JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

J. Agric. Food Chem. **2009,** *57,* 1735–1741 **1735**

Confirmation of Food Origin Claims by Fourier Transform Infrared Spectroscopy and Chemometrics: Extra Virgin Olive Oil from Liguria

SIOBHÁN HENNESSY,*[,†,‡] GERARD DOWNEY,[†] AND COLM P. O'DONNELL[‡]

Teagasc, Ashtown Food Research Centre, Ashtown, Dublin 15, Ireland, and Biosystems Engineering, School of Agriculture, Food and Veterinary Medicine, University College Dublin, Belfield, Dublin 4, Ireland

The aim of this study was to explore the potential of Fourier transform infrared spectroscopy and various chemometric tools for confirming the geographic origin of olive oil from Liguria (northern Italy). Authentic extra virgin olive oil samples ($n = 913$) from three harvests (2004−2007) were collected from Italy, France, Spain, Greece, Cyprus, and Turkey—approximately one-fourth of all samples originated in Liguria. Attenuated total reflectance spectra were recorded at room temperature; the analytical challenge was to confirm that an oil which claimed to be from Liguria originated there. Derivative and standard normal variate data pretreatments were applied to the recorded spectra, which were subsequently analyzed by a number of multivariate procedures—principal component analysis, factorial discriminant analysis, and partial least-squares regression analysis. Prediction models created using samples from all three harvests had sensitivities and selectivities of approximately 0.80.

KEYWORDS: Fourier transform infrared spectroscopy; chemometrics; olive oil; geographic origin; PDO; authenticity; partial least squares regression; factorial discriminant analysis

INTRODUCTION

Virgin olive oils are defined as oils obtained from the fruit of the olive tree solely by mechanical or other physical means under conditions that do not lead to alterations in the oil; oils thus labeled must not have undergone any treatment other than washing, decanting, centrifuging, and filtering (*1*). Extra virgin olive oil is a superior category virgin olive oil that has compositional criteria specified by regulation (*1, 2*). Of particular significance is the free acidity expressed as oleic acid, which must not exceed 0.8 g/100 g of oil. Compared with other edible oils, olive oil is an expensive product typically costing 4−5 times as much as common vegetable oils such as corn or sunflower oil (*3*); this high market price is due to its highly prized organoleptic and nutritional properties. The taste and quality of virgin olive oils from different geographical origins varies as a consequence of distinct local agricultural traditions such as olive variety cultivated, oil extraction technique used, and oil blending practice (*4*). To protect the good name of a product that has been developed by substantial, long-term, collective and individual investment, the European Community (EU) has set regulations on the protection of claims relating to geographic indications and designations of origin for agricultural products and foodstuffs (*5, 6*). This means that packaged virgin olive oils produced in defined oil-producing regions may use a protected designation of origin (PDO) or protected geographical indication (PGI) on their labels. For this reason, geographic origin is an essential element of olive oil authenticity, and food processors, retailers, enforcement agencies, and consumers require an independent mechanism to confirm that any given virgin olive oil sample which claims PDO or PGI status actually complies with all relevant specifications (*7*). The EU-funded TRACE project (www.trace.eu.org) was set up with the strategic aim of developing cost-effective traceability methods and systems to provide consumers with added confidence in the authenticity of European food. The potential of a number of analytical fingerprinting techniques was investigated to achieve this aim.

Various methods have been applied to confirm claims concerning the geographic origin of olive oils. Mass spectroscopy (*8*−*11*) techniques have been used by a number of researchers with varying degrees of success, up to 100% correct classification in some cases. Olive oils from different regions within a single country have been classified using NMR spectroscopy with success rates of up to 90% (*12*−*14*). Fluorescence spectroscopy, which has been used for detecting olive oil adulteration (*15, 16*), has also been successfully applied to the issue of origin confirmation with 27 of 29 samples being correctly classed according to their designation of origin in one reported study (*17*). Chemical and physical characteristics of Ligurian olive oils, measured by HPLC and other methods

* Author to whom correspondence should be addressed (e-mail siobhan.hennessy@teagasc.ie).

[†] Teagasc.

[‡] University College Dublin.

10.1021/jf803714g CCC: $40.75
Published on Web 02/10/2009

Table 1. Number of Virgin Olive Oil Samples from Each Location

	Italy							
	Liguria	other	Spain	France	Greece	Turkey	Cyprus	total
2004/2005	63	163	42	9	25	14		316
2005/2006	79	173	38	10	46		6	352
2006/2007	68	116	34	20	7			245
3 harvests	210	452	114	39	78	14	6	913

defined by EU regulations, were studied (*18*), and chemometric models based on these parameters had sensitivities and selectivities of up to 0.87. These techniques all have specific advantages, but methods with high classification abilities such as NMR can be time-consuming and expensive and require complex sample preparation (*19, 20*). Vibrational spectroscopy methods such as near-infrared (NIR) and Fourier transform infrared (FT-IR) spectroscopy have been used in feasibility studies involving small numbers of samples for geographic origin confirmation of olive oil and have achieved classifications rates of up to 100% (*19*−*26*). These techniques are fast and simple to use, require no sample preparation, and do not destroy the sample (*27*).

Previous work based on two-thirds of the samples used in this study (*28*) employed classification and regression trees (CART) and support vector machines (SVM) to discriminate between oils of different origins. Models based on CART had low (0.11) sensitivities, although the selectivities were high (0.95). Models built using SVM had better sensitivity than CART models (0.59) and also had a high selectivity (0.94).

The objective of this study was to investigate the potential of using FT-IR spectroscopy with germanium ATR sampling and chemometric data processing techniques to confirm the origin of authentic Ligurian extra virgin olive oils using a large set of 913 extra virgin olive oils collected over three harvest periods (2004/2005, 2005/2006, and 2006/2007).

MATERIALS AND METHODS

Samples. Authentic extra virgin olive oil samples ($n = 913$) sourced in Italy, France, Spain, Greece, Cyprus, and Turkey were collected from reliable sources soon after olive harvesting over three harvest seasons (2004/2005, 2005/2006, and 2006/2007). Immediately after collection, samples were placed in the dark at 4 °C to minimize any risk of deterioration. After each harvest, oil samples were accumulated at a single location (Joint Research Centre, Ispra, Italy) and forwarded to Ashtown Food Research Centre by air. Almost one-fourth of all samples ($n = 210$) were from the Liguria region in northern Italy; the number of samples from each country can be seen in **Table 1**. Detailed information on the source of Ligurian olive oils was available including, in most cases, the varietal mix of olives involved and the growing location.

Spectral collection took place within 6 months of sample collection; that is, FT-IR spectral collection was carried out on separate occasions for samples from each olive oil harvest. Samples from each harvest were stored in the dark at 4 °C in headspace or simple screw-cap glass vials in batches of 16 samples from the same region. Two or three batches were selected at random, removed from chilled storage, and placed in a water bath at 25 °C for 1 h prior to spectral collection on each measurement day.

Instrumentation. FT-IR spectra were collected on a Bio-Rad Excalibur series FTS 3000 spectrometer (Analytica Ltd., Dublin, Ireland). Instrument control and spectral collection were performed using WIN-IR Pro (v. 3.0) software supplied by the instrument manufacturer. Samples were applied to an in-compartment benchmark attenuated total reflectance (ATR) trough plate using a 45° germanium crystal with 11 internal reflections (Specac Ltd., Kent, U.K.). One hundred and twenty-eight scans were co-added at a nominal resolution of 4 cm^{-1} with single-beam spectra of each sample being collected and ratioed against a background of air. Spectra were recorded over the frequency range of 600−4000 cm^{-1}.

Between samples, the ATR crystal was cleaned with Triton X-100 solution (1% w/w), rinsed with distilled water, and dried with a soft tissue. The spectral baseline recorded by the spectrometer was examined visually to ensure that no residue from the previous sample was retained on the crystal. All spectra were recorded at room temperature (21 ± 5 °C) without any nitrogen purge of the sample compartment.

Statistical Analysis. Means of the 128 co-added scans for each sample were used for multivariate data analysis. Spectra were exported from WIN-IR Pro as GRAMS files (ThermoGalactic, Salem, NH) and imported directly into The Unscrambler (v 9.7; Camo A/S, Oslo, Norway).

Data pretreatments examined were (i) standard normal variate (SNV) (*29*) and (ii) first and second derivatives using a quadratic Savitzky−Golay method and segment sizes between 5 and 21 datapoints (*30*). Principal component analysis (PCA) (*31*) was performed using The Unscrambler on the entire sample set for preliminary data set examination. Factorial discriminant analysis (FDA) was executed using the SAISIR (*32*) environment for MATLAB developed by Bertrand et al. (*33*). Data files from The Unscrambler were exported as MATLAB files and imported into MATLAB (v 7.2.3.232 (R2006a), The Mathworks Inc., Cambridge, U.K.). FDA was applied in two steps; first, a PCA was carried out on the spectra and then FDA was performed on the principal component (PC) scores. The first step creates a set of orthogonal spectral patterns or principal components, and the second step calculates discriminant factors using a stepwise procedure to identify and incorporate those principal components that best discriminate the samples into the relevant groups, in this case, country of origin (*34, 35*). The key feature of this technique is that the principal components are incorporated into the discriminant model in such a way as to maximize discriminant ability according to the characteristic of interest rather than in numerical order: PC1 alone, PC1 + PC2, etc. Each sample is assigned to one of the classes of interest; in this study there were two classes, that is, olive oils from Liguria and all other olive oils. To eliminate any potential problems arising from sample set imbalance, all models were created using an equal number of Ligurian and non-Ligurian samples; to ensure the non-Ligurian class was as varied as possible, samples were selected at random from each country.

Partial least-squares discriminant analysis (PLS-DA) optimized by leave-one-out cross-validation (*36*) was used to discriminate between Ligurian and non-Ligurian olive oils. Discriminant regression models were created and validated using separate calibration and validation sample sets. Dummy *Y*-variables were used when models were created; a sample was assigned a value of 0 if it was from Liguria and 1 otherwise. PLS models thus developed were used to predict the value of the *Y*-variable for each validation sample; given the values of the dummy *Y*-variables used, an empirical and not entirely arbitrary value of 0.5 was used as a cutoff for identity confirmation with oils with predicted *Y*-values of <0.5 deemed to be from Liguria. Spectra with and without pretreatments were used to create PLS models. Martens' Uncertainty Test (*37*) was used to identify those variables of greatest importance in model development; these variables were then used exclusively to create new models, and the prediction procedure was repeated.

Data analysis was done using two approaches. The first approach involved the application of PLS-DA to each of the three harvests individually. In each harvest season, the calibration set comprised two-thirds of the Ligurian samples selected at random and an equal number of randomly chosen non-Ligurian olive oils. The validation set was made up of the remaining samples from that harvest. In the second approach a PLS-DA was performed that included samples from all three harvests in both the calibration and validation sets.

RESULTS AND DISCUSSION

Recorded ATR spectra (4000−600 cm^{-1}) of five randomly selected extra virgin olive oil samples are shown in **Figure 1**. High-intensity absorption peaks due to triglycerides, a major

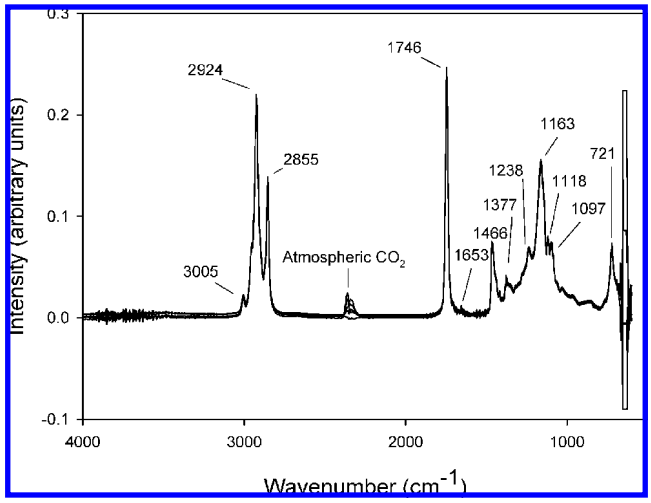

Figure 1. ATR/FT-IR spectrum (4000−600 cm^{-1}) of five randomly selected extra virgin olive oil samples showing the location of major absorption peaks.

Table 2. Functional Groups and Modes of Vibration in Olive Oil Spectra

frequency (cm^{-1})	functional group	mode of vibration
3005	=C−H (*cis*-)	stretching
2924	−C−H (CH_2)	stretching (asymmetrical)
2855	−C−H (CH_2)	stretching (symmetrical)
1746	−C=O (ester)	stretching
1653	−C=C− (*cis*-)	stretching
1466	−C−H (CH_2, CH_3)	bending (scissoring)
1377	−C−H (CH_3)	bending (symmetrical)
1238	−C−O, $-CH_2-$	stretching, bending
1163	−C−O, $-CH_2-$	stretching, bending
1118	−C−O	stretching
1097	−C−O	stretching
721	$-(CH_2)_n-$, −HC=CH− (*cis*-)	bending (rocking)

component in edible oils (*38*), can be seen at 2924, 2855, 1746, and 1163 cm^{-1}. The peaks at 2924 and 2855 cm^{-1} can be assigned to asymmetrical stretching vibrations of C−H bonds, whereas the peak at 1746 arises from C=O bond stretching. Stretching and bending of C−O and $-CH_2-$ account for the maximum at 1163 cm^{-1}. These spectra contain significant structure, particularly in the region between 1500 and 900 cm^{-1}, often referred to as the fingerprint region for edible oil (*39*, *40*). The functional groups and modes of vibration associated with the maxima in **Figure 1** are shown in **Table 2**, which has been adapted from Guillén and Cabo (*41*).

Attenuated frequency ranges (e.g., 1200−800 cm^{-1}) have been deployed for the detection of adulteration in olive oil (*38*, *40*, *42*) but, as the samples in this study were known to be authentic and the property of interest was the more problematic geographic origin, a wider frequency range was examined. The region 3000−700 cm^{-1} was chosen for data analysis as it (a) includes the fingerprint region and (b) avoids noise that was present at both extremes of the spectral range. As there was no nitrogen purge in the sample chamber, adventitious and variable absorption due to atmospheric carbon dioxide could be detected in the frequency range 2400−2250 cm^{-1} (**Figure 1**); this spectral range was not used in any model development or deployment. The complete frequency range studied was therefore 3000−2400 + 2250−700 cm^{-1}.

Principal Component Analysis. A PCA was performed on the complete frequency range of raw spectral data. Principal components 1−5 accounted for 59, 24, 9, 3, and 1% of the total variance in the spectral data set, respectively. Four oil samples

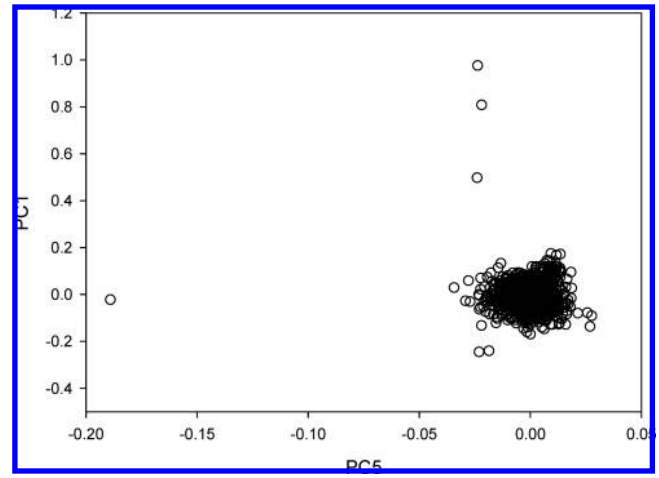

Figure 2. Score plot (PC 1 vs PC 5) from a PCA on raw spectral data of all samples (frequency range of 3000−2400 + 2250−700 cm^{-1}). Atypical points in the lower left and upper right section were excluded from further chemometric analysis.

appear to account for a significant amount of this variance as evidenced by their location at some distance from the main cluster of samples in the PC1 versus PC5 score plot (**Figure 2**). Spectra of all samples (n = 913) were plotted, and three samples showed significant absorption in the region 3000−3500 cm^{-1}; absorption in this frequency range is characteristic of water absorption (*43*), implying the presence of water in the samples. A fourth sample exhibited a baseline offset. These atypical spectra originated from the four samples referred to above and, because none of them originated in the geographic region of interest (Liguria), it was decided to leave them out of all further chemometric analyses.

Factorial Discriminant Analysis. FDA was used to distinguish between samples from Liguria and samples from other areas; results obtained can be seen in **Tables 3** and **4**. In the first step of the procedure, 20 principal components were calculated by PCA and between 3 and 10 components were used for FDA.

Validation results for models developed using olive oils from each harvest applied to other samples from the same harvest may be seen in **Table 3**. Accuracy of models for the 2004/2005 harvest was not as high as for other harvests; sensitivities were approximately 0.5 and selectivities of up to 0.7 were obtained, whereas models for the other two harvests produced sensitivities and selectivities of about 0.8. The model producing best results for the 2006/2007 harvest was developed using raw spectral data and used nine PCs. The first and second PCs accounted for 76.5 and 15.9%, respectively, of the variance in the data set. The rank order of PCs for inclusion in this model was PC4, PC9, PC3, PC18, PC20, PC2, PC12, PC5, and PC15.

Further models were developed using all available Ligurian samples from a particular harvest and an equal number of non-Ligurian samples from that same harvest. These models were applied to samples of the other two harvests (results not shown here), but, in general, when sensitivities were high, selectivities were low and vice versa. This is not unexpected given allegorical evidence concerning differences in weather between these three growing seasons and suggests that a better approach may be to generate a factorial discriminant model by amalgamating calibration samples from many harvests.

To examine the extent of the differences in samples from each harvest season, an FDA was executed on the raw spectral data from the three harvests. In this analysis the class of interest

Table 3. FDA Model Performances on Individual Harvests

	2004/2005			2005/2006			2006/2007	
pretreatment	no. of PC[a]	sensitivity	selectivity	no. of PC[a]	sensitivity	selectivity	no. of PC[a]	sensitivity	selectivity
raw data	5	0.67	0.62	9	0.65	0.77	9	0.87	0.75
first derivative									
5 point gap	8	0.48	0.65	7	0.81	0.77	8	0.65	0.76
9 point gap	7	0.48	0.67	8	0.77	0.74	8	0.78	0.69
13 point gap	4	0.52	0.67	7	0.69	0.76	6	0.78	0.78
21 point gap	5	0.43	0.71	3	0.73	0.73	5	0.78	0.76
second derivative									
5 point gap	4	0.52	0.60	4	0.65	0.66	10	0.70	0.62
9 point gap	6	0.57	0.55	5	0.58	0.74	7	0.74	0.73
13 point gap	6	0.52	0.55	9	0.77	0.74	4	0.83	0.69
21 point gap	8	0.62	0.74	8	0.73	0.77	7	0.83	0.69
SNV	5	0.48	0.65	7	0.73	0.74	9	0.83	0.72

[a] Number of principal components used in FDA model.

Table 4. FDA Model Performances Using Samples from Three Combined Harvests

pretreatment	no. of PC[b]	sensitivity	selectivity
none	7	0.70	0.73
first derivative			
5 point gap	4	0.69	0.72
9 point gap	8	0.69	0.72
13 point gap	8	0.71	0.70
21 point gap	6	0.66	0.72
second derivative			
5 point gap	3	0.69	0.66
9 point gap	3	0.71	0.67
13 point gap	4	0.74	0.72
21 point gap	5	0.61	0.70
SNV	3	0.63	0.68

[b] Number of principal components used in FDA model.

was harvest season rather than origin. A plot of the discriminant scores can be seen in **Figure 3**. Samples from the 2006/2007 harvest are easily distinguished from those of the other two harvests. There is a slight overlap between the samples from the 2004/2005 harvest and the 2005/2006 harvest. All of the 2006/2007 validation samples were correctly classified as well as 99% of the 2004/2005 set. Some samples from the harvest 2005/2006 (21%) were incorrectly classified as belonging to the 2004/2005 set, but the rest were correctly identified. The inclusion order of PCs was PC8, PC7, PC1, PC3, PC11, PC2, PC5, PC16, PC12, and PC6. PC1, despite accounting for 65.6% of the variance in the spectral data set, was ranked as the third most important PC in terms of discriminating samples on the basis of harvest season.

Samples from all three harvests were used as a calibration ($n = 280$) set in an FDA modeling origin (Ligurian vs non-Ligurian), and the prediction results on the corresponding validation sample set ($n = 629$) are shown in **Table 4**. Regardless of pretreatment employed, sensitivities and selectivities of around 0.70 were obtained. As an example, the model developed using spectral data with no pretreatments used seven PCs, in the order PC20, PC6, PC8, PC14, PC7, PC9, and PC2, to obtain the best discrimination of samples according to origin. It is notable that this model does not include PC1, whereas PC2 is ranked seventh. Lower PCs, which model smaller variances in the data, are more important in distinguishing Ligurian and non-Ligurian samples. The spectral pattern of the discriminant factor is shown in **Figure 4**, which also includes a typical olive

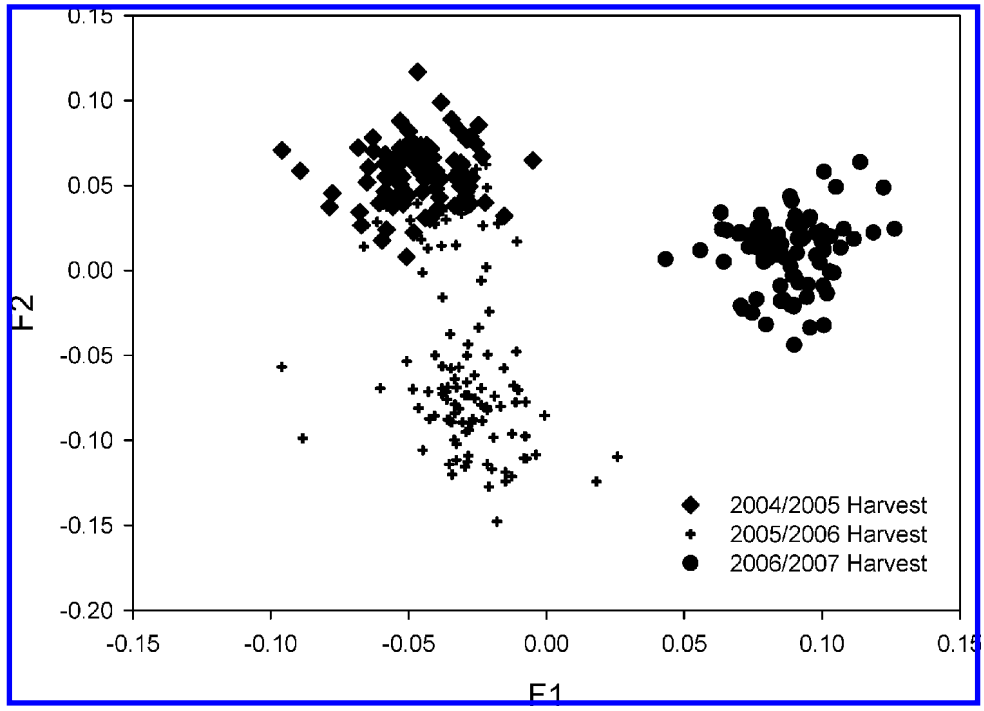

Figure 3. FDA score plot on spectral data (3000−2400 + 2250−700 cm^{-1}) with no pretreatment using harvest season as the class of interest.

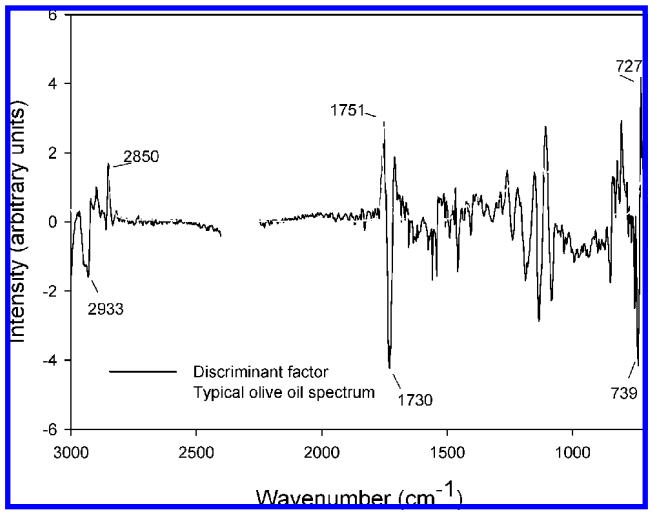

Figure 4. Spectral pattern of discriminant factor (3000−2400 + 2250−700 cm^{-1}, no pretreatment) for three harvest seasons.

oil spectrum. The discriminant profile shows a minimum and maximum at approximately 2933 and 2850 cm^{-1}, which may correspond to peaks at 2924 and 2855 cm^{-1} caused by the stretching of a C—H bond (**Table 2**). There is a strong maximum and minimum at 1751 and 1730 cm^{-1}, which may be due to the stretching of C═O bond in this region of the spectrum. A minimum and maximum can also be seen at approximately 739 and 727 cm^{-1}.

Partial Least Squares Discriminant Analysis. An alternative discriminant procedure applied to this data set was PLS-DA. Models were used to predict the identity of samples, that is, whether the samples were from Liguria. First, the spectra from each harvest were analyzed separately to determine how well samples from a particular harvest could identify other samples from that same harvest (**Table 5**). The sensitivity of the models for the 2004/2005 harvest ranged between 0.33 and 0.62, whereas the corresponding selectivities ranged between 0.55 and 0.74. The model with the highest sensitivity (0.62) was created from second-derivative spectra with a 9 data point gap and 4 loadings; the selectivity of this model was 0.66. The 2005/2006 harvest had slightly better prediction results than the 2004/2005 harvest with a range of sensitivity values between 0.71 and 0.88 and selectivities between 0.69 and 0.80. The best model had a sensitivity of 0.88 and a selectivity of 0.80. This was based on first-derivative spectra with a 13 datapoint gap and 8 loadings. The figures for 2006/2007 were similar to those for 2005/2006. The sensitivity range lay between 0.65 and 0.91, and the selectivity range was between 0.58 and 0.78. The sensitivity of the best model for the 2006/2007 was 0.91, and the corresponding selectivity was 0.77; this model was created using first-derivative data with a 21 datapoint gap and involved 7 loadings.

All models were recalculated after variable selection by Martens′ Uncertainty Test, but no improvement in model performance was noted.

As with FDA, models from one harvest were applied to another, and when sensitivity was high, selectivity was low and vice versa. These models (results not shown) classified the majority of the samples as being from the one class (Ligurian or non-Ligurian). The high or low selectivities varied according to the pretreatment applied, so results based on the same calibration and validation samples were inconsistent.

It is unlikely that samples from a particular year′s harvest will contain all possible variance representative of samples from Liguria so, as was done with FDA, calibration models consisting of Ligurian samples (n = 140) and non-Ligurian samples (n = 140) from the three harvests were used to predict the origin of a separate validation set comprising all remaining samples. Results for this analysis can be seen in **Table 6**. Sensitivities were in the range from 0.59 to 0.79 and selectivities spanned 0.62−0.78 for models developed using frequencies 3000−2400 + 2250−700 cm^{-1}. Models created using variables selected by Martens′ Uncertainty Test gave almost the same results, but, in general, fewer loadings were required. The best overall result, achieved using a model built with first-derivative spectra and

Table 5. Summary of PLS Predictions (Frequency Range of 3000−2400 + 2250−700 cm^{-1}) for Each Harvest

	2004/2005			2005/2006			2006/2007	
pretreatment	no. of L[c]	sensitivity	selectivity	no. of L[c]	sensitivity	selectivity	no. of L[c]	sensitivity	selectivity
raw data	4	0.48	0.59	9	0.85	0.75	7	0.74	0.71
first derivative									
5 point gap	4	0.57	0.68	5	0.81	0.76	3	0.74	0.73
9 point gap	5	0.43	0.72	5	0.71	0.72	6	0.74	0.70
13 point gap	2	0.57	0.55	8	0.88	0.80	8	0.78	0.77
21 point gap	3	0.57	0.59	10	0.85	0.74	7	0.91	0.77
second derivative									
5 point gap	6	0.33	0.71	5	0.85	0.71	3	0.70	0.60
9 point gap	4	0.62	0.66	5	0.77	0.75	7	0.65	0.58
13 point gap	4	0.43	0.67	5	0.77	0.74	8	0.70	0.64
21 point gap	4	0.52	0.64	6	0.85	0.69	11	0.70	0.71
SNV	7	0.48	0.74	9	0.85	0.70	9	0.78	0.78
Models Created with Variables Selected Using Martens′ Uncertainty Test									
raw data	9	0.52	0.78	6	0.85	0.69	5	0.78	0.73
first derivative									
5 point gap	11	0.67	0.82	3	0.85	0.73	7	0.74	0.73
9 point gap	6	0.57	0.75	7	0.88	0.75	6	0.78	0.72
13 point gap	7	0.71	0.84	8	0.88	0.74	6	0.83	0.76
21 point gap	9	0.67	0.83	8	0.85	0.76	6	0.78	0.80
second derivative									
5 point gap	10	0.48	0.80	3	0.81	0.71	7	0.78	0.68
9 point gap	10	0.76	0.73	3	0.85	0.74	4	0.70	0.66
13 point gap	3	0.62	0.76	2	0.81	0.71	6	0.74	0.66
21 point gap	7	0.67	0.86	2	0.73	0.78	7	0.78	0.73
SNV	10	0.52	0.82	8	0.81	0.72	8	0.78	0.77

[c] Number of PLS loadings.

Table 6. Summary of PLS Prediction Results for Samples from Three Combined Harvests

	all variables[a]			selected variables	
pretreatment	no. of L[b]	sensitivity	selectivity	no. of L[b]	sensitivity	selectivity
none	11	0.74	0.69	14	0.70	0.72
first derivative						
5 point gap	10	0.73	0.75	12	0.81	0.77
9 point gap	12	0.79	0.76	9	0.86	0.78
13 point gap	15	0.79	0.78	8	0.81	0.76
21 point gap	14	0.71	0.77	9	0.76	0.79
second derivative						
5 point gap	3	0.59	0.62	3	0.64	0.68
9 point gap	3	0.67	0.70	2	0.62	0.64
13 point gap	13	0.78	0.73	7	0.79	0.76
21 point gap	15	0.75	0.75	6	0.76	0.78
SNV	11	0.70	0.72	10	0.76	0.76

[a] 3000−2400 + 2250−700 cm^{-1}. [b] Number of PLS loadings.

variables selected with Martens' Uncertainty Test (9 data point gap, 9 loadings), was a sensitivity of 0.86 with a corresponding selectivity of 0.78.

For both FDA and PLS-DA, the predictions based on samples from the three harvests together are as good as if not better than predictions based on samples from an individual harvest. Any models generated from these samples will have increased robustness by incorporating variations due to year of harvest. Realistically, it is desirable to have a sample set comprising as many years as possible as there is always the possibility of unusual weather conditions in any particular year that will invariably affect the models characterizing the samples from a particular region, namely, Liguria. Boggia et al. reported that climatic conditions and a *Dacus oleae* infestation affected the prediction ability of models based on Ligurian samples in a classification study of samples from the three main geographical areas specified in the PDO "Riviera Ligure" (*18*).

Information on olive variety was available for most of the Ligurian samples, but not for a large number of the non-Ligurian oils. As PLS-DA gave slightly better results than FDA, Ligurian samples that were classed as non-Ligurian after a PLS-DA were examined to see if this error could be related to olive variety. The olive cultivar most common among the Ligurian samples is Taggiasca, and this cultivar is the main or the sole olive type used to make two-thirds of the Ligurian samples ($n = 140$) from the three harvests. In the PLS-DA using samples from each of the harvests, there are 70 Ligurian samples in the validation set. Of these 70 samples, 46 samples were made from Taggiasca (or mainly Taggiasca) olive oils and 24 samples were not. In this analysis, 20 PLS models were created from spectra with and without pretreatments and variable selection. For the purpose of examining misclassified Ligurian samples, a sample was considered to be misclassified if 5 or more of these models misclassified that particular sample. According to this criterion, 24 Ligurian samples were falsely classified as not being from Liguria, 18 produced from Taggiasca olives. This means that 39% of the Taggiasca olive oils and 25% of the other olive oils were misclassified. This result does not indicate that these classification errors are due to olive cultivar, but no definitive conclusion may be made because the exact proportions of Taggiasca in some oil samples are unknown.

This study demonstrates the potential of FT-IR spectroscopy with ATR sampling and chemometrics to discriminate between Ligurian and non-Ligurian olive oils from three harvests. PLD-DA was shown to be slightly more successful than FDA in classifying this data set with sensitivities and selectivities of approximately 0.80 as opposed to 0.70. The most successful chemometric models from both techniques were those derived from data sets that contained samples from three harvest seasons incorporating variation due to weather conditions in the models. When models created using samples from one or even two harvests were used to classify samples from a different harvest, the models were unsuccessful. In most cases, models that had high sensitivities had low corresponding selectivities, and when selectivities were high, sensitivities were low. A 20% possibility of a model rejecting a Ligurian sample or falsely accepting a non-Ligurian sample means that this technique may not be a commercially viable stand-alone technique for confirming geographic origin, but the speed and low cost at which measurements can be made suggest that it could certainly be used as a screening technique.

LITERATURE CITED

(1) Codex Alimentarius Commission. Standard for olive oils and olive pomace oils. *Codex Stan 33* **2003**, 8.

(2) European Commission. Commission Regulations (EC) No. 1019/2002 of 13 June 2002 on the marketing standards for olive oil. *Off. J. Eur. Communities* **2002**, *L155*, 27−31.

(3) European Commission. Working paper of the Directorate-General for Agriculture: The olive oil and table olives sector. http://ec.europa.eu/agriculture/markets/olive/reports/rep_en.pdf.

(4) European Commission. The olive oil sector in the European Union. *Agriculture and Rural Development—Fact-sheet* **2003**, 1−6.

(5) Barjolle, D.; Sylvander, B. Some factors of success for origin labelled products in agri-food supply chains in Europe: Market, Internal Resources and Institutions. *Econ. Soc., Cah. ISMEA, Ser. Dev. Agroaliment.* **2002**, *25*, 1–31.

(6) European Commission. Council Regulation (EC) No. 510/2006 of 20 March 2006 on the protection of geographical indication and designations of origin for agricultural products and foodstuffs. *Off. J. Eur. Union* **2006**, *L 93*, 12−25.

(7) European Commission. Regulation (EC) No 882/2004 of the European Parliament and of the Council of 29 April 2004 on official controls performed to ensure the verification of compliance with feed and food law, animal health and animal welfare rules. *Off. J. Eur. Union* **2004**, *L165*, 1−141.

(8) Araghipour, N.; Colineau, J.; Koot, A.; Akkermans, W.; Rojas, J. M. M.; Beauchamp, J.; Wisthaler, A.; Mark, T. D.; Downey, G.; Guillou, C.; Mannina, L.; van Ruth, S. Geographical origin classification of olive oils by PTR-MS. *Food Chem.* **2008**, *108* (1), 374–383.

(9) Benincasa, C.; Lewis, J.; Perri, E.; Sindona, G.; Tagarelli, A. Determination of trace element in Italian virgin olive oils and their characterization according to geographical origin by statistical analysis. *Anal. Chim. Acta* **2007**, *585*, 366–370.

(10) Cavaliere, B.; DeNino, A.; Hayet, F.; Lazez, A.; Macchione, B.; Moncef, C.; Perri, E.; Sindona, G.; Tagarelli, A. A metabolomic approach to the evaluation of the origin of extra virgin olive oil: a convenient statistical treatment of mass spectrometric analytical data. *J. Agric. Food Chem.* **2007**, *55*, 1454–1462.

(11) Casale, M.; Armanino, C.; Casolino, C.; Forina, M. Combining information from headspace mass spectrometry and visible spectroscopy in the classification of the Ligurian olive oils. *Anal. Chim. Acta* **2007**, *589*, 89–95.

(12) Petrakis, P. V.; Agiomyrgianaki, A.; Christophoridou, S.; Spyros, A.; Dais, P. Geographical characterization of Greek virgin olive oils (cv. Koroneiki) using H-1 and P-31 NMR fingerprinting with canonical discriminant analysis and classification binary trees. *J. Agric. Food Chem.* **2008**, *56*, 3200–3207.

(13) Vlahov, G.; Del Re, P.; Simone, N. Determination of geographical origin of olive oils using ^{13}C nuclear magnetic resonance spectroscopy. 1. Classification of olive oils of the Puglia region with Denomination of Protected Origin. *J. Agric. Food Chem.* **2003**, *51*, 5612–5615.

(14) Mannina, L.; Patumi, M.; Proietti, N.; Bassi, D.; Segre, A. L. Geographical characterization of Italian extra virgin olive oils

using high-field ^{1}H NMR spectroscopy. *J. Agric. Food Chem.* **2001**, *49*, 2687–2696.
(15) Sayago, A.; Garcia-Gonzalez, D. L.; Morales, M. T.; Aparicio, R. Detection of the presence of refined hazelnut oil in refined olive oil by fluorescence spectroscopy. *J. Agric. Food Chem.* **2007**, *55*, 2068–2071.
(16) Poulli, K. I.; Mousdis, G. A.; Georgiou, C. A. Rapid synchronous fluorescence method for virgin olive oil adulteration assessment. *Food Chem.* **2007**, *105*, 369–375.
(17) Dupuy, N.; Le Dréau, Y.; Ollivier, D.; Artaud, J.; Pinatel, C.; Kister, J. Origin of French virgin olive oil registered designation of origins predicted by chemometric analysis of synchronous excitation−emission fluorescence spectra. *J. Agric. Food Chem.* **2005**, *53*, 9361–9368.
(18) Boggia, R.; Zunin, P.; Lanteri, S.; Rossi, N.; Evangelisti, F. Classification and class-modeling of "Riviera Ligure" extra-virgin olive oil using chemical−physical parameters. *J. Agric. Food Chem.* **2002**, *50*, 2444–2449.
(19) Tapp, H. S.; Defernez, M.; Kemsley, E. K. FTIR spectroscopy and multivariate analysis can distinguish the geographical origin of extra virgin olive oils. *J. Agric. Food Chem.* **2003**, *51*, 6110–6115.
(20) Galtier, O.; Le, D. Y.; Ollivier, D.; Kister, J.; Artaud, J.; Dupuy, N. Lipid compositions and French registered designations of origins of virgin olive oils predicted by chemometric analysis of mid-infrared spectra. *Appl. Spectrosc.* **2008**, *62*, 583–590.
(21) Bendini, A.; Cerretani, L.; Di Virgilio, F.; Belloni, P.; Bonoli-Carbognin, M.; Lercker, G. Preliminary evaluation of the application of the FTIR spectroscopy to control the geographic origin and quality of virgin olive oils. *J. Food Qual.* **2007**, *30*, 424–437.
(22) Galtier, O.; Dupuy, N.; Le Dreau, Y.; Ollivier, D.; Pinatel, C.; Kister, J.; Artaud, J. Geographic origins and compositions of virgin olive oils determinated by chemometric analysis of NIR spectra. *Anal. Chim. Acta* **2007**, *595*, 136–144.
(23) Casale, M.; Casolino, C.; Ferrari, G.; Forina, M. Near infrared spectroscopy and class modelling techniques for the geographical authentication of Ligurian extra virgin olive oil. *J. Near Infrared Spectrosc.* **2008**, *16*, 39–47.
(24) Downey, G.; McIntyre, P.; Davis, A. N. Geographic classification of extra virgin olive oils From the Eastern Mediterranean by chemometric analysis of visible and near-infrared spectroscopic data. *Appl. Spectrosc.* **2003**, *57*, 158–163.
(25) Sinelli, N.; Cosio, M. S.; Gigliotti, C.; Casiraghi, E. Preliminary study on application of mid infrared spectroscopy for the evaluation of the virgin olive oil "freshness". *Anal. Chim. Acta* **2007**, *598*, 128–134.
(26) Gurdeniz, G.; Ozen, B.; Tokatli, F. Classification of Turkish olive oils with respect to cultivar, geographic origin and harvest year, using fatty acid profile and mid-IR spectroscopy. *Eur. Food Res. Technol.* **2008**, *227*, 1275–1281.
(27) Downey, G.; McIntyre, P.; Davies, A. N. Detecting and quantifying sunflower oil adulteration in extra virgin olive oils from the Eastern Mediterranean by visible and near-infrared spectroscopy. *J. Agric. Food Chem.* **2002**, *50*, 5520–5525.
(28) Caetano, S.; Üstün, B.; Hennessy, S.; Smeyers-Verbeke, J.; Melssen, W.; Downey, G.; Buydens, L.; Heyden, Y. V. Geographical classification of olive oils by the application of CART and SVM to their FT-IR. *J. Chemom.* **2007**, *21*, 324–334.
(29) Barnes, R. J.; Dhanoa, M. S.; Lister, S. J. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Appl. Spectrosc.* **1989**, *43*, 772–777.
(30) Savitzky, A.; Golay, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **1964**, *36*, 1627–1639.
(31) Martens, H.; Næs, T. Methods for calibration. In *Multivariate Calibration*; Wiley: New York, 1993; pp 97−100.
(32) Bertrand, D. SAISIR 2008: Package of function for chemometrics in the MATLAB® environment, http://easy-chemometrics.fr.
(33) Bertrand, D.; Courcoux, P.; Autran, J.-C.; Meritan, R.; Robert, P. Stepwise canonical discriminant analysis of continuous digitized signals: application to chromatograms of wheat proteins. *J. Chemom.* **1990**, *4*, 413–427.
(34) Devaux, M. F.; Bertrand, D.; Robert, P.; Qannari, M. Application of multidimensional analyses to the extraction of discriminant spectral patterns from NIR spectra. *Appl. Spectrosc.* **1988**, *42*, 1015–1019.
(35) Downey, G.; Robert, P.; Bertrand, D.; Kelly, P. M. Classification of commercial skim milk powders according to heat treatment using factorial discriminant analysis of near-infrared spectra. *Appl. Spectrosc.* **1990**, *44*, 150–155.
(36) Roussel, S.; Bellon-Maurel, V.; Roger, J.-M.; Grenier, P. Authenticating white grape must variety with classification models based on aroma sensors, FT-IR and UV spectrometry. *J. Food Eng.* **2003**, *60*, 407–419.
(37) Martens, H.; Martens, M. Modified jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR). *Food Qual. Pref.* **2000**, *11*, 5–16.
(38) Yang, H.; Irudayaraj, J.; Paradkar, M. M. Discriminant analysis of edible oils and fats by FTIR, FT-NIR and FT-Raman spectroscopy. *Food Chem.* **2005**, *93*, 25–32.
(39) Van de Voort, F.; Sedman, J.; Russin, T. Lipid analysis by vibrational spectroscopy. *Eur. J. Lipid Sci. Technol.* **2001**, *103*, 815–840.
(40) Baeten, V.; Fernández Pierna, J. A.; Dardenne, P.; Meurens, M.; García-González, D. L.; Aparicio-Ruiz, R. Detection of the presence of hazelnut oil in olive oil by FT-Raman and FT-MIR spectroscopy. *J. Agric. Food Chem.* **2005**, *53*, 6201–6206.
(41) Guillén, M.; Cabo, N. Infrared spectroscopy in the study of edible oils and fats. *J. Sci. Food Agric.* **1997**, *75*, 1–11.
(42) Tay, A.; Singh, R. K.; Krishnan, S. S.; Gore, J. P. Authentication of olive oil adulterated with vegetable oils using Fourier transform infrared spectroscopy. *Lebensm.-Wiss. -Technol.* **2002**, *35*, 99–103.
(43) Stuart, B. Industrial and environmental applications. In *Infrared Spectroscopy: Fundamentals and Applications*; Wiley: West Sussex, U.K., 2004; pp174.

Received for review October 6, 2008. Revised manuscript received January 13, 2009. Accepted January 14, 2009. Funding for this research was provided by the EC via the TRACE FP6 project (www.trace.eu.org). The information contained in this paper reflects the authors' views; the European Commission is not liable for any use of the information contained therein.

JF803714G