Research Article

^{1}H-NMR and isotopic fingerprinting of olive oil and its unsaponifiable fraction: Geographical origin of virgin olive oils by pattern recognition

Rosa M. Alonso-Salces[1,2,3], Nicolas Segebarth[2], Sergio Garmón-Lobato[4], Margaret V. Holland[2], Jose M. Moreno-Rojas[2†], Juan A. Fernández-Pierna[5], Vincent Baeten[5], Sandra R. Fuselli[3,6], Blanca Gallo[4], Luis Angel Berrueta[4], Fabiano Reniero[2], Claude Guillou[2] and Károly Héberger[7]

[1] Consejo Nacional de Investigaciones Científicas y Técnicas, Buenos Aires, Argentina
[2] Institute of Health and Consumer Protection, DG–Joint Research Centre, European Commission, Ispra, Italy
[3] Departamento de Biología, Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Mar del Plata, Mar del Plata, Argentina
[4] Departamento de Química Analítica, Facultad de Ciencia y Tecnología, Universidad del País Vasco/Euskal Herriko Unibertsitatea, Leioa, Spain
[5] Valorisation of Agricultural Products Department, Walloon Agricultural Research Centre, Gembloux, Belgium
[6] Comisión de Investigaciones Científicas de la Provincia de Buenos Aires, La Plata, Argentina
[7] Research Centre for Natural Sciences, Hungarian Academy of Sciences, Budapest, Hungary

^{1}H-NMR spectral data and H and C isotope abundances of virgin olive oils (VOOs) and their unsaponifiable fractions were analyzed by pattern recognition techniques, such as principal component analysis (PCA) and partial-least squares discriminant analysis (PLS-DA). The aim was to develop chemical tools for the authentication of VOOs according to their geographical origin or protected designation of origin (PDO), as well as to detect the mislabeling of the provenance of VOOs, at the regional or national level, or the mislabeling of non-PDO oils as PDO VOOs. The relationship between stable isotope abundances of the VOOs and their unsaponifiable fractions and the latitude of the VOO geographical origin was confirmed, but these criteria were not completely discriminant to differentiate VOOs according to their geographical origin. However, δ^{2}H and/or δ^{13}C data provided complementary geographical information to ^{1}H-NMR data in the PLS-DA binary classification models afforded for VOOs from Greece, Spain, Italy, Izmir (Turkey), Crete (Greece), and the PDOs *Riviera Ligure* (Italy) and *Huile d'olive d'Aix–en–Provence* (France). ^{2}H/^{1}H and ^{13}C/^{12}C ratios of the unsaponifiable fractions of VOOs are reported here for the first time. The present approach for PDO *Riviera Ligure* VOOs, based on ^{1}H-NMR data and C isotope abundance of the bulk oil and its unsaponifiable fraction, outperformed the previously reported classification models. Moreover, the PLS-DA models to authenticate VOOs from Greece and detect non-Greek VOOs achieved over 93% of correct predictions.

Practical applications: The research can be applied in the protection of consumers and honest producers and retailers, and provides potential tools for antifraud authorities and regulatory bodies, which face the challenge of detecting fraudulent practices that do not comply with EU regulations in the trade of VOOs, such as the mislabeling of VOOs produced in a certain geographical origin [Commission Implementing Regulation (EC) no 29/2012 and Commission Implementing Regulation (EC) no 1335/2013] and/or under specific EU quality schemes, named PDO or PGI [Council Regulation (EC) no 510/2006].

Keywords: Multivariate data analysis / NMR / IRMS / Olive oil / Unsaponifiable

Received: May 23, 2014 / Revised: March 12, 2015 / Accepted: March 24, 2015

DOI: 10.1002/ejlt.201400243

Additional supporting information online http://dx.doi.org/10.1002/ejlt.201400243

Correspondence: Dr. Rosa M. Alonso-Salces, Instituto de Investigaciones en Biodiversidad y Biotecnología (INBIOTEC) – CONICET, Vieytes 3103, (7600) Mar del Plata, Argentina
E-mail: rosamaria.alonsosalces@gmail.com
Fax: (+54) 223 4757120

[†]Present address: Instituto de Investigación y Formación Agraria y Pesquera, Área de Tecnología Postcosecha e Industria Agroalimentaria, Córdoba, Spain.

1 Introduction

Virgin olive oil (VOO) is a high added value agricultural product in the European Union (EU) from commercial, nutritional and health-promoting potential points of view. The characterization of the geographical origin of VOO is becoming increasingly important. VOOs are permitted to be marketed under specific EU agricultural product quality schemes, named protected designation of origin (PDO), protected geographical indication (PGI), or traditional specialty guaranteed (TSG) label [Council Regulations (EC) nos. 509/2006 and 510/2006]. The European Commission has already registered 105 PDO and 12 PGI VOOs, produced in Spain, Italy, Greece, France, Portugal, and Slovenia, in the "Register of protected designations of origin and protected geographical indications" (DOOR database). PDO covers VOOs which are produced, processed, and prepared in a given geographical area using recognised know-how. Whereas PGI covers VOOs closely linked to the geographical area; at least one of the stages of production, processing, or preparation takes place in the area. These quality labels promote and protect names of quality agricultural products and foodstuffs; therefore, given the financial benefits associated with these prestigious labels, the likelihood that economic fraud occurs (e.g., false claims of geographical origin on product labels, labeling a non-PDO/PGI product as a PDO/PGI one or adulteration with olive oils that do not fulfil the PDO/PGI requirements) is high. Another fraudulent practice is the mislabeling of the designation of origin of olive oils. The EU has established new labeling rules that make origin labeling compulsory for virgin and extra virgin labeled olive oils [Commission Implementing Regulation (EC) no. 29/2012 and Commission Implementing Regulation (EC) no. 1335/2013]. The designation of origin must refer not only to the olives used but also to the geographical area in which the oil was extracted from the olives; if this is not the same as that where the olives were harvested, this information should be stated on the label. Moreover, oil produced from olives from just one EU Member State or third country has to be labeled with the name of the country of origin. VOO produced from olives from more than one EU Member State has to be labeled as a "blend of olive oils of European Union origin" (or a reference to the Union); while oil produced using olives from outside the EU would be labeled as a "blend of olive oils not of European Union origin" (or a reference to origin outside the Union) or "blend of olive oils of European Union origin and not of European Union origin" (or a reference to origin within the Union and outside the Union). Therefore, analytical methods are urgently needed to guarantee the authenticity and traceability of PDO/PGI olive oils, as well as their country of provenance, to help prevent illicit practices in this sector, and to support the antifraud authorities dealing with these issues.

VOO is characterized by containing fatty aids mostly as triglycerides, a high concentration of oleic acid, and a low concentration of saturated fatty acids in position *sn*-2 [1, 2]. Among the triglycerides, the major ones are triolein (43.5%), 1-palmityl-2,3-diolein (18.4%) and 1-linoleyl-2,3-diolein (6.8%). The unsaponifiable fraction of VOO, which represents 1–2% of the oil, is made up of different minor compounds. Hydrocarbons may be constituted up to 0.7%, mainly squalene, and low quantities of epoxy-squalene isomers and alkanes (C16-C35). Phytosterols make up the main part of the unsaponifiable fraction of olive oil: β-sitosterol is the most abundant, followed by Δ5-avenasterol, and then by campesterol and stigmasterol [1]. Regarding the tracking of commercial fraud, the sterol fraction has many applications, especially where the contamination of some vegetable oils with other cheaper ones is concerned [3]. Of the tocopherols, α-tocopherol comprises about 90% of the total tocopherol fraction. The major phenolic compounds identified and quantified in olive oil belong to three different classes: simple phenols (hydroxytyrosol, tyrosol), secoiridoids, and lignans [1]. Other constituents of the unsaponifiable matter are carotenoids (β-carotene being the most important), chlorophylls, and pheophytins [1]. The alcohol fraction of VOO include aliphatic alcohols, mainly docosanol, tetracosanol, hexacosanol, and octacosanol, and at trace levels, tricosanol, pentacosanol, and heptacosanol. In smaller quantities, triterpenic alcohols (cycloartenol, 24-methylen-cycloartenol, and α- and β-amirines), diterpenic alcohols (fitol and geranylgeraniol), and triterpenic dialcohols (erythrodiol and uvaol) are also present [1]. The composition of the unsaponifiable fraction of VOO is affected by several factors such as olive cultivar, altitude, climatology, agronomical factors, time of harvest, olive storage after harvest, and oil extraction system [1]. The diversity and interrelation between all these factors is reflected in the chemical composition of VOO, and it is highly unlikely that this influence would be the same in different regions. So, the geographical characterization of VOO is closely linked to all these agronomic, pedoclimatic, and botanical parameters that characterize the olive oil of each origin [4]. Therefore, it can be expected that the unsaponifiable fraction of VOOs may contain information, which can be useful for the geographical characterization of olive oils.

A considerable number of sensorial, physical, and chemical approaches combined with statistical analysis have been used to distinguish olive oils of different types, botanical and/or geographical origins, and pedoclimatic conditions [5].

Abbreviations: AF, alcohol fraction; **CV,** cross-validation; **FID,** free induction decays; **HF,** hydrocarbon fraction; **IRMS,** isotope ratio mass spectrometry; **LOO-CV,** leave-one-out cross-validation; **PC,** principal component; **PCA,** principal component analysis; **PDO,** protected designation of origin; **PGI,** protected geographical indication; **PLS-DA,** partial least squares discriminant analysis; **PRESS,** predicted error sum of squares; **RMSEP,** root mean square error of prediction; **SF,** sterol fraction; **TF,** tocopherol fraction; **TSG,** traditional specialty guaranteed; **VOO,** virgin olive oil

For this purpose, fatty acids [6], triglycerides [7], sterols [6], phenolic compounds [8, 9], aldehydes [10], volatiles [9], and pigments [11] have been analyzed by conventional methods [12] that usually require time-consuming pre-treatment methods (solvent extraction, isolation and/or derivatization) followed by chromatographic techniques [3, 13] such as GC-MS and/or GC-FID [14, 15] and HPLC-MS [2, 16].

Fingerprinting techniques, such as nuclear magnetic resonance (NMR), fluorescence, infrared (NIR, FT-IR, FT-MIR), and Raman spectroscopies [17], are particularly attractive since they are non selective, require little or no sample pre-treatment; use small amounts of organic solvents or reagents; and the analysis takes only a few minutes per sample [5]. NIR spectra of VOOs together with chemometric tools allowed to determine their composition and geographical origin [18]. ^{1}H, ^{13}C and/or ^{31}P-NMR analysis of the bulk oil [19–25] or the unsaponifiable fraction of olive oil [19], in combination with multivariate techniques, have been used to distinguish VOOs according to their geographical origin, as well as to detect adulteration of the oil [26]. Mass spectrometry fingerprinting of the volatile profiles of VOO, combined with the National Institute of Standards and Technology Mass Spectral search algorithm for pattern recognition, allowed us to trace the geographical origin of VOOs [27]. IRMS methods have also been used for the authentication of olive oil by analyzing the bulk oil [28, 29]. Isotopic measurements of alcoholic and sterolic fractions of olive oil also proved to be useful for its geographical characterization [30].

In the present work, ^{1}H-NMR and isotopic fingerprinting of VOOs and their corresponding unsaponifiable fractions, in combination with pattern recognition techniques were used to develop protocols to detect the mislabeling of the provenance of VOOs at the regional or national level, or the mislabeling of non-PDO oils as PDO VOOs. In particular, the following approaches for the discrimination of the geographical origin of VOOs were assessed: i) ^{1}H-NMR analysis of the bulk oil, ii) ^{1}H-NMR and $\delta^{13}C$ and $\delta^{2}H$–IRMS analysis of the bulk oil, *iii*) ^{1}H-NMR analysis of the unsaponifiable fraction of the olive oil, iv) ^{1}H-NMR and $\delta^{13}C$ and $\delta^{2}H$–IRMS analysis of the unsaponifiable fraction of the olive oil, v) ^{1}H-NMR analysis of the bulk oil and the unsaponifiable fraction of the olive oil, and vi) ^{1}H-NMR and $\delta^{13}C$ and $\delta^{2}H$–IRMS analysis of the bulk oil and the unsaponifiable fraction of the olive oil. Approaches i, ii, iii, and v had been previously studied separately [19, 20], showing their potential for the aimed purpose. In the present study, all the above-mentioned approaches (i–vi) were evaluated on the same batch of VOOs by multivariate data analysis with the scope of tracing their geographical origin. ^{1}H-NMR spectra of the bulk oil and its corresponding unsaponifiable fraction, as well as the sub-fractions of the unsaponifiable fraction (alcohol, sterol, hydrocarbon, and tocopherol fractions) were studied in the search for the markers that multivariate techniques revealed to be related to the geographical origin of olive oils.

2 Materials and methods

2.1 Chemicals

Deuterated chloroform for NMR analysis (99.8 atom %D), chloroform (p.a.), 1-eicosanol, α-tocopherol, β-sitosterol, stigmasterol, campesterol, and silica gel on TLC plates were provided by Sigma–Aldrich Chemie (Steinheim, Germany), potassium hydroxide (p.a.), anhydrous sodium sulphate (p.a.), hexane (p.a.), and 2′,7′-dichlorofluorescein (TLC grade) by Merck (Darmstadt, Germany), diethyl ether (HPLC grade) by Fluka Chemie (Buchs, Switzerland), methanol (HPLC grade) by Carlo Erba (Rodano, Italy), and erythrodiol by Extrasynthèse (Genay, France). Cycloartenol standard was prepared by extracting the unsaponifiable fraction from flax oil and performing a further purification of the extract by thin layer chromatography (TLC) as described by Alonso–Salces et al. [19].

2.2 Samples

Virgin olive oils (VOOs, 125 samples) from six countries of the Mediterranean basin, namely Italy (29 VOOs), Spain (29 VOOs), Greece (29 VOOs), France (18 VOOs), Turkey (14 VOOs), and Cyprus (6 VOOs), were collected directly from the producers (olive oil mills) at the main producing regions of these countries during two harvests (2004/05 and 2005/06). The sample collection was carried out in the framework of the EU TRACE project (Food Quality and Safety Priority of the Sixth Framework Programe). The true type (virgin or extra virgin) and origin of the olive oils at the national, regional, and PDO level were assured. VOOs produced in the Mediterranean basin are usually defined as multi-varietal because of the presence of several olive cultivars in the same olive grove; from 3 or 4 different varieties to as many as 70, depending on the PDO or production area. Sampling for the present study was planned so as to cover the maximum variability related to the harvests, olive varieties, and production areas. VOOs under the PDOs *Riviera Ligure* and *Huile d'olive d'Aix–en–Provence* were included: Under these PDOs, only extra virgin olive oils produced in Liguria (Italy) and in the French departments of Bouches–du–Rhône and Var, respectively, that fulfil the PDO requirements related to olive varieties, farming practices, oil extraction procedures, bottling, and labeling (PDO *Riviera Ligure*: Dossier Number: IT/PDO/0017/1540, Official Journal of the European Communities 1997, L22; PDO *Huile d'olive d'Aix–en–Provence*: Dossier number: FR/PDO/0005/0111, Official Journal of the European Communities 2000, C297) can be marketed. All the VOO samples were kept frozen (−20°C) until the analysis by NMR and

IRMS analysis of the VOO bulk oil, or the extraction of the VOO unsaponifiable fraction prior to be submitted to NMR and IRMS analysis.

2.3 Extraction of unsaponifiable fraction of olive oil

For the extraction of the unsaponifiable matter, a modification of the humid process (i.e., extraction from an aqueous or alcoholic solution of the soap) recommended by the European Union for olive oil analysis was used [31]. The sample of olive oil is dried under a nitrogen flow and filtered. Then, 50 mL of methanolic potassium hydroxide 2 N is added to an aliquot of 5 g of the dried and filtered olive oil, and heated to a gentle boil in a water bath with continuous vigorous stirring under reflux for 1 h. Then the content is transferred quantitatively into a 500 mL funnel using several rinses of distilled water (about 100 mL), and three successive extractions with ethyl ether (80 mL) are performed. The ether phase is washed with distilled water until the wash water reaches a neutral pH. Once the water is removed, the extract is dried with anhydrous sodium sulphate for 30 min, filtered and the solvent removed under a nitrogen stream to dryness. The unsaponifiable matter of each sample was prepared in duplicate and stored frozen (−20°C) until analysis. The repeatability of the method was evaluated by extracting separately six aliquots of a VOO in-house standard.

Subfractions of the unsaponifiable matter, i.e., alcohol fraction (AF), sterol fraction (SF), tocopherol fraction (TF), and hydrocarbon fraction (HF), of two VOOs of two different origins (Italy and Turkey) were prepared as reported by Alonso–Salces et al. [19], and were also analyzed by ^{1}H-NMR.

2.4 ^{1}H-NMR analysis

Each unsaponifiable fraction of VOO or 40 μL of the bulk oil was dissolved in 200 μL of deuterated chloroform, shaken in a vortex, and placed in a 2 mm NMR capillary. The ^{1}H-NMR experiments were performed on a Bruker (Rheinstetten, Germany) Avance 500 (nominal frequency 500.13 MHz) equipped with a 2.5 mm broadband inverse probe. The bulk oil spectra were recorded at 300 K using a 7.5 μs pulse (90° flip angle), an acquisition time of 3.0 s (32k data points) and a total recycling time of 4.0 s, a spectral width of 5500 Hz (11 ppm), 64 scans (+4 dummy scans), with no sample rotation, and *qsim* acquisition mode. The unsaponifiable fraction spectra were recorded at 298 K using a 7.5 μs pulse (spin-echo pulse sequence, with a 1 ms echo time), an acquisition time of 3.5 s (50 k data points) and a total recycling time of 4.5 s, a spectral width of 7122.5 Hz (14 ppm), 32 scans (+4 dummy scans), without sample rotation, and *DQD* acquisition mode. Prior to Fourier transformation, the free induction decays (FIDs) were zero-filled to 64 k and a 0.3 Hz line-broadening factor was applied. The chemical shifts are expressed in δ scale (ppm), referenced to the residual signal of chloroform (7.26 ppm) [32]. The interesting regions of the NMR spectra are 0–7 ppm for the bulk oil, and 0–10.1 ppm for the unsaponifiable fraction. The spectra were phase- and baseline-corrected manually. 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) for the bulk oil, and 0–10.1 ppm for the unsaponifiable fraction. 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 the application of pattern recognition tools. Eight buckets in the region 4.10–4.26 ppm (reference region) of the bulk oil spectra were excluded in the multivariate data analysis.

2.5 $\delta^{2}$and δ^{13}C IRMS analysis

δ^{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 (Thermo-Fisher, Rodano, Italy). The δ^{2}H signal for the reference peak was 7000 mV; GC column temperature, 80°C; and the glassy–carbon column reactor temperature, 1450°C. δ^{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 Delta Plus mass spectrometer (Thermo Fisher, Rodano, Italy). The δ^{13}C signal for the reference peak was 4000 mV; the oxidation column temperature, 1050°C; the reduction column temperature, 650°C; and GC column temperature, 65°C.

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

$$\delta(\%) = \left[\left(R_{\text{sample}}/R_{\text{reference}}\right) - 1\right] \times 1000.$$

where R is the ratio of the heavy to light stable isotope (e.g., ^{2}H/^{1}H) in the sample (R_{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) IAEA-CH7-Polyethylene ($\delta^{2}H = -100.3\%$) and NBS22-Oil ($\delta^{2}H = -120.0\%$) for H_2 gas cylinder calibration, and ii) IAEA−CH7−Polyethylene ($\delta^{13}C = -32.15\%$) and IAEACH6-Sucrose ($\delta^{13}C = -10.4\%$) for CO_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 that the repeatability of the measurements was acceptable, and to correct for possible

drifts in the measurements. The relative standard deviations ($n = 10$) determined using the corresponding reference gas were 0.8% for δ^2H and 0.05% for δ^{13}C. Each bulk oil and unsaponifiable fraction was analyzed in duplicate, the standard deviations being <2.7% for δ^2H and <0.15% for δ^{13}C.

2.6 Multivariate data analysis

Datasets were made up of the NMR buckets and isotopic δ values (variables in columns) measured on the bulk oils and/or the unsaponifiable fractions of the VOOs analyzed (samples in rows). Firstly, datasets were analyzed by univariate procedures (ANOVA, Fisher index and Box–Whisker plots), and afterwards, by unsupervised and supervised multivariate techniques, such as principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA), respectively [33]. ^{1}H-NMR data of the bulk oil and the unsaponifiable fraction of VOOs were included in several datasets according to the approach studied. Two types of datasets were considered for multivariate data analysis: *a*) a dataset constituted of the original variables, i.e., NMR buckets; and b) a dataset made up with the scores of the principal components with eigenvalues higher than 1, afforded by performing PCA on the original NMR datasets of the bulk oil and the unsaponifiable fraction separately (PCA score matrices). Data analysis was performed by means of the statistical software packages Statistica 6.1 (StatSoft Inc., Tulsa, OK, 1984–2004) and The Unscrambler 9.1 (Camo Process AS, Oslo, Norway, 1986–2004).

The multivariate techniques (PCA and PLS-DA) were applied to the autoscaled (or standarized) data matrix of original variables. PLS-DA was also carried out on the datasets made up with the PCA score matrices. In PLS-DA, PRESS or RMSEP are plotted against the number of PLS components in order to find the optimal number of the latter. 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 predictions in the training-test set are represented in a box and whisker plot in order to define the half of the distance between the quartiles as the boundary.

The datasets were divided into a training set and a test set several times in order to perform cross-validation. The optimization of parameters characteristic of the multivariate technique, i.e., the number of PLS components in PLS-DA, were carried out by cross-validation. The final mathematical models were built using all the samples of the training-test set and the optimized parameters. During the parameter optimization step, the models were validated by threefold cross-validation (threefold CV) and/or leave-one-out cross-validation (LOO-CV). Binary classification models can lead to artefacts if they are not used and validated properly [34]. In order not to have large imbalances in the number of samples of each category, the training-test set used to build binary classification models contained such a number of samples that the sample ratio *class-1*/*class-2* was 1:2; *class-1* was made up with the samples of the origin studied (code 1), and *class-2*, with the samples of the other origins (code 0). PLS-DA is not so sensitive to imbalances in the dataset as other multivariate techniques are [20]; so PLS-DA performed properly using this sample ratio. The total number of samples of each geographical origin or PDO was not large enough to separate samples for a complete external validation. A pseudo-external set was built with the remaining samples of *class-2*, since all samples of *class-1* were required in the training-test set to include all the variability of this class and have a representative sample set. The reliability of the classification models achieved was studied in terms of recognition ability (percentage of the samples in the training set correctly classified during the modeling step), prediction ability in the cross−validation (percentage of the samples in the test set correctly classified by using the model developed in the training step), and the prediction ability in the external validation (percentage of the samples of the external set correctly classified by using the optimized model) [33].

3 Results and discussion

3.1 ^{1}H-NMR spectra of olive oil and its unsaponifiable fraction

^{1}H-NMR fingerprints of 125 VOOs and their unsaponifiable fractions were recorded. In the present study, the ^{1}H-NMR method for the analysis of the unsaponifiable fraction of olive oil was improved in order to overcome the problems encountered previously with regard to the useable spectral region of the unsaponifiable fractions, which was limited to the region 0–5.44 ppm [19]. Whereas the bulk oil spectra were acquired with a classical 90° pulse sequence, a spin-echo sequence was selected for the unsaponifiable fraction of olive oil as it allowed for the suppression of undesirable signals most probably coming from incomplete water removal. The NMR method for the unsaponifiable fractions enabled the use of the whole spectral region, providing all the information contained at the chemical shifts higher than 5.44 ppm, where characteristic ^{1}H signals of aldehydes and phenolic compounds are located [35–37].

The chemical shifts of the ^{1}H signals of the triglycerides are well-known [38]. Minor oil components are only observed by ^{1}H-NMR when their signals are not overlapped by those of the main components and their concentrations are high enough to be detected [19, 39]. Signals of major and some minor compounds typically observed in the ^{1}H-NMR spectra of VOO [19, 24, 36, 39, 40], together with their chemical shifts and their assignments to protons of the

different functional groups are gathered in Table S1 (supplementary information). Minor compound signals, not overlapped by those of the triglyceryl protons, were from cycloartenol at 0.318 and 0.543 ppm, β-sitosterol at 0.669 ppm, stigmasterol at 0.687 ppm, squalene at 1.662 ppm, *sn*-1,2 diglyceryl group protons at 3.71 and 5.10 ppm, *sn*-1,3 diglyceryl group protons at 4.05 ppm, three unknown terpenes at 4.571, 4.648, and 4.699 ppm, and phenolic protons at 5.73, 5.99, 6.55, and 6.75 ppm.

The unsaponifiable constituents of VOOs are mainly sterols, tocopherols, aliphatic alcohols, hydrocarbons, fatty acids, pigments and phenolic compounds. The complete ^{1}H-NMR spectra of the unsaponifiable fraction of VOOs in the region from 0 to 10 ppm have not been previously reported. ^{1}H-NMR analysis of the unsaponifiable matter provides useful information on minor compounds, whose signals are masked by the triglyceride ones in the ^{1}H-NMR spectra of the bulk oil. The ^{1}H-NMR signals of the unsaponifiable fraction, together with their corresponding chemical shifts are listed in Table S2 (supplementary information). Besides the information provided in the references [35–37, 39, 41], the ^{1}H-NMR analysis of the sub-fractions of the unsaponifiable fraction, i.e., alcohol fraction (AF), sterol fraction (SF), tocopherol fraction (TF), and hydrocarbon fraction (HF), and of the standards available allowed us to carry out proton signal assignments by comparing their spectra. The proton signals in the spectra of the unsaponifiable fraction were slightly shifted with respect to the signals in the spectra of the bulk oil (0.013–0.015 ppm to higher chemical shifts). Some proton signals corresponded to a particular sub-fraction of the unsaponifiable matter: AF (*0.141 ppm, 0.333 ppm, 0.38–0.40 ppm, 0.55–0.57 ppm, 0.558 ppm, 0.615 ppm, 0.715 ppm, 0.747 ppm, 0.974 ppm, 1.01–1.04 ppm, 3.10–3.17 ppm, 3.26–3.33 ppm, 3.284 ppm, 3.293 ppm, 3.315 ppm, 3.641 ppm, 3.723 ppm, 4.157 ppm, 4.310 ppm, 4.65–4.76 ppm, 5.256 ppm, 7.05 ppm, 9.401 ppm*), SF (*0.529 ppm, 0.556 ppm, 0.683 ppm, 0.702 ppm, 0.77–0.78 ppm, 0.826 ppm, 0.834 ppm, 0.848 ppm, 0.91–0.94 ppm, 0.921 ppm, 0.934 ppm, 1.009 ppm, 1.80–1.88 ppm, 2.27–2.31 ppm, 2.336 ppm, 2.470 ppm, 3.49–3.58 ppm, 4.771 ppm, 4.821 ppm, 4.964 ppm, 5.30–5.43 ppm, 6.92 ppm*), TF (*0.156 ppm, 0.881 ppm, 0.894 ppm, 2.131 ppm, 2.603 ppm, 3.302 ppm, 3.330 ppm, 3.416 ppm, 3.433 ppm, 4.099 ppm, 4.840 ppm, 4.875 ppm, 4.836 ppm, 4.872 ppm, 4.938 ppm, 5.53–5.63 ppm, 5.988 ppm, 6.367 ppm*) and HF (*0.884 ppm, 1.431 ppm, 2.269 ppm, 3.287 ppm, 3.662 ppm, 4.060 ppm, 4.539 ppm, 4.690 ppm, 4.98–5.01 ppm, 5.020 ppm, 5.703 ppm, 5.745 ppm, 5.87–5.98 ppm, 6.439 ppm, 6.975 ppm, 9.365 ppm, 9.37–9.41 ppm, 9.762 ppm*). Other proton signals were assigned to a certain family of compounds, such as terpenes (*4.609 ppm, 4.648 ppm, 4.694 ppm*), phenolic compounds (*3.487 ppm, 3.855 ppm, 3.950 ppm, 4.125 ppm, 4.55–4.61 ppm, 5.851 ppm, 6.515 ppm, 6.559 ppm, 6.89 ppm, 6.95 ppm, 7.03 ppm, 7.09–7.21 ppm, 7.165 ppm, 7.185 ppm, 7.32 ppm, 7.53 ppm, 7.69–7.73 ppm, 7.81 ppm, 7.84 ppm*), or aldehydes (*6.02–6.18 ppm, 8.025 ppm, 9.383 ppm, 9.539 ppm, 9.739–9.755 ppm, 9.581 ppm, 9.762 ppm, 9.845 ppm, 9.849 ppm, 9.853 ppm, 9.875 ppm, 9.962 ppm, 9.999 ppm*); or to individual compounds, e.g. cycloartenol (*0.333 ppm, 0.558 ppm, and 0.974 ppm*), β-sitosterol (*0.826 ppm, 0.834 ppm, 0.848 ppm, 0.921 ppm, and 0.934 ppm*), stigmasterol (*0.702 ppm*), and squalene (*1.431 ppm*).

3.2 Isotope contents of olive oil and its unsaponifiable fraction

The isotopic ratio ^{2}H/^{1}H (δ^{2}H) and ^{13}C/^{12}C (δ^{13}C) of the bulk oil and the unsaponifiable fraction of the studied VOOs were measured (Table S3 in the supplementary information). The isotope abundance of ^{2}H/^{1}H and ^{13}C/^{12}C in the unsaponifiable fractions of VOOs are here reported for the first time. The δ^{13}C values for both the bulk oil and the unsaponifiable fractions were in the range −31.4 to −27.1‰ and −30.5 to −26.7‰, respectively, consistent with values described for plants following the C3 (Calvin) biosynthetic pathway. The primary source of hydrogen for any organic compound in the biosphere is water and, for plant biomass, water in the leaves. The actual ^{2}H/^{1}H ratio is governed by geographical, climatic and physiological factors, including altitude, latitude, precipitation and plant evapotranspiration. This is the starting point for isotope discrimination during metabolism [42]. The δ^{2}H values measured in the bulk oil (−164.0 to −126.7‰) were in agreement with those previously reported for fats from plants following the C3 cycle: −120 to −196‰ [43]. The δ^{2}H values of the unsaponifiable fractions of VOOs ranged from −178.5 to −143.5‰.

As the isotopic fractionation of C and H is linked to pedoclimatic factors (soil, climate, and latitude) and that of C, also to olive variety [29]; δ^{2}H and δ^{13}C data can contribute to the geographical discrimination of olive oils [20]. As a matter of fact, δ^{2}H and δ^{13}C in olive oils increased according to the olive cultivation latitude, from the North to the South of Italy [29, 30, 44]. We also observed this trend regarding δ^{2}H and δ^{13}C data in VOOs (Fig. S1 in the supplementary information), as well as in δ^{13}C values in the unsaponifiable fractions. The opposite tendency is observed in the δ^{2}H of the VOOs from Liguria–Umbria (central–North) and Molise–Campania (central–South) (Fig. S1). Andalusian VOO isotope abundance were similar or higher than in Sicilian VOOs, as already reported by Armendia et al. [29]. This trend of the increasing δ^{2}H and δ^{13}C with latitude was also observed in Spanish VOOs from Andalucía and the other Spanish regions which are more Northern; as well as in Greek VOOs from Peloponnesus (North) and Crete (South) (Fig. S2 in the supplementary information). At the country level, Cyprus, being at the southern latitude, presented the highest isotope abundances (Fig. S3 in the supplementary information). Despite all these observations, δ^{2}H and δ^{13}C data was not completely discriminant among

VOOs from the geographical origins studied, i.e., H and C isotope abundances cannot differentiate VOOs according to their geographical origin, neither at the regional nor at the country level, by themselves.

3.3 Geographical origin of virgin olive oil

Data analysis of the ^{1}H-NMR spectral data and the δ^{2}H and δ^{13}C measurements of the bulk oil and the unsaponifiable fraction of VOOs was performed in order to develop chemical tools for the authentication of VOOs according to their geographical origin or PDO, and the detection of the mislabeling of VOO provenance or non-PDO oils passed off as PDO VOOs. The univariate analysis (ANOVA, Fisher test, box, and whiskers plots) of the ^{1}H-NMR datasets of the bulk oils and the unsaponifiable fractions disclosed that none of the ^{1}H-NMR variables (buckets) was able to discriminate between VOOs of the different geographical origins by itself. Hence, it was necessary to move on to multivariate data analysis in order to achieve the discrimination required.

The presence of outliers in the bulk oil and unsaponifiable fraction datasets were analyzed by PCA, and extreme samples were removed after having noticed the presence of some irregularities in their ^{1}H-NMR spectra. Then, multivariate data analysis was performed on the autoscaled final datasets (bulk oil dataset: 112×342 matrix, and unsaponifiable fraction dataset: 112×505 matrix). When PCA was carried out on the bulk oil dataset, bidimensional plots of the sample scores in the spaces defined by the four first principal components [accounting for 66% of total system variability: 30.7% (PC1), 13.8% (PC2), 13.5% (PC3), 7.9% (PC4)] did not show any clustering of the samples according to the geographical origin. Samples were distributed in a compact cluster, even though some sub-groupings, partially overlapped, were observed along PC3 according to the harvest year. The most influential variables on PC3, and therefore related to the year, were NMR buckets centered at the following chemical shifts: 1.35–1.47 ppm; 1.65 ppm, due to the β-methylene protons of the acyl groups; 2.07–2.11 ppm, to the allylic protons of the acyl group; 0.91–0.93 ppm, to the methylic proton of the saturated, oleic and linoleic acyl groups; 4.27, 4.33 ppm, to α-methylene protons of the glyceryl group of triglycerides; 2.33–2.37 ppm; 2.29 ppm, to the α-methylene protons of the acyl groups; and 5.25 ppm, to glyceryl protons of triglycerides. Since the samples were all from 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 in order to include the seasonal variability, as well. The direction of maximum variability in the data set did not correspond to the direction of maximum discrimination among the geographical origins. This suggests the presence of other sources of variability; indeed, the year of harvest was confirmed to be one of them as mentioned above. PCA on the VOO unsaponifiable fraction dataset did not show any clustering of the samples due to either the geographical origin or the harvest year.

In order to extract the useful information contained in the ^{1}H-NMR data and the δ^{2}H and δ^{13}C measurements for the geographical origin characterization of VOOs, binary PLS-DA classification models were develop using the following combination of the data available: i) ^{1}H-NMR data of the bulk oil, ii) ^{1}H-NMR, δ^{13}C and δ^{2}H data of the bulk oil, iii) ^{1}H-NMR data of the VOO unsaponifiable fraction, iv) ^{1}H-NMR, δ^{13}C and δ^{2}H data of the VOO unsaponifiable fraction, v) ^{1}H-NMR data of the bulk oil and the VOO unsaponifiable fraction, and vi) ^{1}H-NMR, δ^{13}C and δ^{2}H data of the bulk oil and the VOO unsaponifiable fraction.

3.3.1 Country of origin of VOOs

Most of the PLS-DA models achieved for the case study Greece versus non-Greece VOOs performed similarly, with recognition abilities close to 100% and prediction abilities over 89% in the cross-validation for both classes, and over 82% in the external validation for the non-Greece category (Table 1). Regarding the authentication of Greek VOOs, the best model was obtained with the dataset containing ^{1}H-NMR, δ^{13}C and δ^{2}H data of the bulk oil (OO_NMR-IRMS dataset), since it predicted satisfactory 96.4% of the Greek samples. However, in order to detect fraudulent non-Greek VOOs, ^{1}H-NMR data of the unsaponifiable fraction provided useful information to that of the bulk oil (OO-Unsap_PC_NMR dataset), affording a model that predicted correctly 98.2% of the non-Greek samples in cross-validation and 92.9% in the external validation. This model improved the results previously reported for this case study [20]. These classification results together with the facts that the recognition ability was higher but close to the prediction ability in the cross-validation, and this was higher but close to prediction ability in the external validation, disclosed that the models achieved were feasible and not random, as well as well-represented by the samples in the dataset. The regression coefficients 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 [45]. The variables that presented the highest regression coefficients in the PLS-DA model obtained with OO_NMR-IRMS dataset were: 0.97 ppm, belonging to the methylic proton of the linolenic acyl group; 5.71–5.77 ppm; 4.65 ppm, to terpenes; 0.31–0.33 ppm, to the methylene proton of the cyclopropanoic ring of cycloartenol; δ^{13}C; 4.69–4.71 ppm, to terpenes; 4.05 ppm, to glyceryl protons of *sn*-1,3-diglycerides; 3.61 ppm; 6.45–6.47 ppm and 6.75 ppm, to phenolic compounds; and δ^{2}H. The highest regression coefficients for the PLS-DA model obtained with the OO-Unsap_PC_NMR dataset were of PC9 of the bulk oil dataset and PC15 of the unsaponifiable fraction dataset. The most influential variables in the PCs

Table 1. PLS-DA models for the geographical discrimination of VOOs: Greece versus non-Greece, using ^{1}H-NMR spectral data and/or H and C isotope abundances of the bulk oil and/or its unsaponifiable fraction.[a]

				Greece		Non-Greece[b]		
			N	28		56		28
			prior prob	0.33		0.67		
Dataset	Crossvalidation	PLS comp	boundary	% R	% P	% R	% P	% P-EV
OO_NMR	LOO-CV	5	0.4794	100	92.9	100	87.5	82.1
OO_NMR-IRMS	LOO-CV	5	0.5006	**100**	**96.4**	**100**	**96.4**	82.1
Unsap_NMR	LOO-CV	5	0.5188	100	92.9	100	92.9	78.6
Unsap_NMR-IRMS	LOO-CV	5	0.5148	100	92.9	100	94.6	78.6
OO-Unsap_NMR	LOO-CV	3	0.4768	96.4	92.9	100	89.3	85.7
OO-Unsap_PC_NMR	LOO-CV	1	0.0918	100	89.3	**98.2**	**98.2**	**92.9**
OO-Unsap_PC_NMR	3-fold CV	1	0.0968	100	89.3	98.2	94.6	92.9
OO-Unsap_NMR-IRMS	LOO-CV	3	0.4789	100	92.9	100	87.5	85.7
OO-Unsap_PC_NMR-IRMS	LOO-CV	1	0.0678	100	89.3	98.2	98.2	92.9
OO-Unsap_PC_NMR-IRMS	3-fold CV	1	0.1099	100	85.7	100	98.2	92.9

[a]Abbreviations: PLS-DA, partial least square discriminant analysis; VOO, virgin olive oil; N, number of samples; prior prob, prior probability; OO_NMR, ^{1}H-NMR data of the bulk oil; OO_NMR-IRMS, ^{1}H-NMR, δ^{13}C and δ^2H data of the bulk oil; OO-Unsap_NMR, ^{1}H-NMR data of the bulk oil and the unsaponifiable fraction; OO-Unsap_PC_NMR, dataset made up with the PCA score matrix of the principal components with eigenvalues higher than 1, afforded by performing PCA on the bulk oil and the unsaponifiable fraction NMR datasets separately; OO-Unsap_NMR-IRMS, ^{1}H-NMR, δ^{13}C and δ^2H data of the bulk oil and the unsaponifiable fraction; OO-Unsap_PC_NMR-IRMS, OO-Unsap_PC_NMR dataset and δ^{13}C and δ^2H data of the bulk oil and the VOO unsaponifiable fraction; Unsap_NMR, ^{1}H-NMR data of the unsaponifiable fraction; Unsap_NMR-IRMS, ^{1}H-NMR, δ^{13}C and δ^2H data of the unsaponifiable fraction; LOO-CV, leave-one-out crossvalidation; 3-fold CV, 3-fold crossvalidation; PLS comp, number of PLS components selected; % R, percentage of recognition ability; % P, percentage of prediction ability in crossvalidation; % P-EV, percentage of prediction ability in the external validation. Class codes: Greece, 1; non-Greece, 0.

[b]Non-Greece samples are from Spain (N = 28) and Italy (N = 28) in crossvalidation, and from Turkey (N = 13), France (N = 10), and Cyprus (N = 5) in the external validation.

calculated by PCA were those that presented the highest loading values: in PC9, the glyceryl protons of *sn*-1,3-diglycerides (4.05–4.07 ppm), the buckets at 0.09–0.11 ppm and 0.15 ppm, and the CH glycerol protons of *sn*-1,2-diglycerides (5.07–5.09 ppm); and in PC15, the buckets at 0.75 ppm (AF), 2.75–2.77 ppm (SF, AF), 6.51 ppm (phenolic compounds), 2.61 ppm (TF), and 7.85 ppm and 7.31–7.33 ppm (phenolic compounds).

The best PLS-DA models for both case studies, the Spain versus non-Spain and Italy versus non-Italy, were obtained with the datasets that contained the PCA score matrices of ^{1}H-NMR data, and δ^{13}C and δ^2H values of the bulk oils and the unsaponifiable fractions (OO-Unsap_PC_NMR-IRMS dataset). The Spain versus non-Spain model recognised 100% of the Spanish VOOs, but predicted correctly only 78.6% of the samples (Table 2). This large difference between the recognition and prediction abilities indicated that the classification results were very dependent on the samples included in the training set in the modeling step. In contrast, the recognition and prediction abilities in the cross-validation and external validation for the non-Spain class were close to each other and decreasing in this order (80.4, 76.8, and 71.4% of hits, respectively); pointing out that the results were robust and stable for this category. The classification results achieved for the Spain versus non-Spain case study were similar to those previously reported [20]. The most influential variables on the PLS-DA model were PC7 of the bulk oil dataset, PC8 of the unsaponifiable fraction dataset and the δ^2H values of the unsaponifiable fractions. The highest loadings on PC7 of the bulk oil dataset were the ^{1}H signals at 0.11–0.13 ppm, 6.45 ppm (phenolic compounds), 4.29 ppm (α-methylene protons of the glyceryl group of triglycerides), 0.17 and 0.27 ppm; and on PC8 of the unsaponifiable fraction dataset, 9.75 ppm (aldehydes), 5.59–5.61 ppm (TF), 3.63–3.65 ppm (AF, cycloartenol, 1-eicosanol), 6.9 (phenolic compounds), 2.37, and 4.69 ppm (AF).

The Italy versus non-Italy PLS-DA model (OO-Unsap_PC_NMR-IRMS dataset) predicted properly about 80% of the samples of both categories (Table 3), improving the classifications already described in an earlier work [20], in which the VOO unsaponifiable fractions were not analysed. Indeed, one of the most influential variables on the PLS-DA model was δ^{13}C value of the unsaponifiable fractions. The other variables with the highest regression coefficients were ^{13}C isotopic abundance of the bulk oils and PC10 of the bulk

Table 2. PLS-DA models for the geographical discrimination of VOOs: Spain versus non-Spain, using ^{1}H-NMR spectral data and/or H and C isotope abundances of the bulk oil and/or its unsaponifiable fraction.[a]

				Spain		Non-Spain[b]		
			N	28		56		28
			Prior prob	0.33		0.67		
Dataset	Crossvalidation	PLS comp	Boundary	% R	% P	% R	% P	% P-EV
OO_NMR	LOO-CV	6	0.4205	100	89.3	94.6	85.7	25.0
OO_NMR-IRMS	LOO-CV	6	0.4359	100	89.3	98.2	80.4	32.1
Unsap_NMR	LOO-CV	5	0.4514	100	82.1	96.4	75.0	64.3
Unsap_NMR-IRMS	LOO-CV	5	0.4440	100	82.1	98.2	82.1	60.7
OO-Unsap_NMR	LOO-CV	5	0.4236	100	89.3	98.2	87.5	64.3
OO-Unsap_PC_NMR	LOO-CV	1	−0.0687	100	75.0	75.0	75.0	64.3
OO-Unsap_PC_NMR	3-fold CV	1	−0.0610	100	75.0	75.0	75.0	71.4
OO-Unsap_NMR-IRMS	LOO-CV	5	0.4084	100	89.3	100	83.9	57.1
OO-Unsap_PC_NMR-IRMS	LOO-CV	1	−0.0619	100	75.0	80.4	80.4	71.4
OO-Unsap_PC_NMR-IRMS	3-fold CV	1	−0.0516	**100**	**78.6**	**80.4**	**76.8**	**71.4**

[a]Abbreviations: see Table 1. Class codes: Spain, 1; non-Spain, 0.
[b]Non-Spain samples are from Greece (N = 28) and Italy (N = 28) in crossvalidation, and from Turkey (N = 13), France (N = 10), and Cyprus (N = 5) in the external validation.

oil dataset. In particular, the most influential resonances on PC10 were 5.09–5.15 ppm (CH glycerol protons of *sn*-1,2-diglycerides), 6.75 ppm (phenolic compounds), and 2.71 ppm and 2.77–2.79 ppm (diallylic protons of acyl groups). So, these results disclosed that ^{13}C isotopic measurements of the unsaponifiable fractions of VOOs provided complementary information for the authentication of Italian VOOs and the detection of fraudulent non-Italian VOOs; whereas ^{1}H-NMR data and δ^{2}H values of the unsaponifiable fractions did not.

Table 3. PLS-DA models for the geographical discrimination of VOOs: Italy versus non-Italy, using ^{1}H-NMR spectral data and/or H and C isotope abundances of the bulk oil and/or its unsaponifiable fraction.[a]

				Italy		Non-Italy[b]		
			N	28		56		28
			Prior prob	0.33		0.67		
Dataset	Crossvalidation	PLS comp	Boundary	% R	% P	% R	% P	% P-EV
OO_NMR	LOO-CV	8	0.3851	100	75.0	96.4	75.0	71.4
OO_NMR-IRMS	LOO-CV	8	0.4173	100	82.1	100	83.9	64.3
Unsap_NMR	LOO-CV	1	0.2824	67.9	60.7	69.6	57.1	28.6
Unsap_NMR-IRMS	LOO-CV	9	0.3909	100	85.7	100	80.4	35.7
OO-Unsap_NMR	LOO-CV	7	0.3790	100	71.4	100	73.2	57.1
OO-Unsap_PC_NMR	LOO-CV	1	−0.0676	100	71.4	69.6	69.6	89.3
OO-Unsap_PC_NMR	3-fold CV	1	−0.0489	100	64.3	69.6	67.9	89.3
OO-Unsap_NMR-IRMS	LOO-CV	5	0.4060	100	82.1	98.2	82.1	42.9
OO-Unsap_PC_NMR-IRMS	LOO-CV	1	0.0335	**100**	**82.1**	**78.6**	**78.6**	**78.6**
OO-Unsap_PC_NMR-IRMS	3-fold CV	1	−0.0053	100	78.6	78.6	82.1	78.6

[a]Abbreviations: see Table 1. Class codes: Italy, 1; non-Italy, 0.
[b]Non-Italy samples are from Greece (N = 28) and Spain (N = 28) in crossvalidation, and from Turkey (N = 13), France (N = 10), and Cyprus (N = 5) in the external validation.

3.3.2 PDO *Riviera Ligure* VOOs

In the framework of the EU TRACE project, European VOOs were studied with the aim of developing accurate analytical fingerprinting methods for the certification of the geographical origin of olive oils. In particular, the case study considered was the authentication of VOOs belonging to the PDO *Riviera Ligure* (Liguria, Italy). Several analytical approaches combined with multivariate techniques have been proposed to distinguish VOOs from this PDO and VOOs from other European regions: ^{1}H NMR spectroscopy and PLS-DA [20] or SIMCA [21]; FT-IR spectroscopy and PLS-DA [46], CART or SVM [47]; NIR spectroscopy and PLS-DA [48]; proton-transfer-reaction mass spectrometry (PTR-MS) and PLS−DA [42]; GC × GC-Tof-MS and neural networks [49]; and SESI-MS and kNN or the NIST MS search algorithm [27]. The results of these previous studies [21, 46] afforded models with prediction abilities of 80–84% for the PDO VOOs. Similar results were achieved by those methodologies that analyzed the volatile profiles of VOOs by MS [42, 49]. In contrast, PLS-DA models developed with ^{1}H NMR [20] and NIR data [48], as well as the kNN model or the NIST MS search algorithm with SESI-MS data [27], provided prediction abilities of around 87–92% for the PDO VOOs, and 84–86% for the non-PDO. All these studies were performed on the same samples of VOOs; however some studies considered the whole sample set (941 VOOs), and others examined a subset of the samples, as in the present study. Better results are expected when larger datasets are used [33]. In this work, a novel approach based on ^{1}H-NMR spectral data and ^{2}H and ^{13}C isotope abundances of the bulk oil and the unsaponifiable fraction of VOOs is evaluated for the same purpose. On the one hand, the classification model built with the PCA score matrices of ^{1}H-NMR data of the bulk oils and the unsaponifiable fractions (OO-Unsap_PC_NMR dataset) allowed us to authenticate 100% of PDO *Riviera Ligure* VOOs, and detected more than 86% of the non-Ligurian samples (Table 4). On the other hand, when the H and C isotope abundances were also included in the model (OO-Unsap_PC_NMR-IRMS), 94% of the non-Ligurian VOOs were predicted correctly, providing a better classification model for fraud detection. These classification models outperformed those previously reported for this case study, and cited above [20, 21, 27, 42, 46–49]. The highest regression coefficients on these binary PLS-DA models were PC4 and PC2 of the ^{1}H-NMR datasets corresponding to the bulk oils and the unsaponifiable fractions, respectively. Moreover, δ^{13}C measurements on both, the bulk oils and the unsaponifiable fractions, were also among the most influential variables on the OO-Unsap_PC_NMR-IRMS model. The NMR buckets presenting the highest loadings in PC4 of the bulk oil dataset belonged to the proton signals at 1.55–1.57 ppm and 1.61 ppm (β-methylene protons of the acyl group), 1.21–1.23 ppm (methylene protons of the acyl group), 5.29–5.31 ppm (vinylic protons of the acyl group), 2.23–2.25 ppm (α-methylene protons of the acyl group), 2.01 ppm and 1.95 ppm (the allylic protons of the acyl group), and 4.09 ppm (α-methylene protons of the glyceryl group of triglycerides). The most influential ^{1}H signals in

Table 4. PLS-DA models for the geographical discrimination of VOOs: PDO *Riviera Ligure* (Liguria, Italy) versus non-PDO *Riviera Ligure*, using ^{1}H-NMR spectral data and/or H and C isotope abundances of the bulk oil and/or its unsaponifiable fraction.[a]

				PDO *Riviera Ligure*		Non-PDO *Riviera Ligure*[b]		
			N	10		20		82
			Prior prob	0.33		0.67		
Dataset	Crossvalidation	PLS comp	Boundary	% R	% P	% R	% P	% P-EV
OO_NMR	LOO-CV	6	0.4294	100	90.0	100	85.0	86.6
OO_NMR-IRMS	LOO-CV	6	0.4291	100	90.0	100	85.0	89.0
Unsap_NMR	LOO-CV	4	0.5021	100	80.0	100	75.0	90.2
Unsap_NMR-IRMS	LOO-CV	4	0.5036	100	80.0	100	75.0	90.2
OO-Unsap_NMR	LOO-CV	2	0.5066	100	60.0	100	70.0	92.7
OO-Unsap_PC_NMR	LOO-CV	2	0.1076	**100**	**100**	**95.0**	**90.0**	**86.6**
OO-Unsap_PC_NMR	3-fold CV	2	0.0341	100	80.0	75.0	80.0	79.3
OO-Unsap_NMR-IRMS	LOO-CV	2	0.5046	100	70.0	100	70.0	92.7
OO-Unsap_PC_NMR-IRMS	LOO-CV	2	0.2013	**100**	**90.0**	**100**	**90.0**	**93.9**
OO-Unsap_PC_NMR-IRMS	3-fold CV	2	0.1288	100	80.0	95.0	80.0	89.0

[a]Abbreviations: see Table 1. Class codes: PDO *Riviera Ligure*, 1; non-PDO *Riviera Ligure*, 0.
[b]Non-PDO *Riviera Ligure* samples are from Greece (N = 4), Spain (N = 5), France (N = 2), Turkey (N = 2), Cyprus (N = 1) and other Italian regions (N = 6) in crossvalidation, and from Greece (N = 24), Spain (N = 23), France (N = 8), Turkey (N = 11), Cyprus (N = 4) and other Italian regions (N = 12) in the external validation.

PC2 of the unsaponifiable fractions were resonating at 6.35–6.39 ppm (TF), 5.63–5.71 ppm (TF), 6.43–6.49 ppm (HF), 6.57–6.59 ppm, 6.53 and 6.29 ppm, (phenolic compounds), and 2.45 ppm (SF).

3.3.3 PDO *Huile d'olive d'Aix–en–Provence* VOOs

For the authentication of PDO *Huile d'olive d'Aix–en–Provence* VOOs, the PCA score matrices of ^{1}H-NMR data, and δ^{13}C and δ^2H values of the bulk oils and the unsaponifiable fractions (OO-Unsap_PC_NMR-IRMS dataset) afforded the best classification model with 100% of hits (Table 5). For the non-PDO category, the recognition and prediction abilities of this model provided 80% of correct classifications in cross-validation, and 90% in the external validation. Non-PDO VOOs were much better predicted in the external data set than in the cross-validation, which 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 assured that the training-test set was representative of all the samples (at least of the two harvests studied), however the predictions on the external set could be overoptimistic. The most influential variables in the binary PLS-DA model were the δ^2H values of the unsaponifiable fractions, then δ^2H values of the bulk oil, and PC10 and PC5 of the ^{1}H-NMR datasets of the bulk oils and the unsaponifiable fractions respectively. The main resonances on PC10 of the bulk oil dataset were due to CH glycerol protons of *sn*-1,2-diglycerides (5.07–5.15 ppm), phenolic compounds (6.73–6.75 ppm), diallylic proton of the acyl groups (2.71 ppm), diallylic proton of linolenic acyl group (2.77–2.79 ppm), the allylic protons of the acyl groups (2.05 ppm); and on PC5 of the unsaponifiable fraction, to proton signals of cycloartenol, 1-eicosanol, α-tocopherol, and erythrodiol (1.27–1.35 ppm), to the methylene proton of the cyclopropanoic ring of cycloartenol (0.55–0.57 ppm), to SF, AF, β-sitosterol, campesterol and cycloartenol (0.81 ppm), to AF (4.73 and 4.67 ppm), to SF (5.35 ppm), and to AF (1.03 ppm).

3.3.4 Region of origin of VOOs

VOOs from three of the major European regions that produce olive oils were studied: Andalucía (Spain), Izmir (Turkey) and Crete (Greece). Several PDO/PGI VOOs produced in these regions are registered in the DOOR database. Andalucía is the main VOO producing region in Spain, accounting for ca. 80% of the total Spanish VOO production. Turkey is, together with Tunisia and Syria, one of the main VOO producers outside the EU, accounting each for ca. 5% of the total world production; Izmir being one of the three main producing Turkish regions. Crete, together with Peloponnese, account for over 65% of the total production of Greek VOOs. Considering the importance of each of these regions in the world production of VOOs, ^{1}H-NMR data and C and H isotope abundances of the bulk oils and the unsaponifiable fractions were modeled to get classification tools to authenticate VOOs from these regions and/or detect fraudulent VOOs passed off as belonging to these origins.

Table 5. PLS-DA models for the geographical discrimination of VOOs: PDO *Huile d'olive d'Aix-en-Provence* (France) versus non-PDO *Huile d'olive d'Aix-en-Provence*, using ^{1}H-NMR spectral data and/or H and C isotope abundances of the bulk oil and/or its unsaponifiable fraction.[a]

				PDO *Aix-en-Provence*		Non-PDO *Aix-en-Provence*[b]		
			N	10		20		82
			Prior prob	0.33		0.67		
Dataset	Crossvalidation	PLS comp	Boundary	% R	% P	% R	% P	% P-EV
OO_NMR	LOO-CV	4	0.5541	100	90.0	100	80.0	75.6
OO_NMR-IRMS	LOO-CV	4	0.5463	100	90.0	100	75.0	74.4
Unsap_NMR	LOO-CV	1	0.4860	90.0	70.0	85.0	70.0	69.5
Unsap_NMR-IRMS	LOO-CV	1	0.4886	90.0	70.0	85.0	70.0	70.7
OO-Unsap_NMR	LOO-CV	2	0.5431	100	60.0	100	70.0	78.0
OO-Unsap_PC_NMR	LOO-CV	1	0.0587	100	80.0	85.0	85.0	73.2
OO-Unsap_PC_NMR	3-fold CV	1	0.0905	100	80.0	85.0	85.0	79.3
OO-Unsap_NMR-IRMS	LOO-CV	2	0.5434	100	60.0	100	70.0	79.3
OO-Unsap_PC_NMR-IRMS	LOO-CV	1	0.1313	**100**	**100**	**80.0**	**80.0**	**90.2**
OO-Unsap_PC_NMR-IRMS	3-fold CV	1	0.1425	100	100	80.0	75.0	90.2

[a]Abbreviations: see Table 1. Class codes: PDO *Aix-en-Provence*, 1; non-PDO *Aix-en-Provence*, 0.
[b]Non-PDO *Aix-en-Provence* samples are from Greece (N = 5), Spain (N = 5), Italy (N = 9), and Cyprus (N = 1) in crossvalidation, and from Greece (N = 23), Spain (N = 23), Italy (N = 19), Turkey (N = 13), and Cyprus (N = 4) in the external validation.

For the Andalucía (Spain) versus non-Andalucía case study (Table 6), the best PLS-DA model to authenticate Andalusian VOOs was built with ^{1}H-NMR data of the bulk oils (OO_NMR dataset); presenting 83.3% of recognition ability and 77.8% of prediction ability. To achieve similar classifications for the non-Andalusian VOOs (recognition ability, 77.8%; and prediction abilities in cross-validation, 77.8%, and in external validation, 81.0%), ^{1}H-NMR data of the unsaponifiable fractions was also required (OO-Unsap_PC_NMR dataset), but using PCA score matrices instead of the raw NMR data. The most influential variables, i.e., those with the highest regression coefficients, on the binary PLS-DA model obtained with OO_NMR dataset were the proton signals at the following chemical shifts: 0.11–0.27 ppm; 1.71–1.81 ppm; 4.05–4.07 ppm, due to glyceryl protons of *sn*-1,3-diglycerides; 5.09–5.15 ppm, to CH glycerol protons of *sn*−1,2−diglycerides; 5.71−5.75 ppm; and 6.59−6.61 ppm and 6.71−6.79 ppm, to phenolic compounds. The highest regression coefficients in the PLS-DA model obtained with OO-Unsap_PC_NMR dataset were PC7 of the bulk oil dataset and PC3 of the unsaponifiable fraction dataset. The NMR buckets that presented the highest loadings on PC7 were at 0.11–0.27 ppm, at 6.45, 6.61, 6.69. and 6.77 ppm (phenolic region), at 4.29 ppm (α-methylene protons of the glyceryl group of triglycerides) and at 4.59 ppm (terpene). Regarding PCA on the unsaponifiable fraction dataset, the most influential variables on PC3 were proton signals at 1.37–1.41 ppm (all subfractions and 1-eicosanol and α-tocopherol), 2.01 ppm (HF, TF; and, to a lesser extent, SF and AF), 1.45–1.47 ppm, and 5.39–5.43 ppm (SF).

The Turkey VOOs from Izmir were correctly predicted with 92.3% of hits by a PLS-DA model built with ^{1}H-NMR data and δ^{13}C and δ^{2}H values of the bulk oils (OO_NMR-IRMS dataset) (Table 7). This model presented prediction abilities in both, cross-validation and external validation, of about 85% for the VOOs not produced in Izmir. However, better classifications were achieved for the non-Izmir category by the model developed with the PCA score matrices of ^{1}H-NMR data of the bulk oils and the unsaponifiable fractions (OO-Unsap_PC_NMR dataset), since recognition and prediction abilities in cross validation were of 92.3%, and in the external set, 95.9% of the non-Izmir VOOs were detected (Table 7). The variables that presented the highest regression coefficients, i.e. the most influential variables in the PLS−DA model obtained with OO_NMR-IRMS dataset belonged to terpenes (4.57–4.61 ppm), the diallylic protons of the acyl group of linoleic acid (2.75–2.77 ppm), the ^{13}C/^{12}C isotopic ratio, the allylic protons of the acyl group (2.05 ppm), the methylic proton of the C18-steroid group of β-sitosterol (0.67 ppm), phenolic compounds (6.53–6.57 ppm, 5.97–5.99 ppm, 6.73, and 6.83 ppm), the α-methylene protons of the glyceryl group of triglycerides (4.29 ppm), the β-methylene protons of the acyl group (1.69 ppm), and the ^{13}C satellite of the methylic proton of acyl groups (1.01 ppm). Regarding the binary PLS−DA model built with OO-Unsap_PC_NMR dataset, PC25 of the unsaponifiable fraction dataset and PC3 of the bulk oil dataset

Table 6. PLS-DA models for the geographical discrimination of VOOs, Andalucía (Spain) versus non-Andalucía, using ^{1}H-NMR spectral data and/or H and C isotope abundances of the bulk oil and/or its unsaponifiable fraction.[a]

				Andalucía		Non-Andalucía[b]		
			N	18		36		58
			Prior prob	0.33		0.67		
Dataset	Crossvalidation	PLS comp	Boundary	% R	% P	% R	% P	% P-EV
OO_NMR	LOO-CV	3	0.4328	**83.3**	**77.8**	97.2	75.0	82.8
OO_NMR-IRMS	LOO-CV	3	0.4170	88.9	72.2	91.7	75.0	81.0
Unsap_NMR	LOO-CV	1	0.4263	83.3	66.7	69.4	69.4	82.8
Unsap_NMR-IRMS	LOO-CV	1	0.4283	83.3	66.7	72.2	69.4	82.8
OO-Unsap_NMR	LOO-CV	1	0.4028	88.9	66.7	77.8	72.2	79.3
OO-Unsap_PC_NMR	LOO-CV	1	0.0231	**100**	**77.8**	**77.8**	**77.8**	**81.0**
OO-Unsap_PC_NMR	3-fold CV	1	0.0541	100	77.8	80.6	75.0	84.5
OO-Unsap_NMR-IRMS	LOO-CV	1	0.3986	88.9	66.7	77.8	75.0	79.3
OO-Unsap_PC_NMR-IRMS	LOO-CV	1	−0.0219	100	77.8	69.4	69.4	81.0
OO-Unsap_PC_NMR-IRMS	3-fold CV	1	0.0479	100	77.8	80.6	77.8	84.5

[a]Abbreviations: see Table 1. Class codes: Andalucía, 1; non-Andalucía, 0.
[b]Non-Andalucía samples are from Greece (N=5), Italy (N = 14), France (N = 4), Turkey (N = 3), Cyprus (N = 1), and other Spanish regions (N = 9) in crossvalidation, and from Greece (N = 23), Italy (N = 14), France (N = 6), Turkey (N = 10), Cyprus (N = 4), and other Spanish regions (N = 1) in the external validation.

Table 7. PLS-DA models for the geographical discrimination of VOOs: Izmir (Turkey) versus non-Izmir, using ^{1}H-NMR spectral data and/or H and C isotope abundances of the bulk oil and/or its unsaponifiable fraction.[a]

				Izmir		Non-Izmir[b]		
			N	13		26		73
			Prior prob	0.33		0.67		
Dataset	Crossvalidation	PLS comp	Boundary	% R	% P	% R	% P	% P-EV
OO_NMR	LOO-CV	3	0.4531	100	92.3	96.2	80.8	84.9
OO_NMR-IRMS	LOO-CV	3	0.4574	**100**	**92.3**	**96.2**	**84.6**	**86.3**
Unsap_NMR	LOO-CV	3	0.4665	100	69.2	100	73.1	67.1
Unsap_NMR-IRMS	LOO-CV	3	0.4638	100	69.2	100	65.4	65.8
OO-Unsap_NMR	LOO-CV	2	0.5148	100	76.9	96.2	76.9	84.9
OO-Unsap_PC_NMR	LOO-CV	1	0.1791	100	76.9	**92.3**	**92.3**	**95.9**
OO-Unsap_PC_NMR	3-fold CV	1	0.1322	100	84.6	92.3	88.5	91.8
OO-Unsap_NMR-IRMS	LOO-CV	2	0.5149	100	69.2	96.2	73.1	84.9
OO-Unsap_PC_NMR-IRMS	LOO-CV	1	0.1266	100	76.9	84.6	84.6	91.8
OO-Unsap_PC_NMR-IRMS	3-fold CV	1	0.1098	100	92.3	84.6	88.5	90.4

[a]Abbreviations: see Table 1. Class codes: Izmir, 1; non-Izmir, 0.
[b]Non-Izmir samples are from Greece (N = 5), Spain (N = 5), Italy (N = 13), France (N = 2) and Cyprus (N = 1) in crossvalidation, and from Greece (N = 23), Spain (N = 23), Italy (N = 15), France (N = 8) and Cyprus (N = 4) in the external validation.

were the variables with the highest regression coefficient. In consequence, the NMR signals of the unsaponifiable fractions more influential in the models were at 9.39 ppm (HF), 6.09 and 6.13 ppm (aldehydes), 0.75 ppm (AF), 4.29–4.31 ppm (AF), 7.03 ppm (phenolic compounds), 9.51 ppm (aldehydes), 9.23, 8.41, 6.57 (phenolic compounds), and 2.75 ppm. The variables with the highest PCA loadings in PC3 were 1.35–1.45 ppm (methylene protons of acyl groups), 1.65 ppm (β-methylene protons of the acyl groups), 2.07–2.09 ppm (allylic protons of the acyl groups), 0.91–0.93 ppm (methylic proton of the acyl groups), 4.27–4.33 ppm (α-methylene protons of the

Table 8. PLS-DA models for the geographical discrimination of VOOs: Crete (Greece) versus non-Crete, using ^{1}H-NMR spectral data and/or H and C isotope abundances of the bulk oil and/or its unsaponifiable fraction.[a]

				Crete		Non-Crete[b]		
			N	16		32		64
			Prior prob	0.33		0.67		
Dataset	Crossvalidation	PLS comp	Boundary	% R	% P	% R	% P	% P-EV
OO_NMR	LOO-CV	5	0.4468	100	68.8	96.9	68.8	82.8
OO_NMR-IRMS	LOO-CV	5	0.4541	100	68.8	100	71.9	84.4
Unsap_NMR	LOO-CV	–	–	–	–	–	–	–
Unsap_NMR-IRMS	LOO-CV	–	–	–	–	–	–	–
OO-Unsap_NMR	LOO-CV	–	–	–	–	–	–	–
OO-Unsap_PC_NMR	LOO-CV	1	0.1846	100	75.0	78.1	78.1	95.3
OO-Unsap_PC_NMR	3-fold CV	1	0.1972	100	75.0	78.1	75.0	95.3
OO-Unsap_NMR-IRMS	LOO-CV	–	–	–	–	–	–	–
OO-Unsap_PC_NMR-IRMS	LOO-CV	1	0.1851	100	75.0	78.1	78.1	93.8
OO-Unsap_PC_NMR-IRMS	3-fold CV	1	0.1883	100	75.0	78.1	75.0	93.8

[a]Abbreviations: see Table 1. Class codes: Crete, 1; non-Crete, 0.
[b]Non-Crete samples are from Spain (N = 6), Italy (N = 11), France (N = 2), Turkey (N = 1), Cyprus (N = 2) and other Greek regions (N = 10) in crossvalidation, and from Spain (N = 22), Italy (N = 17), France (N = 8), Turkey (N = 12), Cyprus (N = 3) and other Greek regions (N = 2) in the external validation.

glyceryl group of triglycerides), and 2.29 ppm and 2.33–2.37 ppm (α-methylene protons of the acyl groups).

Regarding Crete (Greece) versus non-Crete VOO datasets (Table 8), PLD-DA failed to build models with datasets containing raw ^{1}H-NMR data of the unsaponifiable fractions. However, the best PLS-DA models were achieved by using the PCA score matrices of ^{1}H-NMR data of the bulk oils and the unsaponifiable fractions (OO-Unsap_PC_NMR dataset). For the Cretan category, the recognition and prediction abilities in cross-validation were 100 and 75%, respectively. This large difference between both abilities meant that the classification results were very dependent on the samples included in the training set in the modeling step. Therefore the model could not be considered satisfactory for this category. In contrast, the model performed better for the non-Crete class, presenting the same recognition and prediction abilities in cross-validation, 78% of hits; and 94% of correct predictions in the external set, which was overoptimistic for the reason already explained above. Similar classifications were obtained if isotope abundances were included in the dataset (OO-Unsap_PC_NMR-IRMS dataset), showing that isotope information was not relevant for the aimed classification. Indeed, δ^{13}C and δ^{2}H were not among the variables that presented the highest regression coefficients. PC19 of the bulk oil dataset and PC14 and PC16 of the unsaponifiable fraction dataset were the most influential variables in both PLS-DA models (OO-Unsap_PC_NMR and OO-Unsap_PC_NMR-IRMS). The highest loadings in PC19 corresponded to 6.47 (phenolic compounds), 6.31, and 4.01–4.03 ppm (TF); in PC14, to 6.13 ppm and 6.15–6.17 ppm (aldehydes), 7.03, and 7.21 ppm (phenolic compounds); and in PC16, to 6.05 (aldehydes), 9.37 (HF), 9.61 (aldehydes), 7.25 (phenolic compounds), 9.05, and 6.01 ppm.

4 Conclusions

^{1}H-NMR and isotopic fingerprinting of VOO and its unsaponifiable fraction contain useful information for the geographical characterization of VOOs. Pattern recognition techniques applied to ^{1}H-NMR spectral data and H and C isotope abundances of the bulk oil and the unsaponifiable fraction, using different approaches, provided chemical tools for the authentication of VOOs according to their geographical origin or PDO, as well as to detect the mislabeling of the provenance of VOOs, at the regional or national level, or the mislabeling of non-PDO oils as PDO VOOs. In this sense, PLS-DA models with ^{1}H-NMR data of the bulk oil allowed the authentication of Andalusian VOOs; and together with the ^{13}C isotopic ratio, the authentication of Izmir VOOs, and with both ^{13}C and ^{2}H isotopic ratios, the authentication of Greek VOOs. ^{1}H-NMR data of the bulk oil and the unsaponifiable fraction of VOOs and PLS-DA provided tools for the detection of fraudulent non-Greek VOOs, non-Andalusian VOOs, non-Izmir VOOs, and non-Crete VOOs, and for the authentication of PDO *Riviera Ligure* VOOs and Cretan VOOs. ^{2}H isotope abundances of the bulk oils and the unsaponifiable fractions provided additional information to the ^{1}H-NMR data to afford PLS-DA classification models for Spain versus non-Spain VOOs and PDO *Huile d'olive d'Aix–en–Provence vs* non-–PDO VOOs. Whereas ^{13}C isotope ratios contained additional information for the classification of VOOs from Italy *vs* non-Italy, and for the detection of fraudulent VOOs passed off as PDO *Riviera Ligure* VOOs. The present approach for PDO *Riviera Ligure* VOOs, based on ^{1}H-NMR spectral data and C isotope abundance of the bulk oil and the unsaponifiable fraction, outperformed the previously reported classification models. H and C isotope abundances of the VOOs and its unsaponifiable fractions confirmed that both isotopes are related to the latitude of the VOO geographical origin.

All these results disclosed that ^{1}H-NMR spectral and stable isotope data of VOO and its unsaponifiable fraction can be a useful tool to assure authenticity and traceability of VOOs regarding the geographical origin of the oil, but further studies should be carried out with a considerably larger balanced sample set for each region, and even for each PDO/PGI, in order to guarantee robust models for both authentication and detection of fraud when VOO is falsely labeled as belonging to a certain origin.

The authors thank the EU TRACE project (Integrated project 006942, under the Food Quality and Safety Priority of the Sixth Framework Programme) for funding this work and the research groups that participated in the collection of the olive oil samples: Dipartimento di Chimica e Technologie Farmaceutiche ed Alimentari – Università degli Studi di Genova (Italia), 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 – Laboratorio di Risonanza Magnetica Annalaura Segre – CNR (Italy), Fondazione Edmund Mach – Istituto San Michele all'Adige (Italy), and Eurofins Scientific Analytics (France).

The authors have declared no conflict of interest. The information contained in this article reflects the authors' views; the European Commission is not liable for any use of the information contained herein.

References

[1] Aparicio, R., Conte, L., Fiebig, H. J., in: Aparicio, R., Harwood, J. (Eds.), *Handbook of Olive Oil, Olive Oil Authentication*, Springer, US 2013, pp. 589–653.

[2] Cañabate-Diaz, B., Segura Carretero, A., Fernandez-Gutierrez, A., Belmonte Vega, A., et al., Separation and

determination of sterols in olive oil by HPLC-MS. *Food Chem.* 2007, *102*, 593–598.

[3] Aparicio, R., Aparicio-Ruiz, R., Authentication of vegetable oils by chromatographic techniques. *J. Chromatogr. A* 2000, *881*, 93–104.

[4] Aparicio, R., Alonso, V., Morales, M. T., Detailed and exhaustive study of the authentication of European virgin olive oils by SEXIA expert system. *Grasas y Aceites* 1994, *45*, 241–252.

[5] Aparicio, R., Morales, M. T., Aparicio-Ruiz, R., Tena, N., et al., Authenticity of olive oil: Mapping and comparing official methods and promising alternatives. *Food Res. Int.* 2013, *54*, 2025–2038.

[6] Matos, L. C., Cunha, S. C., Amaral, J. S., Pereira, J. A., et al. Chemometric characterization of three varietal olive oils (Cvs. Cobrancosa, Madural and Verdeal Transmontana) extracted from olives with different maturation indices. *Food Chem.* 2007, *102*, 406–414.

[7] Aranda, F., Gómez-Alonso, S., Rivera del Álamo, R. M., Salvador, M. D., et al. Triglyceride, total and 2-position fatty acid composition of Cornicabra virgin olive oil: Comparison with other Spanish cultivars. *Food Chem.* 2004, *86*, 485–492.

[8] Romero, C., Brenes, M., Analysis of total contents of hydroxytyrosol and tyrosol in olive oils. *J. Agric. Food Chem.* 2012, *60*, 9017–9022.

[9] García-González, D. L., Tena, N., Aparicio, R., Quality characterization of the new virgin olive oil var. Sikitita by phenols and volatile compounds. *J. Agric. Food Chem.* 2010, *58*, 8357–8364.

[10] Pérez-Camino, M. C., Gómez-Coca, R. B., Moreda, W., Waxy fraction containing long-chain aliphatic aldehydes in virgin olive oils. *Food Chem.* 2012, *132*, 1451–1456.

[11] Cichelli, A., Pertesana, G. P., High-performance liquid chromatographic analysis of chlorophylls, pheophytins and carotenoidsinvirgin olive oils: chemometric approach to variety classification. *J. Chromatogr. A* 2004, *1046*, 141–146.

[12] Ben-Ayed, R., Kamoun-Grati, N., Rebai, A., An overview of the authentication of olive tree and oil. *Compr. Rev. Food Sci. F* 2013, *12*, 218–227.

[13] León-Camacho, M., Morales, M., Aparicio, R., Aparicio, R., in: Aparicio, R., Harwood, J. (Eds.), *Handbook of Olive Oil, Chromatographic methodologies: Compounds For Olive Oil Traceability Issues*, Springer, US 2013, pp. 163–217.

[14] Baccouri, B., Temime, S. B., Campeol, E., Cioni, P. L., et al. Application of solid-phase microextraction to the analysis of volatile compounds in virgin olive oils from five new cultivars. *Food Chem.* 2007, *102*, 850–856.

[15] Vichi, S., Pizzale, L., Conte, L. S., Buxaderas, S., et al. Simultaneous determination of volatile and semi-volatile aromatic hydrocarbons in virgin olive oil by headspace solid-phase microextraction coupled to gas chromatography/mass spectrometry. *J. Chromatogr. A* 2005, *1090*, 146–154.

[16] Zarrouk, W., Carrasco-Pancorbo, A., Segura-Carretero, A., Fernández-Gutiérrez, A., et al. Exploratory characterization of the unsaponifiable fraction of Tunisian virgin olive oils by a global approach with HPLC-APCI-IT MS/MS analysis. *J. Agric. Food Chem.* 2010, *58*, 6418–6426.

[17] Costa, J., Mafra, I., Oliveira, M. B. P. P., Advances in vegetable oil authentication by DNA-based markers. *Trends Food Sci. Tech.* 2012, *26*, 43–55.

[18] Galtier, O., Dupuy, N., Le Dréau, Y., Ollivier, D., et al. Geographic origins and compositions of virgin olive oils determinated by chemometric analysis of NIR spectra. *Anal. Chim. Acta* 2007, *595*, 136–144.

[19] Alonso-Salces, R. M., Héberger, K., Holland, M. V., Moreno-Rojas, J. M., et al. Multivariate analysis of NMR fingerprint of the unsaponifiable fraction of virgin olive oils for authentication purposes. *Food Chem.* 2010, *118*, 956–965.

[20] Alonso-Salces, R. M., Moreno-Rojas, J. M., Holland, M. V., Reniero, F., et al. Virgin olive oil authentication by multivariate analyses of ^{1}H NMR fingerprints and $\delta^{13}C$ and $\delta^{2}H$ data. *J. Agric. Food Chem.* 2010, *58*, 5586–5596.

[21] Mannina, L., Marini, F., Gobbino, M., Sobolev, A. P., et al. NMR and chemometrics in tracing European olive oils: the case study of Ligurian samples. *Talanta* 2010, *80*, 2141–2148.

[22] Petrakis, P. V., Agiomyrgianaki, A., Christophoridou, S., Spyros, A., et al. 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*, 3200–3207.

[23] Dais, P., Hatzakis, E., Quality assessment and authentication of virgin olive oil by NMR spectroscopy: A critical review. *Anal. Chim. Acta* 2013, *765*, 1–27.

[24] Longobardi, F., Ventrella, A., Napoli, C., Humpfer, E., et al. Classification of olive oils according to geographical origin by using ^{1}H NMR fingerprinting combined with multivariate analysis. *Food Chem.* 2012, *130*, 177–183.

[25] Agiomyrgianaki, A., Petrakis, P. V., Dais, P., Influence of harvest year, cultivar and geographical origin on Greek extra virgin olive oils composition: A study by NMR spectroscopy and biometric analysis. *Food Chem.* 2012, *135*, 2561–2568.

[26] Fragaki, G., Spyros, A., Siragakis, G., Salivaras, E., et al. Detection of extra virgin olive oil adulteration with lampante olive oil and refined olive oil using nuclear magnetic resonance spectroscopy and multivariate statistical analysis. *J. Agric. Food Chem.* 2005, *53*, 2810–2816.

[27] Martínez-Lozano Sinues, P., Alonso-Salces, R. M., Zingaro, L., Finiguerra, A., et al. Mass spectrometry fingerprinting coupled to National Institute of Standards and Technology mass spectral search algorithm for pattern recognition. *Anal. Chim. Acta* 2012, *755*, 28–36.

[28] Aramendía, M. A., Marinas, A., Marinas, J. M., Moreno, J. M., et al. Oxygen-18 measurement of Andalusian olive oils by continuous flow pyrolysis/isotope ratio mass spectrometry. *Rapid Commun. Mass Spectrom.* 2007, *21*, 487–496.

[29] Aramendía, M. A., Marinas, A., Marinas, J. M., Sánchez, E., et al. A nuclear magnetic resonance (^{1}H and ^{13}C) and isotope ratio mass spectrometry ($\acute{\imath}^{13}C$, $\acute{\imath}^{2}H$ and $\acute{\imath}^{18}O$) study of andalusian olive oils. *Rapid Commun. Mass Spectrom.* 2010, *24*, 1457–1466.

[30] Angerosa, F., Breas, O., Contento, S., Guillou, C., et al. Application of stable isotope ratio analysis to the characterization of the geographical origin of olive oils. *J. Agric. Food Chem.* 1999, *47*, 1013–1017.

[31] Commission Regulation (EEC) No 656/95 of 28 March 1995, European Communities 1995, Official Journal of the European Union, L 69, Annex XVII.

[32] Hoffman, R. E., Standardization of chemical shifts of TMS and solvent signals in NMR solvents. *Magn. Reson. Chem.* 2006, *44*, 606–616.

[33] Berrueta, L. A., Alonso-Salces, R. M., Héberger, K., Supervised pattern recognition in food analysis. *J. Chromatogr. A* 2007, *1158*, 196–214.

[34] Kjeldahl, K., Bro, R., Some common misunderstandings in chemometrics. *J. Chemom.* 2010, *24*, 558–564.

[35] Perez-Trujillo, M., Gomez-Caravaca, A. M., Segura-Carretero, A., Fernandez-Gutierrez, A., et al. Separation and identification of phenolic compounds of extra virgin olive oil from Olea europaea L. by HPLC-DAD-SPE-NMR/MS. Identification of a new diastereoisomer of the aldehydic form of oleuropein aglycone. *J. Agric. Food Chem.* 2010, *58*, 9129–9136.

[36] Christophoridou, S., Dais, P., Detection and quantification of phenolic compounds in olive oil by high resolution ^{1}H nuclear magnetic resonance spectroscopy. *Anal. Chim. Acta* 2009, *633*, 283–292.

[37] Rotondo, A., Salvo, A., Giuffrida, D., Dugo, G., et al. NMR analysis of aldehydes in sicilian extra-virgin olive oils by DPFGSE techniques. Atti della Accademia Peloritana dei Pericolanti, Classe di Scienze Fisiche. *Matematiche e Naturali* 2011, *89*, 1–8.

[38] Mannina, L., Segre, A., High resolution nuclear magnetic resonance: From chemical structure to food authenticity. *Grasas y Aceites* 2002, *53*, 22–33.

[39] D'Imperio, M., Mannina, L., Capitani, D., Bidet, O., et al. NMR and statistical study of olive oils from Lazio: A geographical, ecological and agronomic characterization. *Food Chem.* 2007, *105*, 1256–1267.

[40] Alonso-Salces, R. M., Holland, M. V., Guillou, C., ^{1}H-NMR fingerprinting to evaluate the stability of olive oil. *Food Control* 2011, *22*, 2041–2046.

[41] Guillén, M. D., Uriarte, P. S., Study by ^{1}H NMR spectroscopy of the evolution of extra virgin olive oil composition submitted to frying temperature in an industrial fryer for a prolonged period of time. *Food Chem.* 2012, *134*, 162–172.

[42] Araghipour, N., Colineau, J., Koot, A., Akkermans, W., et al. Geographical origin classification of olive oils by PTR-MS. *Food Chem.* 2008, *108*, 374–383.

[43] Sternberg, L. D. S. L., D/H ratios of environmental water recorded by D/H ratios of plant lipids. *Nature* 1988, *333*, 59–61.

[44] Camin, F., Larcher, R., Perini, M., Bontempo, L., et al. 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.

[45] Esbensen, K. H., Guyot, D., Westad, F., Houmøller, L. P., *Multivariate Data Analysis - In Practice: An Introduction to Multivariate Data Analysis and Experimental Design*, Camo Process AS, Oslo 2002.

[46] Hennessy, S., Downey, G., O'Donnell, C. P., Fourier transform infrared spectroscopy and chemometrics: Extra virgin olive oil from Liguria. *J. Agric. Food Chem.* 2009, *57*, 1735–1741.

[47] Caetano, S., Üstün, B., Hennessy, S., Smeyers-Verbeke, J., et al. Geographical classification of olive oils by the application of CART and SVM to their FT-IR. *J. Chemom.* 2007, *21*, 324–334.

[48] Woodcock, T., Downey, G., O'Donnell, C. P., Confirmation of declared provenance of European extra virgin olive oil samples by NIR spectroscopy. *J. Agric. Food Chem.* 2008, *56*, 11520–11525.

[49] Cajka, T., Riddellova, K., Klimankova, E., Cerna, M., et al. Traceability of olive oil based on volatiles pattern and multivariate analysis. *Food Chem.* 2010, *121*, 282–289.

