Proposal of statistical procedure to relate aroma chemistry data to aroma sensory data

Marie Auda[1], Nicolas Pineau[2], Frédéric Mestdagh[1], Luigi Poisson[1] & Andréas Rytz[2]

[1] *Nestlé Product Technology Centre Orbe, Switzerland*
E-mail : *marie.auda@rdor.nestle.com*

[2] *Nestlé Research Center, Lausanne, Switzerland*

Abstract

In many R&D projects, it is often desirable to relate the analytical data to the sensory profile. Depending on the type of analytical data, it is critical to account for their nature to fit as best as possible with sensory data. Focusing on the measurement of volatile compounds, it is known that directly relating the raw analytical data to aroma perception, as perceived by a sensory panel, is not meaningful. Indeed, whatever the analytical method used to measure the aroma composition and concentration, the nature of the raw data cannot directly reflect the impact of volatiles on the perceived aroma quality and intensity. In this regard, a novel approach of data standardization procedure is proposed to relate aroma chemistry data (volatiles) with aroma sensory data (profiles). This procedure is an adaptation of a method published in Lindinger et al. (2008).

Keywords: Aroma chemistry data, sensory data

1. Introduction

The aroma composition and concentration are both generally characterized by the use of analytical methods such as Gas chromatography-mass spectrometry (GC-MS) or Proton-Transfer reaction mass spectrometry (PTR-MS). Whatever the analytical method used to measure the aroma chemistry, the nature of the raw data cannot directly reflect the impact of volatiles on the perceived aroma quality and intensity as characterized by a sensory panel. There are three main reasons for that:

1) As described in the perception theory, the aroma concentration is not linearly related to the perceived aroma intensity (sigmoid relation). Some models in literature such as Fechner's (1877) or Stevens' power law (1957) recommend to log-transform the concentration data.
2) The absolute abundance of each volatile compound can impact differently the perceived aroma quality and intensity. There is a need to standardize the log-concentrations for each volatile in order to give them a proper weight.
3) Given the nature of the olfactory receptors in the nasal cavity and the treatment of this signal by the brain (e.g. "The neurobiology of olfaction", Menini, 2009), human nose seems more able to identify volatile molecule patterns due to relative abundance of these molecules rather than identifying aromas through direct relations to absolute (or even standardized) abundances.

It seems therefore necessary to account for these specificities by pre-processing the aroma chemistry data so that the relation with sensory data makes sense. In this regard, a novel approach of data standardization procedure is proposed to relate aroma chemistry data (volatiles) with aroma sensory data (profiles). This procedure is an adaptation of a method published in Lindinger et al. (2008). The sensory data do not require any transformation for this procedure.

2. Procedure to relate sensory & analytical data

The procedure is presented in two sections. The first section describes the transformation steps of the analytical data in order to obtain an estimate of the aroma pseudo-composition which is meaningful to relate to the sensory profiles. In the second section, a multivariate approach is suggested to relate the transformed analytical data to the sensory data.

2.1 Transformation procedure

Consider analytical data of the form X_{ij} with:

- $i = 1, \dots, n$ rows (products)
- $j = 1, \dots, p$ columns (compounds)

Raw data	Compounds 1	.	.	.	Compounds p
Product 1	x_{ij} for $i = 1, \dots, n$ & $j = 1, \dots, p$				
.					
.					
.					
Product n					

Step 1: Log-transformation. Taking into account the non-linear relationship between aroma concentration and perceived aroma intensity, the analytical data should be log-transformed. Before applying the logarithm, it is suggested to add 1 to each raw data to avoid that values inferior to 1 get an overestimated weight due to the log-transformation. Mean and standard deviation of the log-transformed data are then computed for each compound.

Log data	Compounds 1	.	.	.	Compounds p
Product 1	$y_{ij} = ln(x_{ij} + 1)$ for $i = 1, \dots, n$ & $j = 1, \dots, p$				
.					
.					
.					
Product n					
Mean	$\mu(y_{.1})$	.	.	.	$\mu(y_{.p})$
SD	$\sigma(y_{.1})$	.	.	.	$\sigma(y_{.p})$

Step 2: Normalization. The log-transformed data should then be standardized with respect to the mean and standard deviation in order to give the same weight to each compound (i.e. put all compounds on a similar scale of values).

Normalized data	Compounds 1	.	.	.	Compounds p
Product 1	$z_{ij} = \frac{y_{ij} - \mu(y_{.j})}{\sigma(y_{.j})}$ for $i = 1, \dots, n$ & $j = 1, \dots, p$				
.					
.					
.					
Product n					

Step 3: Pseudo-concentration. The pseudo-concentration v_i is the average of all standardized volatile compounds z_{ij} present in a product i which can impact the aroma perception. The aroma pseudo-concentration enables to compare the relative abundance of volatile compounds between products. It gives somehow an estimate of the overall aroma molecule load in the aroma fraction of the product.

Normalized data	Compounds 1	.	.	.	Compounds p
Product 1	z_{ij} for $i = 1, \dots, n$ & $j = 1, \dots, p$				
.					
.					
.					
Product n					

Pseudo-concentration
$v_1 = \mathbf{mean}(z_{1j})$
.
.
.
$v_n = \mathbf{mean}(z_{nj})$

Step 4: Pseudo-composition. The aroma pseudo-composition is obtained by subtracting the pseudo-concentration $\boldsymbol{v}_{\boldsymbol{i}}$ from each standardized compound data $\boldsymbol{z_{ij}}$ for $\boldsymbol{i = 1, \dots, n}$ & $\boldsymbol{j = 1, \dots, p}$. This transformation step gives an estimate of the relative abundance of each aroma compound with respect to the overall aroma fraction load

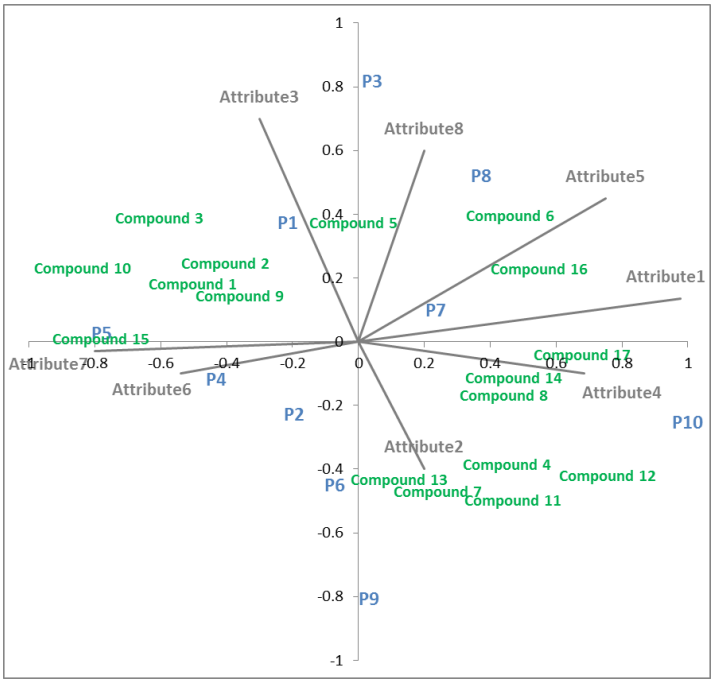
Pseudo-composition	Compounds 1	.	.	.	Compounds p
Product 1	$\boldsymbol{w_{ij} = z_{ij} - v_i}$ for $\boldsymbol{i = 1, \dots, n}$ & $\boldsymbol{j = 1, \dots, p}$				
.					
.					
.					
Product n					

The set of *"pseudo-compounds"* w_{ij} describe the aroma composition in each product in terms of balance of aromas which is more meaningful to relate to the sensory compared to aroma concentrations.

2.2 Procedure to relate transformed data and sensory

Various multivariate techniques can be used to relate the transformed analytical data to the sensory. Among them Principal Component Analysis (PCA) or Partial Least Square (PLS) regressions are very common approaches. The method proposed here is purely descriptive as shown on the plot below. The products are projected on a biplot using Principal Component Analysis (PCA) with the sensory data. Then the aroma pseudo-compounds are overlapped on the map as supplementary variables. This map is the geometrical representation of a Principal Component Regression (PCR). This approach has the advantage to give more importance to the sensory space enabling to identify which aroma pseudo-compounds correlate the best with key sensory dimensions.

Biplot



References

Fechner, G.T. (1877). In Sachen der Psychophysik [In the Matter of Psychophysics]. Leipzig: Breitkopf & Härtel.

Lindinger, C., Labbe, D., Pollien, P., Rytz, A., Juillerat, M. A., Yeretzian, C., & Blank, I. (2008). When machine tastes coffee: Instrumental approach to predict the sensory profile of espresso coffee. Analytical chemistry, 80 (5), 1574-1581.

Menini, A. (Ed.). (2009). The neurobiology of olfaction. CRC Press.

Stevens, S.S. (1957). On the psychophysical law. Psychological Review 64 (3): 153–181.