Using contrails and animated sequences to visualize uncertainty in dynamic sensory profiles obtained from temporal check-all-that-apply (TCATA) data

John C. Castura, Allison K. Baker, Carolyn F. Ross

School of Food Science, Washington State University, Pullman, WA 99164-6376, USA
Compusense Inc., 255 Speedvale Ave. W., Guelph, Ontario N1H 1C5, Canada

Abstract

Approaches for analyzing temporal check-all-that-apply (TCATA) data are further developed and illustrated using data arising from a Syrah wine finish evaluation. Raw and smoothed trajectories are obtained using principal component analysis. Virtual panels are obtained from a partial bootstrap, and the attribute citation proportions are then projected into the solution space to form contrails. Trajectories are overlaid on the contrails, allowing smoothing to be evaluated. Separation between two contrails provides evidence that the trajectories differ. At individual time slices, data concentration ellipses are overlaid on bootstrap scores. Separation of ellipses provides evidence of differences among treatments. Difference trajectories and difference ellipses can also be plotted; if the difference ellipse excludes the origin it indicates a difference between the treatments. Animated sequences summarize changes in product characterization over time in a manner that facilitates review. A glossary of terms introduced in the paper is provided in an appendix.

© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Temporal check-all-that-apply (TCATA; Castura, Antúnez, Giménez, & Ares, 2016) is a temporal sensory method, and its data are used to describe the temporal evolution of sensory attributes in the products under evaluation. Assessors, who could be either trained or untrained, are tasked with checking and unchecking attributes from a list during the evaluation period, such that the attributes that are selected at any given time describe the product at that time. TCATA methodology has been used to investigate temporal sensory evolution in a range of product categories: food, such as yogurt (Castura et al., 2016, Fig. 5), salami, cheese, French bread, and marinated mussels (Ares et al., 2015); beverages, such as chocolate-flavoured milk (Oliveira et al., 2015) and red wine (Baker, Castura, & Ross, 2016); and non-food, such as cosmetic products (Boinbaser, Parente, Castura, & Ares, 2015).

Familiar heuristic approaches for visualizing data from temporal dominance of sensations (TDS; Pineau et al., 2009) studies have been leveraged to show TCATA curves as smoothed attribute citation proportions over time. For each TCATA attribute, the citation rate of a product of interest can be contrasted with the average citation rate of the other products (Castura et al., 2016, Figs. 3 and 4). The data visualization described above can be considered to be a generalization of difference curves, which contrasts attribute citation rates for one product against another product over time, thus emphasizing statistically significant differences (Castura et al., 2016, Fig. 5).

Univariate TCATA curves can be onerous to review when there are many products and many attributes. For this reason it can be useful to consider TCATA data from a multivariate perspective, which provides both data reduction and interpretation advantages (Johnson & Wichern, 2007, chap. 8). Castura et al. (2016, Figs. 6 and 7) submit a contingency table of TCATA citation frequencies (rows: Product * Time; columns: Attributes) to correspondence analysis (CA) and join adjacent time slices to create a separate curve per product. Each curve can then be smoothed so as to show trends without overfitting the data. Each curve is called a trajectory, following terminology for multivariate changes in TDS dominance rates in Lenfant, Loret, Pineau, Hartmann, and Martin (2009). A sense of temporal progression is given by placing markers along the trajectories at specified time intervals (e.g. every 5 s).

A check-all-that-apply (CATA) contingency table can also be analyzed via principal component analysis (PCA) on the covariance matrix (Meyners, Castura, & Carr, 2013). It is possible to conduct...
PCA on TCATA data by organizing the data into contingency tables in the same manner as described by Castura et al. (2016) for analysis using CA. If analyzing data arising from unbalanced experimental designs or in the presence of missing data then PCA can be conducted on the table of mean citation proportions organized with Product * Times in rows and Attributes in columns. Regardless, trajectories are obtained by joining (and optionally, smoothing) adjacent time points. Trajectories describe the evolution in how the products are characterized over time. However, when inspecting trajectories, one might wonder: Do the data provided by assessors discriminate the products based on the trajectories? A plot of the trajectories on its own cannot adequately answer these questions because it lacks information regarding the uncertainty associated with each trajectory.

One way to visualize the stability in product configurations obtained from PCA is to construct confidence ellipses for the products (Husson, Lé, & Pagès, 2005), using an approach derived from the bootstrap technique (Efron, 1979). Assessors are resampled with replacement, resulting in the creation of many virtual panels. With this technique, confidence ellipses tend to decrease as the panel size increases and as the panelist reliability increases. In this manuscript we build on the approach of Husson et al. (2005) to evaluate the stability of trajectories based on TCATA data, and visualize these data both statically and dynamically in a manner that incorporates uncertainty.

In Section 2.1, we describe the TCATA data that will illustrate the methods applied in this paper. In Section 2.2, PCA is conducted on unfolded but untransformed citation proportions to obtain scores, which are joined to form trajectories. The focus of this manuscript is to evaluate the uncertainty associated with these trajectories. In Section 2.3, we resample TCATA assessors with replacement to create virtual panels, and project their data into the multivariate sensory space obtained from the PCA solution. In Section 2.4, we use contrails to evaluate the uncertainty associated with each trajectory. In Section 2.5, we use data concentration ellipses to summarize the region of uncertainty for the panel data for each treatment at each time slice. Non-overlapping ellipses are indicative of treatment differences. In Section 2.6, we use the differences in scores to obtain difference trajectories and difference ellipses to further investigate differences between pairs of treatments. Finally, in Section 2.7, an animated sequence provides a useful dynamic representation of changes in treatment characterization over time. The proposed approaches are intended for exploratory data analysis and data visualization, and can be used to gain insights and to test or generate new hypotheses.

2. Materials and methods

2.1. Data

Methods described in this manuscript are illustrated using a data set collected at Washington State University. Grapes grown in the Columbia Valley of Washington State were hand-harvested by Washington State University students and made into wine using commercial techniques at a pilot plant scale. Starting from the same grape must, sugar and ethanol levels were manipulated to create three wine treatments: a high-alcohol wine ("H"), a low-alcohol wine ("L"), and a low-alcohol wine that was adjusted to have the same alcohol content as the high-alcohol wine ("A"). Each wine treatment was made in 2 fermenters. The wines were evaluated by a trained panel (n = 13) that was recruited, selected, and trained to evaluate Syrah wine using the TCATA method in Compusense at-hand (Compusense Inc., Guelph, Ontario, Canada).

The ballot included 10 attributes (Astringent, Bitter, Dark Fruit, Earthy, Green, Heat/Ethanol burn, Other, Red Fruit, Sour, and Spices) which were randomized in lists allocated to assessors (Meyners & Castura, 2015). Each evaluation was performed in the following manner. Upon taking Sip 1 into the mouth, the assessor clicked Start, which started an onscreen timer. An onscreen prompt appeared at 10 s, cuing the assessor to expectorate the wine sample and begin the evaluation. The assessor could stop the evaluation when all sensations had ceased; otherwise, the evaluation ended when the timer reached 180 s. After a delay of 60 s without palate cleansing, the assessor was presented with another TCATA question, in which Sip 2 was evaluated in the same manner as Sip 1. A complete-block experiment was conducted on the 6 treatment * fermenters with 2 sips per sample as described, in duplicate. There were no missing observations. Other was cited rarely, and, unlike the other sensory attributes, was non-specific. Thus, it was excluded from analyses presented below.

Further details regarding the wine treatments and panel training protocols are omitted for brevity but can be found in Baker et al. (2016).

2.2. Trajectories

Data were unfolded to provide a matrix with the 9 specific sensory attributes in columns, and rows corresponding to unique combinations of treatment and sip at 0.1-s interval time slices from 10.0 s to 180.0 s. (Sip was included to investigate potential differences between the first and second sips.) Each Treatment * Sip is henceforth referred to as a WineSip.

Covariance PCA was conducted on the table of mean attribute citation proportions. Such analysis on an unfolded matrix is also called Tucker-1 (Tucker, 1966; see Dahl & Næs, 2009 for other applications for sensory evaluation data). Scores from adjacent time slices for each WineSip were joined to give raw trajectories, which are presented in Fig. 1. Superimposition of the attributes on this plot was achieved by multiplying each eigenvector by an appropriate scalar. (The choice of scalar is made strictly for display purposes. In this manuscript we obtain an appropriate scalar by dividing the largest absolute score by the largest absolute eigenvector.) The relative positions of the trajectories are given in a space retaining Euclidean distances, i.e., a distance biplot (see Legendre & Legendre, 2012, p. 444).

WineSip trajectories were smoothed in the manner described by Castura et al. (2016). Each WineSip trajectory has its direction indicated with an arrow at the 20-s time slice, and is labelled at its 30-s time slice. Markers are shown at the 40-s time slice and at 10 s-intervals thereafter. Collectively the spacing between markers indicate whether the changes in sensory characterization had occurred rapidly (long segments between markers) or slowly (short segments between markers). Inspection of raw trajectories (Fig. 1) and smoothed trajectories (Fig. 2) help to confirm that the latter successfully capture the main patterns that are present in the data.

It is also possible to plot the difference between two trajectories through simple subtraction. Each difference trajectory summarizes the differences between two trajectories, and will be discussed further in Section 2.6.

2.3. Virtual panels and partial bootstrap

Stratified sampling was conducted, such that assessors were sampled with replacement to create a large number of virtual panels, each the same size as the true panel (n = 13). In this manuscript, 499 virtual panels were used, similar to the quantity used by Husson et al. (2005); results from the analyses in the sections that
follow will be unchanged so long as a sufficiently large number of panels is used. Data from each virtual panel were organized into a table of mean citation rates, matching the organization of the table prepared for the real panel, as described in Section 2.2, with attributes in columns and unique combinations of treatment, sip, and time slices (at 0.1 s intervals) in rows.

Attribute loadings from the real panel and mean citation rates from the virtual panel were used to obtain scores for the time slices along each WineSip. Scores from the 500 panels (1 real panel and 499 virtual panels) are henceforth referred to collectively as bootstrap scores. This stratified bootstrap is similar to the technique that Husson et al. (2005) use to simulate virtual panels.

2.4. Contrails

In Section 2.2, a sensory space is obtained from PCA conducted with active objects (the citation proportions for each WineSip at each time slice by the real sensory panel). In Section 2.3, passive objects (citation proportions for each WineSip at each time slice from each of the virtual panels) are projected onto this space.

Fig. 1. Following analysis of TCATA citation rates by PCA, WineSip trajectories are presented for (a) PC1 vs. PC2 and (b) PC2 vs. PC3. Arrows are provided at 20-s time slices to indicate the directions of the trajectories. (L1 and L2 in (b) have a counterclockwise direction. All other trajectories have a clockwise direction. A label identifying each of the six WineSip trajectories (A1, A2, H1, H2, L1, L2) appears at the 30-s time slice.)
When all active and passive data are projected into a plane showing two principal components, the result is a scatterplot of bootstrap scores. For visualization purposes, symbols used for displaying bootstrap scores are given the same level of semi-transparency, such that the opacity of a symbol increases with each additional point that is overlaid. The linear clouds of bootstrap scores for each WineSip leading up to a particular time slice is here called a contrail. Each contrail provides a perspective on the uncertainty of the corresponding trajectory. Contrails and smoothed trajectories for the Syrah TCATA data introduced in Section 2.1 are presented in Fig. 2.

2.5. Data concentration ellipses

Data concentration ellipses can be used to visualize the probability contours for data. Under the assumption of bivariate normality, a 95% data concentration ellipse encloses 95% of the data. Computation of data concentration ellipses are facilitated by the `dataEllipse` function in the R package `car` (Fox & Weisberg, 2011).

Fig. 3 presents the smoothed trajectory for each WineSip based on data leading up to the time slice at 30.0 s in the plane of PC2 and PC3. Data arising from the real and the virtual panels for each
WineSip at this time slice are projected onto the plane, then summarized with 95% data concentration ellipses.

For brevity and for illustrative purposes, we focus next only on the second sip of each of the three wine treatments using the partial bootstrap procedure at the 30.0-s time slice, and obtain successively wider data concentration ellipses which encompass 40%, 68%, and 95% of the bootstrap scores in the plane of PC2 and PC3 (Fig. 4). These coverage levels are noted by Friendly, Monette, and Fox (2013) to have special relevance. The 40% data concentration ellipse is associated with univariate intervals $\bar{x} \pm s_x$ and $\bar{y} \pm s_y$. The 68% data concentration ellipse encloses approximately 2/3 of bivariate normal data, analogous to the standard deviation for univariate normal data. Finally, the 95% data concentration ellipse encircles 19/20 of bivariate normal data.
2.6. Data concentration ellipses for differences in trajectories

If two trajectories coincide exactly, then all differences in scores will be zero in the multivariate sensory space. Due to both systematic changes and random fluctuation in the data, the difference in trajectories will often depart from the origin. A question arises: are two trajectories considered different at a given time slice? To address this question, bootstrap scores are obtained for each treatment at each time slice. For each pair of treatments, the differences in bootstrap scores are obtained (at each time slice), then plotted and summarized using a difference ellipse. Exclusion of the origin by a difference ellipse provides evidence of a difference between the pair of treatments at that time slice at the coverage level of the ellipse.

In Fig. 5, we illustrate this procedure by investigating the differences between H and A trajectories, between the H and L trajectories, and between the A and L trajectories, in the plane of PC2 and PC3 at the 30-s time slice of the second sip. These trajectories are labelled “HA”, “HL”, and “AL”. A scatter of the differences in the bootstrap scores is presented for each of the pairwise comparisons, each overlaid by its 95% data concentration ellipse, i.e., the difference ellipse.

2.7. Animated sequences

To investigate temporal differences between treatments and sips further, we use an animated sequence. Trajectories leading up to the current time slice are computed and displayed for each WineSip, along with a data concentration ellipse based on bootstrap scores for the current time slice. This approach has the potential of revealing differences in the trajectories at each time slice. It is also possible to see how the smoothed trajectories are updated as new information is made available to the smoothing function.

Animated sequences were prepared using the Syrah wine data discussed in Section 2.1. The timer starts with the start of the evaluation of the wine finish at 10 s. The WineSip trajectories are shown at 0.1-s intervals, and at each time slice, the corresponding 95% data concentration ellipses are shown. The visualization is presented in a video clip that accompanies the online version of this manuscript. Click the play button on the image below to view Video 1 (online version only).

3. Discussion

3.1. Trajectories

In this manuscript, multivariate exploratory analyses were conducted using covariance PCA. Although CA is a natural choice for analyzing count data, there are several advantages to using PCA to obtain trajectories. First, PCA can be applied to either a table of counts or mean proportions; the latter provides a more straightforward analysis for data sets that are unbalanced or contain missing data. Second, CA based on the \( \chi^2 \) metric can yield variables
that have extreme coordinates yet have small mass and contribute little to the interpretation of that component, e.g. in the case of columns with very low counts (Greenacre, 2006), which may be prone to misinterpretation. Third, TCATA citation proportions at the start of the evaluation all start at zero, and ultimately, if sensations are evaluated until extinction, return to zero, and unlike CA, PCA does not require that each row and column have a non-zero sum. Both methods exploit the underlying temporal gradient in the configurations which are visualized by joining adjacent time slices to create the trajectories in the manner described. In either case, the data must be interpreted as related to attribute citation proportions, not to attribute intensities.

As a reviewer pointed out, PCA could be conducted in different ways for these data. For example, PCA could be conducted on the overall average citation proportions for each WineSip, after which WineSip citation proportions for the individual time slices could be projected as passive objects into the space. For these data, the overall average citation rates for the WineSips are similar to the citation rates observed for the WineSips between approx. 75 s and 95 s, thus the proposed approach would end up maximizing the variance via PCA at this particular time segment. The proposal of row centering citation proportion data prior to conducting PCA was also given. Essentially, it is mean subtraction by row, which largely removes citation rate variation that is otherwise captured in PC1 (Figs. 1a and 2a). In mean subtraction, a row with uniformly low citation proportions and a row with uniformly high citation proportions are treated identically after mean subtraction, i.e. both rows would have mean-centered citation rates set to 0. The meaning of the original data–citation proportions, on the same scale–becomes lost. If PCA is conducted on our data after row centering, then PC1 and PC2 from that analysis would explain 41.5% and 30.6% of variance, respectively, and bear resemblance to PC2 and PC3 in Figs. 1b and 2b, both in relative proportions of variances explained and interpretation. Data trimming, or even performing multiple PCAs to investigate dynamics within several time intervals, might also be considered. Time standardization (see Lenfant et al., 2009) is a potential preprocessing manipulation that readers might also consider. This data treatment is intended to remove assessor noise (e.g. assessor differences in oral processing efficiencies), but has the potential to remove real product differences (e.g. product duration differences), so should be considered only after reviewing data. We find that PCA on the untransformed citation proportions is easy to interpret and to communicate, but acknowledge that the best approach may depend on the data and on the objectives of the study.

The PCA distance plots in the figures herein are interpreted by considering the position of coordinates relative to each attribute vector. PC1 is interpretable as a ‘mean citation proportions’ dimension which contrasts the zero and low citation proportions that occur at the start and end of the evaluation with relatively large mean citation proportions, which occur in all WineSips between 28 s and 38 s. For each WineSip, the maximum absolute score either coincides precisely with or overlaps the time slice at which which WineSip receives the largest number of citations across all attributes. PC2 is interpretable as an ‘ethanol impact dimension’. It contrasts the constellation of attributes that tend to occur in the early- to mid-evaluation in low-alcohol wines (Sour, Green, and Red Fruit) vs. high-alcohol wines (Heat/Ethanol burn, Bitter, and Spices). PC3 can be considered a ‘Sour–Astringent contrast’ dimension; it contrasts the Sour gustatory sensation that occurs in early- to mid-evaluation with the Astringent mouthfeel sensation that tends to be perceived in late-evaluation; Sour and Astringent tend to be co-elicited with Red Fruit and Dark Fruit, respectively. Although PC1 explains most of the variability in the data (Figs 1a and 2a), PC2 and PC3 are interpretable and capture important variation in the data (Figs. 1b and 2b).

It is due to variance that PC1 and PC2 are uncorrelated yet not independent. The variance of any binomial proportion is largest at 0.5, and decreases for proportions closer to 0 or 1. (The latter does not occur in this study.) Similar lack of independence also occurs between PC1 and P3 (not shown directly). The relatively small data concentration ellipses that co-occur with very low citation proportions is evident also in Video 1 that is provided with the online version of this manuscript.

Interpretation for the wine data presented here is straightforward. In PC1, the first and second sips of each wine treatment occur at approximately the same times; however, each of the second–sip trajectories reach a more extreme position than the respective first-sip trajectories, suggesting the possibility of sensory build-up of sensations over sips. Furthermore, each sip of L reaches a less extreme position in PC1 when compared to the same sip of H and A. In the plane of PC2 and PC3 (Figs. 1b and 2b), L is characterized initially by the attributes Sour, Green, and Red Fruit, and is increasingly characterized by Dark Fruit and Astringent as the evaluation progresses. H starts off as being characterized by attributes Heat, Spices, and Bitter, and later in the evaluation by Dark Fruit, Astringent, Earthy, and Green. The second sip is characterized by Sour more often than the first sip. A is similar to H, but characterized less often as Bitter, and more often as Sour, increasingly in the second sip.

The question arises whether the trajectories that are visualized in these dimensions are systematic patterns in the data (signal) or random fluctuations (noise). PC2 and PC3, which each explain a low proportion of the variance in the data submitted to PCA, deserve particular scrutiny. We note data truncation, i.e. dropping time slices with a citation rate below some minimum threshold, reduces the variance that is captured in PC1, and variance captured in PC2 and PC3 then accounts for a greater proportion of the variance in the reduced data set. Row centering, discussed above, has a similar effect. However this paper is focused on using the data resampling approach, i.e. partial bootstrap, discussed in Section 2.3, to investigate this question, and the results of this particular investigation, which are presented in Sections 2.4–2.7, will be discussed in the sections that follow.

Chemical composition details associated with the wine treatments are presented and discussed by Baker et al. (2016). As noted there, H and A, which had a higher ethanol concentration than L, also had higher citation rates of Bitterness and Heat/Ethanol burn. L was observed to have the highest Sour citation rate, and A had a higher citation rate than H, likely explained by the higher titratable acidity in L and A than in H, and the sourness masking effects that accompany increasing ethanol concentration (Martin & Pangborn, 1970). All treatments fell into ethanol and pH ranges in which increases might also accompany increases in perceived bitterness (Fischer & Noble, 1994). Lower levels of ethanol and pH are candidate explanations for the higher Astringency citation rates in L (Fontoin, Saucier, Teissedre, & Glories, 2008); however, differences in pH for treatments L, A, and H were extremely slight (3.44, 3.48, and 3.48, respectively). Thus, ethanol was likely driving differences in bitterness and astringency perception.

3.2. Virtual panels and partial bootstrap

One reviewer noted that we used a partial (stratified) bootstrap to obtain contrails and data concentration ellipses, which seems at odds with Cadoret and Husson (2013), who recommend a truncated total bootstrap to obtain confidence ellipses. We provide the following discussion to clarify how and why the partial bootstrap is applicable to these data. When analyzing results from so-called holistic sensory methods (e.g. napping), Cadoret and Husson note that data are typically unfolded to create a two-way matrix with products in rows and assessors * variables in columns.
The number of columns is often large due to the relatively large number of assessors used in such tests, and a Procrustes rotation of data from a virtual panel too often finds a close match with the product configuration obtained from the real panel, even for data in which product labels have been permuted (i.e., random data). The resulting fit is unjustifiably optimistic, and confidence ellipses are small. When investigating uncertainty in descriptive analysis results, which are usually analyzed in a two-way matrix with products in rows and sensory attributes in columns, this dimensionality problem is not present so long as there are relatively few columns. Cadoret and Husson recommend a truncated total bootstrap for all cases for their procedure, but note that the partial bootstrap provides similar outcomes when the number of sensory attributes is reasonably low, such as in their example data set which includes 12 sensory attributes (see: Cadoret & Husson, 2013, Figs. 8 and 9).

The dimensionality problem that motivated Cadoret and Husson (2013) to investigate alternatives to the partial bootstrap does not occur with the procedure that we propose herein, not only because the number of columns (i.e., TCATA attributes) is always low, as in our data (which have only 9 sensory attributes), nor only because we use data concentration ellipses rather than confidence ellipses. Rather, bootstrap scores are obtained without Procrustes rotations or any other manipulations for obtaining a mathematically optimal (and potentially overly optimistic) fit between the product configuration of the real panel and the product configuration of each virtual panel. These assertions seem obvious from the procedure described in Section 2.2, but were nonetheless confirmed empirically by performing the partial bootstrap, then permuting sample labels within each assessor. As expected by construction, all WineSip contrails are centered on the average trajectory, and overlap indistinguishably. Throughout the evaluation period, data concentration ellipses are relatively wide, at each time slice is positioned near the centroid position for the six WineSips, and overlap completely throughout the evaluation. In conclusion, the partial bootstrap provides exactly the output that would be expected for data in which the sample labels are permuted.

### 3.3. Contrails

Contrails show the uncertainty surrounding the data. The contrails in Fig. 2 provide clear evidence that the panel discriminates L from H and A, but it would be difficult to determine whether there are systematic differences between H and A. The sip-to-sip variability that was noted as a potential trend seems to fall within the zone of uncertainty shown by the contrails. Thus, while a future study could be designed to investigate sip effect, the contrails shows that in this experiment, the sip-to-sip differences are not sufficiently pronounced to make strong conclusions regarding such an effect.

Smoothing was conducted (Fig. 2). Its purpose is to avoid overfitting and remove momentary fluctuations in characterizations that distract from the overall temporal pattern. Smoothing has the potential of fitting local segments of data very poorly; intervention (i.e., adjusting smoothing parameters) is sometimes required to ensure that the smoothed curves that are produced are reasonable for the data. Overlaying the smoothed trajectories on the contrails (Fig. 2) reveals that each trajectory passes approximately through the centre of its contrail. The overlay of the trajectory on the contrail provides some assurance that smoothing has been conducted in a manner that summarizes the data well.

### 3.4. Data concentration ellipses

In our manuscript, we obtain data concentration ellipses based on data from the real panel and the (in this case, 499) virtual panels obtained from resampling. In both cases the suitability of distributional assumptions can be subjected to visual assessment. If the response pattern for some assessors is very different from other assessors, the resulting bootstrap scores might be skewed or include influential outliers, and more robust method can be used. Furthermore, it is possible to represent bivariate data nonparametrically via the bagplot (Rousseeuw, Ruts, & Tukey, 1999) or convex hull (De Berg, van Kreveld, Overmars, & Cheong, 2008), which can be applied to bootstrap scores at individual time slices. Alternatively these nonparametric approaches can be used to represent the contours of the contrail bootstrap scores across all time slices.

Husson et al. (2005) obtain 95% confidence ellipses from Hotelling’s $T^2$ distribution based on projections into the multivariate sensory space of either raw scale responses from the assessors, or raw scale responses from a virtual panel obtained from a resampling of the assessors. The size of the confidence ellipse is thus related to the size and precision of the panel, and indicate the uncertainty for the panel mean intensity values arising from descriptive sensory data, as represented in a multivariate sensory space. By contrast, our data consists of projections of proportions (from the real and virtual panels), and the 95% data concentration ellipses enclose 95% of the data (bootstrap scores). These data concentration ellipses can be used to investigate uncertainty within and differences between treatments at particular time slices. Evidence of a significant difference among treatments is indicated by non-overlapping ellipses. (The same approach is appropriate generally, e.g., for comparing sips or other trajectories.) Clearly this investigation will result in many pairwise assessments at many time slices. As the procedure is generally used for exploratory data analysis and hypothesis generation, we do not advise adding any correction for multiplicity.

Data concentration ellipses in Fig. 3 provide no evidence that the panel is able to discriminate the first and second sips within any of the wine treatments at 30 s. Data concentration ellipses shown in Fig. 4 indicate that the panel separated the second sips of H and A at 30 s at a coverage level above 40%. Note that the approach of using overlap in data concentration ellipses as the criterion for concluding there are treatment differences treats the data as if the responses for a given treatment from a particular (virtual) panel is independent from the responses given by the same panel for a different virtual panel.

### 3.5. Data concentration ellipses for differences in trajectories

In Section 2.6, differences in scores are used to compare two treatments. The differences in scores at a particular time slice can be assessed using a single ellipse, which generally has a volume and shape that is smaller than the data concentration ellipses obtained as if all bootstrap scores are independent. An ellipse that excludes the origin indicates a significant difference of the PCs in that plane. In Fig. 5, the panel separated the second sips of H and A at a coverage level at nearly 95%, which is a clear improvement over the separation of these same treatments provided in Figs. 3 and 4. As in Section 3.4 we provide no correction for multiplicity for such assessments.

### 3.6. Animated sequences

A shortcoming of the data concentration ellipses presented in Fig. 3 is that it shows a static cloud around the trajectories. Static contrails (Fig. 2) provide no indication of uncertainty related to the individual time slices. For example, it is not necessarily clear whether two apparently overlapping trajectories are undergoing the same evolution of attributes at the same rate, or at different rates.
Sequential time slices provide static plots which can then be inspected. If animated, as shown in the Video 1 that accompanies the online version of this manuscript, then these static plots can be inspected rapidly. For example, the Video 1 has a duration that matches the original evaluation time; playback is available at a temporal pace equivalent to the original evaluation timeline. It is possible to pause playback to inspect individual frames at particular time slices, and subject these frames to more detailed investigation. Overall, the analysis summarizes how specific characterizations of the Syrah wines increase, plateau, and decrease over time. Trajectories that elongate rapidly (e.g. at the outset of the evaluation) are indicative of rapid sensory changes, whereas trajectories that progress more slowly (e.g. trajectories that return more gradually to their starting point near the end of the evaluation) are indicative of slower sensory changes. The start and the end of the evaluation coincide with very low citation proportions in PC1, as noted earlier, are here observed to coincide in the plane of PC2 and PC3 with relatively small ellipses sizes. Such visualizations of uncertainty provide context and nuance that would otherwise be lost, for example, if aggregating data into time periods.

Results presented in Section 2.7 and in Video 1 further demonstrate that although PC2 and PC3 account for relatively low proportions of variance in these data, the information present in these dimensions is signal, and not merely noise.

3.7. Comparison with finding of Baker et al. (2016)

Data presented in Section 2.1 were first analyzed by Baker et al. (2016). Data were preprocessed using time standardization (Lenfant et al., 2009), and then using attribute citation proportions on a [0, 1] scale divided into three (time-standardized) periods: beginning, middle, and end. The division into three time periods brings communication benefits because the wine industry conventionally discuss wine as having a three-part finish (or as having a short, medium, or long finish). But there are also drawbacks: discretization loses the temporal continuity, and time standardization loses the units of the original time scale (seconds). After preprocessing, data are analyzed by Cochran’s Q test (Cochran, 1950; Meyners et al., 2013), which confirms, e.g., that A is characterized as Sour more often than H in the middle period of the evaluation.

Data are also submitted to PCA to explore the temporal progression (see: Baker et al., 2016, Fig. 1). (The PCA output could be supplemented with data concentration ellipses using the approach presented in the current manuscript.) The temporal progression found in Baker et al. (2016) are broadly in agreement with observations regarding wine treatments made here; however, the temporal progression is much more disjointed compared to the trajectories presented herein (e.g., Figs. 1 and 2) which illustrate clearly the temporal evolution. Nonetheless, the data analyses presented in the two manuscripts could be considered complementary, with the analyses here emphasizing exploratory data analysis, and the approaches of Baker et al. (2016) providing time standardized and discretized view on the same data. Findings can help to inform winemaking decisions related ethanol management, by providing a broader understanding of the impact on sensory perception of Syrah wine finish resulting from fermentation of a must at different starting sugar levels vs. modification of ethanol post fermentation. As noted by Baker et al. (2016), there is a requirement for further research that connects these findings with consumer acceptability and consumer perception of wine quality.

4. Conclusions

In this paper we contribute several ideas for analyzing and interpreting TCATA data. In Section 2.2, we show an analysis of TCATA data based on PCA on the covariance matrix, along with trajectories arising from this analysis. In Section 2.3, we resample TCATA assessors with replacement to create virtual panels, and use their citation proportion data to obtain bootstrap scores that are projected into the multivariate sensory space obtained from the PCA solution for the real panel. In Section 2.4, we introduce contrails to summarize the changes in a treatment over time. TCATA trajectories can be smoothed, and the adequacy of the smoothing can be determined by overlaying the trajectory over its contrail. Separation of contrails from two different treatments provides evidence that the trajectories differ. In Section 2.5, data concentration ellipses are overlaid on bootstrap scores on a plane showing two principal components at a particular time slice, where ellipses summarize the region of uncertainty for the panel data for each treatment. Separation of the ellipses are indicative of treatment differences. In Section 2.6, differences in scores between two treatments are obtained, with bootstrap scores from the same panel are treated as matched data. A single data concentration ellipse can be overlaid; exclusion of the origin from the ellipse indicates a difference between the treatments in the plane of principal components at the coverage level used to construct the ellipse. In Section 2.7, an animated sequence provides a useful summary of changes in treatment characterization over time; the resampled data that is projected into the sensory space is shown at each time slice.

Given that ethanol concentration was manipulated systematically in a designed experiment, it is possible to understand the temporal progression from the context of sugar and ethanol concentrations at fermentation and post-fermentation, respectively. These analyses provide new perspectives on the results presented by Baker et al. (2016), which due to time-standardization cannot be linked directly to time on the original time scale. Communication of results with winemakers could be facilitated by providing wines for evaluation by mouth concurrent with playback of a video that describes the temporal progression of sensations in the finish (e.g. Video 1 that accompanies the online version of this manuscript).

Contrails and animated sequences discussed herein can be applied to data arising from other temporal sensory methods (e.g. TDS) to facilitate interpretation of data. Additionally, the visualization approaches described can be adapted to results obtained in sensory spaces obtained from correspondence analysis, factor analysis, and other approaches for multivariate data analysis.

Acknowledgements

This work was presented at the 14th edition of the AgroStat Symposium on Statistical Methods for the Food Industry in Lausanne, Switzerland in March 2016. An earlier version of this work was presented at the 1st AfroSense conference in Stellenbosch, South Africa in November 2015. We thank the conference participants and the reviewers of this manuscript for constructive feedback.

Appendix A

Glossary of terms

- Active object – an object which participates in the establishment of principal components. See also: passive objects.
- Animated sequence – a dynamic representation of changes in sensory characterization over time. In this manuscript, sensory characterization is visualized in plane of two principal components. Trajectories are shown leading up to the current time slice,
and bootstrap scores are displayed and overlaid by data concentration ellipses. Adjacent time slices are shown in sequence. The frame rate is set to show observed sensory changes in real time, creating the illusion of motion.

**Attribute citation proportion** – the arithmetic average number of citations for an attribute. (E.g. if an attribute is cited in 45 out of 60 opportunities, then its average citation proportion is 0.75.) It is usually calculated independently at many time slices during a TCATA evaluation.

**Bootstrap scores** – scores from the 1 real panel and all virtual panels. Called a contrail if visualized for all time slices simultaneously.

**Citation rate** – less formally, the attribute citation proportion.

**Contrail** – a cloud of bootstrap scores across all time slices, which is used to investigate uncertainty in the associated trajectory, which is often overlaid. (A contrail is a condensation trail created by airplanes due to such factors as water vapour present in engine exhaust and localized pressure changes. The name was selected due to visual similarity.)

**Coverage level** – determines the proportion of data enclosed by a data concentration ellipse. (E.g. a 95% data concentration ellipse encloses 95% of the data.)

**Data concentration ellipse** – probability contours for bivariate data based on the multivariate t-distribution. See: coverage level.

**Difference curve** – differences between two sets of TCATA curves obtained for two products.

**Difference ellipse** – a data concentration ellipse for a set of coordinates that are obtained from differences in scores.

**Differences in scores** – differences between dependent scores in the relevant dimensions, e.g., principal components. (E.g. If data \( x_1 \) and \( x_2 \) arise from a panel that evaluates both A and B, from which scores \( z_A \) and \( z_B \) are obtained, then the differences in scores \( z_A - z_B \).) See also: difference ellipse; difference trajectory.

**Difference trajectory** – a trajectory for a set of coordinates that are differences in scores.

**Passive object** – an object which is projected into a sensory space that was obtained from data from other, active objects. See: projection.

**Projection** – coordinates that position an object in a sensory space, usually obtained from a multivariate analysis involving dimension reduction. (E.g. In principal component analysis, these coordinates are called scores.) Data that are projected can be from an active object or from a passive object.

**Raw trajectory** – a trajectory in which coordinates for adjacent time slices are connected using straight lines. See also: smoothed trajectory.

**Smoothed trajectory** – a trajectory in which coordinates for adjacent time slices are submitted to an approximating function that removes fluctuations in data so that broader trends can be emphasized. See also: raw trajectory.

**Scores** – coordinates of an object in the new space defined by principal components. See also: bootstrap scores.

**Temporal check-all-that-apply (TCATA)** – a temporal sensory method used to investigate the temporal evolution of sensory attributes in products under evaluation. For details, see Castura et al. (2016).

**TCATA curve** – a line that shows the changes in one attribute citation proportion over time. Usually multiple (smoothed) curves are presented in a single plot, permitting review of how attributes characterize the product over time.

**Temporal evolution** – see: temporal progression.

**Temporal progression** – dynamic changes that are observed during the course of a temporal sensory evaluation. When applied to a particular temporal sensory method, this term refers to the specific aspects of the dynamic changes that are observed within the constraints of the temporal sensory method. (E.g. in TCATA studies the temporal progression refers to changes in how the product is characterized over time, whereas in a single-attribute time intensity study the temporal progression refers to changes in the intensity of a single attribute over time.) Also referred to as temporal evolution.

**Time slice** – a point in time at which the state of the temporal progression is evaluated. Calculation of attribute citation proportions and similar operations are conducted at each time slice. Time slices are (usually) separated by very short time intervals (e.g. every 0.1 s).

**Trajectory** – a curve showing the temporal progression of a product within a sensory space obtained from multivariate exploratory data analysis. (The term is retained for continuity in the literature.)

**Virtual panel** – a panel obtained from a partial bootstrap procedure. Panelists from the real panel are selected with replacement to become members of the virtual panel, which is usually the same size as the original panel.

References

Ares, G., Jaeger, S. R., Antúnez, L., Vidal, L., Giménez, A., Coste, B., et al. (2015). Comparison of TCATA and TDS for dynamic sensory characterization of food products. Food Research International, 78, 148–158.

Baker, A. K., Castura, J. C., & Ross, C. F. (2016). Temporal check-all-that-apply characterization of Syrah wine finish. Journal of Food Science. 81, S1521–S1529.

Boinbaser, L., Parente, M. E., Castura, J. C., & Ares, G. (2015). Dynamic sensory characterization of cosmetic creams during application using temporal check-all-that-apply (TCATA) questions. Food Quality and Preference, 45, 33–40.

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

Castura, J. C., Antúnez, L., Giménez, A., & Ares, G. (2016). Temporal check-all-that-apply (TCATA): A novel temporal sensory method for characterizing products. Food Quality and Preference, 47A, 79–90.

Cochran, W. G. (1950). The comparison of percentages in matched samples. Biometrika, 37, 256–266.

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

De Berg, M., van Kreveld, M., Overmars, M., & Cheong, O. (2008). Computational geometry: Algorithms and applications (3rd ed., pp. 2–8). Heidelberg: Springer-Verlag.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. Annals of Statistics, 7, 1–26.

Fischer, U., & Noble, A. C. (1994). The effect of ethanol, catechin concentration, and pH on sourness and bitterness of wine. American Journal of Enology and Viticulture, 45, 6–10.

Fontion, H., Saucier, C., Teissedre, P.-L., & Glories, Y. (2008). Effect of pH, ethanol and acidity on astrignency and bitterness of grape seed tannin oligomers in model wine solution. Food Quality and Preference, 19, 286–291.

Fox, J., & Weisberg, S. (2011). An R Companion to Applied Regression (2nd ed., CA: USA: Thousand Oaks. Sage. URL: http://socserv.socsci.mcmaster.ca/jfox/Books/Companion.

Friendly, M., Monette, G., & Fox, J. (2013). Elliptical insights: Understanding statistical methods through elliptical geometry. Statistical Science, 28, 1–39.

Greenacre, M. (2006). Tying up the loose ends in simple correspondence analysis. Economics Working Papers 940. Department of Economics and Business, Universitat Pompeu Fabra.

Husson, F., Lé, S., & Pagès, J. (2005). Confidence ellipse for the sensory profiles obtained by principal component analysis. Food Quality and Preference, 16, 245–250.

Johnson, R. A., & Wichern, D. W. (2007). Principal component analysis. In Applied multivariate statistical analysis, pp. 430–480. Upper Saddle River, NJ: Pearson Prentice Hall.

Legendre, P., & Legendre, L. (2012). Ordination in reduced space. In Numerical Ecology, pp. 425–521. Amsterdam: Elsevier Science BV.

Lenfant, F., Loret, C., Pineau, N., Hartmann, C., & Martin, N. (2009). Perception of oral food breakdown. The concept of sensory trajectory. Appetite, 52, 659–667.

Martin, S., & Pangborn, R. M. (1970). Taste interaction of ethyl alcohol with sweet, salty, sour, and bitter compounds. Journal of the Science of Food and Agriculture, 21, 653–655.

Meyners, M., & Castura, J. C. (2015). Best practice recommendations for attribute order in CATa and related methodologies. Food Quality and Preference, 48, 210–215.

Meyners, M., Castura, J. C., & Carr, B. T. (2013). Existing and new approaches for the analysis of CATa data. Food Quality and Preference, 30, 309–319.
Oliveira, D., Antúnez, L., Giménez, A., Castura, J. C., Deliza, R., & Ares, G. (2015). Sugar reduction in probiotic chocolate-flavored milk: Impact on dynamic sensory profile and liking. Food Research International, 75, 148–156.
Pineau, N., Schlisch, P., Cordelle, S., Mathonnière, C., Issanchou, S., Imbert, A., et al. (2009). Temporal dominance of sensations: Construction of the TDS curves and comparison with time–intensity. Food Quality and Preference, 20, 450–455.

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.
Rousseeuw, P. J., Ruts, I., & Tukey, J. W. (1999). The bagplot: A bivariate boxplot. The American Statistician, 53, 382–387.
Tucker, L. R. (1966). Some mathematical notes on three-mode factor analysis. Psychometrika, 31, 279–311.