Detection of adulterants in grape nectars by attenuated total reflectance Fourier-transform mid-infrared spectroscopy and multivariate classification strategies

Carolina Sheng Whei Miaw[a,b,c], Marcelo Martins Sena[d], Scheilla Vitorino Carvalho de Souza[a], Maria Pilar Callao [c] * and Itziar Ruisanchez[c]

a Department of Food Science, Faculty of Pharmacy (FAFAR), Federal University of Minas Gerais (UFMG), Av. Antônio Carlos, 6627, Campus da UFMG, Pampulha, 31270-010, Belo Horizonte, MG, Brazil.

b CAPES Foundation, Ministry of Education of Brazil, 70040-020, Brasília, DF, Brazil.

c Chemometrics, Qualimetric and Nanosensors Grup, Department of Analytical and Organic Chemistry, Rovira i Virgili University, Marcel·lí Domingo s/n, 43007 Tarragona, Spain

d Department of Chemistry, Institute of Exact Sciences (ICEX), Federal University of Minas Gerais (UFMG), Av. Antônio Carlos, 6627, Campus da UFMG, Pampulha, 31270-010, Belo Horizonte, MG, Brazil.

* Corresponding author: mariapilar.callao@urv.cat

Abstract

There is no any doubt about the importance of food fraud control, as it has implications in food safety and in consumer health. Focusing on fruit beverages, some types of adulterations have been detected more frequently, such as substitution with less expensive fruits. A methodology based on attenuated total reflectance Fourier-transform mid-infrared spectroscopy (ATR-FTIR) and multivariate classification was applied to detect whether grape nectars were adulterated by substitution with apple juice or cashew juice. A total of 126 samples were obtained and analyzed. Two strategies were proposed: one-class and multiclass approaches. Soft independent modeling of class analogy (SIMCA), partial least squares discriminant analysis (PLS-DA) and partial least squares density modeling (PLS-DM) were used to build the models. Among them, PLS-DA presented the best performance with a sensitivity and specificity of nearly 100%. The multiclass strategy was preferred if the adulterants to be studied are known because it provides additional information.

Highlights

- The detection of grape nectar adulteration with cashew or apple was studied.
- One-class and multiclass approaches were implemented.
- The multivariate classification methods SIMCA, PLS-DA and PLS-DM were compared.
- PLS-DA provided better performance to detect grape nectar adulterations.

Keywords

Food adulteration, Fruit nectar, PLS-DA, SIMCA, One-class classification, Multiclass classification.

1. Introduction

Because of the highly competitive market, drink industries are always searching for product diversification, and in recent years, the largest increase in production was of fruit nectar (Neves, Trombin, Lopes, Kalaki, & Milan, 2012). Nectar is defined as an unfermented beverage produced by the dilution in water of the edible part of fruits or vegetables or their extracts with the addition of sugars, intended for direct consumption (Brazil, 2009). In Brazil, Standards of Identity and Quality (SIQ) are established for fruit nectars and cover the minimum percentages of pulp that must be used in each type of nectar. For some fruits, the minimum parameters include soluble solids (SS), total titratable acidity (TTA), total sugars (TS) and ascorbic acid (AA) (MAPA, 2003, 2013). According to the Brazilian Association of Soft Drinks and Non-Alcoholic Beverages (ABIR), the most consumed nectar in Brazil is grape flavor (ABIR, 2015).

Considering the issue of adulteration of fruit-based beverages, the most frequent practices include substitution with cheaper ingredients, such as simple dilution with water or sugar syrup, and undeclared addition of different species, which can be botanically related, or not, to the main fruit in question (Asadpoor, Ansarin, & Nemati, 2014).

Methods based on spectroscopic techniques are generally rapid, non-destructive, simple and require little or no sample preparation. However, they have the disadvantage of low specificity. Therefore, powerful tools for adulteration testing can be created by combining these techniques with multivariate chemometric methods, while some authors applied just basic statistical techniques (El Darra et al., 2017). Classification methods are particularly suitable for food fraud detection. They can be differentiated in discriminant and class-modeling methods. The most common discriminant method is partial least squares discriminant analysis (PLS-DA), while the most used class-modeling method is soft independent modeling of class analogy (SIMCA) (Bevilacqua et al., 2013).

The necessity of food quality control was reflected in a specific review concerning the development of an effective food traceability system to reduce the numerous cases of food safety incidents and fraudulence. (Dandage, Badia-Melis, & Ruiz-García, 2017). In that sense, reviews have been recently published addressing the use of multivariate classification methods to authenticate or detect adulteration in food (Callao & Ruisánchez, 2018; Esteki, Shahsavari & Simal-Gandara, 2018; Szymańska et al., 2015). Multivariate classification methods have been successfully applied to elucidate specific problems of authenticity or adulteration in different types of food. Examples are wines (Sen & Tokatli, 2016), oils (Georgouli, Del Rincon, & Koidis, 2017), , milk (Gondim,

Junqueira, Souza, Ruisánchez, & Callao, 2017), hazelnut pastes (López, Trullols, Callao, & Ruisánchez, 2014), coffee (Bona et al., 2017), mushrooms (Xu et al., 2016), vinegar (Ríos-Reina, Callejón, Oliver-Pozo, Amigo, & García-González, 2017) and whiskies (Martins, Talhavini, Vieira, Zacca, & Braga, 2017).

Comparatively, the application of these techniques to studies involving authentication or detection of frauds in fruits and derivatives is more limited. For this aim, articles have developed for multivariate classification or calibration models employing different analytical techniques, such as UV-VIS spectroscopy (Boggia, Casolino, Hysenaj, Oliveri, & Zunin, 2013), spectrofluorometry (Ammari, Redjdal, & Rutledge, 2015), nuclear magnetic resonance (NMR) (Cuny et al., 2008) and mid-infrared spectroscopy (He, Rodriguez-Saona, & Giusti, 2007; Miaw et al., 2018; Shah, Cynkar, Smith, & Cozzolino, 2010; Shen et al., 2016;).

In the present study, the detection of grape nectar adulteration with apple and cashew juices was studied by means of attenuated total reflectance Fourier-transform mid-infrared spectroscopy (ATR-FTIR) and classification methods. Apple juice has commonly been used as filler for economic gain by beverage industries (Singhal, Kulkarni, & Rege, 1997), but it is now also being used to replace some of the added sugar. Furthermore, cashew and apple are fruits suspected of being utilized for adulterations by fraudulent industries, justifying the importance of the development of analytical methods to detect these potential adulterants in the most popular beverage products, such as the grape nectar matrix.

In this paper, two approaches were proposed considering their different purposes: one-class and multiclass approaches, utilizing discriminant or class-modeling methods. One-class classification is adequate when the goal is to test whether a sample is adulterated, regardless of which adulterant might be present (López et al., 2014). If the adulterant is known, the multiclass strategy can be chosen, since it gives additional information, such as multiple assignations and samples not assigned to any class (Gondim et al., 2017).

In recent years, some authors have criticized the predominance in the chemometric literature of the use of discriminant methods, such as PLS-DA, to food authentication problems (Rodionova, Oliveri, & Pomerantsev, 2016; Oliveri, 2017). This criticism has noted that classification results will be unreliable when the model is used to predict a new sample from an untrained class. In response, other authors have combined PLS-DA with outlier detection, identifying samples from untrained classes based on large *Hotelling* T^2 and Q residues (Martins et al., 2017). However, as class-modeling models are developed using only the information concerning one-class samples at a time, they are unable to ensure the model specificity for the detection of various food frauds (Xu et al., 2016). Considering all these relevant discussions, it is important to compare the alternatives for developing supervised classification models for detecting food fraud. Thus, SIMCA and PLS-DA, as the most used class-modeling and discriminant methods, respectively, were applied to the authentication of grape nectars. In addition, a recently proposed one-class modeling method, partial least squares density modeling

(PLS-DM) (Oliveri et al., 2014), was also applied. The three classification methods were compared through the evaluation of sensitivity and specificity.

2. Materials and methods

2.1 Formulation of nectars

Grape nectars samples, were prepared starting from reliable raw materials and rigorously meeting the established regulations (MAPA, 2003, 2013), at the Food Science Laboratory and at the Technology Laboratory, both located in the Food Department of the Faculty of Pharmacy of the Federal University of Minas Gerais (UFMG).

Isabel grape samples were obtained from EMBRAPA (the Brazilian Agricultural Research Corporation) Grape & Wine, located in Petrolina, PE, Brazil. Red cashews and Fuji apples were acquired from the Minas Gerais Supply Center (CEASA) in Contagem, MG, Brazil. The selection of fruits took into account the absence of mechanical and phytopathological damage, the degree of maturation and other typical physical characteristics of each fruit, such as size, color and texture (Paltrinieri & Figuerola, 1998). The fruits were stored in the refrigerator at 4-7ºC until the preparation of nectars (EMBRAPA, 2016).

The fruits were sanitized with 100 mg/L of sodium hypochlorite solution (Vetec Química Fina, Ltda, Rio de Janeiro, RJ, Brazil) for 2 min and washed. The juices/pulps of grape, apple and cashew were obtained as described below:

- grapes were heated under constant steam for 1 to 2 h in an autoclave (Fanem, São Paulo, SP, Brazil) at 100 °C, pressed and sieved to obtain the juice;
- apples were peeled and cut into eight pieces, and the seeds were removed. The fruits were scalded in boiling water for 3 min, followed by immersion in water with ice until cooling;
- cashews had their chestnut removed and the fruits were cut into four pieces.

Apples and cashews were individually pulped in an industrial blender (Fisatom 752, São Paulo, SP, Brazil) and sieved (1 mm sieve).

For the formulation of grape nectars, the only SIQ parameter recommended in the Brazilian legislation (MAPA, 2003) is a minimum of 50 % of pulp, which was considered in the formulations of the unadulterated nectars. The amounts of pulp/juice and syrup were estimated as described in Equation (1).

$$\frac{a \times A}{100} + \frac{b \times B}{100} + \frac{c \times C}{100} = \frac{m \times (A+B+C)}{100} \quad (1)$$

where "a" represents pulp Brix, "A" represents percentage of pulp that must be present in the nectar, "b" represents syrup Brix, "B" represents percentage of syrup, "c" represents adulterant pulp Brix, "C" represents percentage of adulterant, "m" represents final nectar Brix, and "A + B + C" is equal to 100 (pulp + syrup + adulterant) (Tressler & Joslyn, 1961).

The quantity of additives added was 0.25 g/100 g, 15 mg/100 g and 0.075 g/100 mL for citric acid, ascorbic acid and guar gum (Pryme Foods, Sorocaba, SP, Brazil), respectively. Syrup at 20 ºBrix was prepared and added to the additives in adequate proportions to produce nectars with 11 to 13 ºBrix. These values were within the ranges permitted by Brazilian legislation and based on preliminary experiments involving commercial nectars (Miaw et al., 2018).

Juices were added to the syrup, homogenized (Fisatom 752, São Paulo, SP, Brazil) and filled in labelled amber glass bottles (250-mL) with plastic screw caps (both previously sterilized by autoclaving at 100 °C for 10 min). Nectars were pasteurized in the autoclave at 100 °C for 10 min. Bottles were hermetically sealed and left at room temperature (Paltrinieri & Figuerola, 1998). After being opened for analysis, the nectar bottles were refrigerated (4 - 7 ºC).

As illustrated in **Fig. 1**, a set of 42 samples of grape nectar were prepared for each of the three studied classes: unadulterated, adulterated with cashew, and adulterated with apple.

First, seven representative batches of each class were prepared according to the following formulations:

a) unadulterated batches were formulated with 50 % of grape, sugar syrup and additives (corresponding to the other 50 %).

b) batches adulterated with cashew were formulated with 40 % grape, 10 % of cashew juice, sugar syrup and additives (corresponding to the other 50 %).

c) batches adulterated with apple were formulated with 40 % grape, 10 % of apple juice, sugar syrup and additives (corresponding to the other 50 %).

Then, to obtain the 42 representative samples of each class, the 7 above described batches were mixed taking 3 of them in almost the same proportion (35/35/30) to give the additional 35 samples. The final number of samples was 126.

2.2 Instrumentation and software

The Brix degrees of each juice/pulp produced was measured using a refractometer (Hanna Instruments Brasil, Barueri, SP, Brazil).

Samples were analyzed by ATR-FTIR in an IRAffinity-1 FTIR (Shimadzu, Kyoto, Japan) spectrophotometer with a DLATGS detector (Deuterated Triglycine Sulfate Doped with L-Alanine) equipped with a horizontal ATR accessory with a ZnSe prism (PIKE

Technologies, Madison, WI, USA) of 20 internal reflections. For each sample, 1.5mL were pipetted onto the ATR cell surface and three readings were recorded with 16 scans, 4 cm^{-1} resolution, generating spectra between 4000 to 937 cm^{-1}. A background correction was performed after each measurement to avoid atmospheric interference and reduce instrumental noise.

Multivariate analysis was conducted using MATLAB software version 8.0.0.783 - R2012b (Natick, MA, USA) and PLS Toolbox 7.0.2 (Eigenvector Research Inc., Wenatchee, WA, USA).

2.3 Data analysis

2.3.1 Pre-processing and exploratory analysis

Multiplicative scatter correction (MSC) (Rinnan, Berg, & Engelsen, 2009) was applied to correct the spectra baseline deviations. Principal component analysis (PCA) was used as an unsupervised exploratory analysis tool to visualize the sample distribution in the multivariate space, to identify any natural clustering in the samples that could influence the subsequent multivariate analysis and to identify possible outliers.

2.3.2 Classification methods

Multivariate classification methods are supervised techniques. They can be divided, among other criteria, into class-modeling and discriminant methods. Discriminant methods define delimiters in the hyperspace of the variables, separating the samples into a number of regions corresponding to the number of predefined classes, and focusing on the differences between the samples from each class. Class-modeling methods build an individual model for each predefined class regardless of the information for the other classes or categories and focusing on the similarities between samples from the same class (Bevilacqua et al., 2013)

SIMCA is a modelling technique based on Principal Component Analysis (PCA) in which each class is modelled independently from all others (Bevilacqua et al., 2013). Each sample is characterized by two scalar statistics, Hotelling T^2 and Q, which measures the information from each sample included or not included in the model, respectively. Class frontiers (*Hotelling* T_{lim}^2 and Q_{lim},) are calculated for each pre-defined class (class model), at a specific significance level (α), usually set at 0.05 (Rius, Callao & Rius, 1997)

Historically, various criteria have been used for the classification of samples in SIMCA models. A common criterion assigns samples to classes based on their reduced values *Hotelling* T_r^2 and Q_r. These values are the ratios between the statistics of sample *i* (T_i^2 and Q_i) and the corresponding statistical limits for each class. A sample must have values lower than 1.0 for both the reduced parameters to be considered within the class model. The most used criterion is a slight variation of the former. A sample *i* is assigned based on its

distance from class *j* ($d_{i,j}$), which is defined as a combination of its reduced parameters (Equation (2)) (Márquez, López, Ruisánchez, & Callao, 2016). In this last case, the class boundary for a sample to be assigned as within the model is a semi-circle with a radius 1.0 (*d* equal to or lower than 1.0), so this criterion is more restrictive than considering *Hotelling* T_r^2 and Q_r statistics independently.

$$\boldsymbol{d_{ij}} = \sqrt{(\boldsymbol{Q_{r,i}})^2 + (\boldsymbol{T^2_{r,i}})^2} \quad (2)$$

PLS-DA is a discriminant method that adapts PLS regression to a classification task. It establishes a linear regression between a matrix of independent variables (**X**) and an array of dependent variables (**Y**). **Y** contains binary dummy variables that indicate the class to which each sample belongs, where 1 indicates membership and 0 does not (Barker & Rayens, 2003). Since this paper aimed to differentiate and classify between three classes, class 1 samples were encoded as (1,0,0), class 2 as (0,1,0) and class 3 as (0,0,1).

The PLS-DA model predicts the class for each sample, assigning values approximately 0 or 1. Bayesian statistics are used to calculate the threshold value above which the sample is considered to belong to the class (Bylesjö et al., 2006). The Bayesian threshold considers that y predicted values of the PLS-DA model are normally distributed, selecting the y value in which the number of false results are minimal (false-negatives and false-positives) (Pulido, Ruisanchez, Boqué, & Rius, 2003). Thus, predicted values above or below this threshold mean that a sample does or does not belong to the class, respectively.

PLS-DM is a one-class method that adapts PLS regression to a classification task. Its particularity is that PLS-DM computes the response vector **y** as an estimation of sample density, based on inter-sample distances in the multivariate space. With the algorithm used in this work (Oliveri, 2017; Oliveri et al., 2014), for each sample in the training set, the response vector **y** is calculated as the sum of Euclidean distances between *k* samples with the lowest distance in the multivariate space. The algorithm applies all possible combinations using the parameter distance of *k* nearest neighbors, smoothing coefficient α (for the definition of the class space in the PLS score domain), the number of latent variables (LV) and the pre-processing suitable for the **X** matrix. Then, the best combination is chosen with the adjustment of the number of LV using efficiency criteria (geometric mean of sensitivity and specificity) and with the evaluation of the other parameters.

For this model, the specificity is calculated in the presence of the non-target class, which can be composed of more than one extraneous class. In this case, the specificity obtained is calculated from the overall alternative class. If the specificity of each specific alternative class is required, it must be calculated for each non-target class separately (Rodionova, Oliveri, & Pomerantsev, 2016).

2.3.3 Performance Parameters

The performance parameters are measurable attributes that indicate the quality of the analytical method (López, Callao, & Ruisánchez, 2015). For qualitative methods, the most common parameters are sensitivity, specificity and the more recently proposed

inconclusive ratio. The first two are based on probabilities regarding four possible binary responses: true positive (TP) (positive response for a sample that is positive), false positive (FP) (positive response for a sample that is negative), true negative (TN) (negative response for sample that is negative) and false negative (FN) (negative response for a sample that is positive). The expressions to calculate these values are presented below.

Sensitivity (SEN) indicates the likelihood of recognizing samples that truly belong to the modeled class (samples from class *j,* $n^o Sj$, that have been properly predicted by the model as belonging to class *j*).

$$\boldsymbol{SEN_j = TP_j / n^o S_j} \quad (3)$$

Specificity (SPE) indicates the likelihood of recognizing samples that are truly different from the modeled class (samples that are not from class *j,* $n^o S\ not\ j$, that have been properly predicted as not belonging to class *j*).

$$\boldsymbol{SPE_j = TN_j / n^o S_{not\ j}} \quad (4)$$

Inconclusive ratio (IR) indicates the percentage of samples that cannot be undoubtedly assigned to class *j*, and thus considers no assignation to any class and the multiple assignation (López et al., 2014).

$$\boldsymbol{IR_j = (NA_j + MA) / n^o S_j} \quad (5)$$

where *NAj* means unassigned samples (samples that are from class *j* that are not assigned to class *j* or any other class); *MA* means multiple assignation samples (samples from class *j* assigned to more than one class) and $n^o S_j$ means the total number of samples that really belong to class *j*.

3. Results and discussion

Fig. 2 shows the mean pre-processed spectra of each predefined class under study. As previously observed (Miaw et al., 2018), the intense band near 3300 cm^{-1} and the sharp peak at 1640 cm^{-1} present in all samples are related to the O-H absorption of water (He et al., 2007; Shen et al., 2016). The region between 1700 and 1000 cm^{-1} incorporates the typical bands for phenolic compounds, such as the C=C-C aromatic ring stretching, the phenol OH bending, the aromatic C-H in-plane bending, and the C-O stretching of phenol (Bureau, Ścibisz, Le Bourvellec, & Renard, 2012). Additionally, in this region, sugars and organic acids are present showing the characteristic bands (between 1500 and 950 cm^{-1}) (Shah et al., 2010; Shen et al., 2016). The low-intensity bands between 1500 and 1200 cm^{-1} were related to the deformations of CH_2, C-C-H and H-C-O (Shah et al., 2010; Vardin, Tay, Ozen, & Mauer, 2008). For the fingerprint region (1200 to 900 cm^{-1}), the stretching vibrations of C-C and C-O bonds correspond to the presence of sugars and organic

acids (He et al., 2007; Shah et al., 2010; Vardin et al., 2008). These described components are present in all the nectars, justifying the similarities among the spectra of the three classes showed in Fig. 2.

First, an exploratory analysis by Principal Component Analysis (PCA) was performed on all the samples from the three classes studied. The scores plot of the first two principal components (PC1 *x* PC2), accounting for 90.32 % of the total variance, are illustrated in **Fig. 3**. It can be seen that PC1 could not distinguish between the 3 classes. Along the PC2, samples adulterated with apple (squares) presented negative scores values and were clearly separated from unadulterated samples, which presented positive score values (triangles). Samples adulterated with cashew (circles) appeared to clearly overlap with the unadulterated samples, and just a few of them appeared to overlap with the apple adulterated samples.

For the supervised classification modeling, each class was separated into training and test sets using the Kennard and Stone algorithm (28 samples for training and 14 for test set) which selects representative and uniformly distributed samples into the multivariate space (Kennard & Stone, 1969).

Initially, a multiclass strategy was implemented by applying SIMCA and PLS-DA classification techniques to establish the three classes: unadulterated (UN), adulterated with cashew (CAS) and adulterated with apple (APP). SIMCA models were independently established for each class using the training set and the optimal numbers of PCs were selected based on the lowest value of RMSECV (root mean square error of cross validation). The models were validated using leave-one-out cross validation as well as predictions of the test set. Three PCs for each class were necessary to build the SIMCA model, accounting for 95.20, 93.79 and 90.58 % of total variance, for UN, CAS and APP classes, respectively.

PLS-DA models were also built with the three classes. The model was validated using leave-one-out cross validation and the number of LV, chosen based on the smallest cross validation classification errors, was 6, accounting for 95.13 % of variance in the **X** block and 82.94 % in the **Y** block. The threshold values were 0.25 for the UN class, 0.14 for the CAS class and 0.09 for the APP class, as can be observed from **Fig. 4**.

The summarized class assignations obtained by applying SIMCA and PLS-DA models are presented in **Table 1**. Regarding the results obtained with SIMCA, as expected considering the PCA model shown in **Fig. 3**, samples from UN and CAS classes were multiply assigned to each other. Almost all unadulterated samples, in both training and test sets, were doubly assigned to their class and to the CAS class. To a lesser extent, samples adulterated with cashew, five from the training set and seven from the test set, were also doubly assigned to their class and as unadulterated (UN class), and seven of the 28 training samples were not assigned to any class. Finally, as expected, samples adulterated with apple were properly recognized by their class model, with no wrong or multiple assignation to other classes. Only six of the 28 samples from the training set were not assigned to any class, while all samples from the

test set were correctly assigned. As a result of the assignations, high inconclusive ratios were obtained for all three classes, and the unadulterated class was the one with the highest ratio (**Table 1**).

For results obtained with PLS-DA (**Table 1** and **Fig. 4**), no incorrect assignments were obtained. In addition, few inconclusive assignments, all corresponding to samples adulterated with cashew, were observed: one sample from the training set that was doubly assigned, and one sample from the training and three from test set that were not assigned to any class. Notably, in no cases were adulterated samples assigned as unadulterated; this outcome means that no false-negative errors were obtained. From the perspective of food fraud, false-negative errors are the most important to control, as they correspond to errors related to not detecting the contaminant when it is present.

The next step was the implementation of a one-class strategy, in which only the target class was established by all three classification methods. The UN class was considered the target class and CAS and APP samples were jointly the non-target class. This SIMCA model was similar to the previous one established for the multiclass approach. The only difference is reflected in the calculation of specificity, since CAS and APP samples were modeled together in a single class. PLS-DA was established for two contrasting classes, encoded as (1,0), with 1 as the UN class and 0 as the CAS+APP class. This model was built as in the multiclass approach, namely, the number of LV 5, which accounted for 94.29% of variance in the **X** block and 37.49% in the **Y** block.

As has been explained in the theory section (2.3.2), PLS-DM implies the optimization of several parameters: the number of nearest neighbors' *k,* from 1 to 6; pre-processing type; smoothing coefficient α of the potential function, from 0.3 to 0.8; and the number of LV, from 1 to 10. The optimization step was applied in the training set and, as a result, a matrix of sensitivity, specificity and efficiency values (data not shown) was obtained for all studied values of these parameters. The optimal combination of these results was evaluated considering the highest efficiency and an odd number of *k* nearest neighbors. Even *k* values can lead to ambiguous classifications, which is the reason why odd numbers are preferred. The optimal parameter values were set as k = 3, mean-center preprocessing, α = 0.6 and LV = 4.

The classification results for these three methods in terms of sensitivity and specificity, according to the one-class strategy, are summarized in **Table 2**. PLS-DA presented the best predictions, since both the sensitivity and the specificity of the training and the test set was 100%. Regarding the results of both SIMCA and PLS-DM, they cannot be considered satisfactory, especially in relation to specificity, since a significant percentage of adulterated samples were predicted as not adulterated (25% for SIMCA and 32% for PLS-DM).

When the two strategies are compared, it can be stated that the multiclass classification would be preferable, because it provides more specific information about the adulterations. Many samples in the one-class strategy were erroneously assigned, and in the multi-class were considered inconclusive; therefore, a confirmatory analysis is required.

Regarding the comparison among the three classification methods, PLS-DA, SIMCA and PLS-DM, the best performance was clearly provided by the discriminant PLS-DA model. This superior performance of discriminant over class-modeling methods is consistent with observations in the chemometric literature (Bylesjö et al, 2006). Class-modeling methods, such as SIMCA, search for data directions of the highest variance, which might be distinct from the variance direction responsible for the separation of classes. A specific explanation for the worse results provided by class-modeling methods (SIMCA and PLS-DM) in our case is the similarity between UN and CAS samples, which was verified by observing their highly overlapped clusters in the PCA model shown in **Fig. 3**.

4. Conclusions

The combination of ATR-FTIR and classification techniques allowed the detection of adulterations of grape nectars with apple and cashew juices. The entire analytical procedure was very simple and rapid, and it did not require sample pretreatment or the consumption of reagents or solvents. All 126 samples used in this study were obtained from reliable raw ingredients and prepared in strict compliance with Brazilian regulations, except for the intended adulterations.

Three different classification models (SIMCA, PLS-DA and PLS-DM) were developed, and two approaches were considered: the one-class approach with all three methods, and the multiclass approach with SIMCA and PLS-DA. The one class approach is adequate if the main interest is only to detect whether a sample is adulterated, regardless of the type of the adulterant. For the problem under study, PLS-DA provided excellent results, classifying all samples correctly. SIMCA and PLS-DM produced less satisfactory results, with specificity for the test set of 75% and 68%, respectively.

The multiclass approach is the proper choice when the main interest is to investigate the possible presence of known adulterants. It provides more specific information, since in addition to the percentage of samples correctly or incorrectly assigned, information related to the inconclusive assignations is also available. Samples inconclusively classified could be submitted in the sequence to undergo confirmatory analyses. Among the multiclass models, PLS-DA also presented the best performance, with no false-negative predictions, i.e., no adulterated samples were classified as unadulterated. In food fraud analysis, it is essential to avoid false-negative results, since the analyst could declare a sample as unadulterated when it is actually adulterated. For the multiclass approach, the

SIMCA model was not able to differentiate unadulterated samples from samples adulterated with cashew. Nonetheless, the apple class was well characterized by SIMCA.

Finally, we can suggest this type of application as a potential tool to assist the beverage industry and regulatory organisms in the field of food quality control, allowing detection in fruit nectars through direct, fast and reliable screening analyzes. Further research could be the implementation of the developed classification techniques to detect grape nectar samples adulterated with blends of more than one adulterant.

Acknowledgements

The authors acknowledge CAPES for providing the sandwich PhD scholarship (Proc., nº 88881. 132172/2016-01); Prof. Adriana Silva França from the Biofuels Laboratory of UFMG for enabling the use of an ATR-FTIR Spectrophotometer and MSc Andréia H. Suzuki for assistance with this equipment; Giuliano Elias Pereira from EMBRAPA Grape & Wine/Semiarid (Petrolina, Brazil) for providing the Isabel grapes; and Prof. Paolo Oliveri (University of Genova, Italy) for providing the PLS-DM routine and assisting with its use.

5. References

ABIR (2015). Associação Brasileira das Indústrias de Refrigerantes e Bebidas não Alcoólicas. Available in https://abir.org.br/, accessed in March 2018.

Ammari, F., Redjdal, L., & Rutledge, D. N. (2015). Detection of orange juice frauds using front-face fluorescence spectroscopy and independent components analysis. *Food Chemistry, 168*, 211-217.

Asadpoor, M., Ansarin, M., & Nemati, M. (2014). Amino acid profile as a feasible tool for determination of the authenticity of fruit juices. *Advanced Pharmaceutical Bulletin, 4*, 359.

Barker, M., & Rayens, W. (2003). Partial least squares for discrimination. *Journal of Chemometrics, 17*, 166-173.

Bevilacqua, M., Bucci, R., Magrì, A. D., Magrì, A. L., Nescatelli, R., & Marini, F. (2013). Classification and class-modelling. In F. Marini (Ed.), *Data handling in science and technology* (Vol. 28, pp. 171-233). Amsterdam: Elsevier.

Boggia, R., Casolino, M. C., Hysenaj, V., Oliveri, P., & Zunin, P. (2013). A screening method based on UV–Visible spectroscopy and multivariate analysis to assess addition of filler juices and water to pomegranate juices. *Food Chemistry, 140*, 735-741.

Bona, E., Marquetti, I., Link, J. V., Makimori, G. Y. F., Arca, V. C., Lemes, A. L. G., Ferreira, J. M. G., Scholz, M. B. S., Valderrama, P., & Poppi, R. J. (2017). Support vector machines in tandem with infrared spectroscopy for geographical classification of green arabica coffee. *LWT-Food Science and Technology, 76*, 330-336.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Decreto nº 6.871, de 04 de junho de 2009. Regulamenta a Lei n. 8.918, de 14 de julho de 1994. Dispõe sobre a padronização, a classificação, o registro, a inspeção, a produção e a fiscalização de bebidas.

Bureau, S., Ścibisz, I., Le Bourvellec, C., & Renard, C. M. (2012). Effect of sample preparation on the measurement of sugars, organic acids, and polyphenols in apple fruit by mid-infrared spectroscopy. *Journal of Agricultural and Food Chemistry, 60*, 3551-3563.

Bylesjö, M., Rantalainen, M., Cloarec, O., Nicholson, J. K., Holmes, E., & Trygg, J. (2006). OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *Journal of Chemometrics, 20*, 341-351.

Callao, M. P., & Ruisánchez, I. (2018). An overview of multivariate qualitative methods for food fraud detection. *Food Control, 86*, 283-293.

Cuny, M., Vigneau, E., Le Gall, G., Colquhoun, I., Lees, M., & Rutledge, D. (2008). Fruit juice authentication by 1H NMR spectroscopy in combination with different chemometrics tools. *Analytical and Bioanalytical Chemistry, 390*, 419-427.

Dandage, K., Badia-Melis, R., & Ruiz-García, L. (2017). Indian perspective in food traceability: A review. *Food Control*, 71, 217-227.

El Darra, N., Rajha, H. N., Saleh, F., Al-Oweini, R., Maroun, R. G., & Louka, N. (2017). Food fraud detection in commercial pomegranate molasses syrups by UV–VIS spectroscopy, ATR-FTIR spectroscopy and HPLC methods. *Food Control*, 78, 132-137.

EMBRAPA. (2016). Sistema de Produção do Caju. Retrieved from https://www.spo.cnptia.embrapa.br/conteudo?p_p_id=conteudoportlet_WAR_sistemasdeproducaolf6_1ga1ceportlet&p_p_lifecycle=0&p_p_state=normal&p_p_mode=view&p_p_col_id=column-1&p_p_col_count=1&p_r_p_-76293187_sistemaProducaoId=7705&p_r_p_-996514994_topicoId=10320

Esteki, M., Shahsavari, Z., & Simal-Gandara, J. (2018). Use of spectroscopic methods in combination with linear discriminant analysis for authentication of food products. Food Control, 91, 100-112.

Georgouli, K., Del Rincon, J. M., & Koidis, A. (2017). Continuous statistical modelling for rapid detection of adulteration of extra virgin olive oil using mid infrared and Raman spectroscopic data. *Food Chemistry, 217*, 735-742.

Gondim, C., Junqueira, R. G., Souza, S. V. C., Ruisánchez, I., & Callao, M. P. (2017). Detection of several common adulterants in raw milk by MID-infrared spectroscopy and one-class and multi-class multivariate strategies. *Food Chemistry, 230*, 68-75.

He, J., Rodriguez-Saona, L. E., & Giusti, M. M. (2007). Midinfrared spectroscopy for juice authentication rapid differentiation of commercial juices. *Journal of Agricultural and Food Chemistry, 55*, 4443-4452.

Kennard, R. W., & Stone, L. A. (1969). Computer aided design of experiments. *Technometrics, 11*, 137-148.

López, M. I., Callao, M. P., & Ruisánchez, I. (2015). A tutorial on the validation of qualitative methods: From the univariate to the multivariate approach. *Analytica Chimica Acta, 891*, 62-72.

López, M. I., Trullols, E., Callao, M. P., & Ruisánchez, I. (2014). Multivariate screening in food adulteration: Untargeted versus targeted modelling. *Food Chemistry, 147*, 177-181.

MAPA (2003). Secretaria de Defesa Agropecuária. Ministério da Agricultura, Pecuária e Abastecimento. Instrução Normativa No. 12, Brazil.

MAPA (2013). Secretaria de Defesa Agropecuária. Ministério da Agricultura, Pecuária e Abastecimento. Instrução Normativa No. 42, Brazil.

Márquez, C., López, M. I., Ruisánchez, I. & Callao, M. P. (2016). FT-Raman and NIR spectroscopy data fusion strategy for multivariate qualitative analysis of food fraud. *Talanta*, *161*, 80-86.

Martins, A. R., Talhavini, M., Vieira, M. L., Zacca, J. J., & Braga, J. W. B. (2017). Discrimination of whisky brands and counterfeit identification by UV–Vis spectroscopy and multivariate data analysis. *Food Chemistry, 229*, 142-151.

Miaw, C. S. W., Assis, C., Silva, A. R. C. S., Cunha, M. L., Sena, M. M., & de Souza, S. V. C. (2018). Determination of main fruits in adulterated nectars by ATR-FTIR spectroscopy combined with multivariate calibration and variable selection methods. *Food Chemistry*, *254*, 272-280.

Neves, M. F., Trombin, V. G., Lopes, F. F., Kalaki, R., & Milan, P. (2012). *The orange juice business: A Brazilian perspective*: Wageningen Academic Publishers.

Oliveri, P. (2017). Class-modelling in food analytical chemistry: Development, sampling, optimisation and validation issues–A tutorial. *Analytica Chimica Acta*, *982*, 9-19.

Oliveri, P., López, M. I., Casolino, M. C., Ruisánchez, I., Callao, M. P., Medini, L., & Lanteri, S. (2014). Partial least squares density modeling (PLS-DM)–A new class-modeling strategy applied to the authentication of olives in brine by near-infrared spectroscopy. *Analytica Chimica Acta, 851*, 30-36.

Paltrinieri, G., & Figuerola, F. (1998). Small-scale processing of native and introduced Amazonian fruits and vegetables. Technical manual. Santiago (Chile): Food and Agriculture Organization (FAO).

Pulido, A., Ruisanchez, I., Boqué, R., & Rius, X. F. (2003). Uncertainty of results in routine qualitative analysis. *TrAC Trends in Analytical Chemistry, 22*, 647-654.

Rinnan, Å., Berg, F. v. d., & Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry, 28*, 1201-1222.

Ríos-Reina, R., Callejón, R. M., Oliver-Pozo, C., Amigo, J. M., & García-González, D. L. (2017). ATR-FTIR as a potential tool for controlling high quality vinegar categories. *Food Control*, 78, 230-237

Rius, A., Callao, M. P., & Rius, F. X. (1997). Multivariate statistical process control applied to sulfate determination by sequential injection analysis. *Analyst*, *122*, 737-741.

Rodionova, O. Y., Oliveri, P., & Pomerantsev, A. L. (2016). Rigorous and compliant approaches to one-class classification. *Chemometrics and Intelligent Laboratory Systems, 159*, 89-96.

Sen, I., & Tokatli, F. (2016). Differentiation of wines with the use of combined data of UV–visible spectra and color characteristics. *Journal of Food Composition and Analysis, 45*, 101-107.

Shah, N., Cynkar, W., Smith, P., & Cozzolino, D. (2010). Use of attenuated total reflectance midinfrared for rapid and real-time analysis of compositional parameters in commercial white grape juice. *Journal of Agricultural and Food Chemistry, 58*, 3279-3283.

Shen, F., Wu, Q., Su, A., TAng, P., ShAo, X., & Liu, B. (2016). Detection of Adulteration in Freshly Squeezed Orange Juice by Electronic Nose and Infrared Spectroscopy. *Czech Journal of Food Science, 34*, 224-232.

Singhal, R. S., Kulkarni, P. R., & Rege, D. V. (1997). Chapter 3 - Fruit and Vegetable Products. In *Handbook of Indices of Food Quality and Authenticity* (pp. 77-130). Oxford: Woodhead Publishing.

Szymańska, E., Gerretzen, J., Engel, J., Geurts, B., Blanchet, L., & Buydens, L. M. (2015). Chemometrics and qualitative analysis have a vibrant relationship. *TrAC Trends in Analytical Chemistry, 69*, 34-51.

Tressler, D. K., & Joslyn, M. A. (1961). *Fruit and vegetable juice processing technology*. Westport: AVI Publishing Company.

Vardin, H., Tay, A., Ozen, B., & Mauer, L. (2008). Authentication of pomegranate juice concentrate using FTIR spectroscopy and chemometrics. *Food Chemistry, 108*, 742-748.

Xu, L., Fu, H. Y., Yang, T. M., Li, H. D., Cai, C. B., Chen, L. J., & She, Y. B. (2016). Enhanced specificity for detection of frauds by fusion of multi-class and one-class partial least squares discriminant analysis: geographical origins of Chinese shiitake mushroom. *Food Analytical Methods*, *9*, 451-458.

Figure 1. Scheme of grape nectar samples formulation.

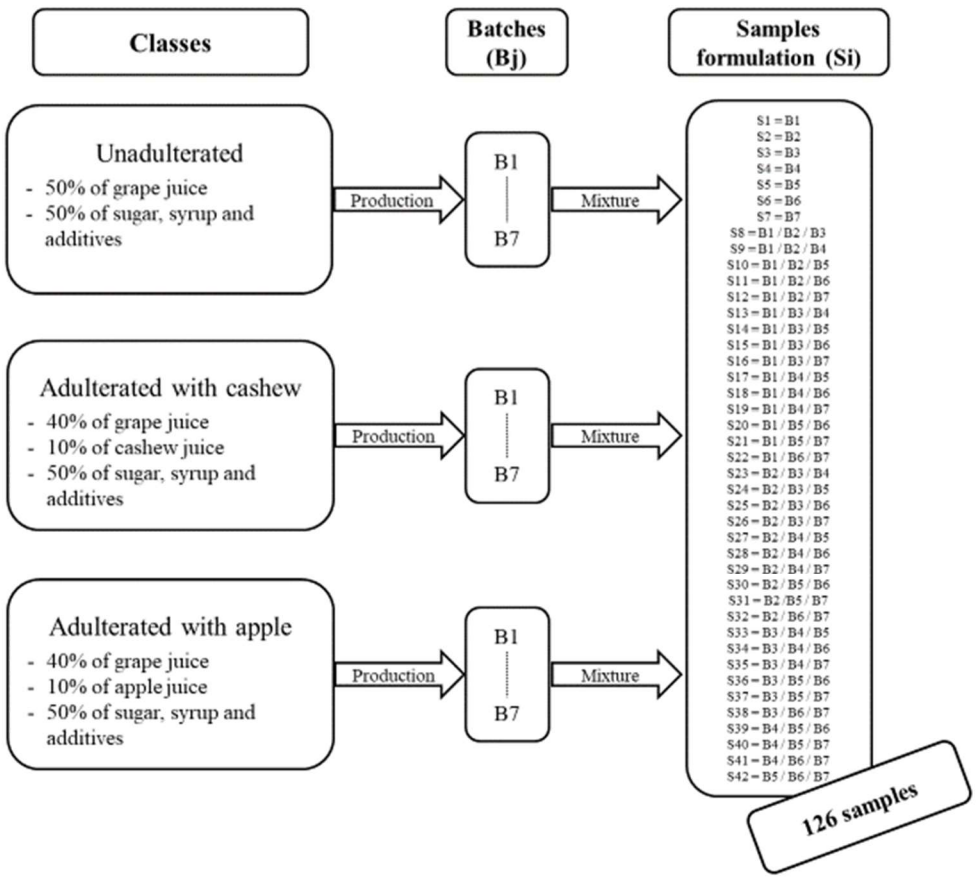


Figure 2. Mean preprocessed spectra of unadulterated class (dashed line), adulterated with cashew class (solid line) and adulterated with apple class (dashed-dot line).

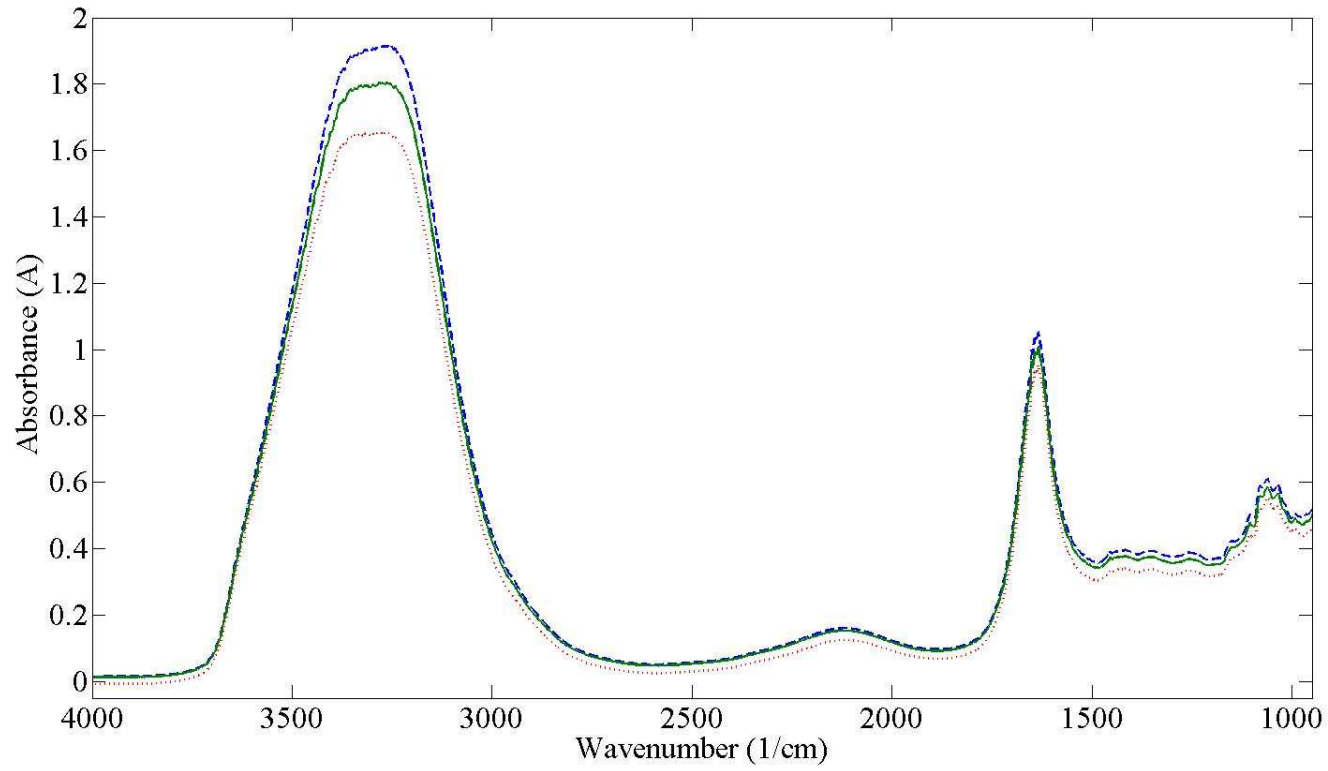


Figure 3. Scores of PC1 *versus* PC2 of unadulterated (down triangles), adulterated with cashew (circles) and adulterated with apple (squares) grape nectar samples.

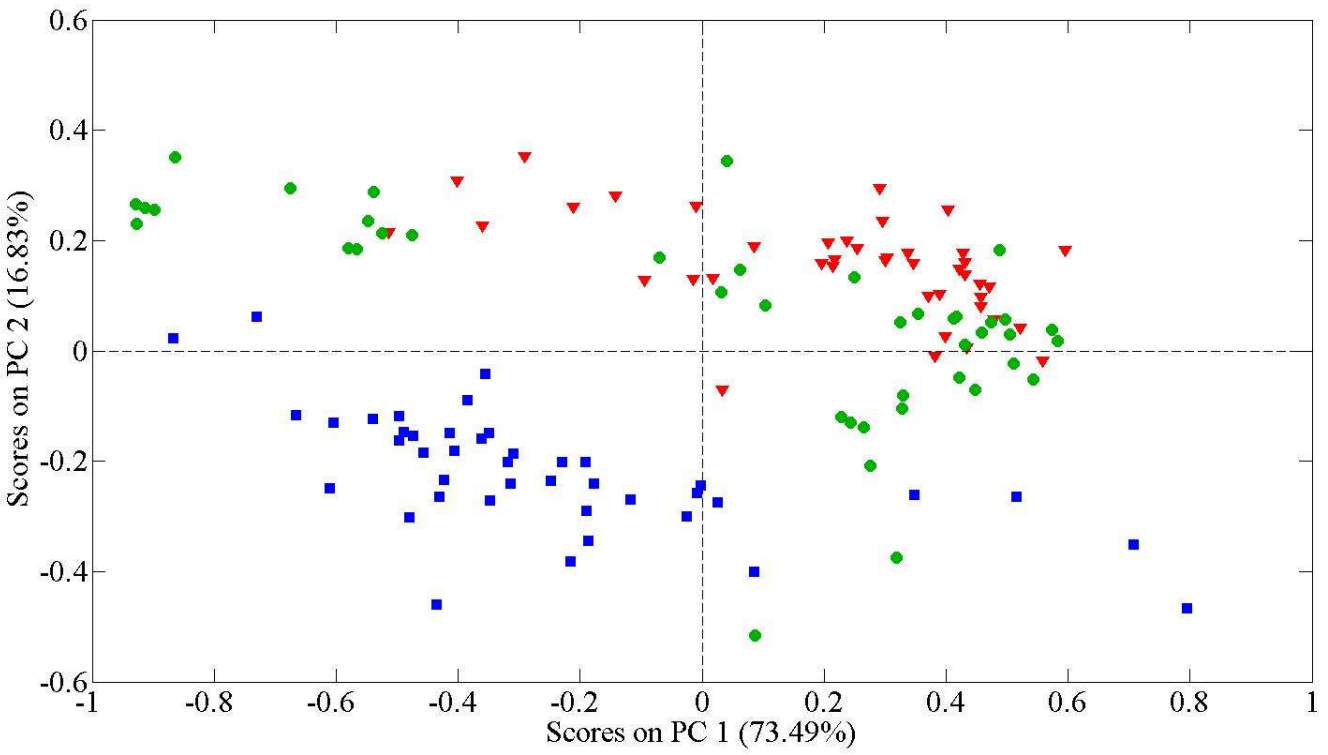
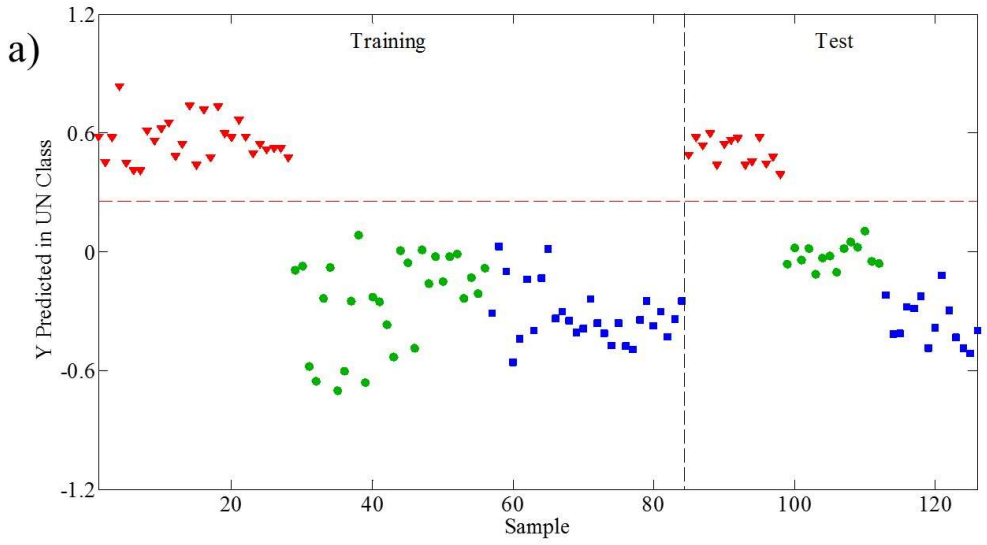
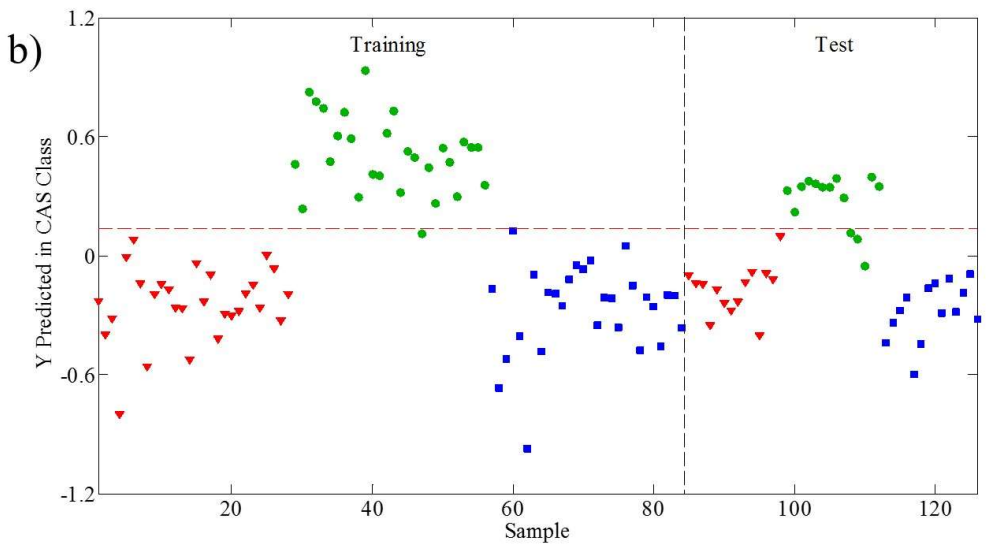
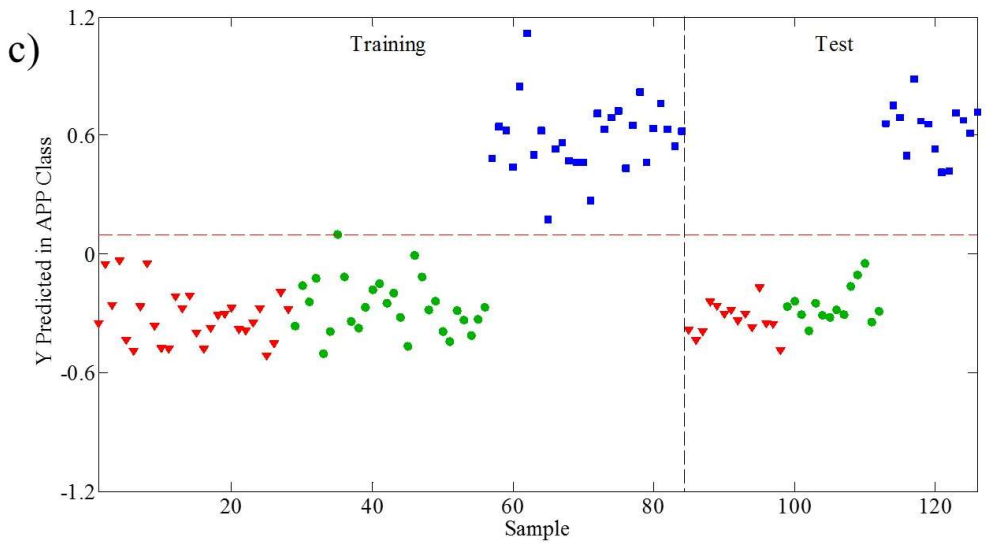


Figure 4. PLS-DA predictions for each class: a) unadulterated (UN), b) adulterated with cashew (CAS) and c) adulterated with apple (APP). Horizontal dashed lines indicate the threshold class and the vertical dashed lines separate training and test samples. Samples symbols: down triangles for unadulterated, circles for adulterated with cashew and squares for adulterated with apple.


a)
Training
Test
Y Predicted in UN Class
1.2
0.6
0
-0.6
-1.2
20
40
60
80
100
120
Sample


b)
Training
Test
Y Predicted in CAS Class
1.2
0.6
0
-0.6
-1.2
20
40
60
80
100
120
Sample


c)
Training
Test
Y Predicted in APP Class
1.2
0.6
0
-0.6
-1.2
20
40
60
80
100
120
Sample

419

Table 1. SIMCA and PLS-DA multi-class predictions of samples from the unadulterated class (UN), the adulterated with cashew class (CAS) and the adulterated with apple class (APP) for training and test set. nºS: number of samples; NA: not assigned; MA: multiple assignments; IR: inconclusive ratio

				Classified as					
Method	**Set**	**Class**	**nºS**	**UN**	**CAS**	**APP**	**NA**	**MA**	**IR (%)**
SIMCA	**Training**	**UN**	28	26	28	0	0	26	92.86
		CAS	28	5	21	0	7	5	42.86
		APP	28	0	0	22	6	0	21.43
	Test	**UN**	14	14	14	0	0	14	50.00
		CAS	14	7	14	0	0	7	25.00
		APP	14	0	0	14	0	0	0.00
PLS-DA	**Training**	**UN**	28	28	0	0	0	0	0.00
		CAS	28	0	27	1	1	1	7.14
		APP	28	0	0	28	0	0	0.00
	Test	**UN**	14	14	0	0	0	0	0.00
		CAS	14	0	11	0	3	0	21.43
		APP	14	0	0	14	0	0	0.00

Table 2. Sensitivities and specificities for the one-class strategy

Method	**Set**	**Sensitivity (%)**	**Specificity (%)**
SIMCA	Training	93	91
	Test	100	75
PLS-DA	Training	100	100
	Test	100	100
PLS-DM	Training	82	91
	Test	100	68