Food Quality and Preference 95 (2022) 104358

Contents lists available at ScienceDirect

Food Quality and Preference

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

Combining hedonic information and CATA description for consumer segmentation

Evelyne Vigneau [a,*], Véronique Cariou [a], Davide Giacalone [c], Ingunn Berget [d], Fabien Llobell [b]

[a] Oniris, StatSC, Nantes, France
[b] Addinsoft, XLSTAT, Paris, France
[c] University of Southern Denmark, Odense, Denmark
[d] Nofima, Ås, Norway

ARTICLE INFO

Keywords:
Liking
CATA
Penalty-lift analysis
Consumer segmentation
Cluster stability
Sensometrics

ABSTRACT

Check-all-that-apply (CATA) has become a popular method for obtaining a consumer-based sensory characterization. In most case studies, consumers are also asked to evaluate the set of products according to a liking scale with the aim to identify the key sensory attributes associated with the most liked, or disliked, products. The common approach consists, first, in the identification of consumer segments based on the preference profiles. Thereafter, the analysis of the CATA responses is performed within each segment. Our purpose herein is to investigate different ways to simultaneously identify clusters of preference profiles while taking into account the CATA attributes. These approaches are derived from strategies already proposed by the different co-authors, namely: Fuzzy Clusterwise Regression (FCR), Clustering around Latent Variables (CLV) approach with external data, CLUSCATA-liking and CLV3W. The first two approaches involve the aggregation of the individual CATA data into a contingency table, while the last two ones deal with the combination of liking and CATA data at the individual level. These four strategies are illustrated on the basis of a real case study. Results are compared with respect to cluster stability together with interpretability of liking profiles within each segment. The stability of the results, assessed by bootstrapping, differed according to the strategy used. Moreover, working at the individual level or with combined data lead to a somewhat different segmentation of the panel of consumers.

1. Introduction

Check-All-That-Apply (CATA) questions are nowadays increasingly used to obtain perceptual product profiles from consumers (Meyners & Castura, 2014). Regularly applied to collect rapid sensory information, CATA questions were also successfully introduced to collect other perceptual measures such as emotional responses (Jaeger et al., 2018) or situational appropriateness (Jaeger, Lee, Jin, Chheang, Rojas-Rivas & Ares, 2019). In a CATA experiment, consumers are simply asked to check all the items of a predefined list of attributes they deem to be appropriate to describe each of the samples. This quick and straightforward task has been shown to provide information about the consumer perception of the sensory characteristics of food products (Ares et al., 2015). Moreover, Jaeger, Chheang, Jin, Roigard, & Ares (2020), among others, showed that despite the simplicity of the task, the average citation frequencies of the sensory CATA attributes reflect to a large extent the average intensity ratings of food products. Therefore, with regard to sensory description of products, the common approach consists in considering the contingency table between products and CATA attributes, that is, the product $\times$ attribute matrix depicting the number of consumers who selected a given CATA attribute to characterize a given product. Different statistical techniques can be further applied to analyze the obtained contingency table. In particular, Correspondence Analysis (CA; Greenacre, 2017) is the factorial method most often advocated to represent, on a low dimensional space, the associations between the rows (*i.e.*, the products herein) and the columns (*i.e.*, the CATA attributes herein) of such a contingency table. This simultaneous representation of both products and CATA attributes, usually onto the first two components, provides a convenient perceptual map summarizing the consumers' sensory description of the products. Besides this factorial exploratory analysis, univariate analyses such as Cochran's Q test are widely used to test product differences for each CATA attribute (Meyners, Castura, & Carr, 2013; Meyners & Castura, 2014).

In addition to the CATA questions ballots, it is usual to ask consumers

* Corresponding author.
E-mail address: evelyne.vigneau@oniris-nantes.fr (E. Vigneau).

https://doi.org/10.1016/j.foodqual.2021.104358
Received 10 March 2021; Received in revised form 2 June 2021; Accepted 9 August 2021
Available online 11 August 2021

to rate the products under study on an overall liking scale (Jaeger & Ares, 2014). In order to relate CATA and liking data, penalty analysis (Ares, Dauber, Fernandez, Gimenez, & Varela, 2014) or penalty-lift analysis (Williams, Carr, & Popper, 2011; Plaehn, 2012) have been proposed. In the former approach, it is required that consumers also check all the appropriate attributes that they would assign to their ideal product, in addition to liking and CATA evaluations of real products. Without this supplementary part in the experimental design, penalty-lift analysis for a given CATA attribute leads to assess the difference of the averaged liking scores depending on whether the attribute was selected or not. Finally, in penalty-lift analysis, the rating value is averaged over all consumers and products (Meyners & Castura, 2014; Meyners et al., 2013). For representation purpose, the difference in liking depending on whether each CATA attribute has been selected or not (also referred to as unweighted CATA penalty), is plotted against the relative proportion of consumers who checked that attribute (Giacalone, 2018). Finally, in a testing hypothesis framework, Monte-Carlo simulation-based procedures have been suggested by several authors, either for penalty-lift analysis (Plaehn, 2012; Meyners, 2016) or for PLS regression models relating CATA responses and a design matrix with regard to external information about products or consumers (Rinnan, Giacalone, & Frøst, 2015).

Up to now, penalty(-lift) analysis appears to be the predominant approach used to highlight relationships between liking and CATA measures on the same set of products. It is worth noting that this analysis lies on an underlying homogeneity assumption considering the consumer panel as a whole. In other words, it assumes that all consumers share the same preference profiles for the same reasons. A potential problem with this approach arises when, for example, subsets of consumers pay attention to the same attributes, but with opposite effects in terms of liking. In such cases, penalty analysis would completely miss this critical information. Furthermore, a CATA attribute rarely selected by the whole panel is likely to be excluded from the analysis of penalties regardless of the impact it might actually have on the liking score of a small subset of consumers.

The analysis of liking data typically encompasses internal preference mapping, with possible consumers segmentation conducted by means of a clustering strategy (MacFie, 2007). Even if a few studies considered segmentation of the panel according to liking measures collected in addition to CATA data, this segmentation was performed independently of which CATA attributes had been selected (*e.g.* Ares & Jaeger, 2015; Spinelli, Monteleone, Ares, & Jaeger, 2019). On the opposite front, a cluster analysis based only on CATA data is also possible with the CLUSCATA method (Llobell, Cariou, Vigneau, Labenne, & Qannari, 2019), but without taking into account possible differences in liking even if the same attributes are chosen.

In this work, we are interested in the segmentation of a panel of consumers according to their differences in liking, while simultaneously considering the description of the products they gave based on a list of CATA attributes. The ultimate goal is identifying the most significant CATA attributes related to the different segments obtained, *i.e.*, within each segment of consumers, the attributes that explain the liking, or disliking associated with the products under study. Herein, several alternatives are investigated to simultaneously identify clusters of preference profiles while taking into account the CATA attributes. To this end, we consider different statistical approaches based on strategies already proposed by the different co-authors of this work, with modifications either in terms of data preparation or of algorithm development.

The rest of the paper is organized as follows. The methodological section (Section 2) is devoted to the presentation of the four considered strategies. Of particular interest is the assessment of the stability of the consumer segments obtained (Section 3). This is an important issue for the choice of an appropriate number of segments. Indeed, the four different approaches are found to generate slightly different points of view which may lead to more or less fine segmentations. The four approaches are illustrated and compared on the basis of a real case study (Section 4).

2. Methods

2.1. Notation and data preparation

In the following, we consider a classical CATA experiment in which consumers are monadically presented a set of products and for each product are first asked to provide their liking score, and then to select all the attributes in the CATA list they deemed appropriate to describe the product.

The total number of products evaluated is denoted by I in the following, each product being identified with the index i ($i = 1, \ldots, I$). The total number of consumers is denoted by J, and j is the index associated with consumer j ($j = 1, \ldots, J$). Let us consider that the total number of CATA attributes is noted Q, each attribute being associated with the index q ($q = 1, \ldots, Q$).

The centred ($I \times J$) matrix of the liking scores is denoted by $\mathbf{Y}$. The value y_{ij} in $\mathbf{Y}$ corresponds to the liking score given by the consumer j to the product i minus the mean of the scores this consumer provided to the I products. This centring task aims at discarding the differences between consumers with respect to their mean level of rating.

Suppose that the description of I products with respect to Q CATA attributes were recorded for J consumers, resulting in an ($I \times J \times Q$) array $\underline{\mathbf{Z}}$. As such, the first mode of $\underline{\mathbf{Z}}$ is associated with products while its second mode is associated with consumers and the third one with the attributes. Thus, the j^{th} lateral slice of $\underline{\mathbf{Z}}$ corresponds to the ($I \times Q$) binary table depicting which CATA attributes were selected for each of the I products by consumer j. In other words, $z_{ijq} = 1$ if consumer j checked attribute q for product i, otherwise $z_{ijq} = 0$.

The ($I \times Q$) contingency table depicting the (absolute) frequencies according to products and CATA attributes is denoted by $\mathbf{F}$. Herein, it is simply obtained by summing the values of $\underline{\mathbf{Z}}$ along its second dimension (*i.e.*, along the consumer mode). It should be noted that the contingency table $\mathbf{F}$ is the data matrix usually considered when analyzing CATA data by correspondence analysis, to describe the similarity and dissimilarity between products and to identify the CATA attributes which are the most often associated with one specific product, or subset of products. Let us also notice that $\mathbf{F}$ refers to information at the whole panel level.

Among the four approaches investigated in this paper, two of them consider CATA data at the individual consumer level. Therefore, a combination of liking and CATA data is required. In practice, the two-way matrix $\mathbf{Y}$ and the three-way array $\underline{\mathbf{Z}}$ are aggregated together to form a new three-way array, denoted $\underline{\mathbf{A}}$, of the same size as $\underline{\mathbf{Z}}$. As $\underline{\mathbf{A}}$ combines CATA and liking data, it differs from $\underline{\mathbf{Z}}$ in the sense that, for each triplet of indices (i,j,q), a_{ijq} is defined as the centred liking score, y_{ij}, that consumer j has given to product i when this consumer j checked attribute q for this product i and zero otherwise. Consequently, if z_{ijq} is equal to zero, then a_{ijq} will be also set to zero. Thus, the three-way array $\underline{\mathbf{A}}$ is made of zeros if an attribute q has not been checked by consumer j for product i. Otherwise, when the attribute q has been considered to be appropriate by consumer j to depict the product i, then the value in $\underline{\mathbf{A}}$ corresponds to the centred liking score of this consumer regarding this product. If the consumer appreciated the product more than his/her mean level of liking, the associated value in $\underline{\mathbf{A}}$ will be positive. Contrariwise, if the consumer liked the product less than his/her own mean level of liking, the associated value in $\underline{\mathbf{A}}$ will be negative. In practice, the j^{th} lateral slice of $\underline{\mathbf{A}}$, say $\mathbf{A_j}$, is defined by:

$$\mathbf{A_j} = Diag(\mathbf{y_j}) * \mathbf{Z_j} \tag{1}$$

with $\mathbf{Z_j}$, the j^{th} lateral slice of $\underline{\mathbf{Z}}$; $\mathbf{y_j}$, the vector of liking scores associated with consumer j; and $Diag()$, the diagonal operator. The structure of different data matrices $\mathbf{Y}$, $\underline{\mathbf{Z}}$, $\mathbf{F}$ and $\underline{\mathbf{A}}$ is illustrated in the first part of Fig. 1.

Both $\underline{\mathbf{A}}$ and $\underline{\mathbf{Z}}$ are often sparse since they are likely to contain a quite large number of zero elements. One can also notice that the averaging of $\underline{\mathbf{A}}$ along the first dimension, *i.e.* over the I products, leads no more to zero values. Indeed, a CATA attribute is rarely selected by a consumer for all the products under study. In the context of our data, the column-wise centring of $\underline{\mathbf{A}}$ along its first dimension, which is a common option, seems to be questionable and is therefore avoided.

2.2. Overview of the investigated approaches

The four approaches evaluated for segmenting consumers with respect to their liking profiles, while taking into account the CATA description of the products are listed in Table 1. The original source from which the method has been tailored for relating liking scores and CATA data is also mentioned. These approaches may be split into two families according to the input data matrices (as defined in section 2.1) involved. Fig. 1 provides an overview of how these data are integrated into each of the four approaches, described in more detail in the following subsections.

2.3. Fuzzy Clusterwise regression

The Fuzzy Clusterwise Regression (FCR) approach was first introduced by Wedel & Steenkamp (1991) and discussed by Berget, Mevik, & Næs (2008), Johansen, Hersleth, & Naes (2010) and Menichelli, Olsen, Meyer, & Næs (2012) in the scope of consumer and sensory studies.

The CATA characterization of the products, synthetized in the contingency table $\mathbf{F}$, is first submitted to a Correspondence Analysis (CA) and the first CA components are retained. For sake of simplicity, we consider herein the two first components, but the procedure could also be applied with only one or more than two components. These components provide the coordinates of the products onto the first CA dimensions and are recorded in a matrix which is denoted by $\mathbf{\Phi}$. $\mathbf{\Phi}$ is used as the dependent matrix in a linear regression model adjusted within each cluster of consumers simultaneously determined using a fuzzy clustering approach.

The optimization process in FCR aims to identify K clusters of consumers, the fuzzy memberships u_{jk}^{m} of each consumer j ($j = 1, \ldots, J$) regarding each cluster k ($k = 1, \ldots, K$) according to the fuzzifier parameter m, as well as the regression coefficients, $\widehat{\mathbf{b}}_{\mathbf{k}}$, within each cluster, so that to minimize:

Table 1
List of the methodological approaches investigated.

Name	Acronym	Source/adapted from	Data matrices involved*
Fuzzy Clusterwise Regression	FCR	Wedel & Steenkamp, 1991	**F**, **Y**
CLV with external data (in row)	CLVr	Vigneau, Endrizzi, & Qannari, 2011	**F**, **Y**
Three-Way Cluster analysis around Latent Variables	CLV3W	Cariou & Wilderjans, 2018	$\underline{\mathbf{A}}$
Clustering of CATA-liking tables	CLUSCATA-liking	Llobell, Cariou, Vigneau, Labenne, & Qannari, 2019	$\underline{\mathbf{A}}$

* **F**: CATA contingency table, **Y**: liking scores matrix, $\underline{\mathbf{A}}$: three-way array combining CATA and liking data.

$$J = \sum_{k=1}^{K} \sum_{j=1}^{J} u_{jk}^{m} \| \mathbf{y}_{\mathbf{j}} - \widehat{\mathbf{y}}_{\mathbf{j}}^{(\mathbf{k})} \|^{2} \quad \text{where} \quad \widehat{\mathbf{y}}_{\mathbf{j}}^{(\mathbf{k})} = \mathbf{\Phi}' \widehat{\mathbf{b}}_{\mathbf{k}} \tag{2}$$

In the core of the algorithm, the vector of the predicted liking scores $\widehat{\mathbf{y}}_{\mathbf{j}}^{(\mathbf{k})}$, within each cluster k ($k = 1, \ldots, K$), is extracted from a weighted regression model, where the weights are the fuzzy memberships of the consumers in cluster k, of the unfolded $\mathbf{Y}$ data on the augmented-$\mathbf{\Phi}$ data matrix (Menichelli et al., 2012). The augmented-$\mathbf{\Phi}$ matrix is obtained by replicating p times, vertically, the matrix $\mathbf{\Phi}$ of the scores of the products on the retained CA components.

The value $m = 2$ is commonly used in various fuzzy clustering applications (Krishnapuram & Keller, 1996; Berget et al., 2008). This value was used by Menichelli et al. (2012), while Johansen et al. (2010) investigated the choice of the fuzzifier and found that the best fit was obtained for m as low as 1.1. Membership values and cluster parameters are updated iteratively.

FCR makes it possible to identify segments of consumers by allocating each consumer to the cluster for which his/her membership has the highest values. In the same time, from the vectors of loadings $\widehat{\mathbf{b}}_{\mathbf{k}}$ (for $k = 1, \ldots, K$), a reconstruction formula to transpose back the CA components space to the CATA attributes space, makes it possible to identify the most important coefficients of regression between liking scores, $\mathbf{Y}$, and CATA description, $\mathbf{F}$. Finally, the predicted liking scores vectors $\widehat{\mathbf{y}}_{\mathbf{j}}^{(\mathbf{k})}$ (for $k = 1, \ldots, K$), represent the expected liking profiles for consumers with highest membership in cluster j.

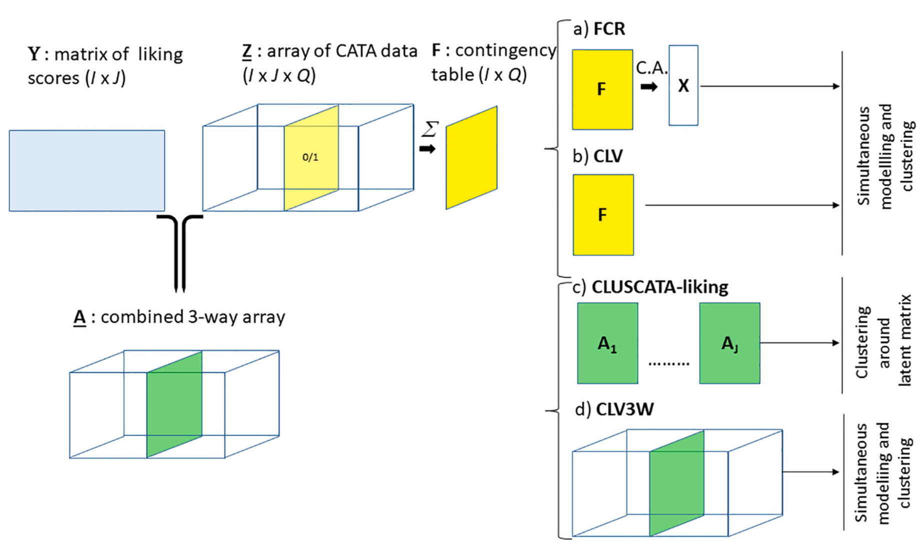

Fig. 1. Schematic representation of the data matrices, $\mathbf{Y}$, $\underline{\mathbf{Z}}$, $\mathbf{F}$ and $\underline{\mathbf{A}}$, and their integration according to the investigated approaches.

2.4. CLV with external data

CLV with external data associated to the rows (*i.e.*, the products), or CLVr, has been introduced in Vigneau & Qannari (2003) at the same time as the Clustering around Latent Variables (CLV) method. This approach was further developed in Vigneau, Endrizzi, & Qannari (2011) for identifying segments of consumers according to their liking scores while taking account of product characteristics data (external data associated to rows of **Y**) or/and consumer background information (external data associated to columns of **Y**).

Herein, besides the liking scores matrix **Y**, the external information collected on the products is the contingency matrix **F** which synthetizes the characterisation of the products given by the consumers according to the CATA attributes. The criterion to be maximized is:

$$S_r = \sum_{k=1}^{K} \sum_{j=1}^{J} \delta_{kj} cov\left(\mathbf{y_j}, \mathbf{t_k}\right) \text{ with } \mathbf{t_k} = \mathbf{F}\mathbf{a_k} \text{and } \mathbf{a_k^t}\mathbf{a_k} = 1 \tag{3}$$

where $\mathbf{a_k}$ ($k = 1, \ldots, K$) is the vector of loadings associated with the CATA attributes in the k^{th} cluster, and δ_{kj}, the (crisp) group membership of consumer j to cluster k (*i.e.* $\delta_{kj} = 1$ if consumer j belongs to cluster k, $\delta_{kj} = 0$ otherwise).

The algorithm used for solving this problem is basically an alternating optimization algorithm. It can be shown that, for a given partition, the latent component $\mathbf{t_k}$ of cluster k ($k = 1, \ldots, K$) is the first PLS regression component of the centroid variable $\bar{\mathbf{y}}_\mathbf{k}$ on **F** ($\bar{\mathbf{y}}_\mathbf{k} = \sum_{j=1}^{p} \delta_{jk}\mathbf{y_j}$ is the mean liking scores profile of the consumers belonging to cluster k). The CLVr approach is in fact a clusterwise one-dimensional PLS regression.

The normalized vectors of loadings $\mathbf{a_k}$ ($k = 1, \ldots, K$) make it possible to identify the most important CATA attributes for the various segments of consumers. By definition, each latent component $\mathbf{t_k}$ ($k = 1, \ldots, K$), which is a linear combination of the attributes in **F**, is expected to have the highest possible covariance coefficient with the centroid variable $\bar{\mathbf{y}}_\mathbf{k}$ in the associated cluster.

2.5. Three-Way cluster analysis around latent Variables

Three-Way Cluster analysis around Latent Variables (CLV3W) is a clusterwise one-dimensional CANDECOMP/PARAFAC model (Carrol & Chang, 1970; Harshman, 1970) proposed by Wilderjans & Cariou (2016) in the scope of conventional sensory profiling analysis. It seeks simultaneously a partition over one mode of a three-way array and a one-rank PARAFAC model associated with each cluster. Cariou & Wilderjans (2018) extended this approach by introducing a Non-Negativity constraint to make it better suited for the analysis of consumers' liking data (as it is desirable to separate into different clusters consumers with negatively correlated patterns of liking).

In contrast to the two previous approaches, FCR and CLVr, CLV3W is applied on the three-way data array $\underline{\mathbf{A}}$, which combines CATA data $\underline{\mathbf{Z}}$ and liking measures $\underline{\mathbf{Y}}$ (see Section 2.1). In this analysis, products ($i = 1, \ldots, I$), consumers ($j = 1, \ldots, J$) and CATA attributes ($q = 1, \ldots, Q$) are respectively associated with the first, second and third modes of $\underline{\mathbf{A}}$.

The aim of CLV3W is to identify K clusters of consumers, and, within each cluster k ($k = 1, \ldots, K$) to determine a latent component $\mathbf{t_k}$ of size ($I \times 1$), a vector of loadings $\boldsymbol{\alpha_k}$ of size ($J_k \times 1$) for the J_k consumers belonging to this cluster, and a vector of weights $\mathbf{w_k}$ of size ($Q \times 1$) associated with the CATA attributes, so that to minimize the loss criterion f:

$$f = \sum_{k=1}^{K} \sum_{j=1}^{J} \delta_{kj} \left\| \mathbf{A_j} - \alpha_{kj}\left(\mathbf{t_k}\mathbf{w_k^t}\right) \right\|^2 \text{ with } \mathbf{t_k^t}\mathbf{t_k} = 1, \mathbf{w_k^t}\mathbf{w_k} = 1 \text{ and } \alpha_{kj} \geq 0 \tag{4}$$

where $\mathbf{A_j}$ is the j^{th} slice of $\underline{\mathbf{A}}$ along its second mode, pertaining to the data of consumer j ($j = 1, \ldots, J$), as defined in Eq. (1). As in Eq. (3), δ_{kj} stands for the group's membership of consumer j to cluster k. The non-negativity constraint on α_{kj} guarantees that consumers, who belong to the same cluster, agree in terms of products' liking according to the CATA attributes they selected. An alternate least squares algorithm is conducted to determine simultaneously the partition and the various parameters associated with clusters.

2.6. Clustering of combined CATA-liking tables: CLUSCATA-liking

Clustering of combined CATA-liking tables (CLUSCATA-liking) stems from the CLUSCATA method (Llobell, Cariou, Vigneau, Labenne, & Qannari, 2019). CLUSCATA makes it possible to cluster a set of individual CATA data matrices, namely the $\mathbf{Z_j}$ matrices corresponding of each slice of $\underline{\mathbf{Z}}$ according to a consumer j ($j = 1, \ldots, J$). Based on a similarity measure, known as Ochiai coefficient (Ochiai, 1957, Llobell et al., 2019), between pairs of individual CATA data matrices, an optimization algorithm has been developed for identifying clusters of consumers such that each individual CATA data matrix related to a consumer is as close as possible to a consensus matrix associated with the cluster, the consumer belongs to. When CATA and liking information are combined, we consider J matrices $\mathbf{A_j}$, rather than the $\mathbf{Z_j}$ ones, with an adapted but similar objective that consists in minimizing:

$$D = \sum_{k=1}^{K} \sum_{j=1}^{J} \delta_{kj} \left\| \frac{\mathbf{A_j}}{\|\mathbf{A_j}\|} - \mathbf{C_k} \right\|^2 \tag{5}$$

where $\mathbf{C_k}$ is the compromise, or latent matrix, associated with cluster k ($k = 1, \ldots, K$), and δ_{kj}, as previously, stands for the group's membership of consumer j to cluster k. It is easy to show that, for a given partition of the consumers, the matrix $\mathbf{C_k}$ is simply the average of the normalized matrices $\mathbf{A_j}$ of the J_k consumers belonging to the cluster k ($k = 1, \ldots, K$).

It is worth to notice that contrariwise to the three other approaches, namely FCR, CLVr and CLV3W, the latent information associated with each cluster extracted with CLUSCATA-liking is no more unidimensional. Indeed, the latent information in cluster k is a matrix $\mathbf{C_k}$ of size ($I \times Q$). Large positive values in $\mathbf{C_k} = \left[c_{k,iq}\right]$ means that consumers in cluster k often selected the attribute q to describe the product i which has been relatively appreciated by these consumers. On the contrary, large negative values reflect that product i has often been associated with the CATA attribute q but that it has not been appreciated by the consumers. Values close to 0 may reflect either that the attribute has not been checked or that the product is moderately liked.

3. Stability assessment

For each of the clustering approaches applied on consumers' liking data, while taking account of the CATA description of the products, the number of clusters is a *meta*-parameter to be *a priori* chosen. If there is an underlying true partition or if clusters are well-separated, choosing the "true" number of clusters is an important issue. A huge number of procedures and criteria have been proposed in this scope, among which 30 procedures tested via Monte-Carlo analysis by Milligan & Cooper (1985). However, in the context of analysing the directions of preference of a set of consumers, the concept of the existence of a true partition of consumers is questionable. The concern is more to identify the main directions of preference, or in other words, to shed light on the directions around which the density of the individual preferences is the highest. Instead of recovering an underlying structure, which is often weak, the concern turns out to assess the stability of the clusters in view of the sampling variability into the population of consumers.

A very usual approach for examining the stability of a partition is to repeatedly split the set of entities to be clustered into two parts (*e.g.*, McIntyre & Blashfield, 1980; Müller & Hamm, 2014; Vigneau, Qannari, Navez, & Cottet, 2016). Among the different splitting methods, the common practice is to perform a split-half partition. The data from the first part are clustered and the clusters' centroids are determined. Thereafter, each entity of the second part is assigned to its 'best' cluster, that is, to the cluster corresponding to the nearest centroid. Finally, the agreement of group memberships of the entities of the second part is considered as a quality measure. However, Krieger & Green (1999) showed some limitations of this rationale on the basis of a simulation study. In particular, they emphasized that such internal replication clustering procedure could be problematic for determining the "correct" number of clusters, especially as the correlation among the entities increases together with an increase of the degree of overlap between clusters. One could also argue that with a set (the panel of consumers in our case study) of modest size, splitting into two parts of equal size is questionable. Actually, our aim is not really to cross-validate the clustering result made on one part of the panel with the other part, but rather to mime what it would occur if consumers were not exactly the same. In a previous work (Vigneau, Cariou, Giacalone, Berget, & Llobell, 2020), the approach adopted was to draw, repeatedly, a large number of subsets of consumers of 80% of the panel size. An alternative Monte-Carlo approach was also investigated herein.

Another strategy suggested by Jhun (1990) or by Hofmans, Ceulemans, Steinley, & Van Mechelen (2015), among others, is to use bootstrap procedures for assessing the stability, or variability, of a k-means clustering. In our case study, instead of clustering the objects (*i.e.*, the products), corresponding to the lines of the data matrix, we are rather concerned by the clustering of a set of consumers. Bootstrap samples of consumers were obtained by drawing, with replacement, J consumers among the panel of size J. As suggested by Hofmans et al. (2015), the b^{th} centroids (latent components) matrix ($b = 1, \ldots, B$) results from the clustering method applied to the b^{th} bootstrap sample, and the b^{th} partitioning matrix (group memberships) is obtained by assigning each entity (consumer) from the full data set to the cluster with the closest centroid. Thus, for the b^{th} trial, the cluster assignment is made for consumers selected to be part of the bootstrap sample but also for consumers, known as "out-of-bag" (OOB) consumers, who had been left out by the random sampling.

In the context investigated herein, both latent components and consumers' partitions were collected for each bootstrap sample. The Adjusted Rand Index (ARI) was considered to measure the similarity between the partition obtained for the whole panel of consumers (reference partition) and each bootstrap-derived partition. An ARI value equal to one indicates a perfect agreement while a value of zero reflects that the similarity is at chance level (Hubert & Arabie, 1985). The stability assessment of the latent components was performed after pairwise alignments between the reference latent components (using the whole panel of consumers) and the bootstrapped ones. This was undertaken by a permutation procedure so that the sum of the similarity indices between matched latent components is maximized. Finally, the average patterns of liking as well as frequencies of selection of CATA attributes were depicted for each bootstrap-derived partition. A simple and meaningful way to compare the FCR, CLVr, CLV3W and CLUSCATA-liking approaches consisted in drawing the barycenter and variability ellipsoids around each attribute derived from all the bootstrap penalty-lift analysis plots.

4. Illustration

The four approaches are illustrated herein on the basis of a case study on rye bread, conducted as part of a larger project about development of protein-enriched products targeted at elderly consumers in Denmark (Giacalone, 2018). The objective of the study was to explore the potential of rye bread, a traditional Danish product, for protein enrichment with whey protein hydrolysates (WPH), as well as to identify an optimal leavening agent. To this end, six samples were developed by systematically varying two experimental factors: leavening agent (sourdough and yeast) and WPH content (0%, 7%, 10% - the 0% WPH samples are referred to as "control products" in the remainder of the paper). All samples were evaluated by a panel of 134 consumers (aged 60 and over) in a central location testing facility. Consumers evaluated the samples monadically in a randomized order. For all samples, they rated the overall liking on a 9-pt hedonic scale and characterized them using a CATA questionnaire with 14 attributes: *dry*, *soft*, *sour*, *moist*, *coarse*, *bitter*, *airy*, *chalky*, *dense*, *metallic*, *off-taste*, *salty*, *yeasty*, and *chewy*. At the aggregated level (Table 2), all products were acceptable (*i.e.*, they all scored at or above the neutral point of the 9-pt scale) although they differed in liking; specifically, the two control samples were liked better than the WPH-enriched ones.

Before getting to the heart of the matter, which concerns the comparison of approaches for simultaneously identifying clusters of preference profiles while taking into account the CATA attributes, an initial exploration of the two parts of collected information (liking scores, on the one hand, CATA data, on the other hand) is proposed in order to better understand their specificities.

The two-dimensional internal preference mapping, on non-standardized liking scores, is illustrated on Fig. 2. The CLV method (Vigneau & Qannari, 2003) applied on the liking scores matrix, made it possible to identify two groups of consumers, denoted G1 (in blue) and G2 (in red), in the following. G1, with 98 consumers, is almost three times larger than G2, which counted 34 consumers. The main group, G1, comprised consumers who preferred the control products, *Ycont* and *Scont*, without any whey protein added. The mean liking scores within these two clusters are provided in Table S1.

The correspondence analysis, performed on the aggregated CATA attributes data (*i.e.* the contingency table **F** shown in Table 2), reveals mainly a one-dimensional configuration (Fig. 3). Globally, the panel of consumers often selected the attributes *Moist*, *Coarse*, *Soft* and *Airy* to describe the control products. In particular, *soft*, which was the most used among the CATA attributes was selected, on average, 62% for the two control products (*Scont* and *Ycont*) and 19% for the breads with 10% of whey protein content (*S10%* and *Y10%*). On the contrary, the higher the whey protein content, the more the products were associated with *dry*, which was the second most used attribute. Relatively to the number of consumers, *dry* was selected 60% for *S10%* and *Y10%* samples and only 9% for *Scont* and *Ycont* samples.

The stability assessment study has been performed on the basis of one hundred bootstrap consumer samples for each of the four approaches. The same bootstrap samples were involved for all of them. Partitions into two, three and four clusters have been systematically investigated. The distributions of the Adjusted Rand Index (ARI) between the reference partition, obtained on the basis of the whole panel data, and each "bootstrap" partition are shown in Fig. 4, for each approach and each number of clusters. It turns out that for the two first approaches, FCR and CLVr, making use of both the liking scores matrix **Y** and the CATA contingency table **F**, the stability of the partitions was better for segmentation into two clusters. For the two approaches, CLV3W and CLUSCATA-liking, which are based on the three-way array $\underline{\mathbf{A}}$, reference and bootstrap-derived partitions were rather different with a two-clusters partition. Regarding CLV3W, a bimodal distribution of the ARI was observed with a two clusters solution. Consequently, it was decided to retain the three-clusters solution. Regarding CLUSCATA-liking, like for FCR and CLVr, a partition into two groups appeared to be more appropriate.

In order to visualize which segments of consumers have been identified, the configuration of the preference mapping based on liking scores, as in Fig. 2, is displayed with group membership identification updated according to the clustering approach used and given the retained number of clusters. These configurations are depicted in Fig. 5.

- **FCR**. For FCR, the two clusters, denoted $G1^{FCR}$, in blue in Fig. 5(a), and G2FCR, in red in Fig. 5(a), are of equal size, with 66 consumers each. The mean liking pattern in $G1^{FCR}$ was very similar to that of the cluster G1 observed on the basis of the liking scores only, but with a

Table 2
Rye Bread data description at the panel level.

Product	factors		CATA attributes (overall number of citation)														Liking
(ID)*	leavening agent	WPH content	dry	soft	sour	moist	coarse	bitter	airy	chalky	dense	Metallic	Off-taste	salty	yeasty	chewy	(overall mean)**
Scont	sourdough	0%	11	73	4	10	28	20	35	10	22	4	15	40	63	57	6.5[a]
S7%	sourdough	7%	52	39	9	13	37	36	22	13	33	6	28	43	25	37	5.6[b]
S10%	sourdough	10%	78	20	3	11	35	26	17	16	33	18	47	33	9	23	5.2[bc]
Ycont	yeast	0%	12	90	2	3	31	12	44	21	15	3	17	12	72	55	6.4[a]
Y7%	yeast	7%	53	54	4	7	40	31	23	11	22	9	50	20	24	25	5.3[bc]
Y10%	yeast	10%	80	31	3	14	24	29	22	24	26	10	46	25	10	23	4.9[c]

* The first column shows products IDs used in the remainder of the paper.
** In the last column, letters indicate result of multiple comparisons Newman-Keuls (SNK) test ($\alpha = 5\%$).

little bit more pronounced differences between the products. On the contrary, the mean liking pattern in G2^{FCR} was very flat due to the fact that this cluster merged together consumers with heterogeneous directions of preference (Table S1).

- **CLVr**. As expected, CLVr led to a solution very similar to that obtained with CLV without external data. Thus, the mean pattern of liking in the cluster G1^{CLVr}, in blue in Fig. 5(b), is almost the same as that of G1, with the highest liking scores for *Scont* and *Ycont* products. G1 and G1^{CLVr} count about one hundred consumers and had 91 consumers in common. The second cluster, G2^{CLVr}, count 26 consumers (20% of the panel). In the G2^{CLVr} cluster, as in cluster G2, a low level of liking for *Scont* is found (Table S1).
- **CLV3W**. Three clusters have been retained when using the CLV3W approach. As it can be observed in Fig. 5(c), the main difference with the segmentation obtained with the other approaches, is that the segmentation is also based on the liking scores given to product *Scont* compared to product *Ycont*, in addition to the opposition in terms of liking between the control product against the others. This fact mainly explains the bimodality observed in the distribution of the similarity indices (*i.e.* ARI shown in Fig. 4(c)) between the reference partition and the bootstrap-derived partitions when a two-clusters partition is considered. According to the bootstrap sample, the algorithm converged towards a solution into two clusters similar to that identified with the other clustering approaches or towards a solution focusing on the distinction between the control products according to the type of yeast used. The three-clusters solution was preferred to the two-clusters partition, even if, at the consumer level, some variability can be observed in terms of cluster's assignment.The mean liking scores within the three clusters from CLV3W are shown in Table S1.
- **CLUSCATA-liking**. Finally, if we consider the two clusters solution obtained using CLUSCATA-liking (CCLik in short) approach, we can notice the similarity between Fig. 5(d) and Fig. 2(b). Accordingly, the partition into the two clusters, denoted for convenience {G1^{CCLik}, G2^{CCLik}} differs from the partition {G1, G2} by only 4 consumers among the 132 consumers of the panel. Thus, the mean liking scores within these two clusters are very similar to those of clusters G1 and G2 (Table S1).

Let us recall that the aim of this study was to identify patterns of liking but also ultimately the associated sensory drivers. Thus, consumers' segments are expected to represent the differences in liking between products but also the differences in CATA attributes that consumers selected to describe the different products. As such, we are also interested in the loadings of the CATA attributes in the definition of, or in relation with, the latent components exhibited within each cluster. For CLVr, the vector of loadings in each cluster, $\mathbf{a_k}(k = 1, \cdots, K)$ in Eq. (3), is directly estimated. In FCR, loadings vectors, $\widehat{\mathbf{b}}_\mathbf{k}(k = 1, \cdots, K)$ in Eq. (2), correspond to the contribution of the retained CA components. However, reconstruction formula may be applied by taking account of the loadings of the CATA attributes in the CA components definition. The loadings of CATA attributes for the two latent components retained with FCR approach, or with CLVr approach, are shown in Fig. 6 (a) and (b). For both approaches, considering the first cluster (G1^{FCR} or G1^{CLVr}) we observe the positive contribution of CATA attributes as *Moist*, *Coarse*, *Soft* and negative contribution of *Dry*. Differences between FCR and CLVr may be outlined for several other CATA attributes associated with G1^{FCR} or G1^{CLVr}, as for *Chalky*, *Sour* or *Salt* for instance. None of the CATA attributes seem to bring information in the second cluster of FCR, *i.e.* G2^{FCR}, which gathered consumers with heterogeneous preferences. For the second group of CLVr, *i.e.* G2^{CLVr}, the CATA attributes loadings are in the opposite sign to G1^{CLVr} which is in relation to the opposite trend in liking.

Regarding CLUSCATA-liking and CLV3W approaches, as the liking and CATA information was combined, it is less straightforward to

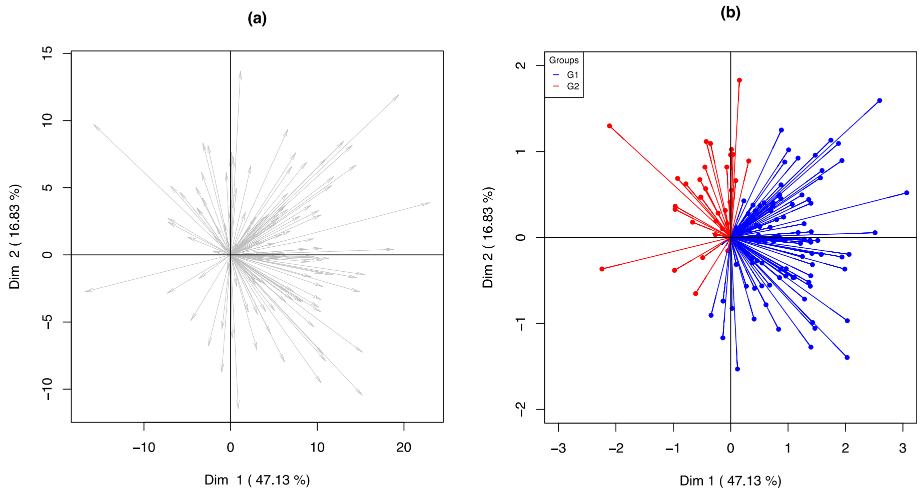

Fig. 2. Internal preference mapping for the Rye Breads case study. (a) PCA biplot. (b) Two clusters of consumers highlighted using the CLV method.

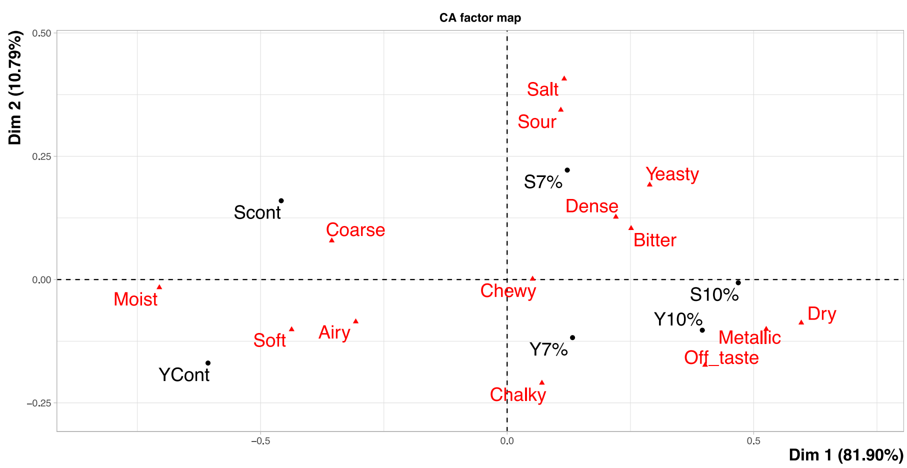

Fig. 3. Correspondence Analysis on the aggregated CATA attributes data.

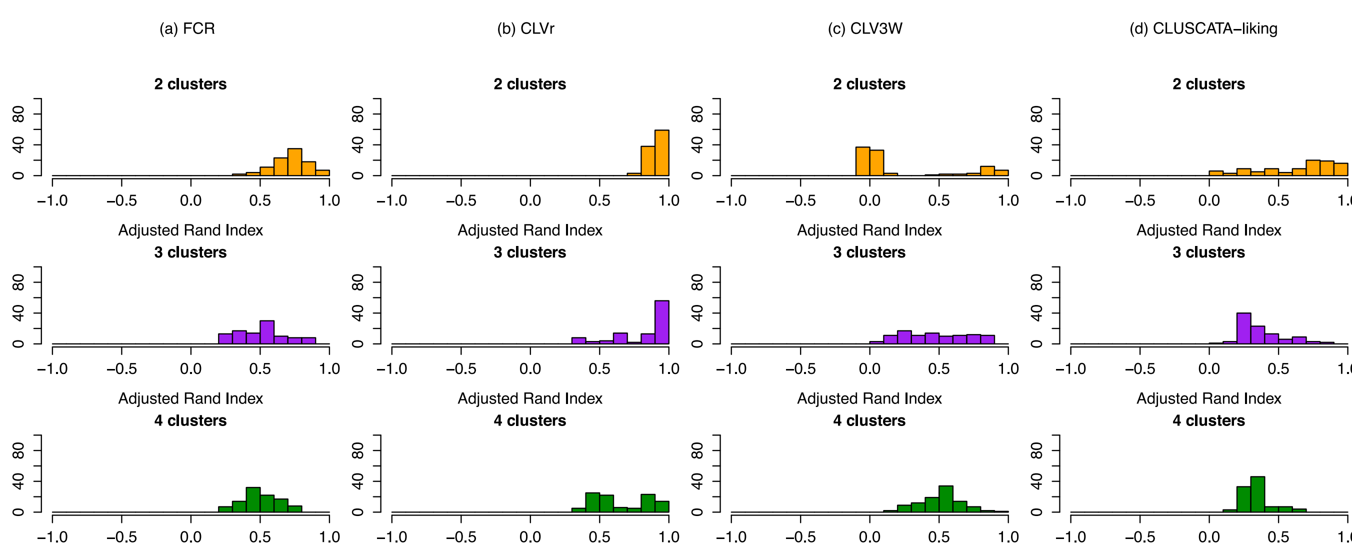

Fig. 4. Stability of the partitions assessed by the Adjusted Rand Index between the reference partition and each of the one hundred bootstrap-derived partitions.

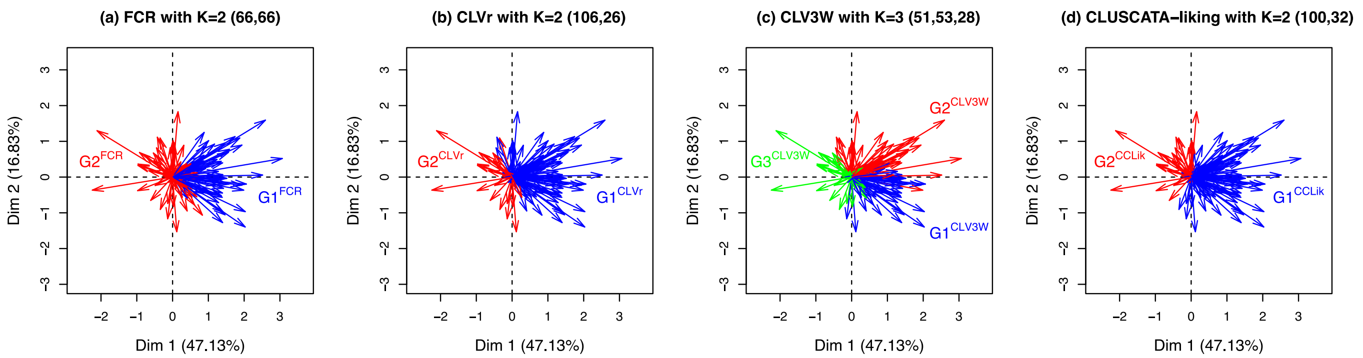

Fig. 5. Internal preference mapping with identification of the segments of consumers highlighted according to the clustering approach used and for the retained number, K, of segments (in parenthesis, the size of the clusters).

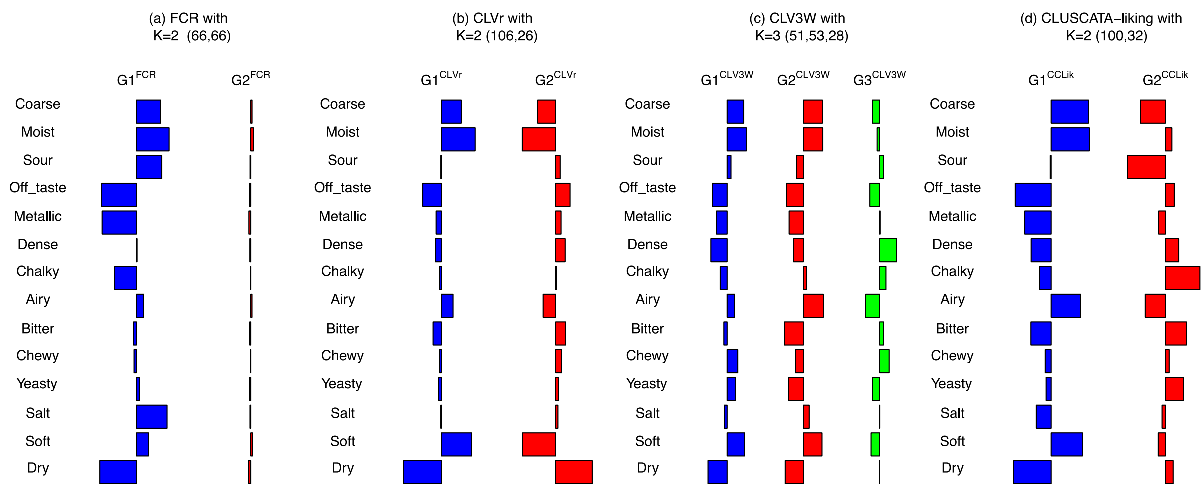

Fig. 6. Loadings of CATA attributes within each consumer segment for FCR, CLVr, CLV3W and CLUSCATA-liking approaches.

identify directly the relative importance of the CATA attributes from the outputs of both methods. Nevertheless, the latent components in CLV3W $\mathbf{t}_k (k = 1, \cdots, K)$ in Eq. (4) depict the pattern of liking within each cluster. The projection of each latent component into the space spanning by the contingency table of the CATA attributes cited by the consumers belonging to the associated cluster, makes it possible to assess the importance of the CATA attributes. The results are depicted in Fig. 6(c). It turns out that in both the first ($G1^{CLV3W}$) and second ($G2^{CLV3W}$) clusters, the CATA attributes such as *Coarse*, *Moist*, *Soft*, *Dry*, but also *Off-taste*, *Metallic*, *Dense*, *Bitter*, *Airy*, can explain the liking of the *Control* products and the disliking of the other products. It can be observed that the relationship between the pattern of citation frequency and the pattern of liking slightly differed between $G1^{CLV3W}$ (*Ycont* preferred to *Scont*) and $G2^{CLV3W}$ (*Scont* preferred to *Ycont*) for the attributes *Salt*, *Yeasty* or *Chewy*.

In the case of CLUSCATA-liking approach, each cluster is associated with a compromise or latent matrix (Eq. (5)). According to the definition of the $\mathbf{A}_j$ matrices, for each consumer j ($j = 1, \ldots, J$), we can observe that the compromise matrices, $\mathbf{C}_k$ ($k = 1, \ldots, K$), are quite unidimensional. By taking the first principal component of $\mathbf{C}_k$ of a given cluster k into account, the same procedure as for CLV3W has been adopted. The relationship between the pattern of citation of the CATA attributes and the pattern of liking within each cluster is shown in Fig. 6(d). In cluster $G1^{CCLik}$, in which *Scont* and *Ycont* were appreciated, the positive importance (that is to say the more the attributes were selected, the more the products were liked) of attributes as *Coarse*, *Moist*, *Airy*, *Soft* is highlighted, whereas the attributes such as *Dry* and *Off-taste* showed negative importance (the less they were used, the more the products were liked). Interestingly, cluster $G2^{CCLik}$, characterized by a marked dislike of *Scont*, gathered consumers who had usually used the attribute *Sour* for describing this product.

In order to compare the various approaches, another point of view was adopted. Indeed, the penalty-lift analysis plots provide an efficient way to illustrate and interpret the different results, taking into account the liking patterns as well as the frequency of CATA attributes citation. Moreover, by consolidating the penalty-lift plots across all bootstrap-derived partitions, variability of the results can be brought to light. The main challenge for this step is to efficiently match the clusters based on their latent structure. The configurations for the four investigated approaches are given in Figs. 7–10.

These configurations allow to clearly identify the most important CATA attributes related to the pattern of liking in a given segment of consumers, especially in case of stability whatever the bootstrap sample involved. This stability is especially marked when using CLVr and for the first and largest cluster, $G1^{CLVr}$ (Fig. 8, left-hand side). This could be explained by the fact that, on the one hand, consumers belonging to $G1^{CLVr}$ had all given high liking scores for *Scont* and *Ycont* products and that, on the other hand, the dependent variables (in the contingency table, **F**) involved an almost uni-dimensional PLS regression fitting model. A mirror symmetry can be observed for the second cluster $G2^{CLVr}$ of CLVr, but with much less clarity (Fig. 8, right-hand side). Regarding FCR, the first cluster configuration (Fig. 7, left-hand side) is more or less similar to those of the first CLVr cluster, with a little more variability which could be induced by the second CA component. As stated before,

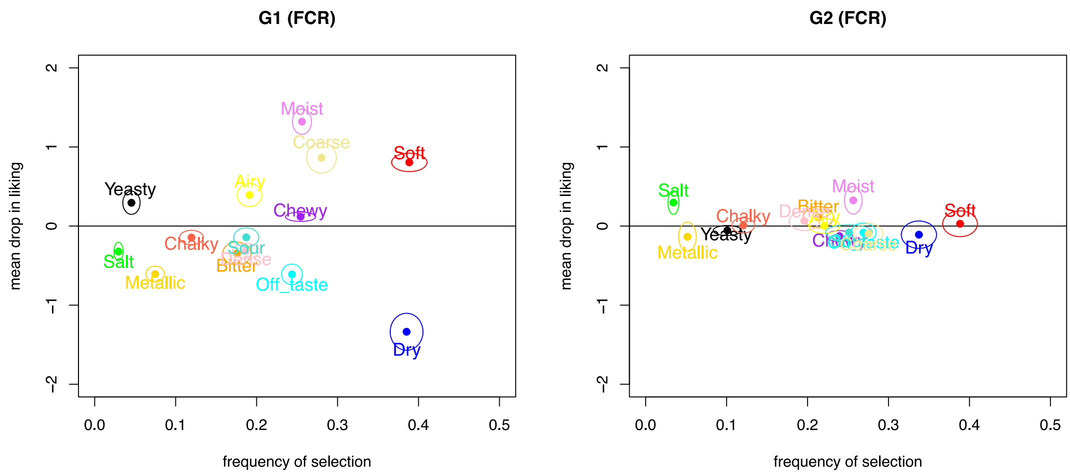

Fig. 7. Penalty-lift analysis plot for 100 bootstrap-derived partitions into two clusters using FCR.

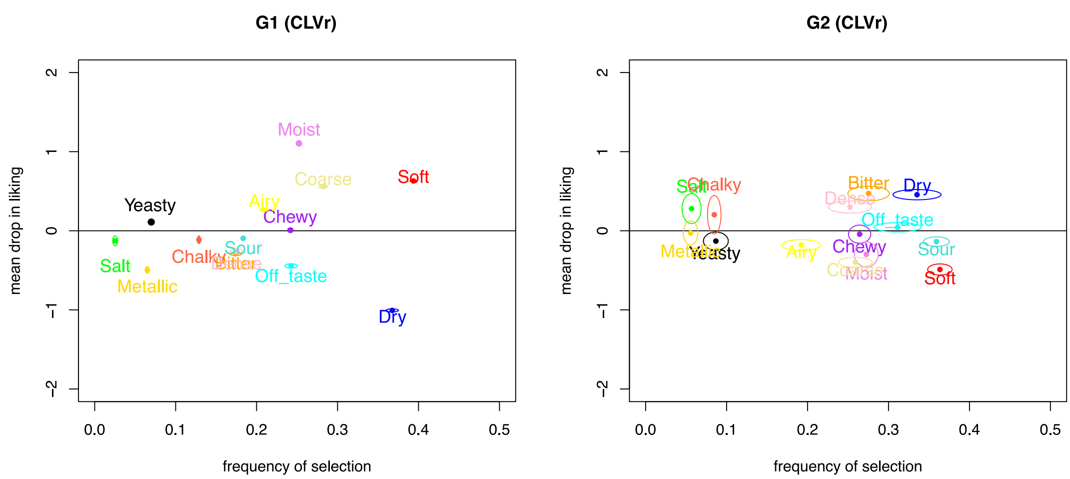

Fig. 8. Penalty-lift analysis plot for 100 bootstrap-derived partitions into two clusters using CLVr.

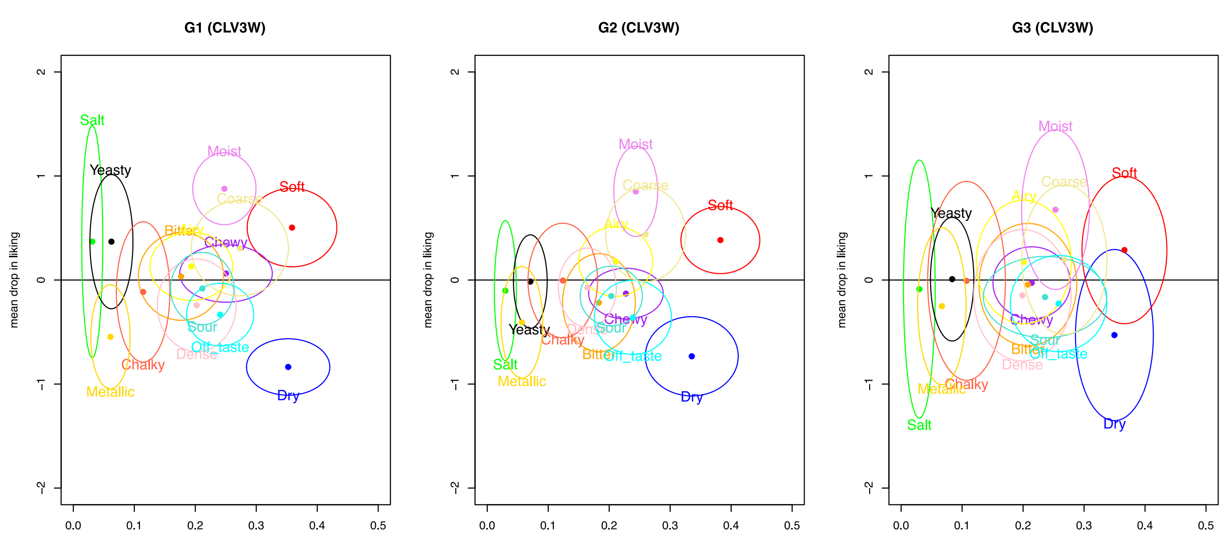

Fig. 9. Penalty-lift analysis plot for 100 bootstrap-derived partitions into three clusters using CLV3W.

the second FCR cluster corresponds to a poorly defined pattern (Fig. 7, right-hand side). This may comprise a group of consumers without any clear relation between CATA attributes and liking or reflect that it is a heterogeneous group. The penalty lift plot associated with the first cluster obtained with CLUSCATA-liking (Fig. 10, left-hand side) presents a quite similar structure than those for the first two approaches. As a

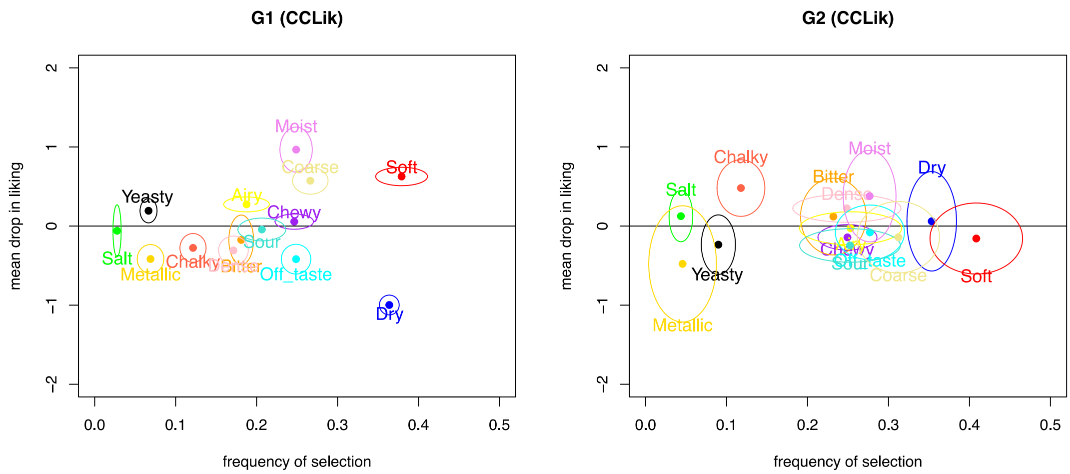

Fig. 10. Penalty-lift analysis plot for 100 bootstrap-derived partitions into two clusters using CLUSCATA-liking.

matter of fact, this cluster is made up of the same type of consumers. However, in contrast to FCR and CLVr, CLUSCATA-liking is based on the three-way array $\underline{\mathbf{A}}$, which led logically to greater variability between bootstrap-derived configurations.

Penalty-lift analysis plots drawn in the case of CLV3W (Fig. 9) presents a relatively high variability, because of the sparsity of the data combining CATA and liking information, stored into the three-way array $\underline{\mathbf{A}}$ (as for CLUSCATA-liking approach). This can involve a greater number of candidate partitions and make the parameters' estimation more sensible to the underlying PARAFAC model when considering the bootstrap samples (Eq. (4)). Even if the latent components associated with the clusters were found in a fairly consistent way regardless of the bootstrap sample (especially for the two first clusters), the assignment of out-of-bag consumers is a source of additional variability. As for the second FCR cluster, the fuzzy configurations for the third cluster exhibited with CLV3W could lead to discard this cluster for interpretation purpose, to mainly focus on the segments with the most salient traits.

5. Discussion and conclusion

In this work, we investigated the segmentation of a panel of consumers according to their differences in liking, while simultaneously considering the description of the products they gave based on a list of CATA attributes. Four alternative clustering approaches were proposed. We compared the clustering solutions as well as the attributes that explain the liking, or disliking, associated with the products under study for each segment. The comparison between the four clustering approaches described in the previous section was based on a single case study. The interest of the Rye Bread case study was that differences between products were controlled thanks to an experimental design. However, it turned out that the description of the products by the means of the CATA attributes was almost always the same for all consumers with an opposition between *soft* and *dry* depending on the addition of whey protein hydrolysates (WPH), or not, in the bread dough. This resulted, at the panel level, in a relatively unidimensional subspace for the product description. The FCR approach, and even more clearly the CLVr approach, which use projection onto this subspace, provided stable and contrasted clustering into a main cluster regarding FCR, or two opposed clusters regarding CLVr. On the contrary, working at the individual level by combining the liking scores of each consumer with the selection of CATA attributes he, or she, made led to an individual information that is finer but more complex to capture and aggregate. The noticeable variability observed between the solutions resulting from the bootstrap samplings is inherent to the specificity of the three-way array of data involved in CLV3W and CLUSCATA-liking.

The first two approaches, FCR and CLVr, which are simpler and more stable, present nevertheless a certain theoretical flaw. If the aim is to segment a panel of consumers, it may seem simplistic, or even biased, to consider the product description obtained at the level of the whole panel. It is questionable whether it makes sense to study in detail, on the one hand, the individual liking assessment provided by each consumer and to aggregate, on the other hand, the choice of CATA attributes over all the panel. However, the main interest of those approaches is to develop algorithms to perform the clustering of the consumers according to their liking profiles while simultaneously estimate the coefficients for the fitting of the cluster's liking patterns as a function of external data such as CATA description. A partition of the consumers is designed to highlight interpretable structures by means of the coefficients of within clusters' models. Compared with the usual strategy of doing clustering on the liking data alone, and then drawing separate penalty-lift plots, it was observed on the basis of the rye bread case study, that the clustering as well as the penalty assessments were not so different. However, the variability estimated using a bootstrap procedure was slightly reduced when imposing subspace projection as in FCR or CLVr approaches. To sum up, working at the panel level regarding the external information (*i.e.* the global contingency table of the CATA attributes) is straightforward and appeared to be well-adapted when the panel has been first checked to be consistent with regard to the CATA description task. In this case, it turns out that FCR or CLVr approaches make it possible to rely consumers' segmentation of preference with sensory drivers as in external preference mapping. But if a certain heterogeneity exists within the panel in terms of the use of CATA descriptors, there is little chance of being able to distinguish between consumers with similar preferences but focusing on different product attributes.

In order to consider, at an individual level, both the CATA attributes selections and the likings given by each consumer, an innovative data integration procedure has been proposed. This led to the definition of a three-way array $\underline{\mathbf{A}}$ on the basis of which CLV3W and CLUSCATA-liking are working. In practice, it can be observed that this three-way array is relatively sparse, as sparse as can be the three-way array $\underline{\mathbf{Z}}$ of the CATA attributes selection. This sparsity is probably the main source of versatility of the solutions obtained with CLV3W or CLUSCATA-liking when a resampling technique is applied. As such, it is rather tricky to make direct comparisons between the stability of the partitions obtained with the first two approaches (FCR and CLVr) and the last two (CLV3W and CLUSCATA-liking). Nevertheless, these two three-way approaches tend to be promising when the panel of consumers is no more consistent with regard to the CATA description of the products. Interestingly, the modification of the "bootstrap" Adjusted Rand Index distribution when increasing the number of clusters shows that managing information at an individual level may lead to highlight different partitions. In the Rye bread case study, CLV3W with two clusters highlighted, according to the bootstrap sample involved, roughly two types of clustering. Thus, a partition into three clusters seemed preferable. At this stage, further investigations are needed to better understand how obtained partitions are merely influenced by CATA description or by liking.

Another point discussed in this paper was the choice of a resampling procedure in order to assess the stability of clustering solutions. Preliminary work was undertaken by repeatedly subsampling a fraction (say 80%) of the consumers' panel and to compare how these consumers were clustered in the initial partition (based on the whole panel) and those obtained using the actual fraction. However, the main drawback of this strategy was the lack of independency between the reference and the bootstrap-derived partitions. From this point of view, a better strategy would be the repeated splitting of the dataset into two equal-sized sub-samples in which cluster analyses are performed separately as in Müller & Hamm (2014). Although interesting, this approach needs to have a large number of entities (*i.e.* consumers) to cluster which is not usually the case for consumer studies where the number of consumers is typically around one-hundred. For this reason, bootstrap re-sampling procedure proposed by Hofmans et al. (2015) was preferred here. Nevertheless, two difficulties arose that remain to be solved: the first one was to assign the out-of-bag consumers according to the similarity criterion of a consumer to the centroids of the clusters. This was specific to each approach. The second one was, if necessary, to define an ad-hoc pairwise matching rule between the latent components exhibited by each of the approaches evaluated herein. Let us mention that the (Adjusted) Rand Index is a between-partitions similarity criterion to be favoured over any other because it does not imply solving the tricky pairwise matching problem. The variability of the outputs discussed herein, especially the one shown in the superimposed bootstrap-derived penalty lift-analysis plots, depends as much on the complexity of the input data as on the inaccuracies in the re-assignation of out-of-bag entities or the misalignment of the clusters' latent components.

At this stage, the purpose of the present work was to discuss alternative clustering approaches in order to integrate both hedonic data and product characterization by CATA in panel segmentation analysis. The application of these approaches was demonstrated on the basis of a single study, and a larger number of case studies would be needed to better understand the respective advantages and disadvantages of each proposed approach. Furthermore, other data coding strategies for combining both types of information at the individual consumer level data can be envisioned and studied further. To date, the four approaches may be applied using already developed algorithms, written in R or Matlab. In particular, CLVr and CLV3W are available in the ClustVarLV R package (Vigneau, Chen, & Qannari, 2015). CLUSCATA-liking, developed in R, is available on request from Fabien Llobell, FCR (Matlab and R code), is available to the site "Software & Downloads - Nofima Data Modelling (nofimamodeling.org)".

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodqual.2021.104358.

References

Ares, G., Antúnez, L., Bruzzone, F., Vidal, L., Giménez, A., Pineau, B., … Jaeger, S. R. (2015). Comparison of sensory product profiles generated by trained assessors and consumers using CATA questions: Four case studies with complex and/or similar samples. *Food Quality and Preference, 45*, 75–86. https://doi.org/10.1016/j.foodqual.2015.05.007

Ares, G., Dauber, C., Fernández, E., Giménez, A., & Varela, P. (2014). Penalty analysis based on CATA questions to identify drivers of liking and directions for product reformulation. *Food Quality and Preference, 32*, 65–76. https://doi.org/10.1016/j.foodqual.2013.05.014

Ares, G., & Jaeger, S. R. (2015). Examination of sensory product characterization bias when check-all-that-apply (CATA) questions are used concurrently with hedonic assessments. *Food Quality and Preference, 40*, 199–208. https://doi.org/10.1016/j.foodqual.2014.10.004

Berget, I., Mevik, B.-H., & Næs, T. (2008). New modifications and applications of fuzzy C-means methodology. *Computational Statistics & Data Analysis, 52*(5), 2403–2418. https://doi.org/10.1016/j.csda.2007.10.020

Cariou, V., & Wilderjans, T. F. (2018). Consumer segmentation in multi-attribute product evaluation by means of non-negatively constrained CLV3W. *Food Quality and Preference, 67*, 18–26. https://doi.org/10.1016/j.foodqual.2017.01.006

Carroll, J. D., & Chang, J.-J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart–Young" decomposition. *Psychometrika, 35*(3), 283–319. https://doi.org/10.1007/BF02310791

Giacalone, D. (2018). Product Performance Optimization. In Ares, G., & Varela, P. (Eds.) Methods in Consumer Research, Volume I (Chapter 7. pp. 159-185), Elsevier.

Greenacre, M. (2017). *Correspondence analysis in practice*. CRC Press.

Harshman, R. A. (1970). Foundations of the Parafac procedure: Models and conditions for an explanatory multi-modal factor analysis. *UCLA Working Papers in Phonetics, 16*, 1–84.

Hofmans, J., Ceulemans, E., Steinley, D., & Van Mechelen, I. (2015). On the Added Value of Bootstrap Analysis for K-Means Clustering. *Journal of Classification, 32*(2), 268–284. https://doi.org/10.1007/s00357-015-9178-y

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification, 2*(1), 193–218. https://doi.org/10.1007/BF01908075

Jaeger, S. R., & Ares, G. (2014). Lack of evidence that concurrent sensory product characterisation using CATA questions bias hedonic scores. *Food Quality and Preference, 35*, 1–5. https://doi.org/10.1016/j.foodqual.2014.01.001

Jaeger, S. R., Lee, P.-Y., Jin, D., Chheang, S. L., Rojas-Rivas, E., & Ares, G. (2019). The item-by-use (IBU) method for measuring perceived situational appropriateness: A methodological characterisation using CATA questions. *Food Quality and Preference, 78*, 103724. https://doi.org/10.1016/j.foodqual.2019.103724

Jaeger, S. R., Chheang, S. L., Jin, D., Roigard, C. M., & Ares, G. (2020). Check-all-that-apply (CATA) questions: Sensory term citation frequency reflects rated term intensity and applicability. *Food Quality and Preference, 86*, 103986. https://doi.org/10.1016/j.foodqual.2020.103986

Jaeger, S. R., Swaney-Stueve, M., Chheang, S. L., Hunter, D. C., Pineau, B., & Ares, G. (2018). An assessment of the CATA-variant of the EsSense Profile®. *Food Quality and Preference, 68*, 360–370. https://doi.org/10.1016/j.foodqual.2018.04.005

Jhun, M. (1990). Bootstrapping K-Means Clustering. *Journal of the Japanese Society for Statistics, 3*(1), 1–14. https://doi.org/10.5183/jjscs1988.3.1

Johansen, S.-B., Hersleth, M., & Naes, T. (2010). A new approach to product set selection and segmentation in preference mapping. *Food Quality and Preference, 21*, 188–196. https://doi.org/10.1016/j.foodqual.2009.05.007

Krieger, A. M., & Green, P. E. (1999). A cautionary note on using internal cross validation to select number of clusters. *Psychometrika, 64*(3), 341–353. https://doi.org/10.1007/BF02294300

Krhisnapuram, R., & Keller, J. M. (1996). The possibilistic C-means algorithm: Insights and recommendations. *IEEE Transactions on Fuzzy Systems, 4*(3), 385–393. https://doi.org/10.1109/91.531779

Llobell, F., Cariou, V., Vigneau, E., Labenne, A., & Qannari, E. M. (2019). A new approach for the analysis of data and the clustering of subjects in a CATA experiment. *Food Quality and Preference, 72*, 31–39. https://doi.org/10.1016/j.foodqual.2018.09.006

MacFie, H. (2007). *Preference mapping and food product development. In MacFie, H., Consumer-led food product development* (pp. 551–592). Woodhead Publishing.

McIntyre, R. M., & Blashfield, R. K. (1980). A nearest-centroid technique for evaluating the minimum-variance clustering procedure. *Multivariate Behavioral Research, 15*(2), 225–238. https://doi.org/10.1207/s15327906mbr1502_7

Menichelli, E., Olsen, N. V., Meyer, C., & Næs, T. (2012). Combining extrinsic and intrinsic information in consumer acceptance studies. *Food Quality and Preference, 23*(2), 148–159. https://doi.org/10.1016/j.foodqual.2011.03.007

Meyners, M., & Castura, J. C. (2014). Check-All-That-Apply Questions. In P. Varela, & G. Ares (Eds.), *Novel techniques in sensory characterization and consumer profiling (Chapter, 11* pp. 271–305). CRC Press.

Meyners, M., Castura, J. C., & Carr, B. T. (2013). Existing and new approaches for the analysis of CATA data. *Food Quality and Preference, 30*(2), 309–319. https://doi.org/10.1016/j.foodqual.2013.06.010

Meyners, M. (2016). Testing for differences between impact of attributes in penalty-lift analysis. *Food Quality and Preference, 47*, 29–33. https://doi.org/10.1016/j.foodqual.2014.11.001

Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika, 50*, 159–179.

Müller, H., & Hamm, U. (2014). Stability of market segmentation with cluster analysis – A methodological approach. *Food Quality and Preference, 34*, 70–78. https://doi.org/10.1016/j.foodqual.2013.12.004

Ochiai, A. (1957). Zoogeographic studies on the soleoid fishes found in Japan and its neighbouring regions. *Bulletin of Japanese Society of Scientific Fisheries, 22*, 526–530.

Plaehn, D. (2012). CATA penalty/reward. *Food Quality and Preference, 24*(1), 141–152. https://doi.org/10.1016/j.foodqual.2011.10.008

Rinnan, A., Giacalone, D., & Frøst, M. B. (2015). Check-all-that-apply data analysed by Partial Least Squares regression. *Food Quality and Preference, 42*, 146–153. https://doi.org/10.1016/j.foodqual.2015.01.018

Spinelli, S., Monteleone, E., Ares, G., & Jaeger, S. R. (2019). Sensory drivers of product-elicited emotions are moderated by liking: Insights from consumer segmentation. *Food Quality and Preference, 78*, Article 103725. https://doi.org/10.1016/j.foodqual.2019.103725

Vigneau, E., & Qannari, E. M. (2003). Clustering of Variables Around Latent Components. *Communications in Statistics, Simulation and Computation, 32*(4), 1131–1150.

Vigneau, E., Endrizzi, I., & Qannari, E. M. (2011). Finding and explaining clusters of consumers using the CLV approach. *Food Quality and Preference, 22*(8), 705–713. https://doi.org/10.1016/j.foodqual.2011.01.004

Vigneau, E., Chen, M., & Qannari, E. M. (2015). ClustVarLV: An R Package for the Clustering of Variables Around Latent Variables. *RJournal, 7*, 134–148. https://journal.r-project.org/archive/2015/RJ-2015-026/index.html.

Vigneau, E., Qannari, E. M., Navez, B., & Cottet, V. (2016). Segmentation of consumers in preference studies while setting aside atypical or irrelevant consumers. *Food Quality and Preference, 47*, 54–63. https://doi.org/10.1016/j.foodqual.2015.02.008

Vigneau, E., Cariou, V., Giacalone, D., Berget, I., & Llobell, F. (2020, October). Combining hedonic information and CATA description for consumer segmentation: new methodological proposals and comparison. Sensometrics 2020, virtual conference, Norway.

Wedel, M., & Steenkamp, J.-B.-E.-M. (1991). A clusterwise regression method for simultaneous fuzzy market structuring and benefir segmentation. *Journal of Marketing Research, 28*(4), 385–396. https://doi.org/10.2307/3172779

Wilderjans, T. F., & Cariou, V. (2016). CLV3W: A clustering around latent variables approach to detect panel disagreement in three-way conventional sensory profiling data. *Food Quality and Preference, 47*, 45–53. https://doi.org/10.1016/j.foodqual.2015.03.013

Williams, A. Carr, B. T., & Popper, R. (2011). Exploring analysis options for check-all-that-apply (CATA) questions. In 9th Rose-Marie sensory science symposium, Toronto, ON, Canada.