

## Research Paper

# Comparison of fatty acid profiles and mid-infrared spectral data for classification of olive oils

Gozde Gurdeniz, Banu Ozen and Figen Tokatli

Izmir Institute of Technology, Department of Food Engineering, Urla-Izmir, Turkey

The composition of olive oils may vary depending on environmental and technological factors. Fatty acid profiles and Fourier-transform infrared (FT-IR) spectroscopy data in combination with chemometric methods were used to classify extra-virgin olive oils according to geographical origin and harvest year. Oils were obtained from 30 different areas of northern and southern parts of the Aegean Region of Turkey for two consecutive harvest years. Fatty acid composition data analyzed with principal component analysis was more successful in distinguishing northern olive oil samples from southern samples compared to spectral data. Both methods have the ability to differentiate olive oil samples with respect to harvest year. Partial least squares (PLS) analysis was also applied to detect a correlation between fatty acid profile and spectral data. Correlation coefficients ( $R^2$ ) of a calibration set for stearic, oleic, linoleic, arachidic and linolenic acids were determined as 0.83, 0.97, 0.97, 0.83 and 0.69, respectively. Fatty acid profiles were very effective in classification of oils with respect to geographic origin and harvest year. On the other hand, FT-IR spectra in combination with PLS could be a useful and rapid tool for the determination of some of the fatty acids of olive oils.

**Keywords:** Classification / Fatty acid profile / Infrared spectroscopy / Olive oil

Received: October 10, 2008; accepted: March 19, 2009

DOI 10.1002/ejlt.200800229

## 1 Introduction

The recent popularity of olive oil could be attributed to not only its sensorial characteristics but also its potential health benefits. These benefits have been related to its well-balanced fatty acid composition, where oleic acid is the main component, and to the presence of minor biomolecules, such as vitamins and natural antioxidants [1].

The composition of olive oils varies with geographic origin due to a number of different factors: regional differences in climate, soil, and agricultural practice. In Europe, certification of extra-virgin olive oils according to the geographical origin is assured by the Designation of Origin regulations [Protected Designation of Origin (PDO) and Protected Geographical Indication (PGI)]. Classification according to the area of production of high-commercial-value oils exploits the information about the content of minor constituents (phenolic

acids, waxes, sterols, hydrocarbons) as well as subtle differences in the concentration of major components [2, 3].

The fatty acid profile determined by gas chromatography supplies extremely useful information for the classification studies. There are several studies involving the use of the fatty acid composition in the classification of olive oils. Differentiation of monovarietal olive oils by their basic classes of compounds including fatty acids was achieved using principal component analysis (PCA) and linear discriminant analysis [4]. It was stated that the genotype has a higher influence on the fatty acid composition than the microclimate [5]. Also, a study on Sicilian olive oils indicated that the effect of the cultivar is predominant in the olive oil classification based on fatty acid composition, but a minor but well-defined geographic effect is also present [6].

Although chromatographic methods supply a high degree of classification, there is an increasing demand for rapid, inexpensive and effective techniques for the determination of the authenticity of olive oils, and infrared (IR) spectroscopy combined with chemometric techniques is one of the promising rapid methods. IR spectroscopy, a vibrational spectroscopic technique, has advantages such as high speed of measurement, moderate instrument cost and relative ease of sample

**Correspondence:** Banu Ozen, Izmir Institute of Technology, Department of Food Engineering, 35430 Urla-Izmir, Turkey.

**E-mail:** banuozen@iyte.edu.tr

**Fax:** +90 232 7506196

preparation, especially for liquid and paste materials [7]. Changes in dipole moment of molecules due to absorbance of energy from a light source in the IR region produce the IR spectrum. Peaks observed in the IR profile of a fat or oil sample correspond to structural or functional groups. When analyzed with multivariate statistical techniques, even subtle differences in the IR spectra of fats and oils would become useful for characterization and classification studies. IR spectroscopy could also provide quantitative information such as *trans* content, iodine value, peroxide value and free fatty acid content about fats and oils [8]. Pure extra-virgin olive oil samples were differentiated from other oils using Fourier-transform (FT)-IR and FT-Raman spectroscopy with multivariate statistical methods [9, 10]. Tapp *et al.* [11] were able to discriminate extra-virgin olive oils from different producing countries with FT-IR in combination with multivariate analysis. 93.9% correct classification of olive oil samples with respect to geographical origin was achieved using near-IR (NIR) spectroscopy in combination with factorial discriminant analysis [12].

FT-IR spectra, which contain information about the complete chemical composition of the analyzed material, have the advantage of rapid analysis time. So, there exist studies involving the use of FT-IR with chemometrics for the determination of analytical parameters such as peroxide value and free fatty acidity value [13, 14]. Another analytically determined parameter, fatty acids are one of the major ingredients that olive oil contains in its chemical structure. Classification studies could also provide the opportunity to check whether there is a correlation between the IR spectra and the fatty acid profile of an olive oil obtained from GC analysis using multivariate statistical analysis. In this respect, Galtier and Dupuy [15] verified the availability of NIR spectroscopy with chemometrics instead of chromatographic determinations such as fatty acids and triacylglycerols for the determination of origin.

In this study, chemometric techniques were applied to the classification of extra-virgin olive oil samples according to geographical origin and harvest year, using two different data sets: (1) fatty acid profiles obtained from GC analysis and (2) spectral data obtained from FT-IR. The discrimination ability of these two methods was compared and discussed. Furthermore, the relation between FT-IR spectra and fatty acid profiles to predict the fatty acid composition of olive oils was studied with multivariate statistical analysis.

## 2 Materials and methods

### 2.1 Samples

Extra-virgin olive oil samples belonging to 30 different locations of the Aegean region of Turkey were obtained from the same olive oil producer for two consecutive harvest years (2005/06 and 2006/07). There were 25 and 22 olive oil samples belonging to the 2005/06 and 2006/07 harvest years, respectively. The olive oil samples were coded either as

north (N) or south (S) according to the region where the olives were obtained from (Tables 1 and 2). Also, numbers beside N or S represent the first or second harvest year. While mainly the Ayvalik variety is cultivated in the North Aegean, Memecik is the dominant variety in the southern part of the Aegean region. As soon as the olive oil samples were received from the processing plant, they were transferred into dark glass bottles and kept at 8 °C in a refrigerator. All analyses were completed within 2 months after receipt of the samples in each harvest year.

### 2.2 Gas chromatographic analysis

Fatty acid methyl esters (FAME) containing C<sub>8</sub>–C<sub>24</sub> (2–11% relative concentration) were used as reference standard (Supelco No. 18918). The olive oil samples were dissolved in hexane and saponified to their methyl esters with the addition of methanolic KOH solution as described in the official journal of the EU [16]. Gas chromatographic analyses were performed on an Agilent 6890 gas chromatograph with a flame ionization detector (FID) and a split/splitless (1 : 50) injector, equipped with a HP 88 capillary column (100 m × 0.25 mm ID × 0.2 μm). The carrier gas was helium with 2 mL/min constant flow. Injection and detector temperatures were 250 and 280 °C, respectively. The oven temperature program was run at 120 °C for 1 min, varied at 3 °C/min to 220 °C and held at this temperature for 5 min. Peak quantification is expressed as percentage of FAME using FAME standards and sample chromatograms. Measurements were conducted in duplicate or triplicate for each olive oil sample.

Fatty acids used in the analysis were myristic acid (14:0), palmitic acid (16:0), palmitoleic acid (16:1), stearic acid (18:0), oleic acid (18:1), linoleic acid (18:2), arachidic acid (20:0), linolenic acid (18:3) and behenic acid (22:0).

### 2.3 FT-mid-IR analysis

All IR spectra were acquired in the 4000–650 cm<sup>-1</sup> range with a Perkin Elmer Spectrum 100 FT-IR spectrometer (Perkin Elmer, Wellesley, MA, USA). This instrument was equipped with a horizontal attenuated total reflectance (HATR) sampling accessory (ZnSe crystal) and a deuterated triglycine sulfate (DTGS) detector. The HATR accessory was used to collect the spectral data of the oils. The resolution was set at 2 cm<sup>-1</sup> and the number of scans collected for each spectrum was 128. The ZnSe crystal was cleaned with hexane in between sample runs. Measurements were conducted in duplicate or triplicate for each olive oil sample and harvest year.

### 2.4 Statistical analysis

Data analysis was performed using multivariate statistical methods with SIMCA software (Umetrics, Sweden). The 3620–2520 and 1875.5–675 cm<sup>-1</sup> regions of the FT-IR spectra were used in the statistical analysis. The discrimination of

**Table 1.** Fatty acid profiles (% of total fatty acids) of commercial olive oil samples belonging to the harvest year 2005/06.

	14:0	16:0	16:1	18:0	18:1	18:2	18:3	20:0	22:0
<i>North (N1)</i>	14:0 <sup>a</sup>	16:0 <sup>b</sup>	16:1 <sup>a</sup>	18:0 <sup>a</sup>	18:1 <sup>a</sup>	18:2 <sup>b</sup>	18:3 <sup>a</sup>	20:0 <sup>a</sup>	22:0 <sup>b</sup>
Ezine	0.02 ± 0.00	13.26 ± 0.01	0.70 ± 0.00	2.77 ± 0.01	71.99 ± 0.09	10.09 ± 0.05	0.58 ± 0.02	0.46 ± 0.03	0.14 ± 0.02
Ezine Gulpinar	0.02 ± 0.01	12.75 ± 0.03	0.57 ± 0.00	2.87 ± 0.03	70.97 ± 0.08	11.57 ± 0.01	0.67 ± 0.02	0.47 ± 0.00	0.11 ± 0.02
Kucukkuyu1	0.02 ± 0.00	13.28 ± 0.03	0.68 ± 0.01	2.81 ± 0.01	72.18 ± 0.10	9.84 ± 0.05	0.60 ± 0.02	0.43 ± 0.03	0.15 ± 0.02
Kucukkuyu2	0.02 ± 0.00	12.79 ± 0.06	0.59 ± 0.01	3.08 ± 0.03	71.38 ± 0.07	10.82 ± 0.04	0.68 ± 0.02	0.46 ± 0.04	0.17 ± 0.05
Altinoluk1	0.02 ± 0.00	13.35 ± 0.02	0.69 ± 0.01	2.78 ± 0.01	72.11 ± 0.02	9.89 ± 0.03	0.62 ± 0.02	0.44 ± 0.05	0.10 ± 0.02
Altinoluk2	0.02 ± 0.01	12.40 ± 0.03	0.60 ± 0.01	3.00 ± 0.03	71.82 ± 0.05	10.87 ± 0.02	0.63 ± 0.01	0.48 ± 0.02	0.17 ± 0.03
Edremit	0.02 ± 0.00	13.32 ± 0.02	0.68 ± 0.00	2.80 ± 0.01	72.04 ± 0.09	9.92 ± 0.04	0.61 ± 0.01	0.45 ± 0.00	0.17 ± 0.04
Havran	0.02 ± 0.00	12.92 ± 0.01	0.59 ± 0.00	3.19 ± 0.01	71.45 ± 0.06	10.50 ± 0.06	0.69 ± 0.00	0.50 ± 0.03	0.14 ± 0.01
Burhaniye	0.02 ± 0.00	13.14 ± 0.05	0.62 ± 0.01	2.93 ± 0.02	71.87 ± 0.02	10.11 ± 0.02	0.65 ± 0.02	0.49 ± 0.03	0.17 ± 0.05
Gomec	0.02 ± 0.01	13.16 ± 0.02	0.63 ± 0.01	2.95 ± 0.02	71.86 ± 0.14	10.12 ± 0.07	0.65 ± 0.02	0.46 ± 0.02	0.14 ± 0.01
Ayvalik	0.02 ± 0.00	13.28 ± 0.09	0.64 ± 0.01	2.93 ± 0.02	71.69 ± 0.70	10.18 ± 0.05	0.67 ± 0.01	0.46 ± 0.00	0.13 ± 0.01
Altinova	0.02 ± 0.00	13.15 ± 0.03	0.66 ± 0.01	2.91 ± 0.02	71.58 ± 0.13	10.48 ± 0.02	0.60 ± 0.01	0.45 ± 0.03	0.14 ± 0.03
Zeytindag	0.04 ± 0.03	12.72 ± 0.10	0.61 ± 0.01	3.07 ± 0.03	71.33 ± 0.15	10.86 ± 0.04	0.68 ± 0.02	0.51 ± 0.01	0.18 ± 0.08
<i>South (S1)</i>	14:0 <sup>a</sup>	14:0 <sup>a</sup>	16:1 <sup>b</sup>	18:0 <sup>b</sup>	18:1 <sup>b</sup>	18:2 <sup>a</sup>	18:3 <sup>b</sup>	20:0 <sup>b</sup>	22:0 <sup>a</sup>
Akhisar	0.02 ± 0.00	13.19 ± 0.22	0.68 ± 0.04	3.28 ± 0.52	71.81 ± 0.20	9.75 ± 0.55	0.68 ± 0.03	0.45 ± 0.00	0.13 ± 0.01
Menemen	0.02 ± 0.00	12.80 ± 0.04	0.75 ± 0.00	3.65 ± 0.02	69.82 ± 0.10	11.69 ± 0.04	0.71 ± 0.01	0.45 ± 0.04	0.11 ± 0.02
Tepecik	0.02 ± 0.00	12.68 ± 0.04	0.83 ± 0.00	3.25 ± 0.01	71.60 ± 0.21	10.40 ± 0.05	0.65 ± 0.32	0.46 ± 0.03	0.11 ± 0.01
Bayindir	0.02 ± 0.00	12.2 ± 0.03	0.65 ± 0.01	2.92 ± 0.01	74.39 ± 0.01	8.42 ± 0.05	0.78 ± 0.00	0.44 ± 0.02	0.12 ± 0.01
Selcuk	0.02 ± 0.00	12.48 ± 0.01	0.73 ± 0.00	3.33 ± 0.02	70.36 ± 0.04	11.76 ± 0.02	0.78 ± 0.00	0.41 ± 0.02	0.14 ± 0.05
Aydin	0.02 ± 0.00	12.63 ± 0.13	0.70 ± 0.02	3.06 ± 0.04	72.60 ± 0.27	9.63 ± 0.04	0.76 ± 0.01	0.47 ± 0.02	0.14 ± 0.03
Ortaklar	0.02 ± 0.00	12.31 ± 0.07	0.76 ± 0.01	3.28 ± 0.01	73.03 ± 0.15	9.34 ± 0.07	0.72 ± 0.02	0.42 ± 0.01	0.12 ± 0.06
Kocarli	0.02 ± 0.00	12.26 ± 0.30	0.69 ± 0.01	3.39 ± 0.18	73.51 ± 0.38	8.87 ± 0.12	0.67 ± 0.01	0.46 ± 0.03	0.12 ± 0.04
Milas	0.02 ± 0.00	12.06 ± 0.05	0.69 ± 0.01	2.70 ± 0.02	72.99 ± 0.11	10.13 ± 0.02	0.83 ± 0.01	0.44 ± 0.01	0.13 ± 0.02

<sup>a,b</sup> Different letters indicate significant differences between north and south olive oils for individual fatty acids.

olive oil samples was achieved with PCA, which is a multivariate projection method designed to extract and display the systematic variation in a data matrix *X*. Both fatty acid profile and spectral data are manipulated with PCA to inspect principal groupings. However, IR spectra are subjected to wavelet analysis to compress and de-noise complicated signals. Daubechies-10 was chosen as wavelet function.

It is important to accurately determine the number of components that should be included in the model since it is linked to the difference between the degree of fit and the predictive ability. The degree of fit increases as the number of components increases but the predictive ability does not necessarily increase after a certain model complexity. So, it is important to reach an optimal balance between fit and predictive ability [17].

Results of PCA are visualized by scores and Coomans' plots. Scores plots were constructed to observe principal groupings among observations. PCA analysis is also used to develop a modeling technique called soft independent modeling of class analogy (SIMCA), which is one of the most commonly used class-modeling tools in chemometrics. Coomans' plot is a useful tool to interpret the results of SIMCA, which shows discrimination of two classes. In SIMCA, PCA is performed for each class separately, and this results in a principal component (PC) model for each class. The class distance can

be calculated as the geometric distance from the PC models. Each group may be bounded by a region of space representing 95% confidence that a particular object belongs to a class. The residual variance of a variable of a class is used for estimating the modeling power of a particular variable [18].

Scores and Coomans' plots were employed for discrimination of samples according to geographical origin and harvest year. Classification studies were performed by setting nine GC variables and FT-IR spectra as independent variable data set.

Estimation of fatty acid profile was performed using partial least squares (PLS) regression analysis, which relates FT-IR data to GC variables. The spectral regions used for PCA were also used for the PLS models. Prior to PLS, orthogonal signal correction (OSC) in combination with wavelet analysis was applied to spectral data. OSC is a signal correction technique which constructs a filter that removes the part from the spectral matrix *X* that is definitely unrelated to *Y* [17]. For PLS analysis, 35 oil samples belonging to the first and second harvest years were used as calibration set, and 12 samples as validation set to test the predictive ability of the model. The error criteria chosen for this study were the standard error of calibration (SEC) and standard error of prediction (SEP), both expressed as percentage, which have the advantage of being dimensionless [19, 20].

$$SEC = \frac{100}{\bar{Y}_i} \sqrt{\frac{\sum_{i=1}^m (\hat{Y}_i - Y_i)^2}{M - 1}}$$

$$SEP = \frac{100}{\bar{Y}_i} \sqrt{\frac{\sum_{i=1}^n (\hat{Y}_i - Y_i)^2}{N}}$$

where  $Y_i$  is the observed and  $\hat{Y}_i$  is the predicted percentage of fatty acids;  $\bar{Y}_i$  is its average value;  $M$  and  $N$  are the numbers of samples in the calibration and prediction sets, respectively. The lower the error values for the selected data set, the better is the performance of the model. As expected, the model tends to have smaller  $R^2$  and larger error values during prediction.

### 3 Results and discussions

The olive oil samples used in this study were obtained from the Aegean coastal region of Turkey which is located in the western part of Turkey (between approximately 36° and 40° North). This region is the leading olive oil producer in the

country. The region is divided into two parts as north (38–40°N) and south (38–36°N) of Izmir. The dominant olive variety grown in the northern part is the Ayvalik variety, and the Gemlik variety is also cultivated in this part. The Memecik variety is widely grown in the south, and Memeli, Uslu and Gemlik are the other main varieties of the south. The fatty acid profiles of the commercial olive oil samples belonging to the 2005/06 and 2006/07 harvest years are provided in Tables 1 and 2, respectively. The fatty acid content of each sample was the average of at least two GC measurements. For the 2005/06 harvest year, the content of oleic acid was the highest for Kucukkuyu1 (72.18%) and lowest for Menemen (69.82%), while for the 2006/07 harvest year it was at its highest amount for Tire (75.29%) and its lowest amount for Kucukkuyu (70.20%). Inspection of Tables 1 and 2 revealed that oleic acid was at higher amounts for the oils obtained from the south than from the north for both the 2005/06 and 2006/07 harvest years. As the most abundant saturated fatty acid, palmitic acid was higher for the north than for the south in both years. Also, the content of linoleic acid was higher for the 2005/06 harvest year.

PCA projects a multivariate data table to a low-dimensional plane, usually consisting of two to five dimensions, such

**Table 2.** Fatty acid profiles (% of total fatty acids) of commercial olive oil samples belonging to the harvest year 2006/07.

	14:0	16:0	16:1	18:0	18:1	18:2	18:3	20:0	22:0
<i>North (N2)</i>	14:0 <sup>a</sup>	16:0 <sup>b</sup>	16:1 <sup>a</sup>	18:0 <sup>a</sup>	18:1 <sup>a</sup>	18:2 <sup>b</sup>	18:3 <sup>a</sup>	20:0 <sup>b</sup>	22:0 <sup>b</sup>
Ezine	0.01 ± 0.00	14.08 ± 0.26	1.36 ± 0.00	2.57 ± 0.06	70.33 ± 0.22	10.72 ± 0.10	0.34 ± 0.01	0.46 ± 0.01	0.14 ± 0.01
Kucukkuyu	0.02 ± 0.00	12.99 ± 0.01	1.08 ± 0.00	2.77 ± 0.01	70.20 ± 0.06	11.93 ± 0.05	0.37 ± 0.00	0.48 ± 0.01	0.16 ± 0.03
Altinoluk	0.02 ± 0.00	13.48 ± 0.12	1.17 ± 0.01	2.72 ± 0.02	71.03 ± 0.16	10.47 ± 0.03	0.43 ± 0.00	0.51 ± 0.01	0.16 ± 0.01
Edremit	0.02 ± 0.00	13.09 ± 0.25	1.10 ± 0.01	2.76 ± 0.09	71.65 ± 0.29	10.34 ± 0.06	0.39 ± 0.00	0.50 ± 0.00	0.15 ± 0.01
Havran	0.02 ± 0.00	12.62 ± 0.05	0.99 ± 0.01	2.87 ± 0.03	71.74 ± 0.05	10.71 ± 0.07	0.39 ± 0.00	0.50 ± 0.01	0.15 ± 0.00
Burhaniye	0.02 ± 0.00	12.92 ± 0.06	1.03 ± 0.00	2.85 ± 0.02	70.92 ± 0.11	11.22 ± 0.05	0.39 ± 0.00	0.50 ± 0.01	0.15 ± 0.01
Gomec	0.02 ± 0.00	13.18 ± 0.22	1.04 ± 0.01	2.92 ± 0.09	70.51 ± 0.26	11.26 ± 0.07	0.39 ± 0.00	0.51 ± 0.02	0.16 ± 0.02
Ayvalik	0.02 ± 0.00	13.49 ± 0.04	1.17 ± 0.02	2.74 ± 0.03	70.73 ± 0.12	10.84 ± 0.01	0.39 ± 0.00	0.49 ± 0.00	0.14 ± 0.01
Altinova	0.02 ± 0.00	13.16 ± 0.26	1.17 ± 0.02	2.78 ± 0.08	71.71 ± 0.52	10.10 ± 0.32	0.42 ± 0.01	0.49 ± 0.01	0.15 ± 0.02
Zeytindag	0.02 ± 0.00	12.61 ± 0.01	1.06 ± 0.02	2.69 ± 0.01	71.99 ± 0.02	10.62 ± 0.01	0.38 ± 0.00	0.48 ± 0.01	0.15 ± 0.00
<i>South (S2)</i>	14:0 <sup>a</sup>	16:0 <sup>a</sup>	16:1 <sup>a</sup>	18:0 <sup>a</sup>	18:1 <sup>b</sup>	18:2 <sup>a</sup>	18:3 <sup>b</sup>	20:0 <sup>a</sup>	22:0 <sup>a</sup>
Tepecik	0.02 ± 0.00	12.76 ± 0.50	1.19 ± 0.01	2.70 ± 0.12	74.30 ± 0.46	7.89 ± 0.15	0.48 ± 0.01	0.50 ± 0.00	0.16 ± 0.00
Bayindir	0.02 ± 0.00	11.49 ± 0.04	1.04 ± 0.00	2.46 ± 0.02	75.17 ± 0.03	8.79 ± 0.04	0.47 ± 0.00	0.44 ± 0.00	0.12 ± 0.00
Odemis	0.02 ± 0.00	11.50 ± 0.08	1.13 ± 0.01	2.75 ± 0.04	74.13 ± 0.12	9.41 ± 0.04	0.48 ± 0.00	0.46 ± 0.01	0.13 ± 0.01
Tire	0.02 ± 0.00	12.03 ± 0.07	1.09 ± 0.01	2.35 ± 0.02	75.29 ± 0.10	8.16 ± 0.04	0.49 ± 0.01	0.44 ± 0.01	0.13 ± 0.01
Selcuk	0.02 ± 0.00	12.36 ± 0.23	1.14 ± 0.03	2.76 ± 0.12	74.21 ± 0.60	8.46 ± 0.69	0.45 ± 0.02	0.47 ± 0.01	0.15 ± 0.03
Kusadasi	0.02 ± 0.00	11.59 ± 0.03	0.90 ± 0.00	2.81 ± 0.01	74.61 ± 0.05	9.07 ± 0.04	0.43 ± 0.00	0.46 ± 0.00	0.12 ± 0.01
Germencik	0.02 ± 0.00	12.77 ± 0.13	1.26 ± 0.01	2.72 ± 0.05	72.86 ± 0.20	9.25 ± 0.06	0.51 ± 0.00	0.48 ± 0.01	0.13 ± 0.01
Aydin	0.02 ± 0.00	12.73 ± 0.81	1.11 ± 0.03	2.99 ± 0.23	73.24 ± 0.89	8.78 ± 0.18	0.47 ± 0.02	0.50 ± 0.03	0.15 ± 0.03
Ortaklar	0.02 ± 0.00	12.62 ± 0.40	1.27 ± 0.15	2.67 ± 0.21	73.76 ± 0.29	8.56 ± 0.61	0.49 ± 0.01	0.47 ± 0.01	0.13 ± 0.01
Koski	0.02 ± 0.00	11.51 ± 0.03	1.12 ± 0.01	2.74 ± 0.02	74.14 ± 0.06	9.40 ± 0.03	0.48 ± 0.02	0.46 ± 0.00	0.12 ± 0.01
Dalaman	0.02 ± 0.00	11.50 ± 0.02	1.06 ± 0.01	2.83 ± 0.01	74.63 ± 0.06	8.89 ± 0.05	0.48 ± 0.00	0.47 ± 0.01	0.13 ± 0.00
Kocarli	0.02 ± 0.00	12.10 ± 0.03	1.09 ± 0.01	2.76 ± 0.00	74.18 ± 0.05	8.80 ± 0.05	0.46 ± 0.01	0.47 ± 0.02	0.12 ± 0.01
Erbeyli	0.02 ± 0.00	11.52 ± 0.07	1.07 ± 0.03	2.75 ± 0.02	74.36 ± 0.1	9.22 ± 0.06	0.48 ± 0.00	0.45 ± 0.02	0.13 ± 0.01
Cine	0.02 ± 0.00	12.08 ± 0.57	1.09 ± 0.03	2.73 ± 0.13	74.28 ± 0.50	8.75 ± 0.02	0.46 ± 0.01	0.47 ± 0.02	0.12 ± 0.00
Milas	0.02 ± 0.00	12.81 ± 0.82	1.15 ± 0.01	2.69 ± 0.17	72.19 ± 0.81	10.06 ± 0.19	0.48 ± 0.02	0.47 ± 0.03	0.12 ± 0.02

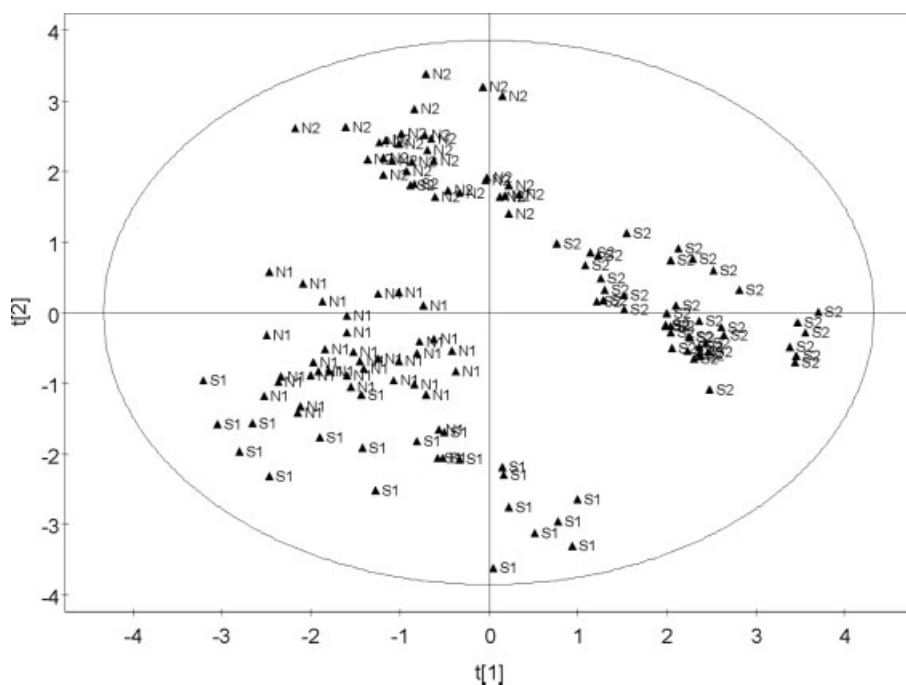
<sup>a,b</sup> Different letters indicate significant differences between north and south olive oils for individual fatty acids.

that an overview of the data is obtained. This new representation of data may reveal groups of observations, trends, and outliers which can be observed from a scores plot. The fatty acid profiles of olive oils coming from various regions including nine GC variables were submitted to PCA to visualize the presence of principal groupings. Figure 1 represents the scores plot of olive oil samples belonging to the north and south regions obtained for two consecutive harvest years (2005/06 and 2006/07). The discrimination of the olive oil samples from the second year with respect to geographical origin is quite clear. However, some southern oil samples belonging to the first harvest year cannot be differentiated from the northern oils in the scores plot. On the other hand, the harvest year seems to have a quite good ability for differentiation. Variation of chemical composition in two consecutive years can be due to changes in climatic conditions (rainfall, temperature, humidity) [21]. D'Imperio *et al.* [6] reported that the harvest year significantly influenced the olive oil fatty acid composition especially when extremely bad climatic conditions occur. Differences in chemical composition between oils obtained in different years were also observed for Calabrian oil cultivars [5].

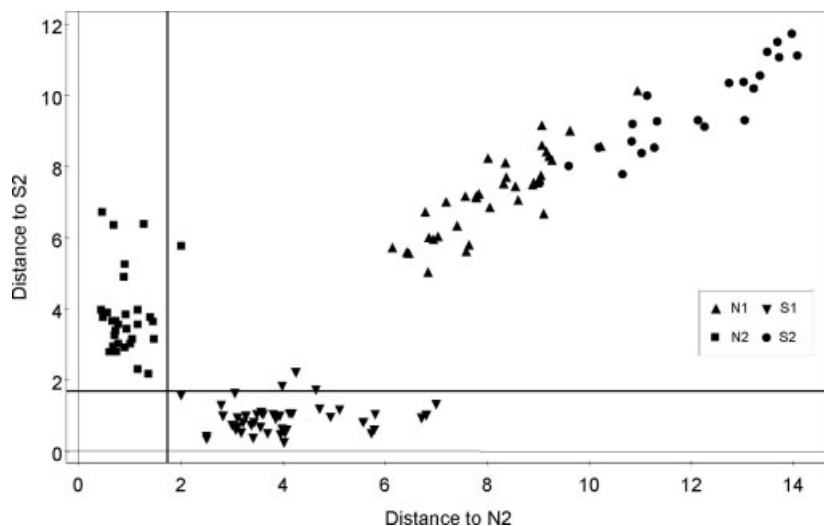
Coomans' plot is a good way for visualizing the SIMCA approach in the two-class case. It is constructed based on class-PCA models calculated independently. Then, two-class models are plotted against each other with the critical levels as straight lines displaying the boundaries. Therefore, a Coomans' plot was constructed for the classification of samples using the fatty acid profiles (Fig. 2). Four PCA classes (N1, S1, N2 and S2) were calculated independently, with N2 and

S2 as the class models placed against each other. Only few samples belonging to N2 and S2 were drawn on the outside of their critical limits. Samples of N1 and S1 were identified as two different classes correctly placed in the outer space of the N2 and S2 models. According to this plot, the analysis of fatty acids with PCA has the ability to discriminate oil samples with respect to geographical origin and harvest year. According to the loading plot (not shown), oleic and linoleic acids could be attributed to geographical classification while palmitoleic and linolenic acids cause discrimination with regard to harvest year.

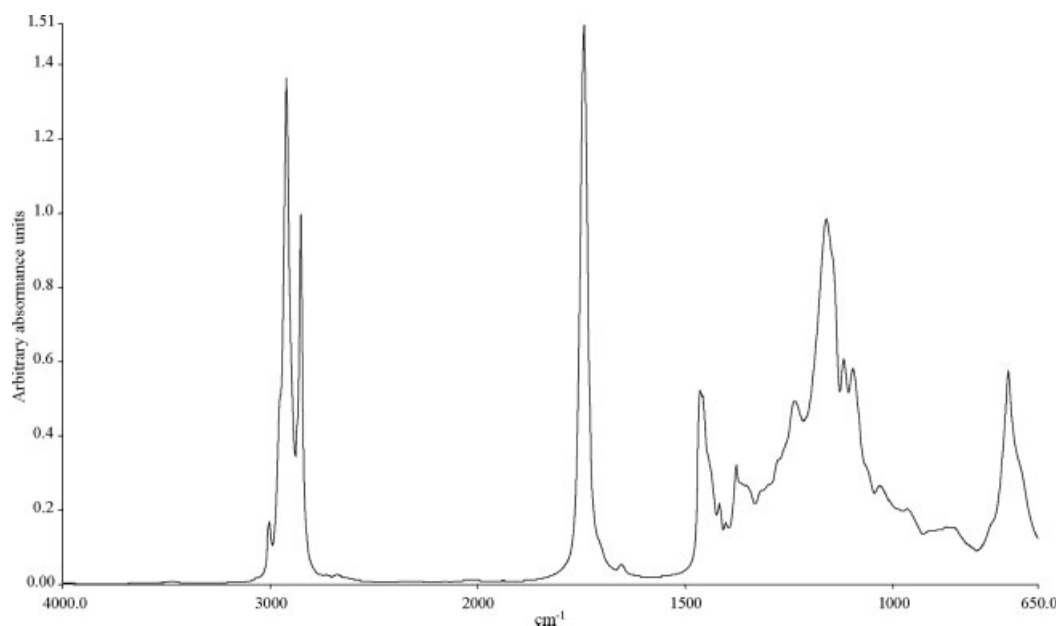
Spectrophotometric methods such as FT-IR have become an emerging tool with the advantage of short analysis time and easy sample preparation for the authentication and classification of olive oils. Figure 3 illustrates typical olive oil spectra obtained in this study. In mid-IR (MIR) spectra, the peaks around the 2950–2800  $\text{cm}^{-1}$  region are due to C-H stretching vibrations of  $-\text{CH}_3$  and  $-\text{CH}_2$  groups. The large peak around 1745  $\text{cm}^{-1}$  results from C=O double bond stretching vibration of carbonyl groups. Peaks of the 1470–1200  $\text{cm}^{-1}$  region corresponds to CH bending of  $-\text{CH}_3$  and  $-\text{CH}_2$ . The fingerprint region lay between 1250 and 700  $\text{cm}^{-1}$ , which is due to stretching vibration of C-O ester groups and  $\text{CH}_2$  rocking vibration [22]. The entire range of spectra appears very similar for each olive oil sample unless one looks very closely. Multivariate data analysis is therefore required to extract the relevant information from these spectra. The FT-IR spectral data between the 3620–2520 and 1875.5–675  $\text{cm}^{-1}$  regions were employed to construct a Coomans' plot (Fig. 4). Analysis of this graph revealed that discrimination of N2 and S2 was



**Figure 1.** PCA scores plot of the two first PC of the fatty acid profile data (N1: olive oils from the north in 2005/06, S1: olive oils from the south in 2005/06, N2: olive oils from the north in 2006/07, S2: olive oils from the south in 2006/07).



**Figure 2.** Coomans' plot for the classification of north and south olive oil samples using fatty acid profiles belonging to two harvest years (the vertical and horizontal lines show the 95% confidence interval).

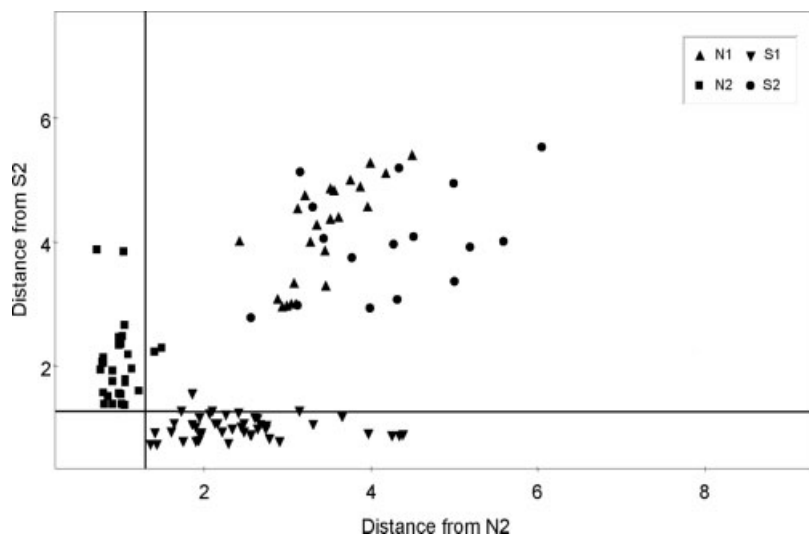


**Figure 3.** Typical olive oil FT-IR spectra.

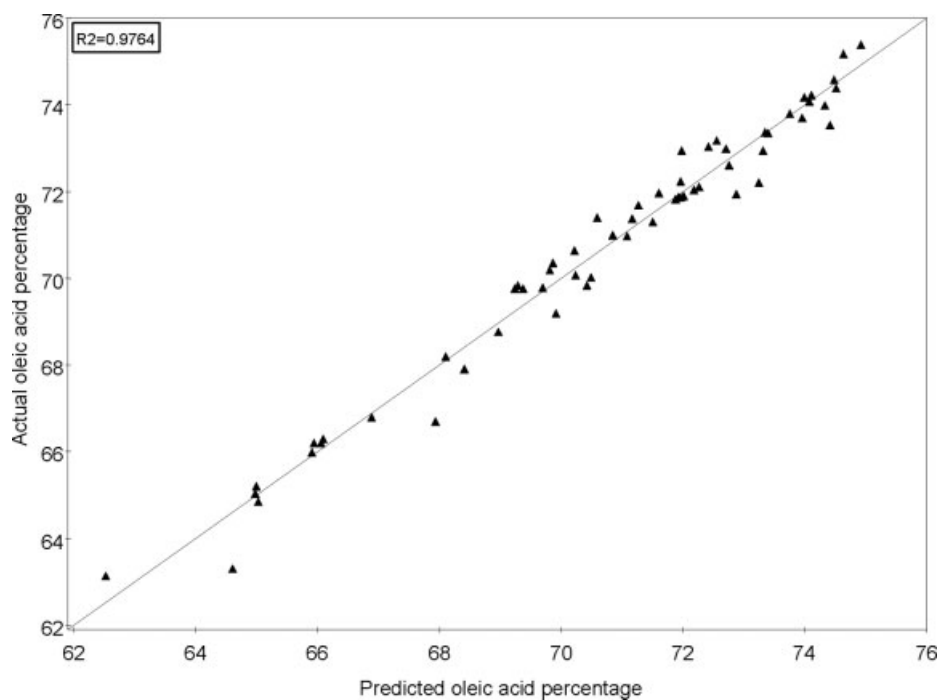
achieved, except for some samples. The samples of N1 and S1 are correctly plotted beyond the critical limits of the N2 and S2 models, but they cannot be differentiated from each other very well. Although the ability of FT-IR for the differentiation of olive oil samples with respect to geographical origin is low, different harvest years can be sufficiently identified. When compared with spectral data, the fatty acid profiles obtained from GC supplied more discrete chemical information than other chromatographic methods for the classification of olive oils with respect to geographical origin. Actually, FT-IR spectra supply much more information that reflects differences among not only fatty acid profiles but also triacylglycerol structures and minor components. However, that much

information results in more variation, which leads to difficulties in the classification studies.

Each olive oil sample has its own fatty acid profile and spectral properties. It is stated that the spectral features of oils vary with the degree of unsaturation [22]. Therefore, PLS analysis was applied to data to check whether there is a correlation between FT-IR spectra and the fatty acid profile of olive oils. Using the PLS algorithm, the fatty acid profile of each olive oil sample is predicted, relating spectral data ( $X$  variable) with fatty acid data ( $Y$  variable). As the major fatty acid of olive oil ranging between 69–75%, the predicted and actual 18:1 percentages of each sample are illustrated in Fig. 5. The correlation coefficient,  $R^2$ , is 0.97, which indicates good pre-



**Figure 4.** Coomans' plot for the classification of north and south olive oil samples using FT-IR spectra belonging to two harvest years (the vertical and horizontal lines show the 95% confidence interval).



**Figure 5.** PLS regression of predicted vs. actual oleic acid composition of the olive oil samples.

diction of the oleic acid percentages from the spectral data. A validation step is applied for each fatty acid to test the predictive ability of the regression model. The results of the PLS model are summarized in Table 3. The error value of the prediction set is lower than 1 for each fatty acid, which demonstrates the success of the model. MUFA are important because of their nutritional implication [23], and 18:1 is the characteristic MUFA of olive oil. The  $R^2$  value of 18:1 and MUFA is relatively high, indicating the predictability of them using FT-IR spectra with PLS. Also, good results were

obtained for other major fatty acids such as 16:1, 18:1 and 18:2. A quite low  $R^2$  value is observed for 14:0, which is present in minor amounts. In overall consideration, MUFA constituting the highest proportion can be predicted best from the MIR data, followed by PUFA and SFA. Galtier and Dupuy [15] reported similar results using NIR data for fatty acid profile determination of virgin olive oils.

In conclusion, the fatty acid profile is more effective in the discrimination of virgin olive oil samples with respect to geographical origin than spectral data. The composition of virgin

**Table 3.** Summary of the proposed PLS regression model for both calibration and validation sets.

Fatty acid	$R^2$ (cal.) <sup>§</sup>	SEC	$R^2$ (val.) <sup>§</sup>	SEP <sup>d</sup>
14:0	0.25	0.0033	0.22	0.0049
16:0	0.85	0.45	0.71	0.66
16:1	0.93	0.13	0.81	0.23
18:0	0.83	0.24	0.55	0.41
18:1	0.97	0.50	0.93	0.97
18:2	0.97	0.40	0.93	0.66
18:3	0.69	0.07	0.64	0.07
20:0	0.83	0.06	0.56	0.07
22:0	0.90	0.34	0.68	0.6
SFA	0.97	0.52	0.79	0.60
MUFA	0.97	0.52	0.94	0.86
PUFA	0.98	0.38	0.94	0.63

<sup>§</sup> PLS regression correlation coefficient for calibration samples.

<sup>§</sup> PLS regression correlation coefficient for validation samples.

SEC, Standard error of calibration; SEP, standard error of prediction;

SFA, saturated fatty acids;

MUFA, monounsaturated fatty acids;

PUFA, polyunsaturated fatty acids.

olives shows significant changes in between two consecutive harvest years, which can be observed from both GC and FT-IR analysis. Moreover, multivariate analysis of FT-IR data has the potential to predict some of the fatty acids of the virgin olive oil samples. Therefore, with the advantage of simple and rapid sample analysis, FT-IR is applicable for the quick determination of characteristic fatty acids of olive oils.

## Acknowledgments

The authors wish to thank to the Union of Taris Olive and Olive Oil Co-Operatives for providing olive oil samples. Fatty acid analyses were carried out in the IYTE R&D Center for Environmental Studies. This study was performed as part of the CODA (MIRG-CT-2005-029134) project supported by an EU Marie Curie Reintegration Grant.

## Conflict of interest statement

*The authors have declared no conflict of interest.*

## References

- [1] L. C. Matos, S. C. Cunha, J. S. Amaral, A. P. José, P. B. Andrade, R. M. Seabra, B. P. P. Oliveira: Chemometric characterization of three varietal olive oils (Cvs. Cobrançosa, Madural and Verdeal Transmontana) extracted from olives with different maturation indices. *Food Chem.* 2007, **102**, 406–414.
- [2] F. Ulberth, M. Buchgraber: Authenticity of fats and oils. *Eur J Lipid Sci Technol.* 2000, **102**, 687–694.
- [3] B. A. Babcock, R. Clemens: Geographical indications and property rights: Protecting value-added agricultural products. MATRIC Briefing Paper 2004, 04-MBP 7, Iowa State University (USA).
- [4] L. Giansante, D. Di Vincenzo, G. Bianchi: Classification of monovarietal Italian olive oils by unsupervised (PCA) and supervised (LDA) chemometrics. *J Agric Food Chem.* 2003, **83**, 905–911.
- [5] S. Lanteri, C. Armanino, E. Perri, A. Palopoli: Study of oils from Calabrian olive cultivars by chemometric methods. *Food Chem.* 2002, **76**, 501–507.
- [6] M. D'Imperio, G. Dugo, M. Alfa, L. Mannina, A. L. Segre: Statistical analysis on Sicilian oils. *Food Chem.* 2007, **102**, 956–965.
- [7] G. Downey: Food and food ingredient authentication by mid-infrared spectroscopy and chemometrics. *Trends Anal Chem.* 1998, **17**, 418–423.
- [8] F. R. van de Voort, J. Sedman, T. Russin: Lipid analysis by vibrational spectroscopy. *Eur J Lipid Sci Technol.* 2001, **103**, 815–840.
- [9] A. Tay, R. K. Singh, S. S. Krishnan, J. P. Gore: Authentication of olive oil adulterated with vegetable oils using Fourier transform infrared spectroscopy. *Lebensm Wiss Technol.* 2002, **35**, 99–103.
- [10] V. Baeten, J. A. F. Pierna, P. Dardenne, M. Meurens, D. L. García-González, R. Aparicio-Ruiz: 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.
- [11] H. S. Tapp, M. Defernez, E. K. Kemsley: FTIR spectroscopy and multivariate analysis can distinguish the geographic origin of extra virgin olive oils. *J Agric Food Chem.* 2003, **51**, 6110–6115.
- [12] G. Downey, P. McIntyre, A. N. Davies: Geographic classification of extra virgin olive oils from the eastern Mediterranean by chemometric analysis of visible and near-infrared spectroscopic data. *Appl Spectr.* 2003, **57**, 158–163.
- [13] S. Armenta, S. Garrigues, S. M. de la Guardia: Determination of edible oil parameters by near infrared spectrometry. *Anal Chim Acta.* 2007, **596**, 330–337.
- [14] F. A. Iñón, J. M. Garrigues, S. Garrigues, A. Molina, M. de la Guardia: Selection of calibration set samples in determination of olive oil acidity by partial least squares–attenuated total reflectance–Fourier transform infrared spectroscopy. *Anal Chim Acta.* 2003, **489**, 59–75.
- [15] O. Galtier, N. Dupuy, Y. le Dréau, D. Ollivier, C. Pinatel, J. Kister, J. Artaud: Geographic origins and compositions of virgin olive oils determined by chemometric analysis of NIR spectra. *Anal Chim Acta.* 2007, **595**, 136–144.
- [16] European Union Commission: Regulation EEC 2568/91 on the characteristics of olive oil and olive-residue oil and on the relevant methods of analysis. *Off J Eur Communities.* 1991, **L248**.



- [17] L. Eriksson, E. Johansson, N. Kettaneh-Wold, S. Wold: *Multivariate Data Analysis. Principals and Applications*. Umetrics, Umea (Sweden) 2001.
- [18] R. G. Brereton: *Chemometrics. Data Analysis for the Laboratory and Chemical Plant*. John Wiley & Sons, Chichester (UK) 2003.
- [19] A. O. Aptula, N. G. Jeliaskova, T. W. Schultz, M. T. D. Cronin: The better predictive model: High  $q^2$  for the training set or low root mean square error of prediction for the test set. *QSAR Comb Sci.* 2005, 24.
- [20] C. Hervás, G. Zurera, J. A. Martínez: Optimization of computational neural network for its application in the prediction of microbial growth in foods. *Food Sci Technol Int.* 2001, 7, 159–163.
- [21] R. Aparicio, G. Luna: Characterization of monovarietal virgin olive oils. *Eur J Lipid Sci Technol.* 2002, 104, 614–627.
- [22] J. Harwood, R. Aparicio: *Handbook of Olive Oil. Analysis and Properties*. Aspen, Gaithersburg, MD (USA) 2000.
- [23] M. P. Aguilera, B. Beltrán, D. Ortega, A. Fernández, A. Jiménez, M. Uceda: Characterisation of virgin olive oil of Italian olive cultivars: 'Frantoio' and 'Leccino', grown in Andalusia. *Food Chem.* 2005, 89, 387–391.