Evaluation of Geographic Origin of Torrontés Wines by Chemometrics

Silvana M. Azcarate¹,², Miguel A. Cantarelli¹,², Eduardo J. Marchevsky³ & José M. Camiña¹,²

¹ Facultad de Ciencias Exactas y Naturales, Universidad Nacional de La Pampa, Av. Uruguay 151 (6300) Santa Rosa, La Pampa, Argentina
² Instituto de las Ciencias de la Tierra y Ambientales de La Pampa (INCITAP), Av. Uruguay 151 (6300) Santa Rosa, La Pampa, Argentina
³ Instituto de Química de San Luis (INQUISAL), Chacabuco y Pedernera (5700), San Luis, Argentina

Correspondence: José M. Camiña, Instituto de las Ciencias de la Tierra y Ambientales de La Pampa (INCITAP), Facultad de Ciencias Exactas y Naturales, Universidad Nacional de La Pampa, Santa Rosa, La Pampa, Argentina. Tel: 54-2954-425-166. Fax: 54-2954-432-535. E-mail: jcaminia@yahoo.com; jcaminia@gmail.com

Received: June 26, 2013   Accepted: August 8, 2013   Online Published: doi:10.5539/jfr.v2n5p1

Abstract
This work discusses the determination of the provenance of commercial Torrontés wines from different Argentinean provinces (Mendoza, San Juan, Salta and Rio Negro) by the use of UV-vis spectroscopy and chemometric techniques. In order to find classification models, wines (n = 80) were analyzed using UV-Vis region of the electromagnetic spectrum. Principal component analysis (PCA), linear discriminant analysis (LDA) and partial least squares discriminant analysis (PLS-DA) were used to classify Torrontés wines according to their geographical origin. Classification rates obtained were highly satisfactory. The PLS-DA and LDA calibration models showed that 100% of the Mendoza, San Juan, Salta and Rio Negro Torrontés wine samples had been correctly classified. These results demonstrate the potential use of UV spectroscopy with chemometric data analysis as a method to classify Torrontés wines according to their geographical origin, a procedure which requires low-cost equipment and short-time analysis in comparison with other techniques.

Keywords: classification, white wine, UV-vis spectroscopy, PLS-DA

1. Introduction
The composition of wine is influenced by many factors related to the specific production area: grape varieties, soil, weather, crop and wine making practices (Rebolo et al., 2000). Argentina is a wine-producing country in which the production of quality wines has a special economic importance. According to the annual statistical report on world viticulture issued by the International Organization of Vine and Wine (OIV) for 2011 (OIV, 2012), Argentina is the 9th largest wine exporter in the world and 5th largest wine producer after Italy, France, Spain and the United States. The determination of wine authenticity is an important challenge to avoid possible frauds and to guarantee their origin and quality (Pérez Trujillo, Conde, Pérez Pont, Câmara, & Marques, 2011).

A wide number of characterization studies have been carried out applying multivariate chemometric techniques to solve problems based on diverse composition data of wines (Kozak & Scaman, 2008; Green, Parr, Breitmeyer, Valentín, & Sherlock, 2011). The same comprehensive and critical reviews have discussed wine authenticity, analyzing the application of different quality control methods by means of multivariate analysis (Medina, 1996; Arvanitoyannis, Katsota, Psarra, Soufleros, & Kallithraka, 1999), and have assessed different strategies for characterizing wines based on compositional profiles and chemometric tools (Saurina, 2010).

Most of these research studies determined chemical data through labor-intensive and high-cost analyses, based on sophisticated instruments that require well-trained operators and are difficult to automate and implement on a regular basis (De Villiers, Alberts, Lynen, Crouch, & Sandra, 2003; Encinar, Sliwka-Kasynska, Polatajko, Vacchina, & Szpunar 2003; Flurer, 2003; Wang, Geil, Kolling, & Padua, 2003). Several spectroscopy methods, combined with multivariate analysis, have also been used as non sophisticated and inexpensive analytical
techniques to classify wine. Some of these studies have used UV-vis spectroscopy and pattern recognition methods to discriminate different origin denominations (Urbano, Luque de Castro, Pérez, García-Olmo, & Gómez-Nieto, 2006; Acevedo, Jimenez, Maldonado, & Dominguez, 2007). Other publications about varietal and geographical discrimination of wines have been based on near-infrared spectroscopy (NIR) combined with multivariate analysis (Liu, Cozzolino, Cynkar, Gishen, & Colby, 2006; Bevin, Dambergs, Fergusson, & Cozzolino, 2008). Often, UV-vis spectral data have been complemented with measurements in the NIR spectrum to classify wines according to their geographical origin (Cozzolino, Cynkar, Shah, & Smith, 2011). Spectroscopy offers a number of advantages in comparison with other analytical determinations such as simplicity, robustness, availability and minimal sample treatment. However, UV-vis and NIR spectra contain a limited number of spectral features, limiting their possibilities in discrimination (Saurina, 2010). For this reason, the application of multivariate data analysis like principal component analysis (PCA), linear discriminant analysis (LDA), and partial least square discriminant analysis (PLS-DA) allow to interpret the spectral properties of the samples and to classify them with the use of the basic chemical information provided by the spectra (Adams, 1995; Reid, O’Donnell & Downey, 2006; Granato, Katayama, Chizuko, & de Castro 2011).

In Argentina, different chemometric procedures have been applied in order to establish quality criteria, genuineness and geographical and varietal differentiation in wine (Fabani, Arrúa, Vázquez, Diaz, Baroni & Wunderlin, 2010; Bentlin, Pulati, Dressler, & Pozebon, 2011; Di Paola-Naranjo et al., 2011; Fanzone et al., 2012), but none of them have applied the spectroscopy methods mentioned above.

Torrontés is a white grape, very important in the production of wines from Argentina. Due to the high quality of winemaking, in 2011 it has been the white wine variety with higher export, which was reported by the National Wines Institute of Argentina (INV, 2011).

In this work, the UV-vis spectroscopy, combined with multivariate data analysis, have been studied with the objective of developing classification models to differentiate between Torrontés wines produced in different Argentinean provinces.

2. Methods

2.1 Wine Samples

A total of 80 (16 trademarks x 5 bottles of each one, declared as mono-varietal wines on their labels) commercially available Argentina Torrontés wine samples from geographically different wine regions were acquired from local wine shops. All samples were from 2011 vintages. All samples were immediately analyzed after being opened and within 1 month after purchase. The alcoholic content ranged from 12.4% to 14.2 % ethanol, and pH ranged from 3.0 to 3.9 (Table 1).

<table>
<thead>
<tr>
<th>Region</th>
<th>N</th>
<th>T</th>
<th>Year</th>
<th>Ethanol content (% v.v⁻¹)</th>
<th>pH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mendoza</td>
<td>7</td>
<td>35</td>
<td>2011</td>
<td>12.4-13.5</td>
<td>3.3-3.7</td>
</tr>
<tr>
<td>Salta</td>
<td>4</td>
<td>20</td>
<td>2011</td>
<td>12.5-13.5</td>
<td>3.4-3.6</td>
</tr>
<tr>
<td>San Juan</td>
<td>3</td>
<td>15</td>
<td>2011</td>
<td>13.5-14.2</td>
<td>3.3-3.9</td>
</tr>
<tr>
<td>Rio Negro</td>
<td>2</td>
<td>10</td>
<td>2011</td>
<td>134-13.5</td>
<td>3.0-3.5</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>16</strong></td>
<td><strong>80</strong></td>
<td><strong>2011</strong></td>
<td><strong>12.4-14.2</strong></td>
<td><strong>3.0-3.9</strong></td>
</tr>
</tbody>
</table>

N: number of trademarks
T: total number of samples analysed

2.2 Sample Preparation

All reagents were of analytical grade. Ultrapure water (Millipore UV Synergy System, Billerica, Mass., U.S.A.), with a resistivity of 18.2 MΩ cm, was used to prepare all solutions. In order to prevent oxidation reactions, all samples were analyzed immediately after the wine bottles were opened and the following treatment was carried out: 2 mL wine and 10 mL Clark and Lubs buffer, pH 10.2 (0.1 mol L⁻¹ KCl and H₂BO₃ -Merck, Darmstadt, Germany), were diluted to 25 mL (Meites, 1982).
2.3 Spectroscopic Analysis

UV-Vis spectroscopy measurements were performed using an Ocean Optics Model CHEMUSB4 UV-Vis spectrophotometer, with a linear Charge-Coupled Device (CCD) array detector (Ocean Optics, Duiven, the Netherlands). The absorbance spectra were collected from 5 different bottles of each label, with a working range from 200 to 500 nm. The pH measurements were performed with a Horiba F42 pH meter (Tokyo, Japan).

2.4 Multivariate Data Analysis and Classification

UV-Vis absorbances were used for a preliminary analysis by principal component analysis (PCA) using the complete dataset. Classification techniques: linear discriminant analysis (LDA) and partial least squares discriminant analysis (PLS-DA) were applied to the pre-selected spectral range from PCA with all the samples included in the training set. The classification obtained before was validated. The 80 objects were randomly divided between a training set and a prediction set. In order to carry out these analyses it was important to know a minimum number of sample characteristics, which provided information for correct classification.

Pattern recognition analysis was performed by means of statistical software packages: Unscrambler 6.11 software (CAMO-ASA, Trondheim, Norway) and Infostat software (Córdoba, Argentina).

3. Results

3.1 Optimisation of pH of Buffer

A study prior to obtaining the multivariate models and to evaluate the optimal pH value with the spectral condition, was performed. All wines were ranged between pH 3-4 (Table 1). However it was necessary to evaluate the effect that pH produces on the spectral data, due that frequently affect the absorbances in the UV-Vis region. Spectral behavior of wines was carried out at 6 different pH values: 1.0, 3.0, 5.0, 7.0, 10.2, and 11.4. The spectral samples at 6 different pH values were collected and subsequently analyzed. Figure 1 shows the results obtained for one sample.

![Figure 1. Spectral curves of one wine sample at six different pH values](image)

The curves at pH 10.2 and 11.4 showed differences between 330 and 420 nm, which were not appreciated in the curves corresponding to lower pH values; these spectral differences were useful for classification. However, the spectral curves at pH 11.4 show a tendency to increase in the range between 220-440 nm, producing a notable difference from the rest of the curves; for this reason these were not used. Therefore, the absorbance values corresponding to pH 10.2 were selected to be included in the multivariate analysis.

3.2 Principal Component Analysis (PCA)

One preliminary step to study the data structure is the search for natural grouping among the samples (Rebolo et al., 2000). Firstly, in the data analysis, a study by principal component analysis (PCA) was done. PCA reduces the dimensionality of the data to a small number of components, in order to examine the possible grouping of samples and visualize the presence of outliers (Otto 2007; Naes, Isaksson, Fearn, & Davies, 2002). For this, it is...
necessary to determine the range of wavelengths in which the greatest differences are found. In order to establish these differences, loadings were analyzed. Loading vectors can be considered as the connection between the initial variables and the PCs, which are a linear combination of all the wavelengths: loadings are the coefficients of this combination and represent the influence of wavelengths in the explanation of the data variance (Urbano et al.; 2006). Figure 2 shows plot loading for first and second principal components (PCs), where the main differences between wine samples were not observed in a specific electromagnetic spectrum range. However, the second principal component in the loading plot, showed that the key wavelengths for the discrimination of groups were in the range 300–350 nm; this range coincides the spectral curve to pH value 10 (Figure 1). This indicates that the pH value for this analysis has been correctly selected. Besides, the main spectral differences observed were established in the ultraviolet region, which might be caused by different wine components absorbing at this range. Absorbances in this wavelength region are basically related to the simplest compounds such as phenolic, benzoic and hydroxycinnamic acids (Cheynier, 2005). For this reason, meaningful relationships between spectral data and enological characteristics can be evaluated.

![Figure 2. Loading plot as wavelength function for first and second PCs](image)

The PCA model was applied to the matrix formed by the UV-Vis spectra corresponding to the white wine samples. Thus, a matrix of 80 x 1185 was built from the autoscaled data, without prior signal pretreatment. The PCA model was built using 242 variables, corresponding to the absorbance values from 300 to 350 nm. Figure 3 shows the score plot of the first two PCs using UV-Vis spectra of the Torrontes wine samples analyzed. The first three PC’s -used to build the model- explained more than 99.5% of the total variance of the spectra in the set of wines analyzed. From this exploratory analysis, it can be observed that four groups could be clearly separated, which correspond to the 4 geographical origins studied, namely: Mendoza (M), San Juan (SJ), Salta (S) and Río Negro (RN).

In accordance with similar results reported by other authors (Cozzolino, Smyth, & Gishen, 2003; Bevin, Fergusson, Perry, Janik, & Cozzolino, 2006; Liu et al., 2008) and, as the PCA plot shows, the replicate samples of the same trademarks are grouped in the same cluster, but without overlapping, which is due to bottle to bottle variation. From this exploratory analysis, the differences and the similarities between these Argentinean wines have been detected in the ultraviolet region, due to differences in the proportions and the compositions of the families of the natural wine components mentioned before, between different geographical origins within same variety.
3.3 Linear Discriminant Analysis (LDA)

LDA is a supervised method, in which classification models are constructed based on pre-established groups of the data. The percentage of correct classification is calculated as the relationship between number of samples correctly predicted and the group population: percentage correct classification = (number of correctly classified individuals/sample population) × 100 (Sperková & Suchánek, 2005). LDA was applied to an initial matrix containing the auto scaled data for 80 objects and the 242 variables, previously selected from the PCA. In this analysis, samples were divided into four categories, according to their geographical origins (M=Mendoza, SJ=San Juan, S=Salta and RN= Río Negro). Fifty-five and twenty five random samples were used for the training set and the prediction set respectively. Table 2 shows the results of the LDA model; the recognition ability for the four groups was highly satisfactory because both in the training and prediction sets, all samples analyzed according to geographical origin were 100% correctly classified.

Table 2. Results of the classification ability of the LDA model for different wines according to their geographical origin

<table>
<thead>
<tr>
<th></th>
<th>Training</th>
<th></th>
<th>Prediction</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Per cent correct</td>
<td>Mendoza</td>
<td>Río Negro</td>
<td>San Juan</td>
</tr>
<tr>
<td>Mendoza</td>
<td>100</td>
<td>25</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Río Negro</td>
<td>100</td>
<td>0</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>San Juan</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>Salta</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>25</td>
<td>7</td>
<td>10</td>
</tr>
</tbody>
</table>

3.4 Partial Least Squares Discriminant Analysis (PLS-DA).

The PLS-DA models were developed using the spectral range selected previously by applying the PCA. Figure 4 shows the score plot of the first three PCs of the PLS-DA model, which contained 98.5% of variance, explained from the original information in the calibration step, and from which it can be observed that the grouping is similar to the PCA score plot; however, separation of wines according to geographical origin seems to be more obvious. This might be explained by the fact that the PLS-DA algorithm maximizes the variance between groups rather than within the group (Kemsley, 1996; Naes et al., 2002; Otto, 2007). Table 3 shows the PLS-DA
classification rates (per cent of classification) for the validation and prediction sets, according to geographical origin. The PLS-DA model produced an overall rate of correct classification of 100%: wine samples belonging to Mendoza, San Juan, Salta and Rio Negro were 100% correctly classified.

Figure 4. PLS-DA 3D score plot of the first three PCs of Torrontés wines samples using UV spectral region for the validation and calibration set from different geographical origins: Mendoza (M), Río Negro (RN), San Juan (SJ) and Salta (S).

Table 3. Results of the classification of the PLS-DA model according to geographical origin

<table>
<thead>
<tr>
<th></th>
<th>Training</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mendoza</td>
<td>Río Negro</td>
</tr>
<tr>
<td>Mendoza</td>
<td>100</td>
<td>28</td>
</tr>
<tr>
<td>Río Negro</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>San Juan</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>Salta</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>28</td>
</tr>
</tbody>
</table>

4. Discussion

This study has shown that a simple technique such as the UV-vis spectroscopy, combined with chemometric analysis were able to discriminate between Torrontés wines produced in the different Argentinean provinces studied. The simple adjustment and selection of pH (10.2) allowed to obtain good classificatory models. The PCA analysis (Figure 1) has verified that the spectra obtained provided important information for the sample classification and were useful to use in LDA and PLS-DA models.

On the other hand, LDA and PLS-DA calibration models, have achieved high classification rates to predict wine samples, obtaining a 100% of correct classification in both cases for all studied samples (Table 2 and 3). Hence, this method is able to classify wines without the need use of more complex techniques.

5. Conclusion

The method here proposed is a powerful analytical tool able to classify Torrontés wines from different regions of Argentina. Due to its low cost and high effectiveness, it can be useful as routine control method in comparison to
other quantitative methods, which require labor-intensive, high-cost and time-consuming analysis, and the use of standards for calibration. For these reasons, the models obtained could be used effectively to evaluate geographical provenance of Torrontés wines in the national and international market. However, due to the nonselective nature of this method, more studies may be necessary to determine the geographical origin of Torrontés wines from other provenances or vintages.

References


**Copyrights**

Copyright for this article is retained by the author(s), with first publication rights granted to the journal. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).