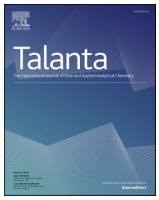
Talanta 253 (2023) 123916

Contents lists available at ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta



One-class model with two decision thresholds for the rapid detection of cashew nuts adulteration by other nuts

Glòria Rovira [a], Carolina Sheng Whei Miaw [b], Mário Lúcio Campos Martins [b], Marcelo Martins Sena [c,d], Scheilla Vitorino Carvalho de Souza [b], M. Pilar Callao [a,*], Itziar Ruisánchez [a]

[a] Chemometrics, Qualimetric and Nanosensors Group, Department of Analytical and Organic Chemistry, Rovira I Virgili University, Marcel·lí Domingo s/n, 43007 Tarragona, Spain
[b] 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
[c] Chemistry Department, 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
[d] Instituto Nacional de Ciência e Tecnologia em Bioanalítica (INCT-Bio), Campinas, SP, 13083-970, Brazil

ARTICLE INFO

Keywords:
Nut adulteration
Multivariate screening
Soft independent modelling of class analogy
Portability
Uncertainty intervals
Decision thresholds

ABSTRACT

A green screening method to determine cashew nut adulteration with Brazilian nut, pecan nut, macadamia nut and peanut was proposed. The method was based on the development of a one-class soft independent modelling of class analogy (SIMCA) model for non-adulterated cashew nuts using near-infrared (NIR) spectra obtained with portable equipment. Once the model is established, the assignment of unknown samples depends on the threshold established for the authentic class, which is a key aspect in any screening approach. The authors propose innovatively to define two thresholds: lower model distance limit and upper model distance limit. Samples with distances below the lower threshold are assigned as non-adulterated with a 100% probability; samples with distance values greater than the upper threshold are assigned as adulterated with a 100% probability; and samples with distances within these two thresholds will be considered uncertain and should be submitted to a confirmatory analysis. Thus, the possibility of error in the sample assignment significantly decreases. In the present study, when just one threshold was defined, values greater than 95% for the optimized threshold were obtained for both selectivity and specificity. When two class thresholds were defined, the percentage of samples with uncertain assignment changes according to the adulterant considered, highlighting the case of peanuts, in which 0% of uncertain samples was obtained. Considering all adulterants, the number of samples that were submitted to a confirmatory analysis was quite low, 5 of 224 adulterated samples and 3 of 56 non-adulterated samples.

1. Introduction

Concerns about food safety from most stakeholders, such as consumers, producers and regulators grow every year. It is well known to these stakeholders that food scandals continue to occur despite national and international regulations [1]. Therefore, there is an increasing demand for developing and improving innovative analytical methods.

The vast potential for food fraud hinders its detection. Often, food fraud detection requires the use of expensive and sophisticated equipment. Currently, alternative and green screening methods, which use less expensive instrumentation, do not consume reagents, and minimize the number of steps and sample manipulation, are gaining relevance [2–9]. Screening methods are very convenient for deciding whether a sample is adulterated or not adulterated (yes/no) and if it is necessary to submit the suspicious samples to a confirmatory analysis.

Due to the high complexity of food, it is difficult to develop screening methodologies based on a sample single property or signal. Therefore, the application of spectroscopic techniques together with multivariate classification models is an alternative that gains importance. Some examples for detecting nut fraud due to the addition of adulterants such as

* Corresponding author.
E-mail address: mariapilar.callao@urv.cat (M.P. Callao).

https://doi.org/10.1016/j.talanta.2022.123916
Received 28 June 2022; Received in revised form 1 September 2022; Accepted 3 September 2022
Available online 14 September 2022

almond [10,11], pistachio [12] and hazelnut [9,13,14] have been referenced.

For screening purposes, obtaining multivariate signals should be a process that requires minimal sample treatment prior to measurement. Methods based on vibrational techniques offer these advantages. In particular, handheld near-infrared (NIR) spectrophotometers present the advantage of portability, allowing in situ measurements, time savings for obtaining analytical results, and increasing the analytical frequency of the method. On the other hand, portable NIR equipment present some disadvantages in comparison to benchtop instruments, such as lower resolution and signal-to-noise ratio, and reduced wavelength range. In general, the optical performance of this type of spectrophotometer has still not reached the level of mainstream commercial instruments [15].

Recently, portable NIR spectroscopy has been successfully applied to food analysis with different aims, such as the varietal discrimination of walnuts [16], the determination of total antioxidant capacity in gluten free grains [17], the prediction of stable isotopes and fatty acids in subcutaneous fat of Iberian pigs [18], and the on-line monitoring of quality parameters in intact olives for determining optimal harvesting times [19].

Regarding chemometric classification tools, one-class modelling has been considered a better option than the most commonly employed discriminant models for food authenticity problems. One-class modelling methods build a class that is focused on authentic/non-fraudulent samples. This is a proper approach since it aims to detect whether new samples belong to the authentic class regardless of the type of fraud being investigated. This strategy has several benefits in relation to multiclass approach, in which, in addition to the non-adulterated class, one or more adulterant classes has also to be modelled, since it is impossible in practice to cover all possible adulterants in a representative way [20].

As with any analytical method, multivariate screening methods should be validated prior to their implementation in the routine of quality control laboratories. This process involves establishing performance parameters; however, the validation of multivariate screening methods is still not fully established. Efforts to develop a harmonized validation procedure have been performed in recent years [21–24]. The main figures of merit, sensitivity and specificity, evaluated from true and false assignments of the samples that belong to the modelled class and those that do not belong to the modelled class, respectively, are currently accepted by the scientific community. Many studies have been carried out aiming to optimize the models and to obtain better performance parameters. Variable selection [25,26] and data fusion [27–29] are notable research topics in this area.

Whether a sample belongs to the modelled class (fits the model) depends on the threshold (limit) established for the class, which in the one-class modelling approach is a distance. As a result, samples having a sample distance lower than the class threshold are the samples that fit the model. Therefore, establishing an optimum class threshold is a key aspect in any screening model. Recently, published articles have discussed alternatives to the establishment of an optimal class threshold [13,14,30,31].

For an analytical method to be considered validated, high sensitivity and specificity values must be obtained, but unless 100% sensitivity and specificity are obtained, there are certain probabilities of error, respectively, for samples belonging to the model class and not belonging to the model class, in the final assignment.

The objective of this article is to propose a new alternative to one-class modelling that estimates two class thresholds (lower- and upper-class model distances) instead of only one class threshold, thus providing three distance intervals: 1) Samples with distance values below the lower-class threshold will fit the model, so they will be unambiguously assigned as authentic/non-fraudulent; 2) Samples with distance values greater than the upper-class threshold will not fit the model, so they will be assigned as fraudulent/adulterated; and 3) samples having distance values between the lower- and the upper-class threshold will be considered inconclusive and, if necessary, will be submitted to a confirmatory analysis.

Attempts have also been made to define an uncertainty region in multivariate qualitative analysis [21,32–34], but these approaches have focused on experimenting at concentration levels above and below the concentration of the cut-off value [34], which is the concentration limit established or specified by the end user or legislation. Nevertheless, even when defining the uncertainty region, most of the published examples do not assure both a sensitivity of 100% and a specificity of 100%. As a case study, the adulteration of cashew nut samples has been considered, and Brazilian nut, macadamia nut, pecan nut and peanut have been studied as possible adulterants with concentrations within the interval between 0.1 and 10.0%. Adulterated and non-adulterated samples were measured by portable NIR spectroscopy (NIRS) and a one class soft independent modelling of class analogy (SIMCA) model of authentic/non-adulterated samples was established.

2. Materials and methods

2.1. Samples

Commercial batches of cashew nut, Brazilian nut (BN), macadamia nut (M), pecan nut (PN) and peanut (P) were obtained from different certified producers. All batches of each nut, individually, were crushed in a sample processor (Arno Magiclean WWBC Blender), homogenized, sieved using a calibrated sieve to size 40 mesh, packed in polyethylene packaging, sealed and kept at room temperature (25 ± 3 °C) until preparation of the non-adulterated and adulterated samples.

Seven formulated batches in eight variations were prepared to form the non-adulterated samples of cashew nut, resulting in a total of 56 samples. Adulterated samples were prepared from each batch of adulterants (Brazilian nut, macadamia nut, pecan nut and peanut). These samples were added in different quantities to the seven formulated batches of non-adulterated samples, to obtain 8 levels of adulteration (10.0; 5.0; 2.5; 1.3; 0.6; 0.3; 0.2 and 0.1%). The total number of adulterated samples was 224 (56 for each adulterant). The non-adulterated samples were divided into 42 samples of training and 14 samples of test set using Kennard-Stone algorithm [35]. The objective of this algorithm was to select the most diverse samples for the training set. This selection should be representative (samples homogeneously distributed throughout the whole composition range) and reproducible, based on a systematic criterion.

2.2. Instrumental measurements

NIR analysis was performed using portable MicroNIR® 1700 equipment from Viavi Solutions (San Jose, CA, USA). A homogenized sample portion was placed on a Petri dish (diameter of 3.5 cm x height of 1.2 cm) until the dish was covered up to its maximum height of 1.2 cm. The plate was placed on the MicroNIR®, and for each sample, a reading was carried out with 20 scans at a resolution of 6.25 nm, providing stable and smooth spectra. NIR spectra were recorded in the wavelength range from 908 to 1676 nm. All 280 sample spectra were recorded in random order on the same day.

NIR analysis was performed using portable MicroNIR® 1700 equipment from Viavi Solutions (San Jose, CA, USA). A homogenized sample portion was placed on a Petri dish (diameter of 3.5 cm x height of 1.2 cm) until the dish was covered up to its maximum height of 1.2 cm. The plate was placed on the MicroNIR®, and for each sample, once reading was carried out with 20 scans at a resolution of 6.25 nm, providing stable and smooth spectra. NIR spectra were recorded in the wavelength range from 908 to 1676 nm. All 280 sample spectra were randomly recorded in random order on the same day.

2.3. Software

The recorded data were processed, and models were built by using MATLAB software, version 8.0.0.783 – R2012b (Natick, MA, USA) and PLS Toolbox 7.0.2 (Eigenvector Research Inc., Wenatchee, WA, USA).

2.4. Data processing

Principal component analysis (PCA) should be performed for a preliminary exploratory analysis of any dataset, even when the final aim is to build a supervised classification model using a class modelling method, such as soft independent modelling of class analogy (SIMCA), for predictive purposes. There is an extensive bibliography that describes theoretical and practical aspects of both PCA and SIMCA. Without being exhaustive in the references, a recent review can be consulted, which provides multiple references [20].

SIMCA is a class modelling method that assumes the main systematic variability characterizing the samples of a category, as retained by a principal component model of opportune dimensionality, and builds the model with training samples of that class.

SIMCA assignments are obtained considering the model distance value from sample "*i*", Eq. (1).

$$d_{r,i} = \sqrt{\left(Q_{r,i}\right)^2 + \left(T^2_{r,i}\right)^2} \quad \text{(Eq. 1)}$$

where $T^2_{r,i}$ and $Q_{r,i}$are the reduced statistics of *Hotelling's* T^2 and *Q, respectively,* of a sample; "*i*" and "*r*" denote for reduced values, which comprise the ratio between the statistics of sample "*i*" and the corresponding statistical class limit (T^2_{lim} and Q_{lim}) at a significance level of significance [36].

Up to now, just one class threshold is set to decide whether a sample fits the model and several criteria have been applied to determining the threshold: 1, $\sqrt{2}$ and a value obtained through experimentation applying receiver operating characteristic (ROC) curves [31]. To evaluate the quality of the classification models the main performance parameters, such as sensitivity, specificity and efficiency, were considered [21,23,24]. These parameters are calculated from the four well-known possibilities of sample model assignment: true positive (TP), false positive (FP), true negative (TN) and false negative (FN). Since TP, FP, TN and FN values depend on the value considered as the class threshold distance, differences in the quality performance parameters can be obtained.

In this work it is proposed to set two class thresholds, upper threshold (d_{upper_th}) and lower threshold (d_{low_th}). d_{upper_th} corresponds to the maximum sample distance value ($d_{r,i}$, Eq. (1)) obtained in the prediction of the non-adulterated samples. Similarly, d_{low_th} corresponds to the minimum $d_{r,i}$ (Eq. (1)) obtained in the prediction of the adulterated samples. As a result, for a given class model, three types of sample assignations can occur: 1) If $d_{r,i}$, $< d_{low\_th}$, the sample will be assigned as non-adulterated (compliant sample) with a 100% probability. 2) If $d_{r,i}$, $> d_{upper\_th}$, it will be assigned as adulterated (non-compliant sample) with a 100% probability. 3) If $d_{r,i}$ falls between the two thresholds, the sample falls into the uncertainty region and should undergo a confirmatory analysis.

3. Results and discussion

Fig. 1 shows the average raw NIR spectra for the non-adulterated cashew nut samples and for adulterated samples with Brazilian nuts (BN), pecan nuts (PN), macadamia nuts (M) and peanuts (P). The largest NIR bands appear between approximately 1170-1300 nm and 1400–1500 nm. The first band can be assigned to the second overtone of the C–H stretching, while the second band can be assigned to the first overtone of the O–H stretching. In particular, the spectral band placed around 1200 nm has been reported as a discriminant of nuts in relation to other food materials (wheat, milk, and cocoa) [37]. The smaller band between 1380 and 1410 nm can be assigned to the combination band of C–H vibrations [38].

Before establishing the classification model, the first derivative followed by mean centering pre-processing was applied to the spectra to eliminate nonlinear baseline deviations caused by multiplicative scatter and to improve the signal-to-noise ratio. Unsupervised PCA was applied before building the supervised classification model, aiming to perform an exploratory analysis with all the samples (non-adulterated and adulterated).

The score plot of the first two PCs for all samples is shown in Fig. 2 (88% of the total explained variance). Regarding the sample distribution along PC 1 (81.4%), no clear grouping was observed in relation to their classes (non-adulterated and adulterated samples with each of the adulterants). This situation is expected because the major chemical components of the samples are identical and, therefore, the most relevant information of the spectra is common to all of them. Regarding PC2 (6.7%) score values, a distinct separation was observed between all non-adulterated samples with positive values and the adulterated samples with mostly negative values. Thus, this lower part of the total variance is responsible for discriminating adulteration. Within the adulterated sample grouping (negative values on PC2), there is no distinct tendency discriminating among the studied adulterants or among the percentages of adulteration.

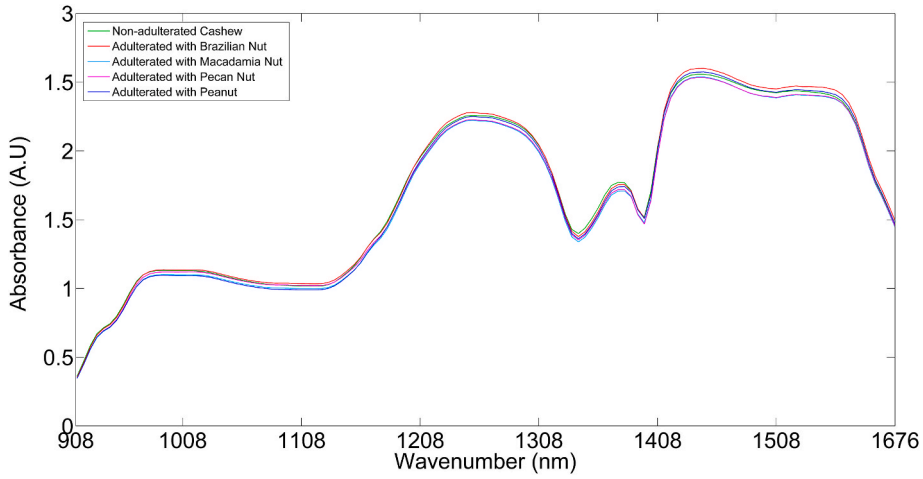


Fig. 1. Average raw NIR spectra. Color code: green for non-adulterated cashew nuts, red for Brazilian nuts, pink for pecan nuts, light blue for macadamia nuts and dark blue for peanuts. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

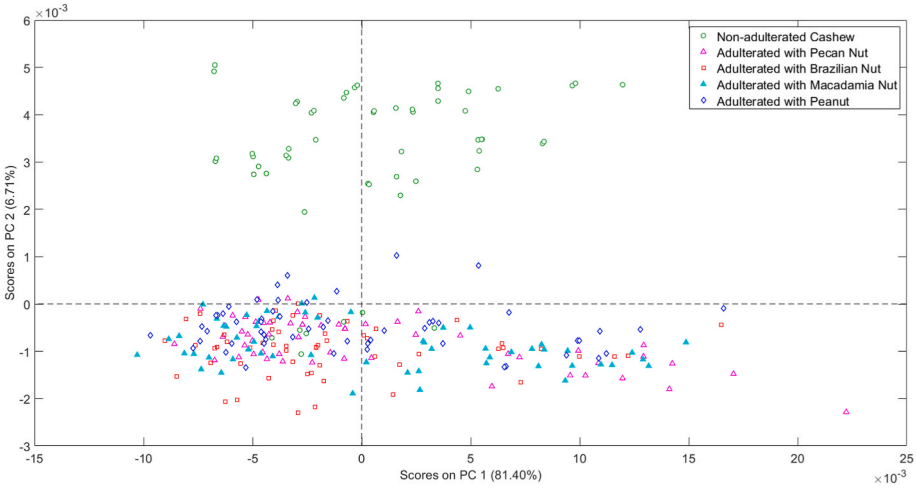


Fig. 2. Score plot of PC1 vs PC2 for NIR data. Color and symbol code: green circle for non-adulterated cashew nuts, red squares for Brazilian nuts, pink triangles for pecan nuts, filled light blue triangles for macadamia nuts and blue diamonds for peanuts. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

In the sequence, a one-class SIMCA model was built for the authentic/non-adulterated cashew nut samples with the 42 training samples selected by the Kennard-Stone algorithm. Five PCs were employed based on the leave-out-one cross-validation classification error (CVCE). For the test set, 14 non-adulterated samples were combined with all adulterated samples containing each of the four studied adulterants (BN, PN, M and P).

It is known that the developed model and therefore its quality parameters, highly depends on the criteria used to obtain the training and test sub-sets. Defining both data sets has been always and continues to be an issue when establishing a multivariate methodology and different algorithms have been proposed together with just do it randomly [39]. Kennard-Stone is a possible algorithm that selects the training samples in such a way that they cover the maximum of the multivariate space defined by the available samples. As a result, K–S emphasizes the training samples and thus could lead to slightly optimistic test set quality parameter results. But it is a well-known and defined algorithm with reliable and reproducible results. If random selection is used, by its very nature, the values of the quality parameters can present more differences, since the possible random sets increases exponentially with the number of samples and may not be very reproducible [40].

In order to evaluate the main quality parameters for the developed model, three criteria to set just one distance threshold (d_{class_th}) have been considered: two criteria fix d_{class_th} at 1.00 and $\sqrt{2}$ and the third criterion fix d_{class_th} at 1.14, calculated by means of a ROC curve. If a sample has a distance value ($d_{r,i}$Eq. 1) lower than d_{class_th}, it is considered to belong to the non-adulterated class (compliant samples); the opposite holds for $d_{r,i}$ greater than d_{class_th}. Fig. 3 shows the reduced distance values calculated from Eq. (1) ($d_{r,i}$) for all analyzed samples (non-adulterated and adulterated), and those for the three considered d_{class_th} values (vertical lines). From these results, the main performance parameters were obtained (Table 1).

As expected, the figure of merit values varied according to the considered threshold. The threshold value obtained through the ROC curve ($d_{class\_th} = 1.14$) is the one that better balances both sensitivity and specificity. When comparing these parameters with each other, as the

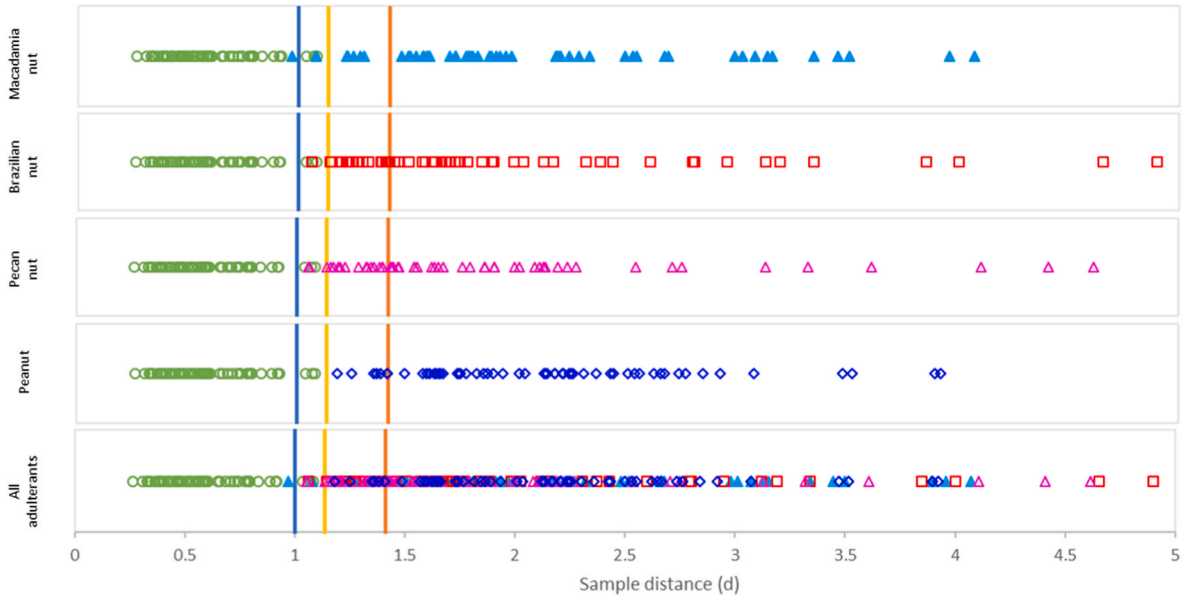


Fig. 3. Distances of all the analyzed samples to the one-class SIMCA model. Color and symbol codes are the same from Fig. 2. Class limit: blue d = 1; yellow d = $\sqrt{2}$; orange d = 1.08, optimized by ROC curves. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 1
Figures of merit for one-class SIMCA models considering differents class limits. NA: non-adulterated cashew nuts; BN: Brazilian nuts; PN: Pecan nuts; M: Macadamia nuts; P: Peanuts; and BN/PN/M/P: all adulterants.

			NA	BN	PN	M	P	BN/PN/M/P
Distance class threshold (d_{class_th})	**d < 1**	**Sensitivity training**	95.24%					
		Sensitivity test	92.86%					
		Specificity		100.00%	100.00%	98.21%	100.00%	99.55%
		Efficiency		98.57%	98.57%	97.14%	98.57%	99.16%
	$d<\sqrt{2}$	**Sensitivity training**	100.00%					
		Sensitivity test	100.00%					
		Specificity		69.64%	71.43%	85.71%	87.5%	78.57%
		Efficiency		74.29%	75.71%	87.14%	88.57%	79.41%
	d(ROC)<1.1359	**Sensitivity training**	100.00%					
		Sensitivity test	100.00%					
		Specificity		98.21%	94.64%	96.43%	100.00%	97.32%
		Efficiency		98.57%	95.71%	97.14%	100.00%	97.49%

threshold increases, sensitivity improves and specificity worsens. Therefore, the choice of the threshold should be based on the practical interest in minimizing the percentage of error associated with the assignment of adulterated or authentic/non-adulterated samples.

Even considering that quite good classification results have been initially obtained, there is still place to improve the model, from the point of view of its application as a screening method. The goal is to identify with certainty whether a sample is compliant or non-compliant, and in the case of a non-conclusive prediction, submit it to a confirmatory analysis. With this idea, the strategy proposed in this article implies defining two thresholds (low and upper).

Fig. 4 shows the results obtained with two thresholds considering the adulterants both individually and concurrently. In the problem under study, the upper threshold (d_{upper_th}) value was equal to 1.08 (green vertical lines, Fig. 4). Since the upper threshold was set from the maximum $d_{r,i}$ of the non-adulterated samples, it is the same regardless of the adulterant that is considered. The lower threshold (d_{low_th}) can be calculated independently for each adulterant considered, resulting in four d_{low_th} values at 0.97, 1.05, 1.06 and 1.18 for macadamia nuts, pecan nuts, Brazilian nuts, and peanuts, respectively. That value changed for each adulterant (one color line for each adulterant, Fig. 4), providing a different uncertainty range for each adulterant (region between both lines). When the differentiation between the four adulterants was not considered, a single d_{low_th} equal to 0.97 was obtained. Note that the lower threshold considering all adulterants coincides with the lowest value individually considering each adulterant and, in that case, this was the threshold for macadamia nut adulteration.

Below the low threshold, there are only non-adulterated samples, and above the upper threshold there are only adulterated samples. Therefore, there is no ambiguity in the assignments. Between the thresholds, there are both adulterated samples and non-adulterated samples, whose assignments were inconclusive.

Table 2 shows the distance values that define the uncertainty region

Table 2
Uncertainty intervals and percentage of samples of uncertain assignment. NA: non adulterated cashew nuts; BN: Brazilian nuts; PN: Pecan nuts; M: Macadamia nuts; P: Peanuts and BN/PN/M/P: all adulterants.

Uncertainty interval (d)		Uncertain assignment (%)	
		Adulterated	Non-adulterated
PN	1.05–1.08	3.6	3.6
BN	1.06–1.08	1.8	3.6
M	0.97–1.08	3.6	5.4
P	–	0.0	0.0
PN/BN/M/P	0.97–1.08	2.2	5.4

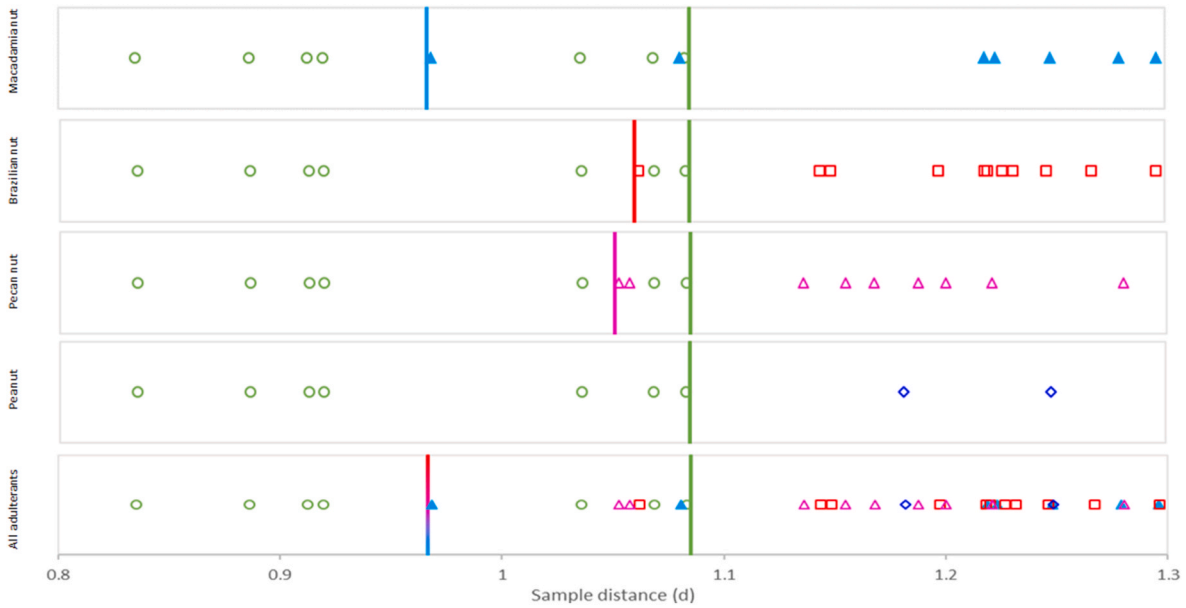


Fig. 4. One-class SIMCA model based on two thresholds, configuring the distances that define uncertainty intervals. Color and symbol codes are the same from Fig. 2. The abscissa axis is shown between 0.75 and 1.25 sample distance aiming to highlight the samples in between and around the uncertainty region. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

and the percentage of samples that fall into it for each adulterant. The number of samples from which the uncertainty assignment percentages were calculated were 56 non-adulterated samples, 56 samples for each individual adulterant (PN, BN, M and P) and 224 when considering all adulterated samples, regardless of the adulterant (PN + BN + M + P). The percentage of samples that should be submitted to a confirmatory analysis is quite low (Table 2): 2.2%, corresponding to 5 out of 224 adulterated samples, and 5.4%, corresponding to 3 out of 56 non-adulterated samples. A comprehensive analysis of these 5 adulterated samples indicates that 3 of them (1 sample from Brazilian nut and 2 samples from pecan nuts) correspond to adulterated samples at very low levels (0.15% and 0.6%), while the two remaining samples have been adulterated with macadamia nuts at levels higher than 1.0 (1.3% and 2.5%).

Particular attention should be given to adulteration with peanuts since the minimum predicted distance of the adulterated samples ($d_{r,i}$ = 1.18) was higher than the upper threshold (1.08). In these cases, there is no uncertainty interval. Moreover, two thresholds are not necessary since with just one threshold, both the sensitivity and specificity were 100%, that is, only the case in which the class distance threshold was set by means of the ROC curve (d_{class_th}, Table 1).

4. Conclusions

A multivariate screening method that jointly uses portable NIR spectroscopy with one-class SIMCA models was developed to determine cashew nut adulteration with Brazilian nut, pecan nut, macadamia nut and peanut. When just one class threshold was estimated for the one-class SIMCA model, values greater than 95% for the optimized threshold were obtained for both selectivity and specificity. The establishment of two threshold values (low and upper limits) generates an uncertainty region. Outside this interval, the samples can be assigned as authentic/non-adulterated (distances less than the lower threshold) or non-authentic/adulterated (distances greater than the upper threshold) with a 100% probability of success. Samples predicted within the interval between these two thresholds are assigned to the uncertainty region and should undergo further confirmatory analysis.

When two class thresholds were defined, it was possible to detect with certainty if a sample was compliant or non-compliant (100% for both sensitivity and specificity) by defining an uncertainty region. In this region, the percentage of samples within the uncertain assignment changed according to the adulterant that was considered. In all cases, the percentage of samples that should be submitted to a confirmatory analysis was quite low, including both non-adulterated samples and adulterated samples, even when they were simultaneously considered regardless of the adulterant (PN + BN + M + P).

The developed analytical methodology is simple, rapid, green (neither consumes reagents or solvents nor generates chemical waste) and non-destructive, and thus is considered suitable for screening analysis. Given the need for constant improvements in food fraud detection, this study represents a contribution to food and analytical scientific communities that can easily be extended to other types of food fraud, or even fraud involving other types of products/matrices.

Author statement

Gloria Rovira: Formal analysis, Chemometric analysis, Investigation, Methodology, writing, Carolina Sheng Whei Miaw: Methodology, Investigation, Formal analysis, Writing – review & editing, Mário Lúcio Campos Martins: Methodology, Formal analysis, Marcelo Martins Sena: Resources, Supervision, Writing – review & editing, Funding acquisition, Scheilla Vitorino Carvalho de Souza: Conceptualization, Resources, Supervision, Writing – review & editing, Funding acquisition. Itziar Ruisanchez: Conceptualization, Investigation, Methodology, Supervision, Validation, writing, reviewing and Editing, Funding acquisition, M. Pilar Callao: Conceptualization, Investigation, Methodology, Supervision, Validation, writing, reviewing and Editing, Funding acquisition

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This study was supported by the research program “Program of research activity (2020PMF-PIPF)” at the Rovira i Virgili University, Tarragona, Spain. The acquisition of a portable NIR spectrometer (Viavi MicroNIR® 1700) was supported by the Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) through project APQ03457-16.

References

[1] A.S. Tsagkaris, J.L.D. Nelis, G.M.S. Ross, S. Jafari, J. Guercetti, K. Kopper, Y. Zhao, K. Rafferty, J.P. Salvador, D. Migliorelli, G.I.J. Salentijn, K. Campbell, M.P. Marco, C.T. Elliot, M.W.F. Nielen, J. Pulkrabova, J. Hajslova, Critical assessment of recent trends related to screening and confirmatory analytical methods for selected food contaminants and allergens, TrAC, Trends Anal. Chem. 121 (2019), 115688, https://doi.org/10.1016/j.trac.2019.115688.

[2] C.S.M. Miaw, M.L.C. Martins, M.M. Sena, S.V. C de Souza, Screening method for the detection of other allergenic nuts in cashew nuts using chemometrics and a portable near-infrared spectrophotometer, Food Anal. Methods 15 (2022) 1074–1084, https://doi.org/10.1007/s12161-021-02184-0.

[3] A. de Girolamo, M.C. Arroyo, V. Lippolis, S. Cervellieri, M. Cortese, M. Pascale, A. F. Logrieco, C. von Holst, A simple design for the validation of a FT-NIR screening method: application to the detection of durum wheat pasta adulteration, Food Chem. 333 (2020), 127449, https://doi.org/10.1016/j.foodchem.2020.127449.

[4] M. Spiteri, E. Jamin, F. Thomas, A. Rebours, M. Lees, K.M. Rogers, D.N. Rutledge, Fast and global authenticity screening of honey using ^{1}H-NMR profiling, Food Chem. 189 (2015) 60–66, https://doi.org/10.1016/j.foodchem.2014.11.099.

[5] B. Quintanilla-Casas, J. Bustamante, F. Guardiola, D.L. García-González, S. Barbieri, A. Bendini, T.G. Toschi, S. Vichi, A. Tres, Virgin olive oil volatile fingerprint and chemometrics: towards an instrumental screening tool to grade the sensory quality, LWT–Food Sci. Technol. 121 (2020), 108936, https://doi.org/10.1016/j.lwt.2019.108936.

[6] C.S. Gondim, R.G. Junqueira, S.V. C de Souza, I. Ruisánchez, M.P. Callao, Detection of several common adulterants in raw milk by MID-infrared spectroscopy and one-class and multi-class multivariate strategies, Food Chem. 230 (2017) 68–75, https://doi.org/10.1016/j.foodchem.2017.03.022.

[7] L.S.A. Pereira, F.L.C. Lisboa, J.C. Neto, F.N. Valladão, M.M. Sena, Screening method for rapid classification of psychoactive substances in illicit tablets using mid infrared spectroscopy and PLS-DA, Forensic Sci. Int. 288 (2018) 227–235, https://doi.org/10.1016/j.forsciint.2018.05.001.

[8] B. Quintanilla-Casas, G. Strocchi, J. Bustamante, B. Torres-Cobos, F. Guardiola, W. Moreda, J.M. Martínez-Rivas, E. Valli, A. Bendini, T.G. Toschi, A. Tres, S. Vichi, Large-scale evaluation of shotgun triacylglycerol profiling for the fast detection of olive oil adulteration, Food Control 123 (2021), 107851, https://doi.org/10.1016/j.foodcont.2020.107851.

[9] M.I. López, N. Colomer, I. Ruisánchez, M.P. Callao, Validation of multivariate screening methodology. Case study: detection of food fraud, Anal. Chim. Acta 827 (2014) 28–33, https://doi.org/10.1016/j.aca.2014.04.019.

[10] M. Esteki, Y. Heyden, B. Farajmand, Y. Kolahderazi, Qualitative and quantitative analysis of peanut adulteration in almond powder samples using multi-elemental fingerprinting combined with multivariate data analysis methods, Food Control 82 (2017) 31–41, https://doi.org/10.1016/j.foodcont.2017.06.014.

[11] G. Campmajó, R. Saez-Vigo, J. Saurina, O. Núñez, High-performance liquid chromatography with fluorescence detection fingerprinting combined with chemometrics for nut classification and the detection and quantitation of almond-based product adulterations, Food Control 114 (2020), 107265, https://doi.org/10.1016/j.foodcont.2020.107265.

[12] H. Eksi-Kocak, O. Mentes-Yilmaz, I.H. Boyaci, Detection of green pea adulteration in pistachio nut granules by using Raman hyperspectral imaging, Eur. Food Res. Technol. 242 (2016) 271–277, https://doi.org/10.1007/s00217-015-2538-3.

[13] M.I. López, E. Trullols, M.P. Callao, I. Ruisánchez, Multivariate screening in food adulteration: untargeted versus targeted modelling, Food Chem. 147 (2014) 177–181, https://doi.org/10.1016/j.foodchem.2013.09.139.

[14] C. Márquez, M.I. López, I. Ruisánchez, M.P. Callao, FT-Raman and NIR spectroscopy data fusion strategy for multivariate qualitative analysis of food fraud, Talanta 161 (2016) 80–86, https://doi.org/10.1016/j.talanta.2016.08.003.

[15] C. Zhu, X. Fu, J. Zhang, K. Qin, C. Wu, Review of portable near infrared spectrometers: current status and new techniques, J. Near Infrared Spectrosc. 30 (2022) 51–66, https://doi.org/10.1177/09670335211030617.
[16] J. Nogales-Bueno, L. Feliz, B. Baca-Bocanegra, J.M. Hernández-Hierro, F. J. Heredia, J.M. Barroso, A.E. Rato, Comparative study on the use of three different near infrared spectroscopy recording methodologies for varietal discrimination of walnuts, Talanta 206 (2020), 120189, https://doi.org/10.1016/j.talanta.2019.120189.
[17] V. Wiedemair, C.W. Huck, Evaluation of the performance of three hand-held near-infrared spectrometers through investigation of total antioxidant capacity in gluten-free grains, Talanta 189 (2018) 233–240, https://doi.org/10.1016/j.talanta.2018.06.056.
[18] M.I. González-Martín, O. Escuredo, M. Hernández-Jiménez, I. Revilla, A.M. A. Vivar-Quintana, I. Martínez-Martín, P. Hernández-Ramos, Prediction of stable isotopes and fatty acids in subcutaneous fat of Iberian pigs by means of NIR: a comparison between benchtop and portable systems, Talanta 224 (2021), 121817, https://doi.org/10.1016/j.talanta.2020.121817.
[19] A.J. Fernández-Espinosa, Combining PLS regression with portable NIR spectroscopy to on-line monitor quality parameters in intact olives for determining optimal harvesting time, Talanta 148 (2016) 216–228, https://doi.org/10.1016/j.talanta.2015.10.084.
[20] P. Oliveri, C. Malegori, E. Mustorgi, M. Casale, Qualitative pattern recognition in chemistry: theoretical background and practical guidelines, Microchem. J. 162 (2021), 105725, https://doi.org/10.1016/j.microc.2020.105725.
[21] M.I. López, M.P. Callao, I. Ruisánchez, A tutorial on the validation of qualitative methods: from the univariate to the multivariate approach, Anal. Chim. Acta 891 (2015) 62–72, https://doi.org/10.1016/j.aca.2015.06.032.
[22] A.L. Pomerantsev, O.Y. Rodionova, New trends in qualitative analysis: performance, optimization, and validation of multi-class and soft models, TrAC, Trends Anal. Chem. 143 (2021), 116372, https://doi.org/10.1016/j.trac.2021.116372.
[23] D. Ballabio, F. Grisoni, R. Todeschini, Multivariate comparison of classification performance measures, Chemometr. Intell. Lab. Syst. 174 (2018) 33–44, https://doi.org/10.1016/j.chemolab.2017.12.004.
[24] L. Cuadros-Rodríguez, E. Pérez-Castaño, C. Ruiz-Samblás, Quality performance metrics in multivariate classification methods for qualitative analysis, TrAC, Trends Anal. Chem. 80 (2016) 612–624, https://doi.org/10.1016/j.trac.2016.04.021.
[25] Y.H. Yun, H.D. Li, B.C. Deng, D.S. Cao, An overview of variable selection methods in multivariate analysis of near-infrared spectra, TrAC, Trends Anal. Chem. 113 (2019) 102–115, https://doi.org/10.1016/j.trac.2019.01.018.
[26] A.A. Gomes, S.M. Azcarate, P.H.G.D. Diniz, D.D.S. Fernandes, G. Veras, Variable selection in the chemometric treatment of food data: a tutorial review, Food Chem. 370 (2022), 131072, https://doi.org/10.1016/j.foodchem.2021.131072.
[27] A. Mir-Cerdà, B. Granell, A. Izquierdo-Llopart, A. Sahuquillo, J.F. López-Sánchez, J. Saurina, S. Sentellas, Data fusion approaches for the characterization of musts and wines based on biogenic amine and elemental composition, Sens 22 (2022) 2132, https://doi.org/10.3390/s22062132.
[28] E. Borràs, J. Ferré, R. Boqué, M. Mestres, L. Aceña, O. Busto, Data fusion methodologies for food and beverage authentication and quality assessment –A review, Anal. Chim. Acta 891 (2015) 1–14, https://doi.org/10.1016/j.aca.2015.04.042.
[29] M.P. Callao, I. Ruisánchez, An overview of multivariate qualitative methods for food fraud detection, Food Control 86 (2018) 283–293, https://doi.org/10.1016/j.foodcont.2017.11.034.
[30] R. Vitale, F. Marini, C. Ruckebusch, SIMCA modeling for overlapping classes: fixed or optimized decision threshold? Anal. Chem. 90 (2018) 10738–10747, https://doi.org/10.1021/acs.analchem.8b01270.
[31] I. Ruisánchez, A.M. Jiménez-Carvelo, M.P. Callao, ROC curves for the optimization of one-class model parameters. A case study: authenticating extra virgin olive oil from a Catalan protected designation of origin, Talanta 222 (2021), 121564, https://doi.org/10.1016/j.talanta.2020.121564.
[32] F.C. Lemyre, B. Desharnais, J. Laquerre, M.A. Morel, C. Côté, P. Mireault, C. D. Skinner, Qualitative threshold method validation and uncertainty evaluation: a theoretical framework and application to a 40 analytes LC-MS/MS method, Drug Test. Anal. 12 (2020) 1287–1297, https://doi.org/10.1002/dta.2867.
[33] C.S. Gondim, R.G. Junqueira, S.V.C. de Souza, M.P. Callao, I. Ruisanchez, Determining performance parameters in qualitative multivariate methods using probability of detection (POD) curves. Case study: two common milk adulterants, Talanta 168 (2017) 23–30, https://doi.org/10.1016/j.talanta.2016.12.065.
[34] I. Ruisánchez, G. Rovira, M.P. Callao, Multivariate qualitative methodology for semi-quantitative information. A case study: adulteration of olive oil with sunflower oil, Anal. Chim. Acta 1206 (2022), 339785, https://doi.org/10.1016/j.aca.2022.339785.
[35] R.W. Kennard, L.A. Stone, Computer aided design of experiments, Technometrics 11 (1969) 137–148, https://doi.org/10.1080/00401706.1969.10490666.
[36] A. Rius, M.P. Callao, F.X. Rius, Multivariate statistical process control applied to sulfate determination by sequential injection analysis, Analyst 122 (1997) 737–741, https://doi.org/10.1039/A607954G.
[37] S. Ghosh, P. Mishra, S.N.H. Mohamad, R.M. de Santos, B.D. Iglesias, P.B. Elorza, Discrimination of peanuts from bulk cereals and nuts by near infrared reflectance spectroscopy, Biosyst. Eng. 151 (2016) 178–186, https://doi.org/10.1016/j.biosystemseng.2016.09.008.
[38] H.E. Genis, S. Durna, I.H. Boyaci, Determination of green pea and spinach adulteration in pistachio nuts using NIR spectroscopy, LWT – food Sci, Technol. 136 (2021), 110008, https://doi.org/10.1016/j.lwt.2020.110008.
[39] N. Shetty, Å. Rinnan, R. Gislum, Selection of representative calibration sample sets for near-infrared reflectance spectroscopy to predict nitrogen concentration grasses, Chemometr. Intell. Lab. Syst. 111 (2012) 59–65, https://doi.org/10.1016/j.chemolab.2011.11.013.
[40] A. Fort, I. Ruisánchez, M.P. Callao, Chemometric strategies for authenticating extra virgin olive oils from two geographically adjacent Catalan protected designations of origin, Microchem. J. 169 (2021), 106611, https://doi.org/10.1016/j.microc.2021.106611.