Næs, T. (2008). Regression models with process variables and parallel blocks of raw material measurements. *Journal of Chemometrics, 22* (8), 443-456.

Martens, H., & Næs, T. (1989). *Multivariate calibration*: John Wiley & Sons Canada, Limited.

Martens, M., Tenenhaus, M., Esposito Vinzi, V., & Martens, H. (2007). 21 - The use of partial least squares methods in new food product development. In H. MacFie (Ed.), *Consumer-Led Food Product Development* (pp. 492-523): Woodhead Publishing.

Menichelli, E., Almøy, T., Tomic, O., Olsen, N. V., & Næs, T. (2014). SO-PLS as an exploratory tool for path modelling. *Food Quality and Preference, 36*, 122-134.

Menichelli, E., Hersleth, M., Almøy, T., & Næs, T. (2014). Alternative methods for combining information about products, consumers and consumers' acceptance based on path modelling. *Food Quality and Preference, 31*, 142-155.

Monecke, A., & Leisch, F. (2012). semPLS: Structural Equation Modeling Using Partial Least Squares. *Journal of Statistical Software, 48* (3), 32.

Næs, T., Tomic, O., Mevik, B. H., & Martens, H. (2011). Path modelling by sequential PLS regression. *Journal of Chemometrics, 25* (1), 28-40.

Nguyen, Q. C., Næs, T., Almøy, T., & Varela, P. (2020). Portion size selection as related to product and consumer characteristics studied by PLS path modelling. *Food Quality and Preference, 79*, 103613.

Pagès, J. (2005). Collection and analysis of perceived product inter-distances using multiple factor analysis: Application to the study of 10 white wines from the Loire Valley. *Food Quality and Preference, 16* (7), 642-649.

Pagès, J., & Tenenhaus, M. (2001). Multiple factor analysis combined with PLS path modelling. Application to the analysis of relationships between physicochemical variables, sensory profiles and hedonic judgements. *Chemometrics and Intelligent Laboratory Systems, 58* (2), 261-273.

Risvik, E., McEwan, J. A., Colwill, J. S., Rogers, R., & Lyon, D. H. (1994). Projective mapping: A tool for sensory analysis and consumer research. *Food Quality and Preference, 5* (4), 263-269.

Roininen, K., Lahteenmaki, L., & Tuorila, H. (1999). Quantification of Consumer Attitudes to Health and Hedonic Characteristics of Foods. *Appetite, 33* (1), 71-88.

Romano, R., Tomic, O., Liland, K. H., Smilde, A., & Næs, T. (2019). A comparison of two PLS-based approaches to structural equation modeling. *Journal of Chemometrics, 0* (0), e3105.

Rönkkö, M., & Evermann, J. (2013). A Critical Examination of Common Beliefs About Partial Least Squares Path Modeling. *Organizational Research Methods, 16* (3), 425-448.

Rönkkö, M., McIntosh, C. N., & Antonakis, J. (2015). On the adoption of partial least squares in psychological research: Caveat emptor. *Personality and Individual Differences, 87*, 76-84.

Rönkkö, M., McIntosh, C. N., Antonakis, J., & Edwards, J. R. (2016). Partial least squares path modeling: Time for some serious second thoughts. *Journal of Operations Management, 47-48*, 9-27.

Sanchez, G. (2013). *PLS Path Modeling with R*: Berkeley: Trowchez Editions.

Sanchez, G., Trinchera, L., & Russolillo, G. (2017). plspm: Tools for Partial Least Squares Path Modeling (PLS-PM). *R package version 0.4.9*.

Schutz, H. G., & Cardello, A. V. (2001). A labeled affective magnitude (LAM) scale for assessing food liking/disliking *Journal of Sensory Studies, 16* (2), 117-159.

Tenenhaus, M. (2008). Component-based Structural Equation Modelling. *Total Quality Management & Business Excellence, 19* (7-8), 871-886.

Tenenhaus, M., Vinzi, V. E., Chatelin, Y.-M., & Lauro, C. (2005). PLS path modeling. *Computational Statistics & Data Analysis, 48* (1), 159-205.

Vinzi, V. E., Chin, W. W., Henseler, J., & Wang, H. (2010). *Handbook of Partial Least Squares: Concepts, Methods and Applications*: Springer Berlin Heidelberg.

Vinzi, V. E., Trinchera, L., & Amato, S. (2010). PLS Path Modeling: From Foundations to Recent Developments and Open Issues for Model Assessment and Improvement. In V. Esposito Vinzi, W. W. Chin, J. Henseler & H. Wang (Eds.), *Handbook of Partial Least Squares: Concepts, Methods and Applications* (pp. 47-82). Berlin, Heidelberg: Springer Berlin Heidelberg.

Wold, H. (1980). Model Construction and Evaluation When Theoretical Knowledge Is Scarce: Theory and Application of Partial Least Squares. In J. B. Ramsey (Ed.), *Evaluation of Econometric Models* (pp. 47-74): Academic Press.

Wold, H. (1982). Soft modeling: the basic design and some extensions. In K. Jöreskog & H. Wold (Eds.), *Systems under Indirect Observation* (Vol. 2, pp. 1-54). Amsterdam: North-Holland.

Table 1. Formulation of the yoghurt samples (3*2 design).

Sample	Viscosity	Particle size	Flavour intensity
P1 (TnFkL)	Thin	Flakes	Low
P2 (TkFkL)	Thick	Flakes	Low
P3 (TnFrL)	Thin	Flour	Low
P4 (TkFrL)	Thick	Flour	Low
P5 (TnFkH)	Thin	Flakes	Optimal
P6 (TkFkH)	Thick	Flakes	Optimal
P7 (TnFrH)	Thin	Flour	Optimal
P8 (TkFrH)	Thick	Flour	Optimal

Table 2. Formulation of the biscuit samples (4*2 design).

Sample	Particle size	Baking powder
P1 (s1w)	Flour (0.05mm)	With
P2 (s1wo)	Flour (0.05mm)	Without
P3 (s2wo)	Flour (2.00mm)	Without
P4 (s2w)	Flour (2.00mm)	With
P5 (s3wo)	Flakes (small size)	Without
P6 (s3w)	Flakes (small size)	With
P7 (s4wo)	Flakes (big size)	Without
P8 (s4w)	Flakes (big size)	With

Table 3. PLS-PM direct, indirect and total effects (Yoghurt data).

Relations	Direct	Indirect	Total
LikingV → SatiationV	0.30 (0.001)	0.00 (1.000)	0.30 (0.001)
LikingV → SatietyP	0.00 (0.996)	0.11 (0.024)	0.11 (0.356)
LikingV → SatietyV	-0.12 (0.272)	0.15 (0.010)	0.03 (0.798)
LikingV → PortionV	0.44 (0.000)	0.03 (0.425)	0.47 (0.000)
LikingP → SatiationP	0.37 (0.001)	0.00 (1.000)	0.37 (0.001)
LikingP → SatietyP	0.13 (0.273)	0.15 (0.033)	0.28 (0.013)
LikingP → SatietyV	-0.29 (0.001)	0.01 (0.939)	-0.28 (0.003)
LikingP → PortionP	0.72 (0.000)	-0.03 (0.556)	0.69 (0.000)
SatiationV → SatietyP	0.18 (0.024)	0.00 (1.000)	0.18 (0.024)
SatiationV → SatietyV	0.41 (0.000)	0.00 (1.000)	0.41 (0.000)
SatiationP → SatietyP	0.48 (0.000)	0.00 (1.000)	0.48 (0.000)

The relations are eliminated when all direct, indirect and total effects are not significant (P-value ≥ 0.05).

V, P denote viscosity, particle-size component.

P-values (obtained by the bootstrap) of effects are given in the parentheses.

Table 4. The SO-PLS-PM cumulative validated (cross-validation) explained variances (Yoghurt data).

	Model 1	**Model 2**	**Model 3**
Liking	10.5 (5)	0.9 (1)	20.6 (5)
Satiation		15.1 (5)	0 (0)
Satiety			0 (0)

Each column is a model and each row is an input.

The number of components per each block are given in the parentheses.

Table 5. The SO-PLS-PM direct, indirect and total effects (Yoghurt data).

Relations	Direct	Indirect	Total
Liking → Satiation	10.45 (2.68)	0 (0.88)	10.45 (2.39)
Liking → Satiety	0 (1.01)	0.86 (1.82)	0.86 (1.57)
Liking → Portion	20.64 (2.60)	0 (1.60)	20.64 (2.37)
Satiation → Satiety	19.23 (3.55)	0 (0.46)	19.23 (3.52)
Satiation → Portion	0 (1.27)	0 (1.27)	0 (0)
Satiety → Portion	0.03 (1.02)	0 (0.62)	0.03 (1.05)

Standard errors (obtained by the bootstrap) of the effects are given in the parentheses.

Table 6. The SO-PLS-PM direct, indirect and total effects (preprocessed Yoghurt data).

Relations	Direct	Indirect	Total
Liking → Satiation	8.93 (3.84)	0 (0)	8.93 (3.84)
Liking → Satiety	1.83 (1.71)	2.56 (2.83)	4.39 (3.86)
Liking → Portion	31.8 (5.03)	0 (0)	31.8 (5.03)
Satiation → Satiety	20.18 (4.61)	0 (0)	20.18 (4.61)
Satiation → Portion	1.01 (3.08)	1.42 (2.04)	2.44 (3.35)
Satiety → Portion	0 (1.91)	0 (1.91)	0 (0)

Standard errors (obtained by the bootstrap) of the effects are given in the parentheses.

Table 7. The PLS-PM direct, indirect and total effects (Biscuit data).

Relations	Direct	Indirect	Total
Liking1 → Satiation1	0.30 (0.032)	0.00 (1.000)	0.30 (0.032)
Liking1 → Satiation2	0.20 (0.060)	0.00 (1.000)	0.20 (0.060)
Liking1 → Satiety1	-0.03 (0.815)	0.19 (0.012)	0.16 (0.341)
Liking1 → Satiety2	0.18 (0.136)	0.11 (0.031)	0.29 (0.022)
Liking1 → Portion1	0.48 (0.000)	0.06 (0.426)	0.54 (0.000)
Liking2 → Satiety1	-0.19 (0.066)	0.01 (0.882)	-0.18 (0.054)
Satiation1 → Satiety1	0.53 (0.000)	0.00 (1.000)	0.53 (0.000)
Satiation1 → Portion1	-0.02 (0.837)	0.20 (0.051)	0.17 (0.174)
Satiation1 → Portion2	-0.27 (0.037)	-0.09 (0.334)	-0.36 (0.000)
Satiety1 → Portion1	0.35 (0.022)	0.00 (1.000)	0.35 (0.022)

The relations are eliminated when all direct, indirect and total effects are not significant (P-value ≥ 0.05). For ease of interpretation, some relations with P-values close to significance are kept.

1, 2 denote the first and second component.

P-values (obtained by the bootstrap) of effects were stored in the parentheses.

Table 8. The SO-PLS-PM cumulative validated explained variances (Biscuit data).

	Model 1	**Model 2**	**Model 3**
Liking	0 (0)	0 (0)	7.1 (2)
Satiation		9.5 (5)	0 (0)
Satiety			0 (0)

Each column is a model and each row is an input.

The number of components per each block are given in the parentheses.

Table 9. The SO-PLS-PM direct, indirect and total effects (Biscuit data).

Relations	Direct	Indirect	Total
Liking → Satiation	0 (2.53)	0 (2.53)	0 (0)
Liking → Satiety	0 (2.22)	0 (2.22)	0 (0)
Liking → Portion	7.14 (3.08)	0 (2.95)	7.14 (1.09)
Satiation → Satiety	15.04 (2.98)	0 (0.84)	15.04 (2.87)
Satiation → Portion	0 (2.13)	0.63 (2.34)	0.63 (1.38)
Satiety → Portion	1.27 (1.27)	0 (1.80)	1.27 (1.53)

Standard errors (obtained by the bootstrap) of the effects are given in the parentheses.

Figure Captions

Fig. 1. Illustration of data organization in Menichelli, Hersleth, et al. (2014).

Fig. 2. Path diagram for the satiety studies.

Fig. 3. An example of PLS path model with LV1 as formative mode, LV2 and LV3 as reflective modes.

Fig. 4. The splitting step in PLS path modelling.

Fig. 5. The averaged ratings between products – Yoghurt data.

The p-values showed the significant differences between products for Liking (<0.001), Satiation (<0.001), Satiety (<0.001) and Portion (<0.001).

Product names are given in Table 1.

Fig. 6. PLS-PM path diagram – Yoghurt data.

The 'blue' lines represent the positive relations, the 'red' lines the negative relations and the thickness of the lines the strengths of the relations. The numerical values represent the path coefficients and their p-values obtained by the bootstrap.

The dashed lines represent no relations between blocks.

V and P represent the viscosity and particle-size dimensions, respectively.

Fig. 7. SO-PLS-PM path diagram – Yoghurt data.

The numerical values together with solid lines represent the significant direct effects.

The dashed lines represent no relations between blocks.

Fig. 8. PCP loading plots of input blocks (a), output blocks (b) of Model 2 – Yoghurt data.

Fig. 9. The averaged ratings between products – Biscuit data.

The p-values showed the significant differences between products for liking (<0.001), satiety (0.012), portion (0.017), but for satiation (0.607).

Product names are given in Table 2.

Fig. 10. PCA on double-centered data for liking (left); portion size (right) – Biscuit data.

Product names are given in Table 2.

Fig. 11. PLS-PM path diagram – Biscuit data.

The ‘blue’ lines represent the positive relations, the ‘red’ lines the negative relations and the thickness of the lines the strengths of the relations. The numerical values represent the path coefficients and their p-values obtained by the bootstrap.

The dashed lines represent no relations between blocks.

1 and 2 here denote the first and second component.

Fig. 12. SO-PLS-PM path diagram – Biscuit data.

The numerical values together with solid lines represent the significant direct effects.

The dashed lines represent no relations between blocks.

Fig. 13a. PCP loading plots for input and output data (Model 2) – Biscuit data.

Fig. 13b. PCP loading plots for input and output data (Model 3) – Biscuit data.

Supplementary Figure Captions

Fig. A1. PCA loading plots for liking (a); satiation (b); satiety (c); portion (d) – Yoghurt data.

Fig. A2. RMSEP plots of Model 1 (a), Model 2 (b), Model 3 (c) – Yoghurt data.

Fig. A3. PCP loading plots of Model 1 (a), Model 2 (b), Model 3 (c) – Yoghurt data.

Fig. B1. PCA loading plots for satiation (a); satiety (b) – Biscuit data.

Fig. B2. RMSEP plots of Model 1 (a), Model 2 (b), Model 3 (c) – Biscuit data.