Exploratory analysis of South America wines using artificial intelligence

Candice N. Carneiro
Universidade Federal do Recôncavo da Bahia, Cruz das Almas

Federico J. V. Gomez
National University of Cuyo

Adrian Spisso
National University of Cuyo

Maria Fernanda Silva
National University of Cuyo

Jorge L. O. Santos
Universidade Federal do Oeste da Bahia, Centro Multidisciplinar de Bom Jesus da Lapa

Fabio S. Dias (fabiosdias@yahoo.com.br)
Universidade Federal da Bahia

Research Article

Keywords: Wine, elemental composition, microwave-induced plasma optical emission spectrometry

Posted Date: October 17th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2155932/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Abstract

The microwave-induced plasma optical emission spectrometer has been used for multielement determination in South American wine samples. The analytes were determined after acid digestion of 47 samples of Brazilian and Argentine wine and the application of logistic regression, support vector machine, and decision tree for exploratory analysis and the comparison of these algorithms in differentiating red wine samples by region of origin. The Brazilian wines samples were clustering in a group, while Argentines wines samples formed another cluster. The quantification limits (mg L\(^{-1}\)) were P (0.06); B (0.08); K (0.17); Mn (0.002); Cr (0.002) and Al (0.02). The accuracy of the method was evaluated analyzing the wine samples by ICP OES for results comparison. The concentration in mg L\(^{-1}\) found for each element in wine samples, were Al (< 0.02 - 1.82), Cr (0.15 - 0.50), Mn (< 0.002 - 0.8), P (97 - 277), B (1.7 - 11.6), Pb (< 0.06 - 0.3), Na (8.84 - 41.57) and K (604 - 1701), in mg L\(^{-1}\).

Introduction

The consumption of wine is recommended and beneficial to human health due to the presence of elements such as calcium, copper, iron, potassium, selenium, manganese and zinc and phenolic compounds [1,2,3 Dias ...].

The presence of inorganic species in wine may be related to the quality of this drink. Some ionic species found in wine can cause changes in their organoleptic characteristics and stability [4]. Obtaining a good quality wine requires knowledge and monitoring of the compounds present in the beverage. The chemical composition of the wine is linked to the variety of grapes, soil type, climate, region where they are grown and the type of treatment by which the grapes were submitted during the wine production and preservation process [1,5], atmospheric contaminations, pesticides addition and fertilizers and even by the own methods and equipment of vinification [6]. The elemental composition of wines is related to geographical regions and by metal characterization; it is possible to make a relation with the origin of the wine [7,8]. Many analytical techniques can be used for the elemental determination in several samples such as flame atomic absorption spectrometry (FAAS) very efficient for monoelementar determination, but present low sensibility [9,10]. Inductively coupled plasma optical emission spectrometry (ICP OES) and inductively coupled plasma mass spectrometry (ICP-MS), their emission techniques are high selectivity and sensitivity, however expensive cost [11]. An analytical technique that has been reported in the literature with a new proposal by multielement analyses is microwave-induced plasma optical emission spectrometry (MIP OES) [12-15]. This technique is inexpensive because use a nitrogen plasma and present good sensibility for multielement determination [16,17]. The significance of geographical origin to wine is greater than in any other food product. Terroir is the natural locality of each plant of grapevine. Each terroir is influenced by several factors: climatic, geological and soil, and by human activity when treating the vineyard. Each grape produced in a specific terroir reflects the locality in its chemical composition [18].
An important tool for the characterization of wines has been the clusters analysis in order to classify the wines according to the geographic origin and the type of grape [19,20]. Nelson and collaborators (2015) determined Sr, Rb, Ca, K, Na, and Mg in samples of Malbec wine and used principal components analysis and partial least squares - discriminate analysis to distinguish between wines from different regions by comparing the relative concentrations of metals, which are characteristic of the soil composition of the region of production [21]. Another work, Ivanova-Petropulos and investigators (2013) determined 42 elements in Macedonian wines in order to study the relationship between elemental concentrations and two variables: wine type (white vs. red) and geographical indication, for this purpose, were used descriptive analysis, factor analysis and principal components [22]. Carneiro and and investigators used PCA and HCA for exploratory of Argentine and Brazilian wines using phenolic compounds and melatonin concentrations for studied geographical origins. The results confirm the difference between the two regions and the influence of environmental and climatic conditions on the phenolic compounds and melatonin concentrations in wines [23]. An alternative tool of clusters analysis has been the artificial intelligence (AI).

The artificial intelligence in analytical chemistry is an area of knowledge that is still few studied and aims to use algorithms that simulate the learning processes of the human brain. Most algorithms involved in machine learning can connect each input data to its corresponding output, but also adapting to the environment, learning and making decisions, providing more accurate and precise answers [24]. Moreira and researchers used support vector machine with linear Kernel as supervised learning methods to exploratory analysis the Salvia officinalis samples treated with different growth regulators (salicylic acid, gibberellic acid, abscisic acid and without regulators) on the mineral and agronomic parameters. The authors compared the results with principal component analysis and the results agreed. The model presented precision of 87.5% for samples classified based on mineral composition and 93.7% for classification based on agronomic variables [14]. The data classification with supervised learning algorithms (eg SVM, DT, LR, LDA, and PLS-DA) allow data labeling and, consequently, the algorithm can predict the output label from the input data [25, 26]. This advantage is very useful when the objective is to discriminate a set of data, but to make the prediction of the label from input data, as is the case with certifications of origin and authenticity of foods from the chemical composition profile.

In this study, compare the performance of some of the machine learning (ML) algorithms most used in binary classification and verify the applicability in discriminating the origin of wines (Brazilian and Argentine samples) from their mineral composition (P, K, Mn, B, Cr, Pb, Al, and Na). The classifiers used in this research were logistic regression (LR), support vector machine (SVM), and decision tree (DT).

**Experimental**

2.1. **Instrumentation**

A microwave-induced plasma optical emission spectrometer model MIP OES 4200 (Agilent Technologies, Santa Clara, CA, USA) was used in all determinations. The sample introduction comprised an Agilent SPS
3 automatic sampler, a double-pass cyclonic spray chamber, and inert OneNeb nebulizer. A liquid N₂ Dewar model 4107 (Agilent Technologies) was used as a plasma gas source.

Background correction was automatically performed using MP Expert software. The peristaltic pump rotation was 15 rpm. The stabilization time was 15 s and the integration time was 3 s. Two critical operating parameters (nebulizer gas flow rate and viewing position) were automatically optimized by the software MP Expert (Agilent Technologies). The Analytical wavelengths (nm) chosen were Al (396.152), Mn (403.449), P (214.915), B (249.772), Pb (283.305), K (766.491), Cr (425.433) and Na (589.592). The nebulizer pressure and viewing position were optimized automatically by the instrument.

To compare the results, the samples were analyzed by ICP OES employing a model 720 series spectrometer (Agilent Technologies, Santa Clara, USA). The equipment was operated under conditions for power (1.10 kW), plasma gas flow (15 L min⁻¹), auxiliary gas flow (1.5 L min⁻¹) and nebulizer gas flow (0.75 L min⁻¹) using a Sturman-Master chamber and a V-Groove nebulizer. The wavelength (nm) were: Al (394.401), Cr (267.716), P (213.618), B (249.733), Mn (259.372), Pb (283.305), Na (568.821) and K (769.897).

2.2. Reagents, solutions

All chemicals used in the experiments were of analytical grade. The aqueous solutions were prepared using ultrapure water (resistivity > 18 MΩ cm) obtained from a Milli-Q® purification system (Millipore, Bedford, USA). All flasks and glassware were soaked in 10% (v/v) HNO₃ for at least 12 h and rinsed with water before use.

Stock solution the Al (III), Mn (II), P(V), B (III), Pb (II), K(I), Cr (VI), Na (I) (1000 mg L⁻¹) (SpecSol, Quimlab, Química e Metrologia, São José dos Campos, Brazil) were used to prepare the standard reference solutions (diluted in HNO₃ 1 mol L⁻¹), used for calibration experiments. Nitric acid and hydrogen peroxide (Merck, Darmstadt, Germany) were used for digestion.

2.3 Samples

The samples of Malbec wines were collected from the Mendoza region of Argentina and Cabernet Sauvignon, Tannat, Alicante, Tempranillo, Petit Syrah, from San Francisco Valley, Bahia, Brazil.

2.3.1 Wine samples digestion procedure

For the sample preparation procedure, 2.0 mL of wine, 1.0 mL concentrated nitric acid (60 % w/v) and 0.5 mL of 30 % (w/v) hydrogen peroxide was digested in a glass vessel and then heated at 150 °C on a digester block. Finally, the contents were transferred to 50 mL volumetric flasks and diluted with water.

2.4. Statistical analysis
In the classification by LR, SVM, and DT use algorithms implemented in Python from sklearn open-source library for machine learning [27]. Predictions were arranged in a confusion matrix to calculate accuracy, sensitivity, and specificity based on true positive (TP), true negative (TN), false positive (FP), and false negative (FN) prediction values. Accuracy is the percentage of the model that is correct in its predictions. Sensitivity refers to the percentage of correct answers about the positive class. Specificity is the percentage of correct answers concerning the negative class. These measurements are calculated as the following equations [28]:

\[
\text{Accuracy} \, (\%) = \frac{TP + TN}{TP + TN + FP + FN} \times 100
\]

\[
\text{Sensitivity (Recall)}, \% = \frac{TP}{TP + FN} \times 100
\]

\[
\text{Specificity}, \% = \frac{TN}{TN + FP} \times 100
\]

Results And Discussion

3.1 Multielement determination in wine samples

Al (III), Mn (II), P (V) B (III), Pb (II), K (I), Na (I) and Cr (VI) were determined in 47 samples of red wines from Brazil and Argentina using a microwave-induced plasma optical emission spectrometer. The range (min-max) of concentration of each analyte found in the wine samples can see in Table 1. Residues from agrochemicals such as insecticides, fungicides and fertilizers that contain metal salts can be considered sources of these elements in wines [29]. Catarino and Curvelo-Garcia state that the mineral elements present in wine come essentially from root absorption, having a constant enrichment during the formation and maturation of the grape berry [30]. The presence of some of these elements can be explained from the agronomic point of view since phosphorus and potassium are applied to the soil as fertilizers [31] widely used in Brazilian soils. According to Rizzon and Miele, potassium is the most abundant cation in wine, it is released by the skin and seeds of the grapes during maceration [32]. The authors also point out that in addition to the soil, the plant, the climatic conditions and the system of winemaking conditions, the concentration of conditions also depends on the ionic balance in the wine. Lead is an element that is usually in the soil in low concentration, so the high concentrations detected in wine samples can be from an environmental factor [33]. According to Galani-Nikolakaki et al., in low amounts the manganese is a natural constituent of grapes and small variations in concentration found can be justified by the difference in soil composition [34]. The aluminum source in wine is considered exogenous and is associated with the use of pesticides, contact with aluminum surfaces [6]. Bentonite clay is also considered source of aluminum contamination in wines [35].

3.2 Analytical performance
The limit of detection defined three times the standard deviation of the blank (3.3σ/b) and limit of quantification ten times the standard deviation of the blank, divide by the slope of the calibration curve, (10σ/b), where σ denotes the standard deviation for measurements of blank and b stands for the linear section slope of the analytical curve considering type I and type II errors (false positives and false negatives). The limit of detection (LOD) found varied from 0.0005 to 0.11 mg L\(^{-1}\), whereas the limit of quantification (LOQ) found varied from 0.002 to 0.17 mg L\(^{-1}\), the values were shown in the Table 2. Calculations were made following the recommendations by IUPAC. The accuracy of the method applied in the multi-element determination in wine samples by MI POES was evaluated by comparing the results with ICP OES (Table 3). The T-test has demonstrated that there was no significant difference between the results found.

### 3.3 Artificial intelligence application

The data were organized in a matrix which have the predictor variables (mineral composition) in the columns and the target variables (classes of wines) in the rows. Although the number of samples (instances) of Argentine wines (n = 26) was higher than the number of samples of Brazilian wines (n = 21), the difference is small. In this way, the dataset can be considered balanced. To identify redundancies between the predictor variables, we made a Pearson correlation matrix and removed from the ML model one of the variables in the pair with strong linear correlation (r > 0.8). In executing the ML models, the algorithm standardized and divided the data into a training set (80%) and a test set (20%). Also, we assess the importance of predictor variables for ML algorithms. For this, use the Recursive Feature Elimination technique [27]. In this technique, the algorithm builds the ranking model with all predictor variables and computes an importance score (F1-score) for each predictor. It then removes the least important predictor; the model is rebuilt and recalculated the importance scores. In each loop, the predictor variables are recursively removed by the algorithm until collinearities and dependencies disappear and the most relevant predictors remain.

In evaluating the predictive performance of the ML models, we use the Cross-Validation (CV) method repeated ten times. In the k-fold cross-validation technique, the input dataset (D) is randomly divided into k subsets \(D_1, D_2, D_3, \ldots D_k\) (fold) of approximately equal sizes. The ML model's training is with all data, except with subset k-1, and then the model is tested with subset k-1. The process is repeated k times and, each time for test different subset (k-1) is reserved. The model's accuracy is obtained based on the correct predictions made in the test set (k-1). The model's performance by the average accuracies obtained in each fold is estimated [28].

As shown in Figure 1, with the Recursive Feature Elimination (RFE) method, we verified that of the eight predictor variables (P, B, Pb, K, Mn, Cr, Al, and Na), for all models, only four are relevant for classification. In the case of the SVM model, the most important predictors were B, Cr, K, and Na. For the LR model, the most important variables for classification were B, Cr, K, and Pb, while the most important variables for the DT model were B, Cr, K, and Mn. The predictor variables Cr, B, and K were the most important for all models studied.
Despite the tendency of the RFE method to remove redundant and weak variables for classification, the variables that proved to be the most important in the classification of wine samples were the most correlated. Of the most important variables for classification models, K with Cr (r = 0.78) and B with Cr (r = -0.79) showed the strong linear correlation, B and K had a moderate linear correlation (r = -0.57). All other predictor variables have a weak linear correlation with each other (Table 4).

The DT model has an internal method to measure the relative importance of predictor variables (“Gini importance”) from the calculation of mean impurity reduction (MDI) of the node, as can be seen in detail in the studies by [36-38]. Although the measurement of the importance of the variables by MDI is very useful, it can be biased favoring variables with high cardinality and continuous. Permutation-based variable importance calculation is used to overcome the limitations of MDI and consists of permuting each variable randomly and calculating the change in model performance [38]. Thus, the variable with the greatest impact on performance is considered the most important [38]. In the DT model, the most important predictors by MDI in the classification of wine samples were Cr, K, B, and Mn (Figure 2-a). This result obtained with the MDI method confirms the result obtained with the RFE method. In the calculations of importance with permutation, only the Cr and K variables proved to be the most important for the DT model (Figure 2-b).

Similar results were found by [39] who pointed out B, K, and Na as the most important elements in the differentiation of sparkling wines from South American and European countries by logistic regression. According to the authors, Brazilian and Argentine wines were completely discriminated against by a single element (Mn). For the Brazilian and Argentine wines analyzed in this work, this fact did not occur.

Perfectly correlated variables are truly redundant so that no additional information is gained by adding them to the classification model [40]. However, a very high correlation between the variables does not mean an absence of complementarity [40]. Especially when it comes to data with few predictor variables, simply removing correlated or weak variables can degrade classification model performance [41]. Two weak variables that are useless alone can provide better ranking performance when used together, and correlated variables can be complementary in ranking [40].

In this sense, we compared the performances of the SVM, DT, and LR models using all predictor variables (P, K, Mn, B, Cr, Pb, Al, and Na) and using only the variables considered most important (Cr, K, and B). In both cases, all classification models exhibited an average accuracy of 100% in the cross-validation (CV=10). However, the LR model presented the problem of “perfect separation” (or perfect prediction) that can occur when a predictor can perfectly classify the target variables without the estimation of a model [42].

In the LR model, the perfect predictor was Cr. In other words, the content of this element is enough to differentiate Argentine wines from Brazilian ones without the need for a classification model. Can verify this fact in simple bivariate plots and marginal distributions of the most important variables (Figure 3), in which we can easily distinguish Brazilian from Argentine wines. Also, can see in Figure 3 that the most striking feature of Argentine wines is the higher content of B, and Brazilian wines are the higher content of
K and Cr. The higher content of B in Argentine wines compared to Brazilian wines was also observed by [39] who identified this element among the most important in the classification of South American wines.

We also compared the performance of the models with the set of variables P, K, Mn, Pb, and Al, which do not include the most important predictor variables. In this case, the DT and LR models showed better performances than the SVM model (Table 4). In the SVM model, for the variables P, K, Mn, Pb, and Al, the radial basis kernel function (RBFK) presented a better ability to discriminate the wine classes, with an average accuracy in the CV of 79% (see Figure 4). With the linear, polynomial (degree 3), and sigmoid kernel function, the SVM model presented an average accuracy of 75%, 77%, and 55%, respectively.

In works involving the classification, certification of origin and authenticity of foods, supervised ML algorithms are more useful than unsupervised ones, because they allow the labeling of data and, consequently, predictions from the input data. The SVM has many advantages, such as the possibility of non-linear classification, it works well with high-dimensional data, it is robust against overfitting and suffers less impact from outliers. The main disadvantages are the possible difficulties in selecting the Kernel function and the algorithm does not perform well on very large datasets [43,44]. In the case of RF, in addition to presenting advantages similar to those of SVM, with RF it is possible to obtain the importance of predictors for classification and the algorithm has good performance in unbalanced data sets [45]. The disadvantages of RF are that it is often relatively slower than other classification algorithms and the classification rules used in the model are difficult to interpret. An alternative to RF is the Decision Tree (DT) which, although similar to RF, is much simpler and the classification model is easier to interpret. However, DT is more susceptible to overfitting than RF, but in practice there are several tools to avoid overfitting, such as pre-pruning and post-pruning [46].

Logistic regression (RL), despite not being among the most commonly used supervised algorithms in food chemistry, is one of the most popular classification methods. RL is one of the simplest and easiest ML algorithms to implement, in addition to not requiring high computational power [47]. The main disadvantages of RL are that it does not perform well on nonlinear data and is limited to binary classification problems [48].

The three classification models tested in this study proved to be efficient in discriminating between Argentine and Brazilian wines. In the SVM model, the classification was successful with the linear Kernel function and RBFK. The parameter C (cost) that determines the balance between allowing training errors and forcing rigid margins to allow flexibility in class separation was C = 1. In the classification by DT, the model performed well with 100 trees, a maximum tree depth of 3, and the Gini function for calculating node impurities.

When applying the LR model, the predictor variables cannot be highly correlated. For this reason, the algorithm fits the data in a PCA and uses the independent variables in the model. Using all predictor variables from the data set, the two principal components (in the PCA) applied in the LR model had a cumulative variance of 68%. With the most important variables identified by RFE (B, Cr, and K), the two principal components used in the LR model had a cumulative variance of 95%. With the variables P, K,
Mn, Pb, and Al, the two principal components in the LR exhibited an accumulated variance of 70%. The regression had LLR (Log-Likelihood Ratio) $p$-value < 0.05 (LLR $p$-value = $8.768 \times 10^{-5}$), demonstrating that the model it’s significant.

**Conclusion**

In this work, we show that Brazilian and Argentine wines are accurately discriminated by different supervised machine learning models (SVM, DT, and RL) based on their amounts of chemical elements. Also, we detail the most important chemical components in the discrimination of wines by the classification models and we verified that the high levels of B were a peculiar characteristic of Argentine wines, while the high levels of K and Cr were of Brazilian wines. All machine learning models studied were efficient in classifying wines using the most important predictor variables (B, Cr, and K) and had accuracy of 100%. When we used only the least important variables in wine classification (P, Pb, Mn, Al, and Na), the models were still able to efficiently discriminate Brazilian wines from Argentinean wines and had an accuracy between 80 and 90%. In the case of the SVM model, the classification of wines was successful using the linear Kernel functions and radial basis kernel function.

Through the multi-element determination of wine samples from the region of the San Francisco Valley and Argentina, it was possible to verify the grouping of 47 samples in 2 distinct groups using Artificial intelligence.

The multielement microwave induced optical emission spectrometry (MIP OES) technique was used to determine the chemical species present in the studied samples due to the ease and low operating cost of the equipment. In order to evaluate the analytical performance, reference samples were analyzed, and the detection and quantification limits were calculated, attesting the analytical accuracy and precision.

**Declarations**

**Acknowledgments**

The authors are grateful to Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for providing grants, fellowships and financial support.

**Availability of data and materials**

The datasets supporting the conclusions of this article are included within the article and its additional files.

**Compliance with Ethical Standards**

**Conflict of Interests**
The authors declare that there is no conflict of interests regarding the publication of this paper.

The author Candice N. Carneiro declares that she has no conflict of interest.

The author Federico J. V. Gomez declares that he has no conflict of interest.

The author Adrian Spisso declares that he has no conflict of interest.

The author Maria Fernanda Silva declares that she has no conflict of interest.

The author Jorge L. O. Santos declares that he has no conflict of interest.

The author Fabio de Souza Dias declares that he has no conflict of interest.

**Ethical approval:** This article does not contain any studies with human participants or animals performed by any of the authors.

**Informed consent:** Not applicable.

**References**


Tables

Table 1. Concentration of each analyte found in Argentinian and Brazilian wine samples
<table>
<thead>
<tr>
<th>Wine Samples</th>
<th>P (min-max)</th>
<th>B (min-max)</th>
<th>Pb (min-max)</th>
<th>Mn (min-max)</th>
<th>Cr (min-max)</th>
<th>Al (min-max)</th>
<th>K (min-max)</th>
<th>Na (min-max)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAN</td>
<td>142-144</td>
<td>1.7-4.0</td>
<td>&lt;LOQ-0.1</td>
<td>&lt;LOQ</td>
<td>0.44-0.49</td>
<td>&lt;LOQ</td>
<td>876-940</td>
<td>20.53-23.02</td>
</tr>
<tr>
<td>CAB</td>
<td>181-194</td>
<td>4.5-5.2</td>
<td>0.06-0.1</td>
<td>&lt;LOQ</td>
<td>0.36-0.39</td>
<td>0.84-1.38</td>
<td>1460-1496</td>
<td>9.5-12.61</td>
</tr>
<tr>
<td>PS</td>
<td>271-277</td>
<td>3.0-4.7</td>
<td>&lt;LOQ</td>
<td>0.03-0.04</td>
<td>0.34-0.45</td>
<td>0.00</td>
<td>1664-1701</td>
<td>8.84-11.49</td>
</tr>
<tr>
<td>TEM</td>
<td>196-208</td>
<td>6.0-9.3</td>
<td>0.1-0.2</td>
<td>0.23-0.29</td>
<td>0.38-0.4</td>
<td>LOQ-16.7</td>
<td>1585-1620</td>
<td>19.02-20.48</td>
</tr>
<tr>
<td>ALI</td>
<td>218-227</td>
<td>4.3-7.0</td>
<td>0.1-0.2</td>
<td>0.07-0.13</td>
<td>0.36-0.4</td>
<td>&lt;LOQ</td>
<td>1314-1373</td>
<td>14.55-16.03</td>
</tr>
<tr>
<td>CAB</td>
<td>157-163</td>
<td>5.0-5.6</td>
<td>0.05-0.06</td>
<td>0.67-0.76</td>
<td>0.35-0.47</td>
<td>0.62-0.70</td>
<td>1179-1182</td>
<td>21.64-26.33</td>
</tr>
<tr>
<td>TEM</td>
<td>156-161</td>
<td>3.9-7.9</td>
<td>0.1-0.11</td>
<td>0.67-0.72</td>
<td>0.35-0.39</td>
<td>0-9.33</td>
<td>1264-1341</td>
<td>23.81-27.89</td>
</tr>
<tr>
<td>2RM</td>
<td>67-69</td>
<td>5.6-7.6</td>
<td>0.06-0.14</td>
<td>&lt;LOQ</td>
<td>0.35</td>
<td>0-22.5</td>
<td>1029-1036</td>
<td>22.03-23.09</td>
</tr>
<tr>
<td>3RM</td>
<td>146-149</td>
<td>5.3-5.5</td>
<td>&lt;LOQ</td>
<td>&lt;LOQ</td>
<td>0.34-0.35</td>
<td>0-9.06</td>
<td>1070-1092</td>
<td>19.11-26</td>
</tr>
<tr>
<td>4RM</td>
<td>92-97</td>
<td>9-11</td>
<td>LOQ-0.3</td>
<td>&lt;LOQ</td>
<td>0.35-0.37</td>
<td>&lt;LOQ</td>
<td>848-919</td>
<td>12.43-18.56</td>
</tr>
<tr>
<td>5RM</td>
<td>149-155</td>
<td>9.3-9.6</td>
<td>&lt;LOQ</td>
<td>0.76-0.77</td>
<td>0.15-0.17</td>
<td>0.76-1.1</td>
<td>723-844</td>
<td>19.02-19.05</td>
</tr>
<tr>
<td>6RM</td>
<td>88-90</td>
<td>5.2-7.5</td>
<td>&lt;LOQ</td>
<td>0.23-0.24</td>
<td>0.15-0.18</td>
<td>LOQ-0.97</td>
<td>636-775</td>
<td>11.58-17.22</td>
</tr>
<tr>
<td>7RM</td>
<td>144-149</td>
<td>7.7-8.0</td>
<td>&lt;LOQ</td>
<td>0.24-0.26</td>
<td>0.15</td>
<td>LOQ-1.35</td>
<td>766-785</td>
<td>22.84-25.52</td>
</tr>
<tr>
<td>8RM</td>
<td>216-221</td>
<td>10-11.6</td>
<td>&lt;LOQ</td>
<td>0.45-0.52</td>
<td>0.15</td>
<td>0.29-1.82</td>
<td>780-791</td>
<td>13.02-17.92</td>
</tr>
<tr>
<td>9RM</td>
<td>116-123</td>
<td>7.7-8.5</td>
<td>&lt;LOQ</td>
<td>0.57-0.66</td>
<td>0.14-0.15</td>
<td>0.06-0.47</td>
<td>666-751</td>
<td>21.07-24.9</td>
</tr>
<tr>
<td>23SU</td>
<td>162-171</td>
<td>7.6-8.3</td>
<td>&lt;LOQ</td>
<td>0.56-0.62</td>
<td>0.15-0.17</td>
<td>0-0.83</td>
<td>604-636</td>
<td>23.63-24.38</td>
</tr>
<tr>
<td>10DP</td>
<td>152-153</td>
<td>9.8-10</td>
<td>&lt;LOQ</td>
<td>0.56-0.58</td>
<td>0.16</td>
<td>0.01-0.85</td>
<td>803-809</td>
<td>32.63-38.25</td>
</tr>
<tr>
<td>13DP</td>
<td>152-160</td>
<td>8.4-9.0</td>
<td>&lt;LOQ</td>
<td>0.29-0.34</td>
<td>0.15-0.19</td>
<td>0.3-0.62</td>
<td>807-855</td>
<td>16.03-18.65</td>
</tr>
</tbody>
</table>
### Table 2. Evaluation of Analytical performance

<table>
<thead>
<tr>
<th>Apple leaves (Nist 1515)</th>
<th>P</th>
<th>B</th>
<th>Pb</th>
<th>K</th>
<th>Mn</th>
<th>Al</th>
<th>Cr</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MIP OES (mg Kg⁻¹)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1418 ± 37</td>
<td>24 ± 4</td>
<td>0.50 ± 0.16</td>
<td>15685 ± 738</td>
<td>57 ± 5</td>
<td>285 ± 31</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td><strong>Certified Value (mg Kg⁻¹)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1593 ± 68</td>
<td>27.6 ± 2.8</td>
<td>0.470 ± 0.024</td>
<td>16080 ± 210</td>
<td>54.1 ± 1.1</td>
<td>284.5 ± 5.8</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td><strong>LODᵃ (mg L⁻¹)</strong></td>
<td>0.02</td>
<td>0.02</td>
<td>0.01</td>
<td>0.05</td>
<td>0.0007</td>
<td>0.005</td>
<td>0.0005</td>
</tr>
<tr>
<td><strong>LOQᵇ (mg L⁻¹)</strong></td>
<td>0.06</td>
<td>0.08</td>
<td>0.03</td>
<td>0.17</td>
<td>0.002</td>
<td>0.02</td>
<td>0.002</td>
</tr>
</tbody>
</table>

ᵃ limit of detection;ᵇ limit of quantification

### Table 3. Multi-element determination in wine samples by MIP OES and comparison of results with ICP OES (n=3).
<table>
<thead>
<tr>
<th>Sample</th>
<th>Al (mg L(^{-1}))</th>
<th>Pb (mg L(^{-1}))</th>
<th>P (mg L(^{-1}))</th>
<th>Mn (mg L(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MIP OES</td>
<td>ICP OES</td>
<td>MIP OES</td>
<td>ICP OES</td>
</tr>
<tr>
<td>Wine 1</td>
<td>0.42 ± 0.04</td>
<td>0.45 ± 0.06</td>
<td>0.12 ± 0.01</td>
<td>0.09 ± 0.03</td>
</tr>
<tr>
<td>Wine 2</td>
<td>0.33 ± 0.03</td>
<td>0.35 ± 0.03</td>
<td>0.10 ± 0.01</td>
<td>0.09 ± 0.02</td>
</tr>
<tr>
<td>Wine 3</td>
<td>0.54 ± 0.01</td>
<td>0.50 ± 0.05</td>
<td>0.05 ± 0.01</td>
<td>&lt; LOQ</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Cr (mg L(^{-1}))</th>
<th>K (mg L(^{-1}))</th>
<th>B (mg L(^{-1}))</th>
<th>Na (mg L(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MIP OES</td>
<td>ICP OES</td>
<td>MIP OES</td>
<td>ICP OES</td>
</tr>
<tr>
<td>Wine 1</td>
<td>0.15 ± 0.01</td>
<td>0.17 ± 0.03</td>
<td>698 ± 7</td>
<td>710 ± 10</td>
</tr>
<tr>
<td>Wine 2</td>
<td>0.44 ± 0.03</td>
<td>0.40 ± 0.02</td>
<td>1110 ± 100</td>
<td>1016 ± 80</td>
</tr>
<tr>
<td>Wine 3</td>
<td>0.35 ± 0.03</td>
<td>0.42 ± 0.04</td>
<td>1200 ± 50</td>
<td>1186 ± 60</td>
</tr>
</tbody>
</table>

**Table 4** - Confusion matrix and classification metrics obtained from SVM, DT, and RL model with the predictor variables most important by RFE (Cr, B, and K) and with the least important variables (P, Pb, Mn, Al, and Na).
## Confusion Matrix (Predictors: Cr, B, and K)

<table>
<thead>
<tr>
<th></th>
<th>RFE-SVM (linear, C=1)</th>
<th>RFE-DT (Gini, nt=100)</th>
<th>RFE-LR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentina(^b)</td>
<td>5</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Brazil(^b)</td>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Argentina(^a)</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Brazil(^a)</td>
<td>5</td>
<td>0</td>
<td>5</td>
</tr>
</tbody>
</table>

## Classification model metrics (Predictors: Cr, B, and K)

<table>
<thead>
<tr>
<th></th>
<th>Accuracy (%)</th>
<th>Specificity (%)</th>
<th>Recall (%)</th>
<th>F-score (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RFE-SVM (linear, C=1)</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>RFE-DT (Gini, nt=100)</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>RFE-LR</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

## Confusion Matrix (Predictors: P, Pb, Mn, Al, and Na)

<table>
<thead>
<tr>
<th></th>
<th>SVM (RBF, C=1)</th>
<th>DT (Gini, nt=100)</th>
<th>LR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentina(^b)</td>
<td>4</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Brazil(^b)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Argentina(^a)</td>
<td>2</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Brazil(^a)</td>
<td>4</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

## Classification model metrics (Predictors: P, Pb, Mn, Al, and Na)

<table>
<thead>
<tr>
<th></th>
<th>Accuracy (%)</th>
<th>Specificity (%)</th>
<th>Recall (%)</th>
<th>F-score (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM (RBF, C=1)</td>
<td>80</td>
<td>100</td>
<td>83</td>
<td>80</td>
</tr>
<tr>
<td>DT (Gini, nt=100)</td>
<td>90</td>
<td>100</td>
<td>87</td>
<td>87</td>
</tr>
<tr>
<td>LR</td>
<td>90</td>
<td>100</td>
<td>90</td>
<td>90</td>
</tr>
</tbody>
</table>

DT= decision tree, SVM=support vector machine, RL= logistic regression, RFR= Recursive Feature Elimination, a=actual class, b=predicted class, C=Cost, nt=tree number

### Figures
Figure 1

Variable importance ranking estimated by recursive feature elimination (RFE) for the SVM (linear), DT and RL models. The most important variables for the model are assigned rank 1, followed by the least important in ascending order of rank.

Figure 2
Relative importance of predictor variables measured by decision tree (DT) with A) mean impurity decrease (MDI) and B) permutation.

**Figure 3**

Bivariate plot and marginal distributions of Brazilian and Argentine wines versus concentration of chemical elements (B, mg L\(^{-1}\); K, mg L\(^{-1}\); Cr, mg L\(^{-1}\)).
Figure 4

Classification decision boundary plots using support vector machine with parameter C (cost) = 1 and different kernel functions. Legend: 1= Argentine wines; 2= Brazilian wines