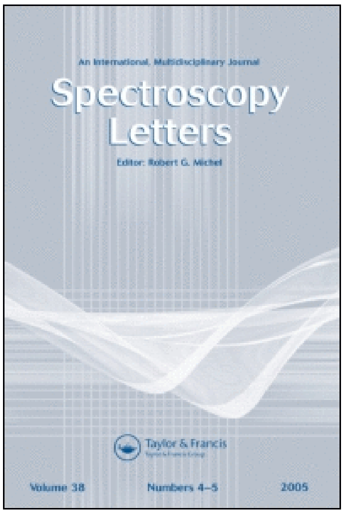

Spectroscopy Letters
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597299

Fourier-Transform Near-Infrared Spectroscopy as a Tool for Olive Fruit Classification and Quantitative Analysis

María José Ayora-Cañada[a]; Barbara Muik[a]; Jose Antonio García-Mesa[b]; Domingo Ortega-Calderón[b]; Antonio Molina-Díaz[a]
[a] Department of Physical and Analytical Chemistry, University of Jaén, Jaén, Spain [b] CIFA Venta del Llano, IFAPA, Jaén, Spain

To cite this Article Ayora-Cañada, María José , Muik, Barbara , García-Mesa, Jose Antonio , Ortega-Calderón, Domingo and Molina-Díaz, Antonio(2005) 'Fourier-Transform Near-Infrared Spectroscopy as a Tool for Olive Fruit Classification and Quantitative Analysis', Spectroscopy Letters, 38: 6, 769 — 785

To link to this Article: DOI: 10.1080/00387010500316106

URL: http://dx.doi.org/10.1080/00387010500316106

Spectroscopy Letters, 38: 769–785, 2005
Copyright © Taylor & Francis, Inc.
ISSN 0038-7010 print/1532-2289 online
DOI: 10.1080/00387010500316106

Fourier-Transform Near-Infrared Spectroscopy as a Tool for Olive Fruit Classification and Quantitative Analysis

María José Ayora-Cañada and Barbara Muik
Department of Physical and Analytical Chemistry, University of Jaén,
Jaén, Spain

Jose Antonio García-Mesa and Domingo Ortega-Calderón
CIFA Venta del Llano, IFAPA, Jaén, Spain

Antonio Molina-Díaz
Department of Physical and Analytical Chemistry, University of Jaén,
Jaén, Spain

Abstract: The potential of diffuse reflectance near-infrared spectroscopy combined with pattern recognition to discriminate between olives (*Olea europaea* L.) of different qualities has been tested. The sample set was formed of sound olive fruits and those showing the most common alterations of olives, which lead to decreased oil quality, namely freeze damages, harvest after falling on the ground, fermentation due to prolonged storage time, and olive tree diseases. Near-infrared (NIR) spectra were recorded between 9900 and 4100 cm^{-1}. Preliminary studies of the data set structure were performed using hierarchical cluster analysis and principal component analysis. Discriminant analysis provided prediction abilities of 100% for sound, 79% for frostbite, 96% for ground, and 92% for fermented olives using a leave-a-fourth-out cross-validation procedure. Quantification of oil and water content in the olives was also approached by

Received 5 August 2004, Accepted 9 February 2005

This paper was by special invitation as a contribution to a special issue of the journal entitled “Quantitative Vibrational Spectrometry in the 21st Century.” This special issue was organized by Professor Miguel de la Guardia, Professor of Analytical Chemistry at Valencia University, Spain.

Address correspondence to Antonio Molina-Díaz, Department of Physical and Analytical Chemistry, University of Jaén, Paraje las Lagunillas s/n, E-23071 Jaén, Spain. E-mail: amolina@ujaen.es

using partial least squares (PLS) regression. Results, in terms of predictive ability using a leave-one-out cross-validation, were compared for calibration using the whole sample set and calibrations for the sound and damaged samples separately. Relative errors of prediction using all samples were 7.2% and 3.4% for oil content and humidity, respectively. Using only sound samples, relative errors of prediction of 3.8% and 2.8% for oil and water content, respectively, were obtained. Thus, better accuracy could be achieved when classification of the olive samples prior to quantitative analysis was performed. These results demonstrate the utility of NIR spectroscopy to differentiate olives of different qualities. Using NIR, a fast selection of sound olives in a quality-orientated production facility becomes feasible.

Keywords: Classification, Fourier-transform near-infrared spectroscopy, olives, partial least squares, pattern recognition

INTRODUCTION

The olive fruit (*Olea europaea* L.) constitutes the raw material in olive oil production. Virgin olive oil is obtained from the fruit of the olive tree only by mechanical or other physical means in conditions, particularly thermal conditions, which do not alter the oil in any way. As one of the few vegetable oils which can be consumed directly without the need of refining, it completely conserves vitamins, antioxidants and other dietary important constituents and can be qualified as a natural product. But to be considered fit for direct consumption virgin olive oil must fulfil different quality criteria that, according to the European Commission regulations,[1,2] include good organoleptic characteristics (taste and aroma), low acidity, low peroxide value, and absence of contaminants. Oil of inferior quality is called *lampante* and must be refined prior to consumption.

From the industrial point of view, the two major parameters that characterize the olive as raw material are oil content and humidity (vegetation water). In routine laboratories, traditional slow analytical methods to determine oil content have been nowadays substituted by direct spectroscopic measurement using nuclear magnetic resonance (NMR).[3] These analyses involve parameters that exclusively refer to the industrial yield (i.e., to the amount of oil which can be extracted from a certain mass of olives). But the production of good quality oil must start necessarily with raw material of high quality. Damage of the olives during growth phase, harvest, and storage leads to decreased oil quality.[4] Analytical methods to evaluate the state of the olives before oil extraction are therefore of crucial importance. The only quality parameter that is actually measured in olives is acidity (free fatty acid content). Sound fruits contain low levels of free fatty acids, but these values are increased by different alterations of the olive fruit. Thus, the measurement of acidity provides a rough estimation of the alterations the olive has suffered. The method implicates the extraction of the oil from the olive. This is normally done by the Abencor method,[5] which reproduces the industrial oil extraction process in laboratory scale. Afterwards, acidity is determined in the oil by an acid-base titration with potassium hydroxide in ethanol/ether and phenolphthalein as indicator.[6] Hence, this method is slow and elaborate. Consequently, only few producers measure acidity in olives, and the quality control of the raw material, if done at all, is mostly reduced to visual inspections. Nevertheless, a separation of olives according to their quality would be necessary if the production of extra virgin olive oil is the goal. To do this, a fast and objective analytical method is desirable.

Near-infrared spectroscopy (NIRS) is a valuable tool for the analysis of complex samples. In many cases, the analysis can be carried out without any sample preparation, thus providing results in an experimental simple way. In the analysis of olives, NIRS has been proposed as a convenient alternative for the direct determination of oil and water content simultaneously in the milled olives.[7] This method avoids the time-consuming step of drying the sample, which is needed in all above-mentioned methods and thus provides results much faster. More recently, the quantitative determination of oil content, humidity, and major fatty acids in intact olive fruits has been approached.[8] These works demonstrate that NIRS can constitute an efficient and flexible solution for the fast analysis of large number of raw material samples.

A further benefit of this technique is that it is possible to bypass classical, quantitative chemical analysis if the desired property is reflected in the overall chemical composition of the sample. Using chemometrics, this information can be extracted from the spectra. Thus, in the analysis of olives, the measurement of traditional parameters like acidity become dispensable, if the information contained in the near-infrared spectra of the olives is sufficient for the discrimination of sound and defective olives.

The aim of the current study is to explore the potential of FT-NIR spectroscopy to provide a fast method to classify olives according to their quality before the extraction of the oil in the production facility. For this, the most common types of alterations that olives can suffer were considered. They included freeze damage, deterioration of the fruit flesh due to the fall from the tree and the permanence on the ground for several days, fermentation as a consequence of excessive storage time in piles, and infestation by olive tree diseases.

Furthermore, it is well-known that parameters like harvest year and variety have great influence on oil content and humidity in the olives. In previous studies, authors report on the difficulties of partial least squares (PLS) calibration transfer between different years.[9] In the current work, we furthermore investigate the influence of olive quality on the determination of oil content and humidity by PLS. This is done by comparing the predictive ability of calibration models that are built on the individual classes using a classification step prior to quantification with the predictive ability of the calibration model that is built on the whole sample set.

EXPERIMENTAL

Olive Samples

A total of 122 olive samples were collected in the period January–March 2004 in the olive culture and process research station CIFA Venta del LLano (Jaén, Spain). They included 36 samples of sound olives (of 8 different varieties) collected from the tree, 28 of olives with frostbite, 6 samples with diseases (Verticillium wilt and Anthracnose), 26 fermented olive samples, and 26 samples of olives that had been collected from the ground. In Table 1, sample details including range, mean, standard deviation of oil content and humidity, and the number of varieties present in each class are presented. In tables, figures, and in the text, samples will be referred to in following five classes: sound, frostbite, disease, fermented, and ground.

Reference Analysis

One kilogram of the sample was milled in a hammer mill (4-mm strainer size). To determine water content, 45 g of the sample was dried for 12 hr in an oven at 105°C. The loss of weight gave the percentage of water and volatile matter of the sample.[10] The dried sample was then used to measure the oil content, employing NMR spectroscopy.[3]

Acquisition of NIR Spectra

Fourier-transform near-infrared spectra were measured in diffuse reflectance mode using an Antaris Near-IR Analyzer (Thermo Nicolet Corporation, Madison, WI, USA). The instrument is equipped with an integrating sphere that contains an internal gold reference.

The milled sample was filled into a sample cup (5-cm diameter, 1-cm depth) that was placed on top of the integrating sphere optics and rotated during measurement (10 rounds per minute). All spectra were obtained between 4100

Table 1. Sample characteristics of the individual classes

	Humidity %			Oil content %			
	(mean)	SD	Range	(mean)	SD	Range	Varieties
Sound	53.2	3.6	46.1–61.8	23.0	2.2	18.9–28.8	8
Frostbite	44.3	3.6	37.7–50.6	26.3	2.6	22.2–32.3	7
Ground	38.9	3.2	33.9–44.4	28.3	3.7	23.5–35.3	4
Fermented	46.3	3.9	32.9–47.4	27.5	0.9	27.2–30.5	—
Disease	37.2	12.1	23.2–53.9	25.6	6.1	18.0–31.9	4

and 9900 cm^{-1} at a resolution of 8 cm^{-1} and were mean of 50 scans, which resulted in an acquisition time of 1 min. A background spectrum average of 50 scans was collected before each sample measurement.

Data Analysis

Data Pretreatment

Mean centering was used to enhance the smaller spectral differences removing the common information from the spectra. Multiplicative signal correction (MSC), which uses a mathematical function originally developed to compensate for variations in light scattering, was also applied.[11] This correction method assumes that the wavenumber dependency of light-scattering or baseline signal is different from that of the chemical information. The MSC treatment is often used in near-infrared diffuse reflectance measurements to compensate for variations in effective sample thickness that are caused by differences in particle size and scattering. First derivatives were calculated with the polynomial method of Savitzky-Golay.[12]

Pattern Recognition Techniques

Pattern recognition techniques can be mainly divided into two categories, namely unsupervised and supervised techniques. The former focus on investigating the structure in the data, detecting similarities among samples or the presence of outliers. They do not require information about class membership because their aim is not to define a classification rule. In contrast, supervised methods use information on class membership to build up a classifier. The models are constructed using the sample characteristics of a training set to establish classification rules, which are then applied to obtain the class-membership of unknown samples.

Unsupervised Pattern Recognition

Hierarchical Cluster Analysis (HCA). For cluster analysis, each sample is treated as a point in an n-dimensional space defined by the measurement variables. Cluster analysis assesses the similarity between samples by measuring the distances between the points in the measurement space. In this work, the similarity matrix for hierarchical clustering was calculated using Euclidean distances, and the Ward algorithm[13] was used to generate the dendrograms.

Principal Component Analysis (PCA). The overall goal of PCA is to reduce the dimensionality of a data set. PCA decomposes the original data matrix into a product of two matrices, the scores and the loadings matrix. The loadings form the new reduced coordinate system in which the samples are represented with their score values. Thus, the scores matrix contains the information of the objects. When PCA is applied for pattern recognition purposes, this information is used to detect groupings among the samples.

Supervised Pattern Recognition

In the discriminant analysis classification technique (TQ Analyst 6.1.1), principal component analysis (PCA) is performed to reduce the number of variables in the analysis. The loading vectors of the principal components are then used as the variables for entry into a quadratic discriminant analysis (QDA) model, which maximizes the variance between groups and minimizes the variance within groups. Each sample is represented by its score values as a data point in the new dimensions, and a multivariate normal distribution is calculated for each class using the information of the training samples. To determine the class membership of an unknown sample, its Mahalanobis distances are measured to the means of all classes and the unknown is assigned to the class with the minimum distance. Furthermore, a threshold limit of 2 Mahalanobis distance units was defined. If the minimum distance of an unknown was above the threshold distance value, the sample was considered not to belong to any class.

Multivariate Calibration: Partial Least Squares Regression

Partial least squares (PLS) regression is one of the most widely used methods for multivariate calibration. It is based on spectral decomposition in which the original variables are replaced by so-called latent variables, which are linear combinations of the original ones.[14] For this calculation, PLS uses the spectral and concentration information and maximizes the covariance between them, so achieving latent variables that are directly related to the constituents of interest. PLS is a powerful full spectrum method with excellent model-diagnostic capabilities, which can account for undefined spectral variations that are not related to the constituents of interest, as long as they are included in the calibration phase.

Software

Hierarchical cluster analysis (HCA) was performed using the Statistics Toolbox (Version 2.2 (R11), 1998) for Matlab (The Mathswork, Inc., Natick, MA, USA). Principal component analysis (PCA), discriminant analysis, and PLS were done using TQ Analyst 6.1.1 (Thermo Nicolet Corp.).

RESULTS AND DISCUSSION

Near-Infrared Spectra of Olives

The fruit of the olive tree (*Olea europaea* L.) consist of three basic parts: the skin (epicarp), the pulp (mesocarp), and the pit (endocarp). The pulp accounts for approximately 75% of the total weight and contains the major part of the oil and water of the fruit, which represent about 20–35% and 30–60%, respectively.[15] In Fig. 1a, the averaged spectra of each class are presented to illustrate the distinct pattern of the olive samples. Two broad bands at 7500–6150 cm^{-1} and 5350–4550 cm^{-1} dominate the spectrum of all olives. These bands due to water can be located at 6950 and 5210 cm^{-1} and are assigned to the first overtone of OH symmetric and antisymmetric

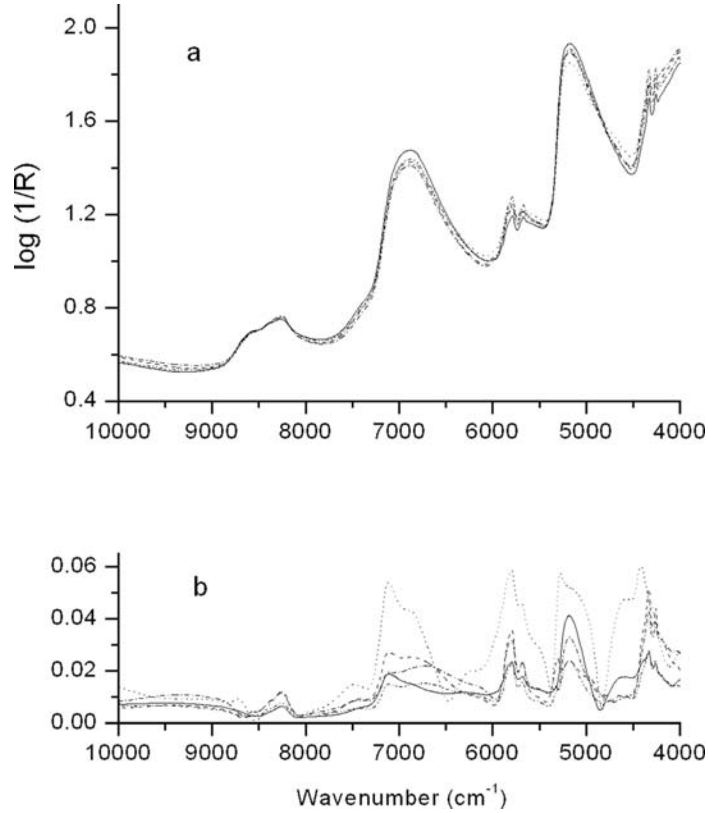

Figure 1. (a) Diffuse reflectance NIR spectra of olive samples: mean spectrum of classes sound (solid), disease (dot), fermented (dash), ground (dash dot), and frostbite (dash dot dot). (b) Variance spectrum of sound (solid), disease (dot), fermented (dash), ground (dash dot), and frostbite (dash dot dot) classes.

stretching ($2\nu_{1,3}$) and to combination of OH stretching and bending ($\nu_{1,3} + \nu_2$), respectively.[16] The significant bands of the olive oil are also clearly visible in all of them. They are located at 4260 and 4370 cm^{-1}, which are characteristic for the combination of C–H-stretching vibrations of $-CH_3$ and $-CH_2$ with other vibrations. The two bands at 5700 and 5750 cm^{-1} correspond to the first overtone of the C–H stretching vibration of $-CH_3$, $-CH_2$, and $-HC{=}CH-$. In the region between 7700 and 9100 cm^{-1}, the second overtone of the C–H stretching vibration of $-CH_3$, $-CH_2$, and $-HC{=}CH-$ can be found.[17,18] Visual inspection of the mean spectra of the classes reveals subtle differences among them. These can be better visualized in Fig. 1b, where the variance spectra of the different classes are presented. The variance within the class sound can be attributed to compositional differences among the eight varieties considered and variations in oil and water content. For the rest of the classes, the different alteration processes lead to the development of characteristic spectral features, with disease samples showing the more distinct pattern.

Unsupervised Pattern Recognition

HCA

After a first inspection of the untreated data, MSC was found to be an appropriate pretreatment. Hierarchical cluster of the mean centered spectra after MSC pretreatment identified five main clusters (Fig. 2). Cluster A was a mixture of ground, fermented, and frostbite samples. It was a large cluster separated into four subgroups. The first subgroup contained samples from fermented and ground classes and two frostbite samples. The second was formed by 10 frostbite samples and 1 sample with disease. The third subgroup included 12 fermented samples and 1 with frostbite. The last one was a mixture of fermented, ground, and frostbite samples. Custer B was formed by sound and frostbite samples and one sample with disease. Cluster C was a pure cluster, constituted by 21 sound samples. D was a very small cluster formed by three samples with disease (Verticillium). Cluster E contained 18 fermented samples, 3 frostbite and 6 ground samples. These results indicate a certain overlap between fermented and ground classes and to a lesser extent with frostbite class. Although overlap between sound and frostbite classes is observed, there is a very good separation among sound samples and fermented or ground.

PCA

With PCA, the data can be plotted in a new coordinate system based on maximum variance. The origin of the new coordinate axes is the center of the data, and the principal components define the variance axes. Each

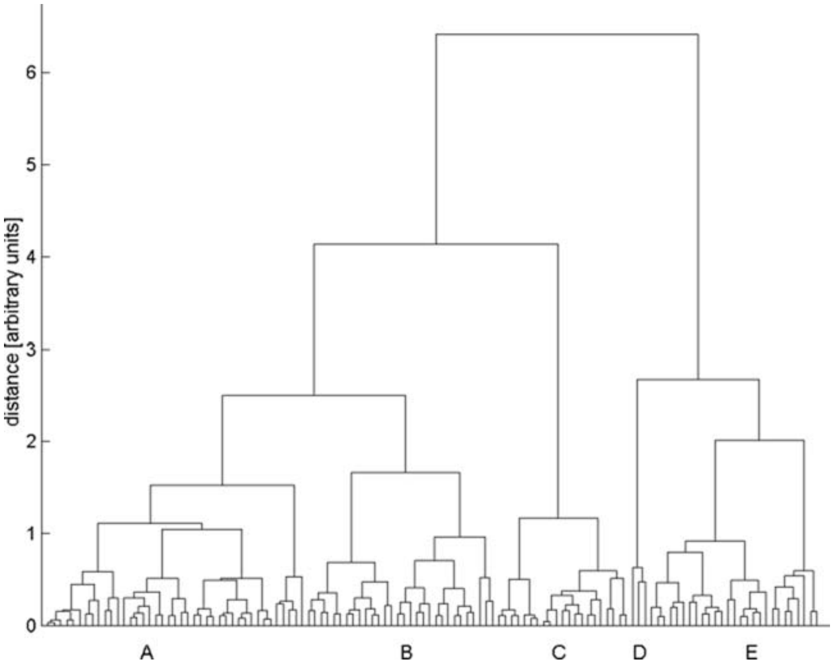

Figure 2. Dendrogram of the mean-centered and MSC corrected data using Euclidean distances and the Ward algorithm.

principal component describes a different direction of variance in the data. By plotting the data in a coordinate system defined by the most significant principal components, it is possible to identify key relationships in the data (i.e., find similarities and differences among objects in a data set.)

First, PCA was done including all the samples. It revealed that the six disease samples were very distant from the others but did not form a clear cluster. Clustering of the rest of the olive samples into distinct groups was not apparent until these six samples were considered outliers and eliminated. The so-achieved grouping can be best seen by plotting the score values in the dimension formed by PC1 and PC4 (Fig. 3). PC1 explained 78.3% of the total variability (tot. var.) and clearly affected a separation between sound and ground or fermented samples. PC4 (2.1% tot. var.) helped to achieve better clustering, although ground and fermented classes still overlapped. Samples with frostbite formed a cluster, which slightly overlapped with sound, fermented, and ground classes. These results agree with those obtained using HCA. With the help of PC3 (3% tot. var.) and especially PC5 (0.9% tot. var.), separation between fermented and ground classes was achieved (not shown). On the contrary, PC2 (15.4% tot. var.) did not contribute to the clustering.

Detailed chemical interpretation of the discriminative information present in the different principal components is not easy because bands in the NIR

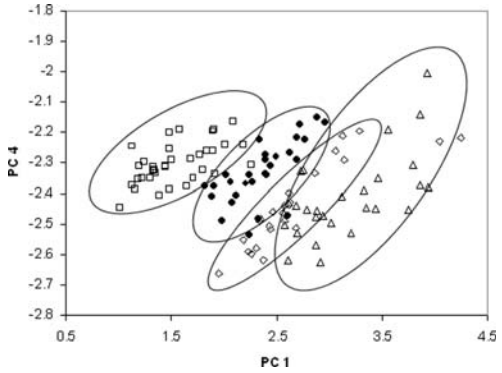

Figure 3. Samples in the space defined by the first and fourth PC of the mean-centered and MSC corrected data. □, sound; •, frostbite; △, ground; ◇, fermented olives.

region arise from overtones and combinations of fundamental vibrations. This results in extensive overlapping in NIR bands. However, some chemical information can be obtained from the PCA results. PC1 loadings vector (Fig. 4a), which shows the maximum discriminative capability, shows spectral features due to oil (bands at 5850 and 5700 cm^{-1}, as well as 4335 and 4260 cm^{-1}). Damaged olives have usually suffered dehydration due to their permanence on the ground or storage. The oil content (referred to humid matter) is therefore higher in these olives. This is reflected in higher values of this PC. All bands of oil can be clearly identified in PC4 and PC5, too. Furthermore, PC1 presents a band in the region of 4400 cm^{-1} that can be seen as shoulder at 4408 cm^{-1}. This region has been reported to show good correlation with titratable fatty acids (free fatty acids).[19] Thus, high score values of PC1 also reflect lipolysis, which takes place extensively in ground and fermented olives due to the action of microbial lipases, increasing free fatty acids content in the oil. Frostbite samples showed scores for this PC more similar to sound samples because dehydration does not take place and free fatty acids are usually not formed as consequence of freezing.[20] PC4 and PC5 show bands in the regions 4800 and 6900, which can be attributed to oxidation products formed during alteration of the olives.[21,22] Other remarkable bands of the PC loadings appear in the water absorption regions and reflect the splitting of the water bands in different contributions. PC2 loading vector (Fig. 4a) exhibits a strong negative band at 5210 cm^{-1}, which can be attributed to a band of water that explains variance in water content not related to class membership (this PC did not contribute to class separation). Nevertheless, the spectrum of water in natural products is very influenced by hydrogen bonding and dissolved organic and inorganic materials. Changes in water spectrum due to water loss of agricultural products such as soybean flour[16] and wheat seeds[23] have been reported. In particular, during drying a band attributed to free water located at

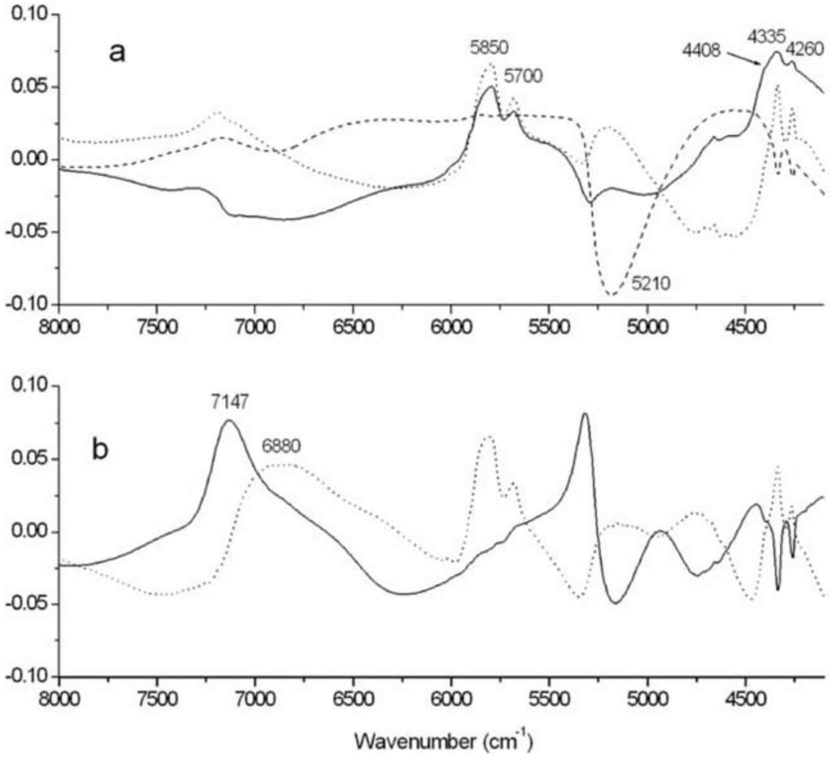

Figure 4. Loading vectors of PCs: (a) PC1 (solid), PC2 (dot) PC4 (dash); (b) PC3 (solid), PC5 (dot).

7120 cm^{-1} decreases, whereas a band at 6990 increases. Such contributions can be identified in PC3 (at 7147 cm^{-1}) and PC5 (6880 cm^{-1}), respectively. It seems that the molecular states of water change with the alterations that the olives suffer, and therefore bands of different water species contribute to the discrimination between the different classes of olives.

Supervised Pattern Recognition

The unsupervised techniques showed that the natural groupings within the NIR spectra of the samples coincide largely with the known class-memberships. Therefore discriminant analysis, a supervised pattern recognition technique, was applied. This classification technique uses a training data set to establish classification rules, which are used to classify samples of unknown origin. A model always delivers smallest residuals for the training set itself. For testing model performance, a leave-a-fourth-out cross-validation procedure was applied. Three-quarters of the samples were assigned to the training set and the other quarter to the test set. This assignment was repeated four times, so that each sample was predicted once. All classification results presented in this section refer to validation results.

The six samples with diseases were too few to define a particular class. Furthermore, they showed no tendency to form a group, as it was seen in the PCA analysis, and only three of them formed a cluster when HCA was applied. It was therefore decided not to include them in the training set as a particular class. However, they were included in the validation procedure to check if they would be misclassified or recognized as not belonging to any class.

Table 2 summarizes the results obtained using a classification model with five principal components for each class. Classification errors were divided into false positive (i.e., how many samples have been wrongly classified to class x) and false negative (i.e., how many samples of class x have been wrongly assigned to other classes). All sound samples were correctly classified, and prediction abilities higher than 90% were obtained for ground and fermented classes. Predictions for frostbite class were not as good, with the highest number of false negatives. This reflects the overlap of this class with sound and ground classes, which can also be seen in Figs. 5a and 5b. In general, freeze damages (frostbite) in the olives are found to affect the quality less than other alterations. In a recent study, the main quality indices of the oil were found not to be strongly affected by freeze damage, especially the acidity did not increase significantly.[20] Furthermore, the extent of freeze damage in the olives depends on the duration of their exposure to chill. Thus, slightly freeze-damaged samples can easily fall within the sound class limits. On the other hand, olives with freeze damage are reported to have lower stability against oxidative degradation. This could justify the overlap with the ground class for seriously damaged samples. Furthermore, it was checked if olives of particular varieties gave

Table 2. Validation results of classification (prediction ability is given in percentage)

Actual class	Predicted class					False negative
	Sound	Frosbite	Ground	Fermented		
Sound	36 (100%)	—	—	—		0/36
Frosbite	1	22 (79%)	4	—	1	6/28
Ground	—	—	24 (96%)	1	1	2/26
Fermented	—	—	1	24 (92%)	2	3/26
Disease	—	—	—	1	5	
False positive	1/86	0/94	5/96	2/96		

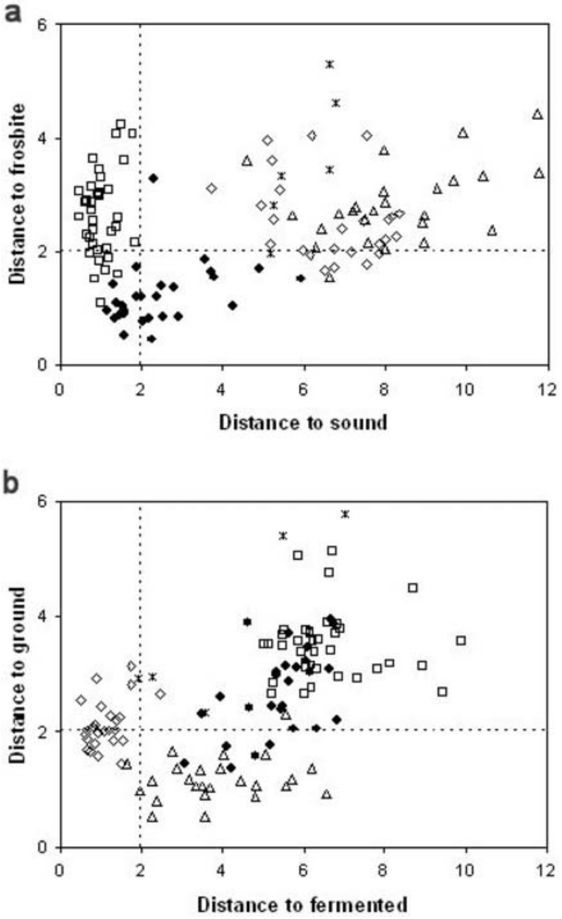

Figure 5. Cooman's plots showing (a) distances of samples to classes sound and frostbite and (b) distances to classes fermented and ground. □, sound; •, frostbite; △, ground; ◆, fermented; ∗, disease olives. Dashed lines mark the threshold limit.

more false negative classifications. But this was not the case. False positive classifications in the class sound are the most critical aspect to be considered, because it can cause a decrease in the quality of the produced oil. Anyway, with the final decision based on the minimum distance, only one of the frostbite samples was classified as sound.

None of the ground and fermented samples was misclassified as sound. As can be seen in Fig. 5a, all samples of these two classes fall very distant of the sound class threshold limit.

The highest number of false positives was observed for the class ground, probably due to the higher diversity of ground samples as a consequence of the manner in which these olives are collected: in contact with the ground, vegetable matter and soil can be incorporated and different alteration processes can take place.

From the six olives with disease, only one was assigned to the class fermented. The others were found to be closest to frostbite or ground, but with large distances to all the classes. One fermented, one frostbite, and one ground sample were also not assigned to any of the four classes: they showed distances slightly higher than the threshold limit for their respective classes.

Quantification of Oil and Water Content in Olives Using PLS Regression

The determination of oil and water content in olives was first approached considering all the samples. The number of latent variables that minimized the prediction error sum of squares (PRESS)[24] during a leave-one-out cross-validation procedure was selected to construct the models. After optimization, best results in terms of prediction ability, evaluated with the root mean square error of prediction (RMSEP)[24] in the cross-validation procedure were obtained using the first derivative of the MSC corrected spectra. A seven-point third polynomial order Savitzky-Golay filter was used for smoothing. Outlier diagnostics were used to ensure the quality of the calibration set by identifying outlier samples with high error and high leverage (large influence on the model). With this diagnostic, no outliers were detected. Results are summarized in Table 3. Relative errors of prediction (REP) (calculated dividing the RMSEP by the mean value of the concentrations) of 7.2% and 3.4% were obtained for oil and water, respectively. These results are less accurate than those previously reported by Jimenez et al.[7] In the mentioned work, relative errors of prediction estimated considering the mean of the concentration range (the exact distribution of the population is not given) represent about 3.3% and 1.8%. The authors used a NIR instrument equipped with optical filters, and the calibration was done by multiple linear regression (MLR). Nevertheless, it should be emphasized that only olives from the variety Picual were considered in the mentioned study. Furthermore, samples were collected from the tree and analyzed within 24 hr after collection. In the current work, much more variance of different origin is present in the data, which could be the reason for the less accurate predictions.

To evaluate if the previous classification of the olive samples could improve the quantification results, samples were divided in two groups (sound and damaged). Here, we focus only in the accurate analysis of sound olives. Results are summarized in Table 3. For the models constructed with damaged samples, results are similar to those with all the samples. Removing the sound samples does not reduce the complexity of the data set. This is reflected in the number of latent variables needed to construct the models, which were the same as in the models using all the samples.

Table 3. Results of the PLS models for oil and water determination in olives[a]

	All samples	Sound	Damaged
Oil			
Concentration range	18.9–35.3	18.9–28.8	22.2–35.3
Mean value	26.0	23.0	27.3
Standard deviation	3.43	2.24	3.07
No. latent variables	6	4	6
R	0.94	0.97	0.95
RMSEP	1.88	0.87	1.76
REP (%)	7.2	3.8	6.4
RPD	1.82	2.57	1.74
Water			
Concentration range	32.8–61.8	46.1–61.83	32.8–50.3
Mean value	46.3	53.2	43.2
Standard deviation	6.99	3.65	5.63
No. latent variables	7	4	7
R	0.98	0.97	0.99
RMSEP	1.71	1.49	1.76
REP (%)	3.4	2.8	4.1
RPD	4.09	2.45	3.20

PLS, partial least squres; RMSEP, root mean square error of prediction; REP, relative errors of prediction; RPD, ratio performance deviation.

[a] Models were constructed with first derivative and MSC correction. Prediction results correspond to a leave-one-out cross-validation procedure.

However, the models for sound samples were less complex: four latent variables were needed for both oil and water. These models gave much better results, with relative errors of prediction of 3.8% and 2.8% for oil and water content, respectively. These results are similar to those from Jimenez et al. or even better considering that in the current study eight different varieties of sound olives were included. To standardize the predictive ability, the ratio performance deviation (RPD)[25] was calculated (Table 3). The RPD values were greater than 2.0 for the models constructed with the sound samples.

CONCLUSIONS

The presented results show that NIR spectroscopy offers an excellent alternative to the traditional methods for olive fruit analysis. Additionally, it makes classification of the fruits possible. Due to the rapidity of the analysis, NIR spectroscopy can provide a tool for establishing a fruit sorting system. Inspection devices could be implemented directly in the production facility providing a means for selecting sound olives in order to produce high-quality virgin olive oil. Furthermore, it has been demonstrated that much more accurate prediction of water and oil content in olives can be obtained if fruits are previously classified in different categories and separate quantification models are created for each class.

REFERENCES

1. *Official Journal of the European Communities*. Commission Regulation (EEC). No. 2568/1991: On the characteristics of olive oil and olive-residue oil and on the relevant methods of analysis. Office for Official Publications of the European Communities, Luxembourg.
2. *Official Journal of the European Communities*. Commission Regulation (EEC). No. 1989/2003 amending Regulation (EEC) No. 2568/1991 on the characteristics of olive oil and olive-pomace oil and on the relevant methods of analysis. Office for Official Publications of the European Communities, Luxembourg.
3. American Oil Chemists'Society (AOCS). *Official Methods and Recommended practices*, Method no. Ba 3-38. Champaign, IL, USA, 2003.
4. Boskou, D. Olive harvesting and olive oil Extraction. In *Olive Oil. Chemistry and Technology*; AOCS Press: Champaign, IL, USA, 1997.
5. Martinez Suarez, J. M.; Muñoz Aranda, E.; Alba Mendoza, J.; Lanzón Rey, A. Informe sobre utilización del analizador de rendimientos "Abencor". *Grasas y Aceites.* **1975**, *26* (6), 379.
6. American Oil Chemists'Society (AOCS). *Official Methods and Recommended practices*, Method no. Ca 5a–40. Champaign, IL, USA, 1997.
7. Jiménez, A.; Izquierdo, E.; Rodríguez, F.; Dueñas, J. I.; Tortosa, C. Determination of fat and moisture in olives by near infrared reflectance spectroscopy. *Grasas y Aceites.* **2000**, *51*, 311.
8. Leon, L.; Rallo, L.; Garrido, A. Analysis of intact olive fruit by near-infrared spectroscopy (NIRS): An useful tool in olive breeding programs. *Grasas y Aceites.* **2003**, *54* (1), 41.
9. Leon, L.; Garrido-Varo, A.; Downey, G. Parent and harvest year effect on near-infrared reflectance spectroscopic analysis of olive (*Olea europaea* L.) fruit traits. *J. Agric. Food Chem.* **2004**, *52*, 4957.
10. IOOC, IUPAC; Method 2601 or ISO 662, Oxford, 1980.
11. Geladi, P.; McDougall, D.; Martens, H. Linearization and scatter-correction for near infrared reflectance spectra of meat. *Appl. Spectrosc.* **1985**, *39*, 491.
12. Savitzky, A.; Golay, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **1964**, *36* (8), 1627.
13. Ward, J. H. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* **1963**, *58*, 236.
14. Martens, H.; Naes, T. *Multivariate Calibration*. Wiley: Chichester, 1989.
15. Hermoso, M.; Uceda, M.; Frias, L.; Beltrán, G. Maduración. In *El cultivo del olivo*, 2nd ed; Fernández-Escobar, R., Barranco, D., Rallo, L., Eds.; Junta de Andalucia, Grupo Mundi-Prensa: Madrid, Spain, 1997.
16. Bünings_Pfaue, H. Analysis of water in food by near infrared spectroscopy. *Food. Chem.* **2003**, *82*, 107.
17. Hourant, P.; Baeten, V.; Morales, M. T.; Meurens, M.; Aparicio, R. Oil and fat classification by selected bands of near-infrared spectroscopy. *Appl. Spectrosc.* **2000**, *54*, 1168.

18. Murray, I.; Williams, P. C. Eds., Chemical principles of near-infrared technology. In *Near-Infrared Technology in the Agricultural and Food Industries*; Am. Assoc. Cereal Chem. Inc: St. Paul, pp. 17–34.
19. Frankel, E. N.; Nash, A. M.; Snyder, J. M. A methodology study to evaluate quality of soybeans stored at different moisture levels. *J. Am. Oil Chem. Soc.* **1987**, *64* (7), 987.
20. Morelló, J. R.; Motilva, M. J.; Ramo, T.; Romero, M. P. Effect of freeze injuries in olive fruit on virgin olive oil composition. *Food Chem.* **2003**, *81*, 547.
21. Yildiz, G.; Wehling, R. L.; Cuppett, S. L. Monitoring PV in corn and soybean oils by NIR spectroscopy. *J. Am. Oil Chem. Soc.* **2002**, *79* (11), 1085.
22. Cho, S. Y.; Kim, J. Y.; Rhee, C. l. Determination of rancidity of soybean oil by near infrared spectroscopy. *J. Near Infrared Spectrosc.* **1998**, *6* (1–4), A349.
23. Gergely, S.; Salgó, A. Changes in moisture content during wheat maturation-what is measured by near infrared spectroscopy? *J. Near Infrared Spectrosc.* **2003**, *11*, 17.
24. Azzouz, T.; Puigdoménech, M.; Aragay, M.; Tauler, T. Comparison of different data pre-treatment methods in the analysis of forage samples using near-infrared diffuse reflectance spectroscopy and partial least-squares multivariate calibration. *Anal. Chim. Acta.* **2003**, *484*, 121.
25. Williams, P. C.; Sobering, D. C. Comparison of commercial near infrared transmittance and reflectance instruments for analysis of whole grains and seeds. *J. Near Infrared Spectrosc.* **1993**, *1*, 25.