Egypt. J. Chem. **Vol. 64,** No. 6 pp. 2765 - 2774 (2021)

Egyptian Journal of Chemistry

http://ejchem.journals.ekb.eg/

Classification of the geographical origin of argan kernels by infrared spectroscopy and chemometrics

Abdelkhalek Oussama, [a] Fatiha Elabadi [a], Noureddine Barka, [b] Fatima Zahra Mahjoubi, [a,*] Fouzia Kzaiber, [c]

[a] *Université sultan Moulay Slimane, Laboratoire de Spectro-chimiometrie appliquée et Environnement, Faculté des Sciences et techniques de Beni Mellal, Morocco*

[b] *Sultan Moulay Slimane University of Beni Mellal, Research Group in Environmental Sciences and Applied Materials (SEMA), FP Khouribga, B.P.145, 25000 Khouribga, Morocco.*

c Université sultan Moulay Slimane, Ecole Supérieure de Technologie de Beni Mellal, Laboratoire d'Ingénierie et Technologies Appliquées (LITA), Morocco

Abstract

The main objective of this study is the use of Fourier transform mid-infrared (FTIR) spectroscopy for discrimination of samples of Argan kernels from three Moroccan regions. The content of Argan oil extracted from the kernel depends on the plant genotype and on the environmental conditions. The origin of argan kernel samples could be distinguished combining the IR spectra measurement and chemometric tools for discriminant analysis. A multivariate analysis procedure based on the combined use of Hierarchical Cluster Analysis (HCA) and Partial Least Squares-Discriminant Analysis (PLS-DA) was tested for this purpose. HCA showed three distinct clusters related to the Moroccan regions and suggested the usefulness of IR spectroscopy to distinguish among kernel origins. Afterwards, PLS-DA was used for the discrimination and classification of the origin of the various argan kernels and good classification models and validation results were obtained. For 9 new unknown samples we attribuate 100% these samples by external validation. The results demonstrated that the combined use of FTIR and chemometric analysis (cluster analysis and discrimination by PLS- DA) can be used to rapidly and simply determine the origine of argan kernels, a relevant task for the agroalimentary Moroccan industry.

Keywords: FTIR; Discriminant analysis; Argan kernel ; Food authetification ; Cluster analysis; PLS-DA.

1. Introduction

Argan oil is made from the Argan tree (*Argania spinosa)*, endemic from southwestern Morocco protected by UNESCO. Argan oil is well known for its cosmetic, pharmaceutical and nutritional applications and has particularly unique organoleptic properties, related to its rich aroma and nutty flavour. Argan oil is considered a luxury product and is exported from Morocco around the world [1].

Authentication of food covers many different aspects, such as adulteration, characterisation, mislabelling and misleading origin. Food industry and regulatory bodies are compelled to put an increasing effort to authenticate the origin of foodstuff and to control the food constituents to ensure the sanitary safety of the product and to respond to the quality demands of the consumer [2, 3, 4]. The origin of argan oil is a relevant information that should be authenticated, since the climatology and conditions of the different regions provide particular properties to this kind of product, as it happens with olive oils. To decrease costs and risks, an important issue could be controlling the origin of the argan kernel before the oil production step to prevent non-autheticity problems before manufacturing the commercial argan oil.

Analytical methods used to characterize argan oil components often rely on chromatographic analysis. These separation techniques (gas chromatography, GC; with mass spectrometry, GC-MS; and high-performance liquid chromatography, HPLC) have

*Corresponding author e-mail: mahjoubi.fatimazahra@gmail.com

Receive Date: 19 December 2020, Revise Date: 09 January 2021, Accept Date: 21 March 2021

DOI: 10.21608/EJCHEM.2021.54218.3125

been extensively used to quantify fatty acids [5], as well as minor compound classes such as sterols, [6] phenolic compounds [7], and hydrocarbons [8] in oil samples. However, routine chromatographic analysis is often time consuming and expensive and does not give straightforward information on the global quality of the sample [9]. Therefore, alternative measurements cheaper, faster and non-destructive that can provide a quick answer on the origin of the sample are worth to be explored [10].

In this sense, mid-IR (MIR) spectroscopy has many applications because the fingerprints of many functional groups display a narrow, specific and intense signal in the MIR region (4000–400 cm^{-1}). Fourier transform infrared (FTIR) spectroscopy with attenuated total reflectance (ATR) or transmission cell accessories has been used to authenticate, identify and classify fats and oils [11]. Therefore, it seems to be a potential good technique for argan oil authentication. However, although studies on oil samples have been carried out, to our knowledge, there have been no attempts to use this technique to prevent future problems of authentication of argan oil by previous analysis of the argan kernel used to obtain this product.

Mid or near IR spectroscopy allows for fast monitoring of oil samples without the need of sample pretreatment. However, the use of this analytical measurement should go accompanied by the use of multivariate data analysis tools to provide good models of calibration or classification of samples that can be later used to predict oil component contents or origins of new unknown samples, respectively. Exploring the existence of natural clustering patterns related to the origin of foodstuff samples can be made by unsupervised methods, which identify the group objects on the basis of similarities among the samples [12, 13]. Common methods for unsupervised pattern recognition are Principal Component Analysis (PCA) and cluster analysis (CA), which can give a first insight about the possibility of using a certain measurement, e.g., FTIR spectra, to detect groups of samples according to a certain property; in this case, the origin of the argan kernel. If these groups are detected, classification tools, which build models that establish the relationship between the analytical measurement and the sample class (e.g., based on the origin) can be built, validated and used to predict the origin of new unknown samples. Diverse chemometric methods for classification have been successfully applied to the authentication of olive fruit and olive oil origin [14, 15, 16, 17, 18]. To comment a few [19] authenticated e.v. olive oil varieties using linear discriminant analysis (LDA) and back propagation artificial neural networks (BP-ANN), or [20] studied the visible and NIR spectra of e.v. olive oils and correctly classified 94% of the samples according to the geographical origin, using PLS-LDA and *k*-nearest neighbour method KNN. Therefore, the importance of this study is focused on the identification of the geographical origin and the determination of the authenticity of natural oils using simple and advanced methods.

The aim of the present work is to develop a new application of the FTIR technique combined with unsupervised and supervised classification methods as a rapid, inexpensive and nondestructive tool, useful to determine the geographic origin of argan kernels, later on used for argan oil manufacturing.
Verifying the declared origin or determining the origin of unidentified argan kernels is a relevant problem because of the increasing use and export of this natural product and the benefit that can report detecting fraud in authentication before the argan oil product is manufactured. The present work will test the combined use of FTIR spectroscopy with multivariate data analysis methods, such as Hierarchical Cluster Analysis (HCA) and PLS-DA for the authentication of the origin of argan kernels samples from three Moroccan regions: Agadir, Guelmim and Ait Baha.

2. Materials and methods

2.1. Samples

30 argan kernel samples were obtained from three Moroccan regions during the 2010 harvest season (10 samples per each region). The collection method (pulping, crushing of nut and obtaining the Argan kernels), the storage conditions and preservation were the same for all the samples used. Natural argan kernel samples were analyzed, without being subject to roasting treatments. Investigation was focused on argan samples picked in the zones named Ait Baha (AITB), Agadir (Agt) and Guelmim (Glm). (Fig. 1).

Geographic characteristics such as altitude, latitude and longitude as well as the mean rainfall of these ecoregions are summarized in Table1.

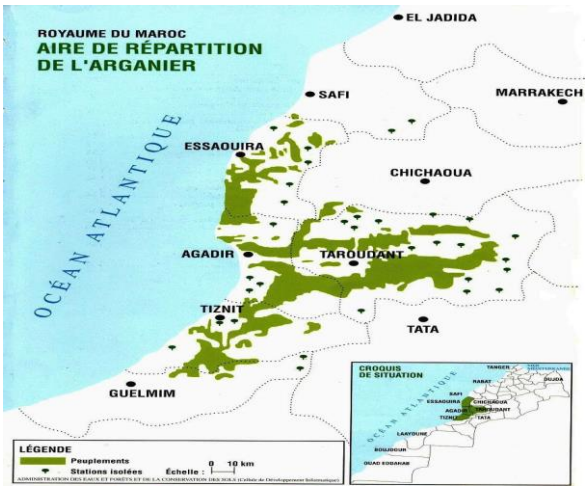


Fig. 1. Chart of the surface of distribution of the argan in Moroccan South-west

Table 1
Climate and geographical properties of the Moroccan regions of argan production.

Ecoregions	Latitude N	Longitude W	Altitude (m)	Rainfall (mm)
Agadir	30°41’	9°33’	150–350	300
Guelmim	28°58’	10°3’	286	217
Ait baha	29°33’	8°58’	340-1380	218

2.2. FT-IR analysis

The MIR measurements were acquired with a Bruker Vector 22 FTIR spectrometer. This instrument was equipped with an Attenuated Total Reflectance (ATR) sampling accessory (Diamant crystal) and a deuterated triglycine sulphate (DTGS) detector. The ATR crystal was carefully cleaned with pure ethanol to eliminate the presence of oil/fat residues between measurements to ensure a clean crystal surface to obtain the best possible sample spectra. Kernel samples were analysed and an FTIR spectrum per sample was recorded in the wavenumber range from 4000 to 600 cm^{-1}. A total sum of 104 scans was collected and averaged per each sample. The spectral resolution was $1cm^{-1}$. The OPUS (5.5) software program was employed for spectral data collection.

2.3. Data analysis

Two data subsets were prepared with the 30 FTIR spectra collected from the argan samples, a calibration set with 21 samples and a validation set with 9. The Kennard-Stone method was used for data subset construction [21]. Both calibration and validation sets were balanced, i.e., the same number of samples per class was selected. For the different data analysis performed, different pretreatments were tested on the FTIR spectra collected, such as Standard Normal Variate (SNV) and 1st and 2nd derivative by Savitzky-Golay method, with different options of data window and polynomial order [22, 23].

Hierarchical cluster analysis (HCA) was first applied to the calibration set to detect the possibility to distinguish natural groups of Argan kernel samples linked to the different Moroccan regions from the FTIR spectra measured. In HCA, each sample is initially assumed to be a lone cluster. Then, samples are joined sequentially to form clusters according to the distance among them. First, the two closest samples are joined to form a cluster and in subsequent hierarchical steps all the rest are clustered until a single group formed by all samples analyzed is formed. The final results are represented as a hierarchical tree structure with all the clustering steps represented the so-called dendrogram [12, 24]. In this work, the average linkage hierarchic agglomerative cluster algorithm with Euclidean distance was adopted to perform the analysis of the FTIR data.

Once seen the possibility to use FTIR spectra as a way to distinguish among argan kernel samples from different regions, Partial Least Squares-Discriminant Analysis (PLS-DA) was applied to build a classification model that could help to assign a geographic origin to unknown argan kernel samples. PLS-DA is a classification method based on the multivariate calibration PLS method, which looks for the model that relates **X** (matrix with predictor information) and **Y** (matrix with properties to be predicted) by using a set of latent variables, calculated to represent the directions of maximum covariance between **X** and **Y**. In the PLS-DA method, the **X** matrix contains the instrumental responses of the samples, e.g. the FTIR spectra, and the **Y** matrix the information on the sample class membership, e.g., origin of the argan kernel. In PLS-DA, samples belonging to a class are coded with 1 and samples out of the class with 0. PLS-DA aims at finding the relationship between **X** and **Y,** i.e., the relationship between the instrumental responses and the class membership of the samples [16, 25].

In this study, PLS-DA was applied using the PLS2 algorithm, which allows setting a **Y** matrix with several variables, one per each of the classes represented, Ait Baha, Guelmim and Agadir. For each class, the **y** variable is set equal to 1 for the samples belonging to the class and to 0 for those not belonging to the class.

The size of the PLS-DA model was chosen by full cross-validation. The number of latent variables was chosen as the best compromise between a low root mean square error of prediction on the cross-validated samples (RMSECV), and a low number of misclassified samples, always trying to keep the smallest possible model. The model was finally validated on an external set of nine validation samples. The UNSCRAMBLER 10 package (CAMO, Norway) was used for all data preprocessing and analysis.

3. Results and discussion

FTIR spectra of 30 argan kernel samples, produced in different locations of Morocco were recorded and divided into a calibration set of 21 samples and a prediction set of 9 samples, as mentioned in the previous section. Fig. 2 shows the mean spectra of the three classes studied. The differences among the FTIR spectra shown were small and occurred only in limited regions of the wavenumber range .The band of water around 2994 cm^{-1} has not been removed for this study. Peaks around 2800–3080 cm^{-1} are due to hydrogen stretching mode, the large peak between 1700 and 1800 cm^{-1} could be attributed to C=O stretching and C–O–C stretching and C–H bending in the region of 900–1400 cm^{-1} could be easily observed [26].

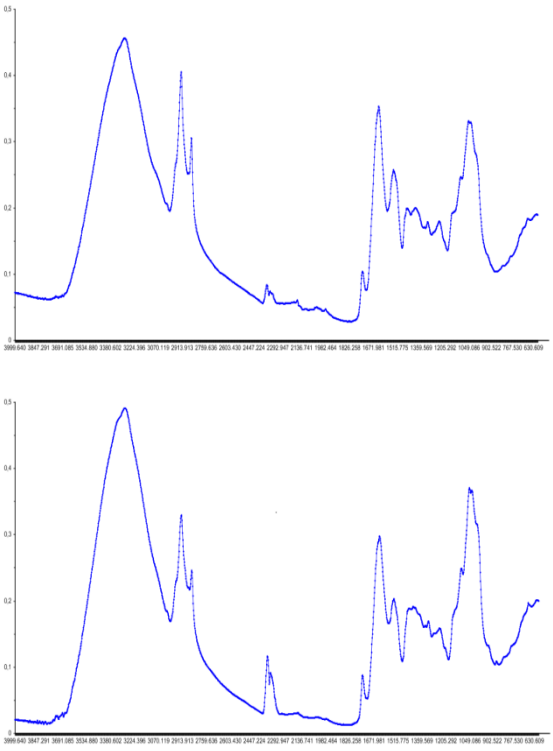


Fig. 2. Mean FTIR spectrum of Agadir, Ait baha and Guelmim, respectively.

With the aim to obtain more clear information from the FTIR spectral data, the spectra were first subject to mathematical pretreatment. In particular, Standard Normal Variate (SNV) and Derivative pretreatments were tested for HCA and PLS-DA purposes [22, 23, 27, 28].

3.1. Hierarchical Cluster Analysis

Hierarchical Cluster Analysis *(HCA)* is an unsupervised technique that uses the information obtained from measured variables to reveal natural clustering structures existing between the studied samples [12].

The results presented for cluster analysis have been obtained on first derivative FTIR spectra obtained applying the Savitzky-Golay method with a 2nd order polynomial and 5-point window smoothing [29]. Figure 3 shows the resulting dendrogram obtained from HCA using the average linkage hierarchic agglomerative cluster algorithm with Euclidean distance on the derivative FTIR spectra.

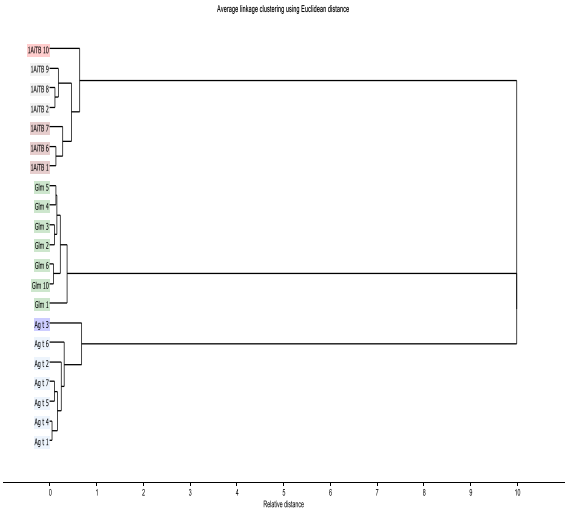


Fig. 3. HCA dendrogram of Ait baha (1AITB), Agadir (Agt), and Guelmim (Glm) argan class and below 0.5 as out of the class.

Three clusters can be clearly identified. The first group contains the samples collected in the territory of Ait baha, the second the samples from Guelmim, an area characterised by hilly uplands and high temperature range, and the last cluster consists of the samples picked in the territory of Agadir. The dendrogram shows that FTIR spectra allow for a very clear definition of the origin of the three types of argan kernels. As can be seen in the dendrogram, the three origins lead to very compact clusters, formed always

by samples from the same origin. These three clusters are clearly separated from one another.

The clear natural clustering pattern of the argan kernels from the three Moroccan origins seen in the dendrogram suggests the possibility to build classification models based on FTIR information from these kernels to predict the origin of unknown argan samples.

3.2. Classification model

For classification purposes, PLS-DA was applied to the set of 21 calibration samples with the aim of building a classification model useful for the later assignment of new argan kernels samples of unknown origin. A PLS2 algorithm was used with the **X** matrix of FTIR spectra and a **Y** matrix containing three variables, one per each class (argan kernel origin). Each class was defined with a binary **y**-variable, i.e., setting 1 for samples in the class and 0 for samples out of the class. Different data pretreatments along the row (spectra) direction were tested, such as Standard Normal Variate (SNV) and 1st and 2nd derivative with different polynomial orders and point window sizes. For these treatments, additional variable centering and autoscaling was also tested. The choice of the best classification model built on the calibration samples was based on the results obtained by full cross-validation, which are summarised in Table 2.

For each of the models built, global parameters such as the number of latent variables in the model and the **Y** explained variance are provided. For each of the three classes in the model: Ait Baha, (-aitb), Agadir (-agt) and Guelmim (-glm), figures such as the RMSEP (root mean square error of prediction) and the squared correlation coefficient between $\mathbf{y_{pred}}$ and $\mathbf{y_{measured}}$ (R^2CV) are given. In addition, the classification quality for each class within each model is evaluated by providing the number of misclassifications, separating false positives (samples identified as in the class, when they are not) and false negatives (samples identified as out of the class when they are in). For this purpose, a threshold value of 0.5 has been chosen and samples with a predicted y value above 0.5 are considered in the class and below 0.5 as out of the class. C and CN: column centered and column autoscaled data, respectively. For derivative pretreatment, e.g., 1D2O5S, 1D or 2D mean 1st or 2nd derivative, respectively, 2O means 2nd order polynomial and 5S, five points of window size.

Irrespective from the pretreatment used, the Agadir class is always the most difficult to model, which can be clearly seen through the RMSEP, r^2 value and the number of misclassifications (false positives and false negatives). The other two classes show satisfactory classification rates with almost all pretreatments tested. Trying to look for the best global results for all classes, the derivative pretreatment seems to be more appropriate than SNV or to the work with raw data, according to the values of RMSEP, R^2CV and number of misclassifications. The pretreatments along the rows, i.e., SNV or derivatives, seem to influence more the quality of the final results than column centering or autoscaling.

Within the different derivative options tested, Table 2 shows that the first derivative using second order polynomial, five points window and autoscaling provides the best results in terms of number of misclassifications, RMSEP and R^2CV and has a high **Y** variance explained. Other options, such as second derivative with second order polynomial and 7 or 9 points window with column centering are also appropriate alternatives in terms of classification rate but seem to be slightly worse when RMSEP and R^2CV are observed.

Focusing on the PLS-DA model finally selected, figure 4 shows the PLS-DA score plot for the calibration samples. Using two latent variables, 72% of Y variance is already explained. Three regions (group) can be clearly identified, each of them representing one type of argan kernels region. This result clearly shows the potential ability of this technique to discriminate among the three argan kernel origins used in this study. At the same time, it is worth noticing that the Agadir class is placed between the other two. This explains the higher difficulty in distinguishing these samples from the rest of the classes.

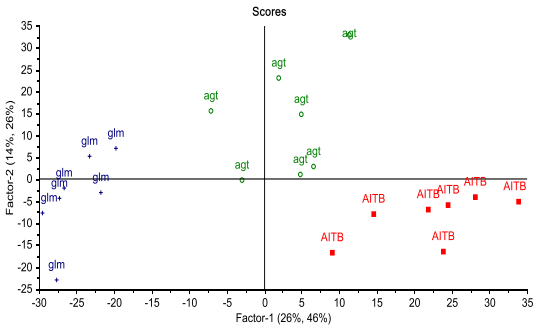
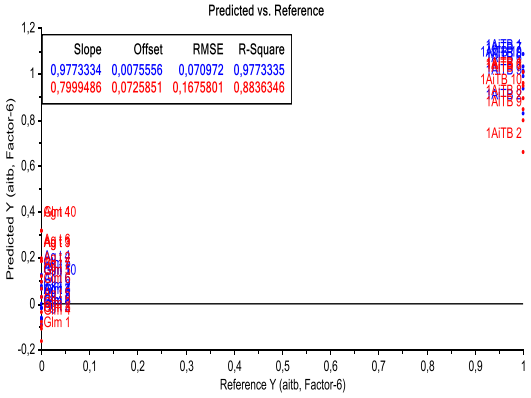


Fig. 4: Score plot of PLS-DA model for argan kernels of Ait baha (AITB), Agadir (agt) and Guelmim(glm) from autoscaled data pretreated with first derivative with five point window.

Table 2. Statistical parameters obtained from full cross-validation on the PLS-DA model.

Pretreatment(*)	Class	Number of factors	Y expl. var. (%)	RMSECV	R^2CV	False +	False -
Raw data (none)	-aitb	3	47.06	0.2524	0.7566	2	0
	-agt			0.4875	0.0215	2	3
	-glm			0.3236	0.6057	1	1
Raw data (CN)	-aitb	6	62.27	0.2266	0.7869	0	0
	-agt			0.3939	0.2940	1	4
	-glm			0.2360	0.7682	1	1
SNV C	-aitb	4	54.42	0.2457	0.7467	1	1
	-agt			0.4733	0.1644	4	2
	-glm			0.3011	0.7069	2	1
SNV CN	-aitb	6	63.54	0.2412	0.7672	0	1
	-agt			0.4098	0.3469	3	2
	-glm			0.2366	0.7859	0	0
2D2O5S C	-aitb	7	77.26	0.1879	0.8697	0	0
	-agt			0.3349	0.6422	2	1
	-glm			0.2366	0.8156	0	1
2D2O5S CN	-aitb	5	66.96	0.2431	0.7565	0	1
	-agt			0.3545	0.4680	2	1
	-glm			0.2235	0.7826	0	1
2D2O7S C	-aitb	7	76.46	0.1722	0.8910	0	0
	-agt			0.3079	0.6348	1	0
	-glm			0.2335	0.7711	0	1
2D2O7S CN	-aitb	5	67.26	0.2024	0.8285	0	0
	-agt			0.3702	0.4709	2	2
	-glm			0.2679	0.7226	0	1
2D2O9S C	-aitb	7	75.47	0.1660	0.8850	0	0
	-agt			0.2990	0.6351	1	0
	-glm			0.2367	0.7527	0	1
2D2O9S CN	-aitb	6	69.29	0.1820	0.8504	0	0
	-agt			0.3493	0.4818	2	2
	-glm			0.2411	0.7485	0	0
1D2O5S C	-aitb	7	64.39	0.2110	0.8129	0	0
	-agt			0.3807	0.4333	1	3
	-glm			0.2694	0.6977	0	1
1D2O5S CN	-aitb	6	80.73	0.1675	0.8836	0	0
	-agt			0.2850	0.6727	1	0
	-glm			0.1881	0.8626	0	0

Figure 5 shows the y predicted vs. y measured results for each one of the classes in this classification model. The red dots indicate the cross validation results for the calibration samples, which are the ones presented in Table 2. The best results are found for the Ait Baha class, where calibration and cross-validation results are very similar. Guelmimclass is also well modeled in the cross-validation results although the y predicted values are more spread (less compact in 'in and out' of the class groups) than the calibration results. The larger worsening when comparing calibration and cross-validation results in the Agadir class, as expected, where slope, RMSE and r^2 worsen significantly from calibration to validation results. Nevertheless, the classification rate is still satisfactory, although some of the samples remain very close to the class separation limit. Within the three classes, Agadir and Guelmim are mainly responsible for miss classifications being sample 2 of Agadir close to be a false – in its own class variable and a false in Guelmim class variable; the opposite situation occurs with sample 1 from Guelmim.



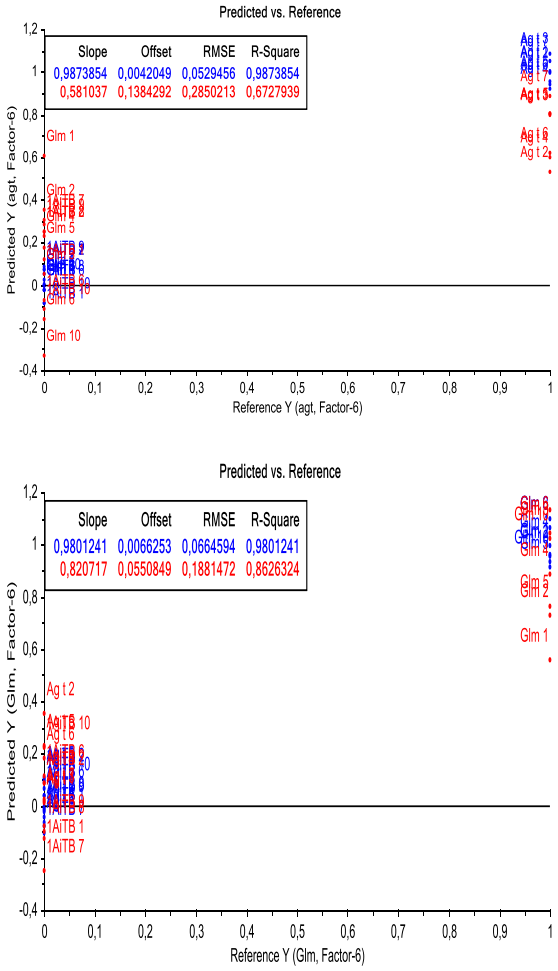


Fig. 5. Predicted value versus measured value of Ait baha, Agadir and Golmume of PLS –DA model.

Table 3: Classification results of argan kernel validation samples. The bold values show the samples that confirm the expected class membership.

	AIT BAHA		AGADIR		GUELMIM	
	Prediction	Reference	Prediction	Reference	Prediction	Reference
Ait b	**1.0026**	**1**	-0.2237	0	0.2211	0
Aitb	**0.9630**	**1**	0.0106	0	0.0262	0
Ait b	**1.0131**	**1**	-0.0174	0	0.0043	0
Agt	0.0203	0	**0.9604**	**1**	0.0192	0
Agt	0.0969	0	**0.5999**	**1**	0.3030	0
Agt	0.0223	0	**0.8913**	**1**	0.0862	0
Glm	-0.0144	0	0.0394	0	**0.9750**	**1**
Glm	0.0713	0	-0.3094	0	**1.2381**	**1**
Glm	0.0556	0	-0.2679	0	**1.2122**	**1**

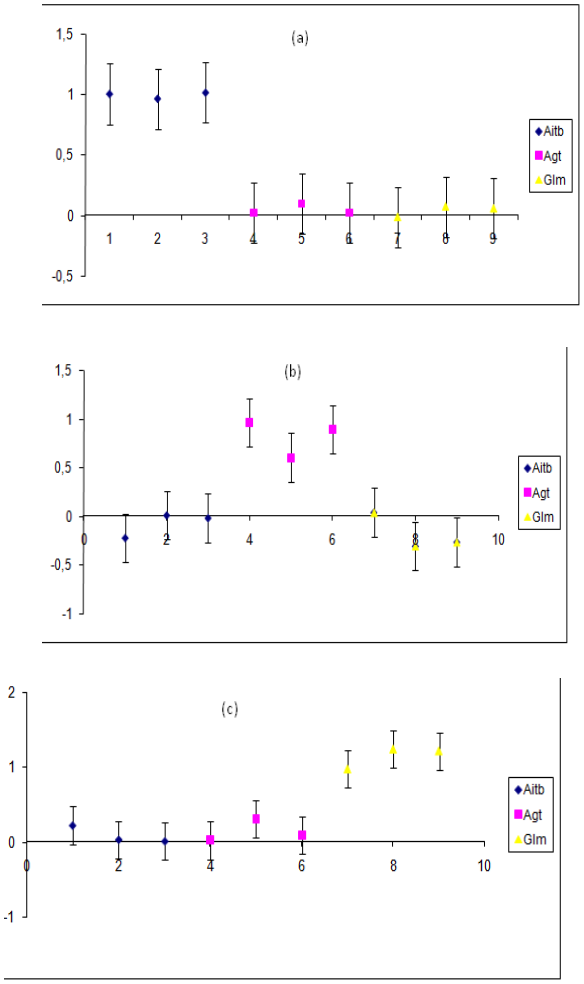


Fig. 6. Predictions for the validation set of argan pulpe: (a) class Ait baha; (b) class Agadir and (c) for class Guelmim.

3.3. Classification of new samples

The chosen classification model was applied on the FTIR external validation set, formed by nine new argan kernels samples, three from each class. This step allowed testing the ability of the built model to classify new samples into the classes previously established. Table 3 shows the classification results with the comparison between the predicted values and the expected theoretical value for each sample in each class variable.

Fig. 6a shows the classification results for the variable related to the Ait baha class. This is the best defined class and samples in the class are well above the threshold value of 0.5, even considering the error associated with the prediction, and samples out of the class are very well recognized too .Figure 6c shows the correct assignment of Guelmim samples into its class. Samples out of the class are classified correctly. In figure 6b, the results for the Agadir class are shown. Although the nominal y predicted values for the Agadir samples classify them in the class, sample 5 goes out of limits when the error associated with the prediction is considered. This confirms the high difficulty linked to the proper modeling of this class. Besides, samples out of the class show a higher spread in the y values predicted as well, showing the lower compactness within the sample groups in and out of the class.

4. Conclusion

In this paper, HCA and later PLS-DA were used in order to build a classification model to check the origin of argan kernels from FTIR measurements. The combined use of this analytical technique and adequate chemometric methods has proven to be an efficient strategy to discriminate the argan kernels according to their geographic origin without the need of further chemical analysis methods. The three regions studied could be clearly classified, although the Agadir region was more difficult to be modeled. This method can be proposed as a suitable alternative for quality control of argan kernels and determination of their geographical origin, with the special value that this authentication strategy can be performed before the manufacturing of the final Argan oil.

Conflicts of interest

There are no conflicts to declare.

References

[1] El Kharrassi Y., Maata N., Mazri M.A., El Kamouni S., Talbi M., El Kebbaj R., Moustaid K., Essamadi A., Andreoletti P., El Mzouri E., Cherkaoui-Malki M., and Nasser B., Chemical and phytochemical characterizations of argan oil (Argania spinosa L. skeels), olive oil (Olea europaea L. cv. Moroccan picholine), cactus pear (Opuntia megacantha salm-dyck) seed oil and cactus cladode essential oil. Food Measure, 12, 747–754(2018).

[2] Marcos-Lorenzo I., Pérez-Pavón J.L., Fernández-Laespada M.E., García-Pinto C., and Moreno-Cordeo B., Detection of adulterants in olive oil by headspacemass spectrometry. Journal of Chromatography A, 945(1-2), 221–230(2002).

[3] Sayazo A., Morales M.T., and Aparicio R., Detection of hazelnut oil in virgin olive oil by a spectrofluorimetric method. European Food Research and Technology, 218 (5), 480–483 (2004).

[4] Reid L.M., O'Donnell C.P., and Downey G., Recent technological advances for the determination of food

authenticity. Trends in Food Science and Technology, 17 (7), 344–353(2006).

[5] Aparicio R.,and Aparicio-Ruiz R., Authentication of vegetable oils by chromatographic techniques. J. Chromatogr,881(1–2), 93-104(2000).

[6] Bassbasi M., De Luca M., Ioele G., Oussama A., and Ragno G., Prediction of the geographical origin of butters by partial least square discriminant analysis (PLS-DA) applied to infrared spectroscopy (FTIR) data. Journal of Food Composition and Analysis, 33, 210–215(2014).

[7] Garcia A., Brenes M., Romero C., Garcia P., and Garrido A., Study of phenolic compounds in virgin olive oils of the Picual variety. Eur. Food Res. Technol, 215 (5): 407-412(2002).

[8] Bortolomeazzi R., Berno P., Pizzale L., and Conte L., S.Sesquiterpene, alkene, and alkane hydrocarbons in virgin olive oils of different varieties and geographical origins. J. Agric. Food Chem, 49 (7), 3278-3283(2001).

[9] Hilali M., Charrouf Z., Soulhi A., Hachimi L., and Guillaume D., Influence of Origin and Extraction Method on Argan Oil Physico-Chemical Characteristics and Composition. J. Agric. Food Chem, 53 (6), 2081-2087(2005).

[10] Hirri A., De Luca M., Ioele G., Balouki A., El Basbassi E., Kziaber F., Oussama A., and Ragno G., Chemometric Classification of Citrus Juices of Moroccan Cultivars by Infrared Spectroscopy. Czech J. Food Sci, 33(2), 137–142(2015).

[11] Jamwala R., Amita., Kumari S., Balan B., Dhaulaniya A.S., Kellyb S., Cannavanc A., and Singha D.K., Attenuated total Reflectance–Fourier transform infrared (ATR–FTIR) spectroscopy coupled with chemometrics for rapid detection of argemone oil adulteration in mustard oil. LWT - Food Science and Technology, 12,108945(2020).

[12] Massart D.L., and Kaufman L., The interpretation of analytical chemical data by the use of cluster analysis. New York, Wiley, (1983).

[13] Forina M., Oliveri P., Lanteri S., and Casale M., Class-modeling techniques, classic and new, for old and new problems. Chemometrics and Intelligent Laboratory Systems, 93(2), 132–148(2008).

[14] Dahlberg D. B., Lee S. M., Wenger S. J., and Vargo J. A., Classification of vegetable oils by FT-IR. Applied Spectroscopy, 51(8), 1118–1124(1997).

[15] Bendini A., Cerretani L., Di Virgilio F., Belloni P., Bonoli-Carbognin M., and LerckerG., Preliminary evaluation of the application of the FTIR spectroscopy to control the geographic origin and quality of virgin olive oils. Journal of Food Quality, 30 (4), 424–437(2007).

[16] De Luca M., Terouzi W., Ioele G., Kzaiber F., Oussama A., Oliverio F., Tauler R., and Ragno G., Dervative FTIR for cluster analysis and classification of moroccan oil olive. Food chemistry, 124 (3), 1113-1118(2011).

[17] Benamar, A., Mahjoubi, F.Z., Barka, N., Kzaiber, F., Boutoial, K., Ali, G.A.M., Oussama, A., Olive mill wastewater treatment using infiltration percolation in column followed by aerobic biological treatment. SN Applied Sciences 2, 655 (2020).

[18] Benamar, A., Mahjoubi, F., Ali, G.A., Kzaiber, F., Oussama, A., A chemometric method for contamination sources identification along the Oum Er Rbia river (Morocco). Bulgarian Chemical Communications 52, 159-171 (2020).

[19] Marini F., Balestrieri F., Bucci R., Magry A. D., Magry A. L., and Marini D., Supervised patternrecognition to authenticate Italian extra virgin olive oil varieties. Chemometric and Intelligent Laboratory System, 73 (1), 85–93(2004).

[20] Downey G., McIntyre P., and Davies A. N., Geographic classification of extra virgin olive oils from the eastern Mediterranean by chemometric analysis of visible and near-infrared spectroscopic data. Applied Spectroscopy, 57 (2), 158–163(2003).

[21] Kennard R.W., and Stone L.A., Computer aided design of experiments.Technometrics, 11 (1), 137-148(1969).

[22] Azzouz T., Puigdomènech A., Aragay M., and Tauler R., Comparison between different data pre-treatment methods in the analysis of forage samples using near-infrared diffuse reflectance spectroscopy and partial least-squares multivariate calibration method. Anal. Chim. Acta, 484 (1), 121-134(2003).

[23] Figueiredo N S., L H C Ferreira., and Dutra O O., An Approach to Savitzky–Golay Differentiators. Circuits, Systems, and Signal Processing, 38, 4369–4379 (2019).

[24] Jolliffe L.T., Allen O.B., and Christie B.R., Comparison of variety means using cluster analysis and dendrograms, Experimental Agriculture, 25 (2), 259-269(1989).

[25] Barker M., and Rayens W., Partial least squares for discrimination. Journal of Chemometrics, 17 (3), 166–173(2003).

[26] Tay A., Singh R.K., Krishnan S.S., and Gore J. P., Authentication of olive oil adulterated with vegetable oils using Fourier transform infrared spectroscopy. LWT-Food Science and Technology, 35(1), 99–103(2002).

[27] Iñón F. A., Garrigues J. M., Garrigues S., Molina A., and de la Guardia M., Selection of calibration set samples in determination of olive oil acidity by partial least squares-attenuated total reflectance-Fourier transform infrared spectroscopy. Analytica Chimica Acta, 489(1), 59–75(2003).

[28] Maggio R. M., Kaufman T. S., Del Carlo M., Cerretanil L., Bendini A., and Cichelli A., Monitoring of fatty acid composition in virgin olive oil by Fourier transformed infrared spectroscopy coupled with partial least squares. Food Chemistry, 114 (4), 1549–1554(2009).

[29] DeNoyer L K., and Jack G. Dodd J G., Smoothing and Derivatives in Spectroscopy. Handbook of Vibrational Spectroscopy,2173-2184 (2006).