Contents lists available at ScienceDirect

Food Quality and Preference

journal homepage: www.elsevier.com

Segmentation in projective mapping

Ingunn Berget[a,*], Paula Varela[a], Tormod Næs[a,b]

[a] Nofima AS, Osloveien 1, P.O. Box 210, N-1431 Ås, Norway
[b] University of Copenhagen, Faculty of Life Sciences, Department Food Science, Rolighedsvej 30, 1958 Fredriksberg, Copenhagen, Denmark

ARTICLE INFO

ABSTRACT

Projective mapping (PM) or napping® has attained much attention in recent literature as a method for fast sensory profiling and measurement of consumer perception. However, little work has been done to understand the consumer's individual differences in these experiments. In this work, segmentation criteria based on the *Procrustes* distance are explored. The Procrustes distance can be applied with hierarchical clustering using the *Procclustrees* method, which consists of doing hierarchical clustering on the pairwise Procrustes distance between consumers. An alternative strategy called sequential *clusterwise rotations* (SCR) is proposed. SCR extracts clusters by a sequentially partitioning obtained by combining fuzzy clustering techniques and general Procrustes analysis. The methods were tested on simulated and real data and compared with clustering based on MFA results. The simulations show that the MFA approach was outperformed by the other methods when the underlying classes were of same size and there are noise configurations present in the data. For the real data, all methods provided at last one cluster similar to the consensus but differed with respect to the number of clusters identified as well as the interpretation of the clusters. Differences between the methodologies point out the need for external cluster validation in such experiments.

1. Introduction

Due to costs of maintaining a trained sensory panel for sensory profiling of food products, alternative methods, also known as rapid methods, have attained much focus in recent literature (Varela & Ares, 2014). One of these methods is projective mapping (PM) (Risvik, McEwan, Colwill, Rogers, & Lyon, 1994), also referred to as Napping® (Pages, 2003, 2005). PM is also similar to the spatial arrangement method (SpAM, (Goldstone, 1994)). In PM the assessors are asked to arrange a set of samples on a sheet of paper according to how similar or dissimilar the samples are.

Because PM can be applied both with trained panels and untrained consumers, the method is useful for studying consumer perception. However, consumers may use diverse criteria for their evaluation, particularly for complex products. As consumers are not trained, neither given instructions on what to focus on, individual differences may exist. In such cases the consensus will not always give the complete picture. Another issue is that similarity can be perceived in different ways for the different modalities. Also, for problems outside the food and drink area, some types of samples may appear similar visually, but have different functionality (Goldstone, 1994). It is therefore important to study individual variations in projective mapping experiments.

To obtain a consensus, PM data are typically analysed by multiple factor analysis (MFA, (Escofier & Pagès, 1994)), INDSCAL (Carroll & Chang, 1970) or Generalized Procrustes Analysis (GPA, (Gower, 1975)). For PM data GPA is constrained to two components, whereas both MFA and INDSCAL can provide more components. It has been shown that when focusing on the two first components, these methods provide similar consensus solutions although they are conceptually different (Næs, Berget, Liland, Ares, & Varela, 2017; Tomic, Berget, & Næs, 2015). Nevertheless, in studies with complex products, more than two components are sometimes necessary to provide a fully adequate consensus solution. More than two components may, however, be conceptually more demanding since the original data are given in two dimensions (x and y coordinates from each sheet). An alternative is to use clustering for the PM data to enhance interpretability, as discussed in (Vidal et al., 2016). In a comprehensive study on several data sets, each with 80–100 consumers they identified 2–4 segments with different descriptions of the samples in all cases. Moreover, they observed that more segments were needed when more components were needed in the MFA model. The study by Vidal et al., clearly shows that it may

* Corresponding author.
Email address: ingunn.berget@nofima.no (I. Berget)

https://doi.org/10.1016/j.foodqual.2018.05.007
Received 21 December 2017; Received in revised form 24 April 2018; Accepted 18 May 2018
Available online xxx
0950-3293/

not be sufficient to consider the consensus only. Little attention has, however, been put on how to do the segmentation in projective mapping.

In (Vidal et al., 2016) the clustering was based on the consumers' correlations to the MFA components from a global MFA model using a hierarchical agglomerative clustering (HAC). Hence, the consumers were considered similar if they had high contributions to the same MFA components, this means that similarity between two assessors will depend on the other assessors present in the data.

In this work, we present two alternative methods which do not depend on a global model, both are based on Procrustes rotations (Gower, 1975; Gower & Dijksterhuis, 2007). Procrustes analysis is particularly suited for this purpose since it removes rotation and scaling effects. The first alternative, is based on a hierarchical clustering strategy, whereas the second utilises a sequential approach for identifying clusters. The hierarchical strategy was developed for outlier detection in traditional sensory panels, under the name Proclustrees (Dahl & Næs, 2004). In this method, the Procrustes distance is computed between all pairs of assessors, and then HAC is applied for the segmentation. This method is expected to be useful for projective mapping as well, although it has not been explored yet.

The second approach is based on the sequential partitioning strategy proposed in (Berget, Mevik, Vebo, & Naes, 2005; Dahl & Næs, 2009). The sequential approach first identifies the most obvious segment, removes this as the first cluster and then repeats this procedure until a maximum number of clusters is identified. For consumer studies this is advantageous as the data often have high noise levels and there may be several assessors without any clear structure in their data. For each step in the sequence a GPA-like criterion is used for identification of the good cluster for each sequence. The proposed method is named *Sequentially Clusterwise Rotation (SCR)*.

In this work Proclustrees and *SCR* are presented as new methods for segmentation of projective mapping data. The proposed methodology is compared with the MFA approach previously applied in the literature, for both for real and simulated data. Results will be summarised, and finally advantages and disadvantages of the different approaches will be discussed.

2. Methodology

2.1. Clustering PM data with MFA and hierarchical cluster analysis

Multiple factor analysis (MFA) (Abdi, Williams, & Valentin, 2013; Escofier & Pagès, 1994) can be used for analysis of several blocks or tables of data and is frequently used for data from projective mapping (PM). In MFA, all the coordinates from the consumers are organized into a wide matrix with N rows and $2 \times M$ columns where M consumers test N products.

Let $\mathbf{X}_k$ denote the data table for consumer k. Then each $\mathbf{X}_k$ is centred and scaled by its first singular value obtaining $\mathbf{Z}_k$. The concatenated matrix $\mathbf{Z} = [\mathbf{Z}_1 \ldots \mathbf{Z}_M]$ is then analysed by PCA according to the model.

$$\mathbf{Z} = \mathbf{T}\mathbf{P}^t + \mathbf{E} \quad (1)$$

The $\mathbf{T}$ represents the scores, $\mathbf{P}$ the loadings each with A components, and $\mathbf{E}$ the residuals. Note that A can be larger than two, even though each matrix $\mathbf{Z}_k$ only has two columns.

Frequency tables of words used to describe samples can be added as supplementary tables, to guide interpretation of the consensus (Abdi et al., 2013). The supplementary variables (word frequencies) can be projected on to the principal components.

There are different ways of using results from a global model such as MFA for clustering consumers. Here we will apply the same approach as in (Vidal et al., 2016) where pairwise distances between consumers were computed as the Euclidean distances between so-called consumer coordinates with A components. The consumer coordinates show how much contribution the consumers have to the components. For component *a,* and consumer *k*, these can be computed as

$$y_{ka} = \sum_{j}^{J_k} \mathbf{P}_{ja}^2 \quad (2)$$

i.e. the sum of squared loadings for the variables in table *k.* For general applications of MFA, the number of variables per block (J_K) may vary, but for projective mapping $J_k = 2$ for all k ($k = 1,\ldots, M$). The distance between consumer k and l, for A components are computed as the Euclidean distance between the vectors $\mathbf{y}_k = [y_{k1} \ldots y_{kA}]$ and $\mathbf{y}_l = [y_{l1} \ldots y_{lA}]$.

Clustering can then be performed by hierarchical clustering (HAC) on the distance matrix computed from the consumer coordinates $\mathbf{y}_1$, $\mathbf{y}_2$, …, $\mathbf{y}_M$… Here we apply the ward linkage as in (Vidal et al., 2016), however, other linkage methods can be applied as well. In the following this approach will be referred to as MFA-HAC.

The MFA based approach is simple, since when the consumer coordinates (Eq. (2)) are computed, the cluster analysis can be performed in any statistical software using standard methods. Potential problems are, however, the dependency on the number of components, as well as on the global model because the consensus may be misleading if there are outliers in the data, or when groups of consumers base their mapping on very different characteristics of the products. In addition, hierarchical methods are known to be prone to noise and outliers, and some studies indicate that partitions methods are better than hierarchical methods (Wajrock, Antille, Rytz, Pineau, & Hager, 2008).

2.2. Proclustrees

A simple way of implementing the Procrustes distance for segmentation of PM data is the Proclustrees method proposed in (Dahl & Næs, 2004). This was originally proposed for outlier detection in descriptive analysis (DA), but is potentially useful for projective mapping as well. The method is based on establishing the Procrustes distance between all pairs of individuals, and then running a hierarchical clustering on the distance matrix. Hence the clustering *strategy* is the same as for MFA-HAC, whereas the *criterion* for similarity is different and only based on raw individual data.

The Procrustes distance between two assessors k and l is given by

$$\widetilde{D}_{kl} = \min \| p\widetilde{\mathbf{X}}_k \mathbf{R} - \widetilde{\mathbf{X}}_l \| \quad (3)$$

where $\widetilde{\mathbf{X}}$ denotes matrix $\mathbf{X}$ after centring and scaling so that $tr(\widetilde{\mathbf{X}}_k^t \widetilde{\mathbf{X}}) = 1$, $\|\cdot\|$is the Frobenius norm, p is an isotropic scaling factor and $\mathbf{R}$ a rotation/reflection matrix that minimises the distance between the two configurations. More details are given in (Dahl & Næs, 2004). When $\widetilde{D}$ is computed for all pairs of k and l, HAC is applied as for any other distance matrix. In the present paper we have applied the ward linkage. Again, this is only one of several options. The group structure can be studied by inspection of the dendrogram, and the consensus for each cluster can be computed after they are identified using MFA, INDSCAL or GPA. In the following, the acronym PT will be used for the Proclustrees method. MFA was applied for computing cluster consensus.

A clear advantage of the Proclustrees method is that is does not rely on any additional parameters, hence it is easy to use. Potential drawbacks related to hierarchical methods is the same as for MFA-HAC described above.

2.3. Sequential partitioning by clusterwise rotations (SCR)

2.3.1. The sequential strategy with fuzzy clustering and the noise cluster modification

The sequential approach was introduced in (Berget et al., 2005), and was also discussed in (Dahl & Næs, 2009). The basic idea, is to extract and shave off one cluster at a time, then repeat this procedure for the remaining data to find the next cluster. The division into cluster and "noise" is done until a maximum number of clusters are identified, or until there are too few observations left in the data. Typically, a part of the data will remain as a final "rest" cluster, comprising consumers that do not form a cluster.

The partition of data into "good" and "noise" cluster is achieved by using the fuzzy C means (FCM) algorithm (Bezdek, 1981) with the noise cluster modification (Dave, 1991). The criterion to be optimised for each round of the sequential procedure is

$$J = \sum_{k=1}^{M} (u_k^m d_k^2 + (1 - u_k)^m \delta) \tag{4}$$

here u_k and d_k are membership and distance from consumer k to the "good" cluster. The membership values are between zero and one, and must sum to one for each consumer (object) to be clustered. Typically the distance is defined as the Euclidean distance between $\mathbf{x}_k$ and the cluster centre $\mathbf{v}$, but specialised distances can be applied in order to find clusters of special properties (Berget et al., 2005).

The δ is the distance to the *noise* cluster and is the same for all observations. The parameter m ($m > 1$) controls the fuzziness of the clustering, which is the degree of sharing objects between clusters. Larger m's provides more fuzzy membership values. Here, we will use $m = 2$. The sum in Eq. (4) is taken over M consumers.

The effect of adding the noise cluster, is that all consumers with a poor fit to the good cluster will have their highest membership in the noise cluster. The value of δ is critical when applying noise clustering. With too small values, all objects fall into the noise cluster, whereas with too large values, no objects fall into the noise cluster. There are some recommendations on setting the δ in (Dave, 1991, 1993; Dave & Krishnapuram, 1997) but in practice trial and error is needed for setting this parameter.

The criterion in Eq. (4) is minimized by iterating between computing cluster centres ($\mathbf{v}$) for given membership values, and then update distances and the membership values (Eq. (5)). The centre of the cluster is the weighted mean using memberships as weights.

$$u_{ik} = \left(\sum_{t=1}^{C} \left(\frac{d_{ik}^2}{d_{it}^2} \right)^{\frac{1}{m-1}} \right)^{-1} \tag{5}$$

2.3.2. Incorporation of the Procrustes distance in FCM

To accommodate the sequential strategy to PM data, a modified FCM algorithm using a GPA like criterion was developed. The new methodology is named *sequential clusterwise rotations*, shortened SCR.

Instead of using the ordinary Euclidean distance in the clustering criterion, we suggest to use the Procrustes distance between consumer k and the "good" cluster, i.e.

$$D_k = \min \| p_k \widetilde{\mathbf{X}}_k \mathbf{R}_k - \mathbf{G} \| \tag{6}$$

here $\mathbf{G}$ is the consensus for the "good" cluster and is computed as a weighted average of all configurations, with the membership values as weights, according to Eq. (7). As before p_k is an isotropic scaling factor and $\mathbf{R}_k$ a rotation/reflection matrix which minimises the distance between the $\mathbf{X}_k$ and $\mathbf{G}$ (see Eq. (3))

$$\mathbf{G} = \frac{\sum_{k=1}^{M} u_k^m \widetilde{\mathbf{X}}_k \mathbf{R}_k}{\sum_{k=1}^{M} u_k^m} \tag{7}$$

As for the Proclustrees method above, all configurations are scaled to be of the same size such that $tr(\widetilde{\mathbf{X}}_k^t \widetilde{\mathbf{X}}) = 1$ applies for all k after scaling of $\mathbf{X}$ to $\widetilde{\mathbf{X}}$. This type of scaling is referred to as configuration scaling, and is recommended for GPA (Gower & Dijksterhuis, 2007). In this paper we will apply $p_k = 1$ for all k, so that there is no scaling except for the configuration scaling done prior to the analyses.

The algorithm for minimisation of the criterion in Eq. (4) with the Procrustes distance (Eq. (6)), is summarised by step A to D below

A. Initialisation:
   a. Centre and scale data such that $tr(\widetilde{\mathbf{X}}_k^t \widetilde{\mathbf{X}}) = 1$ applies for all k
   b. Initialise membership values randomly for the M consumers
B. Optimise rotations for given membership values and **G**
   c. Compute G for given membership values using Eq. (7)
   d. For k = 1:M rotate $\widetilde{\mathbf{X}}_k$ to fit with **G**
   e. Iterate 2a and b until **G** converges
C. Optimise membership values for given **G**
   f. Compute the distances for each consumer (Eq. (6))
   g. Compute memberships (Eq. (5))
D. Compute J (Eq. (4)), and iterate steps B-C until the difference in J between two consecutive iterations is smaller than a given threshold, for instance 1e-6

2.3.3. Sequential clusterwise rotations - SCR

The steps A–D are first applied on the complete data set, then consumers with membership values larger than 0.5 to the good cluster are shaved off as the first cluster, and the procedure is then applied for the remaining consumers. A second cluster is shaved off, and the procedure is repeated, until the desired number of clusters is identified, or the number of consumers left is too small for further partitioning.

Since the "best" clusters are identified first, it is natural to relax the criterion for being in the "good" cluster for each segment that is shaved off. We therefore in each step, increased the value of δ according to the rule $\delta_{new} = s\delta_{old}$, where $s > 1$ is a fixed constant. To avoid extracting small, not meaningful clusters, we added the criterion that the size of a cluster needs to exceed a given number n, depending on the number of consumers in the study. If the number of consumers assigned to the good cluster is smaller than n, no cluster is shaved off, and the δ is increased in the same way as above to secure that a larger cluster can be identified in the next loop of the sequence. For both real and simulated data sets below, we applied $n = 20$. In the simulations, clustering was performed for δ ranging from 0.3 to 0.6 with s = 1.05.

For the real cases, the procedure proposed in the Appendix was applied to set δ. In summary, δ was first varied over a wide range, with fixed s to identify the interval in which one or more clusters could be identified. Secondly, more values of δ within this range were tested, also varying the s. The final parameter values were then selected by using the cluster validation criteria described in Section 3.4. For the yoghurt data more than one cluster was obtained for δ in the range 0.4–0.8, and the final clustering was performed with $\delta = 0.6$. For the wine data, more than one cluster were identified in the range 0.6–0.9, and the final clustering was performed with $\delta = 0.75$. For both cases s = 1.01.

2.4. Validation and comparison of results

2.4.1. Intracluster variation, the similarity index (SI)

Good clusters should have low intraclass variability, intracluster variation was therefore used to select parameters for SCR. More specif-

ically the similarity index (SI) proposed in (Tomic et al., 2015) was applied to measure the deviation between the individual configurations ($\mathbf{X}_k$) and the cluster consensus ($\mathbf{G}_c$) according to

$$SI_{kc} = \frac{\boldsymbol{X}_k - \boldsymbol{G}_c}{\boldsymbol{G}_c} \tag{8}$$

where $\|\cdot\|$is the Frobenius norm. SI_{kc} measures the squared deviation between the configuration of consumer k and the consensus of cluster c, scaled by the size of the consensus configurations which makes the SI for different clusters comparable. The average SI for the cluster is taken as the mean of all assessors allocated to cluster c and is therefore a measure of the intracluster variation.

2.4.2. Similarity between clusters, the SMI index

The consensus of the identified clusters should be different, otherwise the partition is not very meaningful. The similarity between consensus configurations can be measured by for instance the RV coefficient and the similarity of matrix index (SMI) recently proposed (Indahl et al., submitted) and applied in (Næs et al., 2017).

The SMI was proposed as an alternative to the RV coefficient for measuring similarity in data structures since it does not depend on the eigenvalues structure. The SMI can be considered a multivariate R-square, and is equal to the squared correlation in the one-dimensional case. In this paper we focus on SMI because it also offers a significance test for when two configurations are *different*. The test is based on bootstrapping (Indahl et al., submitted).

To measure the similarity between two consensuses $\mathbf{G}_1$ and $\mathbf{G}_2$, the SMI is based on singular vectors of the two blocks, represented by $\mathbf{U}_1$ and $\mathbf{U}_2$, with variance equal to one for all columns. The $\mathbf{U}_1$ is then regressed onto $\mathbf{U}_2$. For the same number of columns in $\mathbf{U}_1$ and $\mathbf{U}_2$, the SMI is defined as

$$SMI = (TSS(\mathbf{U}_1) - RSS(\mathbf{U}_1, \mathbf{U}_2))/TSS(\mathbf{U}_1) \tag{9a}$$

where *TSS* is the total sum of squares defined as

$$TSS(\boldsymbol{U}_1) = trace(\boldsymbol{U}_1^t \boldsymbol{U}_1) \tag{9b}$$

and *RSS* is the residual sum of squares after fitting $\mathbf{U}_1$ to $\mathbf{U}_2$. Different regression methods can be applied for this fitting, but here we will only consider least squares regression. For further details on the SMI we refer to (Indahl et al., submitted).

2.4.3. Comparison of methods

The three approaches for clustering PM data are compared by simulated and real data. The three different methods differ with respect to the data input, clustering criterion, clustering strategy and the outcome of the clustering routing as summarized in Table 3. SCR provides a consensus for the clusters as a part of the algorithm, whereas for the other methods the cluster consensus must be computed for each cluster after the partitioning of the data. In this work we have chosen to use MFA for this. The consensus obtained from SCR is based on GPA and comprises only two components, we therefore apply MFA on these clusters as well when comparing methods. The methods are compared by computing SMI (Similarity of Matrix Index) between consensus configurations. In addition, SI (similarity index) is used for optimization of parameters in SCR.

3. Case studies

3.1. Simulations

Simulations will always be a simplification of the real world but can provide useful insight into data analytical techniques. In projective mapping, the number of samples, the difference between the samples, and the variability within and between classes will obviously have great impact on an automated cluster analysis. In this work, simulations were done to investigate how class composition, i.e. number of consumers for each of the underlying classes affects the clustering results. In addition, four different schemes for intra class variation were tested.

More specifically we consider a situation with $M = 100$ consumers to do PM on $N = 9$ samples. The consumers belong to one of three classes A, B and C, and the number of consumers in each class is denoted M_c (c = 1, 2, 3). The two first classes use different criteria for the PM task, and therefore have different consensuses. The third class (class C), represents consumers that do not share consensus with any other group. This latter group comprises consumers which either have problems detecting differences, or have problems in expressing what the differences are, and therefore organise the samples more or less randomly and, hence, represent noise.

The simulations were done by first generating the consensus for class A and B denoted $\mathbf{X}_1$ and $\mathbf{X}_2$ respectively, then consumers in each class were generated by adding normally distributed noise to the true class configurations. The consensus for each class ($\mathbf{X}_1$ and $\mathbf{X}_2$) can be generated in many ways. We have here chosen to use the same variability for all samples and have not imposed any pattern on the samples with respect to similarities/dissimilarities between them. This means that our results do not depend on any such assumptions. To secure some spread of the nine artificial samples, however, each of them were randomly assigned to one of the four quadrants of the simulated sheet (600 × 400). Coordinates for each sample were then generated from the uniform distribution within the quadrant it was assigned to. This set up give minimal assumptions about similarities/differences between samples and the true configuration for each class.

The vector $\mathbf{x}_{ic}$ gives the true x-y coordinates for sample i in class c, whereas $\boldsymbol{y}_{ick}$ denote the position of sample i for consumer k in class c ($i = 1, 2, N$, $c = 1, 2$, $k = 1,\ldots, Mc$) and is generated according to the model

$$\mathbf{y}_{ick} \sim N(\mathbf{x}_{ic}, v_c \mathbf{S}_0) \tag{10}$$

with

$$\mathbf{S}_0 = \begin{bmatrix} 1 & 0 \\ 0 & 2/3 \end{bmatrix} \tag{11}$$

The v_c determines the variability within each class and is set to low (L = 1000) or high (H = 2000) for class A and B according to a 2 × 2 design. The label HH will be applied when both classes have high variability, whereas if both classes have low variability simulations will be labelled LL. HL and LH will denote simulations were classes have different variability such that HL is when A has high variability and B low variability, whereas LH is the other way around (see Table 1). Note that S_0 was constructed in such a way that the variability in the first direction is higher than in the second, as the sheets normally used for PM has a wide format (A3 sheets). The third class C, with noise configurations, was generated from the uniform distribution within the complete range of the virtual sheet, i.e. $\mathbf{y}_{i3k} \sim$ U([−300, 300], [−200, 200]).

Artificial PM data were generated with different numbers of consumers in class A, B and C, but always with 100 consumers in total, and for different variability within class A and B as described above. One hundred simulations were run for each of the combinations in Table 1. Cases with $N_1 = N_2$ will be referred to as balanced data (B), whereas cases with $N_1 > N_2$ are referred to as imbalanced data (I). For balanced data HL and LH is in principle the same, and only HL is used.

We believe that the described simulations mimic a large range of different situations for projective mapping experiments. An example of simulation number 2 with HH in Table 1 is shown in Fig. 1.

Table 1
Overview of simulation studies. N_1, N_2 and N_3 are the number of assessors in class A, B and C. For each combination of N_1, N_2 and N_3 three or four different combinations of variations in class A and B are run. The four different variability schemes that are tested are HH, HL, LH and LL where H and L denote high and low levels of variability, and the first and second letter correspond to class A and B respectively. For simulations where the number of members is the same for class A and B (balanced data) only HL is run as HL and LH give the same set up. The v in Eq. (10) is 1000 and 2000 for L and H respectively. Case 1–4 are balanced data (i.e. $N_1 = N_2$) whereas in 5–11 there are more members of class A than B. For all cases there are 100 consumers in total.

Simulation ID	N1	N2	N3	Balance	Variability
1	35	35	30	B	HH, HL, LL
2	40	40	20	B	HH, HL, LL
3	45	45	10	B	HH, HL, LL
4	50	50	0	B	HH, HL, LL
5	40	30	30	I	HH, HL, LH, LL
6	50	20	30	I	HH, HL, LH, LL
7	60	10	30	I	HH, HL, LH, LL
8	60	20	20	I	HH, HL, LH, LL
9	70	10	20	I	HH, HL, LH, LL
10	70	20	10	I	HH, HL, LH, LL
11	80	10	10	I	HH, HL, LH, LL

3.2. Case 1: projective mapping on yoghurt packages

The data applied here were also used in (Næs et al., 2017; Varela, Antúnez, et al., 2017). One hundred consumers performed a projective mapping task to describe packages of yoghurt and yoghurt-like products with blueberry, in Norway. The mapping was based on the evaluation of the extrinsic product properties only with no tasting. The criteria for selecting consumers were that they were frequent users of yoghurt (once a week or more).

Twelve commercial blueberry yogurt packages, bought in local supermarkets, were used in the study. Samples were selected to get a wide range of products in terms of type of product, brand, nutritional characteristics, package size (single, two-pack and four pack) and nutritional and health claims on the packages. The samples represented the main characteristics of the blueberry yoghurts and yoghurt-like products available in the Norwegian market. The different products are labelled P1-P12 and are described in Table 2.

After an initial instruction and explanation of the projective mapping technique, the participants received the twelve yoghurt packs and performed the projective mapping test with the use of a computerized data collection software (EyeQuestion, Logic8 BV, Netherlands). They were also asked to provide descriptive words for samples or groups of samples (for more details, see (Varela, Antúnez, et al., 2017).

3.3. Case 2: projective mapping on wine with and affective approach

Projective mapping based on choice or preference was performed with 144 consumers in Chile on eight Sauvignon Blanc commercial wines (2014 harvest). The wines are labelled V1, V1-D1, V1-D2, V2, R1, R1-D1, R1-D2 and R2. Here the V indicates a varietal type (regular bottling), R reserve type (reserve bottling), and the suffixes D1 and D2 indicate that the wine was enriched with natural wine aromas in two different concentrations. Enrichment came from an aromatic condensate recovered from alcoholic fermentation of Sauvignon Blanc wines. The projective mapping was based on choice, which means that

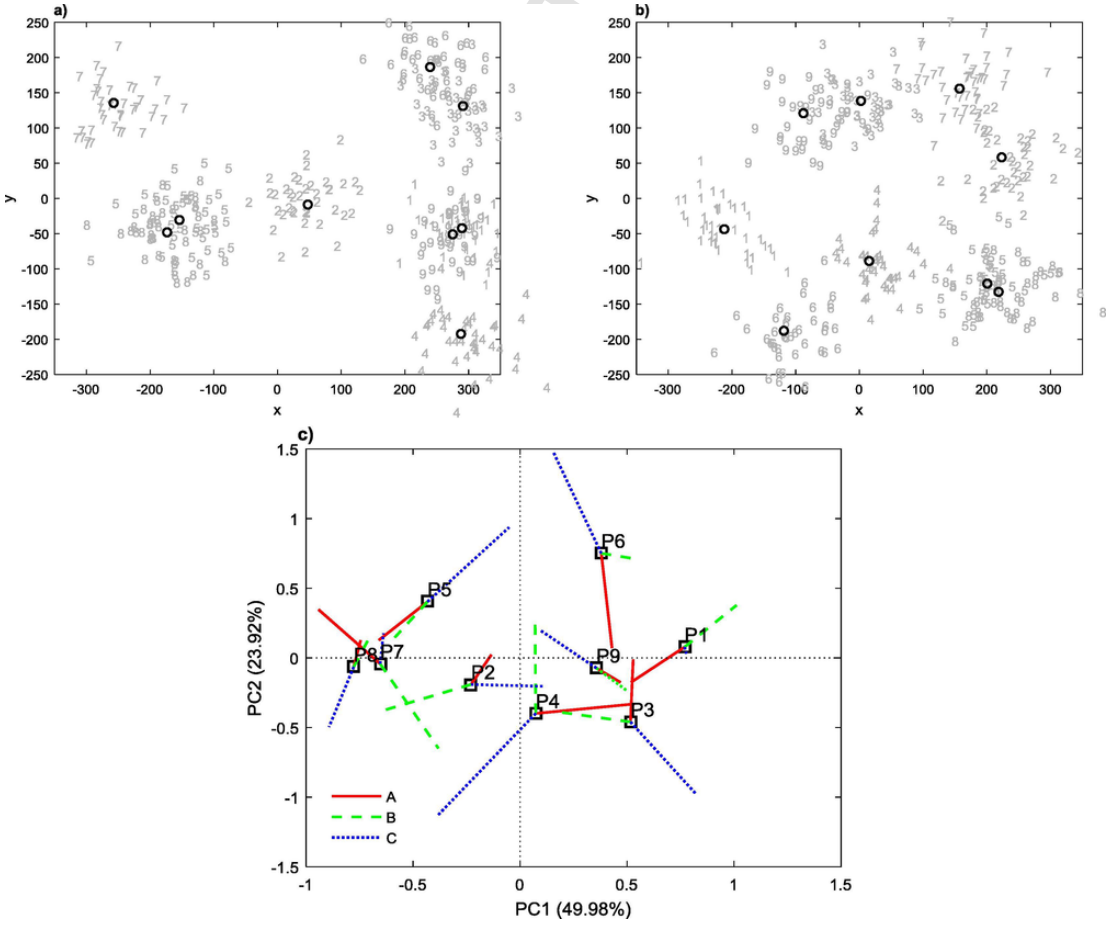

Fig. 1. Example of a simulated data set. a) Data and true configuration for class A, b) data and true configuration class b c) MFA of consensus for A, B and C.

Table 2
Description of the yoghurt packages. Note that P1/P11, P4/P8 and P5/P9 represents similar yoghurt types but differ with respect to topping and/or the packaging (wrapped multi package or single pack).

Label	Yoghurt type	FatInfo	SugarInfo	Other info	Topping	wrapped	Comments on pack
P1	Greek yoghurt	2% fat	Low sugar				source of fibre, 14 g protein, 90Kcal
P2	Yoghurt	wholefat	w/sugar	Lactose free		X	
P3	**cultured milk**	**fat free**	sugar free				16 g protein, original Icelandic cultures
P4	**Yoghurt**	**fat free**	sugar free		X		rich in protein and fibre
P5	Bifidus culture	2.8% fat	w/sugar	Probiotic (actiregularis)	X		
P6	Soy fermented product	2% fat	w/sugar	Lactose free		X	"with yoghurt cultures", "natural lactose free"
P7	Yoghurt	wholefat	w/sugar				"Extra blueberry"
P8	**Yoghurt**	**fat free**	sugar free			X	"fruit yoghurt with fibre"
P9	Bifidus culture	2.8% fat	w/sugar	Probiotic (actiregularis)		X	
P10	**Greek yoghurt**	**fat free**	w/sugar				"a layer of blueberry pieces"
P11	Greek yoghurt	2% fat	Low sugar		X		14 g protein, 90Kcal, source of fibre"
P12	Curd	wholefat	w/sugar				

Table 3
Overview of differences between the discussed methods for clustering PM data.

Method	Input to clustering algorithm	Clustering criterion (distance)	Clustering strategy	Parameters involved	Output
MFA and hierarchical agglomerative clustering MFA-HAC	Coordinates from global MFA model	Euclidean distance	Hierarchical	Linkage (wards method) Number of MFA components Number of clusters	Cluster labels
Proclustrees (PT)	Raw data	Procrustes distance	Hierarchical	Linkage (wards method) Number of clusters	Cluster labels
Sequential clusterwise rotations (SCR)	Raw data	Procrustes distance	Sequential partitioning	n (minimum size of cluster)δ (distance to noise cluster)s (increment of δ) Number of clusters	Cluster labels/Membership values Consensus per cluster

instructions to the consumers differ from the "classic" PM and they are asked to base their categorization on their preferences. Samples that share similar characteristics with respect to choice or preference, should be placed close to each other, while different samples should be placed further apart from each other with regards to each consumers preference. More details of the affective approach are given in (Varela, Berget, et al., 2017). Consumers were also asked to give words to describe samples/groups from the PM. For more details on the experimental set up, we refer to (Lezaeta, Bordeu, Næs, & Varela, 2017).

3.4. Data analyses

For each data set (simulated and real) a global MFA was conducted, clustering was performed with Hierarchical clustering on consumer coordinates from MFA (MFA-HAC), Proclustrees (PT) and sequential clusterwise rotations (SCR). Data analyses and simulations were done in Matlab. For the hierarchical methods, functions within the stats toolbox were applied, whereas for PT and SCR in-house programs were used. Matlab code can be provided by the authors on request. Computation of SMI values and significance testing of SMI were done within R and the package MatrixCorrelation (Indahl et al., submitted).

3.4.1. Computing classification performance

For the simulated data the true underlying classes are known, hence error rates can be computed. Only members from class A and B were used for computing error rates. To compute the error rates, each cluster must be matched to one of the underlying classes (A or B). This was done by computing the SMI (Eq. (9)) between the cluster consensus and the true configurations. If a cluster had the highest SMI to for instance class A, it was assigned as class A. If both clusters had the highest similarity to the same class, the cluster with the largest number of members was assigned to this class, and the simulation was registered as an "overlap". The average error rate was computed for each simulation, and then averaged over all simulations within each set up in Table 1, and for simulations without overlap only. To measure the ability to estimate the underlying consensuses, the SMI values between the cluster consensuses and the underlying classes were computed.

3.4.2. Optimising clustering parameters for simulation studies

Determining the number of clusters is a problematic aspect of cluster analysis and this task usually requires investigation of dendrogram for hierarchical methods, or different validation parameters for partitioning methods (Halkidi, Batistakis, & Vazirgiannis, 2001; Jain, 2010). In simulation studies this is not feasible, and the number of clusters and other necessary parameters must be decided beforehand. Clustering was done with two clusters for each of the methods, however, for SCR, this means two clusters plus the rest cluster.

For MFA-HAC classification was performed with 2–4 components. The minimum size of clusters in SCR was set as $n = 20$, meaning that if the good cluster in one round comprised less than 20 consumers, no cluster was shaved off, but the δ was increased as described in Section 3.3. SCR was performed with δ_0 ranging between 0.3 and 0.6, and the incremental value was $s = 1.01$. SCR was run with 25 random initialisations of membership values for each sequence, and the solution with the smallest value of the criterion function was selected. The results indicated some sensitivity to the initialisation, but stable results were obtained with 25 initialisations. Further research will be needed to inves-

tigate differences in sensitivity in early and later stages of the sequential procedure.

3.4.3. Clustering parameters in real data sets

For the real data sets, the number of clusters was selected by inspecting the dendrogram for the hierarchical methods. A detailed description of the procedure to optimise δ in SCR for each of the cases is described in the Appendix (see also Section 3.3). SCR was run with 50 random initialisations.

4. Results

4.1. Simulations

4.1.1. Overview of the simulations

Fig. 1 shows an example for one of the simulated data sets. This example corresponds to HH of row number 2 in Table 1. In Fig. 1a and b the data and the true underlying consensus for class A and B are shown, whereas in Fig. 1c the differences between the three classes are illustrated by representing the consensus of each class in an MFA plot. Clearly, interpretation of the global consensus may represent a problem.

Across all simulations, the SMI between the true configurations for the two underlying classes varied from 0.02 to 0.48 with a mean of 0.23. This means that the classes were quite different from each other, and probably represent more distinct classes than can be expected for real data.

When doing MFA on the complete datasets, the variance explained by two components ranged from around 50% to 90%, the amount of variation explained by two components was higher when one class (A) was large compared to the other class (B), and when there were few noise configurations in the data. Also, less variation is explained by two components with for instance HH than LL, with LH and HL in between. For the MFA of class A and B separately, more than 90% of the variability could be explained by two components. For comparison, the explained variance by two components for the two real data sets, were 37% (wine) and 45% (yoghurt).

Based on Fig. 1, the explained variation for global MFA and the summary statistics for similarity between the classes, we consider that the simulated data represent the type of problem where cluster analysis will contribute to a better understanding of consumer perception.

The discussed methods were compared with respect to classification performance. Results are first presented for MFA-HAC with one to four components, then for SCR using δ in the range 0.3–0.6 and finally all tree methods are compared. With Proclustrees (PT), no parameters need to be optimised, therefore results for PT will be summarised in the comparison of the three different clustering methods.

For all methods the classification performance differed for balanced (equal size of class A and B) and imbalanced data (class A larger than class B). Performance with respect to classification differed more for the balanced data, therefore results for the balanced data will be in focus.

4.1.2. MFA-HAC for simulated data

For the simulation study, MFA-HAC was applied with one to four components from the global MFA model.

Results with MFA-HAC for balanced data (row 1–4 in Table 1) are summarised in Fig. 2, which shows average error rates for all simulations and for simulations without overlap (3a) as well as the percentage of simulations with overlap (3b) for classification using 2–4 MFA components. The classification ability decreases as the number of noise configurations goes up, both when considering the average error over simulations, and the amount of simulations with overlap. The classification performance did not vary much between LL, HL and HH, although the influence of the number of noise configurations was more evident for HH and HL than LL (results not shown). The classification performance is best with three components.

For the unbalanced data (class A more prevalent than class B) error rates as well as the proportion of overlap were considerably smaller, and the MFA-HAC could be considered as successful in classifying the data correctly with both two and three components, unless for the nearly balanced case with $N_1 = 40$, $N_2 = 30$ and $N_3 = 30$ (results not shown).

Note that Fig. 2 represents results for the average of the simulations, hence it is important to keep in mind that the classification performance varies a lot from perfect to not better than random classification within the same set up in Table 1.

4.1.3. Optimising SCR, setting δ

In the simulation studies, SCR was performed with δ = [0.3, 0.4, 0.5, 0.6]. With balanced data the δ influenced the results considerably, giving more overlap and errors for higher values. For unbalanced data the δ did not have that much influence, except for simulation type 5 (Table 1). For all cases the minimum error rate was obtained with $\delta = 0.3$. For balanced data, only small differences were detected for 0.3 and 0.4, whereas for the unbalanced data also $\delta = 0.5$ worked well. Hence, the methodology does not depend on selecting exactly the best δ. As for MFA-HAC overlapping clusters were observed, but the frequency was much lower, and for $\delta = 0.3$ and 0.4, the frequency was below 0.05 for all simulations.

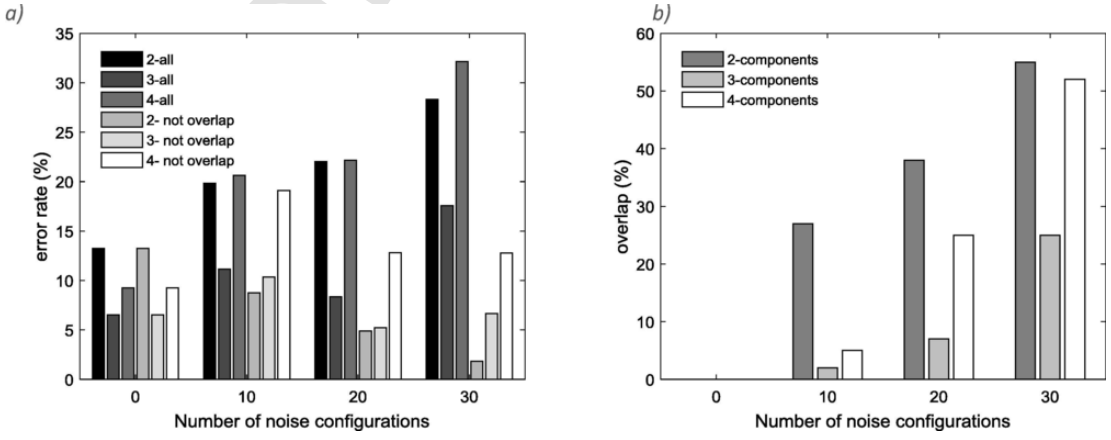

Fig. 2. Classification performance of MFA-HAC method for the simulations with balanced data and variability HH. a) Average error rates with 2–4 components in the MFA for all simulations and for simulations without overlap. The legend indicates the number of components applied for clustering, and if the average is computed for all simulations (-all) or only simulations without overlap (-not overlap). b) The percentage of simulations which have overlapping clusters with 2, 3 and 4 components.

4.1.4. Comparison of methods

Fig. 3 compares average error rates for simulations with balanced data and variability HH. Results from Proclustrees, SCR with $\delta = 0.3$ (SCR-0.3) and MFA-HAC three components (MFA-HAC-3) are shown. Error rates are computed as averages over all simulations, and as averages over simulations without overlap. As can be seen Proclustrees and SCR-0.3 work very well for all cases (error rates < 2%), whereas MFA-HAC-3 have higher error rates (average is above 5% for all four situations). Moreover, the percentage of simulations which have overlap increases with the number of noise configurations present in the data. All methods worked well for unbalanced data (results not shown), unless for the nearly balanced case with much noise ($N_1 = 40$, $N_2 = 30$, $N_3 = 30$), which had results similar to those for balanced cases. Similar patterns were observed for the other variability settings.

Because the overlap between the identified clusters was a much larger problem for MFA-HAC than the other methods, it is of interest to investigate potential causes for this problem. A deeper investigation of the simulated data showed that simulations with overlap were characterised by lower Frobenius norms of both basic configurations, which means that the configurations did not span a large area in the virtual sheet. Consequently, the within class variation (determined by v in Eq. (10)) was large compared to the variability between the samples. We therefore see as a possible explanation that MFA-HAC failed in situations where the underlying classes did not have a very clear consensus in the first place.

In addition to the classification ability, it is of interest to consider how well the true consensus is estimated by the identified clusters. Fig. 4 shows the number of simulations where the true underlying consensus is well estimated with SMI between the cluster consensus and the true configuration exceeding 0.9. In part a) results for the balanced cases are presented. Only counts for class A are shown, since for the balanced cases similar results were observed for both classes. As can be seen the MFA based approach estimates the true consensus less accurate than the other methods, whereas Proclustrees and SCR have similar performance. Moreover, for MFA the number of simulations with "good" values of SMI goes down when the number of noise configuration increases. In part b) results for both classes for the two cases with $N_1 = 60$ are shown. The dominating class (class A) is well estimated by all methods. The smaller class (class B) is, however, well estimated only by SCR.

4.1.5. Conclusions simulation studies

The current simulation studies show that all methods can work well in many cases. Compared to the other methods investigated, MFA-HAC had poor classification ability for simulations where the two underlying classes were of similar size, MFA-HAC was also more sensitive to the number of noise configurations than the other methods. Proclustrees and SCR gave good classification in all cases, although the latter depended on selecting appropriate values of δ. SCR was better than the other methods with respect to the estimation of the consensus for the small class in simulations with unbalanced data.

4.2. Case 1: Yoghurt data

The two first components of the MFA solution for the yoghurt data are displayed in Fig. 5. The twelve samples have been grouped into three groups using HAC with ward linkage. Two of the groups represent two different types of packages (wrapped packages and products with topping/muesli), and the third group comprises the remaining products. The third component (not shown) separates P7 (dessert type) in one direction and Greek-style yoghurts in the other direction. The fourth component is related to the fat content (not shown). The first four components explain 68% of the variation in the data and indicate a high degree of individual variation. Detailed results on samples perception and interpretation can be found in (Varela, Antúnez, et al., 2017).

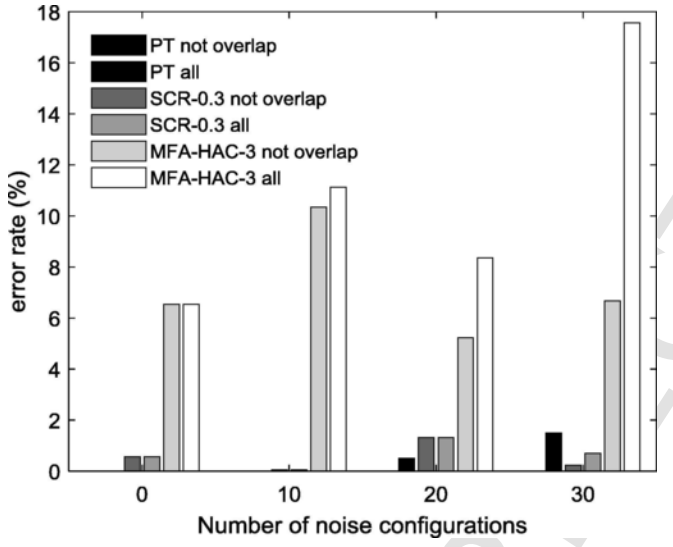

Fig. 3. Error rates for simulations with balanced data and variability HH for Proclustrees, SCR with $\delta = 0.3$ and MFA-HAC with 3 components. Error rates are computed as averages of simulations without overlap (italics in legend), for all simulations (normal letters).

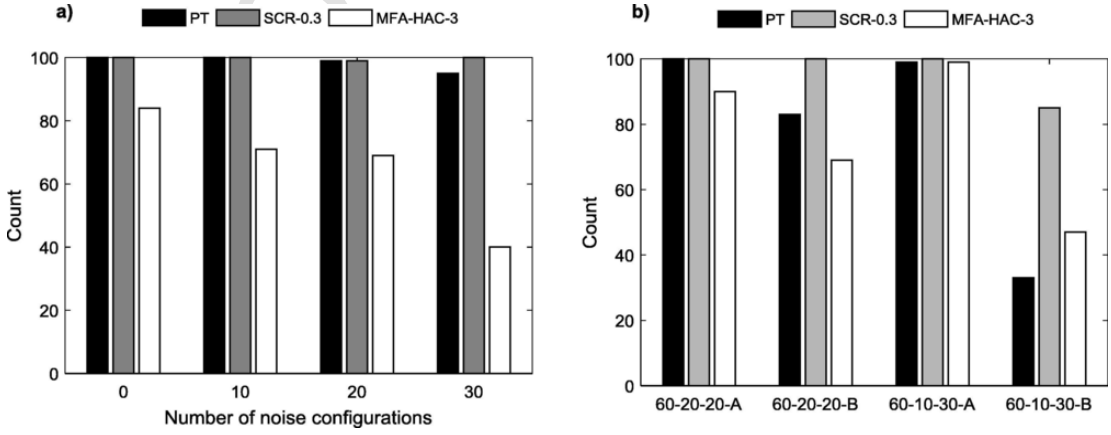

Fig. 4. Number of simulations (out of 100) with SMI between cluster consensus and the true underlying consensus of the class exceeds 0.9. a) Counts for class A for balanced cases with variability HH. b) For imbalanced cases with (60-20-20) and (60-10-30) members of class A, B and C respectively. Counts for both class A and B are shown.

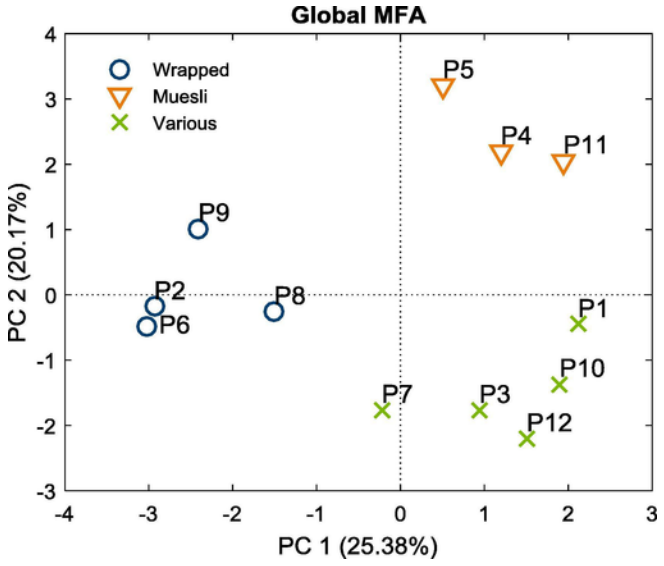

Fig. 5. MFA Consensus Yoghurt. Samples are grouped by HAC and named by common feature of packs in the same group.

4.2.1. Hierarchical methods

With MFA-HAC the solution depends on the number of components applied. When cutting the dendrogram at 70% of the maximum linkage of the tree, MFA-HAC provides three clusters when using 2–3 components and 4 clusters with 5 components. It is therefore natural to focus on three clusters. Fig. 6 shows the dendrograms from MFA-HAC with 2 components (6a) as well as for the Proclustrees method (6b). Hierarchical clustering based on consumer coordinates (MFA-HAC) or Procrustes distances (Proclustrees) provide different results. This is not unexpected and shows the need for studying different criteria for clustering with projective mapping data. Clusters obtained with two components and MFA comprise 28 (C1), 26 (C2) and 46 (C3) members. Clusters obtained with Proclustrees have 47 (C1) and 53 (C2) members.

4.2.2. Optimising SCR

A detailed description of setting δ for the yoghurt data is given in the Appendix. Up to five clusters could be identified for δ in the range 0.4–0.8, but since only the two first had SI considerably lower than the overall average, clustering was limited to two clusters plus the rest cluster. As a compromise between obtaining a dense first cluster, and a second cluster which had a different consensus than the first, the final δ was set to 0.6. The final clustering with $\delta = 0.6$ and s = 1.05, provided two clusters SCR-C1 (n = 42) and SCR-C2 (n = 26) reminding consumers were allocated to the rest cluster labelled SCR-C3 (n = 33). The first cluster (SCR-C1) had high similarity with the global consensus for all δ.

4.2.3. Comparison of methods

SMI and RV coefficients between the consensus for all pairs of clusters, and between clusters and the global solution were computed to compare segments within and across methods. The SMI values are reported in Table 4. Consensuses that are significantly different ($p < 0.05$) are marked with bold in Table 4.

Table 4 shows that MFA-HAC with two to four components, each identified one cluster resembling the global consensus. This was C1, C2 and C2 for two, three and four components respectively. Table 4 also shows that each cluster identified with two components had SMI ≥0.8 and were not significantly different to one of the clusters obtained with three components. When applying four components, however, one cluster is significantly different from all other clusters identified with the other methods. Because it is natural to focus on the major characteristics, results with two components are considered further.

For the two clusters obtained with the Procrustes method, C1 has the highest similarity to the overall consensus, but neither C1 nor C2 were significantly different from the overall consensus (Table 4), although they are significantly different from each other when considering the SMI test.

With SCR and the selected δ (0.6), the first cluster had high similarity to the overall consensus, whereas the consensus of second cluster was significantly different from this. The rest cluster was not significantly different from the first cluster, nor the overall consensus. This reflects that with a higher value of δ, a partitioning with a larger first cluster could have been obtained without changing the consensus considerably, however, the intracluster variation would have been larger.

As can be noted, at least one of the clusters obtained for all methods resembled the overall MFA consensus by being not significantly different from the overall consensus. Hence MFA-HAC-2-C1, PT-C1 and SCR-C1 all provide estimates of the overall consensus. MFA-HAC-2-C2 does not have high resemblance to any of the other clusters, whereas there is some similarity between MFA-HAC-2-C3 and PT-C2 and between SCR-C2 and PT-C2 (Table 4).

As can be expected the clusters with similar consensus were overlapping. The number of consumers common to the main cluster, i.e. the cluster with the highest similarity to the overall consensus, were 21 (MFA-HAC-2 and Proclustrees), 26 (MFA-HAC-2 and SCR) and 32 (Proclustrees and SCR). Altogether 20 consumers were allocated to the main cluster with MFA-HAC-2, Proclustrees and SCR. When considering MFA-HAC with two to four components the number of consumers allocated to the main cluster were 20 (two and three components), 21 (two and four components) and 37 (three and four components). For the remaining clusters, less stability was observed.

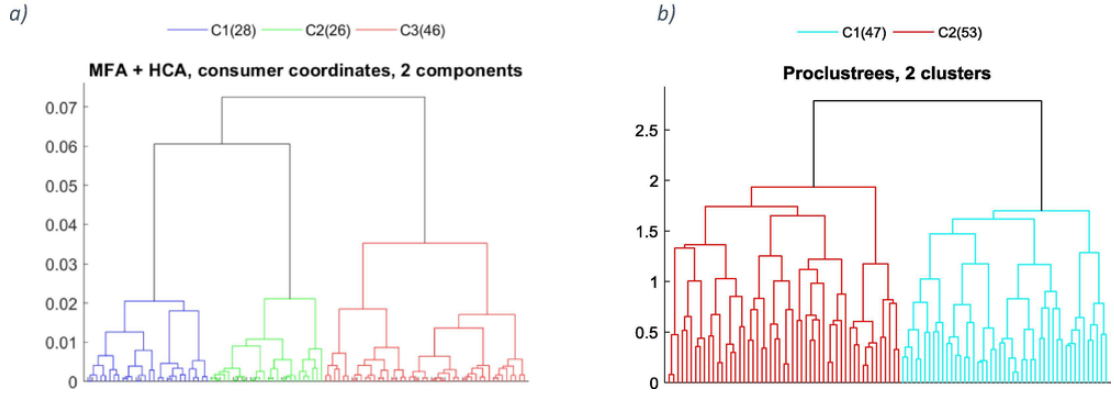

Fig. 6. Dendrogram for Yoghurt, a) MFA-HAC with A = 2 components b) proclustrees.

Table 4
Yoghurt data. SMI between consensus of global MFA model, and MFA model per identified cluster for the different methods. Consensuses that are significantly different are in bold ($p < 0.05$). Grey colour and italic letters indicate SMI to the overall consensus (top row). SMI values between clusters from the same method is also highlighted with grey.

Method		MFA-HAC 2comp			MFA-HAC 3comp			MFA-HAC 4comp			Proclustrees		SCR (δ=0.605,s=1.05)		
	cluster	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C1	C2	rest
global		*0.95*	***0.48***	***0.56***	***0.51***	*0.98*	***0.49***	***0.59***	*0.99*	***0.42***	*0.90*	*0.76*	*0.96*	***0.49***	*0.91*
MFA-HAC 2comp	C1	1.00	**0.42**	**0.48**	**0.43**	0.97	**0.49**	0.63	0.96	**0.32**	0.97	**0.61**	0.99	**0.30**	0.84
	C2		1.00	**0.07**	0.83	**0.44**	**0.13**	**0.16**	**0.48**	**0.42**	**0.37**	**0.35**	**0.43**	**0.40**	**0.47**
	C3			1.00	**0.14**	**0.54**	0.80	0.67	**0.53**	**0.54**	**0.47**	0.75	**0.50**	**0.55**	**0.48**
MFA-HAC 3comp	C1				1.00	**0.47**	**0.05**	**0.10**	**0.53**	**0.49**	**0.35**	**0.46**	**0.44**	**0.43**	**0.43**
	C2					1.00	**0.49**	**0.60**	0.98	**0.38**	0.94	0.71	0.98	**0.42**	0.88
	C3						1.00	0.92	**0.47**	**0.27**	**0.52**	**0.53**	**0.48**	**0.45**	**0.54**
MFA-HAC 4comp	C1							1.00	**0.57**	**0.27**	0.67	**0.49**	**0.60**	**0.38**	0.68
	C2								1.00	**0.36**	0.90	0.74	0.96	**0.45**	0.88
	C3									1.00	**0.30**	**0.60**	**0.34**	**0.43**	**0.43**
Proclustrees	C1										1.00	**0.53**	0.97	**0.26**	0.82
	C2											1.00	**0.61**	0.76	0.68
SCR (δ=0.605,s=1.05)	C1												1.00	**0.32**	0.83
	C2													1.00	**0.44**
	rest														1.00

4.2.4. Interpretation of clusters

When considering two dimensions of the different clusters identified, all methods provided one segment which agrees with the global consensus and has grouped the products into "wrapped", "with muesli/topping" and "the rest". With segmentation, however, the sample groups were more clearly separated than when looking at the complete data set. Hence segmentation applying any of the discussed methods, can help interpretation for the global solution. Clusters resembling the global consensus (MFA-HAC-C1, SCR-C1, PT-C1) are not discussed further.

The remaining clusters from MFA-HAC with 2 components, Proclustrees and SCR that differ from the overall consensus, are shown in Fig. 7. The same type of colour and symbols are used as for the global solution to highlight differences between segments.

For MFA-HAC-C2 (Fig. 7a), the first component spans products typically used for dessert/snack (P3, P7 and P12) on one side, with products with muesli/topping on the other side. Hence this component may be related to usage of the product. The second component spanned allergy friendly products, but also P1 and P11 had high scores on this component. For MFA-HAC-C3 (Fig. 7b) the first component separated wrapped and single packages (as for the global solution) and the second component spanned fat-reduced and non-fat reduced yogurts with P7 (dessert type) as the most extreme of the latter.

In the consensus plot of PT-2 (Fig. 7c), the upper left corner contains lactose-free products (P2, P6), the upper right corner the dessert types (P3, P7, P12), the low right corner Greek-style yoghurts (P1, P10, P11) and the lower left corner contains the yoghurts with probiotics (P5, P9), with fat reduced products in between these groups. These groups represent differences in functionality, but are also confounded with brands, as the pairs (P5, P9), (P4, P8) and (P1, P11) are from three different brands and the products within each of these pairs only differ with respect to packaging (with/without muesli and single/multipack).

Consumers in SCR-C2 (Fig. 7d) seem to have grouped products more according to product properties than the packages. Products on the right side are either fat reduced (P3, P4, P8, P10) or have reduced sugar content (P1, P11) whereas products on the left side all have sugar and no fat reduction. Hence the first component may be related to perceived healthiness. On the left side of the plot, the two products with probiotics (P5, P9) are grouped together in the upper part, whereas the dessert types (P7, P12) are in the lower part. In between these two groups we find the allergy friendly products (P2, P6). Moreover, pairs of product with similar yoghurt but different packaging (multi-, single or with topping, see Table 2, P1/P11 and P4/P8) are coming closer together. Consensus for the rest cluster (SCR-C3, not shown) was not significantly different from the global MFA consensus, although less discrimination among samples was observed.

4.3. Case 2: wine data

4.3.1. Global MFA of the wine data

Wine is a complex product from a sensory point of view, and this is reflected by a large amount of individual variation in the PM data. The two first components of the MFA solution for the wine data are presented in Fig. 8. Aroma intensity and complexity seem to be related to the fist factor with enriched samples towards the right of the map, while the second factor separated varietal (bottom) and reserve wines (top). HAC with ward linkage has been applied to group the samples; One group at the right side of the plot, comprises V1 with enrichment (both doses) as well as R1-D2, the group on the left side comprises V2, R1, R2 and R1-D1. V1 is different from both groups and is in the middle and bottom part of the plot. R1 was less affected by the enrichment than V1, as only R1-D2 was perceived as different from R1. This is somehow logical as reserve wines are already complex in aromatic profile, so a low enrichment might not have such a noticeably effect for a consumer (being an untrained assessor), while the low enrichment level would have a more significant effect in a varietal sample (less aromatically complex). When looking into the third dimension (not shown), R1 with and without enrichment are separated, whereas the fourth component is difficult to interpret. The first four components explained 63.5% of the data.

4.3.2. Clustering of the wine data

Based on the interpretation of the global MFA, both two and three components can be applied for clustering using MFA-HAC. When using

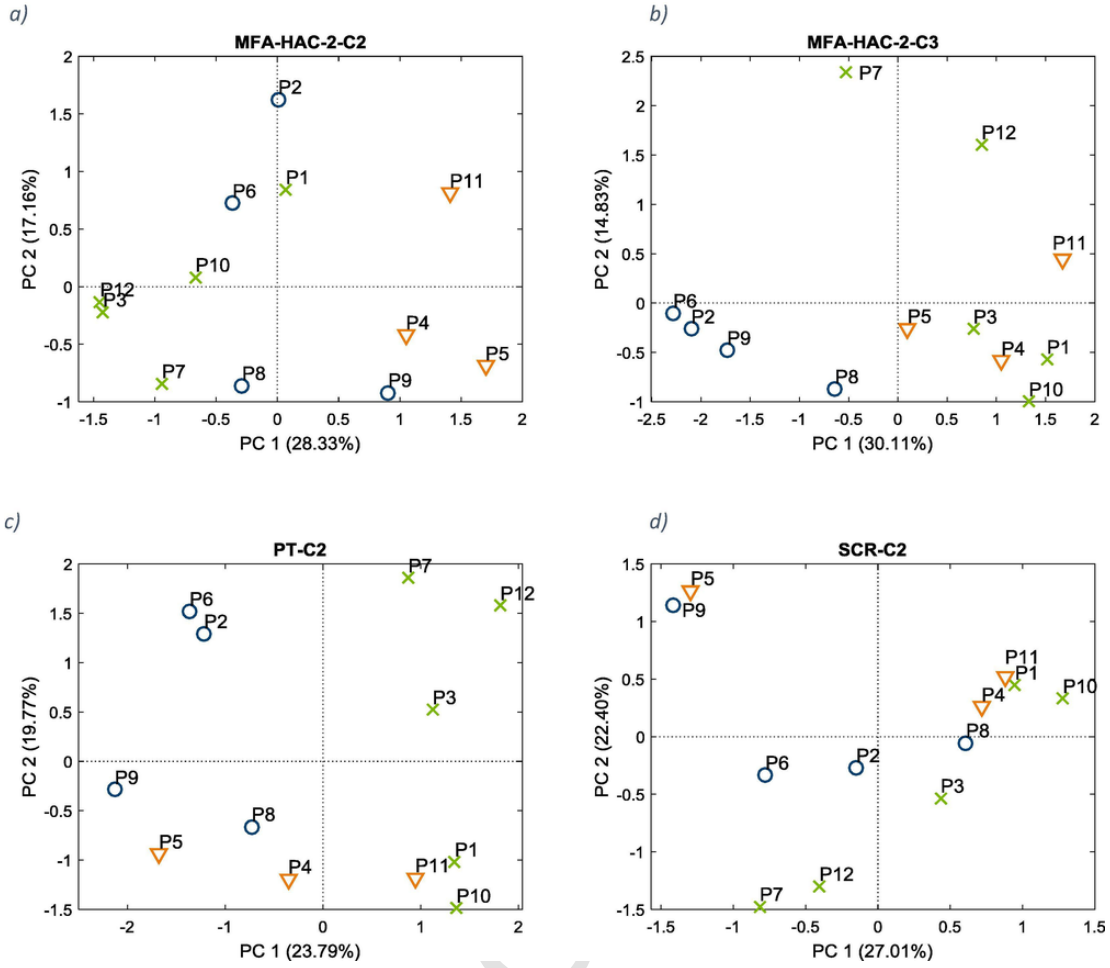

Fig. 7. Consensus configurations for identified clusters a-b) MFA-HAC with two MFA components cluster 2 and 3, c) proclustrees cluster 2, (d) SCR cluster 2 with s = 1.01 and $\delta = 0.6$.

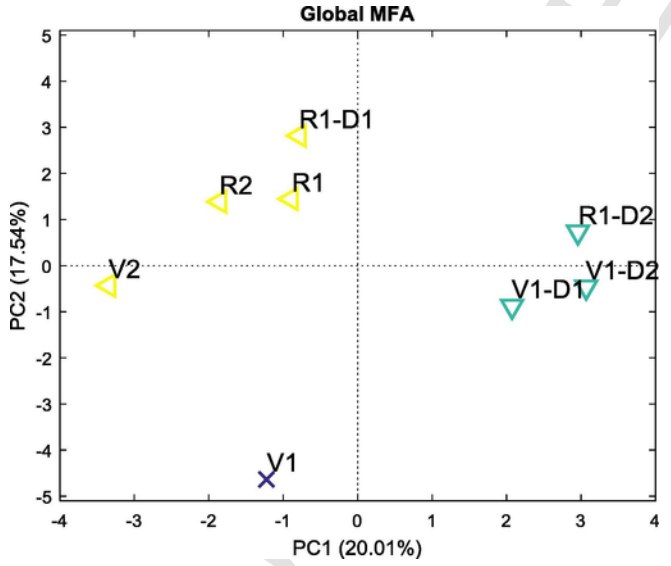

Fig. 8. The two first MFA components from MFA on the wine data. Samples are grouped by HAC, and groups are represented by different symbols and colours.

two components, the dendrogram indicates three clusters (Fig. 9a) whereas with three components, four clusters are indicated (not shown). The partitions obtained were different, as the adjusted rand index (ARI, (Hubert & Arabie, 1985)) was 0.4. The dendrogram produced by the Proclustrees method (Fig. 9b) did, however, not indicate clear segments.

With SCR up to six clusters could be identified for δ in the range 0.6–0.9. A similar procedure as for the yoghurt data (see Appendix) was applied for further optimisation of δ, but as for the Proclustrees method, several small segments with different interpretations could be extracted from the data. The first cluster resembled the overall consensus shown in Fig. 8. This cluster increased in size as δ increased, consequently the intracluster variation also increased. A second cluster had relatively stable consensus until $\delta = 0.7$ but was difficult to interpret. Remaining clusters were small, and their consensus changed considerably between different values of δ and were therefore not considered as stable. It is natural that small groups are not stable as adding/removing a few members can have large impact on the consensus.

In summary, both MFA-HAC and SCR extracted one cluster with a consensus similar to the global consensus. The other clusters were less reliable. It was, however, clear that although the global model indicated that enriched varietal wines may be perceived similar to reserve wines, there are clearly groups of consumers which show the ability to separate such wines. It may therefore be risky to make conclusions based on analysing the complete dataset only. One of the reasons consumers perceived the enrichment very differently, was probably a large variation in the perception of the original wines. All combinations of separation between the two groups of wines (varietal and reserve), and within groups (V1 versus V2 and R1 versus R2) could be

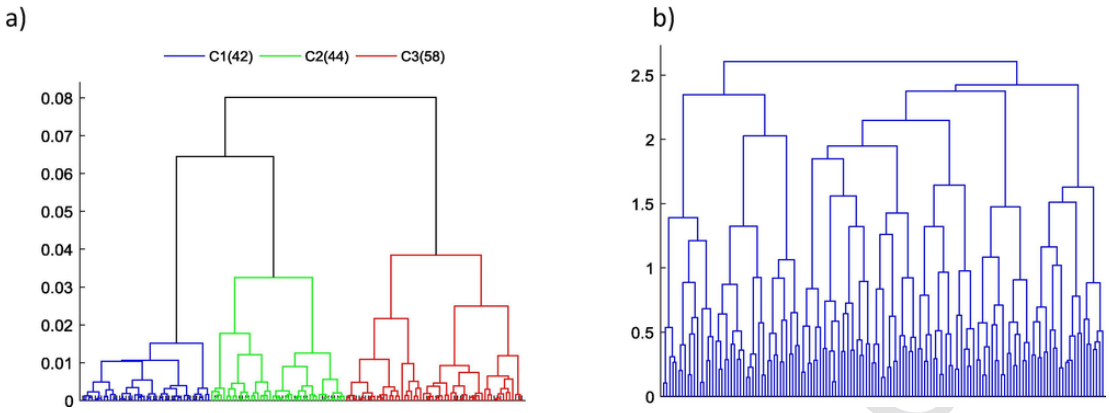

Fig. 9. Dendrograms for the wine data a) MFA-HAC with two components (left) and b) proclustrees.

observed, as well as consumers which seem to group these wines randomly.

5. Discussion

Different methods for clustering projective mapping data have been discussed. Two methods based on the Procrustes distance were compared to a more traditional approach based on MFA. Clustering based on Procrustes can be done hierarchically, using Proclustrees (Dahl & Næs, 2004) or sequentially using a new methodology named *sequentially clusterwise rotations* (SCR). Methods were compared on simulated and real data. Below results are summarised and discussed.

5.1. Simulations

The simulation study showed that performance of methods depended much on the class composition and the number of noise configurations in the data. The level of the intraclass variations did not influence results that much for this study, but for other levels of the intraclass variation (v) this could be different. MFA-HAC did not perform well when the underlying classes where of the same size and with noise configurations present. The two other methods were able to obtain good classification in most cases but had problems when noise configurations comprised 30% of the data. Classification performance was similar for Proclustrees and SCR, but the latter method was slightly better at estimating the underlying consensus of the classes, particularly for the small class in the unbalanced cases.

It is reasonable to expect noise configurations in real data, but it is difficult to know the true extent of this. Simulation results indicate that when such noise is present, MFA-HAC will not provide a good solution if the classes with structure are of similar sizes.

5.2. Real data

For both cases, all methods provided one cluster resembling the consensus. This agrees with previous research (Vidal et al., 2016). The remaining clusters differed, it is therefore important to do an external validation of the segments.

The wine data differed from the yoghurt data in that the projective mapping was based on an affective approach, with the tasting of a sensory complex product. This was reflected in the data, which contained high individual variation, and whereas MFA-HAC indicated a clustering structure, the other methods did not.

MFA-HAC and Proclustrees utilise the same clustering strategy but use different criteria for clustering. For both case studies MFA-HAC was successful in identifying clusters, whereas Proclustrees did not result in interesting clusters for the wine data. Because the ward linkage tends to blow up the distances (Kaufman & Rousseeuw, 1990), MFA-HAC was also tested with other linkage methods, but the resulting dendrograms still indicated clear groups.

A possible explanation of why MFA indicates a clearer grouping structure than the other methods for the real data, is that MFA focus on directions with high variability, whereas Proclustrees sees only pairs of consumers at a time. This is supported by the fact that more groups are indicated with MFA-HAC when more components are used.

5.3. Advantages and disadvantages of the different methods

MFA-HAC has already been applied for several data sets (Vidal et al., 2016). A clear advantage is that it is easy to apply. However, as shown for the case studies, the results may depend on the chosen number of components. A possible modification is to apply a weighted distance, in such a way that the first components contribute more to the distance than the later components. This will make the procedure less dependent on the number of components selected, but also add complexity. In contrast to the other methods discussed, MFA-HAC does not cluster on raw data directly. With this approach the consumers are assumed to be similar if they have high contribution to the same MFA components. This means that the distance between a pair of consumers will change if new consumers are added to the data (or some consumers are removed). Moreover, MFA-HAC was outperformed by the other methods with respect to classification of simulated data with balanced classes and noise configurations.

With Proclustrees, the Procrustes distances must be computed prior to the clustering. This is not available in standard software but can be easily implemented in for instance R and Matlab. A clear advantage is that there are no user defined parameters except from the linkage method. In this work we focused on the ward linkage, but unpublished results indicate that centroid, complete, average and median linkage also work. Proclustrees worked well for the simulated data, however, the method failed to produce interesting clusters for the case studies.

The proposed method SCR is the less user-friendly method as it depends on several parameters that need to be optimised for each dataset. However, a strategy for selecting the parameters is suggested (see the Appendix). As SCR is based on a partitioning strategy, the method is sensitive to initialisations, and therefore needs repeated runs. We found that 25–50 initialisations were enough to secure stable solutions. Because an iterative algorithm is conducted at each step, the method is slower than the two other methods discussed. The speed is not an issue for PM data sets of normal size but was problematic for the simulation studies.

Although the δ needs to be optimised for SCR, results from both case studies indicate that clusters may be stable across a range of δ's. This indicates that the method may not be that sensitive to the para-

meters. The range of δ which provided solutions were 0.4–0.85 for the yoghurt data, and 0.6–0.9 for the wine data. A different choice for *n*, will give other results. Note that selecting the parameters also give insight into the number of clusters that can be identified in the data.

Both MFA-HAC and Proclustrees utilise a hierarchical strategy, which cannot handle noise configurations in a good manner. In our opinion, a situation without any noise configurations cannot be expected for projective mapping studies, but the amount of noise configurations in real data is probably related to the complexity of the products as well as how different they are. In experiments where there are small differences, it is reasonable to expect more variation in the data. We therefore think the sequential partitioning applied in SCR is better than hierarchical approaches. The sequential approach can, however, be applied using a PCA based approach as well, as discussed in (Dahl & Næs, 2009).

5.4. Implications

Both case studies as well as previous studies (Vidal et al., 2016), have shown that there can be considerable individual variation in projective mapping data. Studying the perception from different groups of consumers can be helpful for obtaining a better understanding of consumer behaviour. When looking at consensus solutions only, important information could be missing from groups of consumers that perceive the products differently (e.g. based on sensitivity differences, differences in exposure determined by familiarity, etc.) or that simply give different weights to different sensory attributes (e.g. some consumers may be more concerned about flavours, while others could give more importance to texture when assessing samples), and these differences could be reflected in their choices or preferences.

6. Conclusions

In this work we have highlighted the use of the Procrustes distance for segmentation of projective mapping data, either using a hierarchical strategy with the Proclustrees method, or by sequential partitioning with SCR. The methods have been compared to hierarchical clustering based on an MFA model on the complete data. The three clustering methods applied differ a lot with respect to the criterion used for segmentation (similarity measure), the clustering strategy (hierarchical or sequential) and how the data is applied (raw data or results from global model).

For simulations with one dominating class, all methods could identify and estimate the consensus of this class. SCR could also estimate the consensus of the minor class with 20% noise. When underlying classes were of similar size, MFA-HAC was outperformed by other methods, and performance deteriorated with increasing number of noise configurations

In both case studies the dendrogram produced with MFA indicated distinct groups, whereas the dendrogram produced with Proclustrees did not. Further research is needed to see if the MFA based approach exaggerate the group tendency, or if Proclustrees is less suitable for finding groups for real data. SCR provided interesting clusters for both case studies, and the process of optimising the parameters give additional insight to the data. The suggested procedure for setting parameters worked well, although more research is needed to make this methodology more user friendly. Differences between the methods investigated, point out that external validation of identified segments is important.

A final observation is that although the partitions differed considerably for both real data sets, all methods provided one cluster whose consensus configuration was similar to the global consensus according to the SMI test. Hence, all methods can be applied to identify the consumers fitting well with the global consensus.

Acknowledgements

We would like to thank the FFL "Research Levy on Agricultural Products" and Research Council of Norway for financial support through the RapidCheck project.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.foodqual.2018.05.007.

References

Abdi, H., Williams, L.J., Valentin, D., 2013. Multiple factor analysis: Principal component analysis for multitable and multiblock data sets. Wiley Interdisciplinary Reviews: Computational Statistics 5 (2), 149–179.

Berget, I., Mevik, B.H., Vebo, H., Naes, T., 2005. A strategy for finding relevant clusters; with an application to microarray data. Journal of Chemometrics 19 (9), 482–491.

Bezdek, J.C., 1981. Pattern recognition with fuzzy objective function algorithms. Plenum Press.

Carroll, J.D., Chang, J.J., 1970. Analysis of individual differences in multidimensional scaling via an N-way generalization of Eckardt-Young decomposition. Psychometrika 35 (3), 283–1000.

Dahl, T., Næs, T., 2004. Outlier and group detection in sensory panels using hierarchical cluster analysis with the Procrustes distance. Food Quality and Preference 15 (3), 195–208.

Dahl, T., Næs, T., 2009. Identifying outlying assessors in sensory profiling using fuzzy clustering and multi-block methodology. Food Quality and Preference 20 (4), 287–294.

Dave, R.N., 1991. Characterization and detection of noise in clustering. Pattern Recognition Letters 12 (11), 657–664.

Dave, R.N., Krishnapuram, R., 1997. Robust clustering methods: A unified view. Ieee Transactions on Fuzzy Systems 5 (2), 270–293.

Dave, R. N. (1993). Robust fuzzy clustering algorithms. In Paper presented at the Second IEEE International Conference on Fuzzy Systems, San Fransisco, California.

Escofier, B., Pagès, J., 1994. Multiple factor analysis (AFMULT package). Computational Statistics & Data Analysis 18 (1), 121–140.

Goldstone, R., 1994. An efficient method for obtaining similarity data. Behavior Research Methods Instruments & Computers 26 (4), 381–386. https://doi.org/10.3758/bf03204653.

Gower, J.C., 1975. Generalized Procrustes Analysis. Psychometrika 40 (1), 33–51.

Gower, J. C., & Dijksterhuis, G. B. (2007). Procrustes Problems.

Halkidi, M., Batistakis, Y., Vazirgiannis, M., 2001. On clustering validation techniques. Journal of Intelligent Information Systems 17 (2–3), 107–145.

Hubert, L., Arabie, P., 1985. COMPARING PARTITIONS. Journal of Classification 2 (2–3), 193–218.

Indahl, U. G., Næs, T., & Liland, K. G. (submitted). A similarity index for comparing coupled matrices. Linear Algebra and its Applications.

Jain, A.K., 2010. Data clustering: 50 years beyond K-means. Pattern Recognition Letters 31 (8), 651–666.

Kaufman, L., Rousseeuw, P., 1990. Finding groups in data: An introduction to cluster analysis. John Wiley & Sons, Inc..

Lezaeta, A., Bordeu, E., Næs, T., Varela, P., 2017. Exploration of consumer perception of Sauvignon Blanc wines with enhanced aroma properties using two different descriptive methods. Food Research International 99, 186–197.

Næs, T., Berget, I., Liland, K.H., Ares, G., Varela, P., 2017. Estimating and interpreting more than two consensus components in projective mapping: INDSCAL vs. multiple factor analysis (MFA). Food Quality and Preference 58, 45–60.

Pages, J., 2003. Direct collection of sensory distances: Application to the evaluation of ten white wines of the Loire Valley. Sciences Des Aliments 23 (5–6), 679–688.

Pages, 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.

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.

Tomic, O., Berget, I., Næs, T., 2015. A comparison of generalised procrustes analysis and multiple factor analysis for projective mapping data. Food Quality and Preference 43, 34–46.

Varela, P., Antúnez, L., Berget, I., Oliveira, D., Christensen, K., Vidal, L., … Ares, G., 2017. Influence of consumers' cognitive style on results from projective mapping. Food Research International 99 (Part 1), 693–701.

Varela, P., Ares, G. (Eds.), 2014. Novel techniques in sensory characterization and consumer profiling. CRC Press.

Varela, P., Berget, I., Hersleth, M., Carlehög, M., Asioli, D., Næs, T., 2017. Projective mapping based on choice or preference: An affective approach to projective mapping. Food Research International 100 (Part 2), 241–251.

Vidal, L., Antúnez, L., Giménez, A., Varela, P., Deliza, R., Ares, G., 2016. Can consumer segmentation in projective mapping contribute to a better understanding of consumer perception?. Food Quality and Preference 47 (Part A), 64–72.

Wajrock, S., Antille, N., Rytz, A., Pineau, N., Hager, C., 2008. Partitioning methods outperform hierarchical methods for clustering consumers in preference mapping. Food Quality and Preference 19 (7), 662–669.