DOI:10.1021/jf903989b

JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY
ARTICLE

# Virgin Olive Oil Authentication by Multivariate Analyses of $^1$H NMR Fingerprints and $\delta^{13}$C and $\delta^2$H Data

Rosa M. Alonso-Salces,*,† José M. Moreno-Rojas,†,# Margaret V. Holland,† Fabiano Reniero,† Claude Guillou,† and Károly Héberger†,§

†European Commission − Joint Research Centre, Institute for Health and Consumer Protection, Via Fermi 1, I-21020 Ispra, Italy, and §Chemical Research Center, Hungarian Academy of Sciences, P.O. Box 17, H-1525 Budapest, Hungary. # Present address: Área de Tecnología Postcosecha e Industria Agroalimentaria, Centro "Alameda del Obispo", Instituto de Investigación y Formación Agroalimentaria, Avda. Menéndez Pidal s/n, E-14071 Córdoba, Spain.

$^1$H NMR fingerprints of virgin olive oils (VOOs) from the Mediterranean basin (three harvests) were analyzed by principal component analysis, linear discriminant analysis (LDA), and partial least-squares discriminant analysis (PLS-DA) to determine their geographical origin at the national, regional, or PDO level. Further $\delta^{13}$C and $\delta^2$H measurements were performed by isotope ratio mass spectrometry (IRMS). LDA and PLS-DA achieved consistent results for the characterization of PDO *Riviera Ligure* VOOs. PLS-DA afforded the best model: for the Liguria class, 92% of the oils were correctly classified in the modeling step, and 88% of the oils were properly predicted in the external validation; for the non-Liguria class, 90 and 86% of hits were obtained, respectively. A stable and robust PLS-DA model was obtained to authenticate VOOs from Sicily: the recognition abilities were 98% for Sicilian oils and 89% for non-Sicilian ones, and the prediction abilities were 93 and 86%, respectively. More than 85% of the oils of both categories were properly predicted in the external validation. Greek and non-Greek VOOs were properly classified by PLS-DA: >90% of the samples were correctly predicted in the cross-validation and external validation. Stable isotopes provided complementary geographical information to the $^1$H NMR fingerprints of the VOOs.

**KEYWORDS: Olive oil; NMR; fingerprinting; multivariate data analysis; authentication**

## INTRODUCTION

Olive oil is the oil extracted exclusively from the fruit of *Olea europaea* L. only by means of mechanical methods or other physical procedures that do not cause any alteration of the glyceric structure of the oil and preserve its characteristic properties (*1*). At present, 77% of the global production of olive oil takes place in the Mediterranean basin, mainly in Spain, Italy, and Greece. The characterization of the geographical origin of virgin olive oil (VOO) is becoming increasingly important. VOOs are permitted to be marketed under a Protected Designation of Origin (PDO), Protected Geographical Indication (PGI), or Traditional Specialty Guaranteed (TSG) label, on the basis of their area and methods of production [Council Regulations (EEC) 2081/92 and 2082/92]. The European Commission has already registered in the "Register of protected designations of origin and protected geographical indications" 95 PDO and PGI olive oils, produced in Spain, Italy, Greece, Portugal, France, and Slovenia. As can be expected, given the financial benefits associated with these prestigious labels, it is very likely that economic fraud occurs (e.g., labeling a non-PDO product as a PDO one or adulteration with olive oils that do not fulfill the PDO requirements). Other fraudulent practices that were detected by the state security forces were the adulteration of olive oils with low-grade oils and the mislabeling of olive oils. For instance, olive oil imported into Italy from Tunisia, Greece, and Spain was relabeled as the finest Italian product. Other ploys include labeling inferior quality oil as extra virgin olive oil and claiming European Union (EU) subsidies for growing olives in Italy while actually importing them from elsewhere. The EU is about to establish new labeling rules that will make origin labeling compulsory for virgin and extra virgin labeled olive oil. Therefore, oil produced from olives from just one EU country will have to be labeled with the name of the country of origin. Therefore, analytical methods are urgently needed to guarantee the authenticity and traceability of PDO and PGI olive oils, as well as the country of provenance of the olive oil, to help prevent illicit practices in this sector and to support the antifraud authorities dealing with these issues.

More than 98% of VOO is made up of triglycerides and the remaining 1−2% of minor components such as squalene, α-tocopherol, phytosterols, phenolic compounds, carotenoids, and terpenic alcohols, which constitute the unsaponifiable fraction of the oil (*1*). The chemical composition of this fraction may vary both qualitatively and quantitatively depending on vegetal species, climatic conditions, extraction and refining procedures, and storage conditions, which also greatly influence the

*Author to whom correspondence should be addressed (e-mail rosamaria.alonsosalces@gmail.com; telephone +39-0332789929; fax +39-0332789303).

 Published on Web 04/07/2010 

organoleptic quality and stability of the oil (*1*). The diversity and interdependence among all of these factors makes it highly unlikely that these influences would be the same in different regions. Therefore, the geographical characterization of VOO addresses all of these agronomic, pedoclimatic, and botanical aspects that are unique to the oil of each origin (*2*).

Fingerprinting techniques such as NMR (*3*, *4*), NIR (*5*), MIR (*6*), FT-IR, FT-MIR,and FT-Raman (*7*−*10*) spectroscopies, MS (*11*), GC×GC-Tof-MS (*12*−*14*), and DNA fingerprinting (*15*, *16*) have been used for the determination of food authenticity (*17*). These types of techniques are particularly attractive because they are nonselective, require little or no sample pretreatment, use small amounts of organic solvents or reagents, and typically require only a few minutes per sample. Chemometric analysis of NIR spectra of virgin olive oils allows us to determine its composition and geographical origin (*18*). $^{1}$H, $^{13}$C, and/or $^{31}$P NMR analyses of the bulk oil (*19*, *20*) or the unsaponifiable fraction of olive oil (*21*), in combination with multivariate techniques, have been used to distinguish VOOs according to their geographical origin. $^{1}$H NMR and the more recently developed hyphenated LC-SPE-NMR technique have been applied to study phenolic compounds in the polar fraction of olive oil for authentication purposes (*22*). IRMS methods have also been used for the geographical characterization of olive oil by analyzing the alcohol and sterol fractions (*23*). The study of $\delta^{13}$C variability of olive oils in several harvests showed that it was not dependent on either the degree of ripeness or maturity state of the olives or the olive variety (*24*). The isotopic fractionation of C and H is linked to pedoclimatic factors (soil, climate, and latitude); therefore, these data may contribute to the geographical discrimination of olive oils.

In the present study, the $^{1}$H NMR fingerprints of a statistically significant number of authentic VOOs from seven countries, namely, Italy, Spain, Greece, France, Turkey, Cyprus, and Syria, and from three different harvests (2004/2005, 2005/2006, and 2006/2007) were analyzed by pattern recognition and classification techniques, such as principal component analysis (PCA), linear discriminant analysis (LDA), and partial least-squares discriminant analysis (PLS-DA), to evaluate the best approach to identify the geographical origin at the national, regional, and/or PDO level. Further isotopic measurements of $\delta^{13}$C and $\delta^{2}$H were performed on the samples by isotope ratio mass spectrometry (IRMS) to help with the geographical discrimination of VOOs. This work was developed within the framework of the EU TRACE project (http://www.trace.eu.org) with the aim of supporting antifraud authorities in dealing with the prevention and detection of illicit practices in the olive oil sector. Moreover, this study is also of interest to consumers, honest oil producers, and regulatory bodies because it will contribute to ensure the authenticity and traceability of such a high-value foodstuff.

## MATERIALS AND METHODS

**Chemicals.** Deuterated chloroform for NMR analysis (99.8 atom % D) was provided by Sigma-Aldrich Chemie (Steinheim, Germany).

**Plant Material.** Virgin olive oils (963 samples) from seven countries of the Mediterranean basin, namely, Italy (661 VOOs), Spain (144 VOOs), Greece (97 VOOs), France (39 VOOs), Turkey (14 VOOs), Cyprus (6 VOOs), and Syria (2 VOOs), were collected directly from the producers (olive oil mills) from most of the main producing regions of these countries during three harvests (2004/2005, 2005/2006, and 2006/2007). The sample collection was carried out with the collaboration of Laboratorio Arbitral Agroalimentario (Ministry of Agriculture and Fishery, Spain), General Chemical State Laboratory D'xy Athinon (Greece), General State Laboratory (Ministry of Health, Cyprus), Departamento de Química Orgánica - Universidad de Córdoba (Spain), Istituto di Metodologie Chimiche (CNR, UNAPROL, Dipartimento di Chimica e Technologie Farmaceutiche ed Alimentari, Università di Genova, Italia), Fondazione Edmund Mach (Istituto San Michele all'Adige, Italy), and Eurofins Scientific Analytics (France), within the framework of the EU TRACE project. The true type (virgin or extra virgin) and origin of the olive oils at the national, regional, and PDO level were assured. The Italian samples were representative of the olive oil producing areas, which are markedly influenced by pedoclimatic factors from the north to the south of the country.

**NMR Analysis.** Aliquots of 40 $\mu$L of each VOO were dissolved in 200 $\mu$L of deuterated chloroform, shaken in a vortex, and placed in a 2 mm NMR capillary. The $^{1}$H NMR experiments were performed at 300 K on a Bruker (Rheinstetten, Germany) Avance 500 (nominal frequency of 500.13 MHz) equipped with a 2.5 mm broadband inverse probe. The spectra were recorded using a 7.5 $\mu$s pulse (90°), an acquisition time of 3.0 s (32K data points), a total recycling time of 4.0 s, a spectral width of 5500 Hz (11 ppm), and 64 scans (+ 4 dummy scans), with no sample rotation. Prior to Fourier transformation, the free induction decays (FIDs) were zero-filled to 64K and a 0.3 Hz line-broadening factor was applied. The chemical shifts are expressed in $\delta$ scale (ppm), referenced to the residual signal of chloroform (7.26 ppm) (*25*). The spectra were phase- and baseline-corrected manually. The multivariate data analysis was performed on a region of the NMR spectra between 0 and 7 ppm. The spectra were binned with 0.02 ppm wide buckets and normalized to total intensity over the region 4.10−4.26 ppm (glycerol signal). TopSpin 1.3 (2005) and Amix-Viewer 3.7.7 (2006) from Bruker BioSpin GMBH (Rheinstetten, Germany) were used to perform the processing of the spectra. The data table generated with the spectra of all samples was then used for pattern recognition. Eight buckets in the region 4.10−4.26 ppm (reference region) were excluded in the multivariate data analysis.

**IRMS Analysis.** Isotopic measurements of $\delta^{13}$C were performed by continuous flow IRMS using a Carlo Erba elemental analyzer (EA) EA-1108-CHN (Thermo Fisher, Milan, Italy) coupled to a DeltaPlus mass spectrometer (Thermo Fisher, Rodano, Italy). The $\delta^{13}$C signal for the reference peak was 4000 mV; the oxidation column temperature, 1050 °C; the reduction column temperature, 650 °C; and the GC column temperature, 65 °C. $\delta^{2}$H measurements were carried out by continuous flow IRMS using a total conversion elemental analyzer (TC/EA) coupled to a Delta PlusXP mass spectrometer (ThermoFisher, Rodano, Italy). The $\delta^{2}$H signal for the reference peak was 7000 mV; the GC column temperature, 80 °C; and the glassy-carbon column reactor temperature, 1450 °C.

The results of the carbon ($\delta^{13}$C) and hydrogen ($\delta^{2}$H) isotope ratio analyses are reported per mil (‰) on the relative $\delta$ scale and refer to the international standards V-PDB (Vienna Pee Dee Belemnite) for the carbon isotope ratio and V-SMOW (Vienna Standard Mean Ocean Water) for the hydrogen isotope ratio. All results were calculated according to the equation

$$\delta\ (‰) = [(R_{\text{sample}}/R_{\text{reference}}) - 1] \times 1000$$

where $R$ is the ratio of the heavy to light stable isotope (e.g., $^{2}$H/$^{1}$H) in the sample ($R_{\text{sample}}$) and in the standard ($R_{\text{reference}}$). The calibration of the control gases ($CO_2$ and $H_2$) was performed using the following reference materials: (i) for $\delta^{13}$C measurements, IAEA-CH7-Polyethylene ($\delta^{13}$C = −32.15‰) and IAEACH6-Sucrose ($\delta^{13}$C = −10.4‰) for $CO_2$ gas cylinder calibration; and (ii) for $\delta^{2}$H measurements, IAEA-CH7-Polyethylene ($\delta^{2}$H = −100.3‰) and V-SMOW ($\delta^{2}$H = 0‰) for $H_2$ gas cylinder calibration. An olive oil sample was calibrated with the international reference materials previously mentioned and used as a working standard. The standard was analyzed at regular intervals to control the acceptable repeatabilities of the measurements and to correct for possible drifts in the measurements. The standard deviations ($n$ = 10) determined using the corresponding reference gas were 0.05‰ for $\delta^{13}$C, and 0.8‰ for $\delta^{2}$H. Each olive oil sample was analyzed in triplicate, the standard deviations being < 0.15‰ for $\delta^{13}$C and < 2.7‰ for $\delta^{2}$H.

**Data Analysis.** The data set, made up of the values of the 342 buckets of the $^{1}$H NMR spectra (variables in columns) measured on the 963 VOOs analyzed (samples in rows), was first analyzed by univariate procedures (ANOVA, Fisher index, and box−whisker plots) and, afterward, by the following multivariate techniques, already described in the Literature Cited (*26*): unsupervised ones as principal component analysis (PCA) and supervised ones as linear discriminant analysis (LDA) and partial least-squares discriminant analysis (PLS-DA). Statistical and chemometric data

analyses were performed by means of the statistical software packages Statistica 6.1 (StatSoft Inc., Tulsa, OK, 1984−2004), The Unscrambler 9.1 (Camo Process AS, Oslo, Norway, 1986−2004), and SIMCA-P 11.0 (Umetrics AB, Umea, Sweden, 1992−2005).

In LDA, the variable selection strategy was the following. First, a modified best subset selection was used, which is a variable selection procedure that performs a search for the best subsets of a small number of variables that fulfill the criterion for choosing the best one (Wilks' lambda, rate of misclassification, etc.). This can be computed relatively quickly and in several steps: first, best subset selection is applied to the complete data matrix to obtain the first best (small) subset of variables; then, in a second step, best subset selection is sued on a data set omitting the variables selected in the first step, a second best subset is achieved, and so on. Finally, a refined selection of the variables selected successively in the previous steps was performed using forward stepwise selection (*26*).

In PLS-DA, PRESS or RMSEP are plotted against the number of the principal components to find the optimal number of PLS components. Sometimes there are several almost equivalent local minima on the curve; the first one should be preferred to avoid overfitting (according to the principle of parsimony). The model with the smallest number of features should be accepted from among equivalent models on the training set. Once PLS components are estimated by cross-validation, the classifications in the training-test set are represented in a box−whisker plot to define half of the distance between the quartiles as the boundary.

The supervised techniques were applied to the autoscaled (or standardized) or Pareto-scaled data matrix of the VOO profiles following these steps: (i) the data set was divided into a training-test set and an external data set; (ii) the training-test set was subsequently divided into a training set and a test set several times to perform cross-validation; (iii) the training-test set was used for the optimization of parameters characteristic of each multivariate technique by cross-validation, for instance, the number of PLS components in PLS-DA or for variable selection in LDA; (iv) a final mathematical model was built using all of the samples of the training-test set and the optimized parameters; (v) this model was validated using an independent test set of samples (external data set), that is, performing an external validation. During the parameter optimization step, the models were validated by three-fold cross-validation (3-fold CV) or leave-one-out cross-validation (LOO). The reliability of the classification models achieved in the cross-validation was studied in terms of recognition ability (percentage of the samples in the training set correctly classified during the modeling step) and prediction ability (percentage of the samples in the test set correctly classified by using the models developed in the training step). The reliability of the final model was evaluated in terms of classification ability (percentage of the samples of the training-test set correctly classified by using the optimized model) and the prediction ability in the external validation (percentage of the samples of the external set correctly classified by using the optimized model) (*26*).

## RESULTS AND DISCUSSION

**$^1$H NMR Spectra of VOOs.** $^1$H NMR spectra of the 941 VOOs produced in different PDO areas and/or regions from EU olive oil producing countries, namely, Italy, Spain, Greece, and France, and 22 VOOs from other countries from the Mediterranean basin (Turkey, Cyprus, and Syria) were recorded. Olive oil is mainly made up of triglycerides, differing in their substitution patterns in terms of length, degree, and kind of unsaturation of the acyl groups, and by minor components such as mono- and diglycerides, sterols, tocopherols, aliphatic alcohols, hydrocarbons, fatty acids, pigments, and phenolic compounds (*1*). The chemical shifts of the $^1$H signals of the triglycerides are well-known (*4*). Minor oil components are only observed by $^1$H NMR when their signals do not overlap with those of the main components and their concentrations are high enough to be detected (*21*, *27*−*30*). A typical $^1$H NMR spectrum of a VOO and the common $^1$H NMR signals of the major and some minor compounds together with their chemical shifts and their assignments to protons of the different functional groups are published elsewhere (*21*, *27*−*30*).

**Influence of the Harvest on $^1$H NMR Fingerprint of VOOs.** The data set consisted of a 963 × 342 matrix, in which rows represented the 963 samples of VOO and columns the 342 buckets of the $^1$H NMR spectrum. The presence of outliers in the data set was analyzed by PCA, and 28 extreme samples from different origins and harvests were removed after the presence of some irregularities in their NMR spectra was noted. The four first principal components (accounting for 63% of total system variability: PC1 for 31%, PC2 for 13%, PC3 for 11%, and PC4 for 7%) showed that samples were distributed in a compact cluster, even though some subgroupings according to harvest year were observed. PC2, PC3, and PC4 contained information related to the year; however, **Figure 1** shows that the three clusters partially overlapped. Because 70% of the samples were Italian and the rest from countries in the Mediterranean region, seasonal aspects seem to affect all samples in the same way, independently of their geographical origin. Therefore, in the modeling for the authentication of agricultural food products, it is important to have chemical data of several harvests to obtain general classification models that include the seasonal variability, as well. On the other hand, the PCA score plots did not show any clusters related to the geographical origin or the PDO of the oils. This indicates that the direction of maximum variability in the data set did not correspond to the direction of maximum discrimination among the geographical origins or PDOs. This suggests the presence of other sources of variability. Indeed, the year of harvest was confirmed to be one of them as seen above.

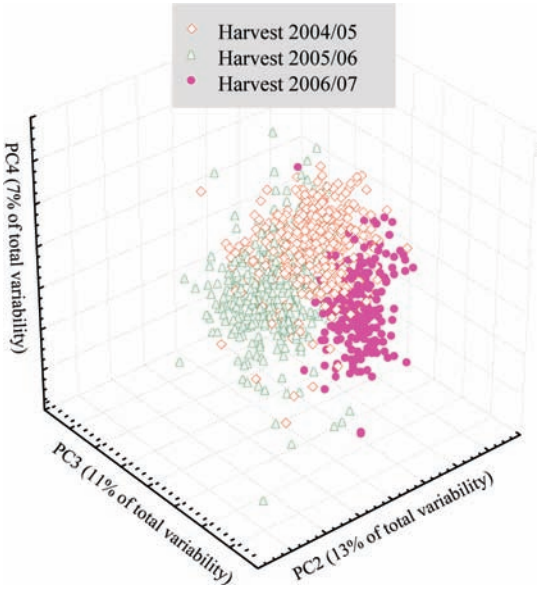


**Figure 1.** PCA score plot of the VOO samples on the space defined by PC2, PC3, and PC4.

**Geographical Characterization of Olive Oil.** The large data set of VOOs was studied regarding the situations that the antifraud authorities and regulatory bodies face. The PDO *Riviera Ligure*, some Italian regions, and the main countries that produce VOOs were used as examples to prove the potential of the tools to detect the mislabeling of non-PDO oils as PDO VOOs or the mislabeling of the provenance of VOOs at the regional or national level. With this purpose, several multivariate data analysis techniques, data sets, types of data scaling, and cross-validation were used. The best classification models were determined for each case study.

*PDO Olive Oils of Riviera Ligure*. Under the PDO of *Riviera Ligure*, extra virgin olive oils produced in Liguria (Italy) that fulfill the PDO requirements related to olive varieties, farming practices, oil extraction procedures, bottling, and labeling (Dossier Number: IT/PDO/0017/1540, *Off. J. Eur. Communities* **1997**, *L22*) can be marketed. The $^1$H NMR data set of VOOs from

**Table 1.** Classification Results Obtained by Supervised Pattern Recognition Techniques for the Authentication of VOO of the PDO *Riviera Ligure* Using $^1$H NMR Spectral Data (Unbalanced Data Set) and $\delta^{13}$C and $\delta^2$H Data[a]

| | | cross-validation | | | | model | | external validation | |
|---|---|---|---|---|---|---|---|---|---|
| | | % recognition | | % prediction | | % classification | | % prediction | |
| | *N*: | 126 | | 466 | | 126 | 466 | 73 | 270 |
| | prior prob: | 0.21 | | 0.79 | | 0.21 | 0.79 | | |
| technique | miscellaneous | validation | Liguria | non-Liguria | Liguria | non-Liguria | Liguria | non-Liguria | Liguria | non-Liguria |
| LDA[b] | 5 NMR buckets selected: 6.61, 5.09, 4.57, 4.05, and 0.33 ppm; autoscaling | 3-fold CV | 56.7 | 93.5 | 54.0 | 93.3 | 56.3 | 93.6 | 45.2 | 92.6 |
| LDA[b] | 5 NMR buckets selected: 6.61, 5.09, 4.57, 4.05, and 0.33 ppm; autoscaling IRMS: $\delta^{13}$C and $\delta^2$H | 3-fold CV | 56.7 | 93.5 | 54.0 | 93.3 | 57.9 | 94.4 | 45.2 | 92.6 |
| PLS-DA[b] | 3 PLS components selected boundary: 0.3180; autoscaling | 3-fold CV | 87.7 | 83.4 | 80.2 | 84.3 | 86.5 | 83.9 | 84.9 | 83.3 |
| PLS-DA[c] | 3 PLS components selected boundary: 0.3180; autoscaling | 3-fold CV | | | 81.7 | 84.1 | 86.5 | 83.7 | 84.9 | 83.3 |
| PLS-DA[c] | 3 PLS components selected boundary: 0.3180; autoscaling | LOO | | | 81.7 | 83.0 | 86.5 | 83.7 | 84.9 | 83.3 |
| PLS-DA[d] | 3 PLS components selected boundary: 0.3180; autoscaling | 3-fold CV/LOO | | | | | 86.5 | 83.9 | 84.9 | 83.3 |
| PLS-DA[d] | 3 PLS components selected boundary: 0.3175; Pareto scaling | 3-fold CV/LOO | | | | | 84.1 | 81.1 | 74.0 | 77.8 |

[a] Abbreviations: *N*, number of samples; prior prob, prior probability; LDA, linear discriminant analysis; PLS-DA, partial least-squares discriminant analysis. class codes: Liguria, 1; non-Liguria, 0. [b] Statistica. [c] The Unscrambler. [d] SIMCA-P.

**Table 2.** Classification Results Obtained by Supervised Pattern Recognition Techniques for the Authentication of VOO of the PDO *Riviera Ligure* Using $^1$H NMR Spectral Data (Balanced Data Set) and $\delta^{13}$C and $\delta^2$H Data[a]

| | | cross-validation | | | | model | | external validation | |
|---|---|---|---|---|---|---|---|---|---|
| | | % recognition | | % prediction | | % classification | | % prediction | |
| | *N*: | 132 | | 135 | | 132 | 135 | 67 | 601 |
| | prior prob: | 0.49 | | 0.51 | | 0.49 | 0.51 | | |
| technique | miscellaneous | validation | Liguria | non-Liguria | Liguria | non-Liguria | Liguria | non-Liguria | Liguria | non-Liguria |
| LDA[b] | 5 NMR buckets selected: 6.61, 5.11, 4.57, 4.05, and 0.33 ppm; autoscaling | 3-fold CV | 84.1 | 85.9 | 84.1 | 83.7 | 82.6 | 85.2 | 86.6 | 79.7 |
| LDA[b] | 4 NMR buckets selected: 5.11, 4.57, 4.05, and 0.33 ppm IRMS: $\delta^{13}$C and $\delta^2$H | 3-fold CV | 88.3 | 84.1 | 85.6 | 80.7 | 87.9 | 83.0 | 89.6 | 79.7 |
| PLS-DA[b] | 5 PLS components selected boundary: 0.540; autoscaling | 3-fold CV | 91.3 | 92.6 | 87.9 | 86.7 | 91.7 | 90.4 | 88.1 | 85.5 |
| PLS-DA[c] | 5 PLS components selected boundary: 0.540; autoscaling | 3-fold CV | | | 86.4 | 85.9 | 91.7 | 90.4 | 88.1 | 85.5 |
| PLS-DA[c] | 5 PLS components selected IRMS: $\delta^{13}$C and $\delta^2$H boundary: 0.547; autoscaling | 3-fold CV | | | 86.4 | 85.2 | 91.7 | 91.9 | 86.6 | 86.0 |
| PLS-DA[c] | 5 PLS components selected boundary: 0.540; autoscaling | LOO | | | 87.1 | 85.9 | 91.7 | 90.4 | 88.1 | 85.5 |
| PLS-DA[c] | 5 PLS components selected IRMS: $\delta^{13}$C and $\delta^2$H boundary: 0.547; autoscaling | LOO | | | 85.6 | 85.2 | 91.7 | 91.9 | 86.6 | 86.0 |
| PLS-DA[d] | 5 PLS components selected boundary: 0.540; autoscaling | 3-fold CV/LOO | | | | | 91.7 | 90.4 | 88.1 | 85.5 |
| PLS-DA[d] | 4 PLS components selected boundary: 0.520; Pareto scaling | 3-fold CV/LOO | | | | | 87.1 | 83.0 | 80.6 | 81.0 |

[a] See abbreviations for **Table 1**. [b] Statistica. [c] The Unscrambler. [d] SIMCA-P.

different geographical origins and PDOs was used to differentiate between VOOs belonging to the PDO *Riviera Ligure* and those not belonging to this PDO. Each VOO was represented in the 342-dimensional space by a data vector made of the 342 NMR variables. Univariate techniques (ANOVA, Fisher index, and box−whisker plots) cannot select a single variable to distinguish between Ligurian (belonging to the PDO *Riviera Ligure*) and non-Ligurian (not belonging to the PDO) samples. Therefore, it was necessary to apply supervised pattern recognition methods to build classification models that can distinguish VOOs of this PDO from the rest. Several multivariate approaches (LDA and PLS-DA) were tested using balanced or unbalanced data sets, different cross-validation methods (LOO and 3-fold CV), and different data scalings (autoscaling and Pareto-scaling) to find the best approach for the authenticity and traceability of PDO olive oils. **Tables 1** and **2** summarize the classification results.

LDA models obtained using an unbalanced training-test set were biased to the class with more representatives, that is, the non-Liguria class, presenting 93% of hits, whereas for the Liguria oils, the recognition and prediction abilities were < 57% (**Table 1**). LDA was very sensitive to imbalances in the number of samples of each category in the data set, as expected from the literature (*26*). In contrast, PLS-DA was not so sensitive to imbalances in the data set and performed better than LDA. The PLS-DA model, using the autoscaled data set, presented recognition and prediction abilities in the cross-validation of 88 and 80%, respectively, for the VOOs from Liguria, and 83 and 84%, respectively, for the non-Liguria VOOs. The percentage of classification of the final model (three PLS components and the boundary at 0.318) and the prediction in the external validation were close to each other, 87 and 85% for Liguria VOOs and 84 and 83% for non-Liguria VOOs, respectively, as well as to the recognition and prediction abilities in the training step. This would be considered indicative of a satisfactory model. However, the fact that the prediction ability was slightly higher than the recognition ability for the non-Liguria class indicated that the model did not perform properly for this class, which was probably due to the imbalances in the data set. Indeed, the percentage of

correct classifications for the non-Liguria category was slightly higher than its a priori probability (79%). The cross-validation used in the training step did not influence the prediction abilities for either category, indicating that the samples were well-represented in the training set. PLS-DA using the autoscaled unbalanced data set attained the same best final model, consisting of three PLS components and the boundary at 0.318 (class codes: Liguria, 1; non-Liguria, 0). PLS-DA applied on the Pareto-scaled unbalanced data set provided a model with four PLS components and the boundary at 0.325, which performed worse.

Both supervised pattern recognition techniques, LDA and PLS-DA, performed better if a balanced training-test set was used. However, PLS-DA still outperformed LDA. LDA achieved classifications of around 85% of hits for both categories. PLS-DA provided a model with five PLS components and the boundary at 0.540, which achieved slightly better results for the Liguria class (prediction ability in the cross-validation, 86−88%; classification ability of the final model, 92%; and prediction ability of the final model in the external validation, 88%) than for the non-Liguria VOOs (86−87, 90, and 86%, respectively). These results together with the facts that in the cross-validation the recognition ability was higher but close to the prediction ability and the classification ability of the final model was also higher but close to prediction ability in the external validation disclosed that the model achieved was feasible and not random, as well as well-represented by the samples in the data set. PLS-DA on the Pareto-scaling balanced data set produced a model with four PLS components and the boundary at 0.520. This model also gave better classifications (classification ability of the final model, 87 and 83% for the Liguria and non-Liguria, respectively; and prediction ability of the final model in the external validation, 81% for both classes) than the model made with the unbalanced data set. With both data sets (balanced and unbalanced), Pareto-scaling led to worse outcomes than autoscaling.

The variable selection used for LDA afforded five NMR buckets centered at the following chemical shifts: 6.61, 5.11 or 5.09, 4.57, 4.05, and 0.33 ppm. These buckets correspond to signals of the following VOO components: phenolic compounds and unsaturated alcohols, which present characteristic resonances in the spectral region 6−7.5 ppm (*31*) and 4.5−5 ppm, respectively; *sn*-1,2-diglycerides (5.09−5.11 ppm) and *sn*-1,3-diglycerides (4.05 ppm), due to their CH glycerol protons; and cycloartenol (0.33 ppm), to the methylene proton of its cyclopropanoic ring (*30*).

The weighted regression coefficients (*32*) of the PLS models indicate the importance of the NMR variables on the model: the larger the regression coefficient, the higher the influence of the variable on the PLS model. The variables selected in LDA were among the variables that presented the highest weighted regression coefficients in the PLS-DA models: 6.85−6.83, 6.75, 6.67, 6.59, and 6.23 ppm belonged to signals of phenolic compounds; 5.15−5.07 ppm was due to the CH glycerol protons of *sn*-1,2-diglycerides; 4.99 ppm was due to unsaturated alcohols; 4.71, 4.65, and 4.57 ppm were due to terpenes; 2.79 ppm was due to diallylic proton of linolenic acyl group; 1.29 ppm was due to methylene proton of linoleic and linolenic acyl group; and 0.33 ppm was due to cycloartenol. Therefore, both pattern recognition techniques arrived at consistent results, each one providing information about the most important features for the characterization of PDO *Riviera Ligure* VOOs.

With the additional information provided by the $\delta^2$H and $\delta^{13}$C isotopes measured by IRMS, the classification results of the LDA model using the unbalanced data set were similar, even though both isotopes were significant variables (at the 5% level) together with the five NMR buckets previously selected (**Table 1**). This was probably due to the imbalances in the data set. However, with the balanced data set, LDA provided a model with four significant NMR variables previously selected and the two isotopes (both significant) that obtained (**Table 2**): (i) slightly better classifications for the Liguria class than the model made only with NMR data; (ii) whereas the classification results for the non-Liguria class were slightly worse. This fact suggested that $\delta^2$H and $\delta^{13}$C isotopes contained some information related to VOOs from Liguria. As a matter of fact, Angerosa et al. (*23*) and, more recently, Camin et al. (*33*) observed that the values of $\delta^{13}$C and $\delta^2$H in olive oils increased according to the olive cultivation latitude, from northern to southern Italy. Thus, olive oils produced in northern, central, and southern Italy could be well differentiated by these isotopes. However, the discrimination of olive oils produced in different regions at similar latitudes is more difficult. Olive oils from Liguria (in northwestern Italy) present a different chemical composition and particular organoleptic characteristics, in comparison with other northern Italian olive oils such as those from the region of Garda lake. This is due to the proximity of Liguria to the sea and the special climate of the region. PLS-DA was also applied on an autoscaled balanced data set that contained 344 variables (342 NMR buckets and the 2 isotopes); however, no improvement was observed in the classification results of the models achieved (neither LOO nor 3-fold CV).

The best model for the distinction between VOOs belonging to the PDO *Riviera Ligure* and other VOOs was afforded by PLS-DA using an autoscaled balanced training-test set.

*Virgin Olive Oils from Other Regions*. The large data set available (963 × 342 matrix) was used to authenticate VOOs produced in selected Italian regions. The regions selected were those best represented in the data set. Models were generated for the following Italian regions: Umbria (which is also a registered PDO, PDO *Umbria*, Dossier IT/PDO/0017/1520, *Off. J. Eur. Communities* **1997**, *L322*); Sicily (six PDOs: *Monte Etna*, *Val di Mazara*, *Valli Trapanesi*, *Valle del Belice*, *Valdemone*, and *Monti Iblei*); Puglia (four PDOs: *Terra d'Otranto*, *Collina di Brindisi*, *Dauno*, and *Terra di Bari*); Lazio (three PDOs: *Tuscia*, *Canino*, and *Sabina*); Garda (three PDOs: *Garda*, *Laghi Lombardi*, and *Veneto Valpolicella, Veneto Euganei e Berici, Veneto del Grappa*); Campania (3 PDOs: *Peninsola Sorrentina*, *Colline Salernitane*, and *Cilento*); and Calabria (three PDOs: *Lametia*, *Alto Crotonese*, and *Bruzio*).

Taking into account the results obtained by the different approaches studied for the PDO *Riviera Ligure* and that the number of samples for each of these regions was considerably smaller than in the Liguria category, the models for these regions were developed using an autoscaled balanced training-test set by PLS-DA and LOO CV. The final models were also evaluated by external validation. The results are summarized in **Table 3**. The model obtained to authenticate VOOs from Sicily recognized 98% of the Sicilian oils and 89% of the non-Sicilian ones and managed to correctly predict in the cross-validation step 93 and 86% of Sicilian and non-Sicilian oils, respectively. Because this model achieved predictions in the external validation (> 85% of hits for both categories) similar to those in the modeling step, it can be considered stable and robust. In contrast, the models created for other regions such as Lazio, Garda, and Calabria were not so satisfactory: although the classification abilities were close to 90% of correct hits or even higher, the prediction abilities in the cross-validation were from 10 to 24% lower, which meant that the classification results were very dependent on the samples included in the training set in the modeling step. This also occurred for Umbria and Campania, but the models achieved about 80% of correct classification for the training set, and predictions on the

**Table 3.** Classification Results Obtained by Supervised Pattern Recognition Techniques for the Authentication of VOO from Certain Italian Regions Using $^1$H NMR Spectral Data and $\delta^{13}$C and $\delta^2$H Data[a]

| | | cross-validation | | | | model | | | | external validation | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | % prediction | | | | % classification | | | | % prediction | | | |
| model | origin | *N* | prior prob | NMR | NMR + IRMS | *N* | prior prob | NMR | NMR + IRMS | *N* | prior prob | NMR | NMR + IRMS |
| Umbria vs non-Umbria | Umbria | 35 | 0.45 | 71.4 | 71.4 | 35 | 0.45 | 82.9 | 82.9 | 12 | 0.014 | 50.0 | 50.0 |
| (a) NMR: 2 PLS components, boundary: 0.525 | non-Umbria | 43 | 0.55 | 74.4 | 72.1 | 43 | 0.55 | 79.1 | 79.1 | 845 | 0.986 | 74.8 | 76.7 |
| (b) NMR + IRMS: 2 PLS components, boundary: 0.5325 | | | | | | | | | | | | | |
| Sicily vs non-Sicily | Sicily | 54 | 0.47 | 92.6 | 92.6 | 54 | 0.47 | 98.1 | 98.1 | 24 | 0.029 | 87.5 | 91.7 |
| (a) 3 PLS components, boundary: 0.460 | non-Sicily | 62 | 0.53 | 85.5 | 85.5 | 62 | 0.53 | 88.7 | 87.1 | 795 | 0.971 | 85.8 | 84.9 |
| (b) NMR + IRMS: 3 PLS components, boundary: 0.452 | | | | | | | | | | | | | |
| Puglia vs non-Puglia | Puglia | 47 | 0.42 | 68.1 | 63.8 | 47 | 0.42 | 72.3 | 72.3 | 22 | 0.027 | 81.8 | 81.8 |
| (a) 2 PLS components, boundary: 0.4435 | non-Puglia | 64 | 0.58 | 62.5 | 70.3 | 64 | 0.58 | 71.9 | 71.9 | 802 | 0.973 | 65.1 | 67.1 |
| (b) NMR + IRMS: 2 PLS components, boundary: 0.451 | | | | | | | | | | | | | |
| Lazio vs non-Lazio | Lazio | 40 | 0.49 | 80.0 | 85.0 | 40 | 0.49 | 97.5 | 97.5 | 19 | 0.022 | 73.7 | 84.2 |
| (a) 4 PLS components, boundary: 0.515 | non-Lazio | 41 | 0.51 | 68.3 | 75.6 | 41 | 0.51 | 90.2 | 90.2 | 835 | 0.978 | 69.3 | 70.5 |
| (b) NMR + IRMS: 4 PLS components, boundary: 0.533 | | | | | | | | | | | | | |
| Garda vs non-Garda | Garda | 36 | 0.46 | 72.2 | 77.8 | 36 | 0.46 | 91.7 | 88.9 | 13 | 0.015 | 69.2 | 76.9 |
| (a) 3 PLS components, boundary: 0.555 | non-Garda | 43 | 0.54 | 74.4 | 74.4 | 43 | 0.54 | 90.7 | 90.7 | 843 | 0.985 | 80.1 | 81.9 |
| (b) NMR + IRMS: 3 PLS components, boundary: 0.538 | | | | | | | | | | | | | |
| Campania vs non-Campania | Campania | 21 | 0.43 | 71.4 | 71.4 | 21 | 0.43 | 81.0 | 81.0 | 7 | 0.008 | 57.1 | 57.1 |
| (a) 2 PLS components, boundary: 0.430 | non-Campania | 28 | 0.57 | 64.3 | 64.3 | 28 | 0.57 | 78.6 | 78.6 | 879 | 0.992 | 62.9 | 63.0 |
| (b) NMR + IRMS: 2 PLS components, boundary: 0.4315 | | | | | | | | | | | | | |
| Calabria vs non-Calabria | Calabria | 17 | 0.38 | 70.6 | 70.6 | 17 | 0.38 | 94.1 | 94.1 | 5 | 0.006 | 60.0 | 60.0 |
| (a) 3 PLS components, boundary: 0.445 | non-Calabria | 28 | 0.62 | 85.7 | 85.7 | 28 | 0.62 | 96.4 | 96.4 | 885 | 0.994 | 79.9 | 80.3 |
| (b) NMR + IRMS: 3 PLS components, boundary: 0.447 | | | | | | | | | | | | | |

[a] See abbreviations for **Table 1**. Models obtained by PLS-DA using autoscaling, LOO, and The Unscrambler. Class codes: region, 1; non-region, 0.

test set were more than 10% lower, except for the oils belonging to the non-Umbria category (5% less). The external validation of some models (only 50% of Umbria, 57% of Campania, and 60% of Calabria VOOs were correctly predicted) confirmed that the classes were not well represented in the modeling step. Puglian VOOs, as well as non-Garda VOOs, were much better predicted in the external data set (82% of hits) than in the cross-validation (68 and 72% of hits, respectively). This was probably due to the way samples were divided into the training-test set and the external set: the PCA scores of all the VOOs were regarded to select samples from the whole cloud of points including the borders. This procedure ensured that the training-test set was representative of all the samples (at least of the three harvests studied); however, the predictions on the external set could be overly optimistic.

The most influential variables, that is, those with the highest weighted regression coefficients, on the binary PLS-DA models achieved for each region are listed in **Table 4**. The signals due to cycloartenol (0.31−0.33 ppm) and *sn*-1,2-diglycerides (5.07−5.15-ppm) were important for all models except for Garda, as well as the resonances in the phenolic region at 6.73−6.79 ppm, which only did not influence the model for Sicily. The acyl group methylene protons of saturated fatty acids (1.23 ppm), $^{13}$C satellite of signal at 4.09−4.32 ppm (α-methylene protons of the glyceryl group of triglycerides) at 3.97 ppm, and the signal at 5.57 ppm were important specifically for the Umbria model; the signals at 0.53 and 0.79 ppm, for the Sicily model; the methylic proton of the C18-steroid group of *β*-sitosterol (0.67 ppm) and the terpene signal at 4.57−4.59 ppm, for the Puglia model; the signal of the cycloartenol at 0.55 ppm, $^{13}$C satellite of signal at 2.26−2.32 ppm (α-methylene protons of the acyl group) at 2.15 ppm, the glycerol proton of *sn*-1,2-diglycerides (3.71 ppm), and signals at 6.19 ppm and 6.15 ppm in the phenolic region, for the Lazio model; signals in the regions 1.35−1.43, 2.35−2.39, and 4.33−4.35 ppm, the α-methylene protons of the acyl group (2.29 and 2.33 ppm), the signal at 3.75 ppm, the α-methylene protons of the glyceryl group of triglycerides (4.27 ppm), and the signal at 6.15 ppm in the phenolic region for the Campania model; and the signal at 5.93 ppm for the Calabria model. The glyceryl protons of *sn*-1,3-diglycerides (4.05−4.07 ppm) and triglycerides (5.25 and 5.29 ppm) were influential for the models of Umbria, Lazio, Umbria, and Campania, respectively; signals in the phenolic region at 6.25−6.29 ppm for the models of Puglia and Calabria; signals in the phenolic region at 6.63−6.65 and 6.69−6.71 ppm for the models of Umbria and Campania; and signals in the phenolic region at 6.45−6.47 ppm for the models of Umbria and Garda.

These results disclosed that $^1$H NMR spectra of VOOs contained information related to the region of provenance of the oil, but further studies should be carried out with a considerably larger sample set for each region, and even for each of their PDOs, to guarantee the detection of fraud when VOO is falsely labeled as belonging to a certain origin. In this regard, Sicily, which is an island at the southernmost point of Italy, produces an olive oil that is markedly influenced by pedoclimatic factors, in accordance with its geographical position. It is therefore coherent that the VOO produced on this island presents a characteristic chemical composition that allows one to distinguish it from all other VOOs from different geographical regions. In contrast, the stable isotopes measured on the samples, $\delta^2$H and $\delta^{13}$C, which are commonly related to pedoclimatic features, did not improve the classification of the Sicilian and non-Sicilian VOOs, even if $\delta^{13}$C was a significant variable in the model. Similarly, these stable isotopes did not enhance the classification results for the Umbria, Campania, and Calabria models. In the Umbria model, both isotopes were among the variables with the highest weighted regression coefficients, whereas they were not in the Campania and Calabria models. However, both stable isotopes provided extra useful information related to the non-Puglia category in the Puglia model, as well as for the Garda class in the Garda model. $\delta^{13}$C data substantially improved the classification for both classes of the Lazio model. Therefore, depending on the case

**Table 4.** Most Important Variables in the Binary PLS-DA Models Achieved for the Geographical Classification of VOOs at the Regional Level by $^1$H-NMR

| Umbria | Sicilia | Puglia | Lazio | Garda | Campania | Calabria | functional group | attribution |
|---|---|---|---|---|---|---|---|---|
| 0.33 | 0.33−0.31<br>0.53 | 0.31 | 0.33−0.31 | | 0.29 | 0.33−0.31 | −C$H_2$− (cyclopropanic ring) | cycloartenol |
| | | | 0.55 | | | | −C$H_2$− (cyclopropanic ring) | cycloartenol |
| | 0.79 | 0.67 | | | | | −C$H_3$ (C18-steroid group) | $\beta$-sitosterol |
| | 0.95 | | 0.97 | | 0.95−0.91 | 0.97 | −C$H_3$ (acyl group) | linolenic (or $\omega$-3) |
| 1.01 | 1.03−0.99 | 1.01−0.99 | | | | | −C$H_3$ ($^{13}$C satellite of signal at 0.87 ppm) | |
| 1.23 | | | | | | | −(C$H_2$)$_n$− (acyl group) | saturated (palmitic, stearic) |
| 1.27 | | 1.25 | 1.25 | 1.27 | 1.25 | | −(C$H_2$)$_n$− (acyl group) | oleic |
| 1.29 | | | 1.29 | | 1.29<br>1.43−1.35 | 1.29 | −(C$H_2$)$_n$− (acyl group) | linoleic and linolenic |
| 1.67, 1.61−1.59 | 1.67, 1.59 | 1.67 | 1.75−1.73 | | 1.65, 1.59, 1.51<br>1.75−1.73 | | −OCO—C$H_2$—C$H_2$− (acyl group) | |
| | 2.07−2.03, 1.99 | | 1.99, 1.95 | 2.05, 2.01 | 2.05−1.99 | | −C$H_2$—CH=CH− (acyl group) | |
| | | | 2.15 | | | | −OCO—C$H_2$− ($^{13}$C satellite of signal at 2.26−2.32 ppm, acyl group) | |
| | | | | | 2.33, 2.29<br>2.39−2.35 | | −OCO—C$H_2$− (acyl group) | |
| 2.75−2.73 | 2.75−2.71 | | | 2.75−2.73 | | | =CH—C$H_2$—CH= (acyl group) | linoleic |
| 2.77 | 2.77 | | | 2.77 | 2.77 | | =CH—C$H_2$—CH= (acyl group) | linoleic and linolenic |
| | 2.79 | | | 2.79 | 2.83 | 2.81−2.79 | =CH—C$H_2$—CH= (acyl group) | linolenic |
| | | | 3.71 | | 3.75 | | −C$H_2$OH (glyceryl group) | *sn* 1,2-diglycerides |
| 3.97 | | | | | | | −C$H_2$OCOR ($^{13}$C satellite of signal at 4.09−4.32 ppm, glyceryl group) | triglycerides |
| 4.07−4.05 | | | 4.07−4.05 | | | | >C$H$—OH (glyceryl group) | *sn* 1,3-diglycerides |
| | | | | | 4.27<br>4.35−4.33 | | −C$H_2$OCOR (glyceryl group) | triglycerides |
| | | 4.59−4.57 | | | | | | terpene |
| 4.65 | 4.65 | 4.65 | 4.65 | | | 4.65 | | terpene |
| 4.71−4.69 | 4.71−4.69 | 4.71 | 4.71 | | | | | terpene |

**Table 4.** Continued

| Umbria | Sicilia | Puglia | Lazio | Garda | Campania | Calabria | functional group | attribution |
|---|---|---|---|---|---|---|---|---|
| 5.15−5.07 | 5.15−5.07 | 5.15−5.07 | 5.15−5.11 | | 5.15−5.07 | 5.15−5.07 | >C*H*OCOR (glyceryl group) | *sn* 1,2-diglycerides |
| 5.29 | | | | | 5.25 | | >C*H*OCOR (glyceryl group) | triglycerides |
| 5.33 | 5.37 | | 5.33 | | 5.43−5.33 | 5.33 | −C*H*═C*H*− (acyl group) | |
| 5.57 | | | | | | | | |
| 5.75 | 5.73 | | | | | | | |
| | | | | | | 5.93 | | |
| 6.01−5.97 | 5.99 | | | 6.01−5.97 | | | | phenolic compounds |
| 6.05−6.03 | | | 6.05−6.03 | 6.05 | | | | phenolic compounds |
| | | | 6.19, 6.15 | | | | | phenolic compounds |
| | | | 6.23 | 6.23 | | 6.23−6.21 | | phenolic compounds |
| | | 6.29−6.25 | | | | 6.29−6.25 | | phenolic compounds |
| | | | | | 6.37 | | | phenolic compounds |
| 6.47−6.45 | | | | 6.47 | | | | phenolic compounds |
| 6.55−6.53 | 6.55−6.53 | | | 6.57−6.51 | | | | phenolic compounds |
| 6.61 | | 6.59 | 6.59 | 6.61−6.59 | 6.61−6.59 | | | phenolic compounds |
| 6.65−6.63 | | | | | 6.63 | | | phenolic compounds |
| 6.71−6.69 | | | | | 6.71−6.67 | | | phenolic compounds |
| 6.75 | | 6.77−6.73 | 6.77−6.75 | 6.79−6.73 | 6.79−6.73 | 6.77 | | phenolic compounds |
| 6.95 | | | 6.95−6.85 | 6.97−6.85 | | 6.87 | | phenolic compounds |

**Table 5.** Classification Results Obtained by Supervised Pattern Recognition Techniques for the Authentication of VOO from the Main Producing Countries, Italy, Spain, and Greece, Using $^1$H NMR Spectral Data and $\delta^{13}$C and $\delta^2$H Data[a]

| | | cross-validation | | | | model | | | | external validation[b] | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | % prediction | | | | % classification | | | | % prediction | | | |
| model | origin | *N* | prior prob | NMR | NMR + IRMS | *N* | prior prob | NMR | NMR + IRMS | *N* | prior prob | NMR | NMR + IRMS |
| Italy vs non-Italy | Italy | 72 | 0.35 | 75.0 | 79.2 | 72 | 0.35 | 88.9 | 90.3 | 568 | 0.78 | 75.7 | 80.1 |
| (a) NMR: 4 PLS components, boundary: 0.4020 | non-Italy | 135 | 0.65 | 77.0 | 85.2 | 135 | 0.65 | 84.4 | 90.4 | 160 | 0.22 | 71.9 | 73.1 |
| (b) NMR + IRMS: 4 PLS components, boundary: 0.4225 | | | | | | | | | | | | | |
| Spain vs non-Spain | Spain | 71 | 0.34 | 78.9 | 77.5 | 71 | 0.34 | 88.7 | 87.3 | 70 | 0.10 | 92.9 | 90.0 |
| (a) NMR: 3 PLS components, boundary: 0.3563 | non-Spain | 136 | 0.66 | 80.9 | 83.8 | 136 | 0.66 | 85.3 | 88.2 | 658 | 0.90 | 67.2 | 71.7 |
| (b) NMR + IRMS: 3 PLS components, boundary: 0.3677 | | | | | | | | | | | | | |
| Greece vs non-Greece | Greece | 64 | 0.31 | 92.2 | 96.9 | 64 | 0.31 | 98.4 | 98.4 | 31 | 0.04 | 96.8 | 96.8 |
| (a) NMR: 5 PLS components, boundary: 0.4725 | non-Greece | 143 | 0.69 | 93.7 | 91.6 | 143 | 0.69 | 97.9 | 96.5 | 697 | 0.96 | 90.0 | 91.1 |
| (b) NMR + IRMS: 5 PLS components, boundary: 0.4582 | | | | | | | | | | | | | |
| Italy vs Spain vs Greece | | | | | | | | | | | | | |
| (a) NMR: 5 PLS components, boundary: 0.4120 | Italy | 72 | 0.35 | 69.4 | 79.2 | 72 | 0.35 | 80.6 | 83.3 | 568 | 0.85 | 65.5 | 73.8 |
| (b) NMR + IRMS: 5 PLS components, boundary: 0.4500 | | | | | | | | | | | | | |
| (a) NMR: 5 PLS components, boundary: 0.3570 | Spain | 71 | 0.34 | 69.0 | 67.6 | 71 | 0.34 | 74.6 | 77.5 | 70 | 0.10 | 78.6 | 78.6 |
| (b) NMR + IRMS: 5 PLS components, boundary: 0.3616 | | | | | | | | | | | | | |
| (a) NMR: 5 PLS components, boundary: 0.4395 | Greece | 64 | 0.31 | 87.5 | 90.6 | 64 | 0.31 | 93.8 | 100.0 | 31 | 0.05 | 80.6 | 83.9 |
| (b) NMR + IRMS: 5 PLS components, boundary: 0.4350 | | | | | | | | | | | | | |

[a] See abbreviations for **Table 1**. Models obtained by PLS-DA using autoscaling, LOO, and The Unscrambler. Class codes: "country", 1; "non-country, 0. [b] The external data set used to evaluate the three-class model consisted of samples from Italy, Spain, and Greece.

study, stable isotopes can provide useful and complementary information to that contained in the $^1$H NMR fingerprint of the VOOs related to their geographical origin. Camin et al. (*33*) found that olive oils produced in the southern regions of Italy had similar isotopic signatures, making a clear discrimination among them difficult. The improvement observed on the Garda model can be attributed to the fact that these isotopes contain information referring to the latitude of this region, that is, the northern part of Italy, and its special environmental conditions, that is, low mean temperatures and rainy weather. Thus, Garda olive oils present lower $\delta^{13}$C and $\delta^2$H than olive oils produced in the rest of Italy.

*Main Producers of Olive Oil: Spain, Italy, and Greece*. With regard to the adulteration of VOOs from a certain country with VOOs produced in another country at a lower cost, or the false labeling of the VOOs as coming from a certain country when they were actually produced in another, the need for chemical approaches to detect these fraudulent activities is evermore apparent.

The $^1$H NMR data of the VOOs from the main olive oil producing countries, that is, Spain, Italy, and Greece, were analyzed by multivariate techniques with the purpose of creating classification models that would allow the distinction between the geographical origins of VOOs from these three countries. **Table 5** shows the results. The model distinguishes VOOs from Greece from all the rest of the VOOs; it classified properly >97% of the samples of both categories, Greece and non-Greece, and predicted correctly >90% of the samples in the test set of the cross-validation, as well as in the external validation. The binary models for Italy and Spain presented classification abilities of 89% for the Italian oils and the Spanish oils, 84% for the non-Italy category, and 85% for the non-Spain category. The prediction abilities in the cross-validation for the model for Spain were ca. 80% of hits for both classes, whereas the predictions in the external validation were considerably different; for the Spanish VOOs it was overly optimistic (92%), and for the non-Spanish VOOs it was considerably low (67%). In the model for Spain, the variability of the non-Spain category was under-represented in the training-test sets. Therefore, this model did not provide good predictions for this category in the external set. The model for Italy provided prediction abilities in the cross-validation of ca. 76% for both classes and in external validation, close to this value. These predictions were substantially lower than the recognition ability of the model, indicating that the model was dependent on the samples included in the training set.

**Table 6** gathers the most influential variables, that is, those with the highest weighted regression coefficients, on the binary PLS-DA models obtained for each country, identifying the functional groups and compounds to which the signals are due. The signals in the phenolic regions at 6.45−6.47 and 6.83−6.85 ppm were important for the three models. In contrast, the model for Spain was particularly influenced by the methylic proton of the C18-steroid group of $\beta$-sitosterol (0.67 ppm), the $\beta$-methylene protons of the acyl group (1.59 and 1.67 ppm), the allylic protons of the acyl group (1.99−2.07 ppm), the diallylic protons of the acyl group of linoleic (2.73−2.77 ppm) and linolenic (2.77−2.81 ppm), the glycerol proton of *sn*-1,2-diglycerides (3.71 ppm), *sn*-1,3-diglycerides (4.05−4.07 ppm), and triglycerides (5.25 and 5.29 ppm), the olefinic protons of the acyl groups (5.37 ppm), the signals in the phenolic region at 6.37, 6.61, and 6.71 ppm, and the signals at 0.53, 1.75−1.77, and 2.35 ppm. Among the most important variables, those that affected only the model for Greece were the methylic proton of the linolenic acyl group (0.97 ppm), the terpene signal at 4.55−4.57 ppm, and the signals at 0.77 and 3.81 ppm. The resonances of cycloartenol (0.31−0.33 and 0.55 ppm) and phenolic compounds at 6.23 and 6.27 ppm were important for the models of Italy and Greece.

A ternary model was developed to classify VOOs from three countries: Italy, Spain, and Greece. It did not classify as well as the binary models created for each country (**Table 5**). Indeed, PLS-DA is known to perform better with a smaller number of classes (*26*). Thus, the model recognized 94% of Greek, 81% of Italian, and 75% of Spanish oils in the training-test sets and predicted correctly 88% of Greek oils and 69% of the samples from Italy and Spain in the cross-validation and 81% of Greek, 79% of Spanish, and 66% of Italian oils in the external validation.

In conclusion, these results show that $^1$H NMR fingerprinting of VOOs can be a useful tool to ensure authenticity and traceability of VOOs at the national level. From this study, a stable model was achieved to distinguish Greek VOOs from oils from other countries. However, for Italian and Spanish

**Table 6.** Most Important Variables in the Binary PLS-DA Models Achieved for the Geographical Classification of VOOs at the National Level by $^1$H NMR

| Italy | Spain | Greece | functional group | attribution |
|---|---|---|---|---|
| 0.33−0.31 | | 0.33−0.31 | −**C**$H_2$− (cyclopropanic ring) | cycloartenol |
| | 0.53 | | | |
| 0.55 | | 0.55 | −**C**$H_2$− (cyclopropanic ring) | cycloartenol |
| | 0.67 | | −**C**$H_3$ (C18-steroid group) | $\beta$-sitosterol |
| | | 0.77 | | |
| | | 0.97 | −**C**$H_3$ (acyl group) | linolenic (or $\omega$-3) |
| | 1.67, 1.59 | | −OCO—$CH_2$—$CH_2$− (acyl group) | |
| | 1.77−1.75 | | | |
| | 2.07−1.99 | | −**C**$H_2$—**CH**=**CH**− (acyl group) | |
| | 2.35 | | | |
| | 2.75−2.73 | | =**CH**—**C**$H_2$—**CH**= (acyl group) | linoleic |
| | 2.77 | | =**CH**—**C**$H_2$—**CH**= (acyl group) | linoleic and linolenic |
| | 2.81−2.79 | | =**CH**—**C**$H_2$—**CH**= (acyl group) | linolenic |
| | 3.61 | 3.63−3.61 | | |
| | 3.71 | | −**C**$H_2$**OH** (glyceryl group) | *sn* 1,2-diglycerides |
| | | 3.81 | | |
| | 4.07−4.05 | | >**C**$H$—**OH** (glyceryl group) | *sn* 1,3-diglycerides |
| | | 4.57−4.55 | | terpene |
| | 4.65 | 4.65 | | terpene |
| | 4.71−4.69 | 4.71−4.69 | | terpene |
| | 5.15−5.09 | 5.13 | >**C**$H$**OCOR** (glyceryl group) | *sn* 1,2-diglycerides |
| | 5.25 | | >**C**$H$**OCOR** (glyceryl group) | triglycerides |
| | 5.37 | | −**C**$H$=**C**$H$− (acyl group) | |
| | 5.75−5.71 | 5.73 | | |
| 6.23 | | 6.23 | | phenolic compounds |
| 6.27 | | 6.27 | | phenolic compounds |
| | 6.37 | | | phenolic compounds |
| 6.47−6.45 | 6.45 | 6.47−6.45 | | phenolic compounds |
| | 6.61 | | | phenolic compounds |
| | 6.71 | | | phenolic compounds |
| 6.79 | 6.79−6.73 | | | phenolic compounds |
| 6.85 | 6.85 | 6.85−6.83 | | phenolic compounds |

VOOs further studies should be performed with a larger balanced data set, in which all categories will be well represented, to obtain robust models. In the present data set, Spain was clearly under-represented, being the main producer (50% of EU production of olive oil); for Italy, even though it was quite well-represented, the numbers of samples were very unbalanced with regard to the other countries and, therefore, few Italian samples were used in the modeling step, so the classification results might be very dependent on the samples in the training-test set.

Stable isotopes considerably enhanced the classification results of VOOs according to their country of origin;, $\delta^2$H and $\delta^{13}$C being among the highest weighted regression coefficients of the PLS-DA models. The results for the model for Italy provided better discrimination for both categories, whereas those of the model for Spain provided better classifications for the non-Spanish category. Higher prediction abilities for the Greek category were obtained in the cross-validation; however, the percentage of classification of the final model and the predictions on the external data set did not improve the classification notably. In contrast, the additional information provided by these isotopes enhances substantially the classification results for the Greek and Italian categories of the ternary model afforded for the three countries. Indeed, $\delta^{13}$C of olive oils already showed potential for the discrimination between olive oils from Greece, Spain, and Italy (*23*, *24*), because the mean value of $\delta^{13}$C increases in the order Italian < Greek < Spanish oils.

## ABBREVIATIONS USED

VOO, virgin and extra virgin olive oils; PDO, Protected Designation of Origin; PGI, Protected Geographical Indication; TSG, Traditional Specialty Guaranteed; NMR, nuclear magnetic resonance; FID, free induction decays; NIR, near-infrared; MIR, middle-infrared; IR, infrared; FT, Fourier transformation; GC, gas chromatography; LC, liquid chromatography; SPE, solid-phase extraction; ANOVA, analysis of variance; PCA, principal component analysis; PC, principal component; LDA, linear discriminant analysis; PLS-DA, partial least-squares discriminant analysis; PRESS, predicted error sum of squares; RMSEP, root-mean-square error of prediction; CV, cross-validation; LOO, leave-one-out cross-validation.

## ACKNOWLEDGMENT

We thank the research groups that participated in the collection of the olive oil samples: Laboratorio Arbitral Agroalimentario (Ministry of Agriculture and Fishery, Spain), General Chemical State Laboratory D'xy Athinon (Greece), General State Laboratory (Ministry of Health, Cyprus), and Departamento de Química Orgánica, Universidad de Córdoba (Spain). We acknowledge N. Segebarth for fruitful discussions.

## LITERATURE CITED

(1) Harwood, J. L.; Aparicio, R. *Handbook of Olive Oil: Analysis and Properties*; Aspen: Gaithersburg, MD, 2000; 620 pp.

(2) Aparicio, R.; Ferreiro, L.; Alonso, V. Effect of climate on the chemical composition of virgin olive oil. *Anal. Chim. Acta* **1994**, *292* (3), 235–241.

(3) Lachenmeier, D. W.; Frank, W.; Humpfer, E.; Schafer, H.; Keller, S.; Mortter, M.; Spraul, M. Quality control of beer using high-resolution nuclear magnetic resonance spectroscopy and multivariate analysis. *Eur. Food Res. Technol.* **2005**, *220* (2), 215–221.

(4) Mannina, L.; Segre, A. High resolution nuclear magnetic resonance: From chemical structure to food authenticity. *Grasas Aceites* **2002**, *53* (1), 22–33.

(5) Woodcock, T.; Downey, G.; O'Donnell, C. P. Near infrared spectral fingerprinting for confirmation of claimed PDO provenance of honey. *Food Chem.* **2009**, *114* (2), 742–746.

(6) Reid, L. M.; Woodcock, T.; O'Donnell, C. P.; Kelly, J. D.; Downey, G. Differentiation of apple juice samples on the basis of heat treatment and variety using chemometric analysis of MIR and NIR data. *Food Res. Int.* **2005**, *38* (10), 1109–1115.

(7) Baeten, V.; Pierna, J. A. F.; Dardenne, P.; Meurens, M.; Garcia-Gonzalez, 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* (16), 6201–6206.

(8) 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* (1), 25–32.

(9) Lopez-Diez, E. C.; Bianchi, G.; Goodacre, R. Rapid quantitative assessment of the adulteration of virgin olive oils with hazelnut oils using Raman spectroscopy and chemometrics. *J. Agric. Food Chem.* **2003**, *51* (21), 6145–6150.

(10) Yang, H.; Irudayaraj, J. Comparison of near-infrared, Fourier transform-infrared, and Fourier transform-Raman methods for determining olive pomace oil adulteration in extra virgin olive oil. *J. Am. Oil Chem. Soc.* **2001**, *78* (9), 889–895.

(11) Vaclavik, L.; Cajka, T.; Hrbek, V.; Hajslova, J. Ambient mass spectrometry employing direct analysis in real time (DART) ion source for olive oil quality and authenticity assessment. *Anal. Chim. Acta* **2009**, *645* (1−2), 56–63.

(12) Cajka, T.; Hajslova, J.; Pudil, F.; Riddellova, K. Traceability of honey origin based on volatiles pattern processing by artificial neural networks. *J. Chromatogr. A* **2009**, *1216* (9), 1458–1462.

(13) Stanimirova, I.; Üstün, B.; Cajka, T.; Riddellova, K.; Hajslova, J.; Buydens, L. M. C.; Walczak, B. Tracing the geographical origin of honeys based on volatile compounds profiles assessment using pattern recognition techniques. *Food Chem.* **2010**, *118*, 171−176.

(14) Vaz-Freire, L. T.; da Silva, M. D. R. G.; Freitas, A. M. C. Comprehensive two-dimensional gas chromatography for fingerprint pattern recognition in olive oils produced by two different techniques in Portuguese olive varieties Galega Vulgar, Cobrançosa e Carrasquenha. *Anal. Chim. Acta* **2009**, *633* (2), 263–270.

(15) Martins-Lopes, P.; Gomes, S.; Santos, E.; Guedes-Pinto, H. DNA markers for Portuguese olive oil fingerprinting. *J. Agric. Food Chem.* **2008**, *56* (24), 11786–11791.

(16) Ranalli, A.; Contento, S.; Marchegiani, D.; Pardi, D.; Pardi, D.; Girardi, F. Effects of "genetic store" on the composition and typicality of extra-virgin olive oil: traceability of new products. *Adv. Hortic. Sci.* **2008**, *22* (2), 110–115.

(17) Reid, L. M.; O'Donnell, C. P.; Downey, G. Recent technological advances for the determination of food authenticity. *Trends Food Sci. Technol.* **2006**, *17* (7), 344–353.

(18) Galtier, O.; Dupuy, N.; Le Dréau, 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* (1−2), 136–144.

(19) Petrakis, P. V.; Agiomyrgianaki, A.; Christophoridou, S.; Spyros, A.; Dais, P. Geographical characterization of Greek virgin olive oils (cv. Koroneiki) using $^{1}$H and $^{31}$P NMR fingerprinting with canonical discriminant analysis and classification binary trees. *J. Agric. Food Chem.* **2008**, *56* (9), 3200–3207.

(20) Rezzi, S.; Axelson, D. E.; Heberger, K.; Reniero, F.; Mariani, C.; Guillou, C. Classification of olive oils using high throughput flow H-1 NMR fingerprinting with principal component analysis, linear discriminant analysis and probabilistic neural networks. *Anal. Chim. Acta* **2005**, *552* (1−2), 13–24.

(21) Alonso-Salces, R. M.; Heberger, K.; Holland, M. V.; Moreno-Rojas, J. M.; Mariani, C.; Bellan, G.; Reniero, F.; Guillou, C. Multivariate analysis of NMR fingerprint of the unsaponifiable fraction of virgin olive oils for authentication purposes. *Food Chem.* **2010**, *118*, 956–965.

(22) Christophoridou, S.; Dais, P.; Tseng, L. I. H.; Spraul, M. Separation and identification of phenolic compounds in olive oil by coupling high-performance liquid chromatography with postcolumn solid-phase extraction to nuclear magnetic resonance spectroscopy (LC-SPE-NMR). *J. Agric. Food Chem.* **2005**, *53* (12), 4667–4679.

(23) Angerosa, F.; Breas, O.; Contento, S.; Guillou, C.; Reniero, F.; Sada, E. Application of stable isotope ratio analysis to the characterization of the geographical origin of olive oils. *J. Agric. Food Chem.* **1999**, *47* (3), 1013–1017.

(24) Royer, A.; Gerard, C.; Naulet, N.; Lees, M.; Martin, G. J. Stable isotope characterization of olive oils. I. Compositional and carbon-13 profiles of fatty acids. *J. Am. Oil Chem. Soc.* **1999**, *76* (3), 357–363.

(25) Hoffman, R. E. Standardization of chemical shifts of TMS and solvent signals in NMR solvents. *Magn. Reson. Chem.* **2006**, *44* (6), 606–616.

(26) Berrueta, L. A.; Alonso-Salces, R. M.; Héberger, K. Supervised pattern recognition in food analysis. *J. Chromatogr. A* **2007**, *1158* (1−2), 196–214.

(27) D'Imperio, M.; Mannina, L.; Capitani, D.; Bidet, O.; Rossi, E.; Bucarelli, F. M.; Quaglia, G. B.; Segre, A. NMR and statistical study of olive oils from Lazio: a geographical, ecological and agronomic characterization. *Food Chem.* **2007**, *105* (3), 1256–1267.

(28) Guillen, M. D.; Ruiz, A. High resolution H-1 nuclear magnetic resonance in the study of edible oils and fats. *Trends Food Sci. Technol.* **2001**, *12* (9), 328–338.

(29) Mannina, L.; Sobolev, A. P.; Segre, A. Olive oil as seen by NMR and chemometrics. *Spectrosc. Eur.* **2003**, *15*, (3), 6−14.

(30) Sacchi, R.; Patumi, M.; Fontanazza, G.; Barone, P.; Fiordiponti, P.; Mannina, L.; Rossi, E.; Segre, A. L. A high-field H-1 nuclear magnetic resonance study of the minor components in virgin olive oils. *J. Am. Oil Chem. Soc.* **1996**, *73* (6), 747–758.

(31) Owen, R. W.; Haubner, R.; Mier, W.; Giacosa, A.; Hull, W. E.; Spiegelhalder, B.; Bartsch, H. Isolation, structure elucidation and antioxidant potential of the major phenolic and flavonoid compounds in brined olive drupes. *Food Chem. Toxicol.* **2003**, *41* (5), 703–717.

(32) Esbensen, K. H.; Guyot, D.; Westad, F.; Houmøller, L. P. *Multivariate Data Analysis in Practice: An Introduction to Multivariate Data Analysis and Experimental Design*, 5th ed.; Camo Process: Oslo, Norway, 2002; 598 pp.

(33) Camin, F.; Larcher, R.; Perini, M.; Bontempo, L.; Bertoldi, D.; Gagliano, G.; Nicolini, G.; Versini, G. Characterisation of authentic Italian extra-virgin olive oils by stable isotope ratios of C, O and H and mineral composition. *Food Chem.* **2010**, *118*, 901−909.

**Received for review November 13, 2009. Revised manuscript received March 12, 2010. Accepted March 13, 2010. We thank the EU TRACE project for funding this work. The information contained in this paper reflects the authors' views; the European Commission is not liable for any use of the information contained herein.**