Comparison of Pivot Profile© to Frequency of Attribute Citation: analysis of complex products with trained assessors.

J. Brand[a,b,*], D. Valentin[c], M. Kidd[d], M. A. Vivier[b], T. Næs[e], H.H. Nieuwoudt[b]

[a]*Department of Viticulture and Oenology, Stellenbosch University, Private Bag X1, Matieland (Stellenbosch) 7602, South Africa*

[b]*Institute for Wine Biotechnology, Department of Viticulture and Oenology, Stellenbosch University, Private Bag X1, Matieland (Stellenbosch) 7602, South Africa*

[c]*Centre des Sciences du Goût et de l'Alimentation, AgroSup Dijon, CNRS, INRA, Univ. Bourgogne, Franche-Comté, F-2100, Dijon, France*

[d]*Centre for Statistical Consultation, Department of Statistics and Actuarial Sciences, Stellenbosch University, Private Bag X1, Matieland (Stellenbosch) 7602, South Africa*

[e]*Nofima ÅS, Raw material and Process Optimization, Oslovegen 1, P.O. Box 210, N-1431, Ås, Norway*

[*]E-mail address: jeanne@sun.ac.za

Abstract

Pivot© profile (PP), a method which compares samples to a reference (pivot), has shown profiling potential for complex matrices. However, various aspects require further investigation. This study's aim was to compare PP to frequency of attribute citation (FC) considering individual judges' data and sample set complexity. A trained panel analysed three wine sets with different within-set product similarity levels. The stability of the PP sensory space was tested by changing the pivot. PP and FC results were compared using RV coefficients. Confidence ellipses on correspondence analysis (CA) plots were constructed to consider individual judges' data. CA plots constructed from different pivot PP data sets, were less similar to each other, than to CA plots of FC data, for the set with medium and the set with high within-set variation. The most profound differences were observed for the set with the high within-set variation. PP configurations of the set with low within-set variation, were more similar to each other than to FC configurations. Higher explained variance was obtained with PP than FC, but confidence ellipses overlapped more frequently indicating fewer significant differences between samples. PP and FC data were

comparable for the set with medium within-set variation. From this study's results PP is recommended for wine profiling if medium within-set variation between samples exist but not when sample sets with low or high within-set variation are profiled. PP is recommended over FC for comparative studies where a reference sample is required for example during benchmarking or for aging and shelf-life studies.

Keywords: Pivot profile, frequency of attribute citation, CATA, trained panel, correspondence analysis

1. Introduction

Describing the intrinsic properties of food products to obtain sensory profiles is a primary need within the food industry. It plays an important role during product development, production, quality control, advertising and marketing. Due to increased pressure from the food and beverage industry to profile products faster, new sensory methods and optimised statistical tools are continuously being developed. These include rapid sensory methods whereby product experts or naïve consumers can do the evaluation without training (Valentin et al., 2012; Varela & Ares, 2012).

One of the recent additions to rapid sensory methods is Pivot Profile© (PP), a frequency-based method proposed by Thuillier et al. (2015). When PP is performed, each sample is compared to a reference sample, also referred to as the pivot. Sensory judges are required to list those attributes that they perceive as, respectively, less or more intense in the sample than in the pivot. PP, therefore, provides an estimation of the intensity of attributes in the samples relative to the pivot. Check-all-that-apply (CATA), (Adams et al., 2007; Lancaster & Foley, 2007) can also provide an estimate of attribute intensities through the assumption that those attributes mentioned by more judges are more intense than those mentioned by fewer judges (Campo et al., 2010). PP could, therefore, potentially be more suitable than CATA for benchmarking applications of complex matrices such as wine (Thuillier et al., 2015) since; (1) relative intensity is captured during the tasting, while with CATA an assumption is made about intensity, and (2) PP involves direct sample comparison and CATA monadic presentation.

Several studies showed that PP is a valuable asset in the rapid sensory method toolbox. Thuillier et al. (2015) profiled champagne, using product experts as sensory judges when the method was introduced. Subsequent research on a set of beer samples showed that the choice of the pivot did not have a significant effect on the product positioning in correspondence analysis (CA) plots (Lelièvre-Desmas et al., 2017). In the field of dairy research, Fonseca et al. (2016) compared PP to comment analysis (Symoneaux et al., 2012) and demonstrated that consumers could profile chocolate ice cream products efficiently with both methods. PP was compared to CATA and projective mapping (PM) (Risvik et al., 1994) in a study on Greek yoghurt samples (Esmerino et al., 2017). The results showed that PP, CATA and PM provided similar results of sufficient quality. Recently, Deneulin et al. (2018) used PP to profile a large number of honey samples from all over the world.

As with all new methods, further studies are needed to investigate and understand the appropriate use and performance of PP when applied to different products. Aspects identified in earlier studies are related to possible effects of the choice of the pivot on the stability of the sensory space (Thuillier et al., 2015) and the performance of the method when applied to sample sets with various degrees of within-set similarity (Lelièvre-Desmas et al., 2017). Lelièvre-Desmas et al. (2017) reported that within-set similarity had a more pronounced impact on the results than the choice of the pivot. However, in that study, the between-sample discrimination power of PP, which is important for benchmarking of wine, was not studied.

Yet another aspect that requires further investigation is the measurement of panel performance. In the studies by Deneulin et al. (2018) and Fonseca et al. (2016), panel performance was not measured. Deneulin et al. (2018) concluded that the vocabulary used required more attention and that calculating panel repeatability and consensus could shed light on these matters. Since Fonseca et al. (2016) used consumers as sensory judges, repeatability could not be measured. However, investigating segmentation could be interesting and could contribute to understanding the sensitivity of PP as a sensory method.

Thuillier et al. (2015) suggested that descriptive analysis (DA) might be more suitable than PP if the objective is to obtain a detailed description of products. In terms of comparing PP to other methods, no study has been conducted to test PP against traditional sensory methods that involve training of a panel to profile complex products such as beer and wine. DA has the limitation that, when assessing complex matrices, sensory judges could experience difficulty in differentiating between different odours by using a line scale (Lawless, 1999).

Frequency of attribute citation (FC) is a method that does not entail rating on a line scale (Campo et al. 2008). FC refers to a profiling method whereby sensory judges are trained using a pre-determined list of attributes and reference standards. Judges are required to select attributes from the list to describe the products under evaluation. FC is an adapted CATA procedure with specific changes and restrictions where: (1) the list contains only sensory attributes: no phrases emotional or hedonic terms are allowed; (2) the sensory attributes are organised into categories such as odour or aroma families; (3) judges are trained with reference standards to use the CATA list; (4) judges can reorganise the CATA list during training through panel consensus; and (5) panel repeatability is measured to ensure quality data. FC was used to analyse wine (Campo et al., 2008) and was compared to DA in a later study in which similar results were obtained with DA and FC (Campo et al., 2010).

The aim of this study was to gain a better understanding of the appropriate application of PP when applied to wine profiling taking sample set complexity, defined as within-set variation, into account. A trained panel was used in this study for both PP and FC to eliminate the panel effect when comparing the two methods and to limit heterogeneity through training. FC, as opposed to DA, was used as reference method, to minimise difficulty experienced by judges in differentiating between odours, particularly experienced when rating intensities on a line scale (Lawless, 1999). Furthermore, comparing continuous

DA data obtained from using a line scale to the categorical data obtained from PP might add extra variation.

Three objectives were formulated: (1) to evaluate the ability of PP to discriminate between different wines using confidence ellipses calculated by bootstrapping; (2) to test the robustness of PP by changing both the pivot sample and the sensory complexity, referred to in this paper as within-set variation; and (3) to compare panel performance for PP and FC in terms of repeatability, consensus and the perceived difficulty of the task. Three sets of wines, one red and two white cultivars, of varying within-set variation, were designed for the investigation.

2. Materials and methods

2.1 Samples

The wines used in this study were selected based on the knowledge acquired in previous research on similar wines (Bester, 2011; Hanekom, 2012; Van Antwerpen, 2012), the knowledge of expert tasters, wine industry professionals and sensory professionals. The following three sets (six wines each) with different within-set sensory variation were subjected to sensory analysis: (1) wooded Pinotage wines with similar characteristics; (2) wooded Chenin Blanc wines of medium within-set variation; and (3) Sauvignon Blanc wines with extreme style differences. For this study, wines were selected in such a way that specific cultivars represented sets with different levels of within-set variation. It is important to note that cultivar *per se* cannot be used as an indication of complexity.

The wines from the set with low within-set variation (Pinotage) had “oaky”, “red berry”, “blackberry”, “spicy”, “caramel” and “dried fruit” notes amongst other. The Chenin Blanc wines, with medium within-set variation, had “citrus”, “tropical fruit”, “yellow apple”, “dried fruit”, “honey”, “caramel” and “woody” aromas. “Tropical” aromas including “guava”, “passion fruit” and “pineapple”, “green” aromas including “green pepper”, “asparagus” and “tomato leaf” as well as “mineral”, “flinty” and “oaky” nuances were used to describe the set with high within-set variation (Sauvignon Blanc wines).

Each set was analysed by FC and PP using the same sensory methodology and workflow, resulting in six separate data sets. Three PP experiments were conducted for each set using different pivot samples, P1, P2 and P3. P1 and P2 were selected to show high sensory characteristics, as described below. P3 was a blend of equal volumes of all the samples in a cultivar set. The assumption was made that P3 of each set was “the average” sample (Thuillier et al., 2015); representative of the set and having no extreme sensory characteristics.

For the set with low within-set variation (Pinotage), P1 was chosen as a predominantly “fruity” sample with “red berries” and “black berries” as the main aroma contributors. P2 had prominent “oaky”, “caramel” and “vanilla” notes.

The dominating aromas characteristics of P1 selected for the set with medium within-set variation (Chenin Blanc), were “fresh green”, “grapefruit” and “citrus”. P2 was characterised by intense “oaky”, “vanilla” and “caramel” aromas, with subtle notes of “dried fruit”, “marmalade” and “honey”.

For the set with high within-set variation (Sauvignon Blanc), P1 was characterised by dominant “mineral” with subtle “tropical” and “green” notes. P2 was predominantly “oaky” with “fruity” attributes.

All wines were commercially available, produced in South Africa and certified by the South African Wine and Spirits Board (Table 1).

/Insert TABLE 1/

2.2 Panel

The panel of sensory judges consisted of three males and 12 females between 24 and 65 years of age (average age: 32). All judges were trained sensory assessors with more than two years of experience in wine sensory analysis and were paid for their participation. The same panel participated in the PP and FC experiments.

2.3 Sensory Methodology

2.3.1 FC and PP methodology

2.3.1.1 Training. Panel training consisted of 15 sessions of one hour each over six weeks. Ballot training on 134 wine aroma attributes using reference standards (Table 2) was conducted according to the frequency of attribute citation training procedure (Campo et al., 2008 and Campo et al., 2010). The list of terms given to the panel was subdivided into aroma categories according to literature (Noble et al., 1987; Campo et al., 2010; Bester, 2011; Hanekom, 2012; Van Antwerpen, 2012). During each training session, judges were presented with 10 to 15 aroma standards to familiarise themselves with the terms on the list (ballot). Two to three wines were presented per session. Attributes used by the panel to describe the wines were discussed and the most frequently cited attributes were highlighted by the panel leader.

The training consisted of two phases; a general phase in which the judges were trained on the initial list of terms, followed by a specific training phase where judges were trained to profile wines similar to those presented during the evaluation. During the specific training, judges could add terms to the initial list and change their categorisation in the separate aroma families to describe the wines accurately. The final aroma attribute list with aroma standards is shown in Table 2 and consisted of 103 attributes. Two specific training sessions, discussing wines from the relevant cultivar and vintages, were performed per cultivar sample set. For this study, judges were trained since detailed descriptions of the wines were required, and panel heterogeneity had to be limited. However, PP could also be performed by industry professional or consumers without training the sensory judges if less detailed profiles are required.

/Insert TABLE 2/

Procedures. Judges had to provide three to five terms from the list to describe the most prominent aromas of each wine. Campo et al. (2010) suggested that the required number of attributes that each judge should use to describe products should be specified with FC to avoid the use of too few or too many descriptors. People have a limited capacity to discriminate between and describe odours in complex samples and using too few descriptors can lead to incomplete descriptions of samples (Laing & Glemarec, 1992). On the other hand, when large numbers of attributes, including many synonyms, are used to describe wines, noise could be added to the data, complicating and adding biases during the statistical analysis of the data.

During PP sessions, judges were asked to write down the attributes that they perceived “less intense” and “more intense” in the sample than the pivot from the list of attributes (Fig. 1). The same list as provided for FC was used. Judges were limited in terms of the number of attributes that they could use during PP to achieve a degree of standardisation between the instructions for PP and FC. No more than the five most prominent attributes per sample were allowed to describe the aromas that they perceived “less intense” in the sample than the pivot. The same rule applied to the attributes perceived “more intense” than the pivot. Finally, judges had to provide at least three attributes in total per sample.

The final task of the sensory evaluation session was to rate the difficulty of performing the sensory methods. Judges were asked to give a score out of 9 on an easiness scale that was derived from the nine-point hedonic liking scale (Peryam & Pilgrim, 1957). The specific words used were: ”extremely easy (1); very easy (2); moderately easy (3); slightly easy (4); neither easy nor difficult (5); slightly difficult (6); moderately difficult (7); very difficult (8); and extremely difficult (9)”.

To minimise panel learning effects, and matrix change due to wine aging, several measures were taken and followed for all three sample sets. Sensory evaluation sessions of a specific set of wines and one pivot, for example P1, were conducted in duplicate by 15 assessors on the same day. The panel did not receive information on the nature of the wines in terms of style, vintage or cultivar and did not know that they evaluated the same wines twice. The same cultivar set with P2 as pivot was only evaluated two to three weeks later. The order in which evaluations, PP-P1, PP-P2, PP-P3 and PP-FC, were performed was randomised within the different sets. The entire set PP-P1, PP-P2, PP-P3 and PP-FC, for example, all the Chenin Blanc evaluations, were done within two and a half months, to ensure that wine ageing did not change sensory characteristics. Since the latter aspect is of particular importance for the white wines, the sets were analysed consecutively. The set with medium within-set variation was analysed first, the set with high within-set variation second and the set with low within-set variation last. The sets were, therefore, not analysed from the lowest to highest, or from highest to lowest within-set variation.

2.3.1.3 Wine evaluation. Wines were evaluated in a well-ventilated, temperature controlled (20 ± 2^{o}C), odour free sensory lab secluded from extraneous noise. The laboratory was equipped with separate off-white individual tasting booths with controlled lighting conditions.

Black (ISO NORM 3591, 1977) tasting glasses labelled with random 3-digit codes were used. Samples were randomised across judges according to a Williams Latin-square design (MacFie et al., 1988). Monadic sample presentation was applied for FC. For PP, samples were presented in pairs. Each pair consisted of a sample and a fresh pivot. Each glass contained 25 mL of wine and was covered with a Petri-dish lid. Wines were poured 20 to 30 minutes before the sensory evaluation session to allow volatile compounds to reach equilibrium in the headspace of the glasses.

Wines were evaluated orthonasally in duplicate for both methods. Duplicates were evaluated on the same day with an enforced 10-minute break in between to limit sensory fatigue. Data were collected using Compusense cloud software (www.compusense.com, Compusense).

2.4 Data analysis

2.4.1 Panel performance

Repeatability. Panel repeatability was calculated for the individual judges using the reproducibility index (R_i) proposed by Campo et al. (2008). Two times the number of common descriptors used in the first and second repeat was divided by the total number of descriptors used in both repeats. This ratio was calculated for every wine and summed over all the wines tasted by one judge to calculate the R_i value for that judge. In addition, a global reproducibility index (R_i) was calculated by computing the average across all judges' R_i values. This measure ranges from 0 to 1. If all the attributes cited during the first and second repeat are the same, then the R_i value will be 1. If entirely different attributes were cited, then the R_i value will be 0. A minimum R_i of 0.2 was proposed by Campo et al. (2008) to deem a sensory judge repeatable enough to record the response as data.

$$R_i = \frac{1}{n}\sum \frac{2 \times des_{com}}{(des_{rep1} + des_{rep2})}$$

Where: $n = number\ of\ wines$

$des_{com} = number\ of\ identical\ descriptors\ chosen\ by\ the\ judge\ in\ both\ replicates$

$des_{rep1} = number\ of\ descriptors\ chosen\ by\ the\ judge\ in\ replicate\ 1$

$des_{rep2} = number\ of\ descriptors\ chosen\ by\ the\ judge\ in\ replicate\ 2$

R_i values were calculated for the FC and PP methods for all the data sets. For PP data the following rule was applied: if a descriptor was cited as "more intense" in one repeat and "less intense" in the other repeat it was not counted as an identical descriptor occurring in both repeats and that descriptor did not contribute to the R_i value. Each PP set obtained from using a different pivot sample was treated as a separate data set.

A three-way mixed model ANOVA with cultivar, method and the cultivar*method interaction as fixed factors and sensory judges as random factors was computed. The ANOVA was used to study the differences between repeatability of the panel in terms of Ri values computed when (1) sample sets with different within-set variation was evaluated and (2) different sensory methods (PP and FC) and pivot samples were used. Sample sets from different cultivars represented sets with different within-set variation, as explained before. Pinotage represented low, Chenin Blanc medium and Sauvignon Blanc large within-set sample variation. The methods used were FC and PP using different pivot samples, P1, P2 and P3. The REML estimation method was used. When significant ANOVA results were found, pairwise comparisons were calculated using the Fisher's LSD *post hoc* test with α set at 5%.

Consensus. Panel consensus was measured calculating Cohen's kappa coefficients for each pair of judges. Cohen's kappa coefficient is a measure of the similarity or agreement between the ratings provided by two individuals. It is commonly used on nominal data as an interrater reliability measure in the field of medical and educational surveying (Cohen, 1960; Altman 1991; McHugh, 2012; Gisev et al., 2013). In this study, Cohen's kappa coefficients (Κ) were calculated using the mathematical equation below:

$$\kappa = \frac{p_0 - p_e}{1 - p_e}$$

Where:

$p_0 = the\ relative\ observed\ agreement\ among\ raters\ (sensory\ judges\ in\ this\ case)$

$p_e = the\ hypothetical\ probability\ of\ chance\ agreement$

In addition, the average panel consensus was calculated for each data set by computing the average of all the Cohen's kappa coefficients across all the judges. Individual data obtained from PP were handled by means of the following rule: if a descriptor was cited as "more intense" by one sensory judge and "less intense" by another the agreement among those two judges for that descriptor was noted as zero as if two different descriptors were used. Each PP sample set obtained from using a different sample as pivot was treated as a separate data set. A three-way mixed model ANOVA similar to the ANOVA computed on the R_i values was computed on the Cohen's kappa coefficients.

Difficulty of the sensory task. A three-way mixed model ANOVA, similar to the ANOVA's applied to assess panel consensus and repeatability, was performed to investigate significant differences between the perceived difficulty of the different FC and PP data sets.

2.4.2 Product characterisation

The descriptors generated to describe each group of wines in the verbalisation phase were captured by constructing a contingency table. The number of attributes used was reduced before statistical analysis. Attributes cited by less than 20% of the panel were combined with similar terms or discarded. Three sensory experts combined similar terms independently by employing lemmatisation and semantic categorisation. Attributes combined differently by the sensory experts were discussed and consensus was reached before the final attribute reduction step. Fig. 1a shows the scheme used for data organisation and analysis.

Correspondence analysis (CA) with confidence ellipses, calculated using bootstrapping (Cadoret et al., 2013; Dehlholm et al., 2012), was performed on the contingency tables and used to visualise the sensory space spanned by the different wines within a data set.

Contingency tables were constructed from FC and PP data in different ways. For FC data, the total number of citations over all the judges for each descriptor per wine was tabulated with the attributes as variables in the columns and the wines as objects in the rows. The number of judges who cited an attribute for a specific wine was tabulated at the intersection of the corresponding column (representing the attribute) and row (representing the wine). This procedure is the same as for standard CATA (Valentin et al., 2012).

PP data sets were compiled by subtracting the citation frequency of “less” from “more” for each attribute for each wine. The pivot sample was added as centre point by including zeros for all the descriptors for the pivot wine. This procedure was followed when P1 and P2 was used as pivot. When P3, the blend, was used as pivot sample this procedure was not followed. The absolute value of the minimum was added to all the values as a translation step. This procedure produced both positive and negative values. Since CA cannot be conducted on a table containing negative values, translation had to be performed to obtain a contingency table consisting of positive values. Through this procedure the relative intensity of the pivot (P1 or P2) relative to the other samples was determined during translation of the data and was reflected in the contingency table on which CA was performed. Consequently, CA plots obtained for P1, P2 and P3 were comparable containing the same samples, which included P1 and P2 but not P3. This procedure is described in detail by Thuillier et al. (2015) and summarised in Fig.1. In order to apply bootstrapping on the PP data, the contingency table was converted into an appropriate data set for CA by repeating each combination of wine and descriptor n_{ij} times where n_{ij} is the frequency of the *i*-th wine and the *j*-th descriptor in the contingency table.

/Insert Fig. 1/

2.4.3 Comparison of methods and testing the stability of the sensory space for PP

The similarities between multivariate plots were assessed by calculating RV coefficients on the first two dimensions. RV coefficients are used to measure the similarity between two matrices or data sets by measuring the amount of variance shared (Robert & Escouffier, 1976; Abdi et al., 2013; Abdi et al., 2014). CA plots generated from PP data sets where different samples were used as the pivot were compared to each other and to the CA plot constructed from FC data (Fig. 1b). This procedure was followed for the set with the low within-set variation (Pinotage), the set with medium within-set variation (Chenin Blanc) and the set with large within-set variation (Sauvignon Blanc) separately. In addition, the repeatability, panel consensus and difficulty perceived by the panellists when performing PP and FC were compared using ANOVA, as described above.

All data organisation and analyses were conducted using Microsoft Excel 2016 (www.microsoft.com, Microsoft), XLSTAT (www.XLSTAT.com, Addinsoft SARL.), Statistica 13 (www.statsoft.com, Statsoft Inc.) and R version 3.4.0, packages “car” and “cabootcrs” (www.R-project.org).

3. Results

3.1 Panel performance

The individual R_i values for all the sensory judges were above 0.2 for both FC and PP, irrespective of which samples were used as the pivot. The highest R_i value was 0.86 and the lowest 0.26. All the judges produced repeatable results, considering that R_i values can range from 0 to 1, and Campo et al. (2008) proposed 0.2 as the lowest acceptable value.

It is clear from the three-way mixed model ANOVA results (Fig. 2a) performed on panel repeatability, with method and cultivar (representing different levels of within-set variation) as fixed factors, that the method*cultivar effect was significant ($p < 0.001$). Therefore, the method*cultivar interaction effect was interpreted using Fisher’s LSD *post hoc* test since the same trend could not be seen for all cultivars or sample sets. Thus, the panel repeatability was influenced by the complexity of the data set analysed. Sensory judges were less repeatable when conducting FC than PP for the data set with medium within-set variation (Chenin Blanc wines). A significant difference between FC and PP with P2 and P3 was seen. In addition, judges were less repeatable when P1 was used than when P2 was used. No significant difference in repeatability was seen when P1 and P3 (the blend of all the samples) and P2 and P3 were used. A significant difference between using P2 and P1 as pivot sample could be seen for the data set with high within-set variation (Sauvignon Blanc wines). In addition, no significant differences between PP when changing the pivot or between PP and FC was observed for the data sets with low within-set variation (Pinotage wines).

In summary, the average panel repeatability was the lowest for the Pinotage wines, which had the least within-set variation and differed significantly from the Sauvignon Blanc wines, (which had high within-set variation).

/Insert Fig. 2/

Panel consensus, measured by Cohen's kappa coefficients, ranged from 0.02 to 0.55. Values below 0.2 are considered poor, 0.4 fair and between 0.4 and 0.6 moderate (Altman, 1991). As with the panel repeatability, the method*cultivar effect was significant with $p < 0.001$. Therefore, the method*cultivar interaction effect's Fisher's LSD *post hoc* test was interpreted since the same trend could not be seen for all cultivar sample sets for all the methods in terms of significant differences between panel consensus.

The ANOVA results (Fig. 2b) clearly show that different trends were observed for the sample sets with different within-sample variation in terms of average panel consensus. The panel consensus for the set with the low (Pinotage) and the set with medium (Chenin Blanc) within-set variation was poor with the average Cohen's kappa coefficient of the panel below 0.2. Interpreting significant differences with such low values would be unwise.

It is interesting to note that the only data set with acceptable average panel consensus coefficients, above 0.2, was the set with high within-set variation (Sauvignon Blanc). Cohen's kappa coefficients above 0.2 were observed for FC and PP except when the blend of the samples was used as a pivot for which a significantly lower value of 0.17 was observed. The best consensus was achieved when P1 was used and was significantly higher than when FC was performed and when other pivot samples were used.

For easiness/difficulty of the task, as with the panel repeatability and consensus, the method*cultivar effect was significant with $p < 0.001$. Therefore, the method*cultivar interaction effect's Fisher's LSD *post hoc* test was interpreted since the same trend could not be seen for all cultivars for all the methods in terms of significant differences in the difficulty of the task. The sensory judges experienced PP as significantly more difficult to perform when compared to FC, irrespective of the within-set variation of the data set and the pivot sample used (Fig. 2c).

Product description and comparison of methods

The RV coefficients calculated between the PP CA configurations when the pivot sample was changed for the set with the lowest within-set variation (Pinotage wines) ranged from 0.52 to 0.83 (Table 3). Since all the RV coefficients were above 0.5, the configurations could be regarded as similar (Louw et al., 2013). However, the similarity between the FC configuration and PP configurations, corresponding to P1 (Fig. 3a) and P2 (Fig. 3b) as pivot samples, indicated low similarity with RV coefficients below 0.35 (Table 3). When a blend of all the samples was used as pivot sample, namely P3 (Fig. 3c), better similarity was observed with an RV coefficient of 0.60.

/Insert TABLE 3/

Furthermore, overlapping confidence ellipses indicated that no significant difference between samples could be observed when PP was conducted on this sample set although the explained variance for the first two factors was well above 60%. The cumulative explained variance for the first two factors was 68% when P1 (Fig. 3a), 75.7% when P2 (Fig. 3b), 69% when P3 (Fig. 3c) and 68.2 when FC (Fig. 3d) was used. Confidence ellipses on the CA plot of the FC configuration indicated that two of the samples were perceived as significantly different from the other four samples (Fig. 3d). It is interesting to note that the cumulative explained variance of factor one and two of the CA plot of PP when P2 was used as pivot sample was higher for PP (Fig. 3b) than for FC (Fig. 3d). This was, however, not the case when P1 and P3 were used as pivot samples.

Descriptors belonging to the same aroma families appeared more scattered on the CA plot and showed less positive correlation with each other for PP data than FC data. The most obvious and prominent cases occurred when extreme samples, P1 and P2, were used as pivot samples (Fig. 3a and b). When the blend P3 (Fig. 3c) was used as pivot, aroma attributes belonging to the same aroma family grouped well together indicating acceptable positive correlation. Examples were: (1) “oaky”, “wooded”, “pencil shavings”, “toasted” and “burnt wood”, belonging to the “wooded” aroma family, and (2) “blackberry”, “blackcurrant”, “black fruit” (including all dark berries except blackberry and blackcurrant), “cherry”, “raspberry” and “strawberry”, belonging to the “berry” aroma family.

/Insert Fig. 3/

The data set with medium within-sample set variation (Chenin Blanc) produced CA plots (Fig. 4) with cumulative explained variances of the first two dimensions above 65%. When P1 was used, the cumulative explained variance of dimension one and two was 71.3%, when P2 was used 68.6%, when P3 was used 84.2% and when FC was conducted it was 66.7%. Furthermore, similar configurations for the PP and FC data sets with RV coefficients ranging from 0.66 to 0.88 (Table 3) were observed. In general, the differences between CA plots from PP data when different pivot samples were used, were more pronounced, with lower RV coefficients, than the differences between PP and FC. The similarity between P1 and P3 with an RV coefficient of 0.75 was an exception and showed good similarity. The RV coefficient between the CA plots constructed using P1 and P2 was 0.44, indicating dissimilarity. P2 had aroma characteristics that could overshadow other aroma nuances since aroma was described by words such as “vanilla”, “wooded”, “oaky”, “buttery” and “caramel” by many of the judges (Fig. 4b). The confidence ellipses on this CA showed frequent overlap between samples. A possible explanation could be that it was difficult for the sensory judges to detect differences between the other samples when comparing samples to P2, which had intense and extreme sensory characteristics. Confidence ellipses overlapped less frequently when a blend between the samples was used as pivot (P3), indicating clearer significant differences between samples (Fig. 4c). It is interesting to note that descriptors from the same aroma family were grouped well together on all CA plots obtained for this set. Examples were: (1) “sweet associated” characteristics such as “vanilla”, “caramel”, “honey” and “toffee” and (2) “oaky”, “wooded” and “planky”, which were positively correlated. Furthermore, higher explained variance could be observed when P3 was used as pivot sample when compared to FC and to the other PP evaluations when P1 and P2 were used.

/Insert Fig. 4/

From the CA plots constructed for the data set with high within-sample set variation (Sauvignon Blanc), the variation explained by dimension 1 and 2 was above 70% (Fig. 5), which is regarded as high for sensory data. When P1 was used, it was 79.9%, when P2 was used 87.1%, when P3 was used 82.4% and when FC was used it was 71.5%. Clear separation between the confidence ellipse of the pivot sample and the other samples was visible, but the overlapping confidence ellipses of the other samples indicated similarity and an inability of the panel to discriminate between those samples. It is possible that the uniqueness of the pivot sample caused the high explained variance and overshadowed the variation between other samples, causing a loss of separation between them.

The RV coefficients between the different sample sets varied from 0.28 to 0.95. Even though the effect of the pivot was overshadowing sensory characteristics, the RV coefficients between the CA maps when the extreme samples were used as pivots, P1 (Fig. 5a) and P2 (Fig. 5b), and the FC CA map were above 0.86 (Table 3). The low RV coefficient of 0.28 between CA maps constructed from P3 and P2, 0.51 between P1 and P3 and 0.36 between FC and P3, originated from the fact that one of the samples, TSL, was profiled differently when P3 was used as pivot sample.

/Insert Fig. 5/

4. Discussion

PP can be a useful technique to use for the profiling of complex products such as wine (Thuillier et al., 2015) and beer (Lelièvre-Desmas et al., 2017). The objective of this study was to evaluate PP critically for the profiling of complex matrices, comparing PP to FC, a well-established descriptive method (Campo et al., 2008). More specifically, the objective was to determine whether one of these techniques offered better discrimination between samples than the other one. To investigate these aspects thoroughly, three wine sample sets with different levels of within-sample set variation were analysed using a trained panel and CA was performed to obtain multivariate sensory maps.

Inspecting these CA plots, the following conclusions were reached. The variance explained by the first two factors when PP was used, regardless of the within-set variation complexity of the data set or the choice of pivot, was higher than 60%, indicating that the differences between samples were described well with PP. Confidence ellipses, calculated with bootstrapping, were added to the CA results as suggested by Lelièvre-Desmas et al. (2017) to understand the significance of product differences described by PP and FC. The confidence ellipses overlapped more frequently for PP than FC, showing that fewer samples were perceived to be significantly different when PP was performed than when FC was performed.

In addition, confidence ellipses shed light on perceived product differences when within-set product variation was varied. It is clear that the lower the within-set variation between samples was, the more frequent the overlap of confidence ellipses of different samples was. Due to the severe overlap of confidence ellipses for the data set with low within-set variation, it is not recommended to use PP to analyse such a set of products, even though it was suggested by Lelièvre-Desmas et al. (2017) that PP might be better suited to more homogenous spaces. However, for the sets with medium and large within-set variation, the confidence ellipses overlapped less frequently when a blend of the samples, rather than a sample with extreme characteristics, was used as pivot sample. It can, therefore, be concluded that more samples were perceived as significantly different when the blend was used as the pivot and the within-set variation was medium or high.

The similarity between sample configurations on the CA plots was tested by means of RV coefficients. Similarity between the different PP configurations, when the pivot sample was changed, and FC configurations differed for data sets with different degrees of within-set variation. Similar product configurations were obtained when the pivot was changed for the data set with low within-set variation, indicating that the choice of the pivot was not crucial. This observation was in line with observations made by Thuillier et al. (2015) when PP was proposed and Lelièvre-Desmas et al. (2017) when the stability of the product space was tested by varying the pivot sample used as well as the within-sample set variation. However, the similarity between PP configurations and the FC configuration was poor, except when a blend of all the samples was used as pivot. Thuillier et al. (2015) proposed using the blend as the pivot to create a centre sample, containing a wide range of sensory properties that spanned the sensory space, to which other samples were compared. Lelièvre-Desmas et al. (2017) noted that the idea of using a blend as pivot might be well suited to profiling of homogeneous spaces, which was confirmed in this study.

It is important to keep in mind that few significant differences between samples were observed for this set when PP was conducted. Even though Lelièvre-Desmas et al. (2017) found that PP might be more suited to homogenous spaces than heterogeneous spaces, this set was probably too homogeneous for profiling using PP. Lelièvre-Desmas et al. (2017), however, did not compute confidence ellipses by means of bootstrapping to validate product discrimination. Furthermore, the lack of quantification of the degree of similarity within a sample set causes subjective interpretation of what low, medium and high within-sample set variation is. Measures to quantitatively determine sample set complexity needs to be developed and can shed light on the performance of many other rapid methods.

If the set, regarded by Lelièvre-Desmas et al. (2017) as the set with low within-sample set variation was compared to the set defined in this study as the set with medium within-sample set variation, remarkably similar results were obtained.

The similarity between FC and PP data sets was good, with RV coefficients above or close to 0.7, regardless of the pivot used for the sample set with medium within-set variation. It is interesting to note that higher RV coefficients, indicating better similarity, were observed between the different PP data sets when different pivot samples were used and FC data than when these PP data sets were compared to each other. This was observed for the data set with large within-set variation as well with an exception when a blend of all the samples was used as pivot. In that case, poor similarity, with low RV coefficients was observed with the FC CA configuration and the PP CA configurations, originating from different pivot samples. Visual inspecting of the CA plots revealed that one sample was described differently and was consequently located differently relative to the other samples. It was noted by El Ghaziri and Qannari (2015) that RV coefficients would not provide a good estimate of the similarity of two spaces if one sample was not in the same position on both maps. In other words, if one sample was perceived differently, the RV coefficient would be low even though all the other samples were perceived similarly and would not provide a reasonable estimate of the overall similarity between two configurations, in this case, sensory spaces.

The question, however, remains why this sample was perceived differently. Two factors could play a role here: a physiological perception factor and a methodological limitation to use vocabulary that would distinguish wines from each other. It was noted by Lelièvre-Desmas et al. (2017) that the vocabulary might change when a different pivot is used. Therefore, they suggested that PP might not always be the best method to obtain a detailed sensory characterisation of samples but should rather be used to compare samples. In order to answer this question, a study could be designed in which sample sets with different complexities are created by substituting some samples with less and more complex wines but keeping to the same wine style and cultivar. Analysing these wines with DA and PP could then shed light on perceived differences due to a change of the pivot sample relative to the DA profile obtained.

The suggestion by Thuillier et al. (2015) to add the pivot sample as centre point by including zeros for all the descriptors in the table of citation frequencies containing +1 for a citation of more intense and -1 for a citation of less intense for individual judges was followed when P1 and P2 was used. The intensity of the pivot relative to the other samples was then determined during translation of the data and was reflected in the contingency table on which CA was performed. When P3, the blend, was used as pivot sample, this procedure was not followed and only the samples evaluated were represented in the CA plots. This should not affect the data, particularly the CA plots, if the assumption that P3 was an average centre sample representing the characteristics of all the samples equally held since all the samples were evaluated relative to the pivot. It, however, cannot be ruled out that the data was affected and, therefore, the RV coefficients describing the similarities between P1, P2 and P3 configurations. It should be noted then that it might be worthwhile testing, by statistically including P3 in the CA plot and comparing the configuration to a CA plot with P3 excluded. Furthermore, a sensory experiment including the pivot as a sample as well and not just a theoretical centre point during the statistical analysis could be insightful.

In the light of what has been discussed, it has to be said that the total number of descriptors allowed for product description was three to five when FC was performed and three to 10 when PP was performed, if the number of descriptors allowed to describe sensory characteristics perceived as less and more intense for PP was taken into account. This could contribute to sensory judges focussing less on the most prominent characteristics of the sample causing more noise, therefore more overlap between confidence ellipses. Furthermore, the chance of choosing the same attribute for more than one sample could also increase the overlap of confidence ellipses around samples on the CA plots. In contrast, richer data might have been obtained since more descriptors per wine were generated, which could explain the higher explained variance observed for PP in comparison to FC. Even though these restrictions might have influenced results, it was considered as the most practical choice for the method when using a trained panel. The choice of the number of allowed attributes was made based on recommendations from the literature but mainly on feedback from the panel during training sessions. These limits were set to ensure that all the panellists used the protocol and a similar approach.

An aspect of PP that still requires attention is the testing of panel performance. In previous studies in which PP was used as a profiling technique, the measurement of panel performance did not receive enough attention. Thuillier et al. (2015) proposed the method but did not propose a strategy to measure panel performance since the focus of that study was on a simulation in which panel heterogeneity was set as a parameter. It would, therefore, not make sense to test panel performance on the simulation data. Fonseca et al. (2016) and Esmerino et al. (2017) performed PP using consumers as panellists without investigating possible segmentation or testing the performance of individuals. Testing panel repeatability was not possible with the data obtained during the consumer studies as judges did not repeat the test. Testing consumers' performance is not common and is deemed irrelevant due to the large number of participants that increases the statistical power of the experiment. However, investigating panel segmentation and individual differences could provide valuable insights into how consumers profile the product when performing PP. Lelièvre-Desmas et al. (2017) proposed a strategy to evaluate global panel consensus and repeatability when performing PP, but the authors also acknowledged that more work needed to be done in this field.

In this study, panel repeatability was measured using the Ri value and consensus using Cohen's kappa coefficients. Both these measures provide useful insights into panel performance but are probably too strict since they only take exact matches of attributes as good consensus between two judges. It could make sense to penalise judges less or not at all when two judges use slightly different attributes that still belong to the same odour family. Weighing contributions to the Ri value could be applied by assigning, for example, 0.5 instead of zero if an attribute from the same aroma family is sited in both the first and second repeat. In order to incorporate this idea into panel performance testing, more work is required in the field of sensometrics.

Critical investigations of panel performance measurements and a proposed workflow to measure consensus and repeatability for PP and FC, similar to the work published by Tomic et al. (2007) and Tomic et al. (2010) for DA, could be valuable additions to the methodology development of rapid methods.

It would be interesting to evaluate the performance of PP when performed by industry professionals or naïve consumers when judges are not trained, and less detailed results might be captured. Industry professionals' sensory perception responses generally reflect the lexicon that they developed during their years of experience taking part in quality/competition-type tastings, keeping the production process in mind. PP was originally proposed by Thuillier et al. (2015) as an alternative to free description when capturing industry professionals' sensory perceptions. Capturing consumers' less detailed descriptions related to styles in general, preferences and emotion could be a new application for PP.

In this study, a single modality, aroma, was assessed. This modality can easily be assessed by methods such as FC and CATA. Mouthfeel and taste might be difficult or unpractical to asses with FC since it often means little if the relative intensity of the attribute in terms of the products cannot be assessed by the individual judges. The assumption that the number of citations will indicate the intensity is not always true when a trained panel profiles wine. From unpublished data, it was found that most wines in a sample set could, for example, be sour and alcoholic but some wines are more sour or less sour than other wines (Brand and O'Kennedy, unpublished research on white wines). Although it was not specifically stated that FC was less suitable for taste attributes than aromas, Campo et al. (2008) only proposed the technique and compared it to DA (Campo et al., 2010) for aroma evaluation of wine. In this case, FC will not be able to detect differences between wines in terms of taste attributes and PP might offer a solution and could be a more suitable option than FC for profiling the taste and mouthfeel properties of wines.

5. Conclusions

PP could be a useful wine sensory evaluation technique when a comparison between products is required either through profiling of individual wines or direct comparison, for example during benchmarking. As a profiling technique, PP could be a viable alternative for FC. However, the results obtained clearly showed that the nature of the samples analysed and particularly the level of variation between samples needs to be considered and that the results could be influenced by the choice of the pivot sample.

From this study, it was clear that when sample sets with very low within-sample set variation were tested, FC was a more sensitive technique to use than PP.

The sensory space generated using PP for a wine sample set with medium within-set variation and using a central sample as the pivot was comparable to results obtained with FC. The most reliable results were obtained from this type of sample set when a blend of all the samples was used as the pivot.

Sample sets with large within-set variation might be less suitable for analysis by PP and FC results will probably be more stable. However, with these sets, good similarity between FC and PP results was obtained when extreme samples were used a pivot samples, whereas poor similarity between PP and FC was observed when a blend of the samples was used as pivot.

The panel repeatability was comparable and good for both PP and FC. PP was experienced by judges as significantly more difficult to perform compared to FC, irrespective of the complexity of the data set and the pivot sample used. Cohen’s kappa coefficients indicated reasonable to moderate consensus for both PP and FC when the sample set with large within-sample variation was analysed, but low values were obtained when a blend of all the samples was used as pivot.

A workflow to test panel consensus and repeatability will add value to the PP methodology. Panel performance testing is currently a shortcoming of the methodology available for PP in the literature. Testing the ability of Cohen’s kappa and related kappa coefficients, for example Fleiss’ kappa, on data sets varying in terms of within-set variation for PP analysis to assess both repeatability and consensus could be a first step in designing such a workflow.

To conclude, for sensory studies where simultaneous sample presentation is required to get an overview of the sample set during profiling, PP could be preferred over FC. This could be the case when product experts, producers or consumers evaluate samples since these judges are generally not trained and might be inconsistent when evaluating samples in a monadic manner. These types of panels are generally not required to evaluate sample sets with small with-in set variation. When FC is used the assumption is made that the larger the number of citations the more intense that attribute might be. In the case of wine fault analysis this assumption might not hold. A method where relative intensity is captured, such as with PP, could be more informative than FC measuring how many judges perceived attributes related to the fault. Another application where PP could be more relevant to use than FC is when a one-to-one comparison between two products is required. The stability of the sensory space will not play a role here since only two products are evaluated directly with each other and not in relation to a common reference. Examples of such cases include benchmarking and shelf-life studies. For these two applications it would be interesting to compare PP to other rapid sensory methods such as sorting and particularly reference-based rapid sensory methods such as polarised sensory positioning (Teillet et al., 2010) and polarised projective mapping (Ares et al., 2013).

Acknowledgements

The authors would like to express their sincere gratitude to Winetech for funding project IWBT 13/02, the National Research Foundation of South Africa and THRIP for funding project Tp14080888680. We thank the Chenin Blanc Association of South Africa and private cellars for their continuous support and for donating and supplying wine to us at reduced prices. A special word of thanks to the sensory technical staff from the Department of Viticulture and Oenology, Stellenbosch University, Valeria Panzeri, Olwethu Fana and Wandisa Maxiti for their technical support.

References

Abdi, H., & Valentin, D. (2014). Encyclopedia of Measurement and Statistics Multiple Factor Analysis. *Encyclopaedia of Measurement and Statistics*, 658–664.

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

Adams, J., Williams, A., Lancaster, B., & Foley, M. (2007). Advantages and uses of check-all-that-apply response compared to traditional scaling of attributes for salty snacks. In: Poster presented at the 7th Pangborn sensory science symposium. Minneapolis, MN, USA (12–16 August). Delegate Manual.

Altman, D.G. (1991). *Practical statistics for medical research*. London England: Chapman and Hall, p404.

Ares, G., de Saldamando, L., Vidal, L., Antúnez, L., Giménez, A., & Varela, P. (2013). Polarized Projective Mapping: Comparison with Polarized Sensory Positioning approaches. *Food Quality and Preference*, *28*(2), 510–518.

Bester, I. (2011). Classifying South African Chenin blanc wine styles. MSc thesis, Stellenbosch University, Private Bag X1, 7602 Matieland (Stellenbosch), South Africa.

Cadoret, M., & Husson, F. (2013) Construction and evaluation of confidence ellipses applied to sensory data. *Food Quality and Preference, 28*, 106-115.

Campo, E., Ballester, J., Langlois, J., Dacremont, C., & Valentin, D. (2010). Comparison of conventional descriptive analysis and a citation frequency-based descriptive method for odor profiling: An application to Burgundy Pinot noir wines. *Food Quality and Preference*, *21*(1), 44–55.

Campo, E., Do, B. V., Ferreira, V., & Valentin, D. (2008). Aroma properties of young Spanish monovarietal white wines: A study using sorting task, list of terms and frequency of citation. *Australian Journal of Grape and Wine Research*, *14*(2), 104–115.

Cohen, J. A. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement. 20*(1), 37–46.

Dehlholm, C., Brockhoff, P. B., & Bredie, W. L. P. (2012). Confidence ellipses: A variation based on parametric bootstrapping applicable on Multiple Factor Analysis results for rapid graphical evaluation. *Food Quality and Preference*, 26(2), 278–280.

Deneulin, P., Reverdy, C., Rébénaque, P., Danthe, E., & Mulhauser, B. (2018). Evaluation of the Pivot Profile©, a new method to characterize a large variety of a single product: Case study on honeys from around the world. *Food Research International*, *106,* 29-37.

El Ghaziri, A., & Qannari, E. M. (2015). Measures of association between two datasets; Application to sensory data. *Food Quality and Preference*, *40*, 116–124.

Esmerino, E. A., Tavares Filho, E. R., Thomas Carr, B., Ferraz, J. P., Silva, H. L. A., Pinto, L. P. F., Freitas, M. Q., Gruz, A. G., Bolini, H. M. A. (2017). Consumer-based product characterization using Pivot Profile, Projective Mapping and Check-all-that-apply (CATA): A comparative case with Greek yoghurt samples. *Food Research International*, 99, 375-384.

Fonseca, F. G. A., Esmerino, E. A., Filho, E. R. T., Ferraz, J. P., Cruz, A. G., & Bolini, H. M. A. (2016). Novel and successful free comments method for sensory characterization of chocolate ice cream: A comparative study between pivot profile and comment analysis. *Journal of Dairy Science, 99(5)*, 3408–3420.

Gisev, N., Bell, J. S., Chen, T. F. (2013). Interrater agreement and interrater reliability: Key concepts, approaches and applications. *Research in Social and Administrative Pharmacy, 9*(3), 330–338.

Hanekom, E. (2012). Chemical, sensory and consumer profiling of a selection of South African Chenin blanc wines produced from bush vines. MSc thesis, Stellenbosch University, Private Bag X1, 7602 Matieland (Stellenbosch), South Africa.

Laing, D. G., & Glemarec, A. (1992). Selective attention and the perceptual attention of odor mixtures. *Physiology and Behavior, 52*, 1047–1053.

Lancaster, B., & Foley, M. (2007). Determining statistical significance for choose-all- that-apply question responses. Seventh Pangborn sensory science symposium, Minneapolis, USA.

Lawless, H. T. (1999). Descriptive analysis of complex odors: Reality, model or illusion? *Food Quality and Preference*, 10(4–5), 325–332.

Lelièvre-Desmas, M., Valentin, D., & Chollet, S. (2017). Pivot profile method: What is the influence of the pivot and product space? *Food Quality and Preference, 61*(May), 6–14.

Louw, L., Malherbe, S., Naes, T., Lambrechts, M., van Rensburg, P., & Nieuwoudt, H. (2013). Validation of two Napping® techniques as rapid sensory screening tools for high alcohol products. *Food Quality and Preference*, 30(2), 192–201.

McFie H. J. H., & Thomson D. M. H. (1988). Preference mapping and multidimensional scaling. In: J. R. Piggott (ed.), Sensory Analysis of Foods. Elsevier Applied Science, New York, pp. 381–409.

McHugh, M.L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, *22*(3), 276 – 282.

Noble, A.C., Arnold, R.A., Buechsenstein, J., Leach, E.J., Schmidt, J.O. & Stern, P.M. (1987). Modification of a standardized system of wine aroma terminology. *American Journal of Enology and Viticulture, 38*, 143- 146.

Peryam, D. R., & Pilgrim, F.J. (1957). Hedonic scale method of measuring food preference. *Food Technology*, *11*, 9-14

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

Robert, P. & Escouffier, Y. (1976). A unifying tool for linear multivariate statistical methods: The RV coefficient. *Applied Statistics*, 25, 257-265.

Symoneaux, R., Galmarini, M. V., & Mehinagic, E. (2012). Comment analysis of consumer's likes and dislikes as an alternative tool to preference mapping. A case study on apples. *Food Quality and Preference*, *24*, 59–66.

Teillet, E., Schlich, P., Urbano, C., Cordelle, S., & Guichard, E. (2010). Sensory methodologies and the taste of water. *Food Quality and Preference*, *21*(8), 967–976.

Tomic, O., Luciano, G., Nilsen, A., Hyldig, G., Lorensen, K., & Næs, T. (2010). Analysing sensory panel performance in a proficiency test using the PanelCheck software. *European Food Research and Technology*, 230(3), 497–511.

Tomic, O., Nilsen, A., Martens, M., & Næs, T. (2007). Visualization of sensory profiling data for performance monitoring. *LWT - Food Science and Technology*, 40(2), 262–269.

Thuillier, B., Valentin, D., Marchal, R., & Dacremont, C. (2015). Pivot© profile: A new descriptive method based on free description. *Food Quality and Preference*, *42*, 66–77.

Valentin, D., Chollet, S., Lelièvre, M., & Abdi, H. (2012). Quick and dirty but still pretty good: A review of new descriptive methods in food science. *International Journal of Food Science and Technology*, 47(8), 1563–1578.

Varela, P., & Ares, G. (2012). Sensory profiling, the blurred line between sensory and consumer science. A review of novel methods for product characterization. *Food Research International, 48(2)*, 893–908.

Van Antwerpen, L. (2012). Chemical and Sensory profiling of dry and semi-dry South African Chenin blanc wines. MSc thesis, Stellenbosch University, Private Bag X1, 7602 Matieland (Stellenbosch), South Africa.

Williams, A.A. & Langron, S.P. (1984). The use of free-choice profiling for the evaluation of commercial ports. *Journal of the Science of Food and Agriculture*, 35, 558–568.