ORIGINAL PAPER

Provenance discrimination of commercial grape juices from Argentina and Brazil based on elemental composition and chemometric methods

Brenda V. Canizo[1] · Ana Laura Diedrichs[2,3] · Agustín Londonio[4,5] · Patricia Smichowski[4,6] · Roberto G. Pellerano[7] · Rodolfo G. Wuilloud[1]

Abstract
The knowledge of a product's origin is an important aspect for consumers who demand quality and authenticity. Grape juice is a beverage that has currently gained more attention due to the tendency of people towards a more natural and healthy lifestyle. The aim of this study was to discriminate the origin of grape juice samples. For this purpose, chemometric techniques were applied to commercial samples of grape juice from Argentina and Brazil, and the feasibility of classification models of predicting the origin of samples based on elemental composition, was investigated. Inductively coupled plasma mass spectrometry (ICP-MS) was used to determine 10 elements (V, Cr, Mn, Fe, Ni, Cu, As, Rb, Y and Mo). Unsupervised methods, such as principal component analysis (PCA) and cluster analysis (CA) and the supervised technique, linear discriminant analysis (LDA) with bootstrapping and k-fold cross-validation were assessed. The best result was achieved for LDA with 4 fold cross validation with a prediction accuracy of 81%.

Keywords Grape juice · Chemometric · ICP-MS analysis · Provenance · Classification

Introduction

The grape juice obtained from the edible part of the grape (Vitis ssp.) through a technological process, is a beverage that has currently gained more attention due to the tendency of people towards a more natural and healthy lifestyle [1–3].

This beverage is an excellent source of a wide range of beneficial components, including carbohydrates, vitamins, minerals, polyphenols, flavonoids among others [3–5]. In the last years, Argentina and Brazil have emerged as important producers of this beverage, being Brazil also a good consumer of this fruit juice. Nevertheless, in the case of

✉ Rodolfo G. Wuilloud
rwuilloud@mendoza-conicet.gob.ar;
rodolfowuilloud@gmail.com

1   Laboratorio de Química Analítica para Investigación y Desarrollo (QUIANID), Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Cuyo, Instituto Interdisciplinario de Ciencias Básicas (ICB), CONICET UNCUYO, Padre J. Contreras 1300, (5500), Mendoza, Argentina

2   Laboratorio DHARMa, Departamento Sistemas de Información, Facultad Regional Mendoza, Universidad Tecnólogica Nacional (FRM-UTN), Mendoza, Argentina

3   Grupo GridTICs, Departamento Electrónica, Facultad Regional Mendoza, Universidad Tecnológica Nacional (FRM-UTN), Mendoza, Argentina

4   Comisión Nacional de Energía Atómica, Gerencia Química, Buenos Aires, Argentina

5   Instituto de Investigación e Ingeniería Ambiental (3iA), Universidad de San Martín, Martín de Irigoyen, San Martín, Buenos Aires 3100, 1650, Argentina

6   Consejo Nacional de Investigaciones Científicas y Técnicas, Buenos Aires, Argentina

7   Instituto de Química Básica y Aplicada del Nordeste Argentino (IQUIBA-NEA), Facultad de Ciencias Exactas y Naturales y Agrimensura, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Universidad Nacional del Nordeste (UNNE), Av. Libertad 5470, (3400), Corrientes, Argentina

Argentina; the grape juice production is principally destined to the exportation as concentrated grape juice [6].

Regarding consumer behavior, the demand for quality and authenticity in food and beverages has risen significantly. Currently, the principal concerns are focused not only on safety products, furthermore on protection against adulteration with low-quality foodstuff and correct labelling with detailed composition, properties and origin of the product. As a result, a precise monitoring system has been developed to ensure the quality and authenticity of products, providing genuine protection for consumers and trustworthy producers [7–11].

Determining analytical parameters in samples is an excellent way to assess their provenance, presence of frauds and adulteration. Many studies have shown that the chemical compounds of grapes and derived-products, principally wines and juices, are influenced by climatic and agricultural conditions related to cultivations of the fruit plants and growth, and by environmental factors such as soil composition, irrigation water composition, sunlight exposure, among others [3, 7, 12–14]. Then, the mineral composition reflects the geochemistry in which grapevines are grown and therefore the grape derived products are influenced by the composition of the soil [15]. Analyzing the concentrations of elements in grape juice samples and combining them with chemometric tools is an excellent way to identify the geographical origin of this beverage. Numerous techniques exist for elemental analysis, among these, inductively coupled plasma optical emission spectrometry (ICP-OES) or mass spectrometry (ICP-MS) are the most used [16, 17].

In this context, statistical treatment is indispensable for extracting information from the data generated, and for identifying similarities and differences between samples. Moreover, the application of chemometric methods has proven to be a versatile and valuable tool for determining the authenticity of food and beverage [3, 8]. In chemometric, the pattern recognition methods are powerful for assessing the geographic origin of food products. These techniques involve an exploratory approach in the initial stage of an investigation using principal component analysis (PCA) or hierarchical cluster analysis (HCA), followed by a discrimination step. These tools are appreciated for ensuring truth-in-labeling, as they help verify the origin of the production region and identify any potential adulteration [4, 18].

Understanding the origin of grape juice is crucial to ensure quality, achieve differentiation in the market, and instill consumer confidence in their product choices. Additionally, it can lead to enhanced market access and increased competitiveness on the global stage. Therefore, the aim of this study was to develop an approach to perform the origin classification of grape juice samples from two important producing countries of South America, based on elements chemical analysis combined principally with the implementation of simple pattern recognition techniques.

Materials and methods

Reagents

Ultrapure water (resistivity of 18 mΩ·cm) obtained from a water purifier equipment OSMOION-APEMA (Buenos Aires, Argentina) was used in all the experiments. Ultrapure concentrated nitric acid 65% (v/v) (Merck, Darmstadt, Germany) was used to perform the mineralization of grape juice samples. For calibration and quality control steps, certified multi-element ICP standard solutions, rhodium (^{103}Rh) and holmium (^{165}Ho) mono-elemental internal standard solutions from Chem-Lab, Ionex Reference Standard (Zedelgem, Belgium) were used. All the glassware was cleaned with 0.5 mol L^{-1} HNO$_3$ solution for 24 h and then rinsed with ultrapure.

Grape juice samples collection and treatment

In this study, a total of 26 commercial grape juices of different red grape varieties were purchased from several markets of Argentina (16) and Brazil (10). The samples were stored at ambient temperature in the original packings and prior to analysis were thoroughly shaken.

For ICP-MS analysis, grape juices were mineralized by wet digestion in open-vessel systems as follows: an aliquot of 5 mL of each sample was put in contact with 2 mL of HNO$_3$ (65% w/w) and left at room temperature for 30 min. After this step, a heat treatment was applied using a heat plate for 45 min at 100 °C. Subsequently, the resulting solutions were left to cool and transferred to volumetric flasks and the volume was completed to 100 mL with ultrapure water [12].

ICP-MS analysis

The elemental analysis of grape juice samples digested was carried out on a Perkin-Elmer model NexIon 300X ICP-MS, provided with a CETAC AXS-520 autosampler, nickel sampler and skimmer cones and a concentric nebulizer. The isotopes, in order of mass number, ^{51}V, ^{52}Cr, ^{55}Mn, ^{57}Fe, ^{60}Ni, ^{63}Cu, ^{75}As, ^{85}Rb, ^{89}Y and ^{98}Mo were determined using operating conditions of RF power of 1250 W; nebulizer and plasma gas flow rate of 0.95 L min^{-1} and 16 L min^{-1} respectively; dwell time of 50 ms; scanning conditions: 20 sweeps per reading, 1 reading per replicate and 3 number of replicates; the data acquisition mode was time resolved

analysis. The detailed instrumental conditions are summarized in Table 1.

The spectrometer was calibrated against external certified standard solutions with concentrations from 0.5 to 100 µg L^{-1}. To correct non-spectral interferences, Rh and Ho (10 µg L^{-1}) were used as internal standards for all determinations. The limit of detection (LOD) for the procedure was calculated with 3 s/m, where m corresponds to the calibration curve slope and s corresponds to the 10 blank solutions standard deviation in the regression of the calibration curve.

Chemometric analysis

The data matrix for the chemometric treatment contained 26 rows and 11 columns. The matrix rows correspond to the number of juice samples and the columns to the concentrations of the ten elements analyzed and one column represents the grape juice geographical provenance as the dependent categorical variable. All samples belonging to the same country were assigned by the same code (AR for Argentina and BR for Brazil). In the data matrix, the values of the element concentrations that could not be detected by ICP-MS in the samples studied were replaced by the corresponding numerical value of the limit of detection (LOD). The dataset was autoscaled as a preprocessing method to avoid wide differences in data dimensionality and misleading results in the interpretation of the generated plots of the chemometrics techniques, providing equal importance to all variables measured.

In order to identify similarities and differences among grape juices produced in Argentina and Brazil, basic exploratory analysis of the samples was made by principal component analysis (PCA) and cluster analysis (CA) as unsupervised pattern recognition methods. Pattern recognition techniques are powerful tools for determining the geographical origin of food and beverages products based on the data provided by analytical instruments [18]. These techniques can be subdivided into supervised and unsupervised methods. In unsupervised tools there is no previous

knowledge of the grouping of the samples is required, and hence produces the grouping itself. These methods are often very advantageous at an initial stage of a study because it presents simple visual plots of the data structure [18, 19]. In contrast, in supervised methods there is prior knowledge of the grouping structure and representative samples of each group are available to establish relationships between given inputs and outputs to enable prediction of the output from new inputs.

PCA was performed to investigate the effect of country of origin on the parameters analyzed in grape juices by showing the data structure and trends in the experimental data obtained. This technique reduces the dimensions of the original data matrix by explaining a large part of the variance using synthetic factors, called principal components (PCs). Moreover, permits the visualization of the original arrangement of grape juices in an n-dimensional space, principally 2 or 3 dimensions, by identifying the directions in which maximum variability is retained and displaying the relationship between variables and observations [17, 18, 20, 21].

The other typical unsupervised method also used to identify groups or clusters of juice samples was CA. Within this method, two clustering algorithms were assessed, hierarchical cluster analysis (HCA) and k-means clustering. When employing HCA, the original data was separated into a few general groups, each of which is further divided into still smaller groups until finally the individual objects themselves remain. The results are presented in the form of dendrograms to facilitate the visualization of sample relationships. For HCA, the Euclidean distance was used as a dissimilarity measure and Ward's method was used to suggest the clustering of grape juices. Ward's method minimizes the square-error increase and minimizes the intra-cluster variance while maximizing the intercluster variance [9, 22]. K-means is a well-known and simple used method for partitioning a dataset of n observations into k clusters in which each observation belongs to the cluster with the nearest mean. Each cluster is represented by its center (i.e., centroid) which corresponds to the mean of points assigned to the cluster and iteratively the center is updated by the mean of its constituent. The algorithm converges when there is no further change in assignment of instance to clusters [23].

A supervised statistical technique was also evaluated since PCA and CA cannot be properly considered as classification methods. Linear discriminant analysis (LDA) was used to further investigate the possible classification of the grape juices based on its country of origin and propose a predictive model. In LDA the knowledge of the grouping structure allows to develop rules which predict the group that a new object belongs to, by determining the so-called discriminant functions [18].

Table 1 Instrument setting for ICP-MS analysis

Instrument	PerkinElmer NexIon 300X ICP-MS con CETAC ASX-520 Autosampler
Solution delivery	0.85 mL min^{-1}
Nebulizer	MEINHARD® Concentric Nebulizer
RF power	1250 W
Argon flows	Nebulizer, 0.95; Plasma, 16; auxiliary, 1.2 (all in L min^{-1})
Dwell time	50 ms
Data acquisition mode	Time resolved analysis
Number of replicates	3
Isotopes monitored	^{51}V, ^{52}Cr, ^{55}Mn, ^{57}Fe, ^{60}Ni, ^{63}Cu, ^{75}As, ^{85}Rb, ^{89}Y y ^{98}Mo

All statistical chemometrics procedures were computed using the statistical R software version 3.6.0 [24].

Results and discussion

Mineral content of grape juices

The elemental concentrations of the grape juice samples analyzed in this study are summarized in Table 2. It shows the mean and standard deviation values expressed in µg L^{-1} of each element determined in juice samples as a function of their country of origin, Argentina and Brazil. In addition, the non-parametric Kruskal-Wallis tests were carried out to highlight significant differences in the trace element contents between the juices from the two countries and the p-values generated is presented (Table 2).

From the analysis of the results presented in Table 2, it is clear that Fe was the most abundant element found in grape juice samples. The highest mean level of this element was found in Argentina (2409 µg L^{-1}) with more than the double concentration found in Brazil (1037 µg L^{-1}). Rb was the highest element found in grape juices from Brazil; this element presented a mean concentration of 1794 µg L^{-1} while in juices from Argentina this element was quantified with a mean concentration of 417 µg L^{-1}. The next elements that follow in order of concentrations were Mn and Cu. The highest mean level of Mn was found in grape juice samples from Brazil (1041 µg L^{-1}), whereas the mean concentration of this element in samples from Argentina was of 680 µg L^{-1}. With respect to Cu element, it was observed that in samples from Argentina the mean concentration was of 515 µg L^{-1} and in Brazil juices, this element has a mean content of 297 µg L^{-1}. The concentration levels of Fe, Mn and Cu elements in grape juices were in agreement to those reported by other authors [16, 25]. With respect to other works on grape juices, differences were found in the levels of these elements [1, 26–31]. These mineral elements in natural levels concentration, participate in important biological functions for the human organism. In the case of Rb, concentrations were in accordance with previous works reported in the literature [4, 25, 27].

The next elements quantified in this study at minor levels of concentration were Cr, Ni, As, V and Mo with mean concentrations that were in accordance with previous works carried out in the literature [16, 25, 31]. Among them, Cr is important because in its trivalent state it is required for humans and plays an important role in glucose, protein and fat metabolism. On the other hand, hexavalent chromium is a human health hazard associated with chronic toxicity. Our methodology is not able to differentiate the oxidation state in which this element is found in grape juice, the levels determined in all samples were under the Recommended Dietary Allowances (RDAs) of chromium set by the U.S. Department of Health and Human Services [32]. Regarding to As, it is an element that is known to be toxic, however the values obtained in our work were lower than the maximum residual level established in juices and fruit nectars by the Argentine food code and the Brazilian legislation [31, 33]. Finally, Y was found in low concentration or were not detected in the grape juices samples.

In order to make visible the differences in the concentrations of the elements measured in juice samples from the two countries under study, Fig. 1 is presented. In this figure, box-plots of the range and distribution of the elements concentrations in grape juice samples from Argentina and Brazil can be compared. It can be observed clear differences in the concentration of Cr, Fe and Rb between the two countries, indicating that they could be variables of interest for the geographic classification problem. This was confirmed with the results obtained by Kruskal-Wallis tests that revealed that As, Cr, Fe, Rb and Y elements were statistically significant ($p < 0.05$) between samples according their geographical origin.

Table 2 Mean and standard deviation of concentration of the analyzed elements in grape juice samples according to their country of origin (concentrations are expressed as µg L^{-1}). *P-values* generated by the non-parametric Kruskal-Wallis test are included

Element	LOD[a]	Argentina		Brazil		p-value[b]
		Mean	SD	Mean	SD	
As	0.01	28.2	47.8	6.7	9.6	0.002*
Cr	0.2	41.3	7.6	21.5	16.5	0.002*
Cu	0.1	515	408	297	222	0.187
Fe	2.9	2409	1315	1037	684	0.004*
Mn	0.03	680	543	1041	678	0.140
Mo	0.01	12.3	16.9	4.8	4.9	0.241
Ni	0.05	20.6	9.7	31.3	17.9	0.155
Rb	0.06	417	338	1794	1372	<0.001*
V	0.05	24.6	33.3	15.8	22.5	0.460
Y	0.003	0.11	0.10	0.01	0.03	0.003*

[a]LOD: limit of detection

[b]p-value < 0.05 denoted with an asterisk*, indicates that the differences between the means are statistically significant
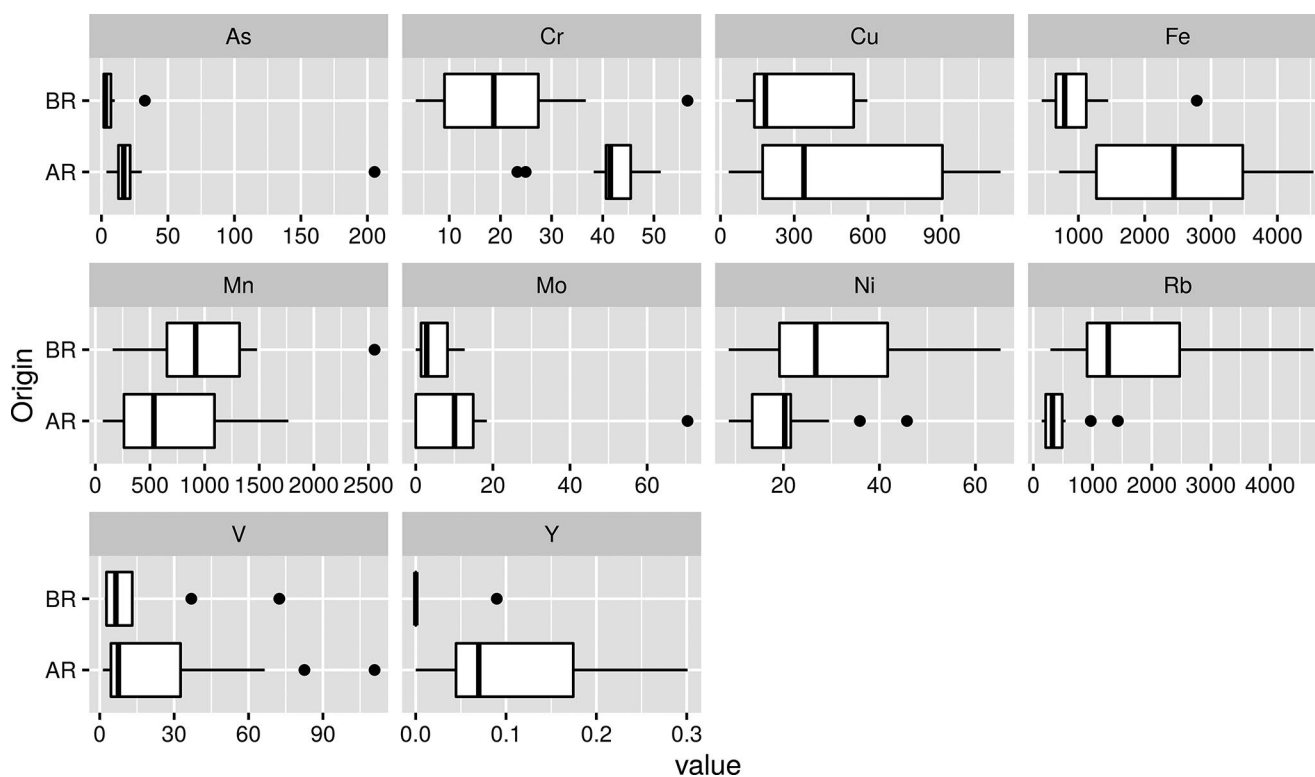
Fig. 1 Box plot of the ten elements analyzed according to grape juice provenance

The nutritional value of fruit juices, in relation to the presence of macro-, micro-, and trace elements, is indispensable for sustaining life. Nonetheless, an excessive presence of these elements can potentially serve as a source of toxic elements, with some of them exhibiting cumulative effects. Essential elements such as Na, K, Ca, and P are vital for human health, whereas metals like Pb, Cd, Hg, and As have been found to induce harmful effects, even at low levels. Other minerals like Fe, Cu, and Zn are essential in specific quantities; however, their excessive ingestion can result in adverse effects [25, 34, 35]. The predominant origin of elements within fruit juices is inherent to the fruits themselves, subject to a multitude of agricultural variables linked to the cultivation and maturation of fruit-bearing plants. These factors encompass the plant variety, maturity of harvested fruits, mineral composition of the soil, types of pesticides and fertilizers employed, as well as prevailing climatic conditions [12, 16]. Another source of elements in fruit juices is associated with post-harvest procedures, encompassing fruit storage, handling, and various juice processing steps. Additionally, the choice of storage conditions and packaging materials can influence the element concentrations in the final fruit juice product [12].

Soil mineralization can significantly influence macroelements like K, Mg, and S, along with certain minor and trace elements such as B, Fe, Mn, Zn, and Rb. Furthermore, the use of fertilizers or synthetic pesticides can notably impact the subsequent concentrations of macroelements like Ca, K, and Mg, as well as microelements including Cu, Fe, Mn, and Zn [17, 31]. During the manufacture of juices, potential sources of element contamination, including Al, Ca, Cu, Fe, Mg, Na, Si, Sr, and Ni, may arise from tap water, technological processes, as well as the packaging and containers employed and storage [31, 34, 36].

In this work, the relationships between concentrations of the elements analyzed in grape juices were studied and Pearson's correlation coefficient (r) was used for this purpose. Figure 2 presents the graphic with the different grade of correlation within the variables, it can be observed a high positive correlation between Mo and As ($r = 0.94$), meaning that increasing the concentration of one increases the another. High positive correlations was also exhibited by the pairs of elements Rb-Mn ($r = 0.7$), and the rest of the variables presented moderate and almost negligible correlations. Relations between elements may be attributed to their similarity in properties of some elements, as well as their coexistence in nature [4, 16]. Moreover, high correlation is also explained by common exposure sources, e.g. water supply, geographical origin, or soil and agricultural practices [37]. In this work the strong positive correlation observed between Mo-As as well as Rb-Mn lends support an assumption that these metals mat share a common origin [38].
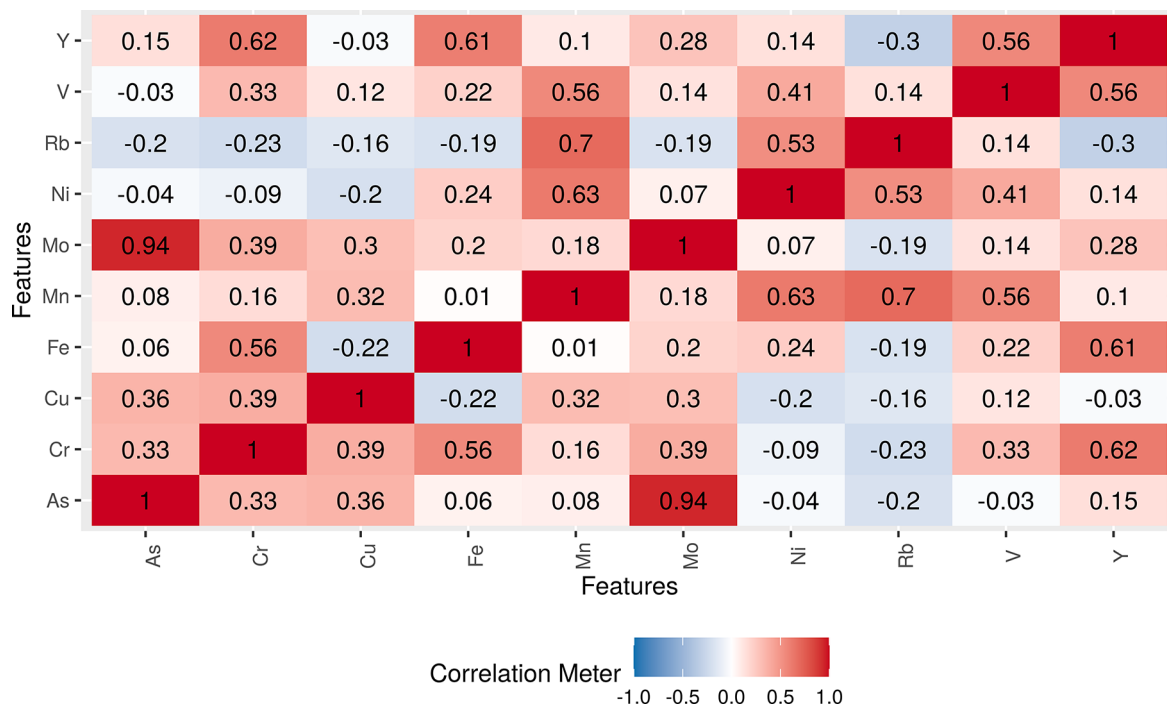
Fig. 2 Pearson's correlation coefficients (r)

Multivariate exploratory analysis

Basic chemometric characterization of the studied grape juice samples was performed by PCA. In this work, the concentration data of 10 elements was used to perform the PCA analysis. The total information content of the given number of principal components was expressed by cumulative percent value of the total variance. The results showed that the first two principal components explained 53.6% of total variance. The first principal component (PC1) represented 30.7%, and the next principal component (PC2) 22.9%, the three first principal components represent 70.3% of the total variability of the data. Figure 3 presents the most important PCA plots, the score plot (Fig. 3a) and the loading plot (Fig. 3b) obtained by using the first two principal components. In the scores plot it can be observed an acceptable separation between the juice samples from Argentina and Brazil. Samples from AR showed principally negative scores on the considered PC1 and were differentiated from the BR samples, which have positive scores on PC1. Figure 3b plots the importance of each variable in the most informative principal components PC1 and PC2. The PC1 was strongly associated with the values of Cr, Y, Mo, Fe and As, indicating higher concentrations in samples that showed negative scores (AR samples). On the other hand, Mn, Rb, Ni, and V were the dominant variables in the second principal component, and samples with positive scores on PC2 (BR juices) correspond to high concentrations of these elements. In order to facilitate the visualization of the

relative importance of each element or variable in the principal components, Fig. 3c is presented. The results obtained by PCA showed that from the analysis of 10 elements, the grape juice samples could be differentiated in terms of their country of origin due to a visible differences in the concentrations of some elements in the samples of the two countries. As previously mentioned, elements such as Cr, Mo, Y, and Fe exhibit a high concentration in juices from AR, whereas BR juices demonstrate a significant presence of Rb, Mn, and Ni. These differences allow for a clear differentiation between the two groups in the score plot.

Another unsupervised exploratory technique such as cluster analysis was also performed to characterize the role of the elements analyzed in classifying grape juices based on their country of origin. In this work hierarchical cluster analysis was used for grouping experimental samples into clusters, based on similarity within class and dissimilarity between different groups, according to the country of origin. Ward's method with the classical approach for distance measure, Euclidean distance, was used. In Fig. 4a is presented the dendrogram of HCA applied to data to observe the grouping feature. By analyzing the dendrogram, it is possible to identify two large clusters. The first was composed of nine samples, in which all samples corresponded to juices from Argentina. The second and largest cluster contained the remaining seven juices from Argentina and the total samples from Brazil. The results achieved were in agreement with those obtained with PCA, in which it can be
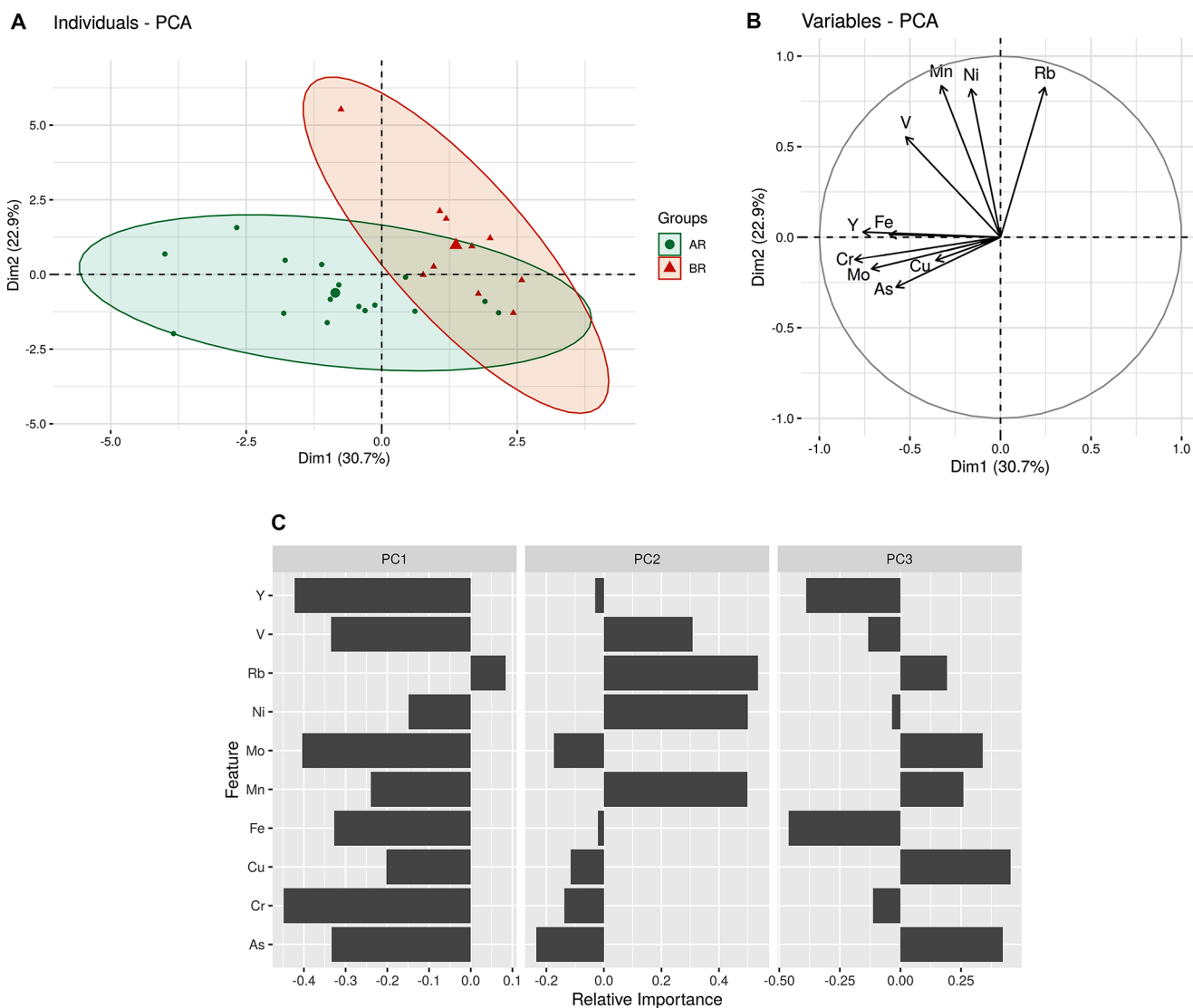
Fig. 3 PCA scores (**a**) and loading (**b**) plots of PC1 versus PC2. Relative importance of the elements for the three first principal components (**c**)

differentiated two large groups, with same samples of AR and BR grouping together.

K-means cluster method was also performed for partitioning the given dataset into a set of k groups (i.e.k clusters) where k represents the number of groups setting by the analyst as a prerequisite for applying the algorithm. In this work k-means clustering was applied considering $k = 2$, since there is only two categories AR and BR. To plot the observations grouped in the two clusters, a 2-dimensional clustering plot, "clusplot" was presented in Fig. 4b. This graphic showed component 1 on the x-axis and component 2 on the y-axis, and the clusters were represented as ellipses that contains the objects (samples). The ellipses are based on the average and the covariance matrix of each cluster, and their size is such that they contain all the points of their group. For higher-dimensional datasets it is applied a dimension reduction technique before constructing the plot.

After analysing this plot, it can be observed that in cluster 2 most of the observations were grape juice samples from Argentina, except one sample (point 17) that was a juice from Brazil. The cluster 1 contained nine samples from Brazil and five from Argentina. These results show a better performance of the algorithm with respect to HCA analysis.

Figure 5 shows the box-plots of the distribution of the elements analyzed for each of the two clusters obtained by k-means. In this figure it can be observed differences in the concentrations of Cr, Fe, Mo, Rb and Y within the two clusters, indicating higher concentrations of the elements Cr, Fe, Mo and Y in cluster 2 in which most of the observations were grape juice samples from Argentina and higher level of Rb in cluster 1, with the majority of samples from Brazil. This confirms the fact presented above, indicating that these variables are good features for the geographic classification problem.
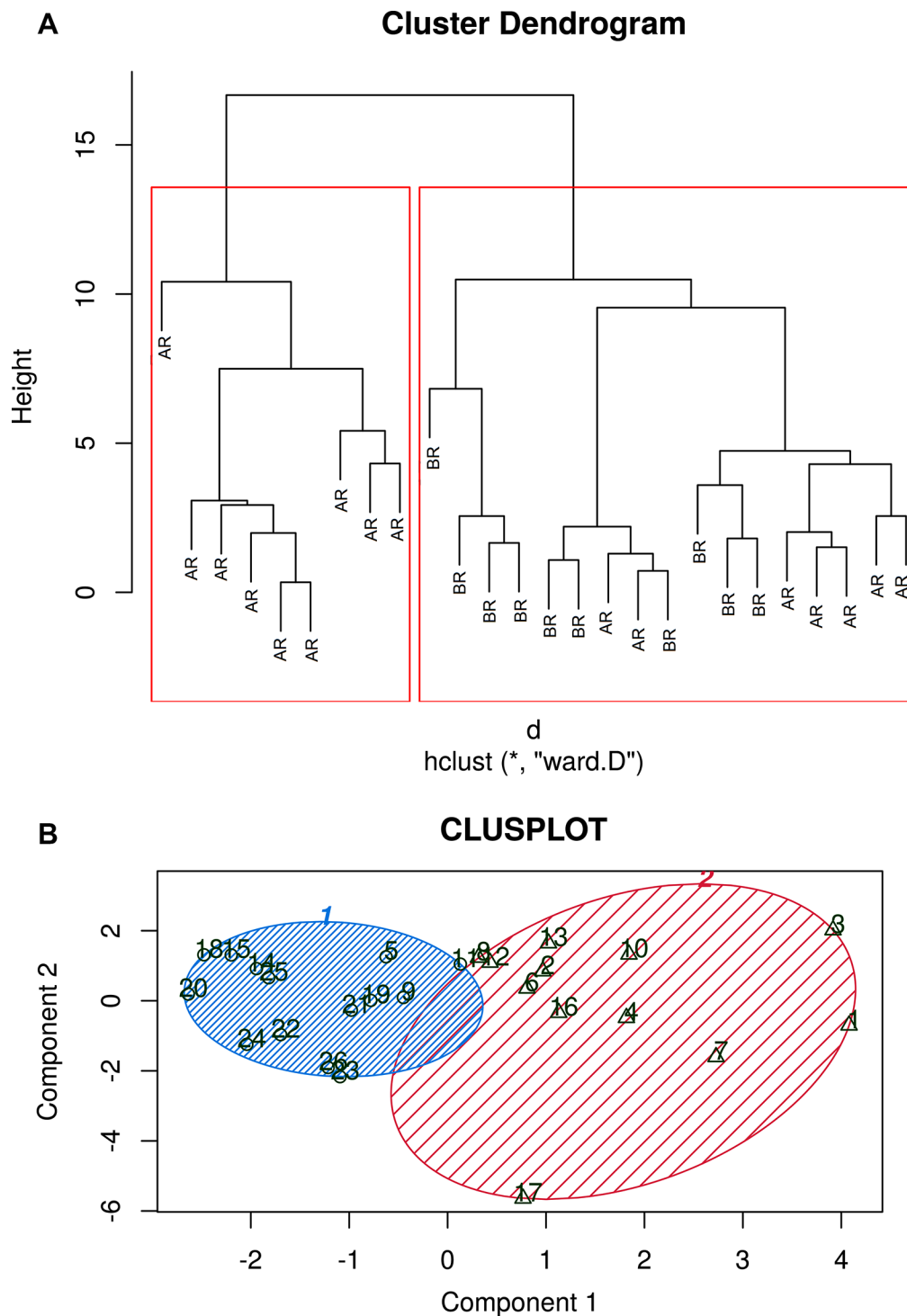
A

Cluster Dendrogram

B

CLUSPLOT

Fig. 4 Hierarchical cluster dendrogram (**a**). Cluster plot for k-means analysis (**b**)

In conclusion, the exploratory analysis based on the elements determined, allowed a separation of the grape juices samples according to their country of origin.

Classification models

PCA and CA exploratory analysis show good results in the geographical differentiation of grape juices analyzed in this work, nevertheless, these techniques cannot be considered as classification methods, since these are unsupervised
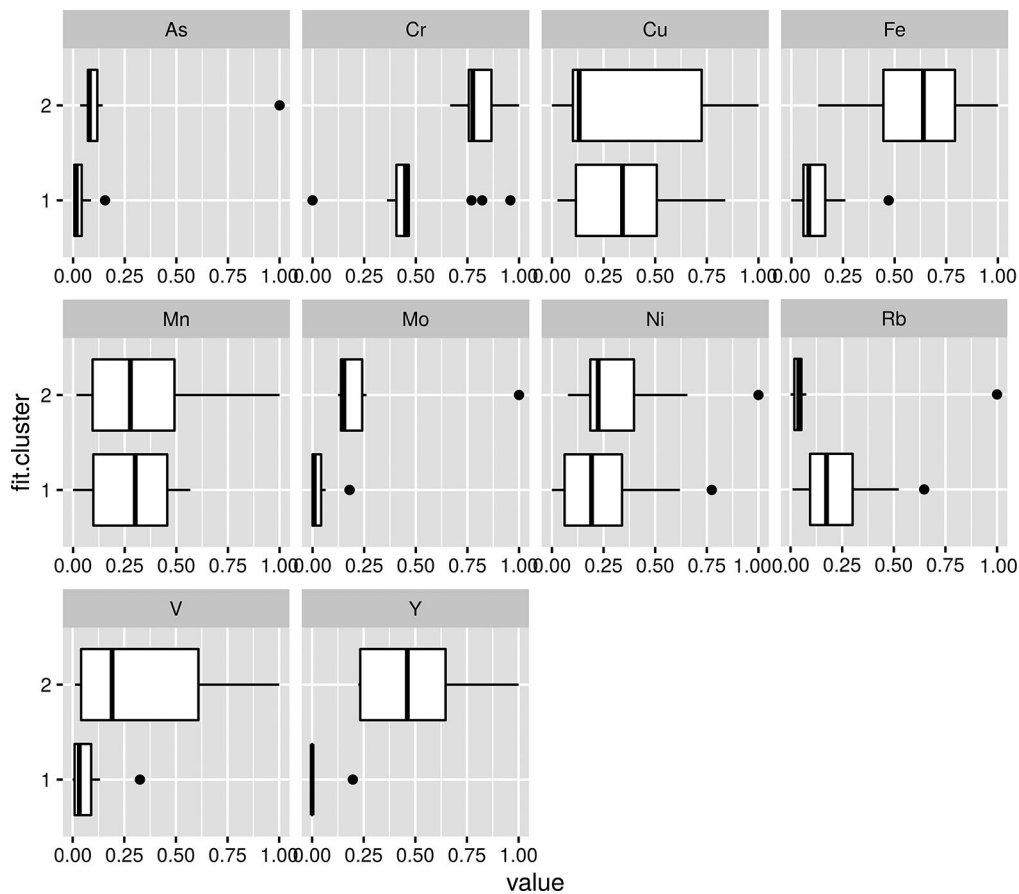
Fig. 5 Box plot of the distributions of the elements in the clusters obtained by k-means

multivariate statistical tools, meaning that they just show the data as they are, and do not use the information to predict classes of objects [39, 40]. For this reason, LDA was used to further assess the possible geographical classifications of grape juices from Argentina and Brazil, based on the different element concentrations analyzed and obtain a predictive model.

Due to the low number of samples available per country of origin, it was no possible to perform model training with data split approach. Therefore, it was necessary to use other approaches, such as bootstrapping (BTS) and cross-validation (CV) especially suitable for use in situations when small data samples are available as validation techniques to reduce variance and overfitting of the models. BTS and CV are methods designed to estimate the classifier performance using the samples available via partitioning (CV) or resampling with replacement (BTS) [41].

The BTS creates bootstrap samples (random subsets of model training and validation data) by iteratively collecting observations from the initial dataset with replacement, i.e. hundreds or thousands of repeats that implies that the same observation can occur more than once. This results in a sample with same size as the original dataset. When using

BTS, is necessary to choose the number of repeats, is this work it was applying BTS with 100 repetitions.

The CV involves partitioning a dataset into subsets, performing the analysis on one subset (called the training set), and validating the analysis on the other subset (called the validation set). In k-fold CV, the dataset is randomly split into k mutually exclusive subsets (the folds) of approximately equal size. The classifier is trained and tested k times, and to estimate the accuracy of the models, the overall number of correct classifications is divided by the number of instances in the dataset. In this work k-fold cross validation was used with k = 4 for the model.

LDA method was applied considering all grape juice samples and their accuracy was analyzed. LDA with bootstrapping gave an overall classification accuracy of 74% and a kappa value of 0.45. In the case of LDA with 4-fold CV the result of accuracy was of 81% and the kappa value was 0.62. It can be observed that by comparing the results, the predictive capability has been increased using CV validation approach for LDA method.

Conclusions

This paper describes the application of ICP-MS analysis combined with chemometric techniques for the geographical discrimination of commercial gape juices obtained from Argentina and Brazil. The concentration levels of ten chemical elements were interpreted using unsupervised methods such as PCA and CA, which provided an exploratory and simple approach for evaluation the origin and hence the authenticity of the grape juice samples. The results of these techniques allowed a separation of the grape juices samples according to their country of origin. In addition LDA model was performed with two validation methods, partitioning (CV) or resampling with replacement (BTS). This method provide good origin prediction ability for samples, the best result was achieved with LDA k-fold CV, which provided an 81% of accuracy.

Finally, achieving a geographical classification of grape juices from Brazil and Argentina through mineral profiling and chemometric analysis is significant. Trace element levels of grape juices are relevant markers of the provenance and have good potential to discriminate the origin of commercial grape juices. Utilizing advanced chemometric techniques enables precise classification, contributing not only to understanding elemental characteristics but also enhancing viticulture and oenology. This classification empowers producers to communicate authentic regional identity, fostering market differentiation and consumer trust. In essence, this pursuit stands at the intersection of science, industry, and consumer preference.

Declarations

Conflict of interest The authors declare that they have no financial interests or personal relationships that could have influenced the work reported in this document.

References

1. R.E.S. Froes, W.B. Neto, N.O.C. Silva, R.L.P. Naveira, C.C. Nascentes, B. B. Da Silva. Spectrochim. Acta Part B at. Spectrosc. **64**, 619 (2009)
2. Codex Alimentarius Commission, Codex Stand. **247**, 1 (2005)
3. C.H. Junges, C.C. Guerra, N.A.P.C. Reis, A.A. Gomes, F.S. Diogo, M.F. Ferrão, J. Food Compos. Anal. **114**, 104793 (2022)
4. A. Szymczycha-Madeja, M. Welna, D. Jedryczko, P. Pohl, TrAC Trends Anal. Chem. **55**, 68 (2014)
5. R.L. Bhardwaj, U. Nandal, A. Pal, S. Jain, Fruits. **69**, 391 (2014)
6. Cámara, Argentina de Fabricantes y Exportadores de Mosto, (2019)
7. D. Granato, M. de Magalhães, V. Carrapeiro, Fogliano, S.M. van Ruth, Trends Food Sci. Technol. **52**, 31 (2016)
8. J. Guo, T. Yue, Y. Yuan, Y. Wang, J. Agric. Food Chem. **61**, 6949 (2013)
9. C. Jiménez-Jorquera, M. Gutiérrez-Capitán, in *Electron. Noses Tongues Food Sci* (Elsevier, 2016), pp. 189–198
10. M. Díaz-García, M.R. Castellar, J.M. Obón, Compr. Anal. Chem. **60**, 439 (2013)
11. J.L.P. Calle, M. Ferreiro-González, A. Ruiz-Rodríguez, D. Fernández, M. Palma, Agronomy. **12**, 683 (2022)
12. P. Pohl, D. Jedryczko, A. Dzimitrowicz, A. Szymczycha-Madeja, M. Welna, P. Jamroz, *Fruit Juices* (Elsevier, 2018), pp. 739–761
13. H. Shimizu, F. Akamatsu, A. Kamada, K. Koyama, K. Iwashita, N. Goto-Yamamoto, J. Food Compos. Anal. **107**, 104377 (2022)
14. M. Dasenaki, N. Thomaidis, Molecules. **24**, 1014 (2019)
15. M. Likar, K. Vogel-Mikuš, M. Potisek, K. Hančević, T. Radić, M. Nečemer, M. Regvar, Sci. Total Environ. **505**, 724 (2015)
16. A. Szymczycha-Madeja, M. Welna, Food Chem. **141**, 3466 (2013)
17. E.M. Borges, D.A. Volmer, E. Brandelero, J.M.L.N. Gelinski, M. Gallimberti, F. Barbosa, Food Anal. Methods. **9**, 362 (2016)
18. V. Uríčková, J. Sádecká, Spectrochim. Acta - Part A Mol. Biomol. Spectrosc. **148**, 131 (2015)
19. M. Kozak, C.H. Scaman, J. Sci. Food Agric. **88**, 1115 (2008)
20. T. Margraf, É.N.T. Santos, E.F. de Andrade, S.M. van Ruth, D. Granato, Food Res. Int. **82**, 145 (2016)
21. A.A.F. Zielinski, C.W.I. Haminiuk, C.A. Nunes, E. Schnitzler, S.M. van Ruth, D. Granato, Compr. Rev. Food Sci. Food Saf. **13**, 300 (2014)
22. G.M. Downs, J.M. Barnard, in *Rev. Comput. Chem. Vol. 18* (John Wiley & Sons, Inc., Hoboken, New Jersey, USA, 2003), pp. 1–40
23. A. Kassambara, *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning* (2017)
24. R Core Team (2019)
25. L. Tormen, D.P. Torres, I.M. Dittert, R.G.O. Araújo, V.L.A. Frescura, A.J. Curtius, J. Food Compos. Anal. **24**, 95 (2011)
26. L.A. Rizzon, M. Link, Ciência Rural. **36**, 689 (2006)
27. L.A. Rizzon, A. Miele, Food Sci. Technol. **32**, 93 (2012)
28. I.M. Toaldo, F.A. Cruz, T.D.L. Alves, J.S. de Gois, D.L.G. Borges, H.P. Cunha, E.L. da Silva, M.T. Bordignon-Luiz, Food Chem. **173**, 527 (2015)
29. M.P. Fabani, M.E. Toro, F. Vázquez, M.P. Díaz, D.A. Wunderlin, J. Agric. Food Chem. **57**, 7409 (2009)
30. C. Dani, L.S. Oliboni, D. Prá, D. Bonatto, C.E.I. Santos, M.L. Yoneama, J.F. Dias, M. Salvador, J.A.P. Henriques, Genet. Mol. Res. **11**, 3154 (2012)
31. A. Miele, L.A. Rizzon, de Queiroz, C. Gianello, Food Sci. Technol. **35**, 120 (2015)
32. Agency for Toxic Substances and Disease Registry (ATSDR), Atlanta, GA U.S. Dep. Heal. Hum. Serv. Public Heal. Serv, (2012)
33. Secretaría, de Estado de Salud Pública (2017)
34. A. Dehelean, D.A. Magdas, Sci. World J. 2013, 1 (2013)
35. I.M. Toaldo, O. Fogolari, G.C. Pimentel, J.S. de Gois, D.L.G. Borges, V. Caliari, M. Bordignon-Luiz, LWT - Food Sci. Technol. **53**, 1 (2013)
36. V.L.C. Bragança, P. Melnikov, L.Z. Zanoni, Biol. Trace Elem. Res. **146**, 256 (2012)
37. I. Coelho, A. Rego, S. Gueifão, A.C. Nascimento, I. Castanheira, J. Chemom. **31**, e2868 (2017)
38. Ş. Tokalıoğlu, Food Chem. **134**, 2504 (2012)
39. D.B.R. Todeschini, in *Infrared Spectrosc. Food Qual. Anal. Control*, edited by D.-W. Sun, 1st Edition (Academic Press, 2009), pp. 83–102
40. M. Kamal, R. Karoui, Trends Food Sci. Technol. **46**, 1 (2015)

41. R. Kohavi, in *Proc. 14th Int. Jt. Conf. Artif. Intell* (Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1995), pp. 1137–1143