

Characterization And Discrimination of Monocultivar Loquat (Eriobotrya Japonica) Fruit From two Areas in Beni Mellal-Khenifra Region

W.Terouzi, A.Oussama¹

¹Laboratory of Spectro- Chemometrics and environment, Faculty of Science and Technology of Beni Mellal, University of Sultan Moulay slimane, Beni Mellal, Morocco

Abstract: The goal of this study was to attempt characterization and classification of moroccan loquat (*Eriobotrya japonica*) according to their geographical origin based on the seed analysis by using Attenuated Total Reflectance Fourier Transform InfraRed (ATR-FTIR) spectroscopy coupled to chemometric tools and find the markers of their authenticity and discrimination. This work was focused on the loquat samples picked in two zones, named Aine-soumaa and Ourbia-foughal, in Beni mellal-Khenifra region, Morocco. The cracterization and classification models were developed by principal component analysis (PCA), support vector machines (SVMs) and linear discriminant analysis (LDA). On the basis of a PCA, two distinct clusters were recognized. The SVMs and LDA procedures were then elaborated. The models resulted able to separate the two classes and classify new samples into the appropriate defined classes with a percentage prediction of 100%. This Result show the capability of ATR-FTIR and the important role of chemometric tools in developing accurate models to discriminate and identify rapidly loquats origin.

Keywords :- characterization, chemometric, discrimination, infrared spectroscopie, loquat.

I. INTRODUCTION

Loquat (*Eriobotrya japonica*) is an important sub-tropical fruit tree native to China [1], [2], [3], [4]. Recently, it has received great attention because of its medicinal uses [5], [6]. Also, It is becoming an important industry in several countries as China, United States, Australia, Spain, Japan, India, Pakistan, Turkey and Lebanon [7], [8].

From a botanical point of view, two types of loquat are distinguishable : the « Chines type », characterized by large fruits that are pear-shaped and yellow-fleshed and the « Japanese type », charaterized by small fruits that are round-shaped and white or pal-yellow-fleshed . From these types a large number of varieties hase emerged in the different countries where loquat is grown [9]. Until now, there are no information about loquat varieties growing in Morocco. But according to morphological side, the « Japanese type » is the type of loquat that is very diffused in farms of Beni mellal-Khenifra region.

In the literature there are some references about the characterization and discrimination of loquat biodiversity [10], [11], [12]. However, this characterization and discrimination by MIR spectroscopy combined with chemometrics has not been reported so far, even though mid-infrared is a region used for quantitative and qualitative analysis of several products.

For qualitative analysis, the mid-IR (MIR) spectroscopy has more applications, because the 'fingerprints' of functional groups can be displayed narrowly and intensely 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 characterize, identify or classify fats and oils [13], [14], [15], [16], [17]. FTIR data have been often treated with multivariate analytical techniques to develop methods of classification and characterization, through the building of relative models. This approach has demonstrated to be very useful in many applications, due to the ability in achieving the spectral resolution of the FTIR signals [18], [19], [20], [21], [22].

In this context, the goal of this study is to identify geographical origin of loquat by using the ATR-FTIR data carried out from analysis of the loquat seed associated with chemometric treatment. This approach could represent a simple and easy technique can be used in research, quality control laboratories and agricultural-food industry. Additionally, to our knowledge, this is the first study on the characterization, discrimination and classification of Moroccan loquat from two different zones by ATR-FTIR.

II. MATERIALS AND METHODS

2.1. Sampling

The study area is confined to province of Beni-Mellal in central Morocco, expanded on a surface of nearly 7100 Km². Investigation was focused on monocultivar loquat samples, picked up at commercial maturity from two zones named Aine-soumaa (AS) and Ourbia-foughal (OS). We created two different collections of samples. The first includes 140 loquats were manually harvested from four farms belong to the two areas. The second consists of 133 loquats were randomly collected from same farms but 10 days before and after first collection harvest time.

273 loquats chosen from the harvested loquats based on their size, maturity and the absence of surface defects. Similar size loquats were chosen to minimize the effect of size on spectral measurements.

A series of 90 samples (from second collection) was used as an external validation set. This last series was used to assess the robustness of the SVM and PCA-LDA models. whilst the remaining 183 samples were selected to build up the calibration models.

The loquat samples were kept in cold storage (10°C) during the nights between the days of measurements. Spectroscopic measurements were taken from the loquats after they had been brought into equilibrium with the room temperature of 25°C.

2.2. ATR-FTIR analysis

ATR-FTIR spectra were obtained using a Vector 22 Bruker FTIR Spectrophotometer equipped with an attenuated total reflectance accessory (ATR single reflexion, Diamond, incident angle 45°, Pike Miracle, Pike Technologies, Madison, USA) with DTGS detector, Globar (MIR) Source and KBr Germanium separator, with a resolution of 4 cm⁻¹ at 80 scans. Spectra were scanned in the absorbance mode from 4000 to 600cm⁻¹ and the data were handled with OPUS logiciel. Seed of loquat samples were directly deposited between two well-polished KBr plates, without preparation on an Attenuated Total Reflectance cell provided with a diamond crystal. For ATR-FTIR measurements, it was necessary to keep a controlled pressure, to ensure good contact between the sample and the diamond surface. Analyses were carried out at room temperature (25°C). The background was collected before every sample was measured. Between spectra, the ATR plate was cleaned in situ by scrubbing with ethanol solution, enabling to dry the ATR.

2.3. Data pre-processing procedures

In this study, a series of pre-processing elaborations were tested on the spectral data prior to the multivariate calibration. The Savitzky–Golay [23] and Norris gap [24] algorithms were tested for data derivatisation. Standard normal variate (SNV) and multiple scatter correction (MSC) [25] were also tested. For data pre-treatment giving best result is the derivative function. In all PCA, SVM and PCA-LDA, second derivative through the Gap algorithm has been applied as preprocessing technique with centered data, in order to correct the spectrum by separating overlapping peaks and to enhance spectral differences.

2.4. Statistical treatment of data

2.4.1. Principal Component Analysis (PCA)

Principal component analysis (PCA) is an unsupervised technique commonly used for characterization and classification of data. It is based on the correlation among variables. It maps samples through scores and variables by the loadings in a new space defined by the principal components. The PCs are a simple linear combination of original variables. The scores vectors describe the relationship between the samples and allow checking if they are similar or dissimilar, typical or outlier, while the loadings vectors describe the importance of each variable [26], [27]. PCA provides a reduction in data set dimensionality and allows linear combinations of the original independent variables that are used to explain the maximum of data set variance [28].

2.4.2. Support Vector Machines (SVMs)

Support Vector Machines (SVMs) is a non-linear supervised learning technique [29]. It is kernel based learning algorithms introduced by Vapnik [30], [31].

The basic concept of SVMs is mapping the original data set into a high or infinite dimensional feature space, and then constructing a hyperplane or set of hyperplanes which can separate the classes of the training set and the unknown sample [29]. The transformation into higher-dimensional space is implemented by a kernel function [32]. Selection of kernel function has a high influence on the performance of the SVMs discrimination model [33]. In a two linearly separable class problem the principal aim of the SVM classifiers, is to find a separating “maximal margin” hyperplane which gives the smallest generalization error among the infinite number of possible hyperplanes. The data on margin and/or the closest ones are called support vectors.

2.4.3. Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) is a classification technique used for supervised pattern recognition using a linear function of variables [34], [35]. The resulting linear combination may be used as a linear classifier or for dimensionality reduction before classification. In a specific sense, the technique maximizes class separability, generating projections where the examples of each class from compact clusters and the different clusters are far from each other [36].

Generally, LDA has been used as supervised linear projection technique to find directions of maximum separation from a set of samples for which class membership is known in order to be able to predict the class membership of unknown samples [37], [38].

Moreover, when the number of variables larger than the number of observations, we cannot use LDA directly. In this work, a strategy of dimensionality reduction through PCA has been used previously to perform LDA in order to overcome the known tendency of this algorithm to overfitting in small-sample size problems, where the dimensionality is higher than the number of vectors. Thus, a combined PCA-LDA process was used as classification technique.

2.5. Software

The pre-treatment procedures and all chemometric models were performed by using the Unscrambler X software version 10.2 from Computer Aided Modelling (CAMO, Trondheim, Norway).

III. RESULTS AND DISCUSSION

3.1. Data acquisition

ATR-FTIR spectra of 273 seed samples were recorded and divided in two sets: a calibration set of 183 samples and an external validation set of 90 samples. A mean spectrum was calculated for each class of calibration set. The resultant mean spectra of two classes are shown in **Fig. 1**.

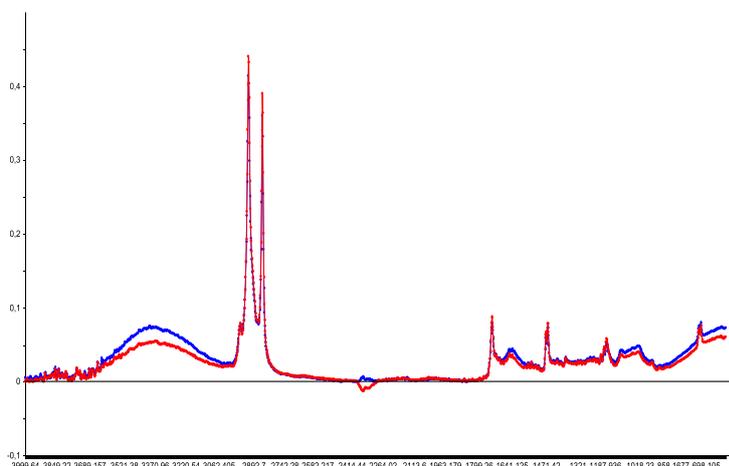


Fig. 1. The mean spectra calculated for each class: Aine-soumaa (AS) and Ourbia-foughal (OS)

Fig.1 shows the mean FTIR spectra of the studied loquats. The differences among them were clearly small and occurred only in limited regions of the spectra. The obtained spectra are dominated by significant bands of water are clearly visible in the seed spectra at 3400 cm^{-1} and 1640 cm^{-1} . The two bands at 2920 cm^{-1} and 2848 cm^{-1} are characteristic of fatty acids composition. The typical infrared pattern of sugar is observed in the region $1200 - 900\text{ cm}^{-1}$, while the range $2400 - 2300\text{ cm}^{-1}$ is due to CO_2 [39], [40], [41], [42].

The use of single peaks or narrow wavelength ranges to obtain information useful to distinguish the loquats according to their origins seemed very hard. These data were so conveniently handled by multivariate statistical techniques. With the aim to obtain more information from the ATR-FTIR spectral data, the spectra were firstly subjected to mathematical elaboration. In particular, derivative transformations were applied [43]. The best improvement in data variance was reached when the derivative function through the Gap algorithm was used. The derivative parameters were optimised and so fixed: 2nd order, gap size 17 with centered data. After this pre-treatment, the data were conveniently handled by appropriate multivariate statistical tools.

3.2. PCA modeling

PCA is a signal processing technique that generates projections along the directions of maximum variance of the analyzed data [37]. The model was built by the multivariate decomposition of the ATR-FTIR data in the ranges 4000–600 cm^{-1} , and validated by full cross-validation procedure [44].

Fig.2 shows the PC1/PC2 score plot, in which the first PC summarised all variation in the data accounting for 99% of the data variance. The two classes resulted perfectly separated from the other ones.

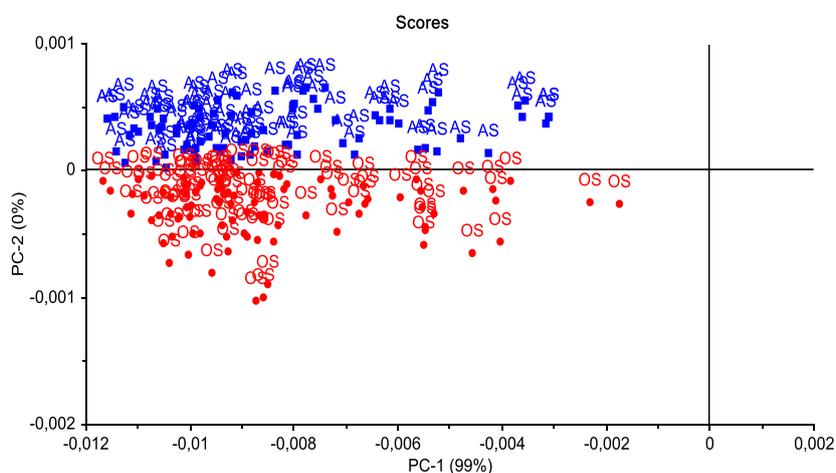


Fig. 2. PC1 / PC2 Score plot by PCA analysis on the calibration set : Aine-soumaa (AS) and Ourbia-foughal (OS)

Fig.3 show the PCA loadings of the first PC. Thus, we see on plot loadings that the first PC has 99% of the variance. In fact, the plotted PC1 loadings values of the spectral data is shown, representing the regions of the spectra where the differences between samples (of two classes) are more evident. The factorial contributions for discriminating the two classes along the positive axis of PC1 were associated with the spectral windows of 3000–2800 cm^{-1} , and the highest loadings are clearly found around 2915 cm^{-1} and 2848 cm^{-1} associated with fatty acids composition.

The results indicated that the particular chemical constituent, such as fatty acids contribute the strongest influences that explain the observed discrimination between the two classes of loquat samples on PCA plot.

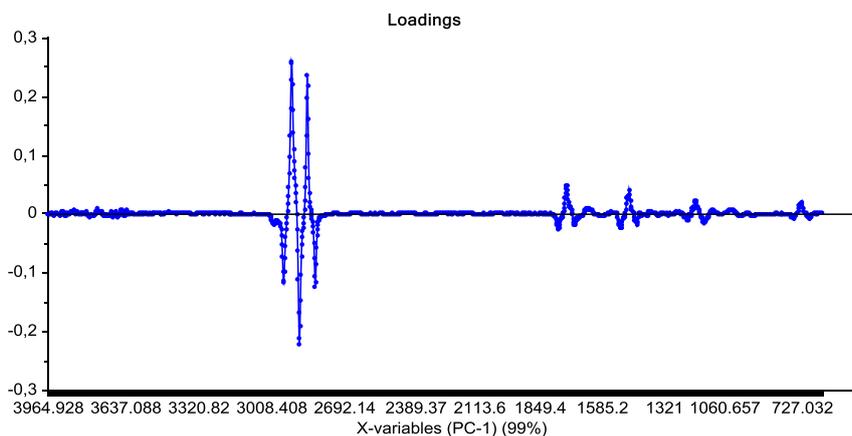


Fig.3. loadings plot of the first principal component (PC1)

3.3. Discrimination and classification models

3.3.1. SVM modeling

The SVMs model was built by considering, as X variables, all spectra; and the classification model was validated by Cross validation with segment = 100.

We constructed SVM model with a nu-SVC classification, different kernels have been tested on these data, and the results showed that the best choice is the linear kernel, to determine the hyperplane that give best separates the classes. The optimal parameter for “nu” which lies in the range 0-1, is then selected as the value that give the maximum correct classification rate, nu = 0.5. Consequently, a larger number of calibration samples are retained as support vectors, it is 100, where 51 of Aine-soumaa samples and 49 of Ourbia-foughal samples.

The main result of the SVMs is the confusion matrix, which indicates how many samples were classified in each class, and the prediction matrix, which indicates the classification determined for each sample in the calibration set. Look at the confusion matrix (**Table.1**). All the samples are perfectly classified.

Table1. Confusion matrix of calibration set, carried out by SVM

Confusion matrix	AS	OS
Predicted	1	2
AS	90	0
OS	0	93

Then, this result is confirmed by classification plot (**Fig.4**). In fact, based on the characteristic wavelengths of fatty acids (2915.842cm^{-1} ; 2848.345cm^{-1}), detected as significant variables responsible for the distinction between the two classes in loadings plot of PCA, all the classes resulted perfectly separated from the other ones.

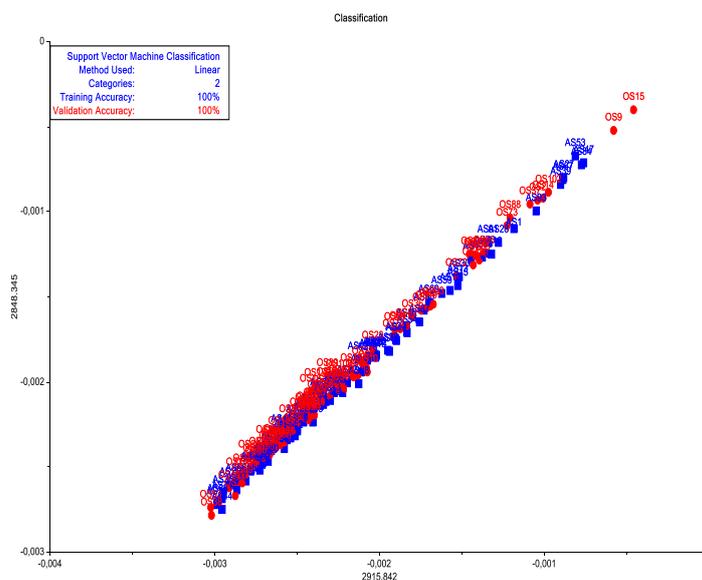


Fig.4. 2 D score plot of classification results by SVM on calibration, with 2 wavelenths $2915.842/2848.345\text{ cm}^{-1}$

3.3.2. Classification of new samples by SVMs model

In this step, the model was subduced to validation procedure by classifying the new samples in to the two classes previously established. The SVMs model was applied to ninety external loquat samples, the results are listed in **Table2**.

Table2 shows the classification results with the comparison between the predicted results of each class and the theoretical reference classes. The rate of correct classification was 100% within the external validation set.

Table2. Prediction result from application of the SVM classification model on the external validation set

Samples	Predicted	Reference class	Samples	Predicted	Reference class
VAS1	AS	AS	VAS25	AS	AS
VAS2	AS	AS	VAS26	AS	AS
VAS3	AS	AS	VAS27	AS	AS
VAS4	AS	AS	VAS28	AS	AS
VAS5	AS	AS	VAS29	AS	AS
VAS6	AS	AS	VAS30	AS	AS
VAS7	AS	AS	VAS31	AS	AS
VAS8	AS	AS	VAS32	AS	AS
VAS9	AS	AS	VAS33	AS	AS
VAS10	AS	AS	VAS34	AS	AS
VAS11	AS	AS	VAS35	AS	AS
VAS12	AS	AS	VAS36	AS	AS
VAS13	AS	AS	VAS37	AS	AS
VAS14	AS	AS	VAS38	AS	AS
VAS15	AS	AS	VAS39	AS	AS
VAS16	AS	AS	VAS40	AS	AS

Samples	Predicted	Reference class	Samples	Predicted	Reference class
VAS17	AS	AS	VAS41	AS	AS
VAS18	AS	AS	VAS42	AS	AS
VAS19	AS	AS	VAS43	AS	AS
VAS20	AS	AS	VAS44	AS	AS
VAS21	AS	AS	VAS45	AS	AS
VAS22	AS	AS	VOS1	OS	OS
VAS23	AS	AS	VOS2	OS	OS
VAS24	AS	AS	VOS3	OS	OS
VOS4	OS	OS	VOS25	OS	OS
VOS5	OS	OS	VOS26	OS	OS
VOS6	OS	OS	VOS27	OS	OS
VOS7	OS	OS	VOS28	OS	OS
VOS8	OS	OS	VOS29	OS	OS
VOS9	OS	OS	VOS30	OS	OS
VOS10	OS	OS	VOS31	OS	OS
VOS11	OS	OS	VOS32	OS	OS
VOS12	OS	OS	VOS33	OS	OS
VOS13	OS	OS	VOS34	OS	OS
VOS14	OS	OS	VOS35	OS	OS
VOS15	OS	OS	VOS36	OS	OS
VOS16	OS	OS	VOS37	OS	OS
VOS17	OS	OS	VOS38	OS	OS
VOS18	OS	OS	VOS39	OS	OS
VOS19	OS	OS	VOS40	OS	OS
VOS20	OS	OS	VOS41	OS	OS
VOS21	OS	OS	VOS42	OS	OS
VOS22	OS	OS	VOS43	OS	OS
VOS23	OS	OS	VOS44	OS	OS
VOS24	OS	OS	VOS45	OS	OS

3.3.3. PCA-LDA modeling

In this work, a combined PCA-LDA process was trained in the calibration step to build the classification model. PCA was applied on the ATR-FTIR spectra, to compress data and to transform the original data set comprising of a large number of inter-correlated variables into a reduced new set of variables [45], [46]. Then, the resulting score matrix was used to make the matrix of the predictors. This matrix and the vector of the membership information related to the loquat samples were then subdued to LDA procedure.

The score plot resulted to the use of the combined PCA-LDA strategy is shown in Fig. 5. The obtained PCA-LDA model was able to discriminate all samples of the calibration set with a correct classification of 100%. This results showed that the samples were perfectly discriminated on the basis of this processing method.

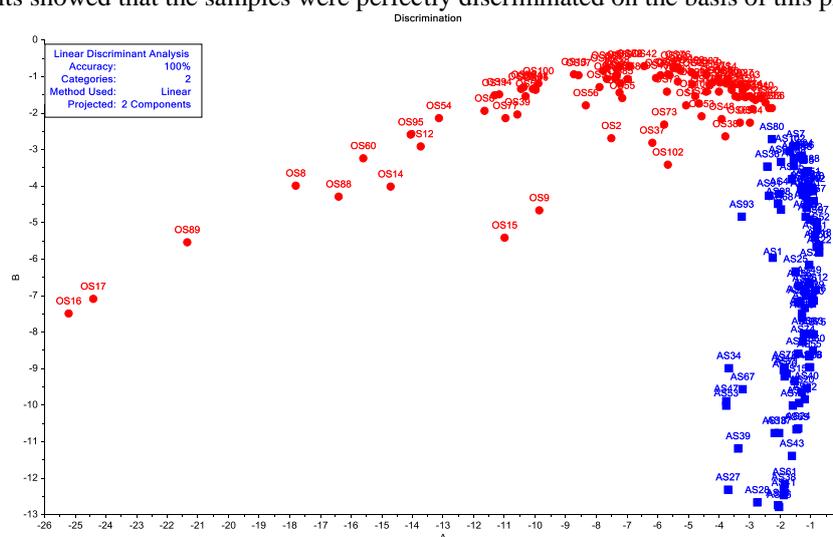


Fig.5. Discriminant value plot for classes AS and OS by calibration set : PCA-LDA application.

3.3.4. Classification of new samples by PCA-LDA model

The PCA-LDA model was applied to ninety external loquat samples. According with the LDA procedure, the prediction matrix (**Table3**) was carried out from the validation process and showed the discriminant values for each class, as well as the predicted class for each sample. The rate of correct classification was 100% within the external validation set.

Table.3. Prediction matrix from application of the PCA-LDA classification model on the external validation set

Samples	AS	OS	Prediction	Samples	AS	OS	Prediction
VAS1	-1.190807	-7.34562	AS	VAS11	-3.683443	-12.33093	AS
VAS2	-1.671835	-3.062262	AS	VAS12	-0.9526265	-7.061384	AS
VAS3	-0.8382414	-6.845025	AS	VAS13	-1.351736	-7.194109	AS
VAS4	-1.558897	-3.255969	AS	VAS14	-1.407449	-7.175128	AS
VAS5	-2.164558	-10.7693	AS	VAS15	-2.010325	-10.77141	AS
VAS6	-0.7204252	-5.624189	AS	VAS16	-3.355397	-11.2038	AS
VAS7	-1.045162	-6.170115	AS	VAS17	-1.115589	-9.548062	AS
VAS8	-1.143899	-4.594952	AS	VAS18	-0.9369592	-4.1452	AS
VAS9	-0.955425	-7.225554	AS	VAS19	-1.861282	-9.043835	AS
VAS10	-1.47424	-6.339116	AS	VAS20	-0.8929427	-7.143092	AS
VAS21	-1.247332	-4.018465	AS	VOS6	-5.898307	-9.764127	OS
VAS22	-1.203892	-3.650162	AS	VOS7	-4.388469	-0.969798	OS
VAS23	-1.08471	-3.962322	AS	VOS8	-8.726705	-9.404175	OS
VAS24	-3.742233	-10.01567	AS	VOS9	-10.98245	-5.421571	OS
VAS25	-1.032434	-8.672784	AS	VOS10	-24.40596	-7.097172	OS
VAS26	-0.8931674	-4.434026	AS	VOS11	-5.474787	-7.842327	OS
VAS27	-1.214143	-8.044779	AS	VOS12	-2.795118	-1.637993	OS
VAS28	-1.83206	-12.17839	AS	VOS13	-3.275799	-1.244627	OS
VAS29	-0.9960029	-8.040246	AS	VOS14	-3.735485	-1.194601	OS
VAS30	-1.445285	-10.68264	AS	VOS15	-6.909424	-7.214074	OS
VAS31	-2.243309	-5.969588	AS	VOS16	-3.577041	-1.376593	OS
VAS32	-1.167345	-4.282441	AS	VOS17	-5.660184	-9.621226	OS
VAS33	-0.8736442	-5.433309	AS	VOS18	-3.772399	-2.639599	OS
VAS34	-1.293902	-7.608792	AS	VOS19	-3.710821	-1.066691	OS
VAS35	-1.511645	-9.34979	AS	VOS20	-9.98174	-1.330098	OS
VAS36	-1.244763	-4.250546	AS	VOS21	-7.31164	-9.831648	OS
VAS37	-1.055324	-8.980998	AS	VOS22	-7.237949	-1.455114	OS
VAS38	-1.270348	-9.639318	AS	VOS23	-5.584757	-9.719366	OS
VAS39	-0.699946	-5.831461	AS	VOS24	-3.353991	-1.586085	OS
VAS40	-1.411625	-10.63767	AS	VOS25	-3.222529	-1.397448	OS
VAS41	-1.179268	-6.875184	AS	VOS26	-13.11095	-2.142332	OS
VAS42	-2.726782	-12.66771	AS	VOS27	-8.330622	-1.797246	OS
VAS43	-0.9597205	-4.056514	AS	VOS28	-6.9555	-1.082285	OS
VAS44	-3.650311	-8.996558	AS	VOS29	-11.62091	-1.947167	OS
VAS45	-2.401477	-3.471408	AS	VOS30	-3.46298	-1.550461	OS
VOS1	-2.626626	-1.599835	OS	VOS31	-4.87422	-9.099355	OS
VOS2	-10.06156	-1.3591	OS	VOS32	-14.04425	-2.590107	OS
VOS3	-3.157716	-1.292513	OS	VOS33	-2.359982	-1.86845	OS
VOS4	-2.87753	-1.907452	OS	VOS34	-7.881501	-1.297829	OS
VOS5	-4.396677	-1.427945	OS	VOS35	-21.32512	-5.546802	OS
VOS36	-4.396737	-0.9283931	OS	VOS41	-10.94303	-2.144359	OS
VOS37	-7.212031	-1.190207	OS	VOS42	-10.43313	-1.357667	OS
VOS38	-2.455892	-1.71728	OS	VOS43	-5.779212	-2.328318	OS
VOS39	-7.390841	-1.044403	OS	VOS44	-4.02602	-1.042499	OS
VOS40	-10.36022	-1.300999	OS	VOS45	-7.359269	-1.031445	OS

IV. CONCLUSIONS

This work proposes a new method for the qualitative analysis of loquat (*Eriobotrya japonica*) by applying Attenuated Total Reflectance-Fourier Transform Mid Infrared Spectroscopy (ATR-FTMIR) associated with chemometric techniques.

In fact, we arrived to characterize and discriminate loquats of two zones by differences in their seed ATR-FTIR spectra. MIR spectroscopy combined with multi-dimensional chemometric techniques PCA, SVM and PCA-LDA are successfully applied to the Characterization and classification of loquats according to their geographical origin (percentage of correct classification is 100%).

These results indicate that ATR-FTIR with chemometric techniques can be a useful tool for rapid control of loquat authenticity. Its application was rapid and simple because no chemical treatment of samples was required. Also, it can be used in food industry for the reliable, cheap and fast quality control of raw material.

REFERENCES

- [1]. Zhang, H.Z., Peng, S.A., Cai, L.H. The germplasm resources of the genus *Eriobotrya* with special reference on the origin of *E. japonica* Lindl. *Acta Hort. Sin.*, 17, 1990, 5–11.
- [2]. Zheng, S.Q. Achievement and prospect of loquat breeding in China. *Proceedings of the Second International Symposium on Loquat. Acta Hort.*, 750, 2007, 85–91.
- [3]. Huang, G.X., Pan, J.C., He, X.L., Yang, X.H., Lin., S.Q. A preliminary report of investigation on genus *Eriobotrya* plants in Guangxi and their characteristics. *Proceedings of the Second International Symposium on Loquat. Acta Hort.*, 750, 2007, 101–105.
- [4]. Polat, A.A., Caliskan, O. Loquat production in Turkey. *Proceedings of the Second International Symposium on Loquat. Acta Hort.*, 750, 2007, 49–53.
- [5]. Hamada, A., Yoshioka, S., Takuma, D., Yokota, J., Cui, T., Kusunose, M., Miyamura, M., Kyotani, S., Nishioka, Y. The effect of *Eriobotrya japonica* seed extract on oxidative stress in adriamycin-induced nephropathy in rats. *Biol. Pharm. Bull.*, 27 (12), 2004, 1961–1964.
- [6]. Kim, S.H., Shin, T.Y. Anti-inflammatory effect of leaves of *Eriobotrya japonica* correlating with attenuation of p38 MAPK ERK, and NF-kappa B activation in mast cells. *Toxicol. in Vitro*, 23, 2009, 1215–1219.
- [7]. Janick, J. Genetic alteration associated with fruit domestication. *Proceedings of the Second International Symposium on Loquat. Acta Hort.*, 750, 2007, 27–35.
- [8]. Badenes, M.L., Janick, J., Lin, S., Zhang, Z., Wang, W., Liang, G.L. Breeding loquat. In: Jules, Janick (Ed.), *Plant Breeding Reviews*, 37, 2013, 259–296.
- [9]. Badenes, M.L., Martínez-Calvo, J. & Liácer, G. Analysis of a germplasm collection of loquat (*Eriobotrya japonica* Lindl.). *Euphytica*, 114, 2000, 187–194.
- [10]. Ferreres, F., Gomes, D., Valentão, P., Gonçalves, R., Pio, R., Alves Chagas, E., M. Seabra, R., B. Andrade, P. Improved loquat (*Eriobotrya japonica* Lindl.) cultivars: Variation of phenolics and antioxidative potential. *Food Chemistry*, 114, 2009, 1019–1027
- [11]. Ercisli, S., Gozlekci, S., Sengul, M., Hegedus, A., Tepe, S. Some physicochemical characteristics, bioactive content and antioxidant capacity of loquat (*Eriobotrya japonica* (Thunb.) Lindl.) fruits from Turkey. *Scientia Horticulturae*, 148, 2012, 185–189
- [12]. Chalak, L., Noun, A., Youssef, H., Hamadeh, B. Diversity of loquats (*Eriobotrya japonica* Lindl.) cultivated in Lebanon as assessed by morphological traits. *Scientia Horticulturae*, 167, 2014, 135–144
- [13]. Lai, Y. W., Kemsley, E. K., & Wilson, R. H. Potential of Fourier transform infrared spectroscopy for the authentication of vegetable oils. *Journal of Agricultural and Food Chemistry*, 42, 1994, 1154–1159.
- [14]. Dahlberg, D. B., Lee, S. M., Wenger, S. J., & Vargo, J. A. Classification of vegetable oils by FT-IR. *Applied Spectroscopy*, 51, 1997, 1118–1124.
- [15]. Dupuy, N., Duponchel, L., Huvenne, J. P., Sombret, B., & Legrand, P. Classification of edible fats and oils by principal component analysis of Fourier transform infrared spectra. *Food Chemistry*, 57, 1996, 245–251.
- [16]. Ozen, B. F., Weiss, I., & Mauer, L. J. Dietary supplement oil classification and detection of adulteration using Fourier transform infrared spectroscopy. *Journal of Agricultural and Food Chemistry*, 57, 2003, 5871–5876
- [17]. Safar, M., Bertrand, D., Robert, P., Devaux, M. F., & Genot, C. Characterization of edible oils, butters and margarines by Fourier transform infrared spectrometry with attenuated total reflectance. *Journal of the American Oil Chemists Society*, 71, 1994, 371–377.
- [18]. Zagonel, G.F., Peralta-Zamora, P., Ramos, L.P. Multivariate Monitoring of Soybean Oil Ethanolysis by FTIR. *Talanta*, 63, 2004, 1021–1025
- [19]. De Luca, M., Terouzi, W., Ioele, G., Kzaiber, F., Oussama, A., Oliverio, F., Tauler, R., Ragno, G. Derivative FTIR spectroscopy for cluster analysis and classification of morocco olive oils. *Food Chem.* 2011, doi: 10.1016/j.foodchem.2010.07.010.
- [20]. De Luca, M., Terouzi, W., Kzaiber, F., Ioele, G., Oussama, A., Ragno, G. Classification of moroccan olive cultivars by linear discriminant analysis applied to ATR–FTIR spectra of endocarps. *International Journal of Food Science and Technology*, 47, 2012, 1286–1292.
- [21]. Pinheiro, P. B. M. and da Silva, J. Chemometric classification of olives from three Portuguese cultivars of *Olea europaea*. *L. Anal. Chim. Acta.*, 544, 2005, 229–235.
- [22]. Terouzi, W., De Luca, M., Bolli, A., Oussama, A., Patumi, M., Ioele, G., Ragno, G. A discriminant method for classification of Moroccan olive varieties by using direct FT-IR analysis of the mesocarp section, *Vib. Spectrosc.*, 2011, doi:10.1016/j.vibspec.2011.01.004.
- [23]. Savitzky, A. & Golay, M.J.E. Smoothing and differentiation of data by simplified least-squares procedures. *Analytical Chemistry*, 36, 1964, 1627–1639.

- [24]. Norris, K.H. & Williams, P.C. Optimization of mathematical treatments of raw near infrared signal in the measurement of protein in hard Red Spring wheat, I: influence of particle size. *Cereal Chemistry*, 62, 1984, 158–165.
- [25]. 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, 2003, 59–75.
- [26]. Ferreira, M. C. M., Morgano, M. A., Queiroz, S. C. N., & Mantovani, D. M. B. Relationships of the mineral and fatty acid contents in processed Turkey meat products. *Meat Science*, 69(3), 2000, 259–265.
- [27]. Granato, D., Castro, I.A., Katayama, F.U. Assessing the association between phenolic compounds and the antioxidant activity of Brazilian red wines using chemometrics. *LWT-Food Science and Technology*, in press, doi:10.1016/j.lwt.2010.05.031.
- [28]. Brown, S.D., Tauler, R., Walczak, B. *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis*, vol. 2. Elsevier, Amsterdam, 2009.
- [29]. Xu, L., Shao, X.G. *Chemometric Methods*. Beijing: Science Press, 2006, 130 -133.
- [30]. Vapnik, V., Chervonenkis, A. *Theory of Pattern Recogn*, Nauka, Moscow, 1974.
- [31]. Vapnik, V. *The nature of statistical learning theory*, Springer Verlag, 1995.
- [32]. Thissen, U. Comparing support vector machines to PLS for spectral regression applications. *Chemom Intell Lab Syst*, 73, 2004, 169–79.
- [33]. Sathiya Keerthi, S., Lin C-J. Asymptotic behaviors of support vector machines with Gaussian kernel. *Neural Comput*, 15, 2003, 1667–89.
- [34]. Massart, D. L., Vandeginste, B.G. M., Buydens, L. M. C., De Jong, S., Lewi, P. L., & Smeyers-Verbeke, J. *Handbook of chemometrics and qualimetrics: Part B*. Amsterdam, The Netherlands: Elsevier, 1998.
- [35]. Naes, T., Isaksson, T., Fearn, T., & Davies, T. A. *User-friendly guide to multivariate calibration and classification*. NIR Publication, 2002.
- [36]. Jain, A.K., Chandrasekaran, B. Dimensionality and sample size consideration in pattern recognition practice in *Handbook of Statistics*, Amsterdam, The Netherlands, 1982.
- [37]. Garrido-Delgado, R., Arce, L., Guamán, A.V., Pardo, A., Marco, S., Valcárcel, M. Direct coupling of a gas-liquid separator to an ion mobility spectrometer for the classification of different white wines using chemometrics tools. *Talanta* 84, 2011, 471–479.
- [38]. Gori, A., M. Maggio, R., Cerretani, L., Nocetti, M., Fiorenza Caboni, M. Discrimination of grated cheeses by Fourier transform infrared spectroscopy coupled with chemometric techniques. *International Dairy Journal* 23, 2012, 115-120.
- [39]. 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, 2003, 59–75.
- [40]. Hafidi, M., Amir, S., Revel, J. C. Structural characterization of olive mill waste-water after aerobic digestion using elemental analysis, FTIR and ¹³CNMR. *Process Biochem.* 40, 2005, 2615–2622.
- [41]. Droussi, Z., D’Orazio, V., Provenzano, M. R., Hafidi, M., Ouattmane, A. Study of the biodegradation and transformation of olive-mill residues during composting using FTIR spectroscopy and differential scanning calorimetry. *J. Hazard Mater.* 164, 2009, 1281–1285.
- [42]. Scibisz, I., Reich, M., Bureau, S., Gouble, B., Causse, M., Bertrand, D., Renard, C. Mid-infrared spectroscopy as a tool for rapid determination of internal quality parameters in tomato. *Food Chemistry*, 125, 2011, 1390–1397.
- [43]. Maggio, R. M., Kaufman, T. S., Del Carlo, M., Cerretani, L., Bendini, A., Cichelli, A., et al. Monitoring of fatty acid composition in virgin olive oil by Fourier transformed infrared spectroscopy coupled with partial least squares. *Food Chemistry*, 114, 2009, 1549–1554.
- [44]. Sathle, L. & Wold, S. Partial least squares analysis with crossvalidation for the two-class problem: a Monte Carlo study. *Journal of Chemometrics*, 1, 1987, 185–196.
- [45]. Scott, I.M., Lina, W., Liakata, M., Wood, J.E., Vermeer, C.P., Allaway, D., Ward, J.L., Draper, J., Beale, M.H., Corol, D.I., Baker, J.M., King, R.D. Merits of random forests emerge in evaluation of chemometric classifiers by external validation. *Analytica Chimica Acta*, 801, 2013, 22–33.
- [46]. Bansal, A., Vikas, C., K. Rawal, R., Sharma, S. Chemometrics: A new scenario in herbal drug standardization. *Journal of Pharmaceutical Analysis*, 2014, <http://dx.doi.org/10.1016/j.jpba.2013.12.001>.