Using partial bootstrap to evaluate the uncertainty associated with TCATA product trajectories

John C. Castura[1], Allison K. Baker[2] & Carolyn F. Ross[2]

[1] *Compusense Inc., 255 Speedvale Ave. W., Guelph, Ontario N1H 1C5, Canada*
E-mail : *jcastura@compusense.com*

[2] *School of Food Science, Washington State University, Pullman, Washington, USA*

Abstract
Temporal Check-All-That-Apply (TCATA; Castura et al., 2016) is a temporal sensory method in which assessors track changes in the applicability of sensory attributes to describe a sample during an evaluation. Data provide information on the complex dynamic profile of products. TCATA curves can be used to show attribute citation proportions over time, or differences in citation proportions between pairs of products. Exploratory data analysis techniques such as correspondence analysis or principal component analysis can uncover underlying structure in the data. Using such analyses, it is possible to show product trajectories, which summarize the different changes that occur in the products over time. Trajectories can be smoothed to avoid overfitting and reveal underlying patterns in the data. Now, resampling techniques are proposed to further investigate the quality of product trajectories. Following an approach similar to Husson et al. (2005) and Cadoret et al. (2009), we use the partial bootstrap to obtain virtual panels. Each virtual panel is identical in size to the original panel but comprised of assessors sampled with replacement. New product coordinates, based on data from these virtual panels, are projected into the sensory space. Resampled data from the virtual panels are shown at each time slice along the product trajectories. These sensory changes are visualized using an animated sequence, with the accumulation forming product contrails. These contrails reveal the variability associated with each product trajectory, and aid in interpretation by illustrating the uncertainty associated with estimates. The proposed approach is illustrated using empirical data from a trained red wine panel in a recent experiment conducted by Baker et al. (2015). Two red wine treatments varying in ethanol content, Low (“L”) and High (“H”), with 10 and 15.5% v/v, respectively, were produced from musts with different initial sugar contents (21 and 27 °Brix, respectively). Post-production, L was manipulated to develop a third wine (Adjusted, “A”) with the same ethanol concentration as H. A trained panel (n=13), using a consensus vocabulary and having practiced the TCATA methodology, used TCATA methodology to characterize the flavours, tastes, and mouthfeels that linger and evolve after swallowing red wine. The panel evaluated the 3 wine treatments in two sips and in quadruplicate. Investigation using the partial bootstrap technique helps to avoid potential over-interpretation and provides a greater level of confidence in the conclusions obtained from the product trajectories.
Keywords: Temporal Check-All-That-Apply; TCATA; contrails; data concentration ellipses; virtual panels

1. Introduction

Temporal Check-All-That-Apply (TCATA; Castura et al., 2016) is a temporal sensory method in which assessors characterize the sample by ensuring that at each moment the checked attributes characterize the sample being evaluated. Multivariate data analyses can be useful in providing a lower dimensional summary that increase interpretability, objectives that can be met by obtaining product trajectories from principal component analysis (PCA) on citation proportions for products on the attributes at multiple time slices, and joining adjacent time points. This presentation deals with how to understand these trajectories in the context of the uncertainty associated with TCATA data.

2. Exploratory data analyses

2.1 Data

Data discussed in this presentation came from a study discussed by Baker et al. (2015; 2016 submitted). Grapes were grown in the Columbia Valley, and used in winemaking on a pilot plant scale to create 3 wine treatments. Two red wine treatments varying in ethanol content, Low ("L") and High ("H"), with 10 and 15.5% v/v, respectively, were produced from musts with different initial sugar contents (21 and 27 °Brix, respectively). Post-production, L was manipulated to develop a third wine (Adjusted, "A") with the same ethanol concentration as H. A trained panel (n=13) was recruited, selected, and trained at Washington State University. The panel used a consensus vocabulary and practiced the TCATA methodology, then used TCATA methodology to characterize the flavours, tastes, and mouthfeels that linger and evolve after swallowing red wine. The panel used 9 sensory attributes, plus the attribute *Other* (not included in the analyses described herein). The panel evaluated the 3 wine treatments in two sips and in quadruplicate. The evaluation protocol for tasting had assessors sip the wine at 0 s, and expectorate and commence the evaluation at 10 s. The evaluation ended at 180 s.

2.2 Trajectories

Data were analyzed using PCA on the covariance matrix. Scores from adjacent time slices were joined to create raw trajectories for each treatment and sip (e.g. H1 refers to sip 1 of treatment H, whereas H2 refers to sip 2 of treatment H). Trajectories were then smoothed. The first two PCs are shown in Fig. 1.

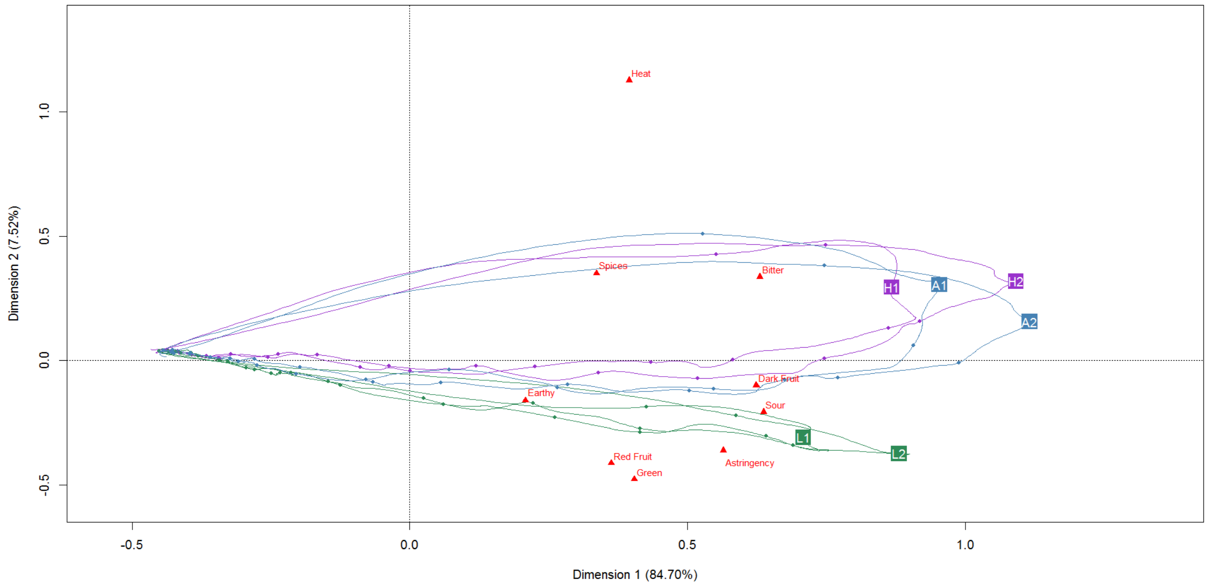

Fig. 1. Smoothed trajectories for 2 sips for each of 3 wine treatments. The trajectories shown move in a clockwise direction, starting at and returning to the point at the far left, corresponding to a citation rate of zero for all attributes. Trajectories are all labelled at the 30-s time slice.

2.3 Virtual panels

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 real panel (n=13). (In this manuscript, 499 virtual panels were used; note that results herein are unchanged provided that a sufficiently large number of virtual panels is used.) Based on PCA from the real panel, scores are obtained from the citation proportions for the virtual panel, and projected into the space. The approach is similar to the one used by Husson et al. (2005) and Cadoret (2009) for sensory data arising from static sensory evaluation methodologies.

2.4 Contrails

A cloud of points from the virtual panels is visualized. (The name "contrail" was given due to the similarity with the condensation trails created by airplanes due to water vapour and localized pressure changes.) Each contrail is overlaid with the corresponding trajectory, providing assurance that the smoothed trajectories summarizes the data well (Fig. 2).

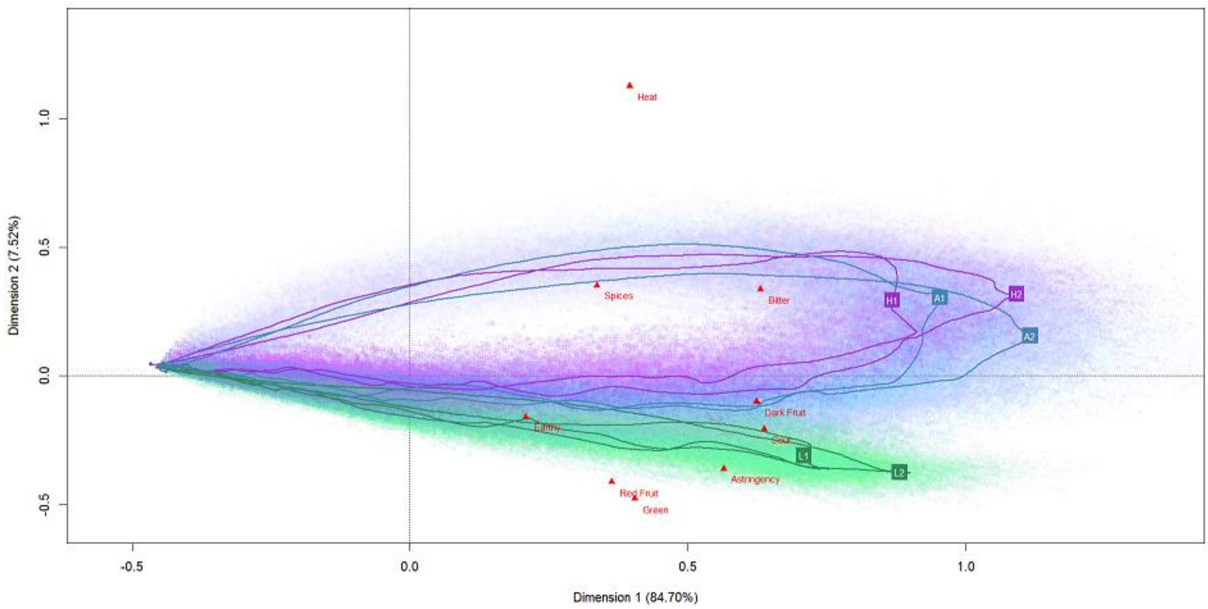

Fig. 2. Smoothed trajectories overlaid on their respective contrails

2.5 Data concentration ellipses

Data concentration ellipses (Fox et al., 2011; Friendly et al., 2013) can be established for the virtual panel's scores for each product and sips at a given time slice such that, under the assumption of bivariate normality, the ellipse covers a pre-specified percentage of the projected scores from the virtual panels.

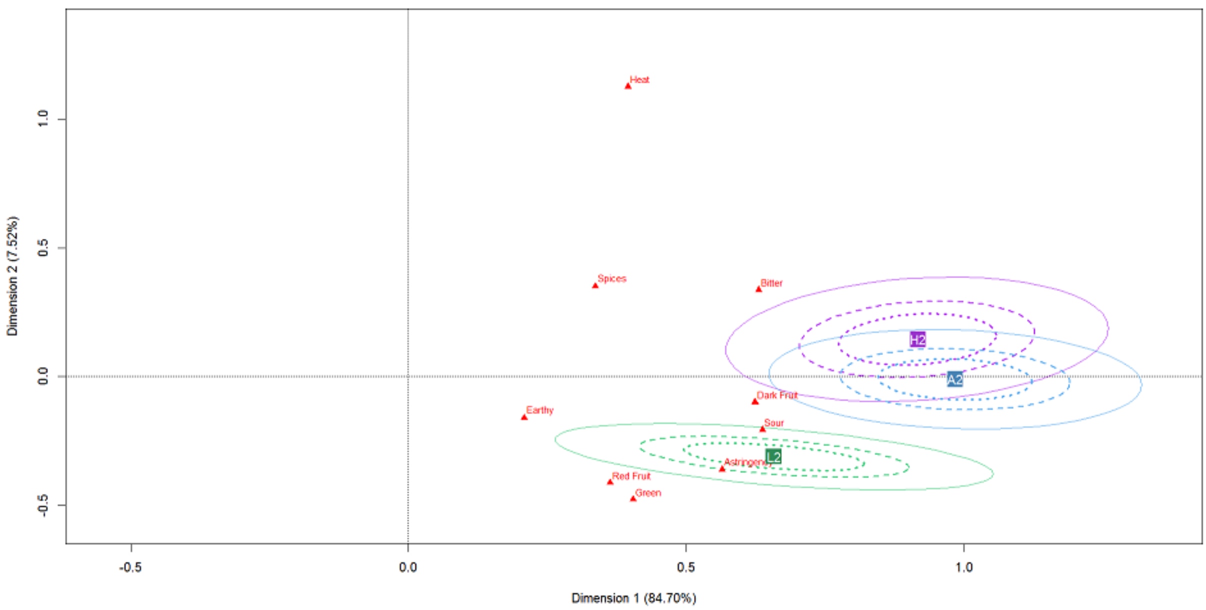

Fig. 3. Data concentration ellipses for Sip 2 data for each of the three wine treatments. The 40%, 68%, and 95% data concentration ellipses are indicated by the dotted inner ellipses, dashed middle ellipses, and solid outer ellipses, respectively.

2.6 Evaluating differences in trajectories

Data from a particular virtual panel are obtained for multiple products. The difference in the projected scores from the virtual panels can be obtained. Data concentration ellipses can be overlaid, such that a 95% of the difference coordinates are enclosed. Exclusion of the origin indicates that the difference between the products is beyond that which might be expected by chance alone.

2.6 Animated contrails

All of the previous plots are static, and either show the uncertainty in the data across all time slices simultaneously (Section 2.4) or show the uncertainty in the data at a particular time slice (e.g. Section 2.5 or Section 2.6). Animated contrails permit a dynamic representation of the products over time. For these data, a sequence of time slices are shown. For each time slice and Product*Sip, the scores associated with the virtual panel data are shown, and overlaid by a 95% confidence ellipsis. The trajectories leading up to the most recent time slice are shown. By allowing for both real-time review of the TCATA evaluation, and also for the animation to be paused, it becomes possible to explore how the products are characterized over time. Such review can further increase interpretability, and promote hypothesis generation.

References

Baker, A. K., Castura, J. C., & Ross, C. F. (2015). Temporal check-all-that-apply characterization of wine finish. In *11th Pangborn Sensory Science Symposium*. 23-27 August. Gothenburg, Sweden. (Poster).

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*, 47, 79–90.

Castura, J. C., Baker, A. K., & Ross, C. F. (2016).s Using contrails to visualize uncertainty in dynamic sensory profiles obtained from Temporal Check-All-That-Apply (TCATA) data. *Food Quality and Preference*, Submitted.

Fox, J., & Weisberg, S. (2011). *An R Companion to Applied Regression*, Second Edition. Thousand Oaks, CA, USA: 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.

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