MOL2NET, International Conference Series on Multidisciplinary Sciences
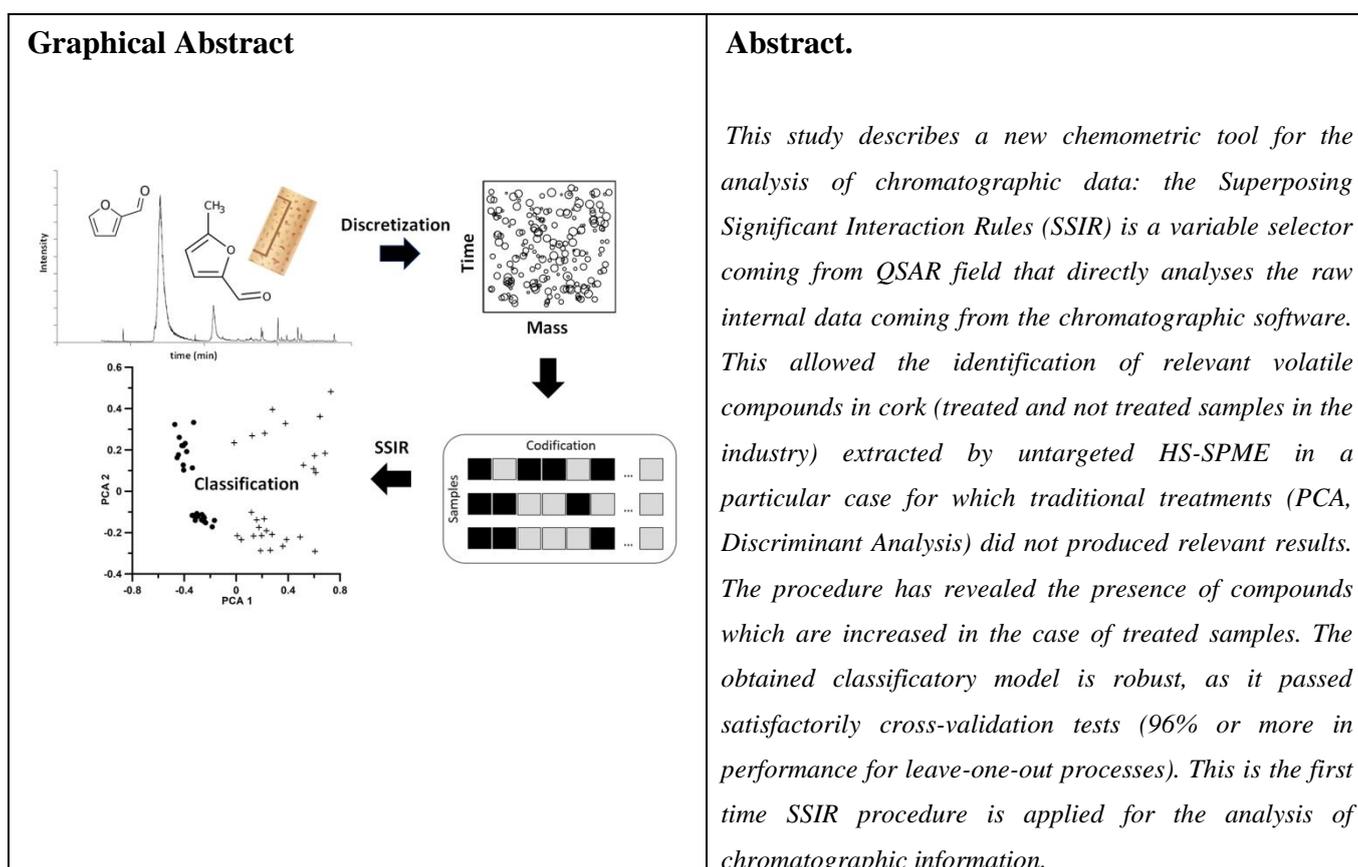
Analyzing chromatographic data from volatiles of cork with the Superposing Significant Interaction Rules (SSIR) chemometric tool

E. Besalú[a], C. Prat[b], E. Anticó[c]

[a] Institute of Computational Chemistry and Catalysis, University of Girona, 17003 Girona Spain
[b] Francisco Oller S.A., 17244 Cassà de la Selva, Spain
[c] Dept. of Chemistry, University of Girona, 17003 Girona, Spain

Graphical Abstract

Abstract.

This study describes a new chemometric tool for the analysis of chromatographic data: the Superposing Significant Interaction Rules (SSIR) is a variable selector coming from QSAR field that directly analyses the raw internal data coming from the chromatographic software. This allowed the identification of relevant volatile compounds in cork (treated and not treated samples in the industry) extracted by untargeted HS-SPME in a particular case for which traditional treatments (PCA, Discriminant Analysis) did not produced relevant results. The procedure has revealed the presence of compounds which are increased in the case of treated samples. The obtained classificatory model is robust, as it passed satisfactorily cross-validation tests (96% or more in performance for leave-one-out processes). This is the first time SSIR procedure is applied for the analysis of chromatographic information.

Introduction

The unique properties of cork, including long-lasting flexibility, hydrophobicity and gas permeability, make it still today the first choice for wine producers as a closure for bottled wine.[1,2] The complex chemical composition of cork together with its importance in the wine industry have resulted in the investigation of extractable compounds that can be released from the cork into the solution having become a subject of great interest. The composition of the volatile fraction, also referred to as the volatile signature, is an informative diagnostic tool for cork quality and sensory qualification.

Analytical methods for the study of volatiles rely on gas chromatography with mass spectrometry detection (GC-MS). A few volatiles in cork present specific odor descriptors and low odor thresholds. The compound 2,4,6-tricholoranisole (TCA) is undoubtedly the one that has been more extensively studied,[3,4] for two main reason: firstly, it is responsible for the musty, or moldy, aroma, a negative sensory attribute that can be transferred to the wine; and secondly, TCA bears an extremely low sensory threshold at around 1 ng L^{-1} in water.[4]

This study describes a new chemometric tool for the identification of relevant variables obtained from gas chromatography with mass spectrometry detection (GC-MS) analysis of the headspace in a set of cork samples. Extraction of volatiles from the headspace of aqueous cork macerates has been carried out using the solid-phase microextraction technique (HS-SPME). The authors were faced to identify volatile compounds in order to classify industrial cork samples. Some of these samples suffered a treatment consisting in a washing procedure based on water and temperature cycles in order to reduce off-flavors and decrease the amount of TCA. Unfortunately, the traditional techniques of Principal Components Analysis (PCA) and Discriminant Analysis (DA) considering common reference compounds revealed to be useless to *a priori* distinguish between both types of samples. Apart, we previously developed a procedure (an automated symbolic model builder and variable selector) called Superposing Significant Interaction Rules (SSIR) which revealed to be a good QSAR tool.[5-7] The application of this last technique permitted the identification of the relevant variables (combinations of ion masses and retention times) which are able to describe and predict the partition of our set of samples in two groups (treated and non-treated ones). The SSIR procedure allowed achieving our classificatory goal from the direct analysis of the bulk (raw, internal) chromatographic data coming from the mass spectrometer software. Even more, despite the investigated industrial procedure being intended to remove undesired compounds from the samples, for example TCA, our methodology allowed to discover that the industrial treatment in fact generated new relevant compounds. This case of success constitutes the first time SSIR procedure has been combined with chromatographic data treatment.

Materials and Methods

The pure standards of compounds (eucalyptol, MDMP, IPMP, Guaiacol, (±)-linalool, (+)-fenchol, veratrol, Camphor, Sec-IBMP, (-)-borneol, IBMP, Menthol, Methylisoborneol, α-terpineol, Benzothiazole, TCA, d5-TCA, and Geosmin) were all purchased from Sigma. Deuterium-labelled 2,4,6-trichloroanisole (d5-TCA) was used as the internal standard (IS). These compounds were employed to confirm the identification of the peaks observed in the chromatograms.

Individual stock solutions of about 500-3000 mg L^{-1} were prepared by weight in methanol (HPLC-gradient grade, Carlo-Erba Reagents) and stored at 4^oC. Other reagents used were of maximum quality available. HS-SPME experiments were manually performed using a 50/30 µm divinylbenzene/Carboxen/polydimethylsiloxane (DVB/CAR/PDMS) fiber from Supelco. Before use, the fiber was conditioned according to the manufacturer's instructions to remove contaminants and stabilize the solid phase. Five mL of sample solution (aqueous cork macerate) were placed in 15 mL amber glass vials containing 1.2 g of NaCl. Next, 100 µL of the IS solution in methanol (3.2 µgL^{-1}) was added. Finally, the vials were closed and introduced in the water bath. The fiber was exposed at 50^oC for 30 min to the headspace above the aqueous solution. Constant stirring was applied during the extraction process. After the completion of sampling, the fiber was pulled into the needle and the SPME device was removed from the vial and inserted into the injection port of the GC for thermal desorption and analysis.

Gas chromatographic analyses were performed with a Trace GC 2000 coupled to a PolarisQ ion trap mass spectrometer detector (Thermo Scientific). A TG-5SIL MS capillary column (30m×0.25 mm i.d.; 0.25 µm film thickness) (Thermo Scientific) was used and the carrier gas was 99.9990 % pure helium (Abelló) at a constant inlet flow rate of 1 mL·min^{-1}. The chromatographic conditions used were adapted from previous studies.[8] MS analyses were conducted in scan mode with two m/z intervals: from 40 to 250 amu (from min 3 to 19 min) and from 100 to 350 amu (from 19 min to the end of the chromatographic run). Ionization was performed in the electron impact mode at 70 eV. The acquisition of chromatographic data was performed using Xcalibur 1.4 software (Thermo Scientific).

The available samples consisted of cork discs obtained before (28 samples) and after (28 samples) the temperature and humidity cycles addressed to eliminate TCA following a particular industrial process. After the treatment, TCA concentration diminished considerably but an increase in vanilla, biscuit and sweet notes was detected after the washing cycle. This is the main effect that is the origin of this study.

The first action taken was to explore the chromatographic results and peak areas associated to the previously mentioned compounds. Relative peak areas (area of the compound divided by the area of IS) were calculated and Principal Component Analysis (PCA) and Discriminant Analysis (DA) was performed using Minitab[9] and in-house software. This traditional procedure did not provided with satisfactory results due to the fact that partition into treated and non-treated samples could not be reproduced. So, another kind of analysis was conducted relying on SSIR procedure, as it is explained next.

The signal was recorded and discrete values were obtained consisting in all the contributions of m/z fragments at the respective times. This information was easily obtained from the ASCII files provided by the Xcalibur software. The time range considered was from 6 to 17 minutes and only ions with amu 49-170.5 were considered. For each sample, all the time and mass registers were reorganized in a rectangular grid of time and mass combination (see the intensities represented in the schematic bubble chart of Figure 1a). The 33 time partitions were 20 seconds (1/3 min) wide. In turn, the mass interval was partitioned into 243 segments centered at half integer and integer mass values. Therefore, the original rectangular grid was made up of 33x243=8019 nodes or variables for each sample, with each node being the combination of a couple of time and ion mass intervals (see a symbolic representation in a simple schematic 10x10 grid in Figure 1b). All the intensity registers were normalized by re-scaling and setting the mean of intensities to an arbitrary value of 1000 units per sample. All these intensity values were then added in the corresponding nodes of Figure 1b. Therefore, the available data descriptors per sample are a two-dimensional grid collecting a sum of intensities in each node.
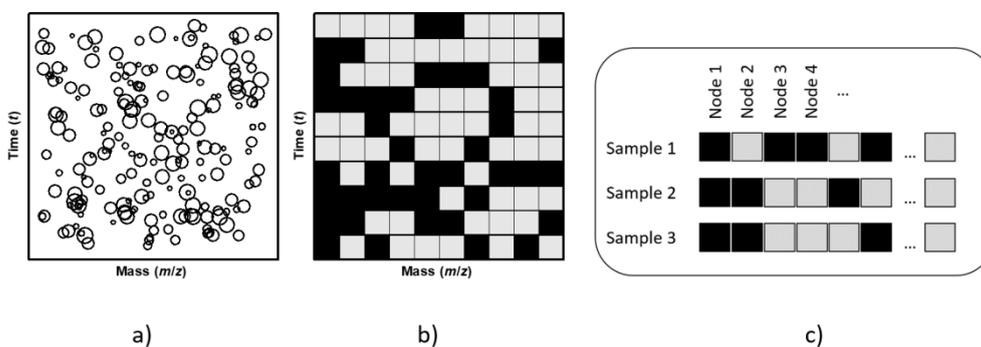

Figure 1. Discretization process of the chromatographic variables: a) bubble graph representation of the peak signals; b) after gridding, each node is assigned to a binary label: low level (grey) or high level (black); c) representation of the discretized grid variables for 3 samples. See text for more details.

The SSIR method deals only with categorical variables or nodes. Here, only binary variables have been considered as it was decided to dichotomize the values of the collected intensities in each node. The top 1% percentile values constituted the high level (black squares in Figures 1b and 1c) and the remaining ones constituted the low level (grey squares). So, each sample is being codified by a string of levels (Figure 1c). The goal is to select those grid nodes that correlate with the previously defined dichotomization label of the samples.

Figure 1c shows a simplified representation of the discretized grid variables and the codification series of low and high binary levels for some samples. After the depuration of irrelevant nodes, 237 were selected as the working ones which act as samples descriptors. The SSIR procedure detects which representative binary nodes are relevant to classify the samples. In general, SSIR algorithm generates

rules of distinct orders. Here rules of order one were selected. This means that all of the 237 discretized variables were inspected one at a time, and a *p*-value was attached to each one according to the considerations that follow.

If *A*=58 is the total number of available samples and *T*=28 is the number of treated ones, given a binary variable or node, it is counted the number of samples that exhibits a particular variable level (value of 0 or of 1), say a, and how many of these samples are treated, say *t*. Then the following probabilistic calculation[10] gives the probability to encounter t or more treated samples if $a \geq t$ are randomly selected from the set of *A*:

$$p = \sum_{i=t}^{min(T,a)} P(i, a; T, A)$$

where

$$P(t, a; T, A) = \frac{\binom{T}{t}\binom{A-T}{a-t}}{\binom{A}{a}} \, ,$$

being $t \leq a \leq A$ and $t \leq T \leq A$.

Each time the *p*-value is equal or lesser to a predefined threshold, the variable receives a vote. After all the variables are inspected, the most voted variables are expected to correlate with the subsets of treated and not treated samples. Consequently, these variables are being selected to enter in an ulterior classification or prediction procedure.

Results and Discussion

The threshold value chosen is this case was of 0.0005, and a total of 13 variables exhibited a *p*-value equal or lesser than it. Table 1 lists the selected variables. After the above described discretization, each variable is a combination of an interval of ion mass and a time interval.

It has been verified that normally there is a correlation with the increase in the intensities of the treated samples with respect to the non-treated ones. Hence, the related ions are to be attached to compounds that were generated after sample treatment. This was surprising as the industrial treatment was conducted with the aim of removing compounds.

Variable #	Time interval (min)	m/z interval (amu)
1	[7.3,7.7)	95
2	[6.7,7.0)	95
3	[10.3,10.7)	110
4	[10.3,10.7)	109
5	[7.0,7.3)	97
6	[7.0,7.3)	67
7	[6.7,7.0)	96
8	[7.3,7.7)	96
9	[10.7,11.0)	109
10	[10.7,11.0)	110
11	[7.0,7.3)	51
12	[8.7,9.0)	51
13	[9.0,9.3)	78

Table 1. Selected variables by SSIR. The mass/charge (m/z) interval is identified by its center, being the radius of 0.25 amu.

The 13 selected variables were considered for modeling purposes. The respective original sum of collected peak intensities per node was entered as descriptors in a PCA calculation. The scores map generated is depicted in Figure 2.
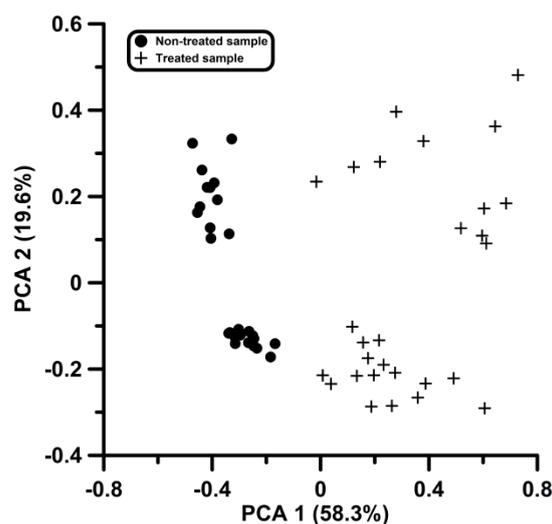

Figure 2. PCA representation of the treated and non-treated samples from the selected variables by SSIR. See text for more details.

Figure 2 reveals that the first component acts as a discriminant function for the classification of both types of samples. The discriminant model correctly classified all but one of the samples. Thus, the sensitivity of the model was about 96% and the specificity was 100%. In order to test for the robustness of such a model, a leave-one-out cross-validation calculation was conducted getting a 96%

performance. Leave-one-out models involving 3 to 5 principal components correctly classified all of the samples (100% performance).

The knowledge of the selected mass and times intervals allowed for the identification of relevant compounds related to the distinct sensorial results attached to both kinds of samples, namely, furfural and 5-methylfurfural. The identity of the compounds was verified by injecting pure standards.

Conclusions

Given a challenging problem of samples classification for which traditional numerical treatment of chromatographic data was inconclusive, it has been demonstrated that the application of SSIR procedure lead to identify a set of 13 relevant variables. These variables allowed for a PCA and DA classification with good figures of merit. The inspection in detail of these variables in the original chromatographs leads to the conclusion that three compounds are responsible for the separation of the 56 cork samples into the treated and non-treated groups presenting different sensory descriptors.

References

1. Pereira, H. The Rationale behind Cork Properties: A Review of Structure and Chemistry. Bioresources, 2015, 10 (3), 1-23.

2. Silva, S. P.; Sabino, M. A.; Fernandes, E. M.; Correlo, V. M.; Boesel, L. F.; Reis, R. L. Cork: properties, capabilities and applications. International Materials Reviews, 2005, 50 (6), 345-365.

3. Soleas, G. J.; Yan, J.; Seaver, T.; Goldberg, D. M. Method for gas chromatographic assay with mass selective detection of trichloro compounds in corks and wines applied to elucidate the potential cause of cork taint. *J. Agric. Food Chem.* **2002**, *50*, 1032-1039.

4. Amon, J. M., Vandepeer, J. M.; Simpson, R. F. Compounds responsible for cork taint in wine. *Wine Ind. J.* **1989**, *4*, 62-69.

5. Besalú, E. Fast Modeling of Binding Affinities by means of Superposing Significant Interaction Rules (SSIR) method. Int. J. Mol. Sci. 2016, 17, 827.

6. Besalú, E.; Pogliani, L.; De Julián-Ortiz, J. V. The Superposing Significant Interaction Rules (SSIR) method in Applied Chemistry and Chemical Engineering, Volume 4 (Experimental Techniques and Methodical Developments). Haghi, A. K.; Pogliani, L.; Castro, E. A.; Balköse, D.; Mukbaniani, O. V.; Chia, C. H. (Editors). Apple Academic Press (AAP), Waretown, New Jersey, 2017. ISBN: 9781771885874. E-Book ISBN: 9781315207636

7. Besalú, E.; Pogliani, L.; De Julián-Ortiz, J. V. Superposing Significant Interaction Rules (SSIR) method: A simple procedure for rapid ranking of congeneric compounds. Croat. Chem. Acta 2016, 89 (4), 481-492.

8. Prat, C.; Trias, R.; Culleré, L.; Escudero, A.; Anticó, E.; Bañeras, L. Off-odor compounds produced in cork by isolated bacteria and fungi: a gas chromatography-mass spectrometry and gas chromatography-olfactometry study. J. Agric. Food Chem. 2009, 57, 7473-7479.

9. MINITAB version 14 for Windows, Minitab Inc., State College, PA, 2004.

10. Mendenhall, W., Sincich, T. (1995). Statistics for Engineering and the Sciences. Prentice-Hall, Englewood Cliffs, NJ.