Peltier Caroline (Orcid ID: 0000-0002-7926-955X)

Automatic pre-treatment and multiblock analysis of flavor release and sensory temporal data simultaneously collected in vivo

Caroline Peltier[1,2], Michel Visalli[1,2], Hélène Labouré[1], Cantin Hélard[1], Isabelle Andriot[1,2], Sylvie Cordelle[1,2], Jean-Luc Le Quéré[1], Pascal Schlich[1,2]

[1]*Centre des Sciences du Goût et de l'Alimentation, CNRS, INRAE, Institut Agro, University of Bourgogne Franche-Comté, F-21000 Dijon*

[2]*CNRS, INRAE, PROBE research infrastructure, ChemoSens facility, F-21000 Dijon, France,*

caroline.peltier@inrae.fr

Abstract

Proton Transfer Reaction-Time-of-Flight- Mass Spectrometry (PTR-ToF-MS) is an analytical chemistry technique that can be used for measuring the concentration of volatile organic compounds directly in the subjects' noses (nosespace, in vivo analysis) during a tasting and over time. It can be combined with temporal sensory methods such as Temporal Dominance of Sensations (TDS) or Temporal Check All That Apply (TCATA) in order to obtain simultaneous sensory and physico-chemical signals.

This paper aims to provide a methodology to analyze in vivo PTR-MS and temporal sensory data and illustrate it on a real dataset.

First, relevant pretreatments of PTR-MS data were established, including breathing correction, blank periods removal and standardization. Then, a statistical multiblock analysis was presented: the Regularized Generalized Canonical Correlation Analysis (RGCCA). The versality of the approach was demonstrated, as it can be used to answer most of problematics (exploratory or supervised). Finally, this methodology is illustrated on a dataset of PTR-MS and TDS or TCATA data collected simultaneously. In this study, 16 semi-trained subjects

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/cem.3450

evaluated 3 chocolates in TDS and TCATA on six flavor attributes (*Spicy, Cocoa, Woody, Fruity, Roasty* and *Dry Fruits*) with 2 replicates for each sensory method. Results showed that TCATA and TDS gave similar results, but TDS was shown to slightly better preserve the PTR-MS observed product configuration than TCATA. All computing tools developed in this work are freely available.

Keywords: Flavor release, PTR-MS, TDS, TCATA, RGCCA

1. INTRODUCTION

Proton Transfer Reaction Time of Flight Mass Spectrometry (PTR–ToF-MS) is a powerful tool for fast, direct and highly sensitive online monitoring of volatile organic compounds (VOCs) that was extensively applied in food science (Biasioli et al., 2011). PTR–ToF-MS uses a soft chemical ionization based on proton transfer from a protonated reagent, most commonly H_3O^+. Sensitive online monitoring capabilities allow performing in vivo aroma release measurements such as nosespace analysis (Biasioli et al., 2011). The addition of a Time of Flight (ToF) mass analyzer to the PTR–MS adds the advantage of providing high mass range, very fast measurement and high mass resolution (Jordan et al., 2009).

After first pretreatments (such as mass recalibration, deconvolution of the peaks…), resulting data can be considered as ion trajectories along time. In vivo, these trajectories are deeply impacted by the breathing of the subject. However, usual procedures do not take the breathing into account and consist of calculating the area under the ion curves (AUCs), measuring the total release of the corresponding aroma molecules, without any further pretreatment. This could be problematic when pairing time-by-time instrumental and sensory measures.

Until now, only a few studies have used in combination sensory and instrumental real-time measurements to investigate aroma release and flavor perception of several sensory descriptors. Déléris et al. (2011) and Mesurolle et al. (2013) studied respectively candies and yogurts with fruit pieces, focusing on the texture effect. Barron et al. (2012) and Charles et al. (2015) paired Temporal Dominance of Sensations (TDS, Pineau et al., 2009) and PTR-MS on espresso coffee for measuring the impact of foam/"crema" on the aroma release or for investigating the impact of roasting degree and sugar addition on aroma release and perception. These studies indicate the existence of some links between the dynamics of flavor perception and aroma release even though it remains challenging to relate these two types of data due to the variety of the phenomena involved (salivation, breathing, chewing, individual sensitivity to volatile and non-volatile compounds, etc.) (Charles et al., 2015). More recently, discontinuous time-intensity was combined with PTR-MS in order to investigate the impact of physiological parameters, ethnicity and gender on flavor perception and flavor release (on mint and sugar) of chewing gums (Pedrotti et al., 2019). As other recent combination of sensory and physico-chemical measurements, Eck et al. (2021) used PTR-MS and time-intensity on lemon aroma on mayonnaises spiked with lemon aroma, showing that when mayonnaises were combined with carriers, aroma release and perception were no longer positively correlated.

In all these studies but the two last ones, TDS was used for measuring real-time sensory responses. It consists in presenting a set of descriptors to the subjects on a computer screen and asking the subjects to select the dominant sensory descriptor (among the predefined list) along the tasting (Pineau et al., 2009). In 2016, Temporal Check All That Apply (TCATA, Castura et al., 2016) was presented as an alternative to TDS. In TCATA, subjects can select several

attributes at the same time and should unselect any attribute when it becomes non-applicable. Ares et al. (2015) compared TCATA and TDS methods through its application to products of varying complexity, using consumers and trained subjects, and concluded that TDS appears to decrease the level of detail in descriptions as well as the discrimination of products when several attributes are perceived, especially when dealing with multiple sensory modalities. Recently, Esmerino et al. (2017) stated in their study that the major part of the variation is common between the two methods, but that there were subtle differences showing better discrimination for TCATA than TDS.

The datasets obtained by pairing sensory and physico-chemical measurements are multiblock data: they are constituted of data from the same individuals (as rows) scoring (or being scored on) different sets of variables called “blocks” (whose columns are the variables). A natural way to analyze such data could consist to compute a PCA of the concatenated blocks, ignoring the block structure. However, this approach would favor the block with the highest variance. For example, if all the variables were scaled (as ions in usual PTR-MS analysis), the block with more variables would have more importance in the analysis than the other ones. Multiblock analyses were proposed to overcome this limitation. The Regularized Generalized Canonical Correlation Analysis (RGCCA) is a multiblock method that was introduced by Tenenhaus & Tenenhaus (2011) as a general framework for a large panel of multiblock methods. It was mainly used for the research of multimodal biomarkers as presented in Xicota et al. (2019) or Lejeune et al. (2022).

This paper aims to propose a methodology to relate in vivo PTR-MS data and temporal sensory data obtained by TDS or TCATA including pretreatment taking breathing into account and statistical analyses based on RGCCA.

After a presentation of the example dataset, relevant pretreatments of PTR-MS data were established, including breathing correction, blank periods removal and standardization. Second, a flexible statistical multiblock analysis was presented: Regularized Generalized Canonical Correlation Analysis (RGCCA). Finally, pretreatments and multiblock analyses were conducted on the example dataset in order to evaluate if PTR-MS data fits better TDS or TCATA data.

All computing tools developed in this work are available on www.github.com/ChemoSens.

2. MATERIAL AND METHODS

2.1. Description of the example dataset

2.1.1. Overview

The experiment took place in CSGA, Dijon, France. Three chocolates were tasted by 16 subjects with TDS and TCATA, while connected to a PTR-ToF-MS device, with 2 replicates. Consequently, each subject tasted the same products 4 times in two separated (TDS or TCATA) sessions. The order of sessions was counter-balanced. We used PTR-TDS (resp. PTR-TCATA) for referring the PTR-ToF-MS experimentation related to TDS (resp. TCATA). TDS and TCATA data were collected via the TimeSens software, developed in the laboratory of the authors.

2.1.2. Samples

Three commercial dark chocolates (hereafter named A, B and C) differing by their origins but also their percentages of cocoa (63% for A and B, 80% for C) were provided by Valrhona, Tain l'Hermitage, France. They were presented according to William's Latin squares (generated on

n=6=3 products*2 replicates) after a warm-up (neutral chocolate in the same format as A, B and C).

2.1.3. Subjects

An internal panel of 16 subjects (4 men and 12 women, 21-63 y.o.) was recruited. They were asked not to smoke, eat or drink coffee at least 1 h before the session. The subjects were familiarized to each method (TDS, TCATA) during two sessions (one for each method).

2.1.4. Descriptors

Six descriptors were selected for the temporal sensory task: *Cocoa, Fruity, Spicy, Woody, Roasted* and *Dry Fruits.* This list was obtained using the panel of 16 subjects presented in 2.1.3. with the following process. First, a free comment task on the chocolates led to an initial list of descriptors composed of 12 attributes. Then, a RATA (Rate-All-That-Apply) method was used on the same products (Ares et al., 2014). The subjects had to tick all the descriptors during the tasting, then to evaluate their intensity on a scale ranging from 0 to 5. The six selected descriptors had a high frequency of citations (higher than 40%), were significant for the product effect. *Spicy* corresponded to the notes of Coconut/Vanilla and the descriptor *Dry Fruits* regrouped nut, almond, hazelnut.

2.1.5. Instrumental conditions

The nosespace of the subject was done through a home-made Teflon nosepiece carried by a light helmet that connected both nostrils of the subject via a heated transfer line to the PTR-MS spectrometer equipped with a Time of Flight analyzer (PTR-ToF 8000, Ionicon Analytik GmbH, Innsbruck, Austria) with H_3O^+ as reagent ion.

Each PTR-MS acquisition corresponds to one nosespace analysis with the PTR-ToF-MS spectrometer and was carried out as follows:

- a blank period during which the breathing of the subject is measured before tasting (25-30 seconds): only the compounds present in the breath are detected.
- a tasting period beginning when the subject puts a product in his mouth and ending when he feels nothing anymore

After acquisition, PTR-MS Viewer software (ver. 3.2.8.0), a software developed by the PTR-ToF-MS manufacturer (Ionicon Analytik, Innsbruck, Austria), was used for preliminary pretreatment steps. They consisted in applying (i) mass recalibration (if necessary), (ii) deconvolution of the peaks, (iii) concentration calculation of each ion at each time, considering the transmission table of ions in function of the mass, and (iv) water clusters ions and ^{13}C isotopologues removal.

Nosespace sampling was performed at a total flow rate of 400 mL/min with the transfer line maintained at 110°C. PTR-ToF-MS parameters were as follows: drift pressure of 2.3 mbar, drift temperature of 80°C, drift voltage of 390 V resulting in an E/N value of 92 Td (1 Td = 10^{-17} V.cm^2). Mass spectra were acquired at a scan speed of 0.1 s for the mass range m/z 1-226. A total of 197 ions was measured.

2.2. Automatization of PTR data pre-treatment

The pre-treatments presented in this paper are based on the TXT output files produced by the software PTR-MS Viewer.

2.2.1. Overview

This paper presents a pipeline of pretreatments including several steps that are detailed in the next subsections and summarized in Figure 1. For each PTR-ToF-MS acquisition file, the breathing is detected on a breathing reference, then corrected for each ion (section 2.2.2). Then, the averaged intensity during the blank period is subtracted for each ion intensity (section 2.2.3). Then, the ions evolving during the tasting are selected (section 2.2.4.). At this stage, the results of whole acquisition files are aggregated in order to obtain a list of relevant ions common to all files (*final ion list*). Finally, the area under curves (AUCs) of each ion of the final ion list are calculated for each file and the data is standardized (section 2.2.5), then aggregated in a single dataset (final dataset).

2.2.2. Breathing correction

A breathing cycle is defined as an interval of time between two consecutive expirations and is calculated as following. For each sample, breathing cycles are detected with picking peaks algorithms on the data of a breath marker (so-called ***breathing reference***) using the MSnbase R package (Gatto & Lilley, 2012). First, the breathing reference curve is smoothed with Moving Average filter. Then a peak picking algorithm (described in the next paragraph) is run on the opposite of the breathing reference curve such as a peak represents an expiration. Only peaks whose intensities are higher than a given threshold (i_{min}) were selected. The breathing cycles are calculated as the intervals between two successive selected peaks. When a breathing cycle duration is lower than (d_{min}) seconds (corresponding to a very high breathing frequency), it is aggregated to the previous one.

Once the breathing cycles based on breathing reference have been obtained, the averaged intensity during each breathing cycle is calculated for each ion p, giving a vector I^p of size C

(C being the number of breathing cycles in the evaluation). As a final step, the averaged intensity by breathing cycle can be plotted against time. At this stage, the data are considered as freed from any breathing effect (steps 1 and 2 of Figure 1).

The peak picking algorithm used in breath cycle detection is the most commonly used peak detection method and is based on finding local maxima with moving windows (Gibb & Strimmer, 2017).

2.2.3. Subtracting the intensity during blank periods

After breathing correction, for each ion, the averaged intensity obtained during the blank period ($i^p_{blank}, scalar$) was subtracted from the signal ($I^p_{corrected} = I^p - i^p_{blank}$). After this correction, potential negative values are replaced by zeros (step 3 of Figure 1).

2.2.4. Selection of relevant ions

The PTR-ToF-MS device can measure all the ions present in the sample.

As all ions are not relevant, only the ions whose maximal value during the tasting is $n_{TimesHigherThanBlank}$ times higher than the maximal value during the blank period are selected (step 4 of Figure 1), that is to say when:

$$max(I^p_{corrected\_during\_tasting}) > n_{\mathrm{TimesHigherThanBlank}} * max(I^p_{corrected\_during\_blank})$$

This selection of ions is done for each evaluation (named an "evaluation-ion selection"). However, a list of ions common to all evaluations is mandatory for any multivariate analysis. Thus, a "panel-ion selection" gathering the ions present in all "evaluation-ion selections" from a given product is established. After this step, all ions in the panel-ion selection of at least one product are selected. Finally, the panel-ion selection for TDS and for TCATA sessions are

merged in order to form a final single list of K ions noted ***final ion list:*** *{sel_ion_1, …, sel_ion_K} (step 5 of Figure 1).*

2.2.5. Calculation of area under curves (AUCs) for the final ion list, standardization and aggregation

After the removal of the averaged intensity of the blank period, the areas under the obtained curves (AUCs) are calculated as the sum of cycle intensities weighted by the cycle durations, given one single value for each evaluation and each ion (step 6 of Figure 1). More formally, it can be written as:

$$AUC_{sel\_ion\_i} = \sum_{c=1}^{C} duration\_cycle[c] * I_{corrected}^{sel\_ion\_i}[c]$$

Then, the data are standardized to give the same weight to each evaluation in the analysis. Indeed, some subjects may keep the product longer in the mouth and consequently have higher AUCs. To solve this issue, each AUC is divided by the sum of AUCs, giving a total weight of 1 for each evaluation (step 7 of Figure 1). Thus, the standardized AUC of *sel_ion_i* ($std\_AUC_{sel\_ion\_i}$) can be obtained with:

$$std\_AUC_{sel\_ion\_i} = \frac{AUC_{sel\_ion\_i}}{\sum_{k=1}^{K} AUC_{sel\_ion\_k}}$$

With this calculation, the quantity information was lost: only relative quantities were kept.

Finally, all standardized AUCs are aggregated in a single table of standardized AUCs with as many lines as pretreated files and K columns (selected ions) (step 8 of Figure 1).

2.3. Pre-treatment of sensory data

In TDS/TCATA data, the durations of dominance/applicability of each descriptor are calculated for each evaluation. These durations are divided by the total duration of the evaluation (giving relative durations) to be aligned to PTR-MS data. Consequently, the total sum of durations in TDS is 1 while the total sum of durations in TCATA is higher than 1 when several descriptors are clicked simultaneously.

2.4. Principle of Regularized Generalized Canonical Correlation Analysis (RGCCA)

Regularized Generalized Canonical Correlation Analysis (Tenenhaus & Tenenhaus, 2011) is a general framework for multiblock analysis that was used to combine chemical and sensory data. This method allows J different blocks of measurement $\mathbf{X}_{j,j=1..J}$ (previously column-centered and scaled to give the same weight to each block in the analysis) with same evaluations as rows sorted in the same order to be co-analyzed in order to find linear combinations maximizing a covariance criterion *crit* under constraints:

$$crit = \sum_{\substack{j,k=1 \\ k \neq j}}^{J} c_{jk}\, g(cov(\mathbf{X}_j a_j, \mathbf{X}_k a_k))$$

$$\forall j \in \{1,..,J\}, \tau_j a_j{}' a_j + (1 - \tau_j)(\mathbf{X}_j a_j)'(\mathbf{X}_j a_j) = 1$$

In this expression, a_j are the coefficients of a linear combinations of the variables (or weights) of block *j* to be found by optimization of *crit* and the other symbols are parameters:

- c_{jk} is a binary value equaling 1 if the blocks *j* and *k* are supposed to be connected, 0 if else

- g is a convex function that can be chosen, generally among identity, square or absolute value
- τ_j is a number between 0 or 1 used for potential regularization: it is called shrinkage constant. The choices τ_j = 1, τ_j = 0 and 0 < τ_j < 1 are, respectively, referred as Modes A, B and Ridge (Tenenhaus & Tenenhaus, 2011). Note that Modes A and B correspond, respectively, to the constraints ${a_j}'a_j$ = 1 and var $(\mathbf{X}_j a_j)'(\mathbf{X}_j a_j) = 1$. Choosing a shrinkage constant τ_j between 0 and 1 ensures a continuum between Modes A and B.

This method presents several advantages presented in Tenenhaus et al. (2017). Almost all sequential multiblock component methods based on an optimization problem are special cases of RGCCA such as Partial Least Square Regression, Canonical Component Analysis, etc. RGCCA algorithm is also very simple and has good convergence properties. It is monotonically convergent and the solution at convergence is a stationary point of the optimization problem (Tenenhaus et al., 2017).

After convergence, $y_j = \mathbf{X}_j a_j$, $\forall j \in \{1, .., J\}$ is named "component" of the block j.

Deflation steps are also included in order to obtain more than one component. The obtained deflated components are orthogonal by construction.

With its flexibility, the RGCCA framework can be used for answering most of research questions, supervised or exploratory, with 2 blocks or more. RGCCA calculations were computed with RGCCA package available on github (https://github.com/rgcca-factory/RGCCA).

2.5. Application of pretreatment and multiblock analysis on the example dataset

2.5.1. Pre-treatment of the dataset

Both PTR datasets (paired with TDS or TCATA) were pre-treated according to the method presented in 2.2. The parameters chosen for breathing correction were: isoprene, $C_8H_9^+$, m/z 69.069 for the breathing reference, a half window size of 5 for Moving Average filter, $n_{TimesHigherThanBlank}$ =3, d_{min} =2s and $i_{min} = i_{max}/5$ for the expiration peak where i_{max} is the maximal intensity during the evaluation. Two examples of pre-treatments were computed for illustration (Figure 2). The first example (E1) corresponds in the subject "S002" for TDS, product C and replicate 1 while the second one (E2) corresponds to the subject "S001", product C replicate 1 in TDS. For a better visualization only some ions (*m/z* 71.085 and *m/z 115.113*, found relevant in the rest of the paper) were represented. The pre-treatment visualization consisted in drawing the raw intensity then the corrected intensity along time. The breathing reference and detected breathing cycles were also drawn.

Thirty-five ions (excluding isoprene) were selected with the method presented in 2.2.3. and described in Supplementary Table 1.

In the next sections, the pre-treated chemical data tables were noted $\mathbf{X_{PD}}$ for PTR-TDS and $\mathbf{X_{PC}}$ for PTR-TCATA while pre-treated sensory data tables were noted $\mathbf{X_D}$ for TDS and $\mathbf{X_C}$ for TCATA. These tables have N lines (with N = number of products* number of subjects * number of replicates) sorted in the same order. $\mathbf{X_{PD}}$ and $\mathbf{X_{PC}}$ have P columns (with P = 35 is the number of relevant ions). $\mathbf{X_D}$ and $\mathbf{X_C}$ have D columns (with D = 6, the number of descriptors).

2.5.2. Pairing sensory data with chemical data

To find out which sensory method fits the chemical results best, RGCCA was conducted for two components by maximizing the following criterion:

$$cov(\mathbf{X_{PD}}a_{PD}, \mathbf{X_D}a_D) + cov(\mathbf{X_{PC}}a_{PC}, \mathbf{X_C}a_C)$$

under the following constraints:

$$\forall j \in \{PD, PC, C, D\}, \qquad {a_j}' a_j = 1$$

With this criterion, the TDS/TCATA durations of sensory attributes were explained by the intensities of relevant ions (independently of products) for all evaluations. This approach is strictly equivalent to run 2 separated PLS but they are computed in one single calculation.

Maximizing this criterion results in getting two components per block that are comparable to principal components in PCA. As in PCA, these components can be plotted on a map. For each j block, the two components are noted, ${y^1}_j$ and ${y^2}_j$. Further components could also be calculated, but we focused here on the first two components, allowing maps to be plotted and the three products to be discriminated (if the inter-product variability was higher than the intra-product variability).

For each combination of two blocks (j and k), univariate Pearson correlations between ${y^1}_j$ and ${y^1}_k$ (respectively ${y^2}_j$ and ${y^2}_k$) were calculated and noted R_1 (resp. R_2). Bivariate RV coefficients (noted RV) between the two individual configurations on 1-2 map were also calculated. They indicated the similarities between PTR-TCATA and PTR-TDS and their related sensory spaces.

It should be noted that RGCCA did not use the product information in its calculations: it only maximized the covariance of sensory data with chemical data. It was expected that the two

first axes (maximizing the covariance between chemical and sensory data) should reflect the product variability rather than inter-subject or intra-subject variability. Therefore, the product discrimination should be similar in sensory and physico-chemical spaces. Consequently, the one-way MANOVA of the product effect on the first 2 axes was also calculated to evaluate whether and to what extent the products were discriminated on the 1-2 map. The F statistic of MANOVA illustrates the product discrimination of the selected first two components. To complete this discrimination analysis of the first 1-2 map, the obtained coordinates were automatically clustered in 3 groups with hclust® algorithm (distance= "Euclidean", method="ward.D2"): theoretically the three groups should be constituted of the three products A, B and C. The rates of well-classified evaluations were calculated for each product group.

Finally, the interpretation of the first RGCCA axis was studied by displaying the weight of each ion on the first component. A high weight (in absolute value) indicates that the ion (or descriptor) strongly contributed to the component. In order to evaluate the stability of these weights, a bootstrap procedure with 1000 RGCCA was run, giving a distribution of weights. Then, 2.5% and 97.5% quantiles of the obtained distribution were plotted. The same work can be conducted for the second (and further) components but were not presented here.

2.5.3. Supplementary statistical analyses

Supplementary statistical analyses were run in order to assess the significance of product effect for the different ions. Two-way mixed ANOVAs with random subject effect were computed for each descriptor in both TDS and TCATA standardized durations in order to assess univariate differences between the three products for the selected descriptors. In the same context, two-way mixed ANOVA with random subject effect were also computed for

each ion in order to assess univariate differences between the products for the selected ions. The F statistics and the number of pairs of products statistically different (with post-hoc Tuckey tests and $p<0.05$) were also calculated. Finally, the temporal curves as defined in Pineau et al. (2009) were plotted in order to visualize the temporality of sensations.

2.6. Software for statistical analysis

All calculations were done with R (4.0.2). All pre-treatments were computed with a specific home-made R package (PTRMSR) available on github (www.github.com/ChemoSens/PTRMR). The multiblock analysis was conducted with RGCCA package available on https://github.com/rgcca-factory/RGCCA. The code required for pre-treatment and multiblock analysis is available in https://github.com/ChemoSens/ExternalCode/ /TCATAvsTDS_paper.r. In the same repository, data for pretreatment is available in ./dataForPretreatementExample and data for RGCCA analysis in ./dataForMultiblockAnalysis.

3. RESULTS

3.1. Pre-treatment of the dataset

Figure 2a and 2d shows that the obtained breathing cycle limits (represented by vertical red bars) are intuitive and suggests the validation of the choice of the parameters of the peak picking algorithm.

Figure 2b and 2e shows raw data for the ions m/z 71.085 and m/z 115.113 that present oscillations due to breathing cycles (as expected). Furthermore, the intensity during the blank period (left of the vertical black line) is clearly not null and different between these two evaluations, thus should be properly removed from the signals.

Figure 2c and 2f shows the result of the calculation of the mean by the breathing cycles obtained with isoprene curves, then corrected by the blank period. Therefore, in Figure 2c and 2f, points - whose abscissa are the centers of the breathing cycles and whose ordinates are the intensities averaged by breathing cycles minus the average intensity during the blank period - are plotted then connected by segments. The signal is still not completely smoothed but represents the averaged intensity measured during each breathing cycle corrected from the blank period.

3.2. Pairing sensory data with PTR-MS

Figure 3 illustrates the individual projections on the RGCCA axes for PTR-TDS, TDS, PTR-TCATA and TCATA. Each point represents an evaluation (product*subject*replicate). In addition to the visual analysis of the maps, the MANOVA F indicates that the product discrimination is better in TDS (48.84) than in TCATA (24.96). Furthermore, the discrimination is also better in PTR-TDS (94.06) than in PTR-TCATA (78.45).

TDS seems to be slightly more correlated to PTR-TDS (R_1=0.71, R_2=0.72, RV=0.51) than TCATA to PTR-TCATA (R_1=0.67, R_2=0.64, RV=0.44). All RV coefficients were found highly significant ($p<0.001$). The two chemical blocks seem to be well correlated (R_1=0.89, R_2=0.4, RV=0.64) whereas the two sensory blocks were clearly less correlated (R_1=0.69, R_2=0.10, RV=0.32).

The first axis of all maps separated C from A and B. Regarding the second axis, it is less clear, but it seems to separate A from B.

Table 1 shows the results of classification in three groups of the evaluations based on their 1-2 coordinates. Over 96 evaluations, 87 (and 86) are well-classified in PTR-TCATA and PTR-TDS. Regarding the sensory results, the classification has less sense in TCATA (68.75% of well-

classified) than in TDS (77.08%). As shown by the maps, this is due to a confusion between A and B (only 19/32 and 15/32 are in the right group) while C is perfectly clustered (32 on 32 evaluations).

Figure 4 displays the weight of each variable for each selected first component of RGCCA, sorted by decreasing absolute value. In TDS, a high score on component 1 indicates a strong score on *Fruity* and *Spicy* and a weak score on *Woody* and *Cocoa*. Same conclusions were obtained in TCATA even if the descriptor order is different. Consequently, C seems to be longer *Cocoa/Woody*, not *Fruity* and not *Spicy*. *Roasted* is significant in TCATA but not in TDS.

In PTR-MS, only the 25 ions with the highest weights on the first component were displayed. Among them, the two highest are *m/z 115.113 and m/z 71.085* in both PTR-TDS and PTR-TCATA. Weights seemed to be similar in PTR-TDS and PTR-TCATA. High weights characterize the linear combination of PTR-TCATA/TDS (as *115.113* or *71.085*).

3.3. Additional statistical analysis

3.3.1. Univariate results

TDS/TCATA ANOVA results (Table 2) showed that F statistics are higher in TDS than in TCATA except for *Roasted* and *Dry Fruits* that were significant only in TCATA. Furthermore, *Spicy* (and not *Cocoa* and *Woody* as selected in RGCCA) were not the most significant attributes for the product effect.

The proportion of pairs of different products were 8/18 for TDS, 10/18 for TCATA, 83/105 for PTR-TDS and 78/105 for PTR-TCATA.

Table 3 shows that most of the relevant ions are significant for the product effect. Only 3/35 ions have slightly different conclusions in significance (in bold).

3.3.2. Kinetics of sensations

Regarding the temporality of the sensations, Figure 5 shows that the sequentiality is not obvious in these chocolates with these flavor descriptors. In both TCATA and TDS, A appeared *Fruity* all the time, B *Spicy* and slightly *Fruity* and C *Cocoa Woody*. Consequently, the RGCCA approach based on the tables of relative durations (without temporality anymore) should not remove much information.

4. DISCUSSION

This paper can be considered as a tutorial to analyze sensory and physico-chemical data conjointly with adapted pre-treatments and a multiblock method. Furthermore, the example dataset allows to compare TDS and TCATA towards their propensity to fit PTR-MS data. This question was never addressed in the literature, although it could be of interest for both flavor chemists and sensory scientists.

4.1. Methodological discussion

4.1.1. Pre-treatment: advantages and limits

Pre-treatment is a primordial step while analyzing PTR-MS data. This paper proposes a pipeline and automatization of the pre-treatments in PTR-MS. The presented method has several advantages:

(i) The breathing effect is properly considered and this makes sense from a physiological point of view. Statistics regarding the breathing cycles of the subjects can also be extracted and can be of interest to study the kinetics of ion release.

(ii) The intensity during the blank period is correctly removed. Indeed, usual practices consist in subtracting the average of intensity during the blank period without any

breathing correction. This practice is misleading as this blank intensity is not the same for inspirations and expirations. The subtraction of the averaged blank signal (whatever the position in the breathing cycle) induces negative signals for all expirations, that are usually replaced by 0. This practice strongly modifies the initial signal and could conduct to misinterpretations. In this paper, this bias is avoided by working on the data after breathing correction instead of before breathing correction.

(iii) This method is automatic and entirely reproducible. Until now, PTR-MS pre-treatment used different home-made procedures and was fastidious. This paper presented a reproducible pipeline that includes many automatizations. This pre-treatment is fast (less than 1 second by PTR-MS file) and flexible. After a first treatment by PTR-MS Viewer, the data can be directly analyzed by R.

A limit of the pre-treatment presented in this paper is the choice of the algorithm parameters (size of smoothing window, smoothing method, minimal duration of breathing cycles, parameters for relevant ions selection...). These parameters are entirely customable and were manually adjusted in this study. The user can check the breathing cycle results thanks to the "report" tool used for producing Figure 2 (available in the R package). Further work would be necessary to investigate the impact of the choice of these parameters on the results and to implement automatic optimization of the choice of these parameters.

4.1.2. Multiblock approach

RGCCA is very flexible and can be used to answer a large number of multiblock problematics. Here, RGCCA was successfully used with a simple approach with 4 blocks. Other multiblock methodologies could have been investigated. For example, the two PTR-MS datasets could

have been averaged in order to provide a "chemical" standard, then a 3-block approach could have been considered. This would have led to more stability of the PTR-MS block, but to the loss of the structure of the experiment. Another analysis could consist in explaining the duration of one single given descriptor (for example *Cocoa*) by the whole physico-chemical block. In this case, the sensory block should be constituted of the single *Cocoa* variable. Sparsity parameters can also be used in order to select a set of ions/variables responsible for this sensation.

A crucial question in multiblock approach is the choice of the algorithm parameters (C, g, τ_j) but also the choice of the number of components. This paper focused on an analysis with the first two components because two axes were sufficient to observe the main differences between the three products (assuming that inter-product variability is greater than intra-product variability). These first two components are the more informative ones, but other components could also contain information and could deserve further investigations. For other datasets, the number of components should also be cautiously chosen. Cross validation or elbow criterion strategies could be considered.

The use of RGCCA allowed several blocks to be related together then correlations between blocks to be measured. It returns a physico-chemical and a sensory component having the highest covariance. The obtained linear combination can be validated with bootstrap procedures indicating which are the ions/sensations with the highest weights in this linear combination. Thus, all the potential links between ions and sensations are considered and a prediction model can easily be built with the obtained components. For example, the linear combination of ions obtained in PTR-TDS could be used as a prediction of the *Cocoa/Woody*

TDS component. The correlations between the predicted values and the real ones would be 0.71 as indicated by the correlation between the two TDS and PTR-TDS components.

As with any multivariate method, the interpretation of the components is not always easy, especially chemical component composed of a large number of compounds. This is why the univariate analysis of variance of each compound is necessary to indicate compounds able to discriminate products.

Another way to observe the relations between sensory and physico-chemical data could be to calculate the univariate correlations between ions and descriptors. This approach leads to univariate conclusions that can be interesting but that are not always appropriated to the multivariate characteristics of sensory or physico-chemical data. For example, in univariate approach, the interactions between attributes or ions are not considered while they can be relevant.

4.2. Discussing the results obtained on the example

4.2.1. Chemical discussion

PTR-TDS and PTR-TCATA gave similar results, but with some potential differences, the more noticeable being a better product discrimination from PTR-TDS. This difference could have several reasons:

(i) TCATA or TDS task could stress more or less the subjects, accelerating their breath and thus affecting ion release. TCATA being a more difficult task than TDS, this could be a reason for its lower product discrimination.

(ii) PTR-MS data natural heterogeneity and repeatability.

These hypotheses could be investigated in further works.

Regarding the ions corresponding to the highest weights in the PTR-TDS and PTR-TCATA linear combinations, the conclusions seemed similar independently of the method. Some ions could be relevant to characterize the sensory perception and could be studied (*m/z 71.085* and *m/z 115.113* for example*)*. These ions were also found significant for product effect in univariate analysis and discriminating C from A and B. However, these specific ions have multiple possible identifications, which stopped further interpretation at the time of submitting the paper. GC-MS could be used to properly identify these compounds.

4.2.2. Sensory discussion

The originality of our approach in this dataset is to use chemistry as a tool to compare TDS and TCATA results. The product configurations were similar in both methods: C was found longer *Cocoa/Woody,* A longer *Fruity* and B longer *Spicy.*

In terms of univariate product discrimination, TCATA returns more significant descriptors than TDS and more pairs of different products, that is in agreement with Ares et al. (2015).

However, the F statistics are clearly higher in TDS for *Cocoa, Woody* (discriminating C from A and B), *Spicy* (discriminating B from A and C) and *Fruity* (discriminating A from B and C) suggesting that the differences between the products were expressed differently in the two protocols: TCATA highlighted more product differences, whereas TDS highlighted stronger product differences. Thus, a subject that would like to express that products are different would not have the same behavior in TDS and TCATA. In TCATA, he would click on more descriptors, while he would choose only one representative descriptor in TDS. Since the level of discrimination is higher in TDS, it means that agreement among panelists was stronger, likely because the task is easier. This is validated by the TDS/TCATA curves showing that TDS

tends to identify one or two single descriptors by product, while the diagnosis is less clear in TCATA.

Regarding the physico-chemical data related to these results, the PTR-MS product structure seemed to be more maintained in TDS than in TCATA for both discrimination and proximity between chemical and sensory maps (better correlations, better multivariate F and better well-clustered rate in TDS). This result may appear counterintuitive: as several ions being released simultaneously, one could argue that a method like TCATA allowing several descriptors to be selected simultaneously would be better suited. However, besides the potential reasons already proposed above, one could also suggest that the product space used in this experiment was too small and too simple: 3 chocolates only with strong differences easy to catch and almost not changing over time. Another explanation is the fact that it could be difficult for a subject to not forget unclicking a descriptor no longer applicable and that could induce noise in the data. Different results could have been obtained by using Fading TCATA, a TCATA variant facilitating the TCATA task by making term deselection automatic and progressive over a period of a few seconds (Ares et al., 2016). Another possible explanation of the better discrimination in TDS could be the better discrimination in PTR-TDS (F=94.06) than in PTR-TCATA (F=78.45): the differences between TDS and TCATA could be due to real differences in the perceived discrimination during the two protocols. This hypothesis appears quite unlikely but not impossible.

The results obtained in this example do not aim to be generalized to sensory analysis. It is specific to this protocol, these products, these subjects (only 16) and these descriptors. Further works would be necessary to reach conclusions on sensory protocols and the methodology presented in this paper could be adapted to other datasets pairing sensory and

physico-chemical methods (with other products, descriptors and more subjects) in this purpose.

Furthermore, here, multiblock approaches were applied to datasets where temporality was synthetized as durations or AUCs. These final datasets contain one single value for each evaluation and descriptor. Consequently, the same multiblock methods can also be used for other sensory/physico-chemical pairings such as the comparison of sensory profiling or CATA to physico-chemical results.

4.2.3. Comparison of mono-block and multiblock results

In this example, univariate and multiblock results agreed regarding the general results: C is different from A and B according to the descriptors *Cocoa* and *Woody* and a list of ions is established as being the potential drivers of these differences (ions *m/z 71.085* and *m/z 115.*113 being the best in that list). However, *Fruity* was the more discriminant descriptor in TDS and TCATA but was not as clearly explained by PTR-MS as *Cocoa/Woody*.

4.3. Perspectives

4.3.1. Practical implications of this work

This work presents a complete pipeline to pair TCATA or TDS data with PTRMS data and makes the R code available for the community. The practical use of the tools presented in this paper by physico-chemists could conduct to further improvements of PTRMSR (more user-friendly interface, default parameters, etc.). The use of multiblock methods was also illustrated. It could be used more widely for any experiment pairing two or more acquisition modalities (in physico-chemistry and/or sensory analysis).

4.3.2. Further steps in pre-treatment

The automatized pre-treatment presented in this paper is a first step that could probably be improved by better algorithms. For example, the parameters could be estimated automatically with the optimization of a criterion to be defined. Further work could also consist in automatizing the PTR-MS Viewer part to work directly with the HDF PTR files (having the "h5" extension), which would lighten the manual treatment.

4.3.3. Further steps in pairing TDS and TCATA: taking temporality into account

The analysis conducted in this paper presents a pairing of sensory and chemical data using durations of sensory descriptors and total intensity of ions during the tasting period. This approach does not take the temporality into account and the time-by-time comparison of ions and sensations. In the example presented in this paper, it was not an issue as no sensorial sequentiality was observed. However, it should certainly be relevant with datasets being more complex and several perspectives can be considered. First, instead of considering matrices (evaluations as lines and descriptors or ions as columns), one could consider tensors (with the time as a third dimension) and the tensorial extension of RGCCA named Multiway Generalized Canonical Correlation Analysis (Gloaguen et al., 2020). This implies some technical issues as to have the same time discretization in PTR and TDS or TCATA, still dealing with breathing correction. Then, functional analysis (Preda et al., 2021; Wang et al., 2016) also exists to take properly the temporality into account. These type of methods does not systematically require to have the same time points and could be another solution to take temporality into account.

5. CONCLUSIONS

This work presents an automatic pipeline of pre-treatment of PTR-MS data and the application of the RGCCA method to analyze sensory and physico-chemical data conjointly. Both pre-treatment and RGCCA were successfully applied on a dataset with TCATA and TDS evaluations paired with nosespace measured by PTR-MS. TCATA and TDS gave similar results, but TDS was shown to preserve the product configuration slightly more than TCATA in this study. As this approach is based on data aggregated over time, the temporality is not considered for itself, further work will consist in using the temporal component for pairing sensory and chemical signals in more details.

CRediT author statement:

Caroline Peltier: formal analysis, software, methodology, conceptualization, visualization, writing – original draft; **Michel Visalli:** software, formal analysis, writing - review and editing; **Hélène Labouré:** project administration, experimental design, supervision, writing - review and editing; **Cantin Hélard**: investigation, data curation; **Isabelle Andriot:** investigation, data curation, writing - review and editing; **Sylvie Cordelle**: investigation, data curation, writing - review and editing;; **Jean-Luc Le Quéré**: project administration, experimental design, supervision, writing – review and editing; **Pascal Schlich**: conceptualization, supervision, writing – review and editing. All authors have read and agreed to the published version of the manuscript.

Acknowledgements:

We would like to thank Pierre Costet, Marie Repoux et Florent Coste from Valrhona for allowing the short stay of Cantin in the company, and for their help in the choice, and gift, of the chocolate samples.

References

Ares, G., Bruzzone, F., Vidal, L., Cadena, R. S., Giménez, A., Pineau, B., Hunter, D. C., Paisley, A. G., & Jaeger, S. R. (2014). Evaluation of a rating-based variant of check-all-that-apply questions: Rate-all-that-apply (RATA). *Food Quality and Preference*, *36*, 87–95. https://doi.org/10.1016/j.foodqual.2014.03.006

Ares, G., Castura, J. C., Antúnez, L., Vidal, L., Giménez, A., Coste, B., Picallo, A., Beresford, M. K., Chheang, S. L., & Jaeger, S. R. (2016). Comparison of two TCATA variants for dynamic sensory characterization of food products. *Food Quality and Preference*, *54*, 160–172. https://doi.org/10.1016/j.foodqual.2016.07.006

Ares, G., Jaeger, S. R., Antúnez, L., Vidal, L., Giménez, A., Coste, B., Picallo, A., & Castura, J. C. (2015). Comparison of TCATA and TDS for dynamic sensory characterization of food products. *Food Research International*, *78*(2015), 148–158. https://doi.org/10.1016/j.foodres.2015.10.023

Barron, D., Pineau, N., Matthey-Dorey, W., Ali, S., Sudre, J., Germain, J. C., Kolokziejczyk, E., Pollien, P., Labbe, D., Jarisch, C., Dugas, V., Hartmann, C., & Folmer, B. (2012). Impact of crema on the aroma release and the in-mouth sensory perception of espresso coffee. *Food & Function*, *3*(9), 223–230.

Biasioli, F., Yeretzian, C., Gasperi, F., & Ma, T. D. (2011). PTR-MS monitoring of VOCs and BVOCs in food science and technology. *Trends in Analytical Chemistry*, *30*(7). https://doi.org/10.1016/j.trac.2011.03.009

Castura, J. C., Antúnez, L., Giménez, A., & Ares, G. (2016). Temporal Check-All-That-Apply (TCATA): A novel dynamic method for characterizing products. *Food Quality and Preference*, *47*, 79–90. https://doi.org/10.1016/j.foodqual.2015.06.017

Charles, M., Romano, A., Yener, S., Barnabà, M., Navarini, L., Märk, T. D., Biasoli, F., & Gasperi, F. (2015). Understanding flavour perception of espresso coffee by the combination of a dynamic sensory method and in-vivo nosespace analysis. *Food Research International*, *69*, 9–20. https://doi.org/10.1016/j.foodres.2014.11.036

Déléris, I., Saint-eve, A., Dakowski, F., Sémon, E., Quéré, J. Le, Guillemin, H., & Souchon, I. (2011). *The dynamics of aroma release during consumption of candies of different structures , and relationship with temporal perception*. *127*, 1615–1624. https://doi.org/10.1016/j.foodchem.2011.02.028

Esmerino, E. A., Castura, J. C., Ferraz, J. P., Tavares Filho, E. R., Silva, R., Cruz, A. G., Freitas, M. Q., & Bolini, H. M. A. (2017). Dynamic profiling of different ready-to-drink fermented dairy products: A comparative study using Temporal Check-All-That-Apply (TCATA), Temporal Dominance of Sensations (TDS) and Progressive Profile (PP). *Food Research International*, *101*(June), 249–258. https://doi.org/10.1016/j.foodres.2017.09.012

Gatto, L., & Lilley, K. (2012). MSnbase - an R/Bioconductor package for isobaric tagged mass spectrometry data visualization, processing and quantitation. *Bioinformatics*, *28*, 288–

289.

Gibb, S., & Strimmer, K. (2017). Mass Spectrometry Analysis Using MALDIquant. *Statistical Analysis of Proteomics, Metabolomics, and Lipidomics Data Using Mass Spectrometry*, 101–124. https://doi.org/10.1007/978-3-319-45809-0_6

Gloaguen, A., Philippe, C., Frouin, V., Gennari, G., Dehaene-Lambertz, G., Le Brusquet, L., & Tenenhaus, A. (2020). Multiway generalized canonical correlation analysis. *Biostatistics*, 1–25. https://doi.org/10.1093/biostatistics/kxaa010

Jordan, A., Haidacher, S., Hanel, G., Hartungen, E., Märk, L., Seehauser, H., Schottkowsky, R., Sulzer, P., & Märk, T. D. (2009). *International Journal of Mass Spectrometry Short communication A high resolution and high sensitivity proton-transfer-reaction time-of-flight mass spectrometer (PTR-TOF-MS). 286*, 122–128. https://doi.org/10.1016/j.ijms.2009.07.005

Lejeune, F.-X., Ichou, F., Camenen, E., Colsch, B., Mauger, F., Peltier, C., Moszer, I., Gilson, E., Pierre-Jean, M., Floch, E. Le, Sabarly, V., Tenenhaus, A., Deleuze, J.-F., Ewenczyk, C., Vidailhet, M., & Mochel, F. (2022). A Multimodal Omics Exploration of the Motor and Non-Motor Symptoms of Parkinson's Disease. *International Journal of Translational Medicine*, *2*(1), 97–112. https://doi.org/10.3390/ijtm2010009

Mesurolle, J., Saint-Eve, A., Déléris, I., & Souchon, I. (2013). Impact of fruit piece structure in yogurts on the dynamics of aroma release and sensory perception. *Molecules*, *18*(5), 6035–6056. https://doi.org/10.3390/MOLECULES18056035

Pedrotti, M., Spaccasassi, A., Biasioli, F., & Fogliano, V. (2019). Ethnicity, gender and physiological parameters: Their effect on in vivo flavour release and perception during chewing gum consumption. *Food Research International*, *116*(November 2018), 57–70. https://doi.org/10.1016/j.foodres.2018.12.019

Pineau, N., Schlich, P., Cordelle, S., Mathonnière, C., Issanchou, S., Imbert, A., Rogeaux, M., Etiévant, P., & Köster, E. (2009). Temporal Dominance of Sensations: Construction of the TDS curves and comparison with time-intensity. *Food Quality and Preference*. https://doi.org/10.1016/j.foodqual.2009.04.005

Pineau, Nicolas, de Bouillé, A. G., Lepage, M., Lenfant, F., Schlich, P., Martin, N., & Rytz, A. (2012). Temporal Dominance of Sensations: What is a good attribute list? *Food Quality and Preference*, *26*(2), 159–165. https://doi.org/10.1016/j.foodqual.2012.04.004

Preda, C., Grimonprez, Q., & Vandewalle, V. (2021). Categorical functional data analysis. The cfda r package. *Mathematics*, *9*(23), 1–31. https://doi.org/10.3390/math9233074

Tenenhaus, A., & Tenenhaus, M. (2011). Regularized Generalized Canonicalcorrelation Analysis. *Psychometrika*.

Tenenhaus, M., Tenenhaus, A., & Groenen, P. J. F. (2017). Regularized Generalized Canonical Correlation Analysis: A Framework for Sequential Multiblock Component Methods. *Psychometrika*, *82*(3), 737–777. https://doi.org/10.1007/s11336-017-9573-x

Wang, J. L., Chiou, J. M., & Müller, H. G. (2016). Functional Data Analysis. *Annual Review of Statistics and Its Application*, *3*, 257–295. https://doi.org/10.1146/annurev-statistics-041715-033624

Xicota, L., Ichou, F., Lejeune, F. X., Colsch, B., Tenenhaus, A., Leroy, I., Fontaine, G., Lhomme, M., Bertin, H., Habert, M. O., Epelbaum, S., Dubois, B., Mochel, F., & Potier, M. C. (2019). Multi-omics signature of brain amyloid deposition in asymptomatic individuals at-risk for Alzheimer's disease: The INSIGHT-preAD study. *EBioMedicine*, *47*, 518–528. https://doi.org/10.1016/j.ebiom.2019.08.051

Table 1: Number of well-classified evaluations by product group.

	Group A	Group B	Group C	Total	Total (%)
PTR-TDS	31	25	31	87	90.625
TDS	22	22	30	74	77.08
PTR-TCATA	29	26	31	86	89.58
TCATA	19	15	32	66	68.75

Table 2: F statistics and significance of ANOVA (attribute duration~product+subject(random)+product*subject) (*** p<0.001, **p<0.01, *p<0.05, .<0.1) for TDS and TCATA. Supplementary columns (A_tcata, B_tcata, C_tcata and A_tds, B_tds and C_tds) indicate the results of multiple comparisons (same letters indicate same group).

Descriptor	TCATA F	TCATA sig.	A_tcata	B_tcata	C_tcata	TDS F	TDS sig.	A_tds	B_tds	C_tds
Cocoa	19.02	***	A	A	B	30.4	***	A	A	B
Dry Fruits	2.76	.	A	A	A	1.74	Ns	A	A	A
Fruity	14.16	***	B	A	A	30.67	***	B	A	A
Roasted	4.85	*	A	AB	B	0.8	Ns	A	A	A
Spicy	23.77	***	B	C	A	33.74	***	A	B	A
Woody	19.9	***	A	A	B	29.6	***	A	A	B

Table 3: F statistics and significance of ANOVA (attribute AUC~product+subject(random)+product*subject) (*** p<0.001, **p<0.01, *p < 0.05) for PTR-TDS and PTR-TCATA. Supplementary columns (A_tcata, B_tcata, C_tcata and A_tds, B_tds and C_tds) indicate the results of multiple comparisons (same letters indicate same group).

ion	TCATA F	TCATA sig.	A_tcata	B_tcata	C_tcata	TDS F	TDS sig.	A_tds	B_tds	C_tds
43.065	25.66	***	A	A	B	27.35	***	A	A	B
45.033	36.92	***	B	B	A	73.67	***	B	B	A
45.043	7.44	**	B	B	A	7.83	**	B	B	A
55.053	4.82	*	A	B	AB	13.51	***	A	B	B
58.041	6.84	**	A	A	B	5.78	**	A	A	B
69.083	0.55	Ns	A	A	A	0.55	ns	A	A	A
71.049	14.66	***	A	A	B	12.34	***	A	A	B
71.085	188.88	***	A	A	B	368.42	***	A	A	B
73.063	13.31	***	A	B	B	17.38	***	A	C	B
73.077	3.73	*	A	B	AB	10.87	***	A	B	B
74.068	9.27	***	A	B	A	29.84	***	A	B	A
74.085	6.52	**	A	B	A	6.9	**	A	B	B
75.043	34.7	***	B	A	A	51.48	***	B	A	A
78.965	100.42	***	B	A	C	95.95	***	B	A	C
78.976	67.86	***	B	A	C	74.24	***	B	A	C
84.083	50.14	***	A	B	C	101.2	***	A	B	C
85.066	77	***	A	B	C	84.48	***	A	B	C
87.044	41.19	***	A	B	C	67.66	***	A	B	C
87.080	182.49	***	A	B	C	259.29	***	A	B	C
87.099	30.64	***	A	A	B	56.85	***	A	B	C
88.084	124.63	***	A	B	C	296.83	***	A	B	C
88.109	11.78	***	A	A	B	34.39	***	A	A	B
94.999	99.48	***	B	A	C	113.51	***	B	A	C
96.997	73.73	***	B	A	C	73.63	***	B	A	C
97.030	19.28	***	A	B	B	19.68	***	A	B	B
101.061	10.37	***	A	A	B	14.51	***	A	A	B
103.076	37.53	***	A	A	B	28.07	***	A	B	C
104.050	72.21	***	A	B	C	93.55	***	A	B	C
105.072	15.28	***	C	A	B	15.08	***	B	A	A
107.050	44.8	***	A	B	B	55.98	***	A	B	B
113.096	69.12	***	B	A	C	91.45	***	B	A	C
115.113	102.26	***	A	A	B	265.91	***	A	B	C
121.066	12.1	***	A	B	B	7.23	**	A	B	B
123.092	18.48	***	B	A	B	17.44	***	B	A	B
173.154	29.99	***	B	A	A	37.13	***	B	A	A

Figure 1: General scheme of pretreatment of PTRMS data

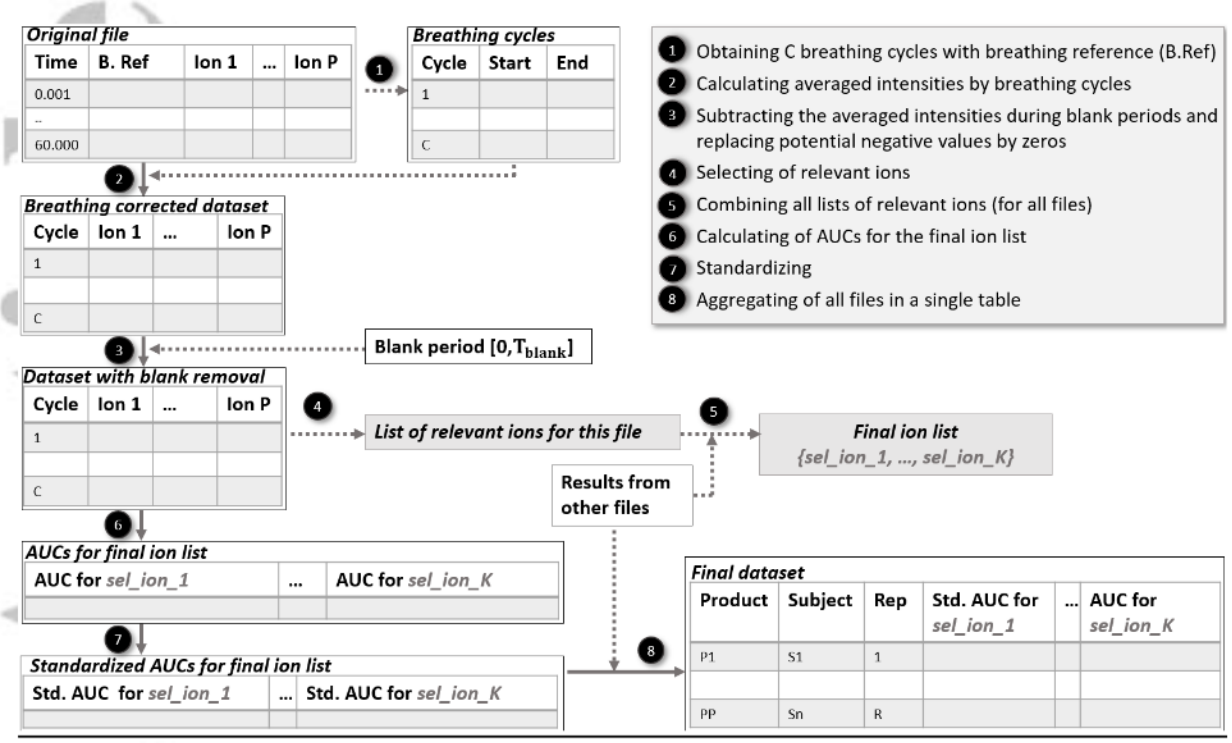

Figure 2: Illustration of the automatized PTR-MS pre-treatment for two evaluations: E1 (a., b., and c.) and E2 (d. e. and f.) with cycle detection on a breathing reference (a, d) then calculation of average by breathing cycle for two ions (m/z 71.085 in blue and m/z 115.113 in green) and noise removal representing the raw intensity before pre-treatment (b, e), while (c, f) represents the obtained pre-treated data.

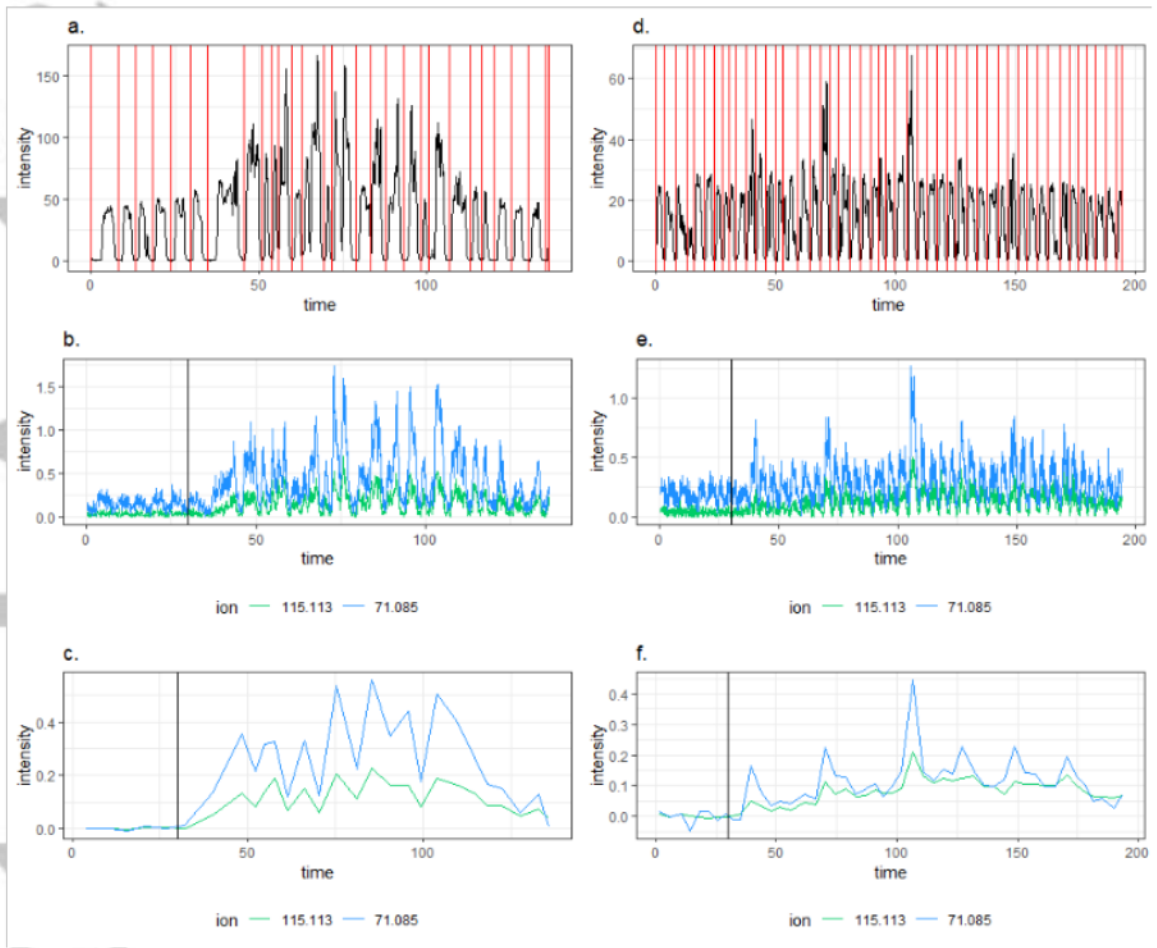

Figure 3: Individual RGCCA maps for raw data approach in PTR-TDS block (a.), PTR-TCATA block (b.), TDS block (c.) and TCATA block (d.). The points are colored according to their related product (red for A, green for B and blue for C).

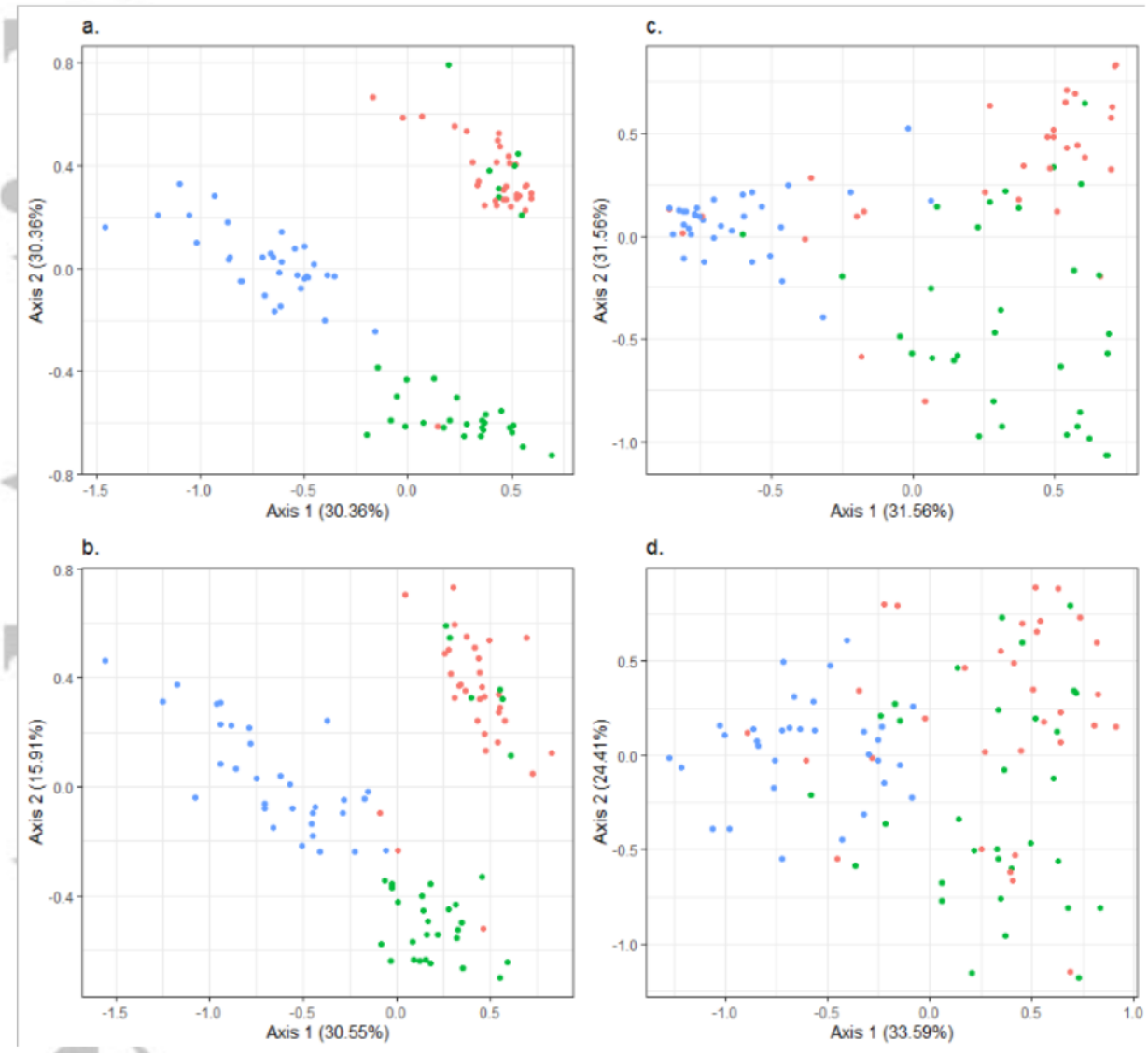

Figure 4: Bootstrap weights on the first component of RGCCA results.

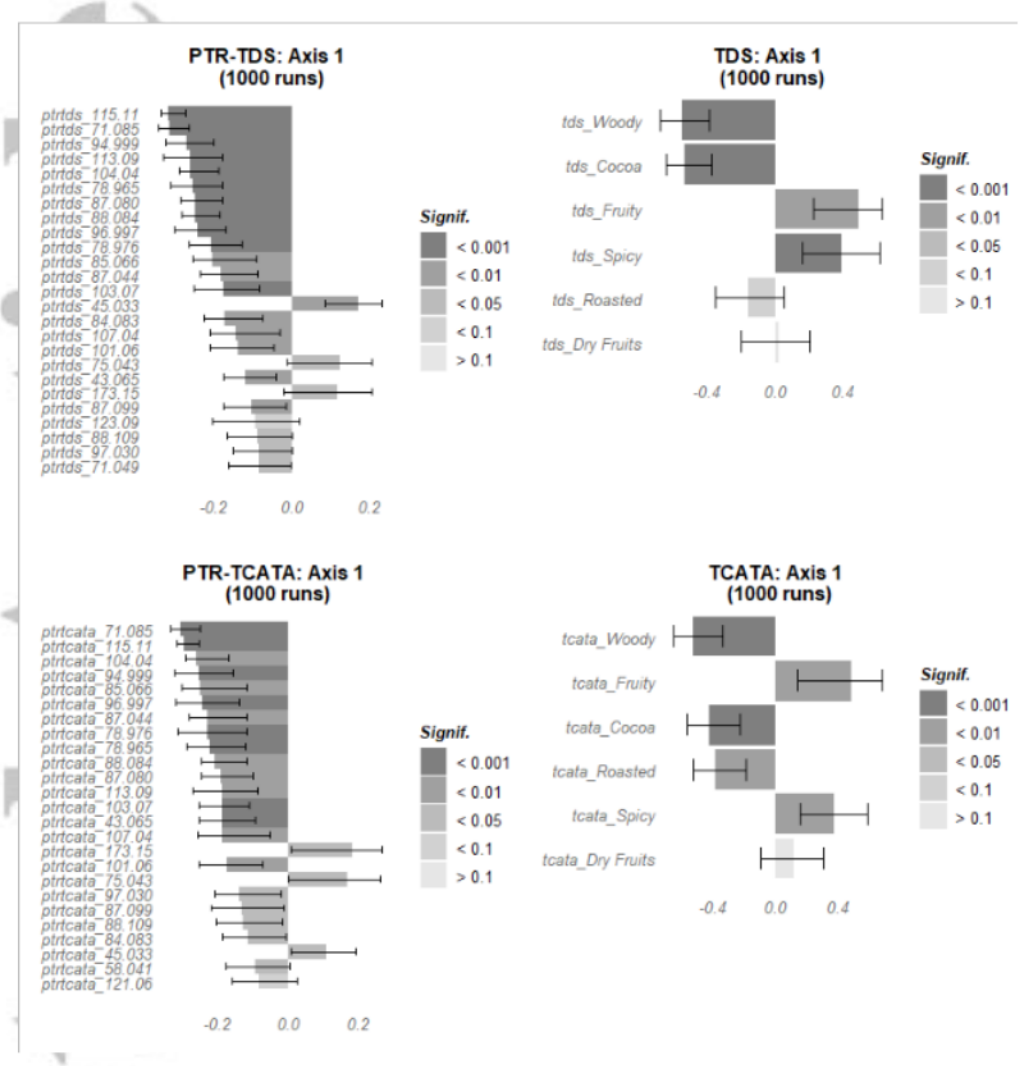

Figure 5: TDS and TCATA curves. The dotted line represents the chance line and TDS dominance rates are significant (p=0.05) above the grey zone.