

The discriminant pixel approach: A new tool for the rational interpretation of GCxGC-MS chromatograms

Jérôme Vial^{a,*}, Benoît Pezous^b, Didier Thiébaut^a, Patrick Sassi^a, Béatrice Teillet^c, Xavier Cahours^c, Isabelle Rivals^b

^a Laboratoire des Sciences Analytiques, Bioanalytiques et Miniaturisation (LSABM), UMR CNRS UPMC PECSA, ESPCI ParisTech, 10 rue Vauquelin, 75005 Paris, France

^b Équipe de Statistique Appliquée, ESPCI ParisTech, 10 rue Vauquelin, 75005 Paris, France

^c SEITA, Imperial Tobacco Group, Science & Stewardship, Centre de Recherche, 4 rue A. Dessaux, 45404 Fleury les Aubrais, France

ARTICLE INFO

Article history:

Available online 1 August 2010

Keywords:

GCxGC-MS
Time-alignment
Multivariate analysis
Correlation analysis
Non-target analytes

ABSTRACT

GCxGC is now recognized as the most suited analytical technique for the characterization of complex mixtures of volatile compounds; it is implemented worldwide in academic and industrial laboratories. However, in the frame of comprehensive analysis of non-target analytes, going beyond the visual examination of the color plots remains challenging for most users. We propose a strategy that aims at classifying chromatograms according to the chemical composition of the samples while determining the origin of the discrimination between different classes of samples: the discriminant pixel approach.

After data pre-processing and time-alignment, the discriminatory power of each chromatogram pixel for a given class was defined as its correlation with the membership to this class. Using a peak finding algorithm, the most discriminant pixels were then linked to chromatographic peaks. Finally, crosschecking with mass spectrometry data enabled to establish relationships with compounds that could consequently be considered as candidate class markers.

This strategy was applied to a large experimental data set of 145 GCxGC-MS chromatograms of tobacco extracts corresponding to three distinct classes of tobacco.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Comprehensive GCxGC appears nowadays to be the prime analytical tool for the study of complex mixtures of volatile compounds [1–4]. Structured 2D-chromatograms (often visualized as color plots) are obtained, where spots having different colors as a function of detector response replace the usual peaks of classical 1D chromatograms. Using an optimized (“orthogonal”) set of columns, compounds are organized in the color plot according to both carbon numbers and chemical structure [5], which facilitates interpretation. This technique has proven its usefulness and reliability in various areas such as the petroleum industry [6,7], flavors and fragrances [8,9], environmental [10,11], food [12,13], etc. However, handling chromatograms with several hundreds of spots, even more for petroleum samples [14,15], is not simple and requires the use of chemometrics [16].

For target analyte analysis, it has been proven that individual or group quantification was quite possible on GCxGC chromatograms and gave results of similar quality as those obtained with GC [7]. In the case of non-target analyte analysis, the situation is much less

favorable: the individual quantification of each compound would be theoretically possible, however the time required for an operator to do so would become rapidly unacceptable. A global comparison of chromatograms is required and should highlight some specific compounds presenting significant concentration differences from sample to sample, or better from one class of samples to another one.

Any global comparison of GCxGC chromatograms requires several pre-processing steps such as background correction, intensity normalization, and especially time alignment [17–22]. As a matter of fact, in a previously published paper, the efficiency of DTW (Dynamic Time Warping) alignment of the second dimension was illustrated on a reduced set of tobacco extract samples [23].

In the same study, the potential of using the loadings of Principal component analysis (PCA) or Independent Component analysis (ICA) in order to discover locations of the chromatograms that could discriminate between classes was investigated [23]. PCA being unsupervised, i.e. not using the knowledge about the classes, the first principal components of PCA might correspond to directions of large within class variance, but not of large class-to-class variance. ICA is also unsupervised, but it looks for components that are as least Gaussian as possible, and is therefore more likely to find directions that display structure in the data [24]. However, ICA

* Corresponding author. Tel.: +33 140794779; fax: +33 140794776.

E-mail address: jerome.vial@espci.fr (J. Vial).

assumes strongly non-Gaussian data, and has the disadvantage of being an iterative method, hence computationally heavy for data of the size of 2D-chromatograms.

Thus, in the present study, we propose a supervised method that evaluates the ability of each pixel of the chromatogram to discriminate between the class of interest and all the other classes in order to find the locations that are characteristic of a given class. A pixel being a detector intensity measured at a given time in the raw acquisition sequence (which corresponds to the retention times along the two dimensions), the most discriminant pixels can be related to chemical markers using mass spectrometry data. This strategy was applied to two extended experimental data sets of GCxGC-MS chromatograms of tobacco extracts corresponding to three distinct types of tobacco (Burley, Oriental, Virginia), and obtained either by supercritical fluid extraction (101 chromatograms) or steam distillation extraction (44 chromatograms) (in the paper focusing on time-alignment [23], only 12 SDE chromatograms were available). Because of their complexity, tobacco extracts are typically suited for GCxGC separations.

2. Experimental

2.1. GCxGC apparatus

A trace GCxGC system from Thermo-Electron Corporation (Courtalboeuf, France) equipped with a Merlin Microseal injector (Merlin Instrument Company, CA, US) was used. It was fitted out with a double jets carbon dioxide cryogenic modulator, and a split/splitless injector. In order to avoid discrepancies related to a poor trapping of the compounds in the modulator, the two jets were placed closer to the column than in the original configuration. The set of columns presenting the best compromise both in terms of separation and resistance of the stationary phases to degradation was as follows. The first column was an apolar capillary column VF-1ms, Varian (Les Ulis, France), 15 m × 0.25 mm, 1.0 µm. This column was connected to a DB 1701 1.5 m × 0.1 m, 0.1 µm from Agilent Technologies (Waldbronn, Germany). Connections between columns were made using deactivated Press-Tight® connectors from Restek (Evry, France). The flow rate of the carrier gas was 1 mL/min and the injector was set at 240 °C. In order to have a preconcentration of the solutes at the beginning of the first column, a cold trapping was applied to the splitless injection. The oven temperature program started at 40 °C for 40 s, then an increase at 60 °C/min was applied up to 70 °C, and after a hold for of 3 min, 2.5 °C/min were applied up to 240 °C. The injected volume was 2 µL. The injection was carried out in splitless mode with a surge of 400 kPa during 40 s. A typical modulation period of 5 s was used. Detection was carried out with the quadripolar mass spectrometer DSQ I (Thermo-Electron). The transfer line was set at 250 °C. Classical electron ionization (70 eV) was used; the mass range was limited to 40–240 m/z so that the acquisition frequency (experimentally measured around 30 Hz) was compatible with GCxGC data. Excalibur software was used for acquisition; then, data were imported into Hyperchrom S/W software for the visualization of 2D-chromatograms. Hyperchrom S/W offers the possibility to export the total ionic current (TIC) as a 2D chromatogram in a text file. This TIC matrix (807 × 161) could then be read with Matlab (R2008b, The MathWorks, Natick, MA, USA) for further data processing.

2.2. Gases

Liquid CO₂ was of industrial grade and purchased from Air Liquide (Le Plessis Robinson, France). Pure gases, i.e. helium

Table 1
Sources of variability and their modalities.

Tobacco type	Burley		Virginia		Oriental	
	SDE	SFE	SDE	SFE	SDE	SFE
Number of different samples	4	4	4	4	4	12
Number of extractions	1	2	1	2–3	1	2
Number of injections	2–3	2–5	2–3	2–3	2–3	2–3

(99.9995%) and CO₂ (99.999%, for supercritical fluid extraction) were purchased from Messer (Asnières, France).

2.3. Tobacco extracts

Three types of tobacco were considered: Burley, Virginia and Oriental. For each type, different samples were available, corresponding to different batches or different origins. The extracts were obtained either by steam distillation extraction (SDE data set) or supercritical fluid extraction (SFE data set).

2.3.1. SDE data set

A first set of tobacco hexane extracts was provided by the Imperial Tobacco group. They were obtained by SDE, i.e. the "Likens Nickerson" process [25] directly from tobacco leaves cut into small pieces. Only one extract was available per sample, and several injections were made for each sample, leading to 44 chromatograms.

2.3.2. SFE data set

Another set of extracts was generated at the LSABM by SFE. Extractions were performed with a Suprex SFE Prepmaster GA apparatus (Pittsburgh, PA, USA) using 5 mL SFE cells. Extractions were performed in static mode for 5 min, and in dynamic mode for 30 min with a CO₂ density of 0.4 at a temperature of 150 °C [26]. Extracted compounds were collected after the pressure was released by bubbling in three successive vials, each filled with 3 mL of hexane:ethyl acetate (50:50, v/v) mixture. The injected sample corresponded to the first 3 mL, as it was checked that no compound was present in the last two vials. Several extracts were available per sample, and several injections were made for each sample, leading to 101 chromatograms.

Table 1 summarizes the sources of variability of the used data sets. In order to limit chromatographic variability, all analyses were performed in the shortest possible period of time for each data set. In [26], the GCxGC repeatability was characterized by a RSD of 2.5%, while extraction to extraction variability (measured by global quantification) ranged generally from 5 to 15%.

3. Data processing

Each data set (SDE and SFE) was processed independently. In fact, the two extraction techniques lead to quite different results, which cannot be pooled together.

Matlab routines were developed for the following processing steps, except for time alignment which was implemented in the C programming language providing faster computation, and integrated in Matlab as a MEX-file.

3.1. Data pre-processing

In order to compensate for the variability inherent to GC injection, a pre-processing of the data was required.

3.1.1. Background correction

Correcting the baseline of a one-dimensional signal is a classic pre-treatment before the global comparison of chromatograms (fingerprinting). In GCxGC, one-dimensional acquired signals being

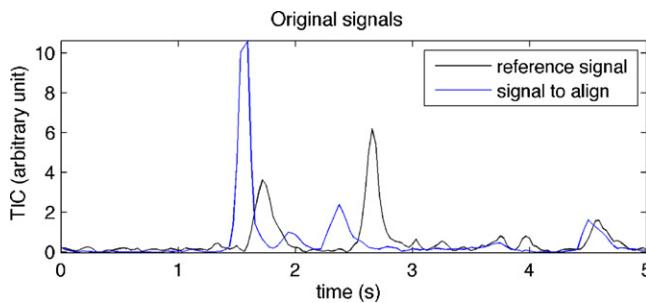


Fig. 1. Unaligned background corrected and normalized signals (second dimension).

turned into images, a baseline correction amounts to a background correction. Our method of background subtraction was inspired by a DNA microarray pre-processing algorithm [27]. First, the image was divided into N rectangular zones (typically $N=60$). The pixels of each zone were ranked and the mean intensity of the lowest 2% was chosen as the background level of the studied zone. Then, a smoothing adjustment was performed: distances between each pixel of the image and the N zone centers were computed and a weighted sum was calculated based on the reciprocal of a constant plus the square of the distances to all the zone centers. The smooth backgrounds obtained in this manner were subtracted to the chromatograms.

3.1.2. Normalization

Differences in chromatograms due to variations in sample amount and/or detector response must be corrected. GCxGC Intensities being approximately linear with respect to amount, normalization could be implemented by a simple multiplicative scaling. The scale factor was set so that the mean intensity of each chromatogram equaled 1.

3.2. Dynamic time warping

The short separation along the second chromatographic dimension (5 s), is very sensitive to experimental conditions (pressure, temperature, etc.). As a result, bias in retention time occurs for a target peak from one analysis to another (see Fig. 1 in the case of two Virginia SDE samples).

Consequently, a comparison of chromatograms with peak misalignment would be meaningless. Therefore, to align the peaks, Dynamic Time Warping (DTW) was used. Originally, this method was developed to address speech recognition issues [28]; Wang and Isenhour first proposed its application to chromatographic signals [29]. In [23], the classic DTW algorithm was applied to the alignment of a small set of SDE extracted tobacco samples. In the present work, we improved the performance of the standard algorithm by adding constraints on the warping path. Indeed, two major pitfalls may occur during the construction of the path.

First, the path may follow a staircase trajectory, presenting long, unrealistic, vertical and horizontal thresholds. In order to solve this problem, a direct constraint on the slope of the warping path has been applied, by simply preventing the succession of two horizontal or two vertical segments in the dynamic programming matrix (slope constraint of [28] $P=1$). Second, even with this first modification, the path may move too far away from the matrix diagonal. As proposed in [28], a windowing on the dynamic programming matrix prevents the path from deviating. It merely consisted in confining the path in a pipe centered on the matrix diagonal. The result of the alignment for the signals of Fig. 1 is shown in Fig. 2.

All the chromatograms have been aligned with the DTW algorithm constrained as described above. The windowing was performed with a rectangular, 40 pixels wide window. This algorithm was first developed with Matlab, and later implemented

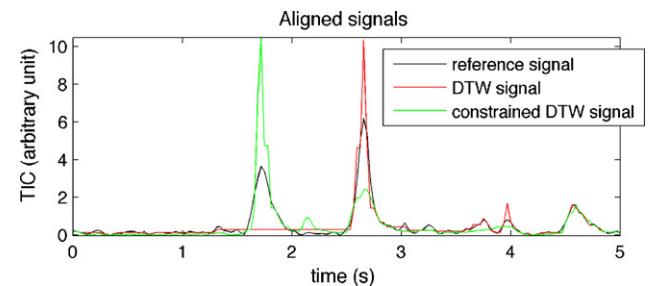


Fig. 2. Aligned signals (second dimension).

in the C language as a MEX-file in order to save computation time.

The alignment was operated on each column of the matrix corresponding to each chromatogram. It was done in a supervised fashion inside each tobacco family: for each tobacco type, one chromatogram was taken as a reference on which all the others samples were aligned. It was checked that the choice of the reference chromatogram inside each group had no decisive influence on the subsequent discriminant pixel identification.

3.3. The discriminant pixel approach

In order to find the locations that are characteristic of a given class, we developed a supervised method that evaluates the ability of each pixel of the chromatogram to discriminate between the class of interest and all the other classes. This idea was inspired by supervised classification algorithms which starts simplifying the c class classification task by dividing it into c easier subtasks: separating each class from all others [30] using the descriptors (also pixels in the case of hand-written recognition for example) that are most relevant for this separation.

Let us denote by $N_1=807$ and $N_2=161$ the number of pixels along the first and second dimensions. We consider the vector \mathbf{p}_{ij} associated to the pixel ij (for $i=1$ to N_1 , $j=1$ to N_2), whose components $\mathbf{p}_{ij}(l)$ are the pixel intensity values for each chromatogram of the data set, hence of dimension $n=n_1+n_2+n_3$, the $\{n_i\}$ being the numbers of chromatograms in each class (class 1 being Burley, class 2 Virginia and class 3 Oriental). For each class $k=1$ to $c=3$, we define a vector \mathbf{v}^k of dimension n coding for the belonging of the corresponding samples to the class of interest, i.e. whose components $\mathbf{v}^k(l)$ equal +1 for the samples belonging to the class of interest and -1 for all other samples.

The ability of the pixel ij to discriminate between class k and the others was defined as Pearson's linear correlation coefficient between \mathbf{p}_{ij} and \mathbf{v}^k :

$$r_{ij}^k = \frac{\sum_{l=1}^n (\mathbf{p}_{ij}(l) - \bar{\mathbf{p}}_{ij})(\mathbf{v}^k(l) - \bar{\mathbf{v}}^k)}{\sqrt{\sum_{l=1}^n (\mathbf{p}_{ij}(l) - \bar{\mathbf{p}}_{ij})^2 \sum_{l=1}^n (\mathbf{v}^k(l) - \bar{\mathbf{v}}^k)^2}}$$

The two vectors are schematically represented in Fig. 3 in the case where the class of interest is the Burley tobacco, i.e. the coding vector is \mathbf{v}^1 . If the correlation is close to 1, the pixel corresponds to a compound over-represented in Burley samples, and if it is close to -1, the pixel corresponds to a compound under-represented in Burley samples.

For each class $k=1$ to $c=3$, the computation was performed for all the chromatogram pixels, i.e. a correlation map r^k of dimension $N_1 \times N_2$ was obtained. A correlation threshold value must then be chosen, or a maximal number of pixels (typically 500 pixels), in order to select the most discriminant ones (i.e. with correlation closest to 1). In practice, the correlation map for a given class looks like a GCxGC chromatogram whose pixel intensities are the values

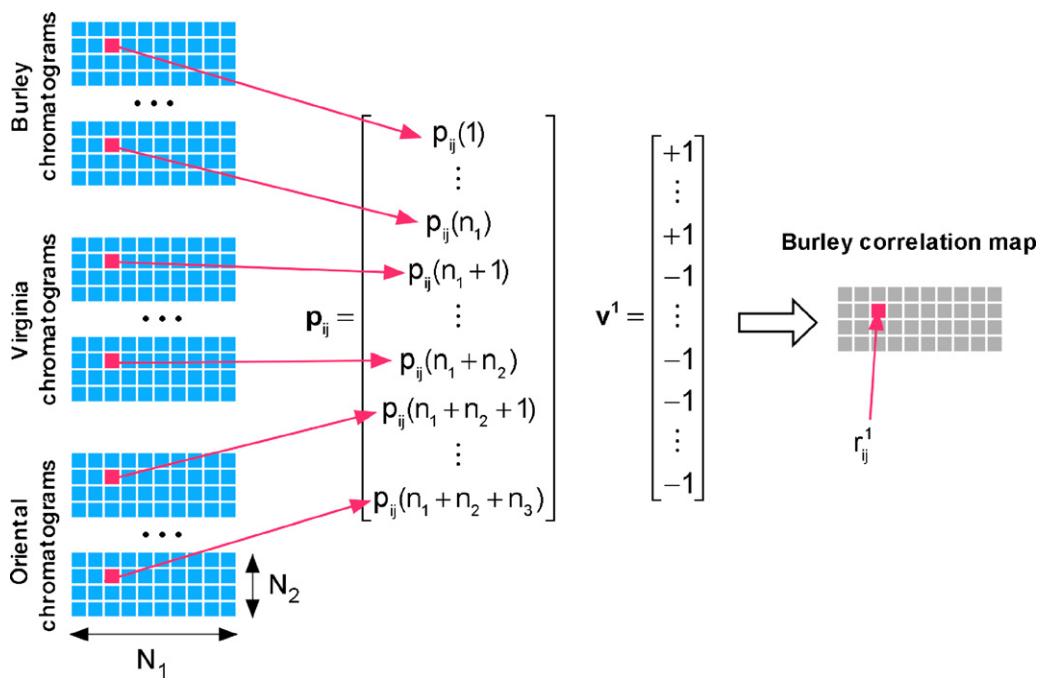


Fig. 3. Computation of the Burley correlation map (for one of its pixels).

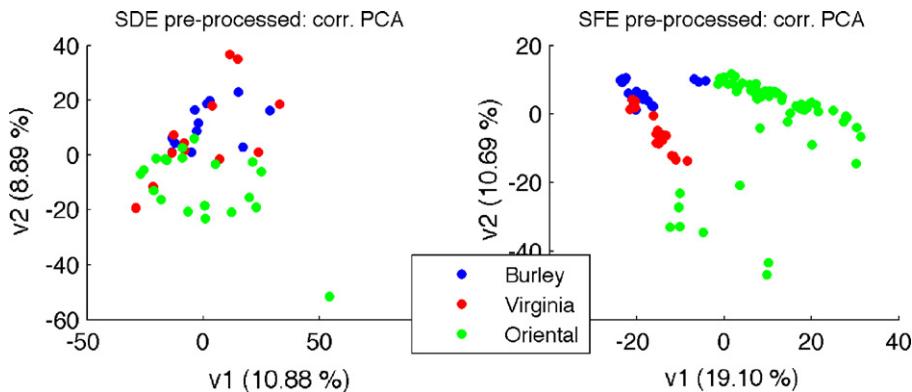


Fig. 4. PCA score plot after data pre-processing. "corr." indicates that PCA was performed with the columns of the experience matrix scaled to have unit norm. v_1 and v_2 denote the first and second principal components, and the percentages in parentheses are the amount of the original variance they account for.

of the correlation coefficients with the vector coding for the class in question.

An analogy could be naturally established between the discriminant pixel approach and the Fisher Ratio method, first proposed by Pierce et al. [31], and further developed by Mohler et al. [32]. The Fisher Ratio consists in performing a one-way analysis of variance (ANOVA) for each pixel, the modalities of the single factor being the different classes, and the Fisher Ratio being the F -statistics of the ANOVA. Pixels exhibiting large Fisher ratios correspond to large class-to-class variance with respect to within class variance, i.e. to compounds whose amount differs significantly between at least two couples of classes. Whereas the Fisher Ratio produces a single F -map, and a list of pixels for which it remains to be determined for which couples of classes they are discriminant (in the same way that an ANOVA that rejects the null hypothesis must be followed by multiple pairwise comparisons), the discriminant pixel approach directly outputs as many correlation maps as there are classes, together with the lists of the pixels corresponding to compounds over-represented (or under-represented) in each of the classes with respect to all others. In the case of only two classes however, the

Pixel Discriminant approach and the Fisher Ratio method coincide (both ANOVA and correlation significance being equivalent to that of a t -test).

3.4. Peak identification

After alignment, the coordinates of the pixels cannot be directly converted into the retention times of the corresponding compounds, because the coordinates of the pixels no longer correspond to the original retention times: the warping breaks the linear relation between pixels coordinates and retention times. Therefore, a reverse warping had to be applied to get back to the original coordinates of the pixels (their coordinates before DTW). This operation was simply made by reversing the warping path of the considered signal calculated during the alignment.

Then, the discriminant pixels could be related to the peaks they belong to (several pixels may correspond to a single peak). A simple search of local maximum around discriminant pixels allowed to find the maximum of the considered peak, and to define a chromatographic peak zone using the Hyperchrom software.

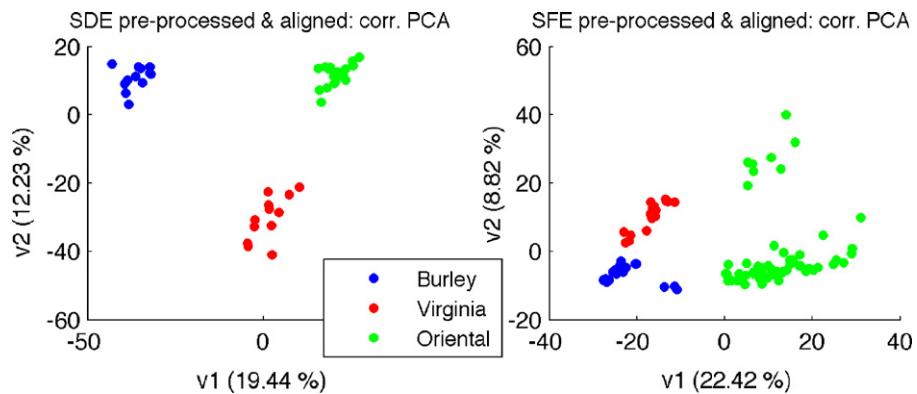


Fig. 5. PCA score plots after the time alignment.

4. Results and discussion

The effect and the relevance of each step of the data processing was evaluated according to the quality of the separation of the tobacco types on PCA score plots.

4.1. Effect of the data pre-processing

The results of PCA applied to the pre-processed data (intensity normalization and background correction) are presented on Fig. 4. The different tobacco types were quite overlapping (particularly for SDE data), which confirmed the need for a peak alignment process.

4.2. Effect of the time alignment

Fig. 5 shows that the alignment enhanced the discrimination between the different kinds of tobaccos: the three classes were quite well separated after DTW.

4.3. Discriminant pixel selection

The correlation maps described previously have been computed for each data set and for each tobacco type; they are shown in Fig. 6 (SDE) and Fig. 7 (SFE).

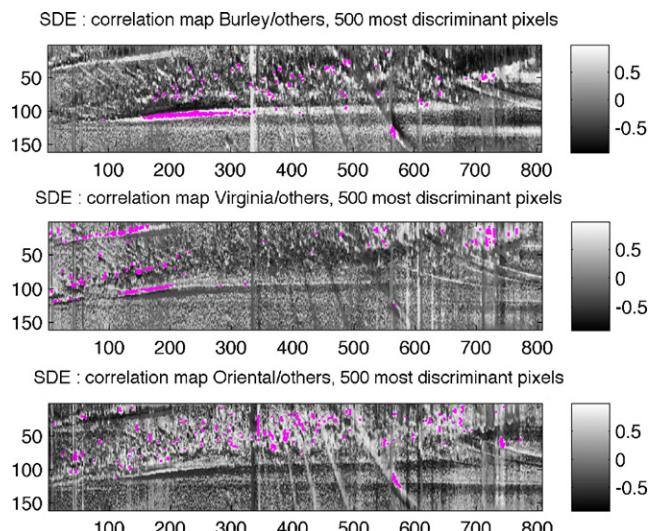


Fig. 6. SDE correlations maps and 500 most discriminant pixels (magenta points). Axis units are given in pixels because the alignment breaks the linear relationship with time along the second dimension.

The color code for the maps is given by the column on the right side of Figs. 6 and 7: white corresponds to a maximum correlation of 1 with the coding vector of the class of interest (i.e. to the pixels that are most discriminant for that class, and correspond to compounds that are over-represented in this class), and black to -1 for a minimum correlation (these pixels are also discriminant, but correspond to compounds that are under-represented in the class). Grey corresponds to correlation values around zero, i.e. non-discriminatory pixels. As opposed to the Fisher Ratio approach, which produces a single map whatever the number of classes, the discriminant pixel approach provides a correlation map for each class (here each of the three tobacco types). Whereas for a significant pixel of the F -map, it remains to be determined for which couples of classes it is discriminant, this information is directly available on the correlation maps.

A first comparison of the discriminant pixels identified on SFE and SDE extracts showed a partial overlap. However, a comprehensive comparison could not be carried out, because the two sets of experiments have been made at quite different times and on different sides.

The selected pixels may further be used as a new set of descriptors for the PCA: Instead of taking the 129927 pixels of the chromatograms as descriptors, only the union of the 1500 (3 kinds of tobaccos \times 500 pixels for each kind) selected pixels was retained (see Fig. 8). The separation between the tobacco types appeared then obvious: the discriminant pixel selection has nat-

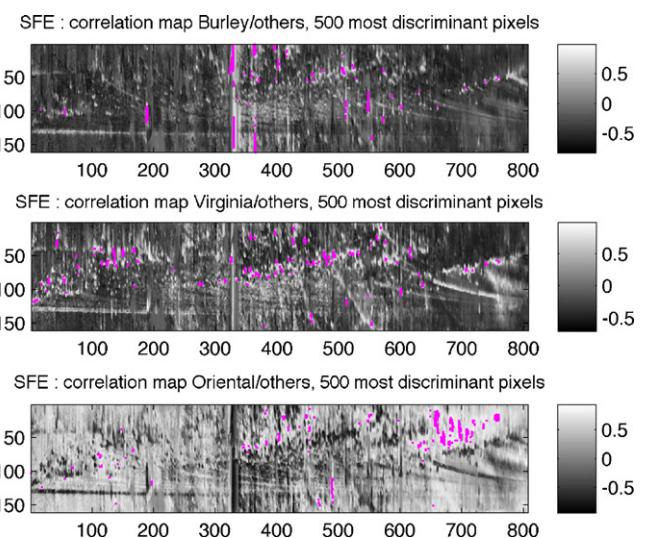


Fig. 7. SFE correlations maps and 500 most discriminant pixels. Axis units are in pixels.

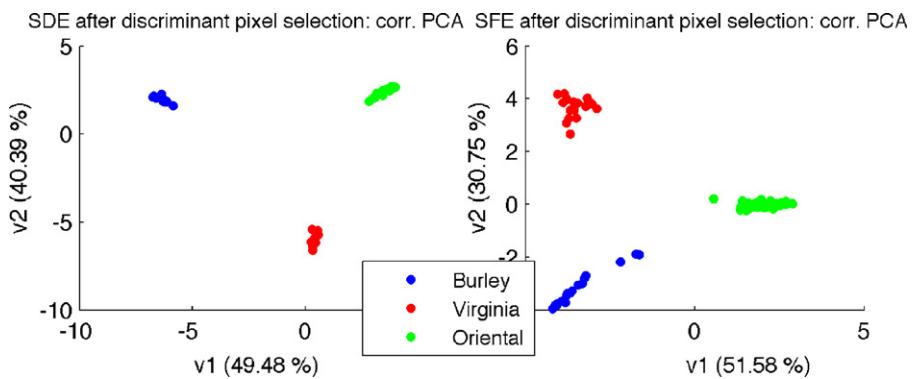


Fig. 8. PCA after discriminant pixel selection.

urally enhanced the discrimination between the chromatograms of each type by focusing on the differences. Moreover the percentage of information brought by the two first PCs increased from 30% before pixel selection to 90% afterwards.

4.4. Discriminant components

After a reverse time warping the discriminant pixels could be transferred back on the original chromatograms, and using the Hyperchrom peak finding algorithm, it led to the definition of a short list of discriminant peaks. This short list could be linked to mass spectrometry data in order to identify the discriminant chemical compounds. An example is shown in Fig. 9 on a SFE Burley sample. These potential chemical markers are among the usual compounds found in tobacco extracts.

4.5. Discussion and outlooks of the discriminant pixel strategy

The whole strategy from the raw data to the correlation maps is fully automated and takes less than 1 min on a standard computer for a set of a hundred chromatograms.

However, the proposed discriminant pixel approach needs to be further optimized. In particular, it is still necessary for the analyst to define the number of the most discriminant pixels manually. In the present paper for example, the same number was chosen for all tobacco classes, but this may not reflect the reality. In fact, when computing the significance of the correlations (by assuming Gaussian data), and accounting for multiple testing with a Bonferroni correction, we observed that on SDE data, similar numbers of pixels were considered discriminant for the three classes, but that they were very different on the SFE data set, with many more discriminant pixels for Oriental tobacco. This is in accordance with the fact

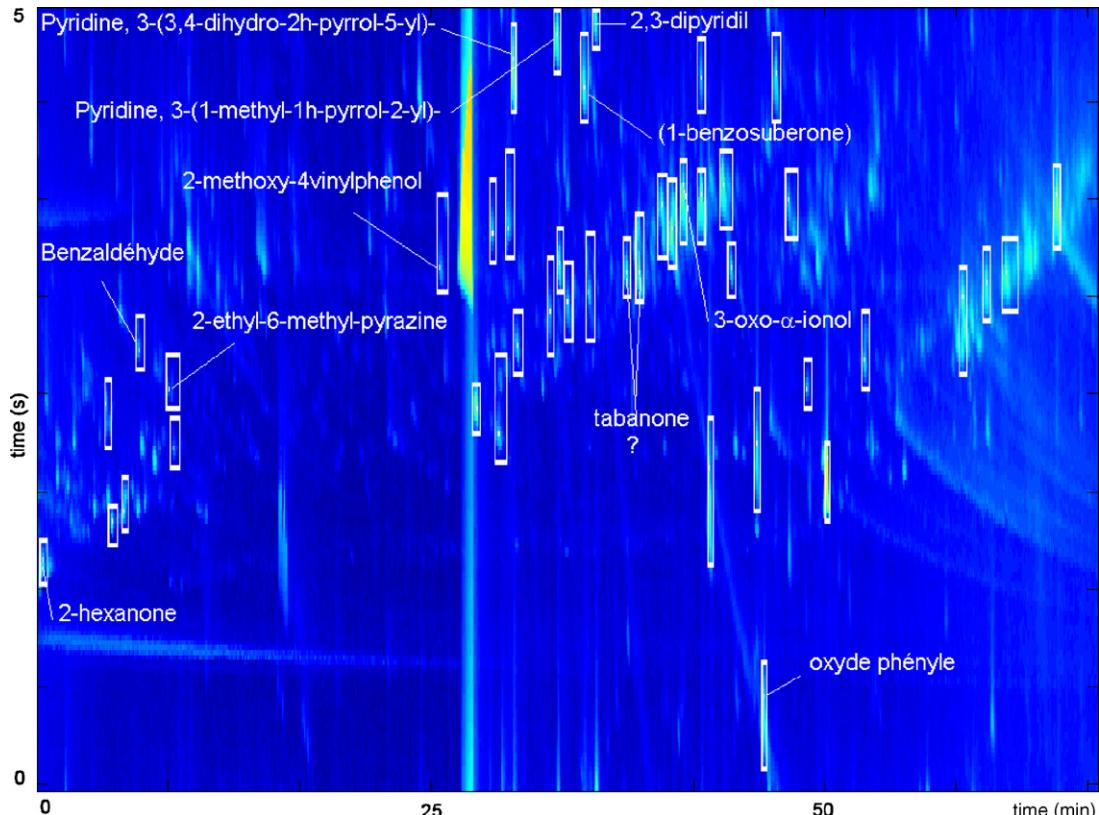


Fig. 9. Discriminant components identified thanks to mass spectrometry data.

that Oriental tobaccos are richer in volatile compounds than the other tobacco types and that SFE is more likely to extract than SDE.

Another issue that needs to be addressed is the way the different variability sources are taken into account. As summarized in Table 1, the variability of the samples has three sources: the origin, the extraction and the injection. Here, all these sources have been put at the same level. A future task will be to evaluate their magnitude, and according to the results, to determine whether and how the discriminatory power evaluation made today should be modified.

5. Conclusion

This study demonstrates the relevance of a chemometric strategy to interpret and compare GCxGC chromatograms in an automatic way. PCA allowed checking the quality and the relevance of the different data processing steps we used. The discriminant pixel approach authorized using a whole data set in a supervised fashion in order to identify areas of the chromatograms that were discriminant for each sample class. Ultimately, the discriminant pixel approach coupled with mass spectrometry data allowed the identification of candidate chemical markers specific of a given tobacco class.

Beyond the discrimination of origin and grades inside a tobacco type, this opens wide outlooks in the improvement of automatic processing strategies to interpret GCxGC chromatograms. In parallel, other analytical problematic requiring the use of comprehensive chromatographic separation, environmental for example, remain to be investigated to demonstrate the flexibility and the wide applicability of the proposed strategy. Moreover, spiking strategies remain to be applied to tobacco samples for validation purposes [31]. The complexity of GCxGC chromatograms of plant extracts requires nevertheless a thorough study of the spiking strategy to ensure that it is compatible both with the sample handling procedure and the dense coverage of the separation space.

References

- [1] M. Adahchour, J. Beens, R.J.J. Vreuls, U.A.T. Brinkman, TrAC Trends Anal. Chem. 25 (2006) 821.
- [2] C. Cordero, P. Rubiolo, B. Sgorbini, M. Galli, C. Bicchi, J. Chromatogr. A 1132 (2006) 268.
- [3] T. Gorecki, J. Harynuk, O. Panic, J. Sep. Sci. 27 (2004) 359.
- [4] R. Shellie, P. Marriott, Flavour Fragr. J. 18 (2003) 179.
- [5] M. Adahchour, J. Beens, R.J.J. Vreuls, U.A.T. Brinkman, TrAC Trends Anal. Chem. 25 (2006) 438.
- [6] M. Adahchour, J. Beens, R.J.J. Vreuls, U.A.T. Brinkman, TrAC Trends Anal. Chem. 25 (2006) 726.
- [7] C. Vendeville, F. Bertoncini, L. Duval, D. Thiébaut, M.C. Hennion, J. Chromatogr. A 1056 (2004) 155.
- [8] B. d'Acampora Zellner, A. Casilli, P. Dugo, G. Dugo, L. Mondello, J. Chromatogr. A 1141 (2007) 279.
- [9] R. Shellie, P. Marriott, A. Chaintreau, Flavour Fragr. J. 19 (2004) 91.
- [10] C. Danielsson, K. Wiberg, P. Korttar, S. Bergek, U.A.Th. Brinkman, P. Haglund, J. Chromatogr. A 1086 (2005) 61.
- [11] J.-F. Focant, A. Sjödin, D.G. Patterson Jr., J. Chromatogr. A 1040 (2004) 227.
- [12] S. de Koning, E. Kaal, H.-G. Janssen, C. van Platerinkd, U.A.Th. Brinkman, J. Chromatogr. A 1186 (2008) 228.
- [13] P.Q. Tranchida, A. Giannino, M. Mondello, D. Sciarrone, P. Dugo, G. Dugo, L. Mondello, J. Sep. Sci. 31 (2008) 1797.
- [14] C. Vendeville, R. Ruiz-Guerrero, F. Bertoncini, L. Duval, D. Thiébaut, M.-C. Hennion, J. Chromatogr. A 1086 (2005) 21.
- [15] C. Vendeville, F. Bertoncini, D. Espinat, D. Thiébaut, M.-C. Hennion, J. Chromatogr. A 1090 (2005) 116.
- [16] K.M. Pierce, J.C. Hoggard, R.E. Mohler, R.E. Synovec, J. Chromatogr. A 1184 (2008) 341.
- [17] N.-P. Vest Nielsen, J.M. Carstensen, J. Smedsgaard, J. Chromatogr. A 805 (1998) 17–35.
- [18] C.G. Fraga, B.J. Prazen, R.E. Synovec, Anal. Chem. 73 (2001) 5833.
- [19] K.M. Pierce, J.L. Hope, K.J. Johnson, B.W. Wright, R.E. Synovec, J. Chromatogr. A 1096 (2005) 101.
- [20] K.M. Pierce, L.F. Wood, B.W. Wright, R.E. Synovec, Anal. Chem. 77 (2005) 7735.
- [21] F. Suits, J. Lepre, P. Du, R. Bischoff, P. Horvatovich, Anal. Chem. 80 (2008) 3095.
- [22] D. Zhang, X. Huang, F.E. Regnier, M. Zhang, Anal. Chem. 80 (2008) 2664.
- [23] J. Vial, H. Noçairi, P. Sassi, S. Mallipattu, G. Cognon, D. Thiébaut, B. Teillet, D.N. Rutledge, J. Chromatogr. A 1216 (2008) 2866.
- [24] A. Hyvärinen, E. Oja, Neural Networks 13 (2000) 411.
- [25] S.T. Likens, G.B. Nickerson, Am. Soc. Brew. Chem. Proc. 22 (1964) 5.
- [26] J. Vial, D. Thiébaut, P. Sassi, M.S. Beldean-Galea, M.J. Gomez Ramos, G. Cognon, S. Mallipattu, B. Teillet, M. Bouzige, J. Chrom. Sci. 48 (2010) 267.
- [27] Affymetrix Statistical Algorithms Description Document: <http://www.affymetrix.com/support/technical/whitepapers/sadd.whitepaper.pdf>.
- [28] H. Sakoe, S. Chiba, IEEE Trans. Acoust. Speech Signal Process 26 (1978) 43.
- [29] C.P. Wang, T.L. Isenhour, J. Anal. Chem. 59 (1987) 649.
- [30] S. Knerr, L. Personnaz, G. Dreyfus, IEEE Trans. Neural Networks 3 (1992) 962.
- [31] K.M. Pierce, J.C. Hoggard, J.L. Hope, P.M. Rainey, A.N. Hoofnagle, R.M. Jacik, B.W. Wright, R.E. Synovec, Anal. Chem. 78 (2006) 5068.
- [32] R.E. Mohler, K.M. Dombeck, J.C. Hoggard, K.M. Pierce, E.T. Young, R.E. Synovec, Analyst 312 (2007) 756.