International Journal of Chemical, Material and Environmental Research 2015, 2 (4): 30-36

*Research article*

INTERNATIONAL JOURNAL OF CHEMICAL, MATERIAL AND ENVIRONMENTAL RESEARCH

An International Open Free Access, Peer Reviewed Research Journal

ISSN 2410-5945
2015, Vol. 2, No. 4:
pg. 30-36

# Classification of Five Olive Varieties Introduced in Morocco Using Mid Infrared Spectroscopy Coupled with Chemometric Analysis

**A. Hirri[1*], A. Boulli[2], A. Oussama[1]**

[1]Laboratory of applied and environmental Spectrochemistry, Faculty of Science and Techniques, Sultan Moulay Slimane University, BeniMellal, Morocco
[2]Environment and Valorization of Agro resources, Faculty of Science and Techniques, Sultan Moulay Slimane University, BeniMellal, Morocco

(Received: November 02, 2015; Accepted: December 10, 2015)

## Abstract

Table olives and olive oils are foods that provide health benefits with great economic relevance. However, qualitative and quantitative aspects of these products can vary among varieties and cultivars. The high biodiversity of olive tree and the economic needs require tools for a rapid, cheap and correct classification and identification of the different varieties. Due to this, simple and rapid methods are in increasing demand. In the present work, FT-MIR spectroscopy associated to chemometric treatment is proposed as a direct and rapid tool to discriminate cultivars according to their olive leaves. A set of 125 samples of olive leaves representative of five varieties (Hojiblanca, Frantoio, Leccino, Manzanilla and Arbequina) cultivated in the same geographical area was analyzed. Discrimination between the five varieties was performed by the chemometric approach, principal component analysis (PCA), based on the FT-MIR spectra data provided by the olive leaves. Furthermore, a correct classification (100%) of the five varieties was obtained by the partial least square discriminate analysis (PLS-DA) method.

**Keywords**: Olive leaves, FT-MIR, principal component analysis, partial least square discriminate analysis, cultivars

## 1. Introduction

Olive trees have a great international importance, both socially and economically [1]. In the Mediterranean area, olive trees present multiple phenotypic expressions, commonly designed by varieties or cultivars that were selected and spread by men. Theses varieties or cultivars resulted of mutations and natural hybridizations manifest a high diversity of morphologic and physiologic characteristics that corresponds to different aptitudes and qualities [2]. Although 2600 different olive varieties and cultivars have been recorded, however the number of olive cultivars throughout the word is uncertain [3]. It is well know that qualitative and quantitative aspects of production of olive oils and table olives vary, among varieties and cultivars [4]. Nowadays, the identification of the olive cultivar become a topic of great economic relevance since the demand of table olives and olive oils is increasing and there is a growing commercial interest in high quality products. Furthermore, correct varieties and cultivar identification can represent a useful tool for nursery owners who need to certify and patent their plant material. Initially olive material discrimination and classification were based on morphological and agro-

* Corresponding author:
azizhirri@yahoo.fr
Published online at www.ijcmer.org

nomic traits [5] and then in using isoenzymes analysis [6]. Recently, chemometric methods have been applied for discrimination and classification of olive cultivars. The analyses have been carried out on a particular olive tissue such as leaves [7], olive fruits [2] or olive oil [8]. Morphological, biochemical and molecular analysis used for discriminating cultivars are time consuming as well as expensive and involve a considerable amount of manual work. Very often, complex chemical treatment of the sample and the use of sophisticated instruments are required. Recently, Fourier transform Infrared (FT-IR) spectroscopy, has become an emerging well accepted analytical technique, due to its simplicity with advantages in terms of cost per sample. It achieves high analysis speed and requires little or no sample preparation. It has been widely used as an analytical tool in various laboratories and industrial sectors [9] such as food, agricultural [10], petrochemical and pharmaceutical ones [11]. Until now, in the same way, a lot of studies have been done using near and mid FT-IR and multivariate statistical analysis of spectra data for authentication, identification or classification of agro-foods products, notably olive oils [12, 13] and table olives [14].

Morocco is a Mediterranean country in which olive tree is widely spread and in continuous use as an alternative culture regarding climate change. To reach their goals, farmers introduced many varieties and cultivars from others countries witch contribute to olive diversity in Morocco. The aim of this study is to develop a direct and rapid tool to discriminate olive tree planted in Morocco using FT-MIR spectroscopy associated to chemometric treatment. In this first step, five olive varieties originating from Spain and Italy cultivated in Kebbaj farm, Morocco were studied.

## 2. Materials and Methods

### 2.1. Olive Leaf Samples

A series of 125 olive leaves were sampled from Kebbaj farm, Marrakech (Morocco) during the harvested season 2014. The collection method, the storage condition and preservation were the same for all samples which were analysed just after collection by FT-IR in a range from 4000 $cm^{-1}$ to 600 $cm^{-1}$. A set of 95 samples was selected at random to build up the calibration model whilst the other 30 samples were used as an external validation.

### 2.2. FT-IR Analysis

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 degree), DTGS detector, Globar (MIR) source and KBr Germanium separator, with a resolution of 4 $cm^{-1}$ at 98 scans. Spectra were scanned in the absorbance mode from 4000 to 600 $cm^{-1}$. Analyses were carried out at room temperature. The background spectrum was collected before each sample was measured.

### 2.3 Chemometric Methods

Two chemometric tools were applied in this study: the principal components analysis (PCA) and partial last square discriminate analysis (PLS-DA). PCA and PLS-DA regression were performed using the Unscrambler version 10.2.

#### 2.3.1. Principal Component Analysis

PCA allows to reducing the analytical data to a few latent variables (PCs) that represent the main information from the original data. The first PCs carry most information whereas, after a number of PCs, the variance can be discarded because of the noise [15].

#### 2.3.2. Partial Least Square Discriminate Analysis

The PLS-DA was applied to find what were the variables which better discriminate between different groups of samples from their FT-IR spectra (X block) according to their maximum covariance with a target class defined in a class pertinence variable (y data block)[16]. It attempts to describe whether a spectrum of a sample belongs or not to a particular class, consisting of zeros and ones. According to the number of simultaneously regressed y vectors, two different PLS-DA approaches are possible. In case of only one class modelled at a time the method is the ordinary PLS1-DA. When several classes are simultaneously modelled at the same time, the PLS2-DA modified method can be used [17].

The selection of optimal number of components in PCA and of latent variables in PLS-DA was done from the lowest prediction error in cross validation (leaving out one sample at a time) related to the value of RMSEP. The model giving the lowest relative prediction errors in external validation is finally chosen. In the classification study of this work, PLS2-DA was preferred. All data (FTIR spectra) were processed for the purposes of PLS-DA by the Unscrambler software, version X (Camo, Norway).

# 3. Results and Discussions

## 3.1. Typical FTIR Spectra of Olive Leaves

The FTIR spectra of 125 olive leaf samples were recorded and divided into a calibration set of 95 samples and a prediction set of 35 samples, as mentioned in the previous section. The differences among the FTIR spectra were small and occurred only in limited regions of the wavenumber range.

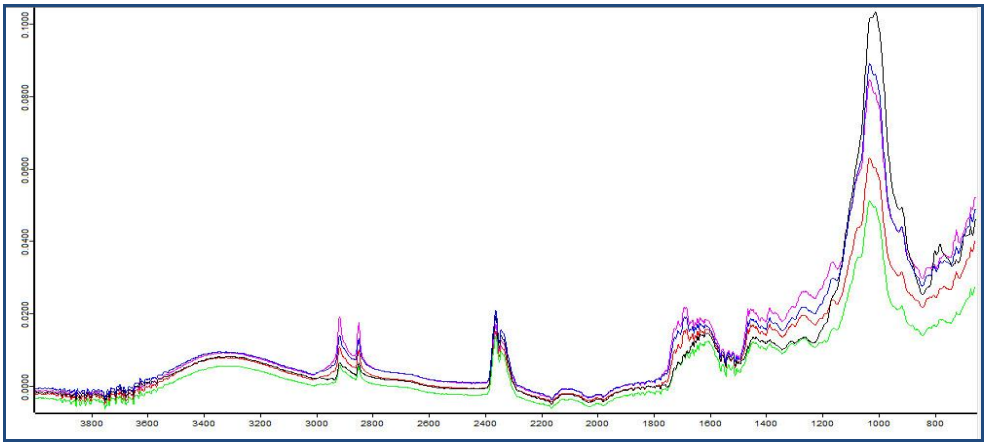


**Figure 1.** Typical Mid FTIR-ATR Spectra of Olive Leaf Samples Recorded for the Five Different Cultivars

The MIR spectra obtained for all samples were similar by visual inspection, indicating that no noticeable qualitative difference existed between the leaf samples. Typical average spectra from each variety were presented in Figure 1. All spectra are characterized by common absorption bands. However, the spectra present some weak differences between 1750 and 700 $cm^{-1}$ in the relative intensities of absorbance from various cultivars. A large band (3700–3000 $cm^{-1}$) is due to OH stretching vibrations (-OH) and the two bands (3000–2800 $cm^{-1}$) from symmetric and asymmetric CH stretching vibrations (-CH, alkyl). The 1800–1500 $cm^{-1}$ region corresponds to C= O and C=C stretching vibrations (esters, acid, carboxylate, aromatic ring). The1500–1200 $cm^{-1}$ range is very complex with especially CH (-CH, alkyl), and OH deformation vibrations as CO stretching vibrations.

The visual examination of the spectral variations did not permit to apprehend clearly the difference chemical structure and chemical species concentration in olive leaves depending on varieties. Chemometric treatments are often applied in order to extract information from the spectral data set.

## 3.2. Chemometric Evaluation

### 3.2.1. Principal Component Analysis

Multivariate analysis is an essential chemometric tool to study data coming from observations made on several variables. Its aim is to resume information contained in data with a reduced number of dimensions to characterize as well as possible the differences or similarities between observations and variables. The spectra data set of olive leaf samples was subjected to the basic tool for PCA data analysis. This statistical method is very important especially in the preliminary steps of a multivariate analysis to perform an exploratory analysis in order to have an overview of data. It allows describing data set without a prior knowledge of the data structure.

After analyzing 95 samples of olive leaves, PCA with full cross validation was applied to the first data set of 95 classification samples exploring the full acquired data.

PCA was used to reduce the data dimensionality in order to obtain a better visualization of the separation in groups according to the varieties. Figure 2 shows the 2D plot of principal component scores. When the model was validated by full cross-validation procedure [18], PCA showed a clear separation of the five classes. In the PCA model, the first two components achieved an explained variance of 94%, which is enough to cluster the samples in the five classes, as can be seen in the score plot of Figure 2. The projection of individuals in the plane generated by the axis 1 and 2, showed the distribution of olive leaves in five main groups( A, B, C, D and F).

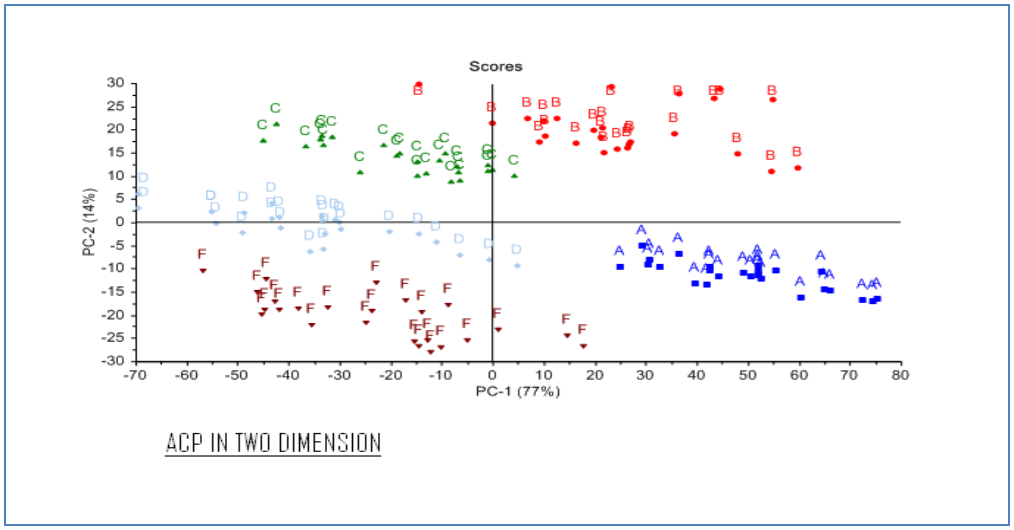


**Figure 2.** PCA Scores Plot (PC1 vs. PC2): Hojiblanca (A), Frantoio (B), Arbequina(C), Leccino(D) and Manzanilla (F)

### 3.2.2. Partial Least Square Discriminate Analysis

This PLS2-DA model was built considering the FTIR spectra as X variables, while the Y variables were associated with the five different olive leaves classes (one different y variable for each group's class, with 1 or 0 depending on whether it belongs or not to the considered data group). The model obtained in this way was able to discriminate among the five classes (A, B, C, D and F), as it can be seen from the PLS2-DA scores plot in Figure 3.

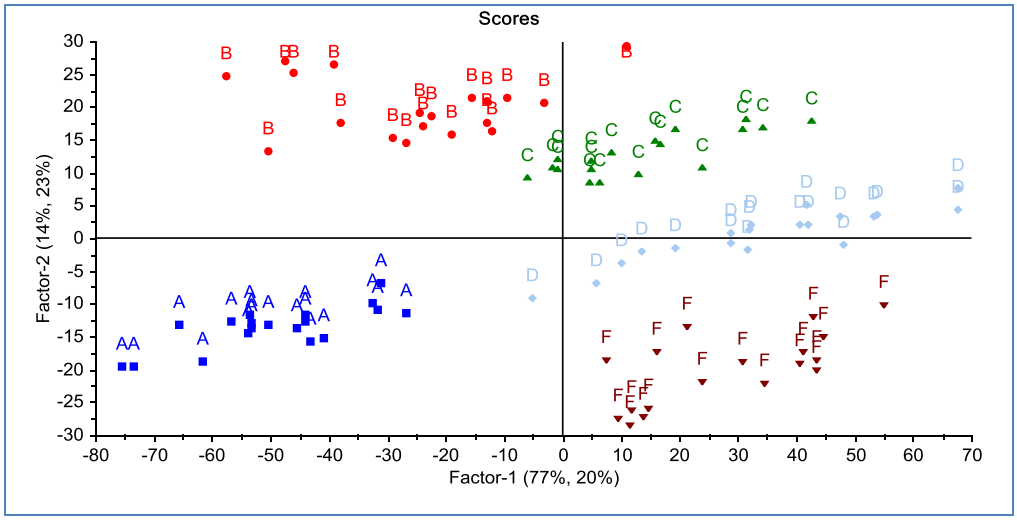


**Figure 3.** PLS2-DA Scores Plot (LV1 versus LV2) in the Analysis of the FTIR Spectra

The first PLS2 latent variable (LV1 explains 77 % of Y variance and 20 % of X variance) discriminates between the type A, B, C and F. The second PLS2 latent variable explains 14% of Y variance and 23% of variance in X, related with the amount and type D.

A calibration model (Figure 3) for predicting the varieties of olive leaves was built using the PLS-DA method from MIR spectra. The calibration (n = 95) set was used for building model, while the validation set (n = 30) was used for testing the robustness of the developed model. The performance of the PLS model was evaluated in terms of standard error of calibration (SEC), root mean square error of prediction (RMSEP) and correlation coefficient (R).

Table 1 shows the calculated figures of merit of the results obtained by the PLS2-DA model using the calibration samples subset. The PLS-DA model is validated by full cross validation. The obtained statistical parameters RMSEC, RMSECV and $R^2$ are summarized in Table 1. A correlation between measured and predicted classes ($R^2$ around 0.92 and 0.96 in all classes), low root mean square error of calibration (RMSEC between 0.07 and 0.09) and a prediction error (RMSEP between 0.08 and 0.10) were obtained. These results demonstrated the ability of the proposed technique to dis-

criminate between the five different cultivars of olive leaf samples used in this study. Seven LVs (latent variables) are necessary to have a good PLS performance. Table 2 lists the explained variances from the developed model.

Table1. Figures of Merit Achieved by PLS2-DA Discrimination of the Five Different Cultivars of Olive Leaf Samples

| Classes[1] | Figures of merit[2] | | | |
|---|---|---|---|---|
| | $R^2c$ | | RMSEC | RMSECV |
| | Calibration | Validation | | |
| A (Hojiblanca) | 0.964 | 0.956 | 0.075 | 0.083 |
| B (Frantoio) | 0.961 | 0.952 | 0.078 | 0.086 |
| C (Arbequina) | 0.949 | 0.937 | 0.089 | 0.099 |
| D (Leccino) | 0.963 | 0.953 | 0.076 | 0.086 |
| F (Manzanilla) | 0.940 | 0.925 | 0.097 | 0.109 |

1: Investigated classes by PLS-DA,
2: Reported model figures of merit: $R^2_c$ - R-square in calibration; RMSEC-Root Mean Squared Error in Calibration; RMSECV-Root Mean Squared Error in cross validation

**Table 2.** Explained Variances (%) of PCs Used in the PLS-DA Model

| Explained | Factor 1 | Factor 2 | Factor 3 | Factor 4 | Factor 5 | Factor 6 | Factor 7 |
|---|---|---|---|---|---|---|---|
| **Calibration** | 19,84 | 43,17 | 65,44 | 88,2 | 94,12 | 95,03 | 96,01 |
| **Validation** | 16,59 | 39,67 | 62,88 | 86,88 | 93,26 | 94,17 | 95,13 |

### 3.3. Classification of new samples

The chosen classification model was applied on the FTIR external validation set, formed by 30 new olive leaf samples, six 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.

Considering the difficulty to calibrate and predict the varieties with binary variables, it was necessary to discriminate the results between the initial values 0 or 1. We considered that all the values between negative value and 0.49 conduce to a non-recognized sample and the ones between 0.51 and high positive ones to a recognized sample.

**Table 3.** Prediction of Olive Leaves Varieties by Chemometric Analysis of MIR Spectra

| | A: Hojiblanc | | B: Frantoio | | C: Arbequina | | D: Leccino | | F: Manzanill | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Samples** | **Y-Pred** | **Y-Ref** | **Y-Pred** | **Y-Ref** | **Y-Pred** | **Y-Ref** | **Y-Pred** | **Y-Ref** | **Y-Pred** | **Y-Ref** |
| A10a | **0 ,9580** | 1 | 0,1016 | 0 | 0,096 | 0 | -0,0286 | 0 | 0 ,0649 | 0 |
| A10b | **0,9914** | 1 | 0,0544 | 0 | -0,0544 | 0 | -0,009 | 0 | 0 ,0176 | 0 |
| A10c | **0,9926** | 1 | 0,0613 | 0 | -0,0687 | 0 | -0,0062 | 0 | 0,021 | 0 |
| A11a | **0 ,9853** | 1 | 0,0219 | 0 | 0,0147 | 0 | 0 ,0355 | 0 | -0,0574 | 0 |
| A11b | **0,993** | 1 | -0,0282 | 0 | 0,0394 | 0 | -0,004 | 0 | 0,0002 | 0 |
| A11c | **1,0397** | 1 | -0,1579 | 0 | 0,1363 | 0 | 0 | 0 | -0 ,0181 | 0 |
| B11a | 0,0194 | 0 | **0 ,9342** | 1 | 0,0227 | 0 | -0 ,0359 | 0 | 0 ,0596 | 0 |
| B11c | -0,0128 | 0 | **0,9457** | 1 | -0,0187 | 0 | -0,0649 | 0 | 0,1507 | 0 |
| B12a | 0,0575 | 0 | **0,871** | 1 | 0,0691 | 0 | 0,0041 | 0 | -0,0018 | 0 |
| B12b | 0,0488 | 0 | **0,9225** | 1 | 0,0428 | 0 | 0,0181 | 0 | -0,0322 | 0 |
| B14a | 0,0232 | 0 | **1,0032** | 1 | -0,0653 | 0 | -0,0282 | 0 | 0,0671 | 0 |
| B1a | -0,0831 | 0 | **1,1897** | 1 | 0,0169 | 0 | 0 ,1O32 | 0 | -0,2267 | 0 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| C10a | -0,0059 | 0 | 0,0778 | 0 | **0,9466** | 1 | -0,0433 | 0 | 0,0247 | 0 |
| C10b | 0,0328 | 0 | -0,0248 | 0 | **0,9893** | 1 | -0 ,0289 | 0 | 0,0315 | 0 |
| C10c | -0,088 | 0 | 0,0274 | 0 | **0,9506** | 1 | -0,0622 | 0 | 0,1722 | 0 |
| C11a | 0,0388 | 0 | -0,0167 | 0 | **1,0026** | 1 | 0,0081 | 0 | -0,0327 | 0 |
| C11b | -0,0017 | 0 | -0,0032 | 0 | **1,0296** | 1 | 0,0375 | 0 | -0,0622 | 0 |
| C11c | -0,0047 | 0 | 0,0068 | 0 | **0 ,9903** | 1 | 0,0452 | 0 | -0,0376 | 0 |
| D10a | 0,0137 | 0 | 0,0299 | 0 | -0,0567 | 0 | **0,9756** | 1 | 0,0374 | 0 |
| D10b | -0,0141 | 0 | -0,0241 | 0 | -0,0507 | 0 | **0,9785** | 1 | 0,1104 | 0 |
| D10c | 0,0053 | 0 | -0,0279 | 0 | -0,0481 | 0 | **0,9751** | 1 | 0,0956 | 0 |
| D11b | 0,0294 | 0 | -0,0343 | 0 | 0,118 | 0 | **1,0749** | 1 | -0 ,1880 | 0 |
| D11c | -0,0582 | 0 | 0,0655 | 0 | -0,0092 | 0 | **0,969** | 1 | 0,0328 | 0 |
| D12a | 0,0116 | 0 | 0,0128 | 0 | 0,003 | 0 | **0,9629** | 1 | 0,0353 | 0 |
| F10a | -0,0178 | 0 | 0,0146 | 0 | 0,0008 | 0 | 0,0088 | 0 | **0,9937** | 1 |
| F10b | -0,0288 | 0 | 0,0486 | 0 | -0,0534 | 0 | -0,0853 | 0 | **1,1189** | 1 |
| F10c | 0,0719 | 0 | -0,0254 | 0 | 0,0437 | 0 | 0,097 | 0 | **0,8129** | 1 |
| F11b | 0,0204 | 0 | -0,0671 | 0 | 0,0872 | 0 | 0,053 | 0 | **0,9066** | 1 |
| F11c | -0,0232 | 0 | 0,024 | 0 | -0,0098 | 0 | -0,0196 | 0 | **1,0286** | 1 |
| F12b | 0,005 | 0 | 0,0221 | 0 | 0,0276 | 0 | 0,0697 | 0 | **0,8754** | 1 |

Pred - predicted; Ref - reference

Table 3 shows that for the 30 validation samples,we noted that the samples A10a, A10b, A10C, A11a, A11b and A11c belong to the class A (Y Predict ≈ Y Reference), the same observation for the samples B11a, B11c, B12a, B12b, B14a and B1a, samples C10a, C10b, C10c, C11a, C11b and C11C, samples d10a, d10b, d10c, d11b, d11c and d12a, and the samples F10a, F10b, F10c, F11b, F12b and F11c, which belong to the class B, C, D and F, respectively .

A 100% correct classification was achieved, i.e. all olive leaves spectra of the validation data set were correctly matched to the five corresponding classes. In Table 3, the predicted values by the PLS2-DA model are always very close to 1. These results confirm that the predictive ability of the developed PLS2-DA model was very good. Therefore, it was established that FTIR spectroscopy together with the application of chemometrics (PLS2-DA) method can be used to discriminate between varieties of olive leaves.

## 4. Conclusions

In this paper, it can be concluded that the discrimination between five introduced varieties of olive in Morocco, cultivated in the same geographical area can be performed by a chemometric approach based on the FT-MIR spectra data provided by olive leaves. Furthermore, MIR spectroscopy presents high potential for varieties and cultivar differentiation and prediction by PLS-DA method. Therefore, a simple, rapid and reliable overall characterization of olive varieties and cultivar may be obtained at a low cost. It might be an application for rapidly classification of olive cultivar. This study demonstrates the great potential in the application of chemometric tools in infrared spectroscopy for the correct classification of food.

## REFERENCES

[1] Ruanoa F., Camposb M., Soler J.J., 2003. Differences in leaves of olive trees under organic, integrated and conventional pest management. Ecosyst. Environ., 97, 353-356.

[2] Paula B.M., Joaquim C.G., Esteves da Silva. 2005. Chemometric classification of olives from three Portuguese cultivars of Oleaeuropaea. Anal. Chim. Acta, 544, 229-235.

[3] Alché J.D., Castro A.J., Jiménez J.C., Morales S., Zafra A., Hamman A.M., Rodríguez M.I., 2007. Differential characteristics of olive pollen cultivars. J. Investig. Allergol. Clin. Immunol., 17, 69-75.

[4] Gómez-Rico A., Fregapane G., Salvador M.D., 2008. Effect of cultivar and ripening on minor components in Spanish olive fruits and their corresponding virgin olive oils. Food Res. Int., 41, 433-440.

[5] Hannachi A., Catherine B., Monji C., El Gazzah A., Andre B., 2008. Differences between native and introduced olive cultivars as revealed by morphology of drupes, oil composition and SSR polymorphisms. Sci. Hortic., 116, 280-290.

[6] Trujillo I., Rallo L., Arus P., 1995. Identifying olive cultivars by isoenzyme analysis. J. Am. Soc. Hortic. Sci., 120, 318-324.

[7] Aouidi F., Nathalie D., Jacques A., Sevastianos R., Monji M.D., Isabelle P.G., Moktar H., 2012. Discrimination of five Tunisian cultivars by Mid InfraRed spectroscopy combined with chemometric analyses of olive Oleaeuropaea leaves. Food Chem., 131, 360-366 .

[8] De Luca M., Wafa T., Giuseppina I., Kzaiber F., Oussama A., Filomena O., Romà T., Gaetano R., 2011. Derivative FTIR spectroscopy for cluster analysis and classification of morocco olive oils. Food Chem., 124, 1113-1118.

[9] Aouidia A., Nathalie D., Jacques A., Sevastianos R., Monji M., Isabelle P., Moktar H., 2012. Rapid quantitative determination of oleuropein in olive leaves (Oleaeuropaea) using mid-infrared spectroscopy combined with chemometric analysis. Ind. Crops Prod., 37, 292-297.

[10] Vergnoux A., Guiliano M., Dréau Y., Kister J., Dupuy N., Doumenq N., 2009. Monitoring of the evolution of an industrial compost and prediction of some compost properties by NIR spectroscopy. Sci. total Environ., 407, 2390-2403.

[11] Jung-Hyun P., HyeJunga J., Jin-Young Y., Hyun-Sook K., 2013. Olive leaf down-regulates the oxidative stress and immune dysregulation in streptozotocin-induced diabetic mice. Nutr. Res., 33, 942-951.

[12] Sinelli N., Cerretani L., Di Egidio., Bendini A., Casiraghi E., 2010. Application of near (NIR) infrared and mid (MIR) infrared spectroscopy as a rapid tool to classify extra virgin olive oil on the basis of fruity attribute intensity. Food Res. Int., 43, 369-375.

[13] Terouzi W., De Luca M., Bolli, A., Oussama A., Patumi M., Ioele G., Ragno G., 2011. A discriminant method for classification of Moroccan olive varieties by using direct FT-IR analysis of the mesocarp section. Vib. Spectrosc., 56, 123-128.

[14] Casale M., Zunin P., Elisabetta C.M., Pistarino E.,Perego P., LanterS., 2010. Characterization of table olive cultivar by NIR spectroscopy. Food Chem., 122, 1261-1265.

[15] Lebart L., Morineau A., Pirlo M., 1997. L'Analyse en composantes principales. In: Statistiques exploratoires multidimensionnelles, Dunod, Bordas, Paris: 32.

[16] Elbassbasi M., Hirri A., Kzaiber F., Ragno G., Oussama A., 2010. Classification of raw milk by infrared spectroscopy (FTIR) and chemometric. Journal of Scientific Speculations and Research, 1, 28-33.

[17] Geladi P., Kowalski B., 1986. Partial least-squares regression: a tutorial. Anal. Chim. Acta, 185, 1-17.

[18] Sathle L., Wold S., 1987. Partial least squares analysis with crossvalidation for the two-class problem: a Monte Carlo study. J. Chemometr., 1, 185-196.