# Monitoring panel performance within and between sensory experiments by multi-way analysis

R. Romano, J. S. Vestergaard, M. Kompany-Zareh, W. L. P. Bredie

**Abstract** In sensory analysis a panel of trained assessors evaluates a set of samples according to specific sensory descriptors. The training improves objectivity and reliability of assessments. However, there can be individual differences between assessors left after the training that should be taken into account in the analysis. Monitoring panel performance is then crucial for optimal sensory evaluations. The quality of the results is strongly dependent on the performance of each assessor and of the panel as a whole. The present work proposes to analyze the panel performance within single sensory evaluations and between consecutive evaluations. The basic idea is to use multi-way models to handle the three-way nature of the sensory data. Specifically, a PARAFAC model is used to investigate the panel performance in the single experiment. N-PLS model is used to test the predictive ability of the panel on each experiment. A PARAFAC model is also used for monitoring panel performance over different experiments.

## 1 Introduction and data description

The present work investigates panel performance in sensory experiments from a project considering organic milk at the University of Copenhagen. One of the objectives of the project is to establish knowledge about production of high quality organic milk. Two different sensory experiments were conducted consecutively in 2007: the first in spring and the second in autumn. Milk samples from seven different farms representing two different breeds (Holstein-Friesland and Jersey) were analyzed by sensory descriptive analysis (Lawless, Heymann, 1998) (12 attributes in the spring experiment and 16 attributes in the autumn experiment). Information

Rosaria Romano, Jannie S. Vestergaard, Wender L. P. Bredie
University of Copenhagen, Denmark, e-mail: rro@life.ku.dk, jve@life.ku.dk, wb@life.ku.dk

Mohsen Kompany-Zareh
Institute of Advanced Studies in Basic Sciences, Iran, e-mail: kompanym@iasbs.ac.ir

on milk production was also provided. The two experiments presented small differences: in spring only 6 samples were evaluated; the panel in both experiments included 9 assessors but some of them differed from one experiment to another.

In the present work, multi-way models (Smilde, Bro, Geladi, 2004) are used in order to evaluate the panel performance within and between the experiments. First, focus is given on each experiment separately: a) the PARAllel FACtor (PARAFAC) model is used to investigate individual differences between assessors; b) the N-way Partial Least Squares (N-PLS) model is used to test the predictive ability of the panel. Then, the model from one experiment is tested using data from the other experiment to investigate the performance of the panel as a whole over the time.

## 2 Modeling assessors' performance by PARAFAC

PARAFAC is a generalization of PCA to higher order arrays. Let $\underline{\mathbf{X}}$ be the three way array holding the scores $x_{ijk}$ given by $K$ assessor, on $I$ products, according to $J$ attributes. The model can be written as:

$$x_{ijk} = \sum_{f=1}^{F} a_{if} b_{jf} c_{kf} + e_{ijk} \tag{1}$$

where $a_{ij}$, $b_{jf}$, and $c_{kf}$ are the elements of the loading matrices **A**, **B** and **C**. The solution to the model can be found by the Alternating Least Squares (ALS), minimizing the sum of squares of the residual $e_{ijk}$.

Using PARAFAC, variation in the products space and in the assessors space can be modeled at the same time. PARAFAC permits to investigate individual differences in sensitivity, reproducibility, and consistency (Bro *et al.*, 2007).

## 3 Modeling panel predictive ability by N-PLS

N-PLS model is a straightforward extension of the bi-linear PLS regression in case of higher order arrays. Focus here is on the tri-linear PLS, where the explanatory variables are collected in a three way array $\underline{\mathbf{X}}$ (IxJxK) and the dependent variables in a two-way array **Y** (IxM). The algorithm aims at decomposing the cube X into a set of triads satisfying a certain criterion. A triad consists of one score vector (**t**), one weight-vector ($\mathbf{w^j}$) one the second order, and one weight-vector ($\mathbf{w^k}$) on the third order. In case of one dependent variable $y$, the algorithm find the vectors ($\mathbf{w^j}$) and ($\mathbf{w^k}$) that satisfies:

$$\max_{w^j w^k} \left[ cov(t,y) | min \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (x_{ijk} - t_i w_j^j w_k^k)^2 \right] \tag{2}$$

In presence of several dependent variables it is possible to use the algorithm in (**??**) to model each dependent variable separately. Alternatively, it is possible to model all the variables simultaneously as in the PLS2 algorithm.

N-PLS is used for the prediction of production data from sensory data. The aim is to test the predictive ability of the panel in the two experiments.

## 4 Modeling panel performance between experiments

The basic aim here is to compare panel performance in the two experiments. As discussed in (**??**), panel compositions are not exactly the same. However, it is possible to consider the panel as a whole. Under this assumption, a PARAFAC model with two components is performed on the autumn experiment data and used to predict the spring data. Residuals from this model are then compared with residuals from PARAFAC with two components on the spring data. If residuals from application of PARAFAC on spring data are very low as compared to the residuals from the prediction, then considerable differences between the two evaluations exist.

## 5 Results and conclusions

Results from PARAFAC model with two components in the two experiments are shown in Fig. **??** and Fig **??**. For sake of space only loadings from assessors' mode are presented. In spring, the first component shows a good agreement of the panel but different sensitivities: assessors 8, 1 and 5 are the most sensitive. On the second factor it seems there is not consistency as the assessors are divided into two groups. Similar results in autumn: good panel agreement on first component with differences in sensitivity and not consistency on the second component. Here there is also disagreement as some assessors have opposite loadings with respect to the panel. Residual analysis provides results on the variability of each assessor with respect to single attributes or over all the attributes together. A comparative analysis of the results in both experiments shows that the panel in autumn performed better than in spring: in autumn there was a group of good assessors, whereas in spring only a reduced number. The improvement can be due to a better panel performance (training effect) but also to differences in the samples in the two experiments.

Results from N-PLS in spring shows that there is no linear relation between the actual values and the values estimated by cross validation, for any number of components and for both dependent variables. In autumn a slight linear trend is observed for $y1$ and a clear linear relation for $y2$. Hence, the panel predictive ability in the autumn experiment is better than in spring. However, it must be stressed that more information was provided in the autumn experiment (4 additional attributes).

Residuals from PARAFAC on the autumn experiment for the prediction of the spring experiment and residuals from PARAFAC on the spring data directly are

compared. The residuals with respect to the samples mode are not very different from each other. This shows the similarity between the autumn and the spring experiment, i.e. the spring experiment had a structure similar to the autumn experiment. There were some differences, but these may be due to the season effect as the two evaluations span from spring to autumn. Thus, the conclusion is that even the assessment in spring was much noisy and it was not possible to build a valid model due to lack of information, it provided a valid sensory evaluation.

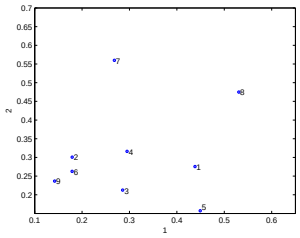


**Fig. 1** PARAFAC assessors' loadings in the spring experiment

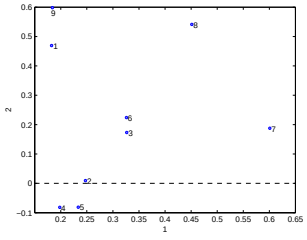


**Fig. 2** PARAFAC assessors' loadings in the autumn experiment

## Acknowledgments

This work is supported financially by the *Organic Milk of High Quality* project. The authors would like to thank Professor Rasmus Bro for his intellectual contributions.

## References

1. R. Bro, E. M. Quannari, H. A. L. Kiers, T. Naes, M. B. Frost (2007). Multi-way models for sensory profiling data. *J. Chemometrics*, 21, 1–10
2. H. T. Lawless, H.Heymann, *Sensory evaluation of food*, (Chapman / Hall, New York, 1998)
3. A. Smilde, R. Bro, P. Geladi, *Multi-way Analysis. Applications in the chemical sciences*, (New York: Wiley PL., 2004)