

## Accepted Manuscript

Characterisation and attempted differentiation of European and extra-European olive oils using stable isotope ratio analysis

Luana Bontempo, Mauro Paolini, Pietro Franceschi, Luca Ziller, Diego L. García-González, Federica Camin

PII: S0308-8146(18)31852-1

DOI: <https://doi.org/10.1016/j.foodchem.2018.10.077>

Reference: FOCH 23738

To appear in: *Food Chemistry*

Received Date: 15 May 2018

Revised Date: 19 September 2018

Accepted Date: 14 October 2018

Please cite this article as: Bontempo, L., Paolini, M., Franceschi, P., Ziller, L., L. García-González, D., Camin, F., Characterisation and attempted differentiation of European and extra-European olive oils using stable isotope ratio analysis, *Food Chemistry* (2018), doi: <https://doi.org/10.1016/j.foodchem.2018.10.077>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



## **Characterisation and attempted differentiation of European and extra-European olive oils using stable isotope ratio analysis**

Luana Bontempo<sup>1\*</sup>, Mauro Paolini<sup>1</sup>, Pietro Franceschi<sup>2</sup>, Luca Ziller<sup>1</sup>, Diego L. García-González<sup>3</sup>,  
Federica Camin<sup>1</sup>

<sup>1</sup>Department of Food Quality and Nutrition, Research and Innovation Centre, Fondazione Edmund Mach, Via E. Mach 1, 38010 San Michele all'Adige, Trento, Italy

<sup>2</sup>Unit of Computational Biology, Research and Innovation Centre, Fondazione Edmund Mach, Via E. Mach 1, 38010 San Michele all'Adige, Trento, Italy

<sup>3</sup>Instituto de la Grasa, Campus University Pablo de Olavide, Sevilla, Spain

### **ABSTRACT**

European law requires a designation of origin for virgin and extra virgin olive oils (at least in terms of EU/non-EU provenance). Stable isotope ratios have been successfully applied to determine the geographical origin of olive oils, but never to distinguish EU and non-EU oils. In this study  $^2\text{H}/^1\text{H}$ ,  $^{13}\text{C}/^{12}\text{C}$  and  $^{18}\text{O}/^{16}\text{O}$  ratios were analysed in bulk olive oils using Isotope Ratio Mass Spectrometry (IRMS) as well as  $^{13}\text{C}/^{12}\text{C}$  and  $^2\text{H}/^1\text{H}$  in the four main fatty acids (linoleic, oleic, palmitic and stearic acids) using IRMS coupled with GC. The isotopic composition of olive oils was successfully used to distinguish samples originating in the two areas. Specifically, when bulk data were combined with fatty acid isotopic data the differentiation power of the method improved clearly. This separation is due to the specific isotopic fingerprint of the individual countries making up the EU and non-EU samples.

Keywords: olive oil, geographical origin, C, H and O stable isotopes, fatty acids

\* Correspondence to: Luana Bontempo, Food Quality and Nutrition Department, Research and Innovation Centre, Fondazione Edmund Mach, Via E. Mach 1, 38010 San Michele all'Adige (TN), Italy. Tel: +39 461 615138

## 1. INTRODUCTION

Extra virgin olive oil (EVOO) is a key product in the European Union (EU) market, since Europe is the main olive oil exporter and consumer worldwide and around 98% of the world's olive trees are concentrated in the Mediterranean Basin. This product enjoys global recognition due to its nutritional value and beneficial effects on health (Tripoli, Giammanco, Tabacchi, Di Majo, Giammanco, & La Guardia, 2005), however due to increasing fraud scandals, consumers have lost confidence in the authenticity of EVOO. To protect ethical producers and consumers, European law requires that the origin of certain premium products, such as EVOOs and virgin olive oils (VOOs), must be declared on the label. These two types of products indeed have extremely restrictive rules about what should be declared on the label (EU Regulations 29/2012 and EU Reg. 1335/2013) and they must bear a designation of origin. Specifically, it must be stated whether the oil originates in a specific European Member State or a non-EU country, and whether it is a blend of European Community and/or non-EU olive oils, as appropriate.

Official olive oil (OO) quality control methods are based on quantitative analysis of specific chemical compounds, *e.g.* fatty acids, sterols, alcohols or stigmastadiene (EEC Reg. 2568/91 and amendments), but these parameters do not allow verification of the geographical origin of EVOOs. Indeed, at the time no official analytical method could verify whether a sample had been produced in the EU or outside the EU and there is increasing demand for the development of an analytical approach capable of effectively verifying claims of origin (García-González, & Aparicio, 2010; Tena, Aparicio-Ruiz, Koidis, & Garcia-Gonzalez, 2018).

Stable isotope ratio analysis (SIRA) offers one of the most promising tools for establishing the authenticity of premium products. Indeed, the content of stable isotopes of bio-elements (H, C, N, O and S) reflects local agricultural practices and geo-climatic characteristics and can be experimentally determined with isotope ratio mass spectrometry (IRMS) (Laursen, Bontempo, Camin, & Roßmann, 2016). In the last two decades, different studies have been published on the isotopic composition of bulk OOs and its correlation with the geo-climatic characteristics of the production area, and it is nowadays widely accepted that SIRA can be a powerful tool for tracing the geographical origin of EVOOs (Camin et al., 2010a; Camin et al., 2010b; Camin et al., 2016; Chiocchini, Portarena, Ciolfi, Brugnoli, & Lauteri, 2016; Portarena, Gavrichkova, Lauteri, & Brugnoli, 2014). Specifically, it has been shown that  $^{13}\text{C}/^{12}\text{C}$ ,  $^{18}\text{O}/^{16}\text{O}$  and  $^2\text{H}/^1\text{H}$  ratios change according to latitude and altitude, suggesting distance from the sea, environmental and climatic conditions as co-factors of variability (Bréas, Guillou, Reniero, Sada, & Angerosa, 1998; Angerosa

et al., 1999; Aramendía et al., 2007; Bontempo et al., 2009; Iacumin, Bernini, & Boschetti, 2009; Aramendía et al., 2010).

In order to obtain additional information for authentication purposes, the isotope composition of Olive oils can be measured not only in bulk samples but also in sub-components (Camin & Bontempo, 2017). OO is mainly composed of triglycerides (about 99%) and small quantities of free fatty acids (FA), glycerol, phosphatides and other compounds (León-Camacho, Morales, & Aparicio, 2013). The triglyceride molecule is an ester derived from glycerol and three fatty acids. The main FAs in olive oil triglycerides are oleic acid (55% to 83% of olive oil), linoleic acid (3.5% to 21% of olive oil), palmitic acid (7.5% to 20%) and stearic acid (0.5% to 5%). The composition of FAs varies widely depending on the variety, maturity of the fruit, geo-climatic conditions of the production location, and several other factors. Despite the great potential of FA triglyceride isotopic analysis, few studies have dealt with the  $\delta^{13}\text{C}$  composition of OO FAs (Spangenberg & Ogrinc, 2001). Two studies in particular focused mainly on the use of these parameters to investigate EVOO adulteration with cheaper or refined oils (Woodbury, Evershed, & Barry Rossell, 1998; Spangenberg, Macko, & Hunziker, 1998). A further two studies were aimed mainly at geographical characterisation of Slovenian (Spangenberg & Ogrinc, 2001) and Italian OOs (Faber et al., 2014). In this last study, despite the wide variability of pedoclimatic characteristics in Italy, it was possible to distinguish three Italian macro areas by combining the  $\delta^{13}\text{C}$  of bulk samples with the  $\delta^{13}\text{C}$  of the four main FAs.

To our knowledge, no study concerning the  $\delta^2\text{H}$  composition of FAs in either OOs or other edible vegetal oils has been published to date, but recently a method for determining this parameter was developed as a potential tool for this purpose (Paolini, Bontempo, & Camin, 2017). It can be reasonably expected that the combination of bulk isotopic data with the  $\delta^{13}\text{C}$  and  $\delta^2\text{H}$  of fatty acids would allow the production of a “profile” that can better characterise the specific product and as a consequence, also its geographical origin. A similar study, based on the determination of  $\text{d}2\text{H}$  and  $\text{d}13\text{C}$  of fatty acids, was successfully applied for the traceability of milk and milk powders according to their regional origin in New Zealand (Ehtesham, Hayman, McComb, Van Hale, & Frew, 2013).

In this study we investigated the potential of SIRA to distinguish EVOOs collected in the main producer countries within and outside Europe. The study was conducted by analysing around a hundred EVOO samples.  $^2\text{H}/^1\text{H}$ ,  $^{13}\text{C}/^{12}\text{C}$  and  $^{18}\text{O}/^{16}\text{O}$  ratios were analysed in bulk oils, as well as  $^{13}\text{C}/^{12}\text{C}$  and for the first time  $^2\text{H}/^1\text{H}$  determined in the four main FAs (linoleic, oleic, palmitic and stearic acids).

## **2. MATERIALS AND METHODS**

### **2.1 Sampling**

In this study a total of 101 authentic EVOOs from different cultivars were examined, sampled during 2015-2016. The olive oil samples were collected worldwide in the main olive oil-producing regions of Argentina (1), Australia (9), France (6), Greece (8), Italy (28), Morocco (1), Peru (1), Portugal (12), Spain (13), Tunisia (12), Turkey (4), Uruguay (2) and the USA (4). They were collected in 30 ml dark glass bottles and kept at 4°C until analysis.

### **2.2 Isotopic analysis of bulk EVOO**

The analysis of bulk EVOO was performed in duplicate using an isotope ratio mass spectrometer (visION, Isoprime Ltd, UK) coupled with an elemental analyser (Vario Isotope Cube, Elementar Analysensysteme GmbH, Germany) for  $^{13}\text{C}/^{12}\text{C}$  measurement. For  $^2\text{H}/^1\text{H}$  and  $^{18}\text{O}/^{16}\text{O}$  measurement, an isotope ratio mass spectrometer (Finnigan DELTA XP, Thermo Scientific, Bremen, Germany) was used, coupled with a pyrolyser (Finnigan  $^{\text{TM}}$ TC/EA, high temperature conversion elemental analyzer, Thermo Scientific) equipped with an autosampler (Finnigan AS 200, Thermo Scientific) and interfaced with the IRMS through a dilutor (Conflo III, Thermo Scientific), dosing the sample and reference gases. To separate the gases, the pyrolyser was supplied with a Molecular Sieve 5A (0.6 m) GC column.

Aliquots of 0.2 - 0.3 mg of sample were weighed into tin capsules to determine  $^{13}\text{C}/^{12}\text{C}$ , and silver capsules for quantification of  $^{18}\text{O}/^{16}\text{O}$  and  $^2\text{H}/^1\text{H}$ . For  $^{18}\text{O}/^{16}\text{O}$  and  $^2\text{H}/^1\text{H}$  analysis, the weighed samples were stored in a desiccator above  $\text{P}_2\text{O}_5$  for at least 4 days before analysis, then put into the auto-sampler equipped with a suitable cover. During measurement, dryness was guaranteed by flushing nitrogen continuously over the samples. Before measuring  $^2\text{H}/^1\text{H}$ , the  $\text{H}_3$  factor, which allows correction of the contribution of  $[\text{H}_3]^+$  to the  $m/z$  3 signal, was verified to be lower than 8.

### **2.3 Isotopic analysis of individual FAMES using Gas Chromatography-Combustion\Pyrolysis-Isotope Ratio Mass Spectrometry (GC-C\Py-IRMS)**

For the preparation of FAMES via the transesterification of triglycerides, 0.1g of OO sample was weighed into a 10 mL vial with 4 mL of hexane, then 1 mL of 2 M methanolic sodium hydroxide

solution was added and the vial shaken for 1 min at room temperature (Paolini, Bontempo, & Camin, 2017). The mixture was allowed to stratify until the upper layer became clear and 1 mL of the hexane solution was filtered and then injected into the GC-C\Py-IRMS.

Individual FAME isotopic analysis was carried out in triplicate using Trace GC Ultra (GC IsoLink + ConFlo IV, Thermo Scientific) interfaced with an isotope ratio mass spectrometer (DELTA V, Thermo Scientific), through an open split interface and with a single-quadrupole GC-MS (ISQ Thermo Scientific) to identify the compounds. 1.0  $\mu$ L of each sample was injected in split mode (1:10) with an auto-sampler (Triplus, Thermo Scientific) and a BPX-70 capillary column (60 m  $\times$  0.32 mm i.d.  $\times$  0.25  $\mu$ m film thickness; SGE) with He as carrier gas (at a flow of 1 mL/min) was used (Paolini, Bontempo, & Camin, 2017).

To monitor instrumental performance, heneicosane was added to each sample as an internal standard and its  $\delta^{13}\text{C}$  and  $\delta^2\text{H}$  values checked. The carbon and hydrogen isotopic values of pure heneicosane (-28.8‰ and -190‰, respectively) were determined with EA-IRMS and TC/EA-IRMS, and the differences with GC-C\Py-IRMS values were at most  $\pm 0.2\%$  and  $\pm 2.0\%$  respectively for  $\delta^{13}\text{C}$  and  $\delta^2\text{H}$ . More details of the analytical methods are carefully described in Paolini et al. (2017).

## 2.4 Data analysis and corrections

The isotopic values were expressed according to equation 1:

$$\delta_i E = \frac{(i R_{SA} - i R_{REF})}{i R_{REF}} \quad (1)$$

where  $i$  is the mass number of the heavier isotope of element  $E$  (i.e.  $^2\text{H}$ ,  $^{13}\text{C}$ ,  $^{18}\text{O}$ ),  $R_{SA}$  is the respective isotope ratio of a sample (for H: number of  $^2\text{H}$  atoms/number of  $^1\text{H}$  atoms or as approximation  $^2\text{H}/^1\text{H}$ , for C: number of  $^{13}\text{C}$  atoms/number of  $^{12}\text{C}$  atoms or as approximation  $^{13}\text{C}/^{12}\text{C}$ , for O: number of  $^{18}\text{O}$  atoms/number of  $^{16}\text{O}$  atoms or as approximation  $^{18}\text{O}/^{16}\text{O}$ ), and  $R_{REF}$  is the relevant internationally recognised reference material. The delta values are multiplied by 1000 and expressed in units “per mil” (‰).

The  $\delta^{13}\text{C}$  values were reported relative to Vienna-Pee Dee Belemnite on a scale normalised by assigning a value of -46.6‰ to LSVEC lithium carbonate (IAEA International Atomic Energy Agency, Vienna, Austria) and +1.95‰ to NBS 19 (IAEA).  $\delta^2\text{H}$  and  $\delta^{18}\text{O}$  values were normalised relative to the Vienna Standard Mean Ocean Water–Standard Light Antarctic Precipitation (IAEA) standard scale with consensus values of 0‰ and -428‰.

The  $\delta^{13}\text{C}$  and  $\delta^2\text{H}$  values determined in FAMEs were calculated against two international reference materials by building a linear relationship, injected separately before and after each analytical run: Icosanoic Acid Methyl Esters USGS70 ( $\delta^{13}\text{C}$  value:  $-30.53\text{‰}$ ,  $\delta^2\text{H}$  value:  $-183.9\text{‰}$ ) and USGS71 ( $\delta^{13}\text{C}$  value:  $-10.5\text{‰}$ ,  $\delta^2\text{H}$  value:  $-4.9\text{‰}$ ) (Paolini, Bontempo, & Camin, 2017). Similarly, the  $\delta^{13}\text{C}$ ,  $\delta^2\text{H}$  and  $\delta^{18}\text{O}$  values of bulk samples were calculated against USGS70 and USGS71 (for  $\delta^{13}\text{C}$  and  $\delta^2\text{H}$ ), and the benzoic acids IAEA-601 and IAEA-602 (IAEA) (for  $\delta^{18}\text{O}$ ) international reference material.

## 2.5 Statistical analysis

The data were analysed with the Statistica version 9 (StatSoft Inc., Tulsa, OK, USA) package and in the R statistical environment (R Core Team, 2014).

All plots and data management were performed using tidyverse (Wickham, 2017) and cowplot (Wilke, 2017). PCA analysis was performed and visualised using FactoMineR (Le, Josse, & Husson, 2008) and factoextra (Kassambara & Mundt, 2017) packages. Random Forest classification (Hastie, Tibshirani, & Friedman, 2001) was performed with the randomForest package (Liaw & Wiener, 2002) and the validation scheme was implemented in caret (Kuhn et al., 2018).

Briefly, the complete dataset was split into training and test sets (1/1), maintaining the class balance (stratified split). The RF model was optimised by 5-fold repeated cross validation (3 times) on the training set and its performance was evaluated on the test set in terms of receiver operating characteristic curves (ROCs), calculated with the pROC package (Robin et al., 2011). ROC curves are commonly used to evaluate the trade-off between sensitivity and specificity as a function of the probability cut-off for class assignment. A model unable to classify will result in a ROC curve lying on the diagonal line, while a perfect classifier will show a vertical increase with a knee at the [0,1] point. The area under each ROC curve (AUC) is commonly used to assess the performance of the classifier. Random classification results in an AUC close to 0.5. A perfect classifier is associated with an AUC of 1. In order to estimate the variability of prediction, the overall classification scheme was repeated 100 times, resulting in 100 ROC curves. As an additional validation step, the classification pipeline was also applied with 100 random shuffles of class labels.

Furthermore, to investigate which variables are more important in separating the two groups of samples we used the Gini index. Indeed, Gini index (Wehrens, 2001) is commonly used in decision tree learning as a measure of "purity" of classification. The change of this index when a variable is included into the model is also used as a measure of the classification importance.

### 3. RESULTS AND DISCUSSION

#### 3.1 Bulk analysis of olive oils

Figure 1 shows the distribution of the  $\delta^{13}\text{C}$ ,  $\delta^{18}\text{O}$  and  $\delta^2\text{H}$  values determined in all the bulk olive oil samples, grouped according to the country of production. In the strip plots the EU countries were listed according to geographical longitude, from Portugal to Greece. The plot clearly shows that the sampling was not balanced for the different countries, which results in different class numbers for EU and extra-EU samples. The extra-EU samples were collected from a larger number of countries, but the total number of samples was lower than for the EU (67 versus 34 samples).  $\delta^{13}\text{C}$  values measured in bulk samples seemed to vary in Europe from Portugal to Greece according to longitude, whereas  $\delta^{18}\text{O}$  and  $\delta^2\text{H}$  values showed an inverse trend. The behaviour of these last two parameters can be explained by the 'continental effect'. Indeed, most water evaporation occurs in sub-tropical areas of oceans, and due to the increasing distance from the primary source of vapour, a decreasing  $\delta^{18}\text{O}$  and  $\delta^2\text{H}$  value can be expected from Portugal to Greece, due to depletion related to previous precipitation (Clark, & Fritz, 1997). In the case of  $\delta^{13}\text{C}$ , the inverse trend related to longitude is probably related to the combination of regional climatic characteristics such as mean temperature, relative humidity and drought stress suffered by plants (O'Leary, 1995).

In the data obtained, it is worth noting that some countries have particular values for one or more isotope ratios. Australian samples showed particularly low  $\delta^{13}\text{C}$  values and high  $\delta^2\text{H}$  values. The high  $\delta^2\text{H}$  values were expected, because of the climatic characteristics of this country, affecting the  $\delta^2\text{H}$  values of water (waterisotope.org), the only source from which hydrogen lipids originate (Schmidt, Werner, & Eisenreich, 2003). The particularly low  $\delta^{13}\text{C}$  values (from -29.4‰ to -31.6‰), on the other hand, can be explained by the fact that the orchards were irrigated in order to produce consistently high yields of high quality fruits (Australian Government, Guide to Efficient Olive Harvesting, 2008) and therefore did not suffer from drought. In contrast, Uruguayan and Peruvian samples showed particularly low values for both  $\delta^2\text{H}$  and  $\delta^{18}\text{O}$ . This result could be rationalised by taking into account the fact that in both these countries most of the rivers that flow through the two regions come from mountain areas (the Andes for Peru, the Andes and southern Brazilian mountains for Uruguay), probably bringing water with lower  $\delta^2\text{H}$  and  $\delta^{18}\text{O}$  values originating at high altitude. North American samples showed particularly low values for  $\delta^2\text{H}$ , but not for  $\delta^{18}\text{O}$ , the latter instead being consistent with the average values of the other samples. The different behaviour of the two isotopic parameters can be explained by the fact that the hydrogen in plant lipids originates solely from source water, and therefore has an isotopic composition more closely



correlated with that of rainfall water than the oxygen in lipids, which also derives from atmospheric CO<sub>2</sub> and O<sub>2</sub> (Schmidt, Werner, & Rossmann, 2001; Schmidt, Werner, & Eisenreich, 2003).

Despite the particular characteristics of the isotopic values of some countries, a general and similar pattern for  $\delta^2\text{H}$  and  $\delta^{18}\text{O}$  is evident throughout the different countries, verified by correlation analysis (Figure 2).  $\delta^{18}\text{O}$  and  $\delta^2\text{H}$  values in bulk olive oil were significantly correlated, as occurs in water (Clark & Fritz, 1997). The calculated correlation equation ( $\delta^2\text{H}_{\text{bulk}} = 3.09 * \delta^{18}\text{O}_{\text{bulk}} - 226$ ,  $p < 0.001$ ,  $R^2 = 0.5$ ) is similar to those reported in other studies (Bontempo et al., 2009; Camin et al., 2010a; Camin et al., 2010b). No significant correlation was found between  $\delta^{13}\text{C}$  and the two other isotopic ratios.

### 3.2 Fatty acids analysis in olive oils

Figure 1 shows the strip plots for the  $\delta^2\text{H}$  and  $\delta^{13}\text{C}$  values determined in linoleic, oleic, palmitic and stearic acids extracted from olive oils. The  $\delta^{13}\text{C}$  values measured in the FAs of EU olive oils, according to the  $\delta^{13}\text{C}$  determined in bulk samples, seem to vary in relation to the longitude with a positive gradient, although with a less pronounced trend. Similarly to the  $\delta^2\text{H}$  of bulk oils, the  $\delta^2\text{H}$  of FAs from EU olive oils showed a negative trend according to longitude. However, the trend for the two isotopic ratios was less marked for fatty acids than for bulk samples, indicating that they are affected also by other and different factors. For this reason, the level of correlation between the different isotopic parameters measured in bulk olive oil samples and fatty acids was determined and the relevant correlation matrix is shown in Figure 2. A positive correlation was observed between the  $\delta^{13}\text{C}$  determined in bulk samples and the  $\delta^{13}\text{C}$  of the four fatty acids. A similar pattern can be noted for the  $\delta^2\text{H}$  of bulk and fatty acids. Specifically, the strongest correlations were found between the  $\delta^{13}\text{C}$  of bulk samples and the  $\delta^{13}\text{C}$  of oleic acid and between the  $\delta^2\text{H}$  of bulk samples and the  $\delta^2\text{H}$  of oleic acid, in line with the fact that oleic acid is the main fatty acid in olive oil. The correlation was also significant when considering the  $\delta^{18}\text{O}$  value of bulk samples and the  $\delta^2\text{H}$  values of the four fatty acids, as expected, due to the correlation previously highlighted between the  $\delta^2\text{H}$  and  $\delta^{18}\text{O}$  of bulk samples.

The  $\delta^{13}\text{C}$  values determined in the four fatty acids were positively correlated with each other. In particular, the  $\delta^{13}\text{C}$  of palmitic acid correlated with the  $\delta^{13}\text{C}$  of stearic, oleic and linoleic acids, furthermore the  $\delta^{13}\text{C}$  of stearic acid correlated with the  $\delta^{13}\text{C}$  of oleic and linoleic acids. The strongest correlation was found between the  $\delta^{13}\text{C}$  values of oleic acid and linoleic acid. As regards  $\delta^2\text{H}$ , this parameter determined in palmitic acid was correlated with the  $\delta^2\text{H}$  of linoleic acid, and strongly correlated with the  $\delta^2\text{H}$  of stearic and oleic acids. The  $\delta^2\text{H}$  of stearic acid was strongly

correlated with the  $\delta^2\text{H}$  of oleic acid and to a lesser extent with the  $\delta^2\text{H}$  of linoleic acid. A strong correlation was also found between the  $\delta^2\text{H}$  of oleic and linoleic acids.

Figure 3 shows the dispersion and mean values for the  $^{13}\text{C}/^{12}\text{C}$  and  $^2\text{H}/^1\text{H}$  ratios measured in palmitic, stearic, oleic and linoleic acids, across all 101 extra-virgin olive oil samples. The  $\delta^{13}\text{C}$  trend in fatty acids was in agreement with the results obtained when analysing vegetable oils reported in other studies (Royer, Gerard, Nault, Lees, & Martin, 1999; Spangenberg & Ogrinc, 2001; Faberi et al., 2014). The variability observed in the carbon isotopic composition of the four fatty acids considered is linked to intrinsic isotopic effects along the synthesis pathway. A multienzyme complex catalyses the reaction sequences of chain elongation, through the addition of acetyl groups (e.g. from palmitic to stearic acid) and desaturation of fatty acids (e.g. from stearic to oleic and linoleic acid) (Spangenberg & Ogrinc, 2001). It seems that there is an initial isotopic fractionation occurring from palmitic to stearic acid related to the first elongation step, with generally lower  $\delta^{13}\text{C}$  values in stearic acid. On the other hand, there is a second phase of isotopic fractionation with the opposite effect, with a first desaturation step leading to an increase of  $\delta^{13}\text{C}$  values in oleic acid compared to stearic acid. The second desaturation step leads to a slight decrease in the  $\delta^{13}\text{C}$  values of linoleic acid, probably related to the fact that this step happens in chloroplasts and no longer in the cytosol, as takes place for the other three fatty acids.

The  $\delta^2\text{H}$  values obtained for the four fatty acids (Figure 3) fall into the natural field of variability of lipids in vegetable matrices. Lipids from living plants can be divided into two groups: saponifiable lipids and nonsaponifiable lipids (fatty acids) with  $\delta^2\text{H}$  values between -80‰ and -250‰ (Estep & Hoering, 1980). The hydrogen isotope value of fatty acids is influenced by the isotopic composition of the primary hydrogen source (source water through leaf water), the characteristics of precursors and the biosynthetic pathway (Schmidt, Werner, & Eisenreich, 2003). The relationship between the  $\delta^2\text{H}$  values of fatty acids is the opposite of the situation found for  $\delta^{13}\text{C}$  in the first three (palmitic, stearic and oleic acids), whereas in the case of the desaturation step from oleic to linoleic acid, we found a decrease in the  $\delta^2\text{H}$  values and  $\delta^{13}\text{C}$  values.

### **3.3 Combination of bulk and fatty acids data**

Principal component analysis (PCA) was used to display the samples in an unsupervised map (score plot) to check whether differentiation of geographical origin was the main component in dataset variability when using all the isotopic parameters determined in olive oil samples (bulk and FAs). In Figure 4a, the samples are 'labelled' according to the country of origin, whereas in Figure 4b the different colours indicate EU and non-EU provenance. The first two factors of PCA explained 61.1% of variance in the 11 original variables. The first component (36.8% of variance) was mainly

positively loaded by the  $\delta^{13}\text{C}$  of bulk and FAs, whereas the second component (24.3% of variance) was positively loaded by the  $\delta^2\text{H}$  and  $\delta^{18}\text{O}$  of bulk and the  $\delta^2\text{H}$  of FAs. As previously noted, many of the variables are correlated with each other, in particular all the  $\delta^2\text{H}$  isotopic ratios are correlated with each other and with  $\delta^{18}\text{O}$ , which is the main contribution to separation along PC1. On the other hand, the  $\delta^{13}\text{C}$  of bulk and fatty acids contributed primarily to separation along PC2. The PCA shows partial separation of EU and non-EU sample groups along the diagonal, but it is clear that it is not possible to clearly distinguish between them (Figure 4b). The higher heterogeneity of non-EU samples is confirmed by the PCA plot, in line with the data that can be observed in the strip plots in Figure 1. However, it is worth noting that despite the overlapping of EU and non-EU samples, there is a rough organisation of samples according to the country of origin (Figure 4a). This is particularly clear for Australian, American and Tunisian samples, but a tendency to cluster is also noticeable for Spanish, Portuguese, Italian and French samples. It is therefore reasonable to presume that separation between the two EU and non-EU macro areas is actually due to the different characteristics of the specific countries of origin.

A more powerful classification approach with Random Forest was applied to investigate the potential of the set of isotopic parameters to discriminate between EU and non-EU samples, and the results are summarised in Figure 5.

Figure 5a shows the ROCs and the AUCs obtained in three different scenarios. In the first case, only bulk stable isotope ratios were considered (green curve). In the second scenario, the data obtained for the FA ( $\delta^{13}\text{C}$  and  $\delta^2\text{H}$ ) were included in the classifier (purple curve). The last set of curves was obtained for the validation run with shuffled class labels (yellow lines).

The plot clearly indicates that all the classifiers with the correct class labels performed markedly better than randomly, showing that isotopic data can be used to distinguish EU and extra-EU samples with good accuracy. Notably, the inclusion of FA data in the classifier resulted in a significant improvement in performance, visible both in the ROC and the AUC.

As already anticipated in the discussion of PCA analysis, it would be interesting to understand if the separation between EU and extra-EU can be really be interpreted on the basis of differentiation between different countries of origin. EU and extra-EU labels can indeed be seen as specific “administrative” macro areas that separate potentially similar nations such as Tunisia and Portugal. In order to investigate this aspect, the previously described classification approach was applied with a “random region” validation scheme, in which the different nations were randomly assigned to EU and extra-EU macro areas. The results are shown in Figure 5b. The gray ROC curves graphically display the performance of the different classifiers. It is clear that even with randomly assigned countries, the isotopic data were able to correctly classify the test samples almost all the time. The

variability in efficiency was wide, but in general we are far from the behaviour of a fully randomised classifier (yellow lines in Figure 5a). These results strongly suggest that the driving force of EU and extra-EU separation is not the specific “isotopic” characteristics of EU samples, but rather the combination of geographical patterns which characterise the different countries in terms of the geographical, climatic, pedological and geological features of the place where the olive trees were grown.

The proposed machine learning approach has shown strong potential for the classification of EVOOs on the basis of their isotopic fingerprint. Furthermore, to clarify which variables are more important in separating the two groups of samples we determined the Gini index. In the case of Random Forest, and for all tree-based approaches, the decrease in the Gini Index can be used to evaluate to what extent the classification “purity” is reduced when a variable is removed from the classifier. This type of information is presented in Figure 6. Specifically, this analysis considers the mean decrease in accuracy related to the proportion of observations that are incorrectly classified by removing a specific variable from the model. What emerges from Figure 6 is that the variables with the strongest effect on classification accuracy are (in order of importance) the  $\delta^{13}\text{C}$  of linoleic acid, followed by the  $\delta^2\text{H}$  and  $\delta^{13}\text{C}$  of oleic acid. These three parameters have a much higher classification capacity than the  $\delta^{13}\text{C}$ ,  $\delta^{18}\text{O}$  and  $\delta^2\text{H}$  of bulk samples, explaining the better ‘success rate’ of models including FA data. Furthermore, it is interesting to note that the most influential variables are the isotopic values of linoleic and oleic acids, which are the last acids biosynthesized by the plant. We can therefore surmise that the characteristics of the place of growth (e.g. water availability, temperature range between day and night, etc.) may also affect the biosynthetic pathway of fatty acids, and in particular the isotopic values of the two last biosynthesized fatty acids, in some way.

#### 4. CONCLUSION

The  $\delta^{13}\text{C}$ ,  $\delta^2\text{H}$  and  $\delta^{18}\text{O}$  determined in bulk samples showed a particular isotopic fingerprint for some countries, but were not capable of completely separating EU and non-EU oils. Therefore, to try and improve the classification capability of the method,  $\delta^{13}\text{C}$  and  $\delta^2\text{H}$  were determined in the four main FAs extracted from olive oils. These parameters showed a particular trend for the biosynthetic pathway. Specifically, for  $\delta^{13}\text{C}$  there was an initial decrease in the values from palmitic to stearic acids, followed by a major increase in oleic acid and a final decrease in linoleic acid. A different trend was found for  $\delta^2\text{H}$ , with an initial marked increase in the values in palmitic and stearic acids and two following decreasing steps going from oleic to linoleic acids.

The combination of isotopic data from bulk samples and those from FAs enabled a meaningful improvement in the differentiation capacity of the method, confirming that isotopic analysis of specific compounds can open up promising new prospects for food traceability and authenticity. At a more detailed level, our results showed that separation was not attributable to the different characteristics of EU and non-EU samples, but most probably to the specific isotopic fingerprint of the individual countries and their particular climatic, geographical and pedo-geological characteristics. Starting from these initial results, it would be desirable for the study to be expanded in the future, considering a larger number of samples and possible combination with other analytical techniques.

## ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement No. 613688.

## REFERENCES

- Angerosa, F., Bréas, O., Contento, S., Guillou, C., Reniero, F., & Sada, E. (1999). Application of stable isotope ratio analysis to the characterization of the geographical origin of olive oils. *Journal of Agricultural and Food Chemistry*, 47(3), 1013–1017.
- Aramendía, M. A., Marinas, A., Marinas, J. M., Moreno, J. M., Moalem, M., Rallo, L., & Urbano, F. J. (2007). Oxygen-18 measurement of Andalusian olive oils by continuous flow pyrolysis/isotope ratio mass spectrometry. *Rapid Communications in Mass Spectrometry*, 21(4), 487–496.
- Aramendía, M. A., Marinas, A., Marinas, J. M., Sánchez, E., Urbano, F. J., Guillou, C., ... Rallo, L. (2010). A nuclear magnetic resonance ( $^1\text{H}$  and  $^{13}\text{C}$ ) and isotope ratio mass spectrometry ( $\delta^{13}\text{C}$ ,  $\delta^2\text{H}$  and  $\delta^{18}\text{O}$ ) study of Andalusian olive oils. *Rapid Communications in Mass Spectrometry*, 24(10), 1457–1466.
- Bontempo, L., Camin, F., Larcher, R., Nicolini, G., Perini, M., & Rossmann, A. (2009). Coast and

- year effect on H, O and C stable isotope ratios of Tyrrhenian and Adriatic Italian olive oils. *Rapid Communications in Mass Spectrometry*, 23(7), 1043–1048.
- Bréas, O., Guillou, C., Reniero, F., Sada, E., & Angerosa, F. (1998). Oxygen-18 measurement by continuous flow pyrolysis/isotope ratio mass spectrometry of vegetable oils. *Rapid Communications in Mass Spectrometry*, 12(4), 188–192.
- Camin, F., Larcher, R., Nicolini, G., Bontempo, L., Bertoldi, D., Perini, M., ... Hoogewerff, J. (2010). Isotopic and elemental data for tracing the origin of European olive oils. *Journal of Agricultural and Food Chemistry*, 58(1), 570–577.
- Camin, F., Larcher, R., Perini, M., Bontempo, L., Bertoldi, D., Gagliano, G., ... Versini, G. (2010). Characterisation of authentic Italian extra-virgin olive oils by stable isotope ratios of C, O and H and mineral composition. *Food Chemistry*, 118(4), 901–909.
- Camin, F., Pavone, A., Bontempo, L., Wehrens, R., Paolini, M., Faberi, A., ... Mannina, L. (2016). The use of IRMS, <sup>1</sup>H NMR and chemical analysis to characterise Italian and imported Tunisian olive oils. *Food Chemistry*, 196, 98–105.
- Camin, F., & Bontempo, L. (2017). Edible Vegetal Oils. In *Food Forensics Stable Isotopes as a Guide to Authenticity and Origin* (pp. 257–272).
- Chiocchini, F., Portarena, S., Ciolfi, M., Brugnoli, E., & Lauteri, M. (2016). Isoscapes of carbon and oxygen stable isotope compositions in tracing authenticity and geographical origin of Italian extra-virgin olive oils. *Food Chemistry*, 202, 291–301.
- Clark, I. D., & Fritz, P. (1997). Precipitation. In: *Environmental Isotopes in Hydrogeology* (pp. 67–70).
- Estep, M. F., & Hoering, T. C. (1980). Biogeochemistry of the stable hydrogen isotopes. *Geochimica et Cosmochimica Acta*, 44(8), 1197–1206.
- Ehtesham, E., Hayman, A.R., McComb, K.A., Van Hale, R., & Frew R.D. (2013). Correlation of geographical location with stable isotope values of hydrogen and carbon of fatty acids from New Zealand milk and bulk milk powder. *Journal of Agricultural and Food Chemistry*, 61,

8914–8923.

Faberi, A., Marianella, R. M., Fuselli, F., La Mantia, A., Ciardiello, F., Montesano, C., ...

Compagnone, D. (2014). Fatty acid composition and  $\delta^{13}\text{C}$  of bulk and individual fatty acids as marker for authenticating Italian PDO/PGI extra virgin olive oils by means of isotopic ratio mass spectrometry. *Journal of Mass Spectrometry*, 49(9), 840–849.

García-González, D.L., & Aparicio, R. (2010). Research in olive oil: Challenges for the near future. *Journal of Agricultural and Food Chemistry*, 58, 12569–12577.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The Elements of Statistical Learning*.

Iacumin, P., Bernini, L., & Boschetti, T. (2009). Climatic factors influencing the isotope composition of Italian olive oils and geographic characterisation. *Rapid Communications in Mass Spectrometry*, 23(3), 448–454.

Kassambara, A. & Mundt, F. (2017). Factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package version 1.0.5. <https://CRAN.R-project.org/package=factoextra>

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., the R Core Team, Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., & Hunt, T. (2018). caret: Classification and Regression Training. R package version 6.0-79. <https://CRAN.R-project.org/package=caret>

Laursen, K. H., Bontempo, L., Camin, F., & Roßmann, A. (2016). Advances in Isotopic Analysis for Food Authenticity Testing. In *Advances in Food Authenticity Testing* (pp. 227–252).

Le, S., Josse, J., & Husson, F. (2008). FactoMineR: An R Package for Multivariate Analysis. *Journal of Statistical Software*, 25(1), 1-18.

León-Camacho, M., Morales, M.T., Aparicio, R. (2013). Chromatographic Methodologies: Compounds for Olive Oil Traceability Issues. In: *Handbook of Olive Oil* (pp. 163–217).

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News* 2(3), 18-22.

O’Leary, M. (1995). Environmental effects on carbon isotope fractionation in terrestrial plants. In:

*Stable isotopes in the biosphere* (pp. 78-91).

- Paolini, M., Bontempo, L., & Camin, F. (2017). Compound-specific  $\delta^{13}\text{C}$  and  $\delta^2\text{H}$  analysis of olive oil fatty acids. *Talanta*, *174*, 38–43.
- Portarena, S., Gavrichkova, O., Lauteri, M., & Brugnoli, E. (2014). Authentication and traceability of Italian extra-virgin olive oils by means of stable isotopes techniques. *Food Chemistry*, *164*, 12–16.
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez J.-C., & Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, *12*, p.77.
- Royer, A., Gerard, C., Naulet, N., Lees, M., & Martin, G. J. (1999). Stable isotope characterization of olive oils. I—Compositional and carbon-13 profiles of fatty acids. *Journal of the American Oil Chemists' Society*, *76*(3), 357–363.
- Schmidt, H.-L., Werner, R.A., Rossmann, A. (2001). 18O Pattern and biosynthesis of natural plant products. *Phytochemistry*, *58*, 9–32.
- Schmidt, H.-L., Werner, R. A., & Eisenreich, W. (2003). Systematics of  $^2\text{H}$  patterns in natural compounds and its importance for the elucidation of biosynthetic pathways. *Phytochemistry Reviews*, *2*(1-2), 61–85.
- Spangenberg, J. E., Macko, S. A., & Hunziker, J. (1998). Characterization of Olive Oil by Carbon Isotope Analysis of Individual Fatty Acids: Implications for Authentication. *Journal of Agricultural and Food Chemistry*, *46*(10), 4179–4184.
- Spangenberg, J. E., & Ogrinc, N. (2001). Authentication of vegetable oils by bulk and molecular carbon isotope analyses with emphasis on olive oil and pumpkin seed oil. *Journal of Agricultural and Food Chemistry*, *49*(3), 1534–1540.
- Tena, N., Aparicio-Ruiz, R., Koidis, A., García-González, D.L. Analytical tools in authenticity and traceability of olive oil. In: *Food Traceability and Authenticity: Analytical Techniques* (pp. 232-260).



- Tripoli, E., Giammanco, M., Tabacchi, G., Di Majo, D., Giammanco, S., & La Guardia, M. (2005). The phenolic compounds of olive oil: structure, biological activity and beneficial effects on human health. *Nutrition Research Reviews*, 18(1), 98–112.
- Wehrens, R. (2011). *Chemometrics with R: multivariate data analysis in the natural sciences and life sciences*. Springer Science & Business Media.
- Wickham, H. (2017). Tidyverse: Easily Install and Load the 'Tidyverse', R package version 1.2.1, <https://CRAN.R-project.org/package=tidyverse>.
- Wilke, C.O. (2017). Cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2', R package version 0.9.2, <https://CRAN.R-project.org/package=cowplot>
- Woodbury, S. E., Evershed, R. P., & Barry Rossell, J. (1998). Purity assessments of major vegetable oils based on  $\delta^{13}\text{C}$  values of individual fatty acids. *Journal of the American Oil Chemists' Society*, 75(3), 371–379.

## Figure captions:

**Figure 1.** Strip plots for the  $\delta^{13}\text{C}$ ,  $\delta^{18}\text{O}$  and  $\delta^2\text{H}$  values determined in bulk extra virgin olive oils and for the  $\delta^2\text{H}$  and  $\delta^{13}\text{C}$  values determined in linoleic, oleic, palmitic and stearic acids extracted from olive oils from the EU (blue dots) and outside the EU (red dots). EU countries are listed according to geographical latitude, from Portugal to Greece

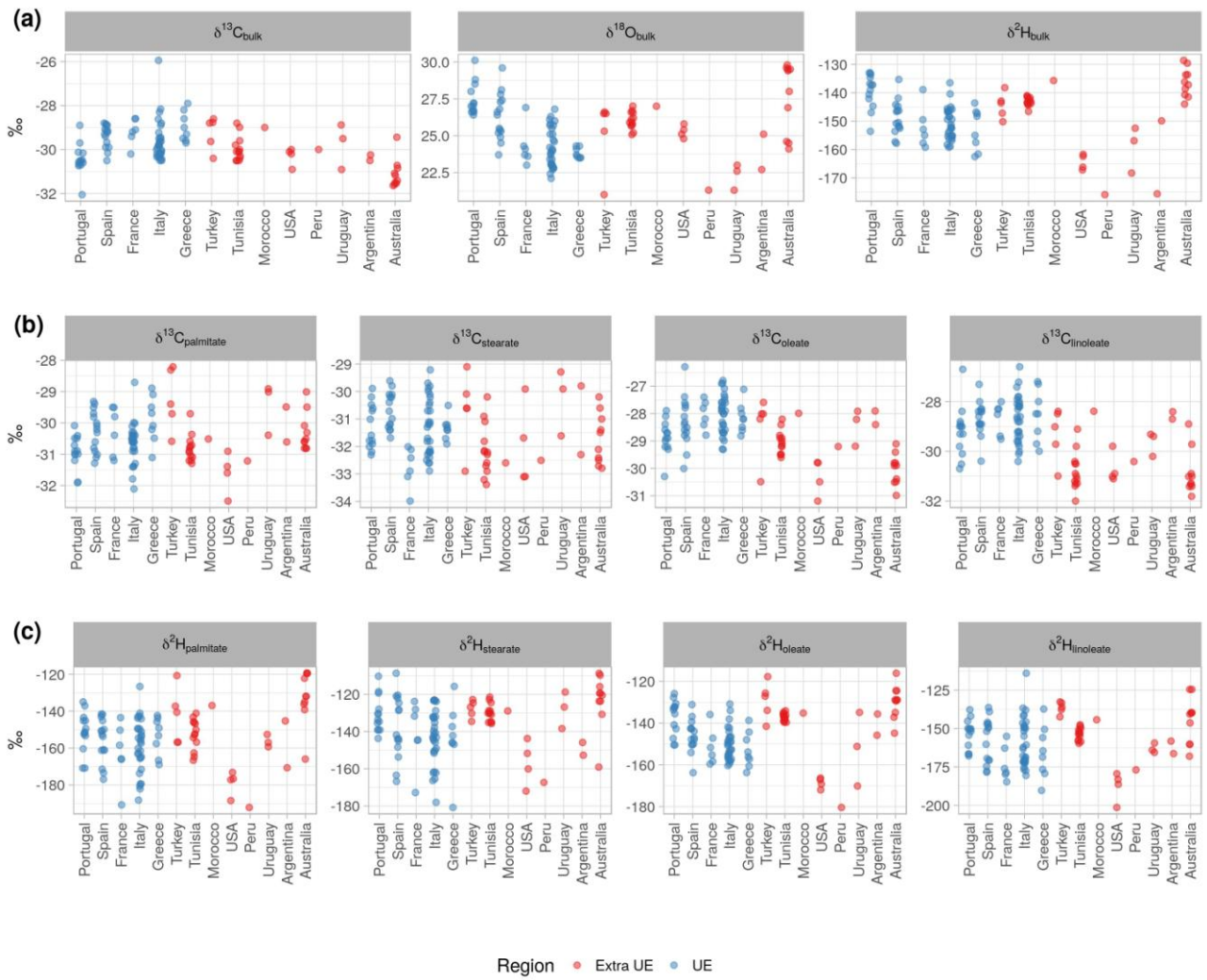
**Figure 2.** Correlation matrix of  $\delta^{13}\text{C}_{\text{bulk}}$ ,  $\delta^{18}\text{O}_{\text{bulk}}$ ,  $\delta^2\text{H}_{\text{bulk}}$ ,  $\delta^{13}\text{C}_{\text{FAs}}$  and  $\delta^2\text{H}_{\text{FA}}$  determined in extra virgin olive oils from the EU and outside the EU. The bigger and more marked are the circles and stronger is the correlation between two variables. The X indicate no statistical significance of the correlation.

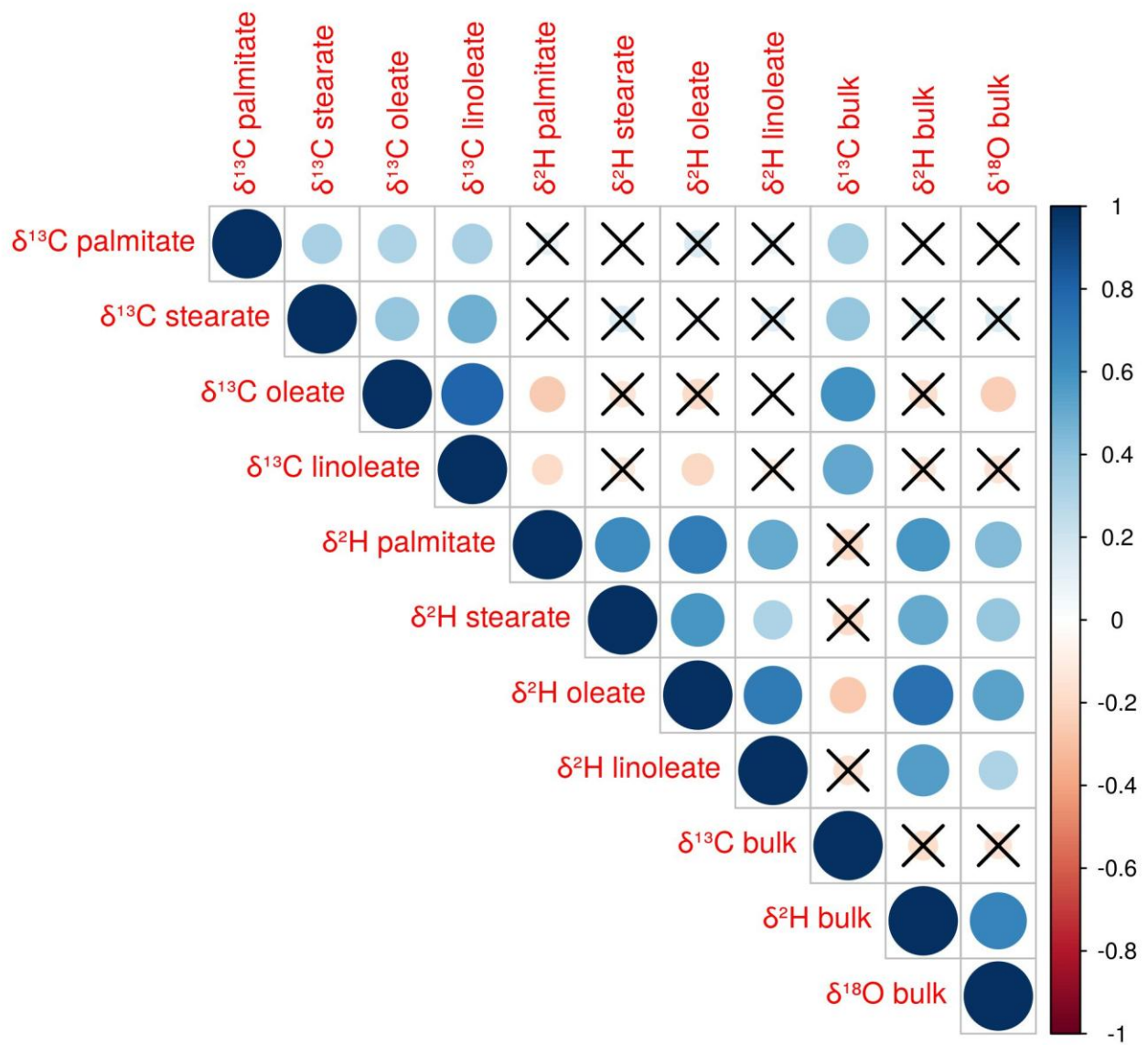
**Figure 3.**  $\delta^{13}\text{C}$  trends and  $\delta^2\text{H}$  values determined in the four main fatty acids, listed in biosynthetic order. The line connects the mean values of each fatty acid between each other.

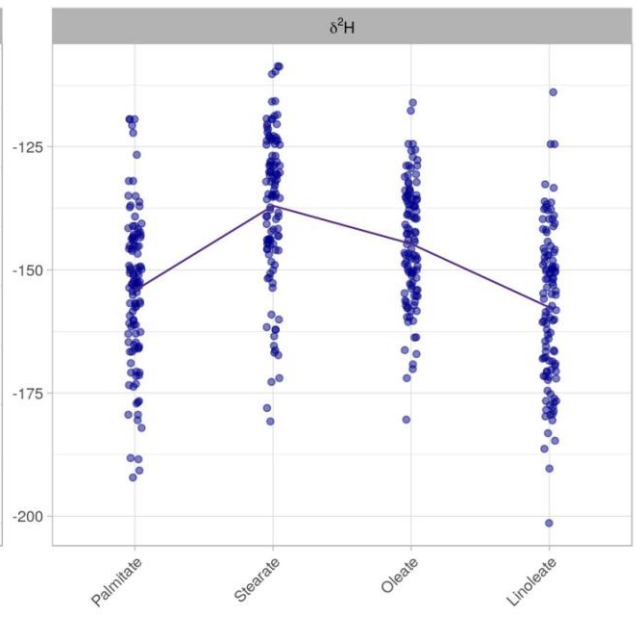
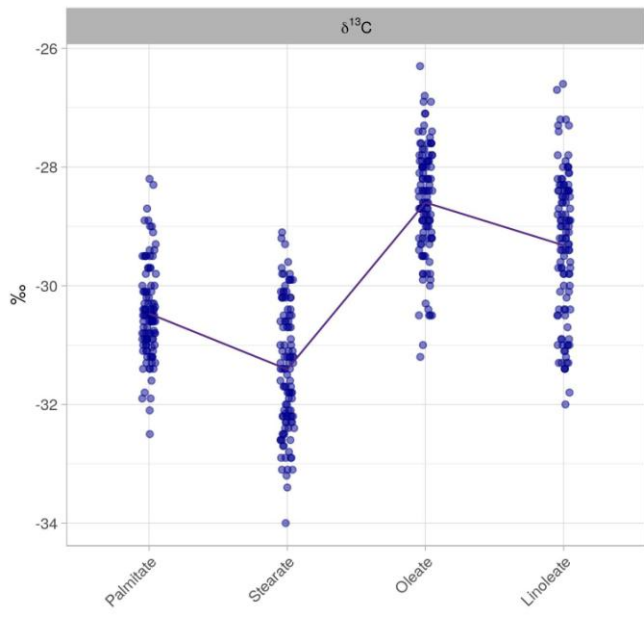
**Figure 4.** PCA of the  $\delta^{13}\text{C}_{\text{bulk}}$ ,  $\delta^{18}\text{O}_{\text{bulk}}$ ,  $\delta^2\text{H}_{\text{bulk}}$ ,  $\delta^{13}\text{C}_{\text{FAs}}$  and  $\delta^2\text{H}_{\text{FAs}}$  of extra virgin olive oil samples. In Figure 4a the samples are labelled according to the country of origin, whereas in Figure 4b according to the EU and non-EU provenance. In the plot is represented also the variable correlation plot with the contribution of each variable.

**Figure 5a.** ROC curves of extra virgin olive oil samples, analysing  $\delta^{13}\text{C}_{\text{bulk}}$ ,  $\delta^{18}\text{O}_{\text{bulk}}$ ,  $\delta^2\text{H}_{\text{bulk}}$  ('Bulk'),  $\delta^{13}\text{C}_{\text{bulk}}$ ,  $\delta^{18}\text{O}_{\text{bulk}}$ ,  $\delta^2\text{H}_{\text{bulk}}$ ,  $\delta^{13}\text{C}_{\text{FAs}}$  and  $\delta^2\text{H}_{\text{FAs}}$  ('Full'), with the labels of the countries randomly assigned to the samples ('Random'). Figure 5b. ROC curves of extra virgin olive oil samples, analysing  $\delta^{13}\text{C}_{\text{bulk}}$ ,  $\delta^{18}\text{O}_{\text{bulk}}$ ,  $\delta^2\text{H}_{\text{bulk}}$ ,  $\delta^{13}\text{C}_{\text{FAs}}$  and  $\delta^2\text{H}_{\text{FAs}}$  data with EU and non-EU randomly assigned to the countries considered ('Random region')

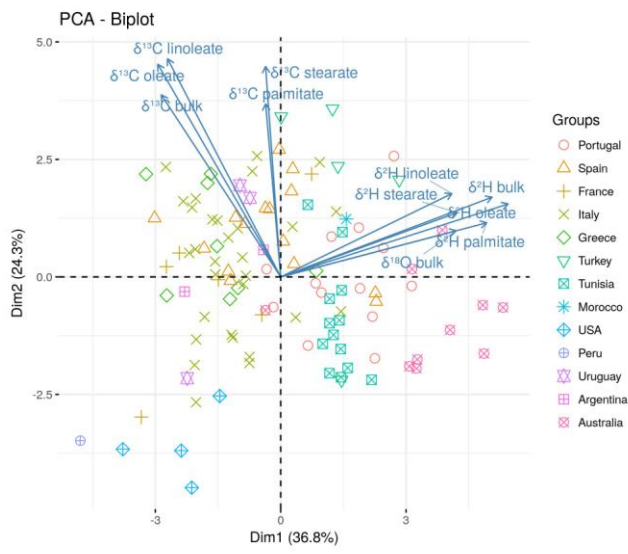
**Figure 6.** Gini index calculated for the isotopic variable determined in extravirgin olive oils collected in the EU and outside the EU



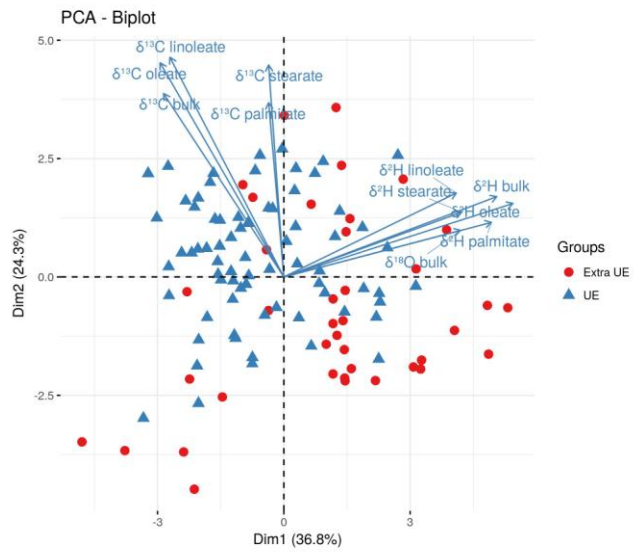


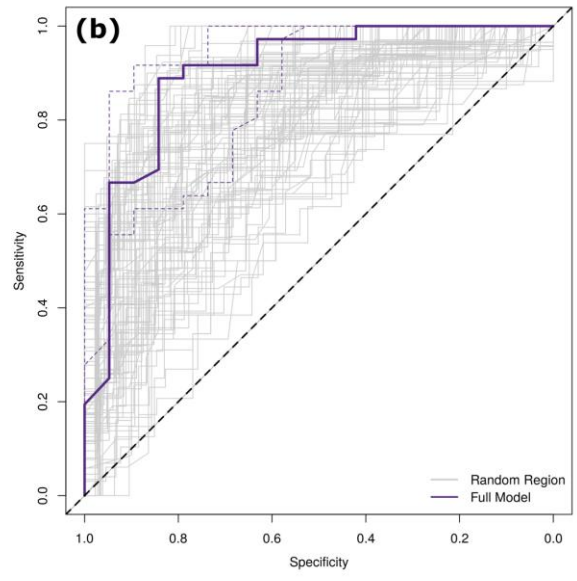
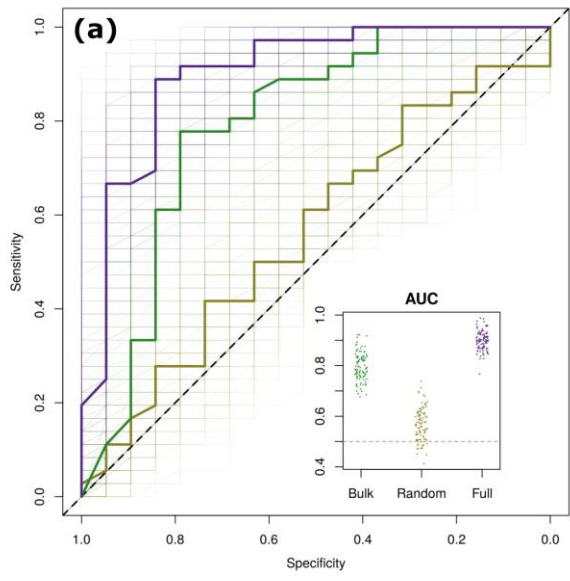


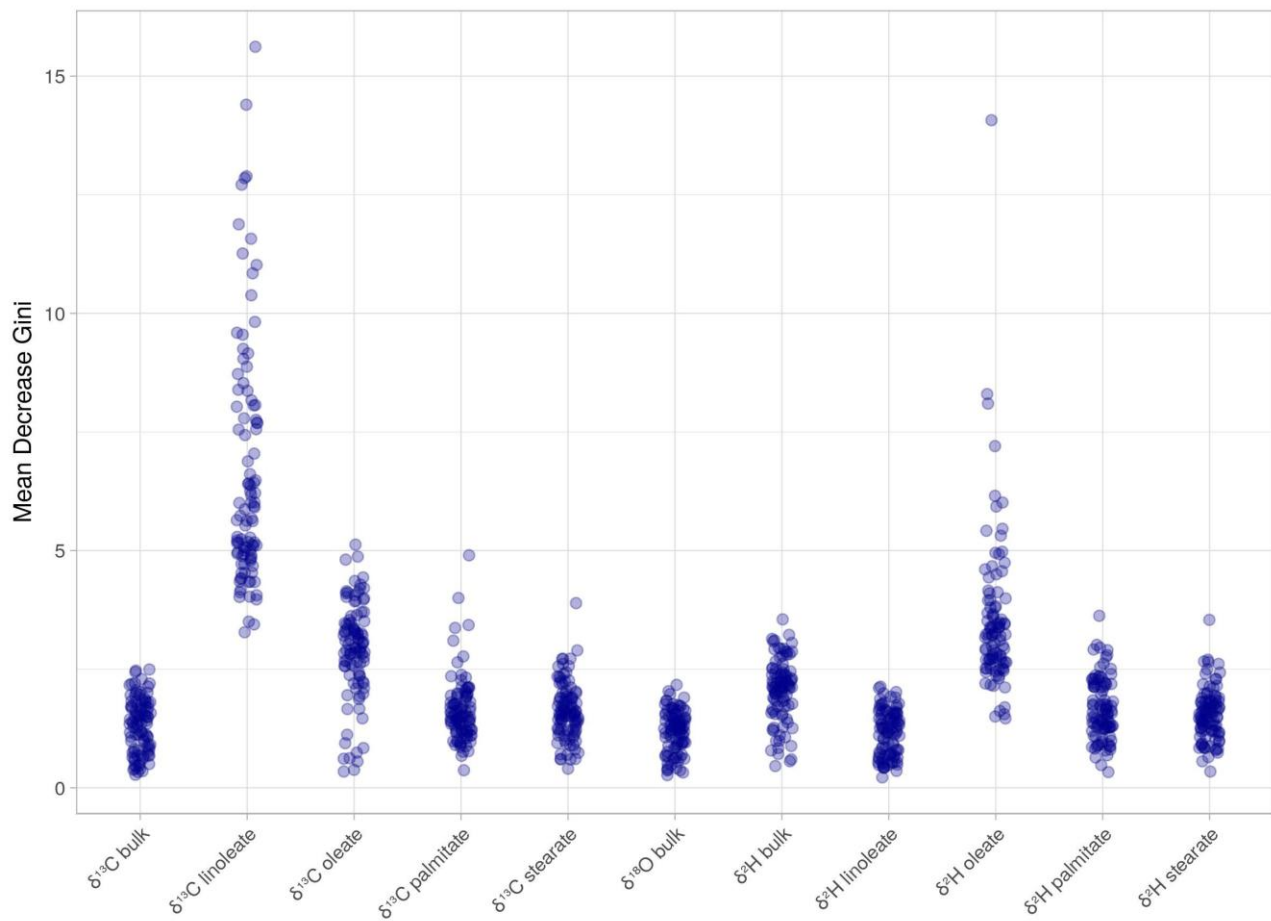
(a)



(b)









Highlights:

- $^2\text{H}/^1\text{H}$ ,  $^{13}\text{C}/^{12}\text{C}$ ,  $^{18}\text{O}/^{16}\text{O}$  of bulk was combined with  $^{13}\text{C}/^{12}\text{C}$  and  $^2\text{H}/^1\text{H}$  of fatty acids
- Isotopic composition was used to distinguish EU and non EU extra virgin olive oils
- Application of Random Forest classification and model performance were evaluated