



Geographical Classification of Authentic Moroccan Argan Oils and the Rapid Detection of Soya and Sunflower Oil Adulteration with ATR-FTIR Spectroscopy and Chemometrics

Fouad Taous¹ · Tibari El Ghali¹ · Hamid Marah¹ · Khalid Laraki¹ · Marivil Islam² · Andrew Cannavan² · Simon Kelly²

Received: 1 November 2021 / Accepted: 23 May 2022 / Published online: 4 July 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

For the purpose of implementing low-cost, field-deployable analytical techniques to ensure the authenticity and traceability of argan oil, a comprehensive approach that combined mid-infrared spectroscopy data with discriminant and modeling classification methods—including principal component analysis (PCA), soft independent modeling of class analogy (SIMCA), and data-driven soft independent modeling of class analogy (DD-SIMCA)—was applied to classify and check the authenticity of 78 argan oil samples according to their geographical origins and distinguish them from two sets of 24 argan oil samples that contained 5–100% w/w soya or sunflower oil. Optimal models were selected as combinations of many wavelength ranges and data pre-processing methods, thus leading to maximum efficiency for cross-validation. The discrimination approach provided satisfactory classification results with good efficiency for determining argan oil authenticity and detecting adulteration. In addition, an adulteration quantification study was performed with the help of partial least square (PLS) regression of binary mixture, with this demonstrating good linear regression for actual values against predicted ones. The coefficient of determination (R^2) was 0.999, while the root mean square errors of calibration (RMSEC) were low at 0.389% and 0.685% w/w and the root mean square errors of validation (RMSEV) were 0.639 and 0.863% w/w for soya and sunflower adulteration, respectively. Moreover, the PLS models best predicted adulterant content, with the R^2 and root mean square error of prediction (RMSEP) being 0.998 and 1.067% (w/w), respectively, for soya and 0.997 and 1.199% (w/w) for sunflower.

Keywords Argan oil · Classification · Adulteration · Quantification · SIMCA · Partial least square

Introduction

Argan oil is extracted from the fruits of *Argania spinosa*, a tree endemic to a specific area of southwestern Morocco. During recent decades, the specific characteristics of argan oil and its beneficial effects on human health have attracted intensive, rigorous scientific, and marketing efforts. It is now widely recognized as one of the most expensive edible and cosmetic oils in international markets. Its composition has been extensively studied in numerous studies,

with them reporting that it is especially rich in long-chain monounsaturated fatty acids—principally oleic (49%) and linoleic acid (36%)—and natural antioxidants, such as sterols, polyphenols, and tocopherols (Madawala et al. 2012; Sour et al. 2012, 2015; Aabd et al. 2013; Lopez et al. 2013; Kharbach et al. 2018; Maleš et al. 2018; Miklavčič et al. 2020; Gharby and Charrouf 2021; Simões et al. 2021). Its high nutritional value and biochemical properties have led to it being believed to have therapeutic value for treating and preventing several diseases (Drissi et al. 2004; Cherki et al. 2006; Charrouf and Guillaume 2008; El Monfalouti et al. 2010; Guillaume and Charrouf 2011; Sour et al. 2015). While the growing demand and soaring prices for Moroccan argan oil present economic and social opportunities, it is unfortunately also tempting for fraudsters and dishonest traders to make quick financial gains by mislabeling and/or adulterating products.

The most common form of such fraud is combining argan oil with cheaper oils, such as soya, sunflower, or hazelnut

✉ Fouad Taous
f_taous@yahoo.fr

¹ National Center for Energy Sciences and Nuclear Techniques, B.P. 1382 R.P. 10001, Rabat, Morocco

² International Atomic Energy Agency, Vienna International Center, PO Box 100, Wagramer Strasse 5, 1400 Vienna, Austria

oil. Furthermore, the introduction of strict quality standards for argan oil as a product with protected designation of origin as an additional assurance of quality and authenticity has increased consumer interest in learning about the geographical origin and/or variety of argan oil that they buy. However, authenticating the varietal or geographical origin, as well as detecting adulterants, often represents a real analytical challenge. Thus, in order to monitor and control the authenticity of argan oil, several research studies have recently tried using various physicochemical determinations associated with chemometrics data processing (Rueda et al. 2014; Kharbach et al. 2017, 2018, 2019; Farres et al. 2019; Gunning et al. 2020; Mohammed et al. 2020; Elgadi et al. 2021). Among these, González et al. (2010) obtained better classification rates for edible oils (i.e., argan, sunflower, olive, seed, and soya oils) from their elemental content based on the type or variety using inductively coupled plasma atomic emission spectrometry (ICP-AES) combined with different chemometric approaches. What is more, new approaches have been developed to detect the adulteration of argan oil with vegetable oils through new techniques and instruments (Hilali et al. 2007; Ourracha et al. 2012; Oussama et al. 2012; Mohammed et al. 2013; Addou et al. 2016; Stokes et al. 2018; Çelik et al. 2019; Farres et al. 2019; Joshi et al. 2019). For instance, Salghi et al. (2014) used high-performance liquid chromatography coupled with evaporative light scattering to evaluate the authenticity of argan oil—as well as detect adulteration with vegetable oils such as sunflower, soya, and olive oils up to the 5% level—by studying the triacylglycerol profiles of argan oil. However, despite the high resolution and reliability of these methods, some are impractical for widespread use due to various reasons, such as the time and cost required, the need for skilled operators, the generation of hazardous chemical waste, sample preparation, preprocessing, and various technological parameters, such as the harvest conditions and the extraction, preservation, storage, and transport of the oil.

Today, the most established technique in this area is based on infrared spectroscopy, which is now widely used in food studies. It has become a particularly powerful analytical tool for studying the authenticity of edible oils and detecting any adulteration, especially when combined with techniques from chemometrics (Basri et al. 2017; Vanstone et al. 2018; Yuan et al. 2020; Bragolusi et al. 2021; Du et al. 2021; Sota-Uba et al. 2021). Economically speaking, infrared spectroscopy technology is rapid, inexpensive, and sensitive, and it can be used for high-throughput analysis of food-based components with real-time measurement at all stages of production without requiring specially skilled operators (Prieto et al. 2017).

In this paper, the infrared spectra of pure and adulterated argan oils were determined using Fourier transformed IR (FTIR) spectroscopy. New multivariate statistical analyses

were then applied for the obtained chemical data, such as the principal component analysis (PCA) exploratory method for unsupervised pattern recognition and supervised classification methods like the soft independent modeling class analogy (SIMCA) and the data-driven soft independent modelling of class analogy (DD-SIMCA), in order to classify different argan oil samples according to their geographical origins and detect any adulteration with soya and sunflower oils. In addition, the degree of any adulteration was quantified using the most commonly used regression methods, namely partial least square (PLS) and principal component regression (PCR).

Materials and Methods

Sample Collection

A total of 78 argan kernel samples from five Moroccan argan forest regions (Agadir, Essaouira, Chtouka Aitbaha, Taroudant, and Tiznit) were collected by visiting producers, cooperatives, markets, and industries (Fig. 1). The provenance and quality grade of all the samples had been guaranteed by the suppliers. Argan oil was then mechanically extracted from unroasted kernels using an automatic oil press machine in the laboratory. The oils were then filtered and preserved in 250-mL dark-glass bottles in a refrigerator at +4 °C. These had been filled with nitrogen to avoid

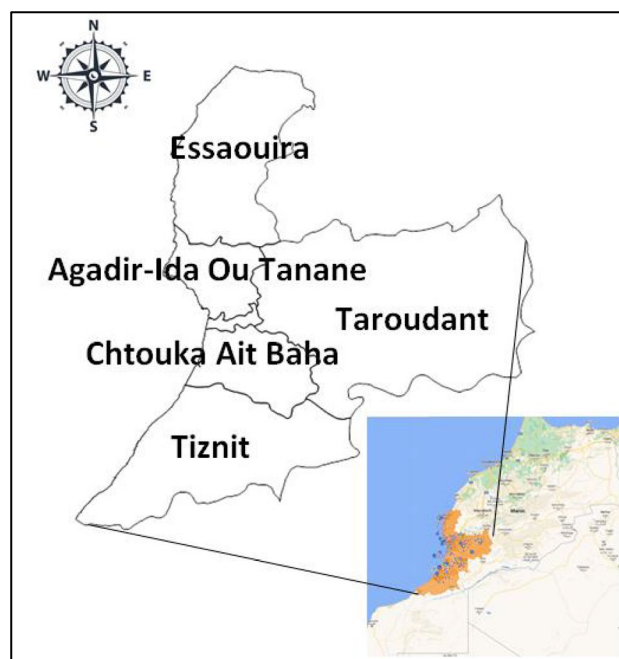


Fig. 1 Moroccan regions from the where argan kernel samples were obtained

oxidation until the day of analysis. Prior to analysis, the oils were left to stand at room temperature for at least 24 h.

Sample Preparation

Soya and sunflower oils were procured from the local supermarket as adulterants. Binary mixtures of argan–soya and argan–sunflower oils were prepared by combining an individual authentic argan oil with varying proportions of adulterant oil (i.e., 0%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, and 100% weight/weight). The resulting 24 mixtures were used as a calibration set to construct the regression models. Next, in order to evaluate the prediction performance of the regression models, 10 further mixtures of soya and sunflower oil with argan oil at 3%, 12%, 23%, 31%, 42%, 49%, 55%, 66%, 74%, and 83% proportions were prepared.

FTIR Spectra Acquisition

All the argan oil samples, both authentic and adulterated mixtures, were subjected to FTIR spectral measurements using a PerkinElmer Spectrum Two equipped with a UATR. All spectra were recorded from 4000 to 400 cm^{-1} at a resolution of 1 cm^{-1} . Using a Pasteur pipette, approximately 0.25 mL of each sample was placed in direct contact with single bounce-attenuated total reflectance using the Smart ARK accessory. The diamond ATR crystal was cleaned between samples using laboratory tissue and a small amount of hexane. Each FTIR spectrum was subsequently subtracted using reference spectra (air) as background. The sample spectra were measured for three replicates and displayed as an average spectrum.

Multivariate Analysis

Raw spectral files were exported from the spectrometer acquisition software to The Unscrambler (v 7.6; CAMO A/S, Oslo, Norway) software for data analysis. The original spectra data were scrutinized for unusual or outlying samples through principal component analysis. Next, a preliminary analysis of the data set for natural groupings was performed through PCA, while classification was done using single-category SIMCA (Wold and Sjöström 1977; De Maesschalck et al. 1999; Bevilacqua et al. 2014; Oliveri et al. 2021). The latter is a popular pattern-recognition technique that is based on principal component analysis, which is currently considered the most commonly used method for class modeling classification in chemometrics contexts. It involves first constructing an appropriate PCA model for each class in the training samples and calculating the residual variance of each model. The residual variances are then calculated for the unknown samples in the prediction sample set and compared to those of

the developed PCA class models to determine which classes of the calibration sample set to assign. A full cross-validation and nonlinear iterative partial least square (NIPALS) algorithm was used to develop the PCA models. Samples were assigned to either a calibration or prediction set based on their positions in the main spectral file. More specifically, even-numbered samples were used for model development, while odd-numbered samples were used for model evaluation. In order to estimate similarity, samples were assigned or refused based on their reduced distance from the class space (d) in the class-model, which was calculated by taking into account the Mahalanobis distance of samples from the center of the score space and their orthogonal distance from their bilinear projection. Classification model quality was evaluated by calculating the models' sensitivity and specificity based on true and false class assignments. Sensitivity refers to the proportion of positive cases captured, whereas specificity refers to the proportion of negatives that are correctly identified. These statistical parameters were calculated using the equations below:

$$\text{sensitivity} = \frac{TP}{TP + FN}$$

where, TP is the true positive (the number of target samples correctly attributed as positive) and FN is the false negative (the number of target samples incorrectly attributed as negative).

$$\text{specificity} = \frac{TN}{TN + FP}$$

where, TN is the true negative (the number of non-target samples correctly attributed as negative) and FP is the false positive (the number of non-target samples incorrectly attributed as positive).

Unless otherwise stated, all classification results are given at the 95% confidence level.

In order to extend the application range of the chemometric tools, the data were exported to a modified version of the SIMCA method, namely DD-SIMCA (Pomerantsev and Rodionova 2014), to establish a rapid identification model. In the present study, calculations were performed using standard Excel functions and a chemometric add-in for Microsoft Excel (SIMCA template.xlsm) (Pomerantsev 2014). The parameter α ($0 < \alpha < 1$) was used to manage the size of the acceptance area, which corresponded to the green boundary line. A lower α means a wider acceptance area, so samples are more likely to be accepted. Meanwhile, the parameter γ ($0 < \gamma < 1$) was used to manage the size of the outlier area corresponding to the red boundary line. A lower γ means a wider area, so fewer outliers will be discovered. The samples are colored according to their status, such that regular objects (green dots) belong to the acceptance area,

outliers (red squares) were outside the outlier area, and extreme samples (yellow diamonds) were located in between these. In DD-SIMCA classification, sensitivity is calculated for both calibration and validation sets, whereas the specificity is only determined from prediction set.

Quantitative spectral decomposition techniques—such as PLS and PCR, which are typically used for predictive linear modeling—were used to build models to determine the concentration of soy or sunflower oil in adulterated argan oil. No outliers were removed during calibration. The root mean square error of calibration (RMSEC), leave-one-out cross validation (RMSECV), and prediction (RMSEP) were used to assess the predictive ability of the models.

Results and Discussion

FTIR Spectral Analysis

The typical spectra for pure argan, soya, and sunflower oils are shown in Fig. 2, and it is readily apparent that no significant visible differences exist among these spectra. The specific frequency of a given vibration mode depends upon the specific arrangement of the carbonyl groups and the lipid hydrocarbon backbone, while the intensity of the specific vibration mode depends on the relative concentration of the related functional group. Argan oil and the soya/sunflower adulterated oils differ in composition principally in their linoleic and oleic acid contents. Thus,

transmittance minima are clearly evident at 2922, 2853, 1744, and 1161 cm^{-1} , together with smaller transmittance bands at 1464, 1377, 1109, 1097, and 723 cm^{-1} . From the literature (Guillen and Cabo 1997), it can be surmised that the bands occurring in the region between 1744 and 2922 cm^{-1} are due to stretching vibrations of the =C-H group, while the band at 1300 and 1465 cm^{-1} is due to the bending vibration of the C-H bonds that are present in triglyceride molecules, and the band at 1097 cm^{-1} relates to the stretching vibrations of -C-O groups. The

Table 1 Functional groups and modes of vibration in the spectra with the assigned functional group, the mode of vibration, and the intensity (Guillen and Cabo 1997)

Frequency (cm^{-1})	Functional group	Mode of vibration	Intensity ^a
3006	$\equiv\text{C-H}(\text{cis-})$	Stretching	m
2924	$-\text{C-H}(\text{CH}_2)$	Stretching (asym)	vst
2853	$-\text{C-H}(\text{CH}_2)$	Stretching (sym)	vst
1746	$-\text{C=O}(\text{ester})$	Stretching	vst
1465	$-\text{C-H}(\text{CH}_2, \text{CH}_3)$	Bending (scissoring)	m
1377	$-\text{C-H}(\text{CH}_3)$	Bending (sym)	m
1163	$-\text{C=O-CH}_2-$	Stretching, Bending	st
1097	$-\text{C-O}$	Stretching	m
723	$-(\text{CH}_2)_n-$, $-\text{HC=CH-}(\text{cis})$	Bending (rocking)	m

^aw, weak; m, medium; st, strong; vst, very strong

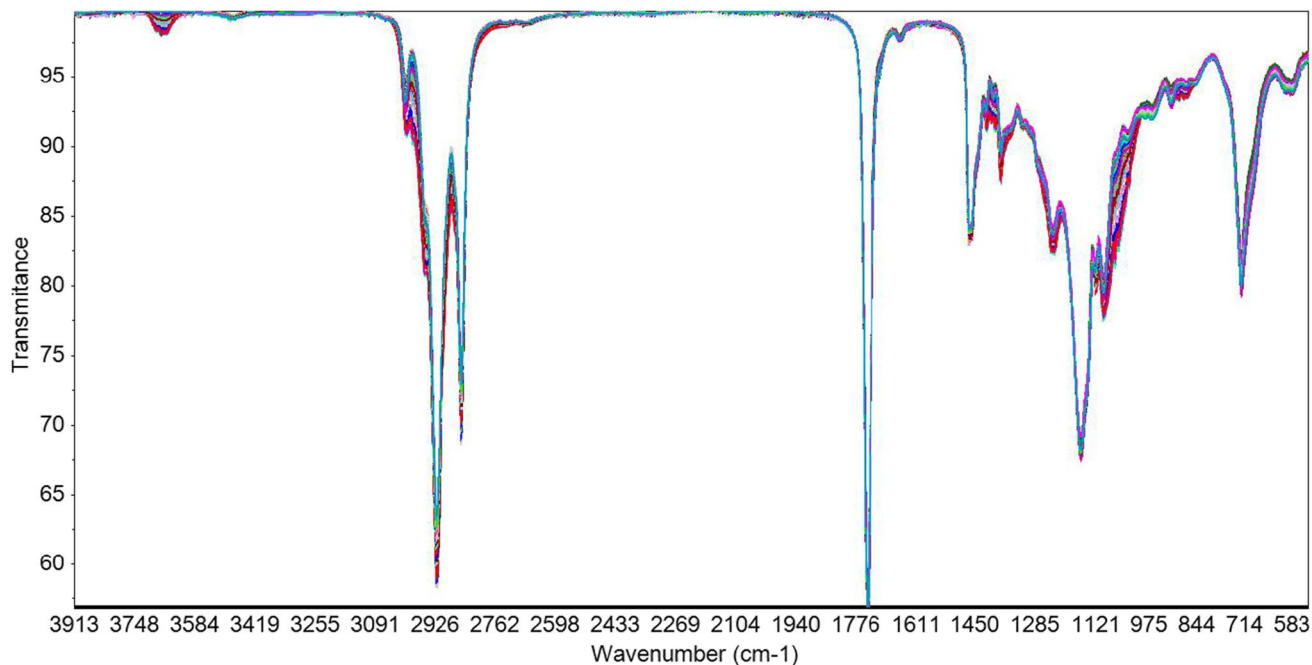


Fig. 2 Plot of spectra for different pure argan, soya, and sunflower oil samples

band at 723 cm^{-1} is connected with the bending vibration of $-\text{HC}=\text{CH}-$ groups. Table 1 summarizes the infrared band assignments that were reported for the considered edible oils. The prominent features in the spectrum include the bands in the region of 2922 cm^{-1} due to the $-\text{CH}-$ stretch of the hydrocarbon chains and those in the region of 1744 cm^{-1} , which are due to the carbonyl groups in the triglyceride.

Discriminant Analysis

The similarity among the spectra reflects the similarities in the chemical compositions of the oils. Although the FTIR spectra of all the argan oil samples are similar in appearance, minor differences manifest in a more detailed analysis through small band shifts and small corresponding changes in their relative intensity (Fig. 2). The authentic argan oil samples were split into a calibration (60%) and a validation set (40%), with these being randomly selected from each region (Table 2). In order to achieve a good classification for authentic argan oil, different spectral pre-processing methods were tested in different spectral regions in combination with the PCA technique, with the best results being obtained using mean centering of the combined raw spectral data from 3044–2943, 2904–2870, 1423–1212, 1130–1001, and 913–853 cm^{-1} . The PCA results for the calibration and validation datasets are illustrated in Fig. 3. The two-dimensional score plot of the two first PCs (PC1 and PC2), which describe 100% of the total variance, reveals that most of the authentic argan oils are clearly separated in the space, with samples generally being grouped according to their geographical origins for the five classes both in the calibration and the validation set. All validation samples are clearly plotted close to their proper classes. However, the closeness that can be observed between the oils from Agadir and Essaouira—as well as among the oils from Chtouka, Taroudant, and Tiznite—can be explained by the fact that these regions neighbor each other and therefore have similar climates, as well as by the fact that a limited number

of representative samples were selected in certain regions. More specifically, Agadir and Essaouira are oceanic coastal regions with a hot, humid, and arid bioclimate, while Chtouka and Tiznite are sub-Saharan regions with climatic conditions that are geographically close to arid Mediterranean. The Taroudant region, meanwhile, is surrounded by two mountain ranges (High Atlas and Anti-Atlas) and has a hot, semi-arid bioclimate.

In order to obtain a classification rule for the set of n known groups of argan oil, such that new samples can be assigned, the soft independent modeling of class analogy (SIMCA) technique was applied. When constructing the model with SIMCA, the similarities between samples with defined categories were addressed, and a PCA was performed individually for each class, resulting in its own class model. As SIMCA is based on PCA, the individual PCA model for each group was created using data from the training set (Table 2). In the SIMCA model, the optimal numbers of PCs were retained for each class based on the explained variance and the predicted residual error sum of squares (PRESS). The classification results are summarized in Table 3 in terms of the sensitivity and specificity rates for the calibration and validation sets from the five regions. For the SIMCA models constructed from the Agadir and Essaouira calibration and validation sets, it can be seen that no target samples were assigned wrongly, with them showing excellent prediction performance with both a sensitivity and specificity of 100%, meaning that the samples were correctly distributed in their respective predefined groups. For the Chtouka, Taroudant, and Tiznite calibration and validation sets, the constructed SIMCA models gave very satisfactory results, with the sensitivity varying between 73 and 92% because some target samples were assigned to more than one group or assigned as non-target samples. In addition, the classification results for these groups show that some non-target samples were not correctly rejected by the model and instead wrongly assigned as target samples, resulting in specificity between 90 and 96%.

Another experiment for geographic discrimination was carried out using DD-SIMCA chemometric analysis of the authentic samples from the different regions in order to build models capable of determining whether the samples from other regions can be excluded as aliens. Taking into account the optimal number of PCAs used in SIMCA for each region (Table 3), a one-class target classification model was developed using the authentic samples from a particular region and applying mean-centering preprocessing methods to the entire set of available spectra. The sizes of the acceptance and outlier areas were set at $\alpha=0.01$ and $\gamma=0.05$ for all cases. The results of the DD-SIMCA classification models are illustrated in Fig. 4a–e, while their results in terms of sensitivity and specificity are summarized in Table 4. Since only target samples were used in the calibration and

Table 2 Geographical origins of the samples and their occurrence in the different data sets

Geographical origin	Number of samples	Calibration	Validation
Agadir	10	7	3
Chtouka	14	8	6
Essaouira	15	9	6
Taroudant	14	9	5
Tiznite	25	14	11
Total	78	47	31
%	100%	60%	40%

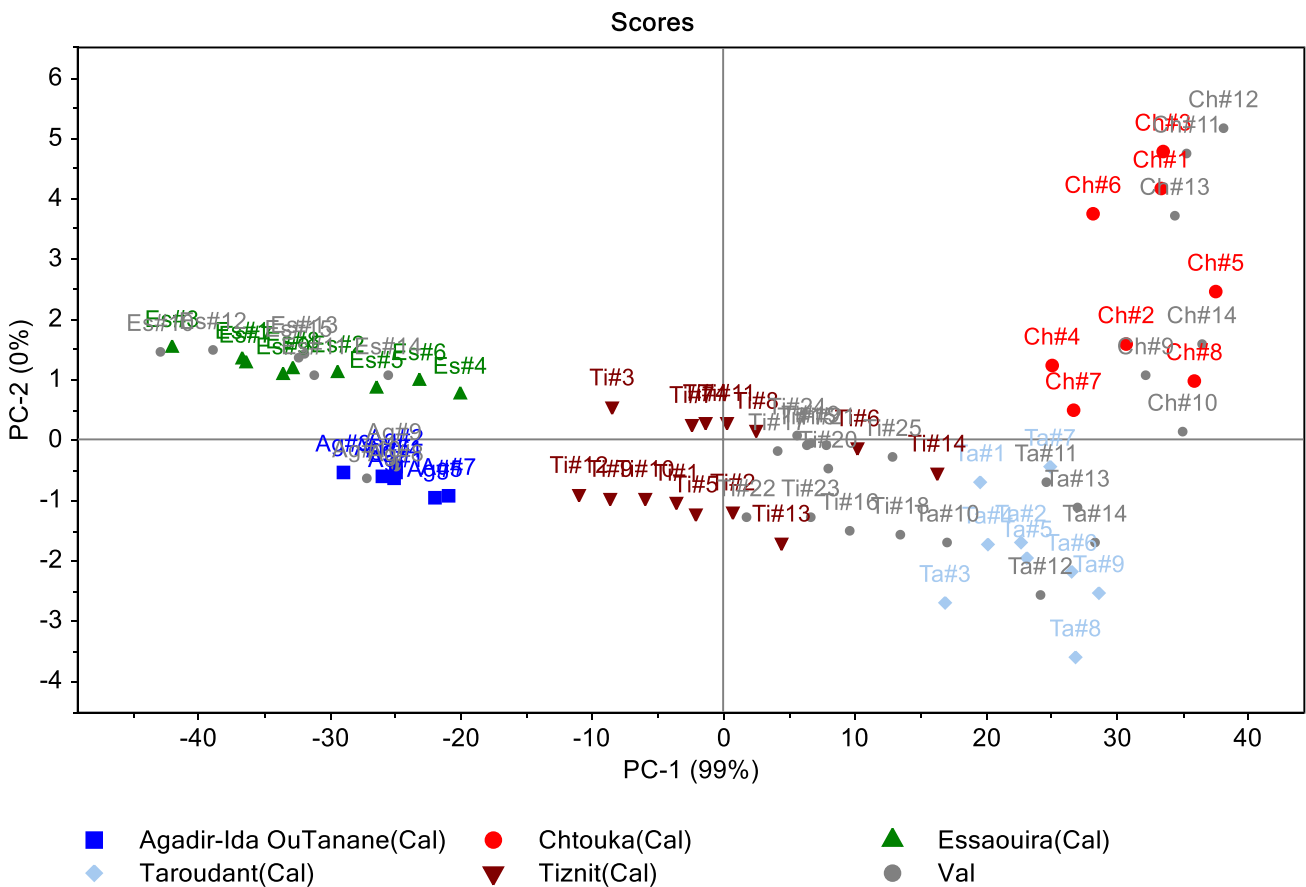


Fig. 3 PCA score plot of the mid-infrared spectra for authentic argan oils (PC1 vs. PC2) for calibration and validation sets

Table 3 Results of the SIMCA classification models when using the near-infrared data for the argan oil samples

Argan oil origin	Data set	TP	FN	TN	FP	Sensitivity (%)	Specificity (%)	Number of PCs
Agadir	Cal	7	0	46	0	100	100	2
	Val	3	0	25	0	100	100	
Chtouka	Cal	8	3	39	2	73	95	4
	Val	4	1	24	1	80	96	
Essaouira	Cal	9	0	38	0	100	100	2
	Val	6	0	22	0	100	100	
Taroudante	Cal	9	2	38	4	82	90	3
	Val	5	1	23	2	83	92	
Tiznite	Cal	11	1	33	0	92	100	2
	Val	10	1	18	0	91	100	

Cal, calibration; Val, validation, Pred, prediction; TP, true positive; FN, false negative; TN, true negative; FP, false positive

validation step (i.e., objects of the modeled class were not used for building the model), the sensitivity rate was the only parameter that could be calculated. For the calibration sets, all target samples fell correctly in line for all models,

resulting in sensitivity rates of 100%. For the validation sets, the Agadir and Essaouira models demonstrated 100% sensitivity, indicating that all target samples were correctly identified. However, the Chtouka, Taroudant, and Tiznite

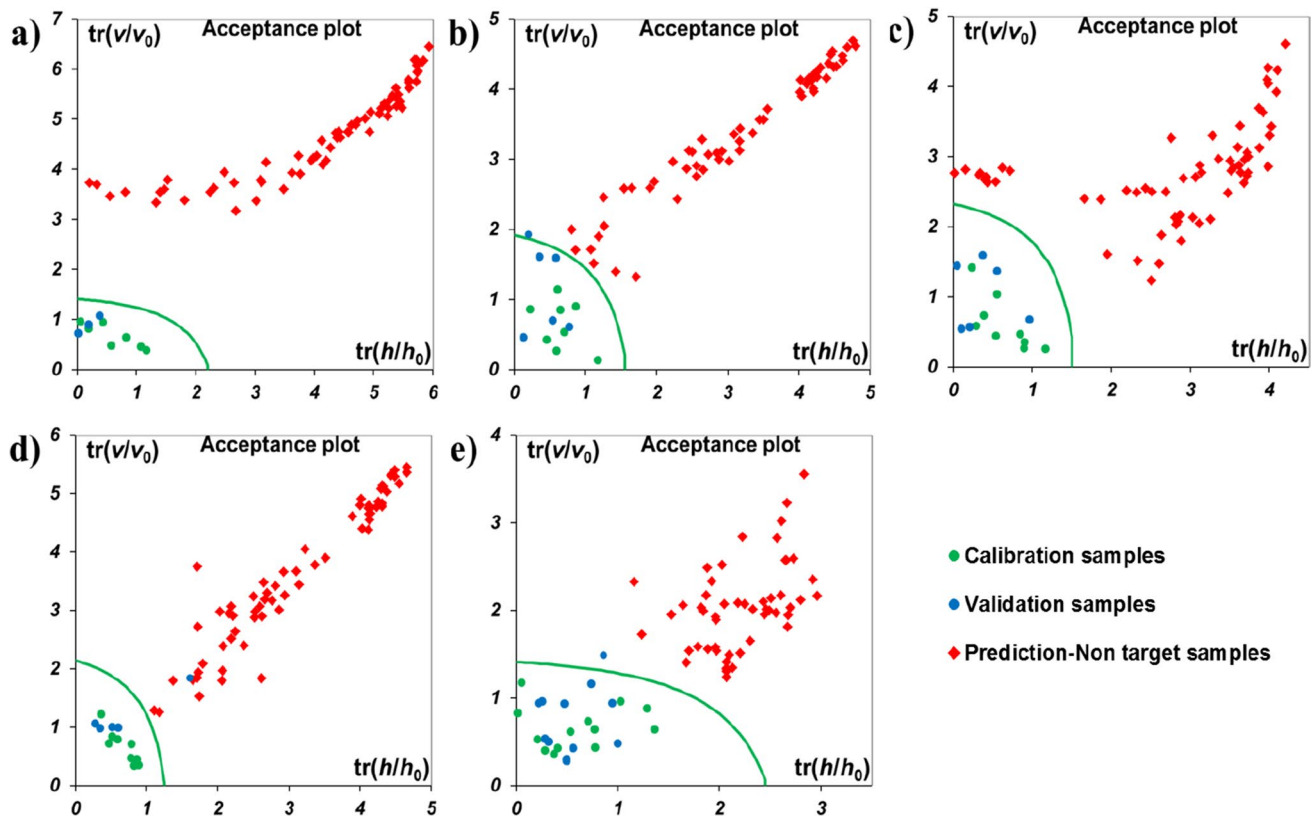


Fig. 4 Results of DD-SIMCA for authentic argan oils from a targeted geographical region vs. other regions: **a** Agadir, **b** Chtouka, **c** Essaouira, **d** Taroudante, **e** Tiznite

Table 4 Sensitivity/specificity achieved by the DD-SIMCA models when applied to pure argan oils from different geographical regions

Argan Oil	Data set	TP	FN	TN	FP	Sensitivity (%)	Specificity (%)	Number of PCs
Agadir	Cal	7	0	-	-	100	-	2
	Val	3	0	-	-	100	-	
	Pred	0	0	68	0	-	100	
Chtouka	Cal	8	0	-	-	100	-	3
	Val	5	1	-	-	83	-	
	Pred	0	0	64	0	-	100	
Essaouira	Cal	9	0	-	-	100	-	2
	Val	6	0	-	-	100	-	
	Pred	0	0	63	0	-	100	
Taroudante	Cal	9	0	-	-	100	-	4
	Val	4	1	-	-	80	-	
	Pred	0	0	64	0	-	100	
Tiznite	Cal	14	0	-	-	100	-	2
	Val	10	1	-	-	91	-	
	Pred	0	0	53	0	-	100	

Cal, calibration; *Val*, validation, *Pred*, prediction; *TP*, true positive; *FN*, false negative; *TN*, true negative; *FP*, false positive

models sometimes assigned a target sample as a non-target, resulting in 83%, 80%, and 91% sensitivity, respectively. For the prediction sets, a 100% specificity rate was evident for all classes, indicating that the samples that did not belong

to the relevant region were correctly rejected by the models, as evidenced by how they are plotted outside the threshold line. Despite the overall sample set being limited in size, the relatively high sensitivity/specificity rates (90–100%)

obtained for certain regions demonstrate the potential for using near-infrared spectroscopy and DD-SIMCA to classify argan oils according to their geographic origin.

Detecting and Quantifying Adulteration by Soya and Sunflower Oils in Argan Oil

Detecting Adulteration

In this section, the IR spectra of the binary mixtures were investigated as a means for detecting the adulteration of argan oil by allowing low levels of adulteration to be more easily detected through multivariate calibrations. To classify the oil IR spectra and quantify the discrepancy between pure and adulterated oil, the multivariate statistical PCA was performed after applying Savitsky–Golay smoothing, a second-order derivative and mean-centering pre-treatment procedure, at the combined frequency regions of 3044–2943, 2904–2870, 1423–1212, 1130–1001, and 913–853 cm^{-1} . As shown in Fig. 5, 92% of the variability in the data was explained by PC1 and 6% was explained by PC2. The pure argan, soya, and sunflower oils were located in well-separated regions in the PCA distribution, with the binary mixtures occupying intermediate positions depending on the particular adulterant and the percentage adulterant content in the argan oil. Indeed, the distance from the pure argan oil increases with increasing adulterant content from 10 to 100%. From these data, it is evident that the adopted spectral parameters depend on the soya or sunflower oil present in the oil mixture, with there being a linear correlation between the proportion of adulterant oil and the PCA score. This confirms that FTIR

spectroscopy can be successfully used to detect the presence of soya and sunflower oils in argan oil samples.

Adulterant detection was studied further to demonstrate the usefulness of NIR spectral data for discriminating between different varieties of argan oil types and soya and sunflower oils by using an open-source data-driven soft independent modelling of class analogy (DD-SIMCA) chemometrics tool, with the data from the 78 authentic argan oil samples listed in Table 2 being used to define the target class. The 47 samples randomly selected for the calibration set were used to establish the model, while the remaining 31 samples were selected for the validation set and to evaluate the robustness of the model. Finally, the prediction set comprised the 42 adulterated samples. The sizes of the acceptance and outlier areas were given by $\alpha=0.05$ and $\gamma=0.05$. Optimum classification with DD-SIMCA was obtained by considering four PCA components, with good sensitivity rates being obtained for both calibration (94%) and validation (97%). As illustrated in Fig. 6, only three samples from the calibration set were located outside the acceptance area, but they were considered as extremes (orange dots), meaning that all training objects were members of the target class as expected. Applying the classification model to the validation set resulted in all samples (blue dots) being within the acceptance area (i.e., inside the green line); thus, they were correctly classified as pure argan oil. The classification model was then applied to the entire set of adulterated samples (i.e., no target samples), ranging from 3 to 100% w/w adulteration with soya or sunflower oil. Figure 6 reveals that these samples (the red dots) are plotted out of the acceptance area, so they were identified as impure argan samples, thus implying the model has 100% specificity and could potentially be used for verifying argan oil authenticity.

Fig. 5 Plot of principal component analysis (PCA) scores for a single sample of argan oil together with adulterated mixtures with 10–100% soya or sunflower oil

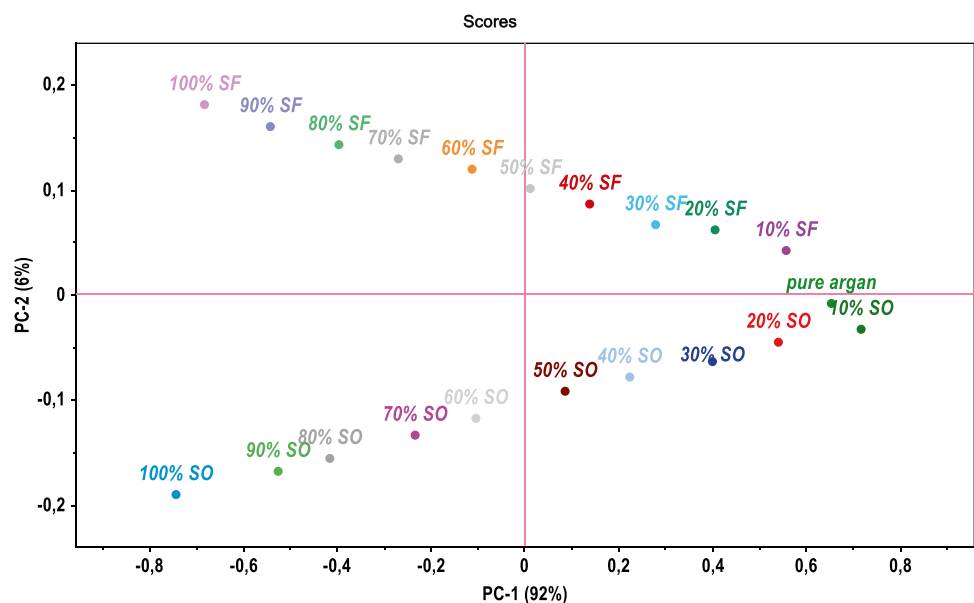
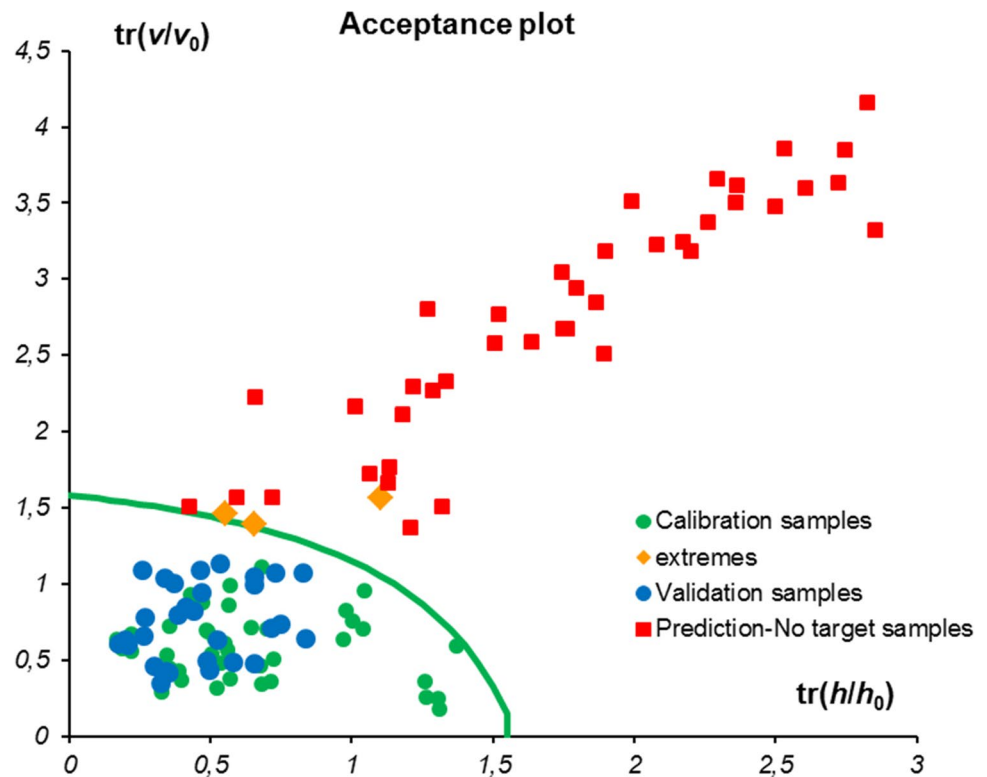


Fig. 6 The DD-SIMCA acceptance plot for the calibration, validation, and prediction samples used for the adulteration model



Quantification

Quantifying the amount of soya or sunflower oil mixed into argan oil was performed by applying multivariate calibrations, namely partial least square (PLS) and principal component regression (PCR), on the spectral data for 11 binary mixtures of each adulterant.

In the PLS and PCR calibration models, the linearity method was evaluated in order to show a proportional relationship between responses (transmittances) and analyte concentrations. The PLS and PCR models were developed using the FTIR mean-centered spectra in the frequency region of $3030\text{--}2860\text{ cm}^{-1}$, where the variations were observed and because of its ability to give the highest values of R^2 and the lowest values of RMSEC, relative to other frequencies. In order to validate the developed model, cross validation using the “leave one out” technique was achieved by excluding one of the calibration samples and then constructing a model for the remaining samples. Subsequently, the model was evaluated using the previously excluded samples, while the error values for the predicted observations were computed. The new samples were then excluded from the model set and a new model was constructed. This process was repeated until all samples had been excluded once in the PLS and PCR models (Rohman and Che Man 2011). The results obtained from the PLS and PCR models in terms of the R^2 , RMSEC, RMSEV, and RMSEP are presented in Table 5.

Based on Table 5, we can say that quantifying SO and SF adulterants in AO was carried out well due to the high coefficient of correlation ($R^2=0.999$) for the calibration and validation models, both with PLS and PCR. However, when compared to PCR, PLS gives a more adequate fit and offers the best calibration and validation models. For the soya PLS model, a low error of calibration was obtained for the root mean square error in calibration (RMSEC = 0.389% w/w), as well as a low validation error as expressed by the root mean square error of validation (RMSEV = 0.639% w/w) when using four factors. Similarly, the presence of SF in AO was quantified using the PLS calibration model, which offers the best results in terms of the high R^2 values for the calibration (0.9994) and validation (0.9993) data and the lowest errors both for RMSEC (0.685% w/w) and RMSEV (0.863% w/w) when using only two factors. Graphical representations of the regressions produced using these models are shown in Fig. 7a and b. The PLS calibration model was also used to calculate the adulterant content of samples in the prediction set. Plots of the relationship between the actual and FTIR-predicted values of adulterant content in the prediction samples are shown in Fig. 7c and d, with the R^2 and root mean square error of prediction being 0.998 and 1.067% (w/w), respectively, for soya and 0.997 and 1.199% (w/w) for sunflower.

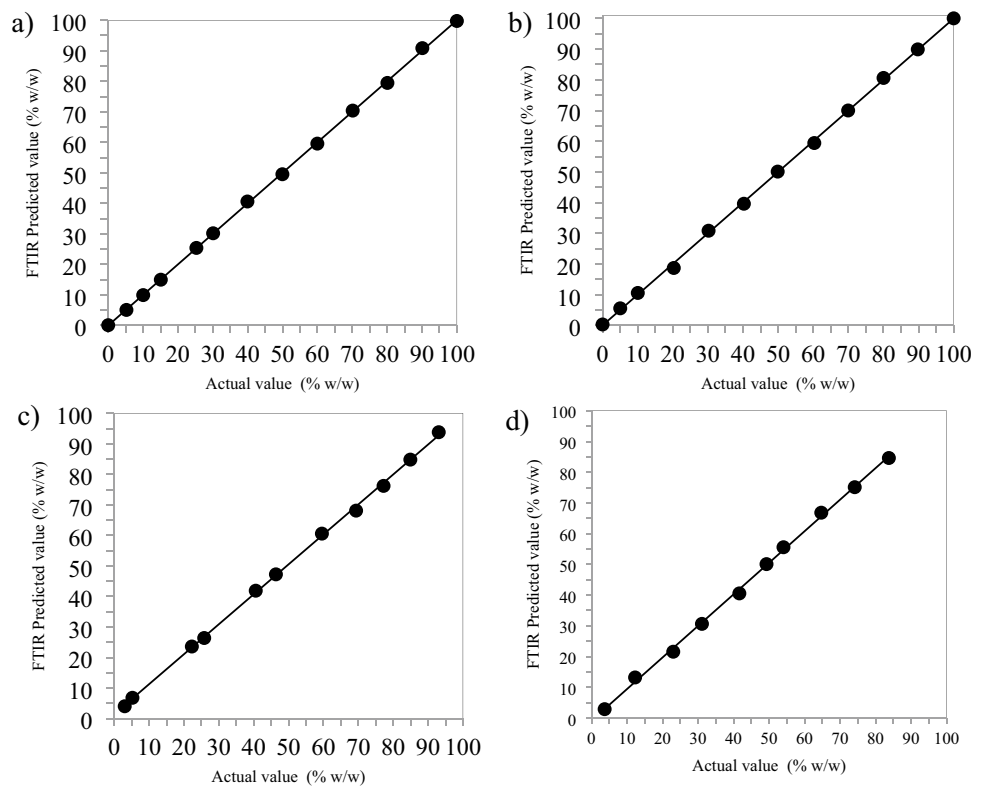
In the PLS calibration model, the verification and justification of the optimum principal component and spectra regions retained for each class and used for constructing the models were carried out by calculating the difference

Table 5 Performance of the multivariate calibration, validation, and prediction for analyzing soya and sunflower oil content in argan oil

Analysis	Factor	Model	Equation	R^2	RMSE (% w/w)
Adulteration with soya oil					
PLS	4	Cal	$y=0.9998x+0.006$	0.999	0.389
		Val	$y=0.998x+0.100$	0.999	0.639
		Pred	$y=0.980x+1.478$	0.998	1.067
PCR	2	Cal	$y=0.999x+0.006$	0.999	0.393
		Val	$y=0.998x+0.101$	0.999	0.64
		Pred	$y=0.980x+1.478$	0.998	1.067
Adulteration with sunflower oil					
PLS	4	Cal	$y=0.999x+0.018$	0.999	0.658
		Val	$y=1.006x-0.272$	0.999	0.863
		Pred	$y=1.027x-0.859$	0.997	1.199
PCR	2	Cal	$y=0.999x+0.019$	0.999	0.662
		Val	$y=1.007x-0.349$	0.999	0.894
		Pred	$y=1.027x-0.835$	0.997	1.205

Cal, calibration; Val, validation, Pred, prediction

Fig. 7 The PLS calibration and prediction models for the relationship between the actual and FTIR-predicted values for AO adulteration with SO (a and c) and SF (b and d)



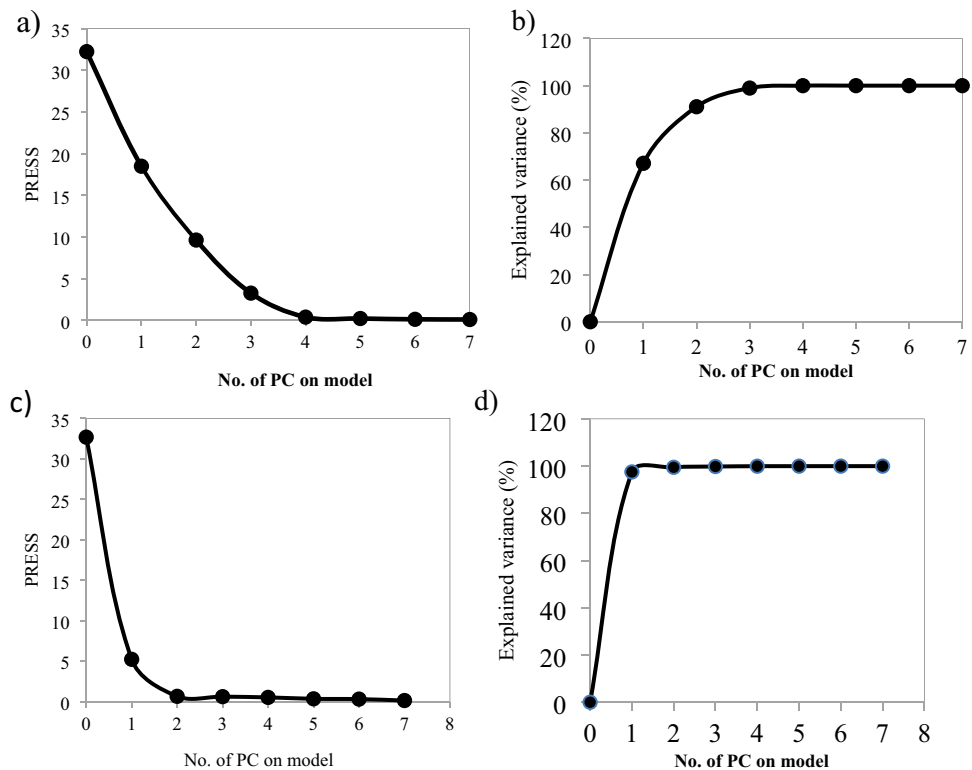
between the real and predicted values. The sum of the squares of the discrepancies is called the predicted residual error of sum of squares (PRESS) value, and this was calculated for different principal component (PC) values (Rohman and Che Man 2011). Figure 8 shows that a stable RMSECV value was minimally obtained after four and two factors for adulteration by soya and sunflower oil, respectively. Based on these results, we can suggest that the proposed models might be of practical use for detecting and

quantifying soya and sunflower oil adulteration in argan oil at levels as low as 3%.

Conclusions

Using a valuable database of 78 authentic argan oils, this study has demonstrated that FTIR spectroscopy combined with multivariate analysis—including principal component

Fig. 8 Predicted residual error sum of squares and explained variance versus number of PCs in the SO **a** and **b** and SF **c** and **d** model



analysis (PCA), soft independent modeling of class analogy (SIMCA), and data-driven soft independent modelling of class analogy (DD-SIMCA)—represents a potentially reliable tool for classifying argan oil samples according to their geographical origin. In addition, it also appears to be a good candidate for rapidly detecting and identifying soya and sunflower oil adulterants in argan oil products for quality control purposes. Moreover, once adulteration has been detected, the degree of adulteration with soya and sunflower oil can be quantified through partial least squares regression, which gives a good model for calibration, validation, and prediction within acceptable errors. The developed PLS and PCR calibration models could be successfully used to accurately determine a low concentration of soya or sunflower oil in argan oil. Being able to distinguish between argan oils and other cheaper oils will help reinforce the quality and commercial value of this plant product.

Acknowledgements The authors wish to thank the women's cooperatives markets in Ait-Baha, Agadir, Essaouira, Tiznit, and Taroudant for providing the argan oil samples. We also thank Mr. Abdelwahed Elaazou and Hassan Allali from AMIGHA (Association Marcaine Indication Geographies) for their support in providing cooperative and geological information. This research was jointly funded through the FAO/IAEA Joint Division's Coordinated Research Project D52040 and CNESTEN.

Data Availability Statement The datasets generated and/or analyzed for the current study are available from the corresponding author on reasonable request.

Declarations

Consent to Participate Informed consent was not applicable.

Conflict of Interest F. Taous declares no competing interests. T. El Ghali declares no competing interests. H. Marah declares no competing interests. K. Laraki declares no competing interests. M. Islam declares no competing interests. A. Cannavan declares no competing interests. S. Kelly declares no competing interests.

References

- Asian Wold Svante and Sjöström Michael (1977) SIMCA: a method for analyzing chemical data in terms of similarity and analogy. ACS Publications 1947–5918. <https://doi.org/10.1021/bk-1977-0052.ch012>
- Aabd NA, El Asbahani A, El Alem Y, El Finti A, Msanda F, El Mousadik A (2013) Variation in oil content and fatty acid composition in preselected argan trees with morphological characters and geographical localization. *Mediterr J Nutr Metab* 6(3):217–225
- Addou S, Fethi F, Chikri M, Rhioua A (2016) Detection of argan oil adulteration with olive oil using fluorescence spectroscopy and chemometrics tools. *J Mater Environ Sci* 7(8):2689–2698
- Basri KN, Hussain MN, Bakar J, Sharif Z, Khir MFA, Zoolfakar AS (2017) Classification and quantification of palm oil adulteration via portable NIR spectroscopy. *Spectrochim Acta Part A Mol Biomol Spectrosc* 173:335–342
- Bevilacqua M, Nescatelli R, Bucci R, Magrì AD, Magrì AL, Marini F (2014) Chemometric classification techniques as a tool for solving problems in analytical chemistry. *J AOAC Int* 97(1):19–28

- Bragolusi M, Massaro A, Zacometti C, Tata A, Piro R (2021) Geographical identification of Italian extra virgin olive oil by the combination of near infrared and Raman spectroscopy: a feasibility study. *J near Infrared Spectrosc* 29(6):359–365
- Çelik SE, Asfoor A, Şenol O, Apak R (2019) Screening method for argan oil adulteration with vegetable oils: an online HPLC assay with postcolumn detection utilizing chemometric multivariate analysis. *J Agric Food Chem* 67(29):8279–8289
- Charrouf Z, Guillaume D (2008) Argan oil: occurrence, composition and impact on human health. *Eur J Lipid Sci Technol* 110(7):632–636
- Cherki M, Berrougui H, Drissi A, Adlouni A, Khalil A (2006) Argan oil: which benefits on cardiovascular diseases? *Pharmacol Res* 54(1):1–5
- De Maesschalck R, Candolfi A, Massart D, Heuerding S (1999) Decision criteria for soft independent modelling of class analogy applied to near infrared data. *Chemom Intell Lab Syst* 47(1):65–77
- Drissi A, Girona J, Cherki M, Godàs G, Derouiche A, El Messal M, Saile R, Kettani A, Solà R, Masana L (2004) Evidence of hypolipemiant and antioxidant properties of argan oil derived from the argan tree (*Argania spinosa*). *Clin Nutr* 23(5):1159–1166
- Du Q, Zhu M, Shi T, Luo X, Gan B, Tang L, Chen Y (2021) Adulteration detection of corn oil, rapeseed oil and sunflower oil in camellia oil by in situ diffuse reflectance near-infrared spectroscopy and chemometrics. *Food Control* 121:107577
- El Monfalouti H, Guillaume D, Denhez C, Charrouf Z (2010) Therapeutic potential of argan oil: a review. *J Pharm Pharmacol* 62(12):1669–1675
- Elgadi S, Ouhammou A, Taous F, Zine H, Papazoglou EG, Elghali T, Amenzou N, El Allali H, Aithhaj A, El Antari A (2021) Combination of stable isotopes and fatty acid composition for geographical origin discrimination of one argan oil vintage. *Foods* 10(6):1274
- Farres S, Srata L, Fethi F, Kadaoui A (2019) Argan oil authentication using visible/near infrared spectroscopy combined to chemometrics tools. *Vib Spectrosc* 102:79–84
- Gharby Said, Charrouf Zoubida (2021) Argan oil: chemical composition, extraction process, and quality control. *Frontiers in Nutrition* 8:2296–861X
- González A, Armenta S, De la Guardia M (2010) Adulteration detection of argan oil by inductively coupled plasma optical emission spectrometry. *Food Chem* 121(3):878–886
- Guillaume D, Charrouf Z (2011) Argan oil. *Altern Med Rev* 16(3):275–279
- Guillen MD, Cabo N (1997) Characterization of edible oils and lard by Fourier transform infrared spectroscopy. Relationships between composition and frequency of concrete bands in the fingerprint region. *J Am Oil Chem Soc* 74(10):1281–1286
- Gunning Y, Jackson AJ, Colmer J, Taous F, Philo M, Brignall RM, El Ghali T, Defernez M, Kemsley EK (2020) High-throughput screening of argan oil composition and authenticity using benchtop 1H NMR. *Magn Reson Chem* 58(12):1177–1186
- Hilali M, Charrouf Z, Soulhi AEA, Hachimi L, Guillaume D (2007) Detection of argan oil adulteration using quantitative campesterol GC-analysis. *J Am Oil Chem Soc* 84(8):761–764
- Joshi R, Cho B-K, Joshi R, Lohumi S, Faqeerzada MA, Amanah HZ, Lee J, Mo C, Lee H (2019) Raman spectroscopic analysis to detect olive oil mixtures in argan oil. *Korean J Agric Sci* 46(1):183–194
- Kharbach M, Kamal R, Marmouzi I, Barra I, Cherrah Y, Alaoui K, Vander Heyden Y, Bouklouze A (2019) Fatty-acid profiling vs UV-Visible fingerprints for geographical classification of Moroccan Argan oils. *Food Control* 95:95–105
- Kharbach M, Kamal R, Bousrabat M, Mansouri MA, Barra I, Alaoui K, Cherrah Y, Vander Heyden Y, Bouklouze A (2017) Characterization and classification of PGI Moroccan Argan oils based on their FTIR fingerprints and chemical composition. *Chemom Intell Lab Syst* 162:182–190
- Kharbach M, Kamal R, Mansouri MA, Marmouzi I, Viaene J, Cherrah Y, Alaoui K, Vercammen J, Bouklouze A, Vander Heyden Y (2018) Selected-ion flow-tube mass-spectrometry (SIFT-MS) fingerprinting versus chemical profiling for geographic traceability of Moroccan Argan oils. *Food Chem* 263:8–17
- Lopez LC, Cabrera-Vique C, Venegas C, Garcia-Corzo L, Luna-Sánchez M, Acuna-Castroviejo D, Escames G (2013) Argan oil-contained antioxidants for human mitochondria. *Nat Prod Commun* 8(1):1934578X1300800111
- Madawala S, Kochhar S, Dutta P (2012) Lipid components and oxidative status of selected specialty oils. *Grasas y Aceites* 63(2):143–151
- Maleš Ž, Mišković G, Bojić M, Ćurak I (2018) Arganovo ulje. *Farmaceutski glasnik* 74(11): 817–824
- Miklavčič MB, Taous F, Valenčič V, Elghali T, Podgornik M, Strojnik L, Ogrinc N (2020) Fatty acid composition of cosmetic argan oil: provenience and authenticity criteria. *Molecules* 25(18):4080
- Mohammed F, Guillaume D, Abdulwali N, Harhar H, Al-Jobory HJ (2020) Argan Oil element content is a powerful marker of the quality of the fruit used for its preparation. *Plant Foods Hum Nutr* 75(2):230–235
- Mohammed FA, Bchitou R, Bouhaouss A, Gharby S, Harhar H, Guillaume D, Charrouf Z (2013) Can the dietary element content of virgin argan oils really be used for adulteration detection? *Food Chem* 136(1):105–108
- Oliveri P, Malegori C, Mustorgi E, Casale M (2021) Qualitative pattern recognition in chemistry: theoretical background and practical guidelines. *Microchem J* 162:105725
- Ourracha I, Radab M, Pérez-Caminob M, Benaissaa M, Guindab A (2012) Detection of argan oil adulterated with vegetable oils: new markers. *Grasas y Aceites* 63:4
- Oussama A, Elabadi F, Devos O (2012) Analysis of argan oil adulteration using infrared spectroscopy. *Spectrosc Lett* 45(6):458–463
- Pomerantsev Alexy (2014) *Chemometrics in excel*, John Wiley & Sons 336:978-1-118-60535-6
- Pomerantsev AL, Rodionova OY (2014) Concept and role of extreme objects in PCA/SIMCA. *J Chemom* 28(5):429–438
- Prieto N, Pawluczyk O, Dugan MER, Aalhus JL (2017) A review of the principles and applications of near-infrared spectroscopy to characterize meat, fat, and meat products. *Appl Spectrosc* 71(7):1403–1426
- Rohman A, Che Man Y (2011) Simultaneous quantitative analysis of two functional food oils, extra virgin olive oil and virgin coconut oil using FTIR spectroscopy and multivariate calibration. *Int Food Res J* 18(4):1231–1235
- Ascensión Rueda, Isabel Seiquer, Manuel Olalla, Rafael Giménez, Luis Lara, Carmen Cabrera Vique (2014) Characterization of Fatty Acid Profile of Argan Oil and Other Edible Vegetable Oils by Gas Chromatography and Discriminant Analysis. *J Chem Article ID 843908*, 8 pages, 2014, 2090–9063. <https://doi.org/10.1155/2014/843908>
- Salghi R, Armbruster W, Schwack W (2014) Detection of argan oil adulteration with vegetable oils by high-performance liquid chromatography–evaporative light scattering detection. *Food Chem* 153:387–392
- Simões T, Ferreira J, Lemos MF, Augusto A, Félix R, Silva SF, Ferreira-Dias S, Tecelão C (2021) Argan oil as a rich source of linoleic fatty acid for dietetic structured lipids production. *Life* 11(11):1114
- Sota-Uba I, Bamidele M, Moulton J, Booksh K, Lavine BK (2021) Authentication of edible oils using Fourier transform infrared spectroscopy and pattern recognition methods. *Chemom Intell Lab Syst* 210:104251
- Sour S, Belarbi M, Sari N, Benammar C, Baghdad C, Visioli F (2015) Argan oil reduces, in rats, the high fat diet-induced metabolic effects of obesity. *Nutr Metab Cardiovasc Dis* 25(4):382–387

- Sour S, Belarbi M, Khaldi D, Benmansour N, Sari N, Nani A, Chemat F, Visioli F (2012) Argan oil improves surrogate markers of CVD in humans. *Br J Nutr* 107(12):1800–1805
- Stokes TD, Foteini M, Brownfield B, Kalivas JH, Mousdis G, Amine A, Georgiou C (2018) Feasibility assessment of synchronous fluorescence spectral fusion by application to argan oil for adulteration analysis. *Appl Spectrosc* 72(3):432–441
- Vanstone N, Moore A, Martos P, Neethirajan S (2018) Detection of the adulteration of extra virgin olive oil by near-infrared spectroscopy and chemometric techniques. *Food Qual Saf* 2(4):189–198
- Zhe Yuan, Liangxiao Zhang, Du Wang, Jun Jiang, Peter de B. Harrington, Jin Mao, Qi Zhang, Peiwu Li (2020) Detection of flaxseed oil multiple adulteration by near-infrared spectroscopy and nonlinear one class partial least squares discriminant analysis. *LWT-Food Science and Technology* 125:109247

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.