ORIGINAL ARTICLE

Journal of
Sensory Studies

WILEY

Application of check-all-that-apply and non-metric partial least squares regression to evaluate attribute's perception and consumer liking of apples

Thais Mendes da Silva | Cristiana Peano | Nicole Roberta Giuggioli

Department of Agricultural, Forest and Food Sciences (DISAFA), University of Turin, Turin, Italy

Correspondence
Thais Mendes da Silva, Department of Agricultural, Forest and Food Sciences (DISAFA), University of Turin, Largo Paolo Braccini, 2, Grugliasco (TO) 10095, Piedmont, Turin, Italy.
Email: thais.mendesdasilva@unito.it

Abstract

Attribute's perception and impact on liking are crucial in quality assessments. An approach with analytical and sensory methods is often necessary to ensure a reliable description of quality perception without neglecting product's characterization. Moreover, the presence of nonlinear patterns demands appropriate models. A methodology is proposed to assess quality and attributes of apples cultivated at different orchard's elevation. Physicochemical, sensory and consumer tests with check-all-that-apply and penalty analysis were performed. A non-metric partial least square model (NM-PLS) was applied with a new coding system. The methodology highlighted similar results at all assessments for taste, with no significant differences among samples. Texture was evaluated different among assessments. Differences were found for hardness and astringency only at the panel level, while there was an agreement considering crispness, better described by the analytical index "average drop." The NM-PLS model confirmed sweet, intense odor and the latent attribute "notes" as the most related to liking. The methodology confirmed that differences found among trained panelists my not be relevant for consumers, therefore, an integrated approach is needed. It shows a new system to create a regression model that provides information at a specific and a global attribute level to deal with multicomponent parameters.

Practical Applications

This work describes a new consumer test methodology to investigate attributes perception and product penalties, therefore, it can be used by industries and researchers to have a deeper understanding of quality and liking of new products. It also highlights advantages and drawbacks of different quality assessment types and suggests an integrated approach is key to assess quality, consumer liking and attribute's perception. Finally, it proposes a new index to investigate crispiness.

1 | INTRODUCTION

The investigation of an attribute's perception and its impact on liking of fruit products is a main step in most of the quality assessments performed in the research and industrial fields. When a quality assessment is performed to obtain a deeper understanding between the quality profile (represented by attributes) and liking, the use of a consumer driven approach is an essential step to obtain more reliable

information since most of the assessments involving analytical methods and trained panelists can lead to a poor description of consumer's perception (Ares, Dauber, Fernández, Giménez, & Varela, 2014; Blaker, Plotto, Baldwin, & Olmstead, 2014). This phenomenon is even more felt for products that are described by multicomponent attributes, such as apples. In this case, the attribute texture is very important to understand the liking of the product (Charles et al., 2018; Corollaro et al., 2014) and involves complex parameters such as crispness and juiciness, for which there is still a lack of consensus regarding which analytical method to be used (Chen & Opara, 2013). Therefore, to describe an apple's quality, the use of trained panelists may be more advantageous due to their ability of integrating different aspects of an attribute to formulate a global judgment (Contador, Shinya, & Infante, 2015), such as sound and hardness to describe apple's crispiness. A drawback of this approach is the fact that trained assessors may describe the product differently and detect differences among samples that may be irrelevant for consumers (Chen & Opara, 2013). On the other hand, a quality assessment based exclusively on the consumer response is not often exhaustive due to consumers not being able to describe why they like a product (ten Kleij & Musters, 2003). Therefore, an integrated approach and proper techniques at the consumer level are key to perform a successful quality assessment.

To improve the consumer test assessment, 1–5-point scales such as Just-about-right (JAR) scales are widely applied to evaluate an attribute's perception (Ares et al., 2014; Estiaga, 2015; Hampson & Quamme, 2000; Robertson, Meredith, Senter, Okie, & Norton, 1992; Zamzami & Ariyawardana, 2019). JAR became a very popular tool to investigate the optimum intensity of sensory attributes (Estiaga, 2015) and is often combined with the Penalty analysis in order to obtain mean drop estimates of liking caused by suboptimal attributes (Pagès, Berthelo, Brossier, & Gourret, 2014). Currently, in order to reduce bias linked to the use of JAR scales, where consumers may focus on sensory characteristics that they would not normally do, the Penalty analysis has been applied also on Check-all-that-apply (CATA) questions (Ares et al., 2014) to evaluate liking and apple's attribute perception. In Both JAR scales and CATA questions approaches, the results of penalty analysis can be combined with regression techniques to predict liking and this has been done by converting JAR or CATA attributes scales into dummy variables (Ares et al., 2014; Li, 2013; Xiong & Meullenet, 2006). Multivariate or regression techniques are needed since the one factor analysis of variance model applied on penalty analysis results does not take into account correlations among attributes (Pagès et al., 2014) and, therefore, the approach is not appropriate to predict liking from JAR or CATA data (Xiong & Meullenet, 2006). The partial-least-square (PLS) model, a linear method that deals with multicollinearity of parameters (Mendes da Silva, Briano, Peano, & Giuggioli, 2020), has been proposed to fulfill these goals (Ares et al., 2014; Xiong & Meullenet, 2006).

Another important concern in the assessment of attributes and liking is the presence of nonlinear relationships among data due to the use of ordinal scales. Different authors have pointed the incorrect use of 1–5-point scales as interval scales rather than ordinal ones in consumer research (Li, 2013; Xiong & Meullenet, 2006). When using JAR scales, little or no behavioral differences may be present between consumers who rate a product at "just about right" and "too little," while the difference between "too little" and "much too little" could be huge. This could be linked to the fact that consumers who assume "stronger positions" concerning their attitude toward an attribute are more prone to penalize the liking of products than consumers that assume "weaker positions" (De Langhe, Puntoni, & Larrick, 2017). Moreover, it is well known that different attributes may display several types of relationships with consumer liking, as explained in the Kano model (Rivière, Monrozier, Rogeaux, Pagès, & Saporta, 2006). For example, many attributes may display mainly a linear relationship with the degree of liking or disliking, such as sweetness perception in apples, apricots, strawberries and peaches (Lado et al., 2019; Mendes da Silva, Peano, & Giuggioli, 2019; Molina et al., 2006; Stanley, Prakash, Marshall, & Schröder, 2013), however, more complex attributes regarding texture or aroma may stimulate consumer satisfaction when they are perceived in the product but not necessarily stimulate disliking when they are absent (Lado et al., 2019; Mendes da Silva, Peano, et al., 2019). Furthermore, this situation may vary across horticultural crops since consumer's expectations on attributes depends on the type of product. Thus, the use of nonlinear models and division of attributes according to their influence on liking are essential to properly predict liking. The first goal of this work is to conduct an integrated quality assessment using analytical methods, a trained panel and a consumer driven approach in order to evaluate differences between apples cultivated at two orchards with different levels of elevation. The second goal of this work is to build a prediction model of liking using the information obtained at all levels of the quality assessment.

2 | MATERIALS AND METHODS

2.1 | Materials

"Gala" apples (*Malus domestica* Borkh. "Brookfield") were cultivated in two areas with different altitude 500 m a.s.l. "Area 1" and 300 m a.s.l. "Area 2" in the northern part of Italy. Meteorological information for each chosen site were obtained from stations near the orchards (meteorological regional information center) and are reported in the Table 1.

Apples were subjected to physicochemical analysis, sensory evaluation with a trained panel and a consumer test. Both samples were

TABLE 1 Information of the meteorological measurements from June to August for the two apple orchards Area 1 and Area 2

Metereological measurements	Area 1 500 a.s.l.	Area 2 300 a.s.l.
The lowest temperature from June to August (°C)	9.6	9.9
The highest temperature from June to August (°C)	33.6	34.8
Total precipitation from June to August (mm)	2.08	3.75

provided by the same Italian supplier and subjected to the same post-harvest storage. Apples were harvested in the last week of august and stored under cold temperature (0–1°C), relative air humidity of 85% and under controlled atmosphere (0.5% O_2 and 1% CO_2) until the last week of November, when they were brought to the Department of Agricultural, Forest and Food Sciences (DISAFA) at the University of Turin and Cadir lab SRL (Alessandria, Italy) to be subjected to the analysis. Samples were placed at room temperature 24 hr before the assessment.

2.2 | Physicochemical assessment

Twenty replicates for each apple sample were analyzed for total solid solubles (TSS) with a digital refractometer (Atago, mod. PAL-1, Tokyo, Japan) and the titratable acidity (TA) of apple's juice was determined in triplicate by titration with 0.1 N NaOH to pH 8.1 and expressed as g 100 g^{-1} of malic acid according to Organization for Economic Cooperation and Development (OECD) guidelines (2009). The ratio between TSS and TA (TSS/TA ratio) values was calculated. Color was measured with a colorimeter (Konica Minolta, mod. CR-400, Tokyo, Japan) upon two sides (the more colorful and the less colored side) in the equatorial zone of 20 fruits. The L^*, a^* and b^* were recorded with Konica Minolta software (Spectra Magic, NX). Deriving from L^*, a^* and b^*, other color indexes previously tested were calculated in order to enhance sensitiveness of color evaluation (Bosco et al., 2015; Pathare, Opara, & Al-Said, 2013):

$$Chroma^* : \sqrt{a^2* + b^2*}$$

The CIEDE2000 Color-Difference index (DE 2000) was used in order to assess color difference of samples. The DE 2000 is more complicated and more accurate since it takes into account the different sensitivities of the eyes to different colors. For brevity, the calculation is illustrated in Sharma, Wu, and Dalal (2005).

Fifteen replicates of each apple sample were compressed with a texturometer TA.XT2+ (Stable Micro Systems, Surrey, United Kingdom) with a compression platen (diameter 75 mm). Since shape and dimension of samples may strongly influence compression tests, the replicates were cut longitudinally in two halves and each half was laid down and compressed at a pre-test speed of 5 mm/s, test speed of 10 mm/s and post-test speed of 10 mm/s. The distance was set to 8.0 mm and the trigger force was 5 g. The first peak of the force-distance curve was registered as hardness of samples. The software compression's script was modified in order to register the jaggedness of the force-distance curve. A high jagged line is expected for crispy food since the peaks produced are a result of the fracture events that have occurred during the test (Campbell, 2016). Therefore, the number of positive and negative peaks after the first force peak from the curve were registered, as illustrated in Figure 1, and the average drop parameter was automatically calculated by the software. This function calculates the average drop in force between consecutive peaks and
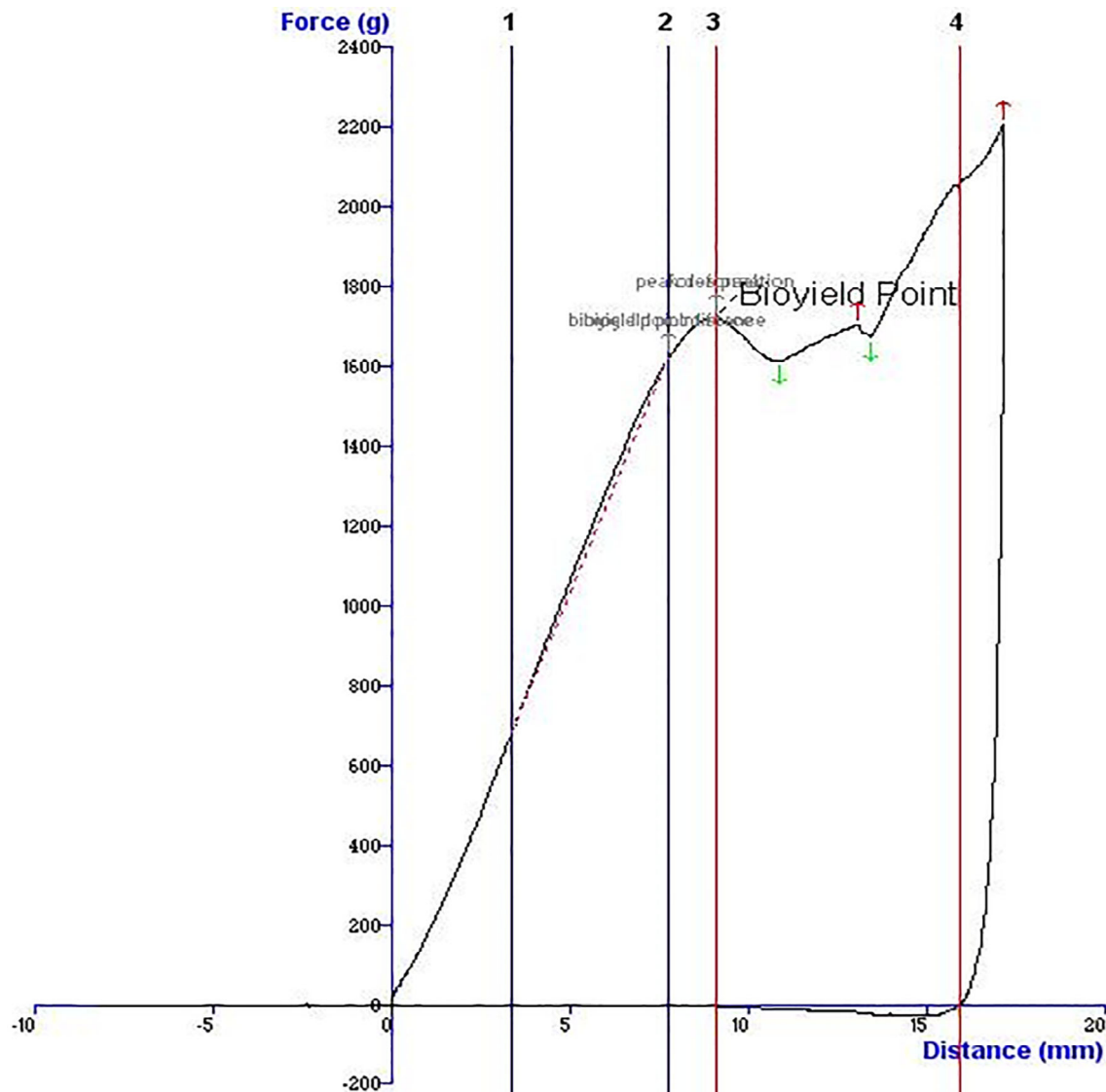
troughs over a selected region. It has been suggested that crunchiness or crispness are associated with the mechanical force required to compress food until it yields suddenly and fractures into small pieces along with a pleasant acoustic output (Tunick et al., 2013). Therefore, by measuring the number of peaks and the average drop parameter it is expected to measure to some extend the crispness of apple samples.

2.3 | Sensory assessment of apples

Ten panelists from Sata SRL (Alessandria, Italy) were selected and trained in sensory evaluation of apples as recommended by ISO 8586 (ISO, 2012). The training program was carried out weekly for 3 months and was divided in theoretical and practical sessions. Panelists were trained by discussing the definition of quality parameters selected for sensory evaluation, explaining the score sheet and method of scoring. The practical training first involved the use of recognition threshold tests, followed by discrimination tests, such as ranking, "duo-trio" and triangle tests to recognize differences in intensity of the following standards: fructose and sucrose (sweet taste), citric acid (sour taste), sucrose octaacetate (bitter taste), tannic acid (astringency), ethyl-hexanoate (apple odor) and 4-(p-acetoxyphenyl)-2-butanone (berry-like odor). The standards were prepared at different concentrations according to the panelists' preparation level. All standards were supplied by AROXA (Cara Technology, Leatherhead, United Kingdom). Panelists were trained to use the variety clone "Royal Gala," as a reference standard in order to evaluate samples, as recommended by ISO 8586 (ISO, 2012). Before the training, the clone was assessed in order to verify that values of TSS and TA were adequate to set the product as a benchmark of excellence, as it is recognized in the northern part of Italy. Two different continuous scales compliant to ISO 412 1-2003 (ISO, 2003) were used: a 1–9 continuous intensity scale with one end, "too far away from the reference," and the opposite end, "very close to the reference," for an overall quality assessment is important to note that the term indicated as "overall-quality" was used to describe how close to the reference the tasted samples were overall (summarizing all the sensory descriptors used in this work), as described in Mendes da Silva, Peano, et al. (2019) work. It was not used to express the hedonic linking of samples, A continuous intensity 1–9 scale with one end, "extremely low intensity," and the opposite end, "extremely high intensity," was used to assess descriptive sensory attributes. In this work, panelists were asked to not consider aspects and color of the product, but only to focus on taste, texture and aroma. Therefore, the descriptive sensory attributes were: ripe apple odor, berry-like odor, hardness, crunchiness, astringency, sweetness, sourness, and aroma. Panelists were presented with four apple slices from different fruits.

2.4 | Consumer test assessment

Forty consumers of a specific target were recruited from Cuneo, Italy, where around 60% of apple's cultivation is represented by red apples

FIGURE 1 Example of a force-distance curve obtained by the modified compression test's script. The number of positive (green) and negative (red) peaks and the average drop were automatically calculated between the third and fourth anchors (anchors in red)

such as "Gala" and "Red Delicious." Their ages were 18–25 years old and around 60% were female. It was decided to work with this specific target since young-adults are more prone to place a greater value to secondary factors such as the vocationality of the territory. All samples were tasted monadically in random order to take into account position and carry over effects. Samples were presented in plastic containers labeled with three-digit random numbers, at room temperature. Water was available for rinsing between samples. Consumers were first asked to score their overall liking using a 9-point hedonic scale anchored at "dislike very much" (1) and "like very much" (9) for each product. A CATA question with 17 attributes related to sensory characteristics of apples was compiled. Consumers were instructed during a 1-hr lesson in order to ensure the meaning of these terms were well-understood. During the test, they were asked to check only the attributes they considered to be appropriate descriptors of each samples. The attributes were selected based on previous literature for

apple (Ares et al., 2014; Mendes da Silva, Marinoni, Peano, & Giuggioli, 2019; Plotto, 1998): intense apple odor, firm, crispy, mealy, fibrous, astringent, juicy, sweet, sour, aromatic, ripe apple notes, green apple notes, grassy notes, honey notes, citric notes, berry notes and woody notes. After testing each sample, consumers were asked to complete the CATA question to describe their ideal apple.

2.5 | Statistical analysis

2.5.1 | Physicochemical and panel test assessments

Differences among apple samples concerning the physicochemical data were assessed with the student t test or in the case of lack of normal distribution of sample's data, the Mann–Whitney test. The t test for paired means was used to assess the significant differences

between the means of apple samples for all sensory attributes scored by the panel test.

2.5.2 | Consumer test assessment

The consumer test methodology was similar to Ares et al. (2014) work. Overall liking scores were analyzed using the paired *t* test considering sample as fixed source of variation and consumer as a random effect. Frequencies of each sensory attributes from CATA questions were determined by counting the number of consumers that used that term to describe each sample and the ideal. Only attributes that were cite at least by 20% of consumers were considered to be important for the product's perception. The McNemar test was used to assess significant differences between samples for all terms included on the CATA question.

Penalty analysis

The penalty analysis was carried out on consumer responses to determine the drop in overall liking scores associated with a deviation from the ideal for each attribute from the CATA question. However, coding of CATA frequencies of samples and ideal CATA questions was done differently from the binary coding proposed in Ares et al. (2014) work, where from each attribute derived two coded categories: the agreement between samples and ideal CATA questions was coded as 1 for each consumer (e.g., attribute checked or not checked for both CATA's questions) and the disagreement was coded as 0 (e.g., attribute checked in ideal but not for sample's CATA question and vice-versa). In this work, the coding was done splitting the disagreement situations in two codes (Table 2): -1 if an attribute was checked in ideal but not for sample's CATA question and 1 if the opposite was true. The 0-code remained represented by the agreement between both ideal and sample CATA's question (for checked and non-checked attributes). Therefore, as similarly proposed in Plaehn (2012), the 0 response for CATA is analogous to the JAR response and the 1 and -1 responses are analogous to an "out-of-JAR response." The penalty analysis was applied following the same principles of penalty analysis with JAR scales, with the one-dimensional model, as proposed in Pagès et al. (2014) work and the *p* values of penalty coefficients were assessed to determine if they were significantly different from zero ($\leq$.05). The analysis was done using the R Package SensomineR (Husson, Le, & Cadoret, 2020).

Regression model: Prediction of liking

The non-metric partial least square model (NM-PLS), as described by Russolillo (2010), was applied to Penalty analysis categories in order to estimate liking and to individuate the weight of deviation of each attribute from the CATA's question. CATA's attributes were used as predictors and overall liking was used as the dependent variable. In order to use attributes as predictors in the model, the penalty's categories of all attributes were transformed into dummy variables following a similar approach used in Xiong and Meullenet (2006) and Ares et al. (2014), however, considering the out-of-JAR variables could not necessarily lead to an overall mean drop, as proposed by Plaehn (2012), each Penalty categories' mean of attributes was assessed in order to first classify attributes as "must be," "reversal" and "attractive" attributes, and then coding was performed specifically for each type of attribute. "Must-be" attributes are those which consumers that expect to but cannot perceive them in the product, will penalize the product greater than consumers that do not expect the attribute on their ideal but perceive it. Considering Gala apples, sweetness is a must-be attribute since it is highly expected to be found in the product among consumers that prefer sweet apples (Plotto, 1998). "Reversal" attributes are those which consumers that do not expect to but perceive them in the tasted product will penalize it greater than consumers that expect to but do not perceive them in the tasted product. This can be applied to aroma notes that are not expect, such as green apple notes in sweet cultivars of apples. Finally, "attractive" attributes are those which consumers that expect to but cannot perceive them in the tasted product will penalize it while consumers that do not expect to but perceive them in the product will not penalize it since they are positive even though when they are not expected. An example of an attractive attribute is crispness, which is nearly always appreciated in the product since it is related to product's freshness, despite genotype differences and consumer preference. Therefore, for each type of attribute the penalty categories were re-coded in -2 (greater penalization), -1 (smaller penalization) and 0 (close to the ideal) as shown in Table 3.

Once the penalty categories were transformed, the NM-PLS model was applied with the R package "plspm" (Sanchez, Trinchera, & Russolillo, 2017). Only attributes that were considered as deviated from the ideal for at least 20% of the consumers were used. They were grouped in latent variables that represent a global quality attribute: "texture," "taste" and "notes." In this way, it will be possible to assess how a specific attribute impact overall liking, such as sweet,

TABLE 2 Illustration of penalty analysis coding process and the resulting penalty analysis category (0, 1 and -1)

Consumers	Attribute1—Sample 1	Attribute1—Ideal1	Penalty category	...	Attribute17—Sample 1	Attribute17—Ideal1	Penalty category
Consumer1	x		1		x	x	0
Consumer2	x	x	0				0
...							
Consumer40		x	-1		x		1

TABLE 3 Coding process of penalty categories according to attributes classification in "must-be" attributes, "reversal" attributes and "attractive" attributes

	Must-be attributes (attributes with the lower likingmean on category −1)	Reversal attributes (attributes with the lower likingmean on category 1)	Attractive attributes
Penalty category coded for penalty analysis as −1 (attributes checked in the ideal CATA question but not in the product)	−2	−1	−2
Penalty category coded for penalty analysis as 1 (attributes checked in the product CATA question but not in the ideal)	−1	−2	0
Penalty category coded as 0 (attributes checked or non-checked for both ideal and product CATA's question)	0	0	0

Abbreviation: CATA, check-all-that-apply.

without neglecting the assessment of a global quality attribute, such as taste. This approach is more in agreement with the human way of thinking, which integrates all parameters to formulate a global judgment of quality rather than consider each parameter separately (Giuggioli, Peano, & Mendes da Silva, 2020). The PLS was created for the estimation of a model with metric data as it is needs numerical variables in order to estimate the relationships between latent variables and manifest variables (Russolillo, 2010). However, in this work the NM-PLS was applied since it is assumed that the new coding originated variables (attributes) that are composed by ordinal states (0, −1 and −2) depending on the degree of mean drop. Moreover, it is also expected that mean drops between the −1 and −2 codes in comparison to the 0 code are different, which denotes the presence of non-linearity as the scale is not wide enough to assume continuity for those variables (Nappo, 2009). Therefore, all predictors were introduced as ordinal variables while overall liking scores were introduced as a continuous dependent variable. Cronbach's alpha and Dillon–Goldstein's values were assessed in order to classify the global latent variables as reflective or formative, along with loadings and communalities of single predictors of the outer model as performed in Mendes da Silva et al. (2020). Due absence of distribution hypothesis on data the PLS inferential tools are usually based on resampling techniques (Nappo, 2009), therefore, in this work the model was validated through a bootstrap technique and the confidence intervals of path coefficients and loadings were assessed.

3 | RESULTS AND DISCUSSION

3.1 | Physicochemical and sensory analysis of apple samples

There were significant differences considering the color of apples as shown by the colorimetric analysis (Figure 2). In particular, apples from Area 1 were redder (as shown by higher a^* values) and presented a more intense color (as 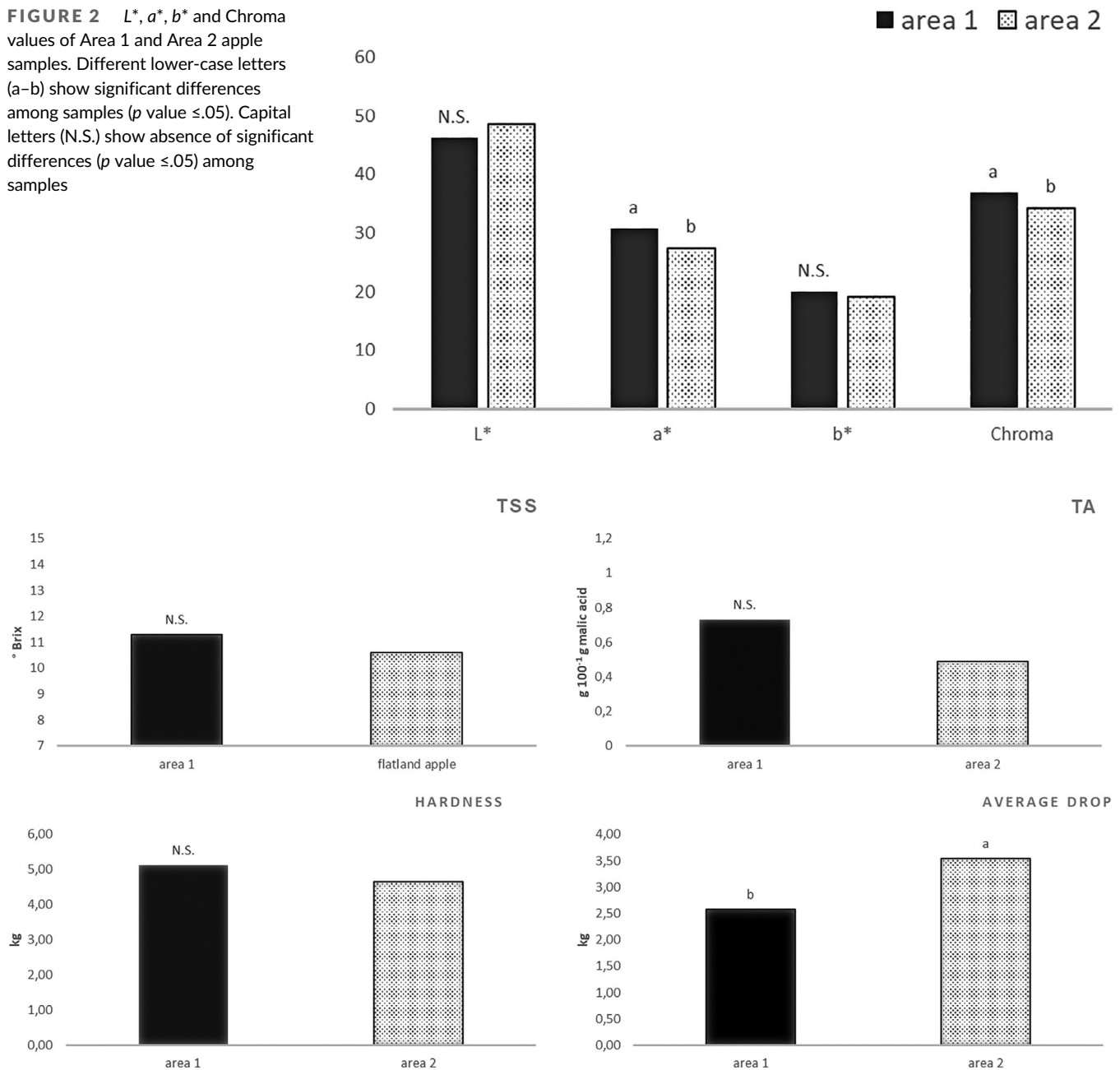shown by higher Chroma values). The DE 2000 index was equal to 2.79, a value that is higher than 2.50, the threshold value used to indicate visible differences between two products (Sharma et al., 2005). Despite color differences, both samples presented very similar quality attributes considering most of the parameters assessed and there was a good agreement between physicochemical data and the sensory scores given by panelists. There were no significant differences among Area 1 and Area 2 samples considering TSS and TA values (Figure 3), however, when taking into account the TSS/TA ratio values, the apple fruits would be classified in two distinct classes of taste, as the samples cultivated in Area 2 would be considered as sweet apples due to a TSS/TA ratio value over 20 (21.6) and Area 1 samples would be considered as sour apples as they presented a TSS/TA value below 20 (15.5), according to previous researches (Almeida & Gomes, 2017; Mendes da Silva, Marinoni, et al., 2019).

Concerning the TPA analysis, there were no differences among products considering the parameter hardness. Both samples produced very jagged lines, with an average number of peaks equals to 18.8 for the Area 2 sample and 17.4 for the Area 1 sample. However, only the average drop of peak's force displayed significant differences among samples (Figure 3), with the Area 2 samples presenting a higher value. Therefore, results from the sensory analysis are important in order to verify which one of the two parameters (the average drop or the number of peaks) is a better indication of perceived crispness.

Results from sensory test (Figure 4) confirmed TSS and TA values: no differences in sweetness or sourness perception were found among samples. Concerning the texture parameters, Area 1 samples were found to be less hard and less crispy than Area 2 samples. Therefore, the average drop parameter was more in agreement with the sensory test than the number of peaks and it is suggested to be a better indicator of crispness. Up to our knowledge, very few studies have assessed crispness using the same compression test applied in this work, therefore, these results still need to be confirmed in other sensory trials and for other cultivars, as each apple cultivar may respond differently to a same type of texture test (Corollaro et al., 2014). Moreover, it is also necessary to extend the type of assessment used in this work to other horticultural crops.

FIGURE 2 L^*, a^*, b^* and Chroma values of Area 1 and Area 2 apple samples. Different lower-case letters (a–b) show significant differences among samples (p value ≤.05). Capital letters (N.S.) show absence of significant differences (p value ≤.05) among samples
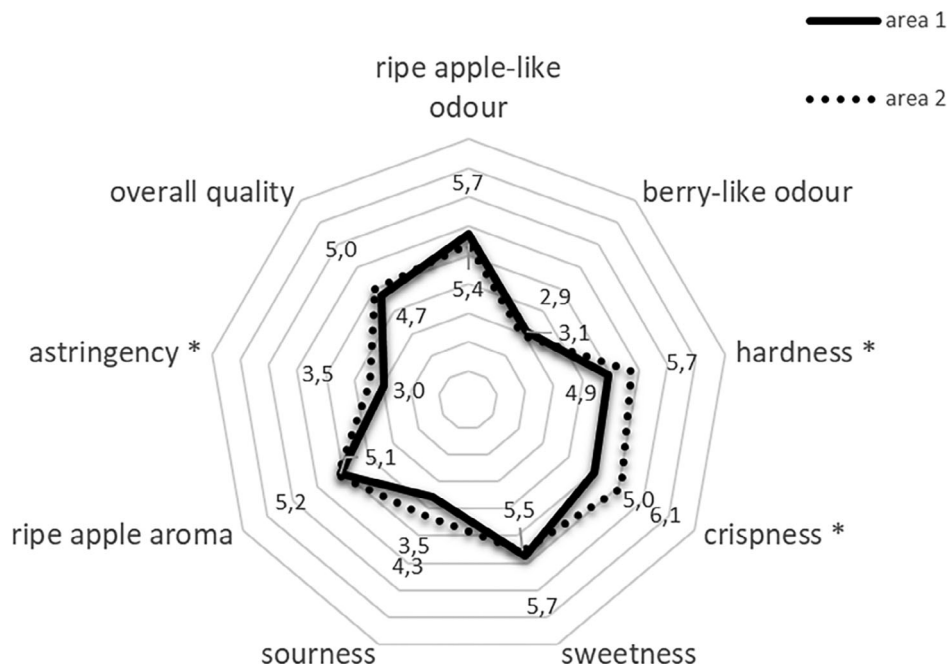
■ area 1 ▨ area 2

FIGURE 3 Total soluble solid content (TSS), titratable acidity (TA), hardness and average drop values of Area 1 apple and Area 2 samples. Different lower-case letters (a–b) show significant differences among samples (p value ≤.05). Capital letters (N.S.) show absence of significant differences (p value ≤.05) among samples
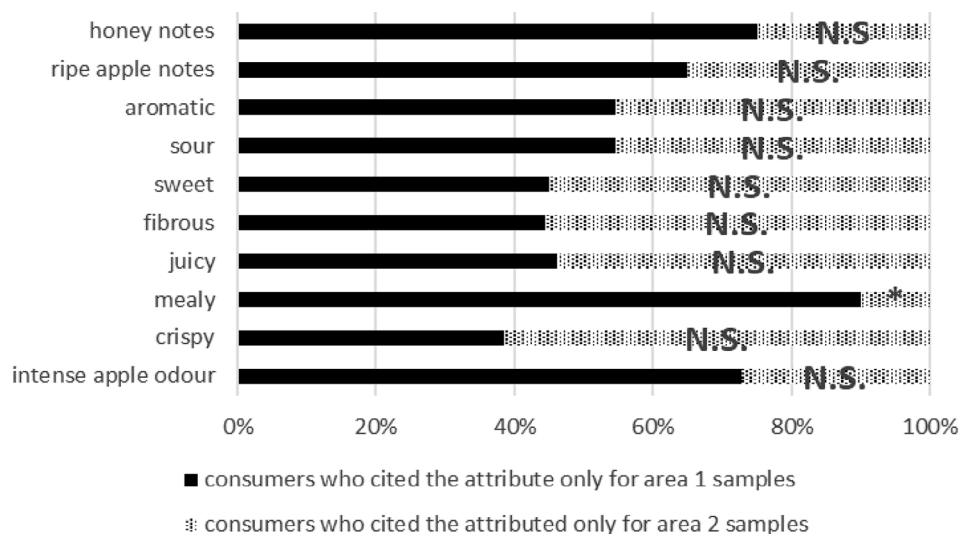
The results of sweetness perception did not confirm the results of the analytical parameter ratio TSS/TA, which have suggested samples differ in terms of sweetness perception. In the sensory test, the parameter was found to be not significantly different among samples. This could be explained by the differences found in texture properties, especially in terms of crispness, which could have influenced the sweetness perception of samples. It is well known that interactions between flavor and texture attributes can lead to poor prediction of sweet taste by instrumental data (Corollaro et al., 2014).

Finally, Area 1 apples were perceived as less astringent by panelists than the Area 2 samples, even though both scores were below

the scale average, indicating the sensation was not intense for both samples. Scores were low also for berry-like odor, a common descriptor of "Gala" apples (Plotto, 1998). Even though there were no differences in the overall quality scores among samples, it is important to note how Area 2 samples presented better quality characteristics considering texture attributes. Those results are in agreement with literature. Apples that are from orchard at higher altitudes usually display better quality characteristics than lowland apples, especially considering the color development (Charles et al., 2018; Neri, 2004), however, many articles showed that differences between mountain and hillside cultivations are not consistent and may vary across different apple

FIGURE 4 Sensory analysis of Area 1 and Area 2 samples. The "*" indicates attribute's values that were significantly different among samples (p value ≤.05)

FIGURE 5 Percentage of consumers who cited a specific check-all-that-apply (CATA) attribute only for one of the apple samples. The "*" indicates attributes that resulted significant in Mc Nermar test (p value ≤.05)
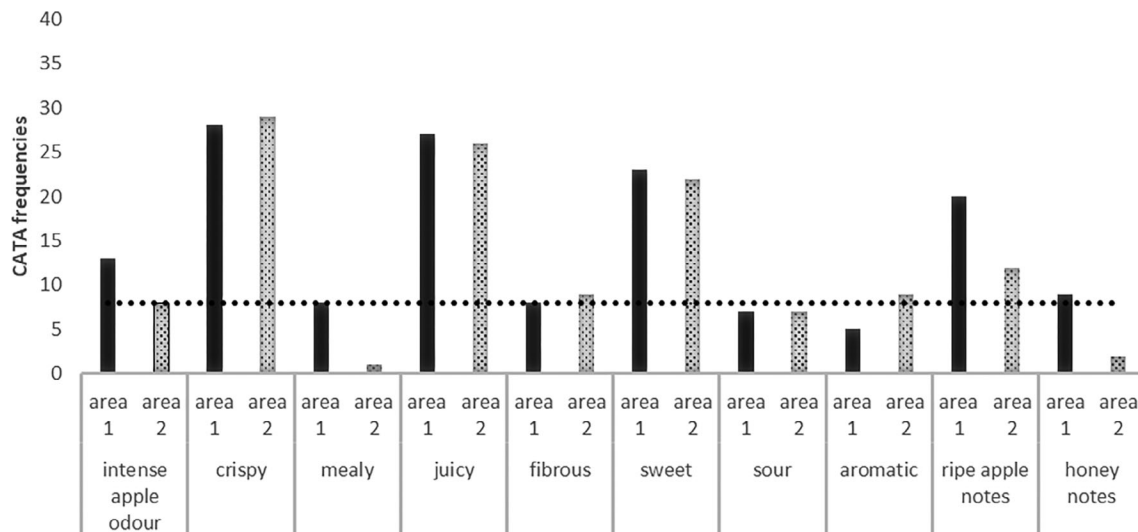
varieties (Gregori, Bergamaschi, Berra, & Folini, 2016). For what concerns Gala, superior quality of mountain apples in the Italian territory is not clear and, differently from other cultivars such as Golden delicious and Fuji, Galas are still advised for all cultivation systems (Bassi & Fideghelli, 1995; Charles et al., 2018).

3.2 | Consumer test assessment

There was a good agreement between the panel test results and the consumer test for many quality attributes. For what concerns the overall liking scores, there were no differences among samples, with 6.61 and 6.89 for Area 1 and Area 2 samples, respectively (data not shown). The McNemar test pointed differences among samples considering the number of CATA frequencies only regarding mealiness (Figure 5), with Area 1 samples resulting more cited than Area 2 samples, even though both products presented a higher and similar number of CATA frequencies for the crispy attribute, while mealiness had fewer citations. Therefore, it can be concluded that some consumers might have perceived differences among samples regarding the crispness of products, which is expressed in terms of mealiness in the CATA questionnaire. It is important to note that even though some consumers have perceived differences among samples in terms of mealiness, the most part of them considered both products to be crispy. Therefore, this fact could be linked to the absence of significant differences among overall liking scores of samples as liking is known to be driven mainly by textural factors, as described in previous works (Charles et al., 2018; Corollaro et al., 2014),

FIGURE 6 List of attributes cited at least by 20% of consumers on check-all-that-apply (CATA) questions of Area 1 and Area 2 samples

The Figure 6 indicates the attributes that were cited by at least 20% of consumers and it is possible to note how samples presented similar number of citations for most cited descriptors such as crispy, juicy, sweet, and ripe apple notes. It is also clear that for attributes that were cited least, there is a greater difference on the number of citations, even though not significant.

3.2.1 | Penalty analysis results

Considering products were very similar, the penalty analysis was calculated for each attribute from all products. This approach is advisable since the main interest of this work is to evaluate attributes that affect Gala apples as a whole instead of apples from a specific cultivation system. With this approach the results obtained in this article are more extendable to the population (Pagès et al., 2014).

From Table 4, it is possible to observe that only the penalty coefficients of sweet and intense apple odor were significantly different from zero, and in both cases for the disagreement category −1. The disagreement category −1 indicates that both attributes were checked for the ideal product but were not checked for the tasted samples in the CATA questionnaire. Therefore, the fact that the penalty coefficient was found significant for both attributes indicates that consumers who did not perceive the samples as sweet and with an intense apple odor have penalized the overall liking scores significantly. The percentage of consumers that penalized the samples under the two attributes was high, especially in the case of Area 2 samples with regard the attribute intense apple odor: in this case the percentage was almost half of participants (44.74%). It is interesting to note that, even though the attribute sweet was highly cited on CATA question for both products, the remaining part of consumers that did not cite this attribute on the sample's CATA question had a huge impact on the overall linking mean drop. Therefore, it could be assumed that even though samples were considered to be sweet, they were not sweet enough, at least for an important part of the participants.

By assessing the coefficients, it is possible to identify many attributes with a negative penalty coefficient, such as crispy, aromatic, citric notes and honey notes concerning the disagreement category 1. The disagreement category 1 indicates consumers that have not checked the attributes for the ideal product but have checked them for the samples. A negative penalty coefficient indicates that the mean of overall liking scores of consumers who represent this category is higher than the mean of consumers that represent the agreement categories. This means that, even though consumers did not expect to find those attributes in their ideal apple, the presence of these attributes on the samples did not lead to a drop of overall scores. This is in agreement with Plaehn (2012) work, where the author stated that "out-of-JAR categories" (1 and −1 in this work) are not necessarily associated with mean drops ("penalties") for the CATA penalty case. Therefore, these attributes are considered to be attractive, as described in the Section 2, and must be codified differently from the others when creating the PLS model.

Even though many penalty coefficients were not significantly different from zero, there was a huge percentage of times where a disagreement between the samples and the ideal citations occurred, especially in the case of the attributes: crispy, fibrous, sweet, sour, intense apple odor, aromatic, ripe apple notes, green apple notes and honey notes. For these attributes, at least 20% of times a disagreement occurred, and therefore, following Ares et al. (2014) work, those attributes were chosen to be used as independent variables of the PLS model. It is interesting to note that the number of consumers that penalize samples for not being sour (the −1 category) was higher than the number of consumers that did not choose this attribute in their ideal but found it on the samples. It is well described in literature how sourness of apples may be perceived as a negative attribute (Molina et al., 2006), however, it is also well described how this attribute may enhance the flavor of product when it is perceived along with the sweetness taste (Harker et al., 2002; Mendes da Silva, Marinoni, et al., 2019).

TABLE 4 The one-dimensional penalty results: the CATA attributes, the penalty categories that represent the disagreement (-1 and 1), the percentage of times the disagreement penalty categories are given for each sample (frequency of consumers columns), the penalty coefficient and its associated p value ($\leq$.05)

CATA attributes	Penalty categories	Frequence of consumers Area 2 sample (%)	Frequence of consumers Area 1 sample	Penalty coefficients (all samples)	p value
Intense apple odor	1	2.63	0.00	0.130	.922
	-1	44.74	33.33	0.997	.002
Hard	1	7.89	7.69	0.797	.179
	-1	2.63	2.56	-0.203	.838
Crispy	1	10.53	7.69	-0.318	.563
	-1	13.16	20.51	0.671	.115
Mealy	1	2.63	17.95	0.261	.618
	-1	5.26	0.00	-0.239	.812
Juicy	1	10.53	7.69	0.629	.260
	-1	15.79	17.95	-0.151	.723
Fibrous	1	23.68	20.51	0.573	.132
	-1	0.00	0.00	/	/
Sweet	1	15.79	15.38	0.379	.389
	-1	23.68	30.77	0.903	.013
Sour	1	13.16	12.82	0.678	.159
	-1	28.95	25.64	0.550	.131
Astringent	1	2.63	0.00	1.757	.211
	-1	2.63	2.56	-0.243	.807
Aromatic	1	13.16	5.13	-0.448	.433
	-1	39.47	46.15	0.323	.332
Ripe apple notes	1	21.05	25.64	0.117	.769
	-1	34.21	20.51	0.466	.220
Green apple notes	1	23.68	7.69	0.585	.191
	-1	21.05	20.51	0.418	.295
Grassy notes	1	13.16	12.82	0.810	.084
	-1	2.56	7.89	-0.690	.328
Honey notes	1	7.89	23.08	-0.348	.418
	-1	12.82	13.16	1.018	.031
Citric notes	1	0.00	5.13	-0.747	.452
	-1	5.26	0.00	1.253	.208
Berry notes	1	0.00	2.56	-1.253	.369
	-1	0.00	0.00	/	/
Woody notes	-1	0.00	2.56	1.747	.211
	1	5.26	5.13	0.517	.469

Abbreviation: CATA, check-all-that-apply.

3.2.2 | Prediction of liking with the NM-PLS

As a result of the Penalty analysis assessment, the prediction model of liking was created using the NM-PLS model. It was created by grouping the attributes crispy and fibrous under the latent variable "texture," sweet and sour were grouped under "taste" and intense apple odor, aromatic, ripe apple notes, green apple notes and honey notes under the group "notes." As suggested by Sanchez (2013), all
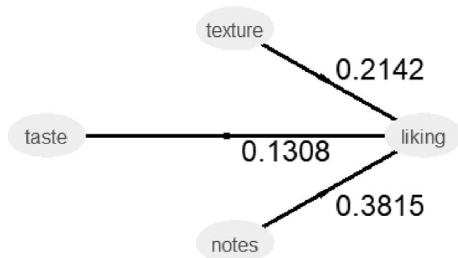
NM-PLS selective ratios indicated "texture" and "taste" variables should be classified as reflective latent variables, with both presenting Dillon-Goldstein's rho values >0.70 (0.707 in both cases), while the group "notes" was classified as formative as communality's values of all attributes were <0.50. This was expected since sour and sweet penalties are known to be highly related in fruit products and represent the same concept (sweetness) (Pagès et al., 2014), therefore, the attributes are highly explained by the latent variable "taste," with high

communality values (>0.50) and loadings (>0.70). On the other hand, the attributes in the "notes" group represent different aspects of fruit metabolism and might be judged by consumers individually, therefore, as each "notes" attribute will contribute to form a different aspect of the latent variable, the latter will not sufficiently explain the variance of each single attribute and communalities and loadings values are low.

In Figure 7 it is possible to assess the NM-PLS inner model, which shows the relationship among the latent global attributes and the dependent variable liking (the overall liking scores). At this level it is possible to see how the group "notes" has a higher impact on the liking scores and the mean drop of this parameter by assessing its higher path coefficient, which expresses the relationship between the group and the liking variable (Nappo, 2009). The latent predictor was significant at a 0.001 level while the "texture" group was significant at a 0.05 level and both predictors were validated by the bootstrap procedure. The group "taste" was not significant. These results might be attributable to the presence of complementary attributes in the

formative group "notes," which contribute to a higher explained variance than groups composed by different variables that underlie the same concept. This explains also differences between "taste" and "texture" coefficient's p values even though both latent variables are reflective: sweet and sour are usually more correlated than crispy and fibrous. In fact, those texture attributes can be strongly related, since horticultural products may be perceived as harder and, therefore, crispier but only to some extension since crispiness is a multicomponent attribute that depends also on the sound produced while eating, especially in the case of apples.
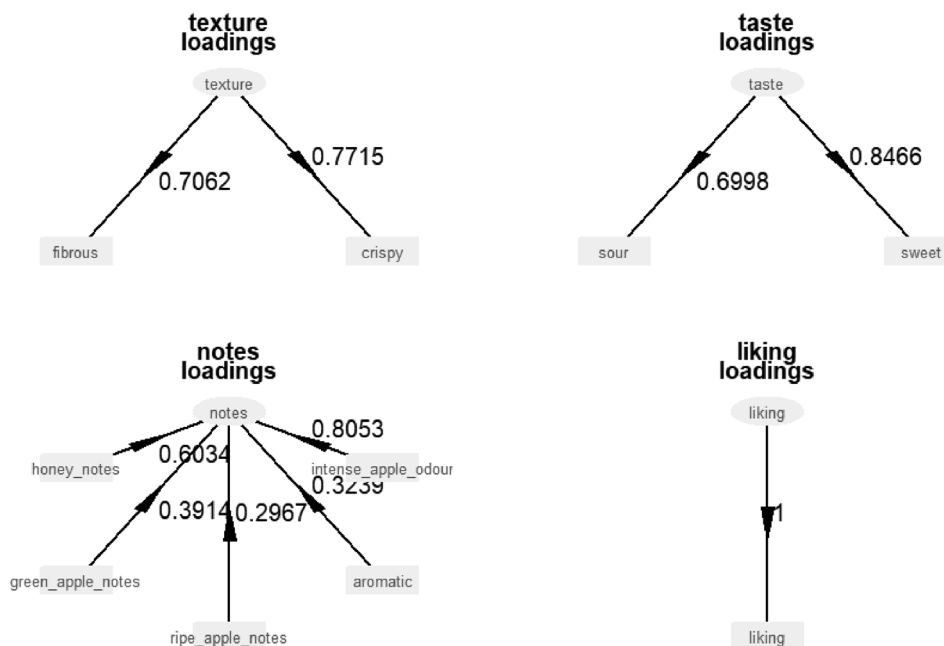
In Figure 8 it is demonstrated the outer model of the NM-PLS model. By assessing loadings of each attribute individually, it is possible to observe that parameters that were more related with their own latent variable were: sweet and intense apple odor, which were the parameters that caused significant mean drops on the penalty analysis, followed by crispy for what concerns the "texture" group. This means those parameters had the highest weight in the formation of their latent variable, and therefore, are the ones indirectly affecting most the liking scores of samples. On the other hand, the attributes green apple notes, aromatic and ripe apple notes presented loadings lower than 0.70. This means that variability in those parameters is not sufficiently capture by the latent variable "notes" (Sanchez, 2013) and therefore, probably only the parameter honey notes is actually contributing to the higher path coefficient displayed by the "notes" group. In other words, when considered together, the attributes intense apple odor and honey notes displayed a higher impact in the liking scores.

FIGURE 7 Non-metric partial least square inner model: the path coefficients of relationships between each independent latent variable ("texture," "taste" and "notes") and the dependent variable "liking"

4 | CONCLUSIONS

In this work an integrated methodology proposed to assess the quality of Gala apples cultivated in at two different altitude was successfully

FIGURE 8 Loadings between texture, taste, notes and liking and their relative parameters of the non-metric partial least square model

developed. Few differences were detected among samples, mainly concerning color and texture attributes. Most of the results at the three levels of the assessment were in agreement, such as the similarities among samples concerning the taste perception, which was confirmed by similar levels of TSS and TA, other than the lack of significant differences concerning the attributes scores of sweetness and sourness in the panel test assessment and the CATA frequencies of sweet and sour attributes at the consumer level. Interestingly, slight differences concerning the texture properties and perception were detected among the three levels of the assessment. The panel test pointed Area 2 samples as harder, crispier and more astringent, while the consumer test suggested the Area 2 samples were less mealy but not harder (which was in agreement with the TPA test) and a higher astringency of Area 2 samples was not detected. These results confirm how important is to implement an integrated assessment, where differences that are perceived by panelists but not by consumers are detected and valuable information from samples characterization is not neglected. Finally, a NM-PLS regression model applied on penalty analysis results with a new system of codification is proposed by taking into account penalty results. It successfully confirmed sweet and intense apple odor as the attributes with the highest impact on liking scores and its mean drop, however, the assessment also allowed to obtain valuable information on the liking impact of global and complex quality attributes, such as "notes," "taste" and "texture." The integration of attributes showed a higher impact of "notes" on liking compared to "texture" and "taste" and these results are not in agreement with other works where texture is claimed to be the most important driver on consumer liking of apple. However, the methodology was applied with a very small number of consumers and therefore, the consumer test results still need to be validated on broader population sample. At the same way, the average drop on force calculated by the compression test need to be tested on a higher number of samples and on different horticultural crops in order to better investigate its relation with the crispness perception.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Thais Mendes da Silva https://orcid.org/0000-0003-4382-9292

REFERENCES

Almeida, D. P. F., & Gomes, M. H. (2017). Physicochemical quality, macronutrients, and dietary fiber in apples from the protected geographical indication 'Maçã de Alcobaça', Portugal. *European Journal of Horticultural Science*, 82(5), 239–243. https://doi.org/10.17660/eJHS.2017/82.5.3

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

Bassi, G., & Fideghelli, C. (1995). La scelta delle cultivar di melo adatte alla pianura Si sta delineando con maggior chiarezza una diversità. In *Melicoltura Di Pianura: Quale Futuro* (pp. 67–77). Verona, Italy: Informatore Agrario.

Blaker, K. M., Plotto, A., Baldwin, E. A., & Olmstead, J. W. (2014). Correlation between sensory and instrumental measurements of standard and crisp-texture southern highbush blueberries (*Vaccinium corymbosum* L. interspecific hybrids). *Journal of the Science of Food and Agriculture*, 94(13), 2785–2793. https://doi.org/10.1002/jsfa.6626

Bosco, L. C., Bergamaschi, H., Cardoso, L. S., de Paula, V. A., Marodin, G. A. B., & Nachtigall, G. R. (2015). Apple production and quality when cultivated under anti-hail cover in Southern Brazil. *International Journal of Biometeorology*, 59(7), 773–782. https://doi.org/10.1007/s00484-014-0893-6

Campbell, G. (2016). Bubbles in Food 2: Novelty, Health and Luxury (Elsevier, Ed.). Retrieved from https://books.google.it/books?id=XG1bDAAAQBAJ

Charles, M., Corollaro, M. L., Manfrini, L., Endrizzi, I., Aprea, E., Zanella, A., ... Gasperi, F. (2018). Application of a sensory–instrumental tool to study apple texture characteristics shaped by altitude and time of harvest. *Journal of the Science of Food and Agriculture*, 98(3), 1095–1104. https://doi.org/10.1002/jsfa.8560

Chen, L., & Opara, U. L. (2013). Texture measurement approaches in fresh and processed foods—A review. *Food Research International*, 51(2), 823–835. https://doi.org/10.1016/j.foodres.2013.01.046

Contador, L., Shinya, P., & Infante, R. (2015). Texture phenotyping in fresh fleshy fruit. *Scientia Horticulturae*, 193, 40–46. https://doi.org/10.1016/j.scienta.2015.06.025

Corollaro, M. L., Aprea, E., Endrizzi, I., Betta, E., Demattè, M. L., Charles, M., ... Gasperi, F. (2014). A combined sensory-instrumental tool for apple quality evaluation. *Postharvest Biology and Technology*, 96, 135–144. https://doi.org/10.1016/j.postharvbio.2014.05.016

De Langhe, B., Puntoni, S., & Larrick, R. (2017). Linear thinking in a nonlinear world. *Harvard Business Review*, 95(3), 130–139.

Estiaga, C. (2015). Bootstrapping penalty analysis in sensory evaluation of pizza products. *Philippine Statistician*, 64(2), 69–94.

Giuggioli, N. R., Peano, C., & Mendes da Silva, T. (2020). Multivariate factor analysis (MFA) approach to evaluate the quality of stored cavendish banana sourced from different geographical areas. *Emirates Journal of Food and Agriculture*, 32(3), 204–212. https://doi.org/10.9755/ejfa.2020.v32.i3.2078

Gregori, R., Bergamaschi, M., Berra, L., & Folini, L. (2016). La qualità delle mele in diversi ambienti: pianura, collina e montagna L'. *Frutticoltura*, 5, 64–68.

Hampson, C. R., & Quamme, H. A. (2000). Use of preference testing to identify tolerance limits for fruit visual attributes in apple breeding. *HortScience*, 35(5), 921–924. https://doi.org/10.21273/HORTSCI.35.5.921

Harker, F. R., Marsh, K. B., Young, H., Murray, S. H., Gunson, F. A., & Walker, S. B. (2002). Sensory interpretation of instrumental measurements 2: Sweet and acid taste of apple fruit. *Postharvest Biology and Technology*, 24(3), 241–250. https://doi.org/10.1016/S0925-5214(01)00157-0

Husson, F., Le, S., & Cadoret, M. (2020). SensoMineR: Sensory data analysis. R package version 1.25. Retrieved fom. https://CRAN.R-project.org/package=SensoMineR

ISO. (2003). *Sensory analysis. Guidelines for the use of quantitative response scales*. Geneva, Switzerland: Author.

ISO (2012). Sensory analysis. In *General guidance for selection, training and monitoring of assessors and expert sensory assessors*. Geneva, Switzerland: Author.

Lado, J., Moltini, A. I., Vicente, E., Rodríguez, G., Arcia, P., Rodríguez, M., … Ares, G. (2019). Integration of sensory analysis into plant breeding: A review. *Agrociencia Uruguay*, *23*(1), 101–115. https://doi.org/10.31285/agro.23.1.16

Li, Q. (2013). A novel Likert scale based on fuzzy sets theory. *Expert Systems with Applications*, *40*(5), 1609–1618. https://doi.org/10.1016/j.eswa.2012.09.015

Mendes da Silva, T., Briano, R., Peano, C., & Giuggioli, N. R. (2020). The use of a new explanatory methodology to assess maturity and ripening indices for kiwiberry (*Actinidia arguta*): Preliminary results. *Postharvest Biology and Technology*, *163*, 111–122. https://doi.org/10.1016/j.postharvbio.2020.111122

Mendes da Silva, T., Marinoni, D. T., Peano, C., & Giuggioli, N. R. (2019). A new sensory approach combined with a text-mining tool to create a sensory lexicon and profile of monovarietal apple juices. *Food*, *8*(12), 608. https://doi.org/10.3390/foods8120608

Mendes da Silva, T., Peano, C., & Giuggioli, N. R. (2019). A novel statistical approach to assess the quality and commercial viability of a retail branded perishable fruit. *CyTA - Journal of Food*, *17*(1), 581–592. https://doi.org/10.1080/19476337.2019.1621389

Molina, D., Alegre, S., Casero, T., Calsals, M., Bonany, J., Carbó, J., … Recasens, I. (2006). Quality indexes for 'Golden Smoothee' apples in relation to consumer evaluation. *Journal of Fruit and Ornamental Plant Research*, *14*, 39–51.

Nappo, D. (2009). *SEM with ordinal manifest variables—An alternating least squares approach*. Napoli-Italy: Università degli Studi di Napoli Federico II.

Neri, D. (2004). Low-input apple production in Central Italy: Tree and soil management. *Journal of Fruit and Ornamental Plant Research*, *12* (Special ed.: Orchard management in sustainable fruit production), 69–76.

Pagès, J., Berthelo, S., Brossier, M., & Gourret, D. (2014). Statistical penalty analysis. *Food Quality and Preference*, *32*, 16–23. https://doi.org/10.1016/j.foodqual.2013.07.008

Pathare, P. B., Opara, U. L., & Al-Said, F. A.-J. (2013). Colour measurement and analysis in fresh and processed foods: A review. *Food and Bioprocess Technology*, *6*(1), 36–60. https://doi.org/10.1007/s11947-012-0867-9

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

Plotto, A. (1998). *Instrumental and sensory analysis of "Gala" apple (Malus domestica, BORKH) aroma*. Oregon (USA): Oregon State University.

Rivière, P., Monrozier, R., Rogeaux, M., Pagès, J., & Saporta, G. (2006). Adaptive preference target: Contribution of Kano's model of satisfaction for an optimized preference analysis using a sequential consumer test. *Food Quality and Preference*, *17*(7–8), 572–581. https://doi.org/10.1016/j.foodqual.2006.04.002

Robertson, J. A., Meredith, F. I., Senter, S. D., Okie, W. R., & Norton, J. D. (1992). Physical, chemical and sensory characteristics of Japanese-type plums grown in Georgia and Alabama. *Journal of the Science of Food and Agriculture*, *60*(3), 339–347. https://doi.org/10.1002/jsfa.2740600311

Russolillo, G. (2010). The non-metric partial least squares approach. 42èmes Journées de Statistique, Marseille, France, France. inria-00494817

Sanchez, G. (2013). PLS path modeling with R. Retrieved from www.gastonsanchez.com/PLS_Path_Modeling_with_R.pdf

Sanchez G., Trinchera, L., & Russolillo, G. (2017). plspm: Tools for partial least squares path modeling (PLS-PM). R package version 0.4.9. Retrieved from. https://CRAN.R-project.org/package=plspm

Sharma, G., Wu, W., & Dalal, E. N. (2005). The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Research and Application*, *30*(1), 21–30. https://doi.org/10.1002/col.20070

Stanley, J., Prakash, R., Marshall, R., & Schröder, R. (2013). Effect of harvest maturity and cold storage on correlations between fruit properties during ripening of apricot (*Prunus armeniaca*). *Postharvest Biology and Technology*, *82*, 39–50.

ten Kleij, F., & Musters, P. A. D. (2003). Text analysis of open-ended survey responses: A complementary method to preference mapping. *Food Quality and Preference*, *14*(1), 43–52. https://doi.org/10.1016/S0950-3293(02)00011-3

Tunick, M. H., Onwulata, C. I., Thomas, A. E., Phillips, J. G., Mukhopadhyay, S., Sheen, S., … Cooke, P. H. (2013). Critical evaluation of crispy and crunchy textures: A review. *International Journal of Food Properties*, *16*(5), 949–963. https://doi.org/10.1080/10942912.2011.573116

Xiong, R., & Meullenet, J. F. (2006). A PLS dummy variable approach to assess the impact of jar attributes on liking. *Food Quality and Preference*, *17*(3–4), 188–198. https://doi.org/10.1016/j.foodqual.2005.03.006

Zamzami, L., & Ariyawardana, A. (2019). Consumer insights that drive value creation opportunities in the Indonesian domestic apple supply chain. *Journal of Agriculture and Rural Development in the Tropics and Subtropics*, *120*(1), 33–44. https://doi.org/10.17170/kobra-20190613554