Tutor

Dr. Xavier Saurina Purroy
Departament d'Enginyeria Química i Química Analítica

Grau de Química

Treball Final de Grau

Characterization of sparkling wines by liquid chromatography
Caracterització de Caves mitjançant cromatografia de líquids

Berta Torres Cobos

June 2018

UNIVERSITAT DE BARCELONA

B:KC Barcelona Knowledge Campus
Campus d'Excel·lència Internacional

I could not live without Champagne. In victory I deserve it, in defeat I need it.

Winston Churchill

En primer lloc vull agrair al meu tutor, Xavier Saurina, tota l'ajuda i coneixements que m'ha transmès, el temps dedicat, la paciència davant els errors i la proximitat en el tracte. Sense la seva dedicació i entrega aquest treball no hagués estat possible.

També vull donar les gràcies a tots els companys del departament, els quals des del primer moment m'han fet sentir com a casa i sempre m'han proporcionat ajuda i recolzament.

Finalment, agrair a la meva família i amics per estar sempre al meu costat, en els bons i mals moments, al llarg d'aquests 4 anys.

REPORT

CONTENTS

1. SUMMARY

Cava is an exceptional kind of sparkling wine with a protected designation of origin (PDO) status from Spain. This beverage is made by the Champanoise method, which requires a secondary fermentation in the bottle.

Polyphenols are one of the principal components of Cava. These substances not only are responsible for some of the unique properties of Cava but, also, can be used as chemical descriptors to discriminate among samples of different varieties and compositions.

In the present project, 10 types of Cavas from different varieties of grapes were characterized according to their polyphenolic profile, which was obtained by high performance liquid chromatography with UV-visible detection at different wavelengths (250, 280, 310, 370 nm). A total of 18 polyphenols were studied.

Two different approaches (chromatograms and integrated peak areas, also referred to as *fingerprinting* and *profiling* approaches, respectively) were used to build principal component analysis (PCA) models focused on the differentiation among the samples. Both models presented an excellent discriminant ability. Besides, with the profiling model the most characteristic polyphenolic compounds of each kind of Cava were identified.

Finally, the chromatograms registered at 310 nm were used to build partial least squares discriminant analysis (PLS-DA) models which allowed the classification of the Cava samples according to their type, and "unknown" samples belonging to a new test set were correctly assigned to their corresponding classes.

Keywords: Polyphenols, Cava, PCA, PLS-DA, HPLC, UV-visible

2. RESUM

El Cava és un tipus de vi escumós de qualitat amb denominació d'origen protegida (DOP) procedent d'Espanya. Aquesta beguda és produïda pel mètode Champanoise, el qual requereix una fermentació secundària a l'ampolla.

Els polifenols són un dels principals components del Cava. Aquestes substàncies no només són responsables de moltes de les propietats úniques del Cava, sinó que també es poden emprar per a discriminar entre mostres de diferents varietats i composicions.

En el present treball, es van caracteritzar 10 tipus de caves procedents de diferents varietats de raïm, segons el seu perfil polifenòlic, obtingut per cromatografia líquida d'alta resolució amb detecció UV-visible a diferents longituds d'ona (250, 280, 310, 370 nm). Es van estudiar un total de 18 polifenols.

Es van usar dos enfocaments diferents (cromatogrames i àrees de pic integrades, també anomenats *fingerprinting* i *profiling,* respectivament) per construir models d'anàlisi de components principals (PCA), per a la diferenciació entre les mostres. Ambdós models van presentar una excel·lent capacitat discriminant. A més, amb el model PCA de les àrees, es va aconseguir una identificació dels compostos polifenòlics més característics de cada classe de Cava.

Finalment, es van utilitzar els cromatogrames enregistrats a 310 nm per a construir models d'anàlisi discriminant per mínims quadrats parcials (PLS-DA) que van permetre la classificació de les mostres de Cava segons la varietat, i la correcta assignació d'un nou conjunt de mostres "desconegudes" a les classes corresponents.

Paraules clau: Polifenols, Cava, PCA, PLS-DA, HPLC, UV-Visible

3. INTRODUCTION

Sparkling wines are exceedingly outstanding beverages, produced in extremely specific conditions for the sole purpose to obtain a high-quality product that pleases the most exquisite palates.

Cava is a sparkling wine with a protected designation of origin (PDO) status from Spain. Precisely obtained after a second fermentation of wine, in the bottle itself, which confers it its peculiar fizzy texture and characteristic aromas [1].

The grape varieties designated to produce Cava are [2]:

- White grape varieties:

  - Macabeu (Viura): Traditional variety from Catalonia characterized by its large and compact bunches, thin skin and white-golden colour. It originates a Cava balanced in acidity and delicate aroma.
  - Parellada: Variety cultivated at high altitudes in Penedès, Tarragona and Conca de Barberà, with distinctive golden-green grapes which are big and compact. It is the last one to be picked because of its late maturation. It produces soft and moderated Cava.
  - Xarello: Mainly cultivated in Penedès and Tarragona, characterized by the big size of its grapes and the thickness of the skin. Rich in sugars, it provides wines with high alcoholic content. It is perfect for long aging Cava.
  - Chardonnay: The grape is small, spherical, compact and of yellow and fine skin. It is a variety of great aromatic power and high alcoholic content, also with high potential for aging.

- Red grapes varieties:

  - Trepat: Variety of early sprout and late harvest, vigorous, with thick skin and large size grapes. It produces rosé base wines with moderate alcohol content and balanced acidity.

- Pinot noir: Its grapes are small and compact, with thick skin and intense colour. It can be used to make white wine called "Blanc de Noirs".
- Garnacha: Dark purple, compact and medium-sized grapes with thin skin and juicy pulp. It originates a well-balanced, aromatic Cava with moderate acidity.

3.1. CHAMPANOISE METHOD

The traditional method used to produce Cava is known as *Champanoise*. It is original from the Champagne region in France. Once the base wine is obtained, further elaboration steps are as follows [2,3]:

1) Tirage: It is the set of operations culminating in the bottling of the wine, after having added the tirage liquor - a mix of sugar, wine and yeast - intended to cause the second fermentation in the bottle and the resulting formation of the foam that characterizes it.
   Next, the bottles must be placed in cool underground stores at a constant temperature below 15ºC, in horizontal position, and, during this phase, the second fermentation and subsequent aging occurs. During the second fermentation, the yeast acts on the sugar, thus generating tiny bubbles of carbon dioxide.
   The legal minimum requirement for this stage of fermentation and aging is nine months, although the majority of Cavas remain in the underground cellar far more time.

2) Remuage: When the aging has finished, the lees (the sediment resulting from the second fermentation) have settled and must be removed through the remuage process. The bottles are placed on inclined surfaces known as *pupitres* and are given an eighth of a full turn every day increasing gradually the angle of inclination. This process usually takes around 30 days, after which the bottles end up in an almost vertical position with the sediment having moved towards the cork.

3) Disgorging: Then, the bottles are removed from the *pupitre* and placed in an inverted position, known as “sur pointe” (upside down). The “sur pointe” bottles are placed in a machine which freezes the necks, and the frozen sediment is easily expelled.
   To obtain the different varieties of Cava, dosage, a mixture of wine and sugar, may be added in greater or lesser amounts.
   Finally, the bottles must be sealed with the final cork labelled and packed.

3.2. POLYPHENOLS

The grape skin contains a set of compounds known as polyphenols, whose main function is to act as a natural shield against the environmental factors and contribute to the sensory qualities of the fruit such as colour, astringency, bitterness, aroma...

Furthermore, polyphenols are one of the principal components of Cava and are responsible for many of its representative properties.

3.2.1. Polyphenol classification

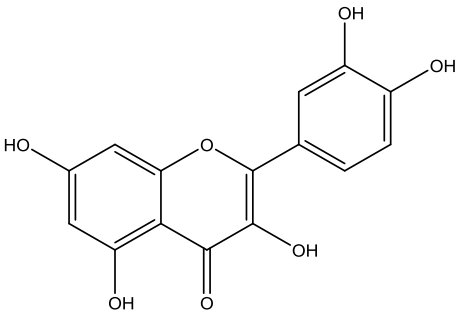
Polyphenols present a characteristic chemical structure with one or more phenolic groups per molecule. They can be classified in several classes according to the number of phenolic rings they have and the structural elements that conform these rings [4,5].

The main polyphenolic groups are flavonoids and not flavonoids.

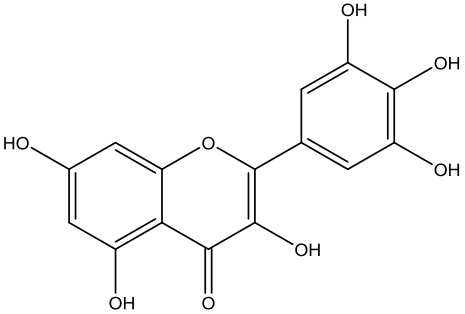
Flavonoids

Flavonoids are low molecular mass compounds that share a diphenyl pyran skeleton, that is constituted by two phenilic rings linked together through a heterocyclic pyran ring. These compounds are usually found as single molecules (aglycones) or combined with sugars (glycosides) [4,5].

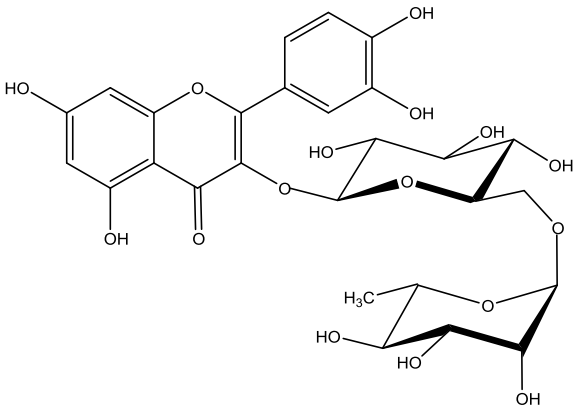
- **Flavonols:** are a subgroup of flavonoids which protect the grapes from the solar light and confer them antioxidant properties. The flavonols studied here are: quercetin, myricetin and rutin [5,6].



1



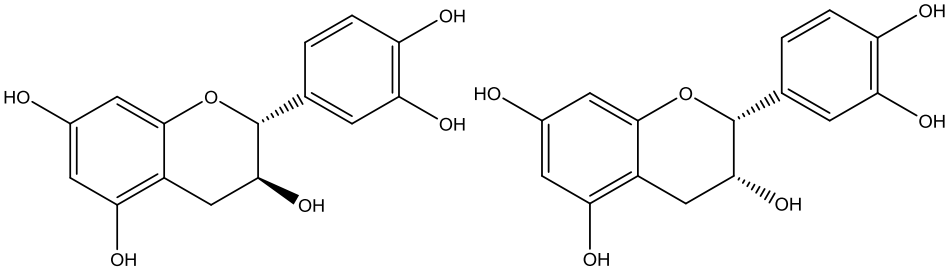
2



3

Figure 1. Chemical structures of flavonols: Quercetin (1), Myricetin (2) and Rutin (3).

- **Flavanols:** are a subgroup of flavonoids that contribute to the perception of bitterness in wine. Additionally, the studied ones, (+)-catechin and (-)-epicatechin, have an essential role in the microbial defence of the skin of the grape, exhibiting a great antioxidant power [5].

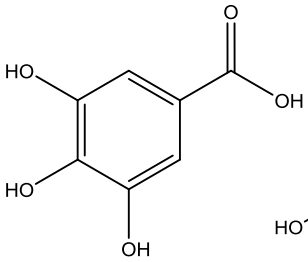


4 5

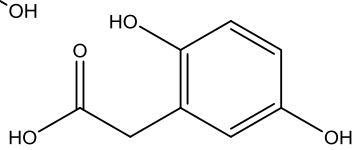
Figure 2. Chemical structures of flavanols: (+)-Catechin (4) and (-)-Epicatechin (5).

Not flavonoids

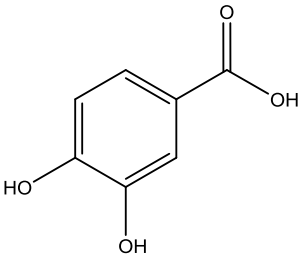
- **Phenolic acids:** are a set of compounds that contain a phenolic ring and an organic function of carboxylic acid [7]. These acids are divided into two main subsets: hydroxybenzoic acids and hydroxycinnamic acids. These compounds, which are highly abundant in red wines, have antioxidant and preservative properties. Hydroxycinnamic acids are mainly found in the skin of white grapes and do not have a contribution as important as the hydroxybenzoic acids to the wine flavour. Compounds studied in this work include: gallic acid, homogentisic acid, protocatechuic acid, gentisic acid, vanillic acid, syringic acid, ethyl gallate, caffeic acid, caftaric acid, p-coumaric and ferulic acids.



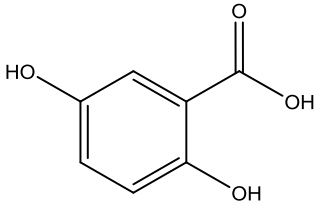
6



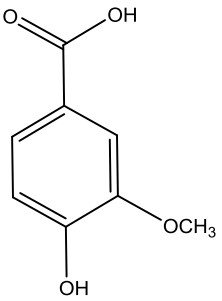
7



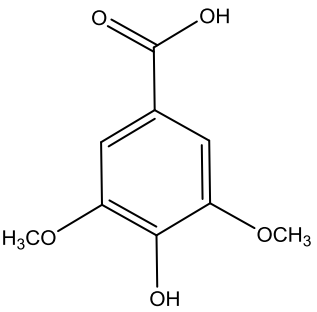
8



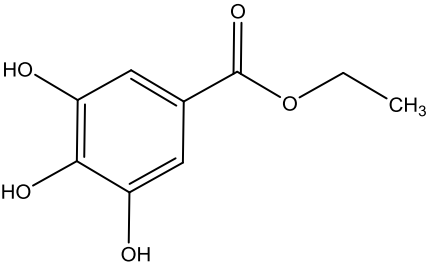
9



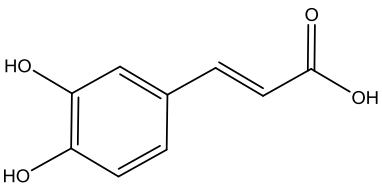
10



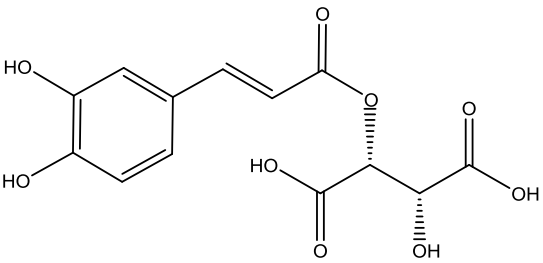
11



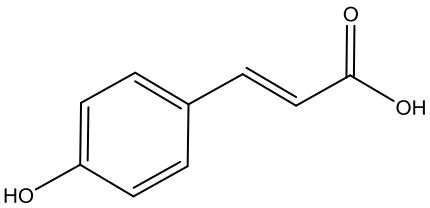
12



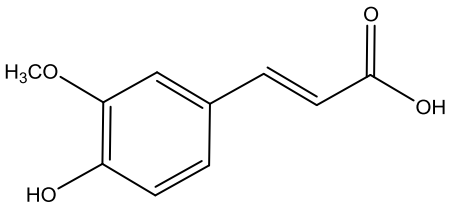
13



14



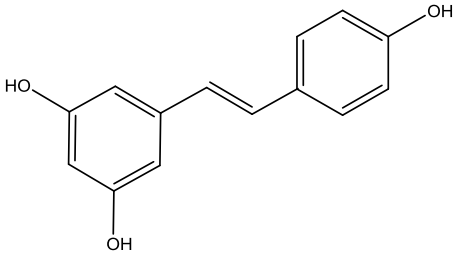
15



16

Figure 3. Chemical structures of phenolic acids. Hydroxybenzoic acids: Gallic acid (6), Homogentisic acid (7), Protocatechuic acid (8), Gentisic acid (9), Vanillic acid (10), Syringic acid (11) and Ethyl gallate (12). Hydroxycinnamic acids: Caffeic acid (13), Caftaric acid (14), p-Coumaric acid (15) and Ferulic acid (16).

- **Stilbenes:** are polyphenolic compounds displaying two aromatic rings linked by an ethene bridge [8]. The most important one in wines is resveratrol which can be found in the grape seeds and skin. However, its concentration is greater in red wines than in white ones.



17

Figure 4. Chemical structures of stilbenes: Resveratrol (17).

3.3. DETERMINATION OF POLYPHENOLS

For separation and determination of polyphenols in wine samples, high-performance liquid chromatography (HPLC) is the established technique since it allows a fast and accurate separation of the main wine polyphenolic compounds and its quantification without the need for preliminary derivatization [9-13].

Usually, separations are carried out with HPLC systems equipped with reversed phase (RP) columns, generally packed with spherical particles of silica bonded to octadecyl chain (C_{18}), and diode array detector (DAD). In Ultraviolet (UV) detection, polyphenols have characteristic absorption bands in the ranges 240-285 and 300-550 nm [9-13]. However, in the recent years, the mass spectrometry (MS) detection has gained importance thanks to its excellent performance [14,15]. The elution systems most commonly used are binary solvent systems which consist of an aqueous acidified solvent and a less polar organic solvent such as methanol or acetonitrile.

Peak identification can be made comparing retention times and UV-visible spectra of the analytes in the samples with those of pure compounds (standards). Anyway, if an exact and precise peak identification is needed, MS will be a more convenient choice [14,15].

Other less extended analytical methods for the analysis of polyphenols are capillary electrophoresis (CE) [16], infrared spectroscopy [17] and spectrophotometric methods [12,18].

4. OBJECTIVES

The aim of this work is to study whether the polyphenolic profiles obtained by HPLC with UV detection can be used to discriminate amongst sparkling wines (Cavas) made of different grape varieties. With this objective in mind, 10 different Cava blends have been analysed and compared by chemometric methods.

In order to achieve this goal, several studies have been carried out, focused on:

-Optimize data pre-treatment so variability of the analysis method does not affect the sample distribution.

-Study the sample distribution by principal component analysis (PCA) in order to identify possible discriminating factors among sparkling wines.

- Evaluate whether discrimination among classes is possible by partial least squares discriminant analysis (PLS-DA). In particular, classification of samples according to their polyphenolic composition and chromatographic profile, and assignation of new samples to their corresponding class.

5. EXPERIMENTAL SECTION

5.1. REAGENTS, STANDARDS AND SOLVENTS

The reagents and solvents used for the preparation of the HPLC mobile phases and samples were:

-Methanol ultra-high performance liquid chromatography (UHPLC) Supergradient, ACS (99,9%, CAS 67-56-1, AppliChem, Panreac, ITW Companies, Castellar del Vallès, Barcelona, Spain)

-Formic acid ACS reagent (>96%, CAS 64-18-6, Sigma-Aldrich, St. Louis, Germany)

-Mili-Q water purified with a Mili-Q system (Bedford, USA)

-Dimethyl sulfoxide (Reag.Ph.Eur.) for analysis, ACS (99,9%, CAS 67-85-5, AppliChem, Panreac, ITW Companies, Castellar del Vallès, Barcelona, Spain)

The polyphenolic standards (all of them of analytical quality) used in this work were obtained from the following sources:

- Homogentisic acid (>99%, CAS 451-13-8, Fluka)
- Vanillic acid (>97%, CAS 121-34-6, Fluka)
- Gallic acid (CAS 149-91-7, Sigma-Aldrich)
- Protocatechuic acid (>97%, CAS 99-50-3, Fluka)
- Protocatechualdehyde (97%, CAS 139-85-5, Sigma-Aldrich)
- Caftaric acid (>97%, CAS 67879-58-7, Sigma-Aldrich)
- Gentisic acid (>98%, CAS 490-79-9, Sigma-Aldrich)
- (+)–Catechin hydrate (>98%, CAS 225937-10-0, Sigma-Aldrich)
- Caffeic acid (>99%, CAS 331-39-5, Fluka)
- Syringic acid (>97%, CAS 530-57-4, Fluka)
- Ethyl gallate (>96%, CAS 831-61-8, Sigma-Aldrich)

- (–)–Epicatechin (>90%, CAS 490-46-0, Sigma-Aldrich)
- Chlorogenic acid (>95%, CAS 327-97-9, Sigma-Aldrich)
- p-Salicylic acid (>99%, CAS 99-96-7, Sigma-Aldrich)
- p-Coumaric acid (CAS 501-98-4, Sigma-Aldrich)
- Ferulic acid (>99%, CAS 537-98-4, Fluka)
- Resveratrol (>99%, CAS 501-36-0, Sigma-Aldrich)
- Quercetin hydrate (CAS 522-12-3, Sigma-Aldrich)

The great majority of standard polyphenolic solutions were prepared in methanol with the only exception of flavonols which were dissolved in a dimethyl sulfoxide (DMSO) since this kind of compounds have a greater solubility in this medium. The polyphenolic compounds studied here were selected on the basis of an initial bibliographic research of the main polyphenols reported in the literature and some preliminary works in which several white wine and Cava samples were analysed [9-15, 19].

5.2. SAMPLES

The samples analysed in this work were different types of white and rose sparkling wines.

White sparkling wine samples were:

Blend 1, which is composed of a high percentage of Chardonnay variety (70%) and of Macabeu, Xarello, Parellada varieties (30%). Blend 2 that is only made of the three classic varieties, Macabeu, Xarello, and Parellada. Blend 3, which is also composed of the three classic varieties plus a small percentage of Chardonnay. Blend 4, whose composition is 100% Chardonnay and blend 5, whose composition is 50% Chardonnay variety and 50% Macabeu, Xarello, Parellada varieties.

Rose sparkling wines samples were:

Blend 6, whose composition is 100% Pinot noir, blend 7, which is composed of 70% Pinot noir variety and 30% Chardonnay and blend 8 made of Garnacha, Trepat and Pinot noir varieties.

Finally, there were two sample types, blend 9 and blend 10, based on Blanc de Noirs varieties. These are white coloured sparkling wines produced using red grapes varieties after removing its skin. As the difference between these classes, blend 10 has made the malolactic fermentation and the other one has not.

All the samples were previously degasified and analysed without filtration.

5.3. INSTRUMENTATION AND CHROMATOGRAPHIC CONDITIONS

The identification and quantification of the polyphenolic compounds present in the different sparkling wine samples was done through high performance liquid chromatography. The system used was an Agilent Series 1100 HPLC (Agilent Technologies, Palo Alto, California, USA) equipped with a quaternary pump (G1311A), a degasser (G1322A), an autosampler (G1329A), a diode-array detector (DAD) (G1315B) and a fluorescence detector (G1321A).

The chromatographic system was controlled through the Agilent ChemStation computer software for liquid chromatography (LC) systems. This software also has an offline option that is mainly used to study and process the chromatographic data obtained.

The chromatographic separation of the different polyphenolic compounds was carried out by a reverse-phase mode with a Kinetex C18 column (100.0 mm x 4.6 mm and particle size of 2.6 µm, Phenomenex, Torrance, CA, USA). To increase the life of the analytical column, a precolumn Security Guard C18 (4.0 mm x 3.0 mm, Phenomenex, Torrance, CA, USA), was placed to eliminate the matter in suspension and solvent pollutants. An aqueous acidified phase (solvent A), consisting of H_2O Mili-Q and 0.1% (v/v) formic acid, and an organic phase (solvent B) of methanol, were used as the components of the mobile phase. The injection volume was 5 µL, the mobile phase flow rate was 0.4 mL/min and the overall runtime was 40 minutes per injection. The elution gradient, which was previously optimized by A. Larrauri [20], is shown in Table 1.

Time (min)	A[%]	B[%]	Elution mode
0→30	97→25	3→75	Lineal
30→32	25→5	75→95	Lineal
32→34	5	95	Isocratic
34→34.20	5→97	95→3	Lineal
34.20→40	97	3	Isocratic

Table 1. Elution gradient used for the chromatographic separation of polyphenols in Cava wines.

Direct UV absorption detection was carried out in the range 190 to 720 nm. Additionally, chromatograms were acquired at 5 different wavelengths (250, 280, 310, 370 and 520 nm) for UV-visible data analysis. Later, the obtained data was analysed using the Excel computer

program (Microsoft, Redmond WA, USA) and the mathematical computing software MATLAB (MathWorks, Natick, Massachusetts, USA).

5.4. IDENTIFICATION AND QUANTIFICATION OF POLYPHENOLS

Stock standard solutions of 1000, 250, 200 and 100 mg/L of the different polyphenols, were employed to prepare the working standard solutions of 20 mg/L, used to characterize the diverse sparkling wines samples and identify the polyphenolic compounds present. Retention time and detection features are summarized in Table 2.

Moreover, a quality control (QC) solution was prepared from a mixture of 100 µL of different samples, in order to be representative of the total of the sparkling wines. This QC was regularly injected along the series of samples (each 10 sample injection).

Compound	**Retention Time** [min]	**Optimum wavelength** [nm]
Gallic acid	7.8	280
Homogentisic acid	9.3	280
Protocatechuic acid	11.6	280
Caftaric acid	13.5	310
Protocatechualdehyde	13.3	280
Gentisic acid	14.4	310
(+)-Catechin	14.9	280
Caffeic acid	17.3	310
p-Coumaric acid	20.4	310
Vanillic acid	16.8	280
Chlorogenic acid	16.5	310
Syringic acid	17.9	280
(-)-Epicatechin	18.0	280
Ethyl gallate	18.6	280
Ferulic acid	21.1	310
p-Salicylic acid	22.3	280
Quercetin	24.5	370
Resveratrol	24.1	310

Table 2. Polyphenols studied in the present work.

In order to identify the different polyphenolic compounds, present in the analysed Cavas, the samples chromatograms were compared with those of the standard polyphenolic solutions, aiming at finding the coincident peaks between samples and standards. For the polyphenolic quantification, the areas of the polyphenols identified were used.

5.5. CHEMOMETRIC MODELS

Several chemometric models were built based on the obtained data that was arranged in the format of table or matrix as follows:

-Matrix A: contains the chromatographic data, retention times and absorbance, of 121 samples and a QC (injected 12 times) at 280 nm in the time window between 9.30 and 17.96 minutes. Thus, 172900 values constitute the whole matrix.

-Matrix B: contains the chromatographic data, retention times and absorbance, of 79 samples and a QC (injected 9 times) at 310 nm in the time window between 9.30 and 27.30 minutes. Hence, the whole matrix is constituted by a total of 237600 values.

Matrix B is subdivided in two smaller matrices for calibration (matrix C) and validation (matrix D) which will be used to build PLS-DA prediction models as follows:

-Matrix C: contains the chromatographic data, retention times and absorbance, of 54 of the 79 samples and the QC at 310 nm in the time window between 9.30 and 27.30 minutes, this matrix includes a total of 170100 values. This data matrix is the calibration matrix and will be used to build the PLS-DA model.

-Matrix D: is constituted of the remaining samples and variables of matrix B that are not contained in matrix C. Therefore, it is formed of the chromatographic data, retention times and absorbance, of 25 samples. As it is also limited to the same time lapse, the one comprised between 9.30 and 27.30 minutes, this matrix includes a total of 67500 values. It will be used as a validation/prediction matrix.

6. RESULTS AND DISCUSSION

6.1. CHROMATOGRAPHIC PROFILES

First of all, the chromatographic profiles of the samples must be carefully studied in order to determine at which wavelength the information provided is more complete and useful. As an example, the chromatograms at 4 different wavelengths (250, 280, 310 and 370 nm) of a 100% Chardonnay sample are shown in Figure 5.

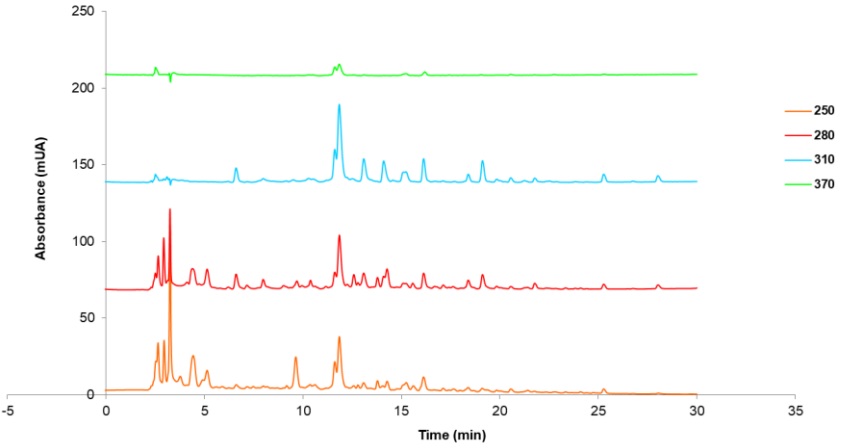


Figure 5. Chromatographic profiles of a pure Chardonnay sample at different wavelengths. Chromatograms were displaced on purpose aiming to ease comparison among the different wavelengths.

For further chemometric analysis, the initial and final parts of each chromatogram, shown in Figure 5, were cut, as these parts correspond to the phase mobile front and the column cleaning so do not provide substantial information.

As it can be seen in Figure 5, the most suitable wavelengths for the study of the polyphenols in the samples are 280 and 310 nm since, at these wavelengths, the chromatographic profiles are clearer, so it is easier to identify the different peaks. Furthermore, even though at 310 nm the

number of peaks is lesser, the ones shown are quite significant because they correspond to some of the most characteristic polyphenolic compounds found in white wines and Cavas.

In the second place, a visual qualitative analysis of the pure blend samples could be done when their chromatographic profiles at 310 nm are overlapped as shown in Figure 6.

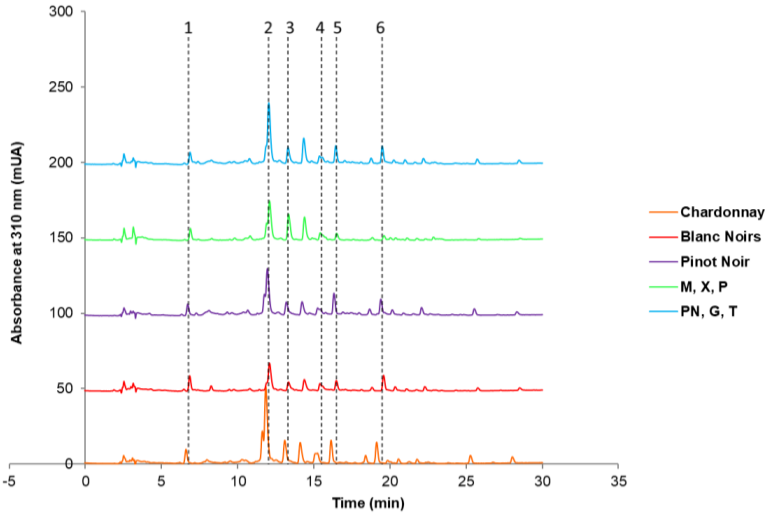


Figure 6. Chromatographic profiles of pure samples at 310 nm. Peak identification: (1) Gallic acid, (2) Caftaric acid, (3) Gentisic acid, (4) Vanillic acid, (5) Caffeic acid, (6) p-Coumaric acid.

Peak identification was done comparing the chromatographic profiles of the samples with the standards of the different polyphenolic compounds. The most characteristic peaks present in Figure 6 could thus be identified, however, there are some main unknown signals still pending of assignation.

The pure Chardonnay sample, blend 4, was the one with the highest concentration values of most of the studied polyphenols. Sample 8 composed of Pinot Noir, Garnacha and Trepat and the pure Pinot Noir sample were also rich in polyphenols.

On the other hand, the classic varietals sample, Macabeu, Xarello and Parellada, blend 2, shows lower polyphenolic concentration levels. However, this sample is abundant in gentisic acid and caftaric acid. Finally, the Blanc de Noirs sample, which has not done the malolactic fermentation, is the one with the lower concentration of polyphenolic compounds.

6.2. PCA MODELS OF THE CHROMATOGRAMS

6.2.1. Chromatograms at 280 nm

The chromatograms of 121 samples, 10 samples of the blend 1, 10 samples of the blend 2, 13 of the blend 3, 14 of the blend 4, 20 of the blend 9, 12 of the blend 6, 12 of the blend 5, 9 of the blend 7, 9 of the blend 8 and 12 of the blend 10, were recorded at 280 nm. Besides, a QC (see experimental section) was evenly injected within the random sample sequence in order to monitor the method performance and the time drift.

Chromatograms were used to build PCA models, focused on the characterization of the samples.

The data matrix studied, matrix A, contains the chromatographic data of all the samples and QCs taken in the time window between 9.30-17.96 min. This time range was selected as it shows a greater discrimination performance between samples. Additionally, an autoscaling pre-treatment was applied to the data. PC1 explains 44% of the samples variability, PC2 explains the 16% and PC3 explains the 13%.

The scores plot of the PCA model, obtained when plotting principal component (PC) 1 vs principal component (PC) 3 is shown in Figure 7.

The scores plot displays information about the samples in the PCA model, and it shows the projected locations of the samples in the PCs space. In this plot, the samples appear grouped in a certain way depending on their compositional characteristics. Those samples that are closer to each other, supposedly, have similar features, i.e. have a similar polyphenolic composition and chromatographic profile.

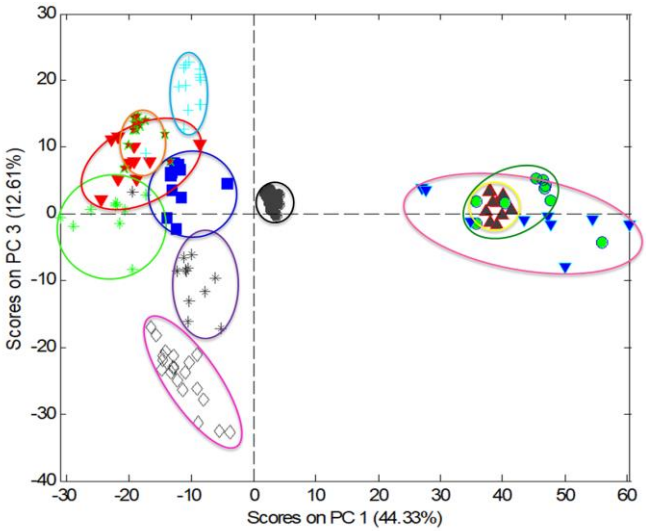


Figure 7. Scores plot of PC1 vs PC3,built from the chromatograms at 280 nm, previously pretreated. Samples code: ● QC, ▲ Blend 8 (Garnacha, Trepat, Pinot Noir), ● Blend 7 (30% Chardonnay, 70% Pinot Noir), ▲ Blend 6 (Pinot Noir), + Blend 10 (Blanc de Noirs), ◇ Blend 9 (Blanc de Noirs), + Blend 2 (Macabeu, Xarello, Parellada), + Blend 4 (Chardonnay), ★ Blend 5 (50% Macabeu, Xarello, Parellada, 50% Chardonnay), ▲ Blend 1 (70% Chardonnay, 30% Macabeu, Xarello, Parellada) and ■ Blend 3 (Macabeu, Xarello, Parellada and a small proportion of Chardonnay).

In the scores plot (Figure 7), 11 different groups of samples are observed. In the first place, the QCs form a compact group in the central area of the scores plot, therefore, it is an evidence of the excelling reproducibility of the PCA model.

Furthermore, an evident tendency is noticeable, since, on the right side of the plot, the four sets of samples correspond to Rosé Cavas. The white Cava samples are located at the top left of the plot. Finally, the Blanc de Noirs samples are situated at the left bottom of the plot in two distinguished groups. Therefore, the three main types of Cavas could be clearly differentiated from each other in this model.

On the other hand, although some groups seem to be overlapped, the pure blends are quite separated of each other. Blend 4, corresponding to the pure Chardonnay Cavas (encircled in light blue), is at the upper part of the white Cavas zone while blend 2 (the three classic varieties, encircled in green) is at the other extreme of this zone. The mixture blends of those two varieties, 1 (encircled in red), 3 (encircled in dark blue) and 5 (encircled in orange) can be found in the

middle. This observed trend agrees with the expected results as the samples, whose composition is similar, are found together or in between the two pure classes that compound them.

On the opposite side of the plot, the pure Pinot Noir samples 6, are a small compacted group (encircled in yellow) whereas, the other two Rose blends, 8 (Pinot Noir, Garnacha and Trepat) (encircled in pink) and 7 (70% Pinot Noir and 30% Chardonnay) (encircled in dark green) are surrounding the Pinot Noir group. Hence, theses sample groups were not completely separated, and a considerable sample dispersion can be observed.

Regarding to Blanc the Noirs samples, both are quite defined sets, isolated from the others. However, blend 9 Cavas (encircled in fuchsia), the ones that have not done the malolactic fermentation, are even more apart from the majority. This fact agrees with the conclusions extracted from analysing of the chromatographic profiles as the 9 samples were the ones that showed lower levels of polyphenolic compounds.

6.2.2. Chromatograms at 310 nm

The chromatograms of 79 samples, 6 samples of the blend 1, 11 samples of the blend 2, 5 of the blend 3, 10 of the blend 4, 10 of the blend 9, 11 of the blend 6, 5 of the blend 5, 5 of the blend 7, 6 of the blend 8 and 10 of the blend 10, recorded at 310 nm were studied in a similar way as given in 6.2.1.

The corresponding data matrix, matrix B, contains the chromatographic data of all the samples and QCs taken in the time window between 9.30-27.30 minutes. An autoscaling pre-treatment was applied to the data prior to PCA.

PC1 explains 53% of the samples variability, PC2 explains the 17% and PC3 explains the 7%. The scores plot of the PCA model, obtained when plotting PC1 vs PC3, is shown in Figure 8.

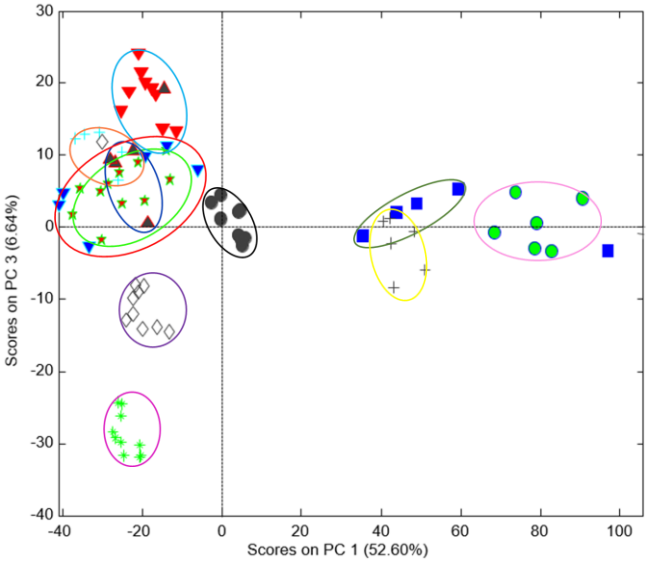


Figure 8. Scores plot, PC1 vs PC3, of the different samples, built from the chromatograms at 310 nm, previously pretreated. Samples code: ● QC, ● Blend 8 (Garnacha, Trepat, Pinot Noir), ■ Blend 7 (30% Chardonnay, 70% Pinot Noir), + Blend 6 (Pinot Noir), ◇ Blend 10 (Blanc de Noirs), + Blend 9 (Blanc de Noirs), ★ Blend 2 (Macabeu, Xarello, Parellada), ▲ Blend 4 (Chardonnay), + Blend 5 (50% Macabeu, Xarello, Parellada, 50% Chardonnay), ▲ Blend 1 (70% Chardonnay, 30% Macabeu, Xarello, Parellada) and ▲ Blend 3 (Macabeu, Xarello, Parellada and a small proportion of Chardonnay).

As in the previous case (Figure 7), three main sets of samples can be easily distinguished, corresponding to the principal classes of Cava, namely: White, Rose and Blanc de Noirs.

The QC samples were also grouped in a small set in the central area of the scores plot (Figure 8), thus confirming the excellent reproducibility of the PCA model.

Nevertheless, the white blends samples were less discriminated among each other, since the 5 groups are significantly close in space and, some of them, are even overlapped or included inside another group zone. Even though, the two pure blends samples, blend 2, composed of Macabeu, Xarello, Parellada varieties (encircled in light green) and blend 4, a pure Chardonnay Cava (encircled in red) were more distant in the PCA plot.

Rose and the Blanc de Noirs blends were better differentiated in the present plot (Figure 8) than in the previous one (Figure 7).

Blends 9 (encircled in fuchsia) and 10 (encircled in purple) appeared compacted in two independent sets far away from the others. Here, an anomalous sample of the 10 class (10-16) was detected as it was located in the white blends zone. It could be a wrongly labeled sample, therefore, it would not fit properly in the PCA model.

Finally, in this model, the three kinds of Rose Cavas are contained in three separated groups. The 8 samples, made of Garnacha, Trepat, Pinot Noir (encircled in pink) were visibly isolated from the other two. The 6, 100% Pinot Noir (encircled in yellow) and the 7, 30% Chardonnay and 70% Pinot Noir (encircled in dark green) were slightly overlapped, although this fact is in agreement with their composition since blend 7 has a considerable proportion of Pinot Noir.

6.3. PCA MODELS OF THE AREAS

As an example, a chromatogram of a pure Chardonnay sample (blend 4) registered at 310 nm is shown in Figure 9 in which the most relevant peaks are marked and labelled.

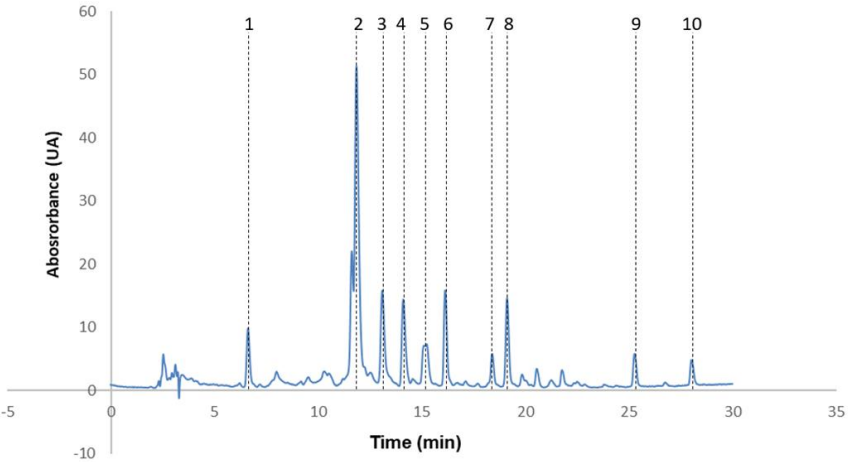


Figure 9. Chromatographic profile of a pure Chardonnay Cava sample (4), at 310 nm. Peak identification: (1) Gallic acid, (2) Caftaric acid, (3) Gentisic acid, (4) Unknown 1, (5) Vanillic acid and Chlorogenic acid, (6) Caffeic acid, (7) Unknown 2, (8) p-Coumaric acid, (9) Unknown 3, (10) Unknown 4.

There are some relevant peaks that could not be identified so are labeled as unknown 1, 2, 3 and 4. According to the study with the standards, chlorogenic and vanillic peaks are overlapped in a single peak so their determination was mutually interfered. However, since chlorogenic acid

was a residual compound in this kind of wines, it was assumed that peak 5 in Figure 9 mainly corresponded to vanillic acid.

In Table 3, the integrated areas of those peaks for each type of sample are gathered:

Compound	**4** [AU]	**6** [AU]	**10** [AU]	**9** [AU]	**2** [AU]	**8** [AU]	**1** [AU]	**3** [AU]	**5** [AU]	**7** [AU]
Gallic acid	100.2	84.9	109.3	77	82.7	103.4	91	87.9	100.7	86.1
Caftaric acid	728.1	447.4	318	110.3	381.4	671.2	626.2	612.3	631.9	645.5
Gentisic acid	232.7	147.5	105.4	74.1	220.8	180.5	235.4	229.4	202	203.2
Unknown 1	202.1	123.4	121.6	86.1	189.6	240.6	192.7	236.1	181	204.1
Vanillic acid	156.3	99	118.8	57.3	101.7	118.6	114.7	116.4	124.4	107.6
Caffeic acid	168.4	169.6	82.3	43.5	56.3	149.4	99.6	96.4	121.8	146.7
Unknown 2	65.9	49	31.3	15.4	25.1	56.8	49.4	36.2	43.2	52.9
p-Coumaric acid	160.6	140.9	123.2	51.5	33.9	156.8	97.3	89.9	104.8	133.7
Unknown 3	60.5	48.8	22.3	9.3	11.2	35.1	24.9	25	34	33.2
Unknown 4	49.6	28	21	0	9.2	27.3	23.8	26.9	34.4	30.5

Table 3. The integrated peak areas for each type of sample and each compound at the wavelength of 310 nm. Some main compounds could not be identified, so were labelled as unknown 1, 2, 3 and 4.

The data previously summarized in Table 3 was used to build PCA models to try to discriminate between samples and identify the most characteristic polyphenolic compounds of each class. To build the model, 4 PCs were selected and an autoscale pretreatment was applied to the raw data.

PC1 explains 68% of the samples variability, PC2 explains the 17% and PC3 explains the 9%. The scores and loadings plots of the PCA model, obtained when plotting PC1 vs PC3 are shown in Figure 10.

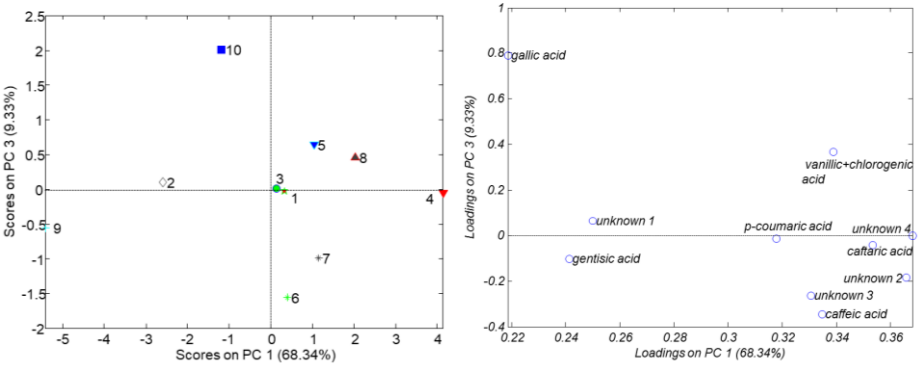


Figure 10. Scores plot, PC1 vs PC3, of the different samples blends (left) and loadings plot, PC1 vs PC3, of the studied variables (right), built from the data of the integrated peak areas of each type of sample at 310 nm.

The scores plot displays information about the samples in the PCA model, while the loadings plot displays information about the variables in the PCA space. In this plot, the variables correlation can be easily studied since close variables are expected to be highly correlated. In contrast, those located far away from each other have no correlation at all.

Analyzing the scores and loadings plots shown in Figure 10 some conclusions could be extracted. First and foremost, almost all the types of samples were visibly differentiated from each other since they were positioned in separated zones of the plot. Moreover, the samples whose composition is a mixture of two or more pure varieties, such as 1, 3, 5 and 7, were located on the space between the corresponding pure samples. Sample 7, which is compounded of 70% Pinot Noir and 30% Chardonnay, was found between sample 6, which is 100% Pinot Noir, and sample 4, a pure Chardonnay Cava. Samples 1, 3 and 5, which are a mixture of the classic varieties Macabeu, Xarello and Parellada, with Chardonnay at different percentages were situated in the middle of samples 4 and 2, this last one composed of the three classic varieties. Although in this model, samples 1 and 3 could not be distinguished easily, when plotting PC1 vs PC2 or PC2 vs PC3 (Appendix 1) their differences became more visible and their locations were in agreement with the Chardonnay percentages in the blends.

In the loadings plot, the variables situated nearby are positively correlated. Hence, caftaric acid, p-coumaric acid, unknown compound 2 and unknown compound 4 are closely correlated. Additionally, caffeic acid and unknown compound 3 also seem to be significantly correlated. On

the opposite side of the plot, gentisic acid and unknown 1 can be found pretty close to each other. This fact is an evidence of the correlation between those two variables. Lastly, gallic acid and vanillic acid does not seem to be correlated with any other variable.

When the scores plot is compared to the loadings plot is possible to determine the variables that define or are more characteristic of each Cava sample. Then, sample 4 is characterized by having high levels of caftaric acid, p-coumaric acid, vanillic acid and caffeic acid, whereas sample 2 has more predominance of gentisic acid and an unknown compound labelled as unknown 1.

On the other hand, the Rose pure Cava samples also have quite a different polyphenolic composition if compared. Sample 8 has higher levels of p-coumaric, caftaric acid and vanillic acid, while sample 6 is more characterized by having greater values of caffeic acid.

Finally, Blanc de Noirs samples are found fairly separated from the others, which is an undeniable evidence of the differences in their composition. However, whereas sample 9, the one that has not done the malolactic fermentation, seems to have greater levels of gentisic acid and unknown compound number 1, sample 10 which has done the malolactic fermentation, is rich in gallic acid.

6.4. PLS-DA MODELS

PLS-DA was performed in order to sharpen the separation between samples of different classes and to understand which variables could be considered as class markers.

PLS-DA consists of a classical Partial least squares (PLS) regression where the dependent variable expresses the class membership of the different samples. PLS components are built by trying to find a proper compromise between two purposes: describing the set of explanatory variables and predicting the response ones [21].

Two PLS-DA models were built in order to sort out the samples following two different criteria. Model 1 was focused on discrimination among sample classes corresponding to pure Chardonnay blends or not. In Model 2 classes were Blanc de Noirs blends or not. A new set of samples was used to make predictions of the corresponding classes.

The data matrix B was divided in two smaller matrices, matrix C and matrix D to be used to establish the models and to make predictions, respectively.

Matrix C contained 54 samples and the QC (injected 9 times), 7 of blend 4, 7 of blend 9, 7 of blend 10, 8 of blend 2, 8 of blend 6, 4 of blend 8, 4 of blend 1, 3 of blend 3, 3 of blend 5 and 3 of blend 7. This data matrix was the calibration matrix and was used to build the prediction model.

Matrix D constituted of 25 samples which are the remaining ones from the original matrix B; 3 samples of blend 4, 3 of blend 9, 3 of blend 10, 3 of blend 2, 3 of blend 6, 2 of blend 8, 2 of blend 1, 2 of blend 3, 2 of blend 5 and 2 of blend 7. This data was used to make predictions. The PLS-DA assigned each "unknown" sample to a class (Chardonnay/Non-Chardonnay in Model 1 and Blanc de Noirs/Non-Blanc de Noirs in Model 2). Then, the real assignation was compared with the estimated by the model and its predictive ability was evaluated.

6.4.1. Discrimination between Chardonnay and Non-Chardonnay Cavas

In Figure 11 the class assignation of the PLS-DA Model 1, built from the data gathered in matrix C and the prediction matrix D is shown. 6 latent variables (LV) were selected by a cross validation method (Appendix 2) to build this model.

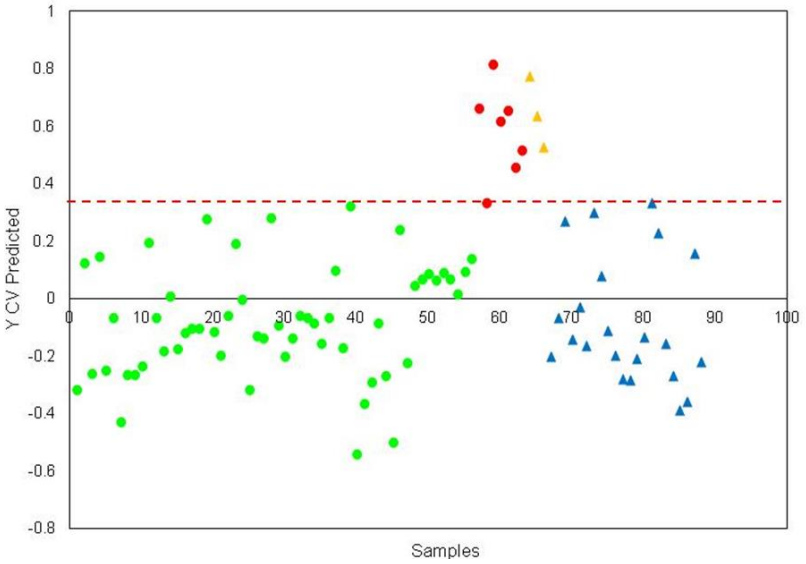


Figure 11. Assignation plot (Samples vs class predicted) of Model 1 for the discrimination between Chardonnay and Non-Chardonnay Cavas. Dashed line indicated the class threshold with upper samples belonging to Chardonnay and lower samples belonging to Non-Chardonnay classes. Samples code: ● Non-Chardonnay calibration, ● Chardonnay calibration, ▲ Chardonnay validation, ▲ Non-Chardonnay validation.

Results in Figure 11 demonstrated the ability of the Model 1 to discern among samples and assign new samples to their corresponding class. Even though, there were a few samples placed at the limit zone between both groups, all the samples were correctly classified. This fact is a strong evidence of the quality of the model.

6.4.2. Discrimination between Blanc de Noirs and Non-Blanc de Noirs Cavas

In Figure 12 the class assignation plot of the PLS-DA Model 2, built from the calibration matrix C and the prediction matrix D is shown. 4 latent variables (LV) were selected by cross validation (Appendix 2) to build this model.

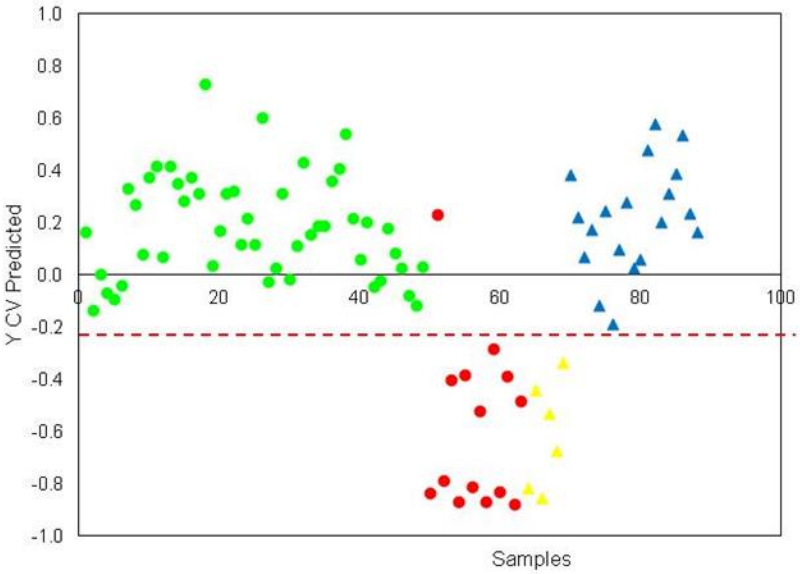


Figure 12. Assignation plot (Samples vs class predicted) of Model 2 for the discrimination between Blanc de Noirs and Non-Blanc de Noirs Cavas. Dashed line indicated the class threshold with upper samples belonging to Non-Blanc de Noir and lower samples belonging to Blanc de Noir classes. Samples code: ● Non-Blanc de Noirs calibration, ● Blanc de Noirs calibration, ▲ Blanc de Noirs validation, ▲ Non-Blanc de Noirs validation.

Results in Figure 12 showed, that most of the samples were assigned correctly to their corresponding group. It can be noted that one sample from Blanc de Noirs calibration group was wrongly classified into the set of Non-Blanc de noirs Cavas. This anomalous sample (10-16) was the same one that previously had been erroneously grouped with the white Cava blend in the

PCA model (Figure 8). As a possible explanation to this situation, this sample could be incorrectly labelled as a Blanc de Noirs Cava when, in fact, should be a white blend Cava.

As a result, it can be concluded that the Model 2 presents an excellent predictive ability.

7. CONCLUSIONS

In this work, the polyphenolic profile of various Cava samples, obtained by HPLC with UV-visible detection, have been thoroughly studied for characterization and classification purposes. From this evaluation several conclusions can be drawn:

- The most suitable wavelengths for the study of the polyphenolic compounds in the sparkling wines samples are 280 and 310 nm.
- The principal polyphenols found in the samples are: gallic acid, caftaric acid, gentisic acid, vanillic acid, caffeic acid and p-coumaric acid. The pure Chardonnay Cava presents the highest concentration of most of the studied polyphenols. It is characterized by having high levels of caftaric acid, p-coumaric acid and vanillic acid. Cavas made of the three classic varieties show greater values of gentisic acid. In the Rose sparkling wines composed of Pinot Noir, caffeic acid has been recognized as a characteristic species. Blanc de Noir samples present lower values of the studied polyphenols, however, those subjected to malolactic fermentation are rich in gallic acid, being the Cava class that has the highest values of this compound.
- The chromatographic data must be pre-treated before being used to build the models if a superior discrimination among samples is desired.
- A successful differentiation among Cava samples of diverse classes based on the polyphenolic composition was achieved.
- PCA models have been built from the chromatographic profiles (fingerprinting approach) and from the integrated areas of the principal peaks of the chromatogram (profiling approach). Both provided an excellent differentiation among Cava samples according to their polyphenolic composition.
- PLS-DA model was successful in classifying the sparkling wines depending on their varieties. In a first study, pure Chardonnay and non-pure Chardonnay classes were defined. The predictive ability was certainly proved since samples of a new external set were correctly assigned to their corresponding classes.

- The second PLS-DA model successfully classified the Cava samples according to their polyphenolic composition in two classes Blanc de Noirs or not. This model also showed an exceling predictive ability which was proved by the accurate classification of the new data set of samples.

8. REFERENCES AND NOTES

1. Institut del cava. Identidad. http://www.institutdelcava.com/identidad/ (accessed May 1, 2018).
2. Denominación de origen cava. El Cava. Variedades y Elaboración http://www.docava.es (accessed May 1,2018).
3. Celler Vell. Elaboration of Cava http://www.cellervell.com/en/contenido/21-elaboracio-del-cava (accessed May 1,2018).
4. Quiñones, M.; Miguel, M.; Aleixandre, A. Los polifenoles, compuestos de origen natural con efectos saludables sobre el sistema cardiovascular. *Nutr. Hosp*. 2012, 27(1), 76-89.
5. Tsao, R. Chemistry and Biochemistry of Dietary Polyphenols. *Nutrients*. 2010, 2(12), 1231-1246.
6. Gutiérrez, A. Vino, polifenoles y protección a la salud. *Revista Cubana Aliment Nutr*. 2002, 16(2), 134-141.
7. Los compuestos fenólicos del vino. https://www.sabrosia.com/2012/10/los-compuestos-fenolicos-del-vino/ (accessed May 3, 2018).
8. Huang, W.;Cai, Y. ;Zhang, Y. Natural Phenolic Compounds From Medicinal Herbs and Dietary Plants: Potential Use for Cancer Prevention. *Nutr. Cancer*. 2009, 62(1), 1-20.
9. Paixão, N.; Pereira, V.; Marques, J.; Câmara, J. Quantification of polyphenols with potential antioxidant properties in wines using reverse phase HPLC. *J. Sep. Sci*. 2008, 31(12), 2189-2198.
10. Pereira, V.; Câmara, J.; Cacho, J.; Marques, J. HPLC-DAD methodology for the quantification of organic acids, furans and polyphenols by direct injection of wine samples. *J. Sep. Sci*. 2010, 33(9), 1204-1215.
11. Serrano, D.; Saurina, J.; Hernández, S.; Checa, A. Classification and characterisation of Spanish red wines according to their appellation of origin based on chromatographic profiles and chemometric data analysis. *Food Chem*. 2012, 135(3), 1425-1431.
12. Stefenon, C.; Bonesi, C.; Marzarotto, V.; Barnabé, D.; Spinelli, F.; Webber, V.; Vanderlinde, R. Phenolic composition and antioxidant activity in sparkling wines: Modulation by the ageing on lees. *Food Chem*. 2014, 145, 292-299.
13. Bosch, J.; Sartini, E.; Flores, C.; Caixach, J.; López, E.; Buxaderas, S. Viability of total phenol index value as quality marker of sparkling wines, “cavas”. *Food Chem*. 2009, 114(3), 782-790.
14. Weingerl, V.; Strlic, M.; Kocar, D. Comparison of Methods for Determination of Polyphenols in Wine by HPLC-UV/VIS, LC/MS/MS and Spectrophotometry. *Acta Chim. Slov*. 2009, 56(3), 698-703.
15. Jaitz, L.; Siegl, K.; Eder, R.; Rak, G.; Abranko, L.; Koellensperger, G.; Hann, S. LC-MS/MS analysis of phenols for classification of red wine according to geographic origin, grape variety and vintage. *Food Chem*. 2010, 122(1), 366-372.
16. Peres, R.; Micke, G.; Tavares, M.; Rodriguez, D. Multivariant optimization, validation, and application of capillary electrophoresis for simultaneous determination of polyphenols and phenolic acids in Brazilian wines. *J. Sep. Sci*. 2009, 32(21), 3822-3828.
17. Bevin, C.; Dambergs, R.; Fergusson, A.; Cozzolino, D. Varietal discrimination of Australian wines by means of mid-infrared spectroscopy and multivariate analysis. *Anal. Chim. Acta*. 2008, 621(1), 19-23.
18. Mitic, M.; Kostic, D.; Paunovic, D.; Stojanovic, B.; Pavlovic, J. A New Kinetic Spectrophotometric Method for Total Polyphenols Determination in White Wines. *S. Afr. J. Chem*. 2014, 67, 104-108.
19. Chamkha, M.; Cathala, B.; Cheynier, V.; Douillard, R. Phenolic Composition of Champagnes from Chardonnay and Pinot Noir Vintages. *J. Agric. Food Chem*. 2003, 51(10), 3179-3184.

20. Larrauri, A. Determination of polyphenols in wines by liquid chromatography: Application to the study of wine degradation. Treball de final de grau. Universitat de Barcelona. Gener 2016.
21. Camo Software: Leading Multivariate Data Analysis and Design of Experiments. Resources. PLS-DA. http://www.camo.com/resources/pls-da.html (accessed May 18, 2018).

9. ACRONYMS

CE Capillary Electrophoresis

DAD Diode Array Detection

DMSO Dimethyl sulfoxide

HPLC High Performance Liquid Chromatography

LV Latent Variable

MS Mass Spectrometry

PC Principal Component

PCA Principal Component Analysis

PDO Protected Designation of Origin

PLS Partial Least Squares

PLS-DA Partial Least Squares Discriminant Analysis

QC Quality Control

RMSECV Root Mean Square Error of Cross Validation

RP Reversed Phase

UHPLC Ultra High Performance Liquid Chromatography

UV Ultraviolet

APPENDICES

APPENDIX 1: PCA MODELS OF THE AREAS

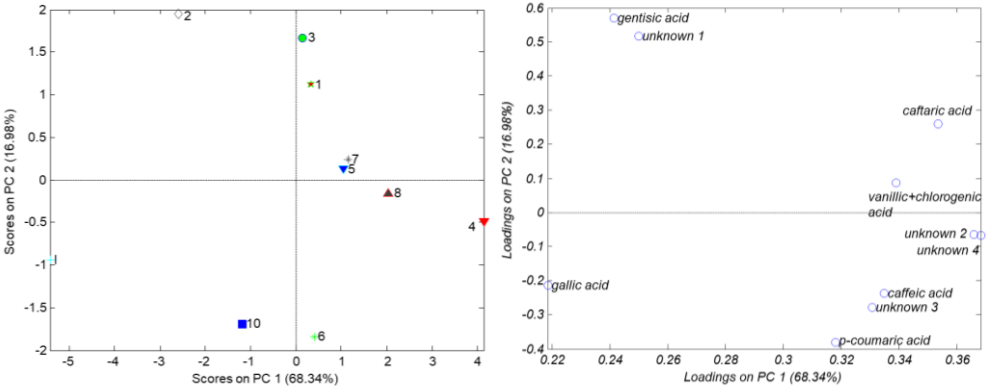


Figure 13. Scores plot, PC1 vs PC2, of the different samples blends (left) and loadings plot, PC1 vs PC2, of the studied variables (right), built from the data of the integrated peak areas of each type of sample at 310 nm.

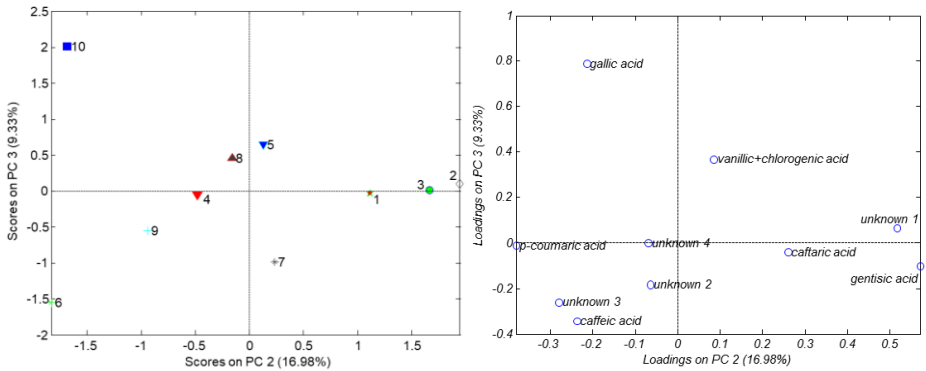


Figure 14. Scores plot, PC2 vs PC3, of the different samples blends (left) and loadings plot, PC2 vs PC3, of the studied variables (right), built from the data of the integrated peak areas of each type of sample at 310 nm.

APPENDIX 2: CROSS VALIDATION OF THE PLS-DA MODELS

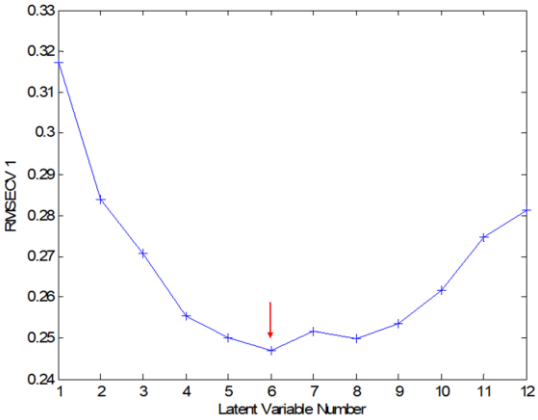


Figure 15. Cross validation plot of the Model 1 (Chardonnay/Non-Chardonnay), LV Number vs RMSECV (Root Mean Square Error of Cross Validation), used to determine the selected LVs. The red arrow points at the minimum which is also the number of LVs selected to build the model.

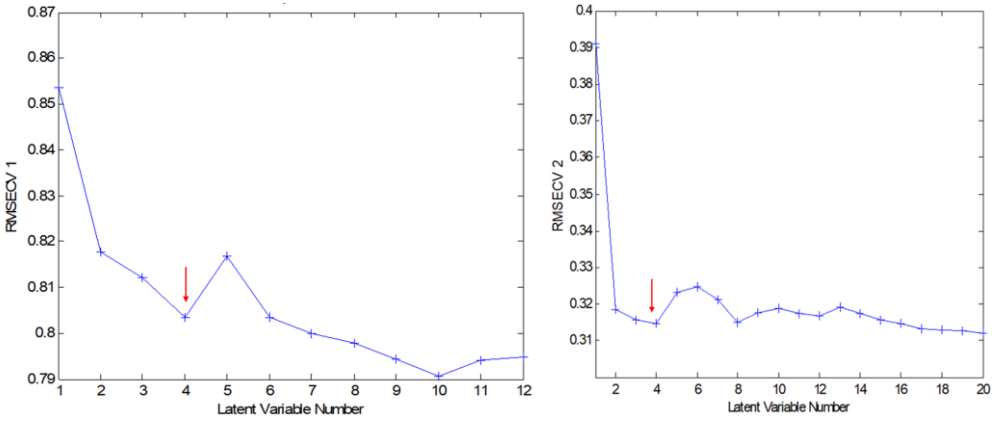


Figure 16. Cross validation plots of the Model 2 (Blanc de Noirs/Non-Blanc de Noirs), LV Number vs RMSECV, used to determine the selected LVs. The red arrow points at the minimum which is also the number of LVs selected to build the model.