Food Control 112 (2020) 107156

Contents lists available at ScienceDirect

Food Control

journal homepage: www.elsevier.com/locate/foodcont

Authentication of the geographical origin of virgin olive oils from the main worldwide producing countries: A new combination of HS-SPME-GC-MS analysis of volatile compounds and chemometrics applied to 1217 samples

Lorenzo Cecchi[a], Marzia Migliorini[b], Elisa Giambanelli[b], Adolfo Rossetti[b], Anna Cane[b], Nadia Mulinacci[a,*], Fabrizio Melani[a]

[a] Dipartimento di NEUROFARBA, Università degli Studi di Firenze, Via Ugo Schiff 6, 50019, Sesto F.no (Firenze), Italy
[b] Carapelli Firenze S.p.A., Via Leonardo da Vinci 31, 50028, Tavarnelle Val di Pesa (Firenze), Italy

ARTICLE INFO

Keywords:
Volatile compounds
LDA
Authentication
Geographical origin
Olea europaea L.
Frauds

ABSTRACT

Authentication of geographical origin of virgin olive oils is necessary to protect consumer and producers from frauds. A method able to classify virgin olive oils from the main worldwide producing countries is still missing. In this work, we developed 3 chemometric approaches for classification of virgin olive oils from Italy, Spain, Greece, Portugal, Tunisia and other countries all over the world. The approaches were developed starting from a data-set containing fatty acid composition and the amount of 72 volatile compounds, evaluated by a never applied HS-SPME-GC-MS quantitation method, of 1217 oil samples from three different olive oil campaign. The approach that gave the best predictive results is based on Linear Discriminant Analysis run on quantitative data from only 25 volatile compounds selected by one-way ANOVA as the most capable in discriminating between the diverse origins. The method was built and internally validated using a training-set of 1000 samples and externally validated with a test-set of 217 independent samples. The method was able to classify the geographical origin of 94.5% samples, with a percentage of correct classification even higher than 97% for some origins. Preliminary studies also suggested the proposed approach is able to correctly classify the geographical origin of binary mixtures of oils from different origins. The approach proposed in this manuscript is easily applicable in testing laboratories and represents a very useful tool for the olive oil field, helping in protecting consumers and producers from frauds.

1. Introduction

Many are the factors affecting the type and the concentration of volatile organic compounds (VOCs) in virgin olive oils (VOOs), including varietal origin, pedoclimatic conditions during olive growing, harvesting period, olive processing conditions, type of filtration and storage conditions, and geographical origin (Angerosa, 2002; Campestre, Angelini, Gasbarri, & Angerosa, 2017; Choe & Min, 2006; Kalua et al., 2007; Lukic, Carlin, Horvat, & Vrhovsek, 2019; Trapani et al., 2017; Vichi, Pizzale, Conte, Buxaderas, & Lopez-Tamames, 2003). The typical fruity and green notes of extra virgin olive oils (EVOOs) are mainly due to a series of C5 and C6 aldehydes, alcohols and esters, originated by the lipoxygenase (LOX) pathway (Angerosa, Mostallino, Basti, & Vito, 2001; Campestre et al., 2017). On the other side, sensory defects can originate, among other, from microbiological and oxidative activities and several studies have been carried out in order to define the molecules responsible for the different sensory defects (Angerosa, Lanza, D'Alessandro, Marsilio, & Cumitini, 1999; Aparicio, Rocha, Delgadillo, & Morales, 2000; Morales, Rios, & Aparicio, 1997; Morales, Luna, & Aparicio, 2005; Cecchi et al., 2019).

At the same time, it is still necessary a chemical/analytical method for authentication of VOOs according to the geographical origin (Bajoub et al., 2018; Berlioz et al., 2006). Indeed, consumers demand and are available to pay more for EVOOs with specific characteristic linked to the geographical origin, and producers need to be able to correctly communicate to buyers and consumers the specific characteristics giving added value to their product (European Community, 2006); EU Regulation 1151/2012; Fregapane and Salvador, 2019). The price of these products is often higher than EVOOs with no specific characteristics, thus economic frauds regarding false claim of geographical origin of the product on the label are spreading on the market and, to date, cannot be fully avoided (Bajoub et al., 2018; Cajka et al.,

* Corresponding author.
E-mail address: nadia.mulinacci@unifi.it (N. Mulinacci).

https://doi.org/10.1016/j.foodcont.2020.107156
Received 27 November 2019; Received in revised form 2 February 2020; Accepted 3 February 2020
Available online 03 February 2020
0956-7135/

2010; Garcia, Martins, & Cabrita, 2013).

Several approaches, based on different analytical techniques and/or statistical approaches, have been proposed for assessing the geographical origin. The use of Nuclear Magnetic Resonance (^{1}H NMR) has been rewieved, and the authors concluded that studies with high number of samples from different olive oil crops and robust statistical analysis are still necessary (Dais & Hatzakis, 2013). In this sense, more recently, a ^{1}H-NMR-chemometric characterization of 383 EVOOs from 9 Italian regions collected in three different harvesting years, was carried out for authentication of their geographical origin, with prediction capability ranging 60–100% for the different regions after applying Hierarchical Cluster Analysis (HCA) (Ingallina et al., 2019). Analysis of trace elements by ICP-MS and δ^{13}C isotope ratio by Isotope Ratio Mass Spectrometry (IRMS), followed by Principal Component Analysis (PCA) and HCA, were applied for assessing the geographical origin of a total of 49 EVOOs from 6 different Turkish areas, with promising but preliminary results (Gumus, Celenk, Tekin, Yurdakul, & Ertas, 2017). Triglyceride fingerprinting data acquired by both reverse and normal phase HPLC coupled to Charged Aerosol Detector (CAD) and High Temperature-GC-FID analysis of 65 EVOOs were analyzed by PCA, PLS-DA and SIMCA for assessing the geographical origin of samples of Arbequina cv from three Spanish regions: the proposed model gave satisfactory results only after applying data fusion for combining data obtained from the three applied techniques (Vera, Jimenez-Carvelo, Cuadros-Rodriguez, Ruisanchez, & Pilar Callao, 2019). Phenolic and sterolic fingerprinting by UHPLC-ESI/QTOF-MS was recently applied for discriminating between Tunisian and Italian EVOOs (Ben Mohamed et al., 2018) and between EVOOs from 6 different Italian regions (Ghisoni et al., 2019): the two studies are based on a total of 15 and 26 samples, respectively, with one part of samples purchased on the market and the other part extracted at laboratory scale. The authors concluded that, despite the encouraging results of the new proposed approach, further investigations based on higher numbers of samples, from several origins are advisable (Ghisoni et al., 2019). Selected Ion Flow Tube Mass Spectrometry (SIFT-MS) was applied to a total of 130 EVOOs from six regions belonging to 4 countries in the Mediterranean area and the proposed PLS-DA model gave excellent prediction accuracy (Bajoub et al., 2018); the main limit of the study is the data-set, constituted by no more than 25 EVOOs for each origin from the same olive oil crop. More recently, Laser Induced Breakdown Spectroscopy (LIBS) has been proposed in combination with Linear Discriminant Analysis (LDA) for assessing the geographical origin of VOOs, but again the main weakness of the work is given by the small number of samples (8), all from different regions of the same country (Greece) (Gazeli, Bellou, Stefas, & Couris, 2020).

Besides these and other approaches for authentication of the geographical origin proposed in the literature and not discussed herein (Camin et al., 2010; Mallamace et al., 2017; Persuric, Saftic, Masek, & Pavelic, 2018; Portarena, Baldacchini, & Brugnoli, 2017; Quintanilla-Casas et al., 2020; Sayago, Gonzalez-Dominguez, Beltran, & Fernandez-Recamales, 2018; Souayah et al., 2017; Techer, Medini, Janin, & Arregui, 2017), the volatile profile of VOOs has been reported as strongly affected by the growing area and used for proposing approaches for authentication of the geographical origin (Lukic et al., 2019; Ouni et al., 2011; Pizarro, Rodriguez-Tecedor, Perez-del-Notario, & Gonzalez-Saiz, 2011; Pouliarekou et al., 2011; Zunin, Boggia, Salvadeo, & Evangelisti, 2005). In one of these recent studies, the volatile fraction of a total of 30 Croatian samples was in depth studied by GC × GC-TOF-MS, and used for discriminating the samples according to the geographical origin (Lukic et al., 2019); Pouliarekou et al. (2011) analyzed the volatile fraction of a total of 51 samples from 6 Greek regions by HS-SPME-GC-MS, and classified them using LDA, while Pizarro et al. (2011) used the same analysis followed by LDA for classifying a total of 40 samples from 3 Spanish regions. Some researchers also combined VOCs profile and fatty acid (FAs) composition aiming at an improved geographical differentiation working on a total of 74 samples from two olive oil crops and four areas in Greece: the proposed LDA model allowed obtaining 81.1% of correct classification, slightly higher than the percentage obtained only using the volatile profile (79.7%) (Kosma et al., 2017).

As above explained and confirmed in other studies (Ben Mansour, Chtourou, Khbou, Flamini, & Bouaziz, 2017), the most of the papers present in the literature are based on small numbers of samples, often from one single olive oil crop. Other authors, working with larger numbers of samples, aimed at distinguish between one geographical origin and all the other ones, e.g. “Ligurian” from “non-Ligurian” samples (Cajka et al., 2010) or “100% Italian” from “non-100% Italian” samples (Melucci et al., 2016). As also reported in some of the above mentioned papers (Bajoub et al., 2018; Ghisoni et al., 2019), pluri-annual studies including a very high number of samples are strongly advisable in order to obtain powerful, robust and reliable models suitable to classify virgin olive oils according to the geographical origin.

For these reasons, in this work, we aimed at proposing a model suitable for authentication of the geographical origin of virgin olive oils from five different countries, including the main worldwide producers (Spain, Greece, Italy, Tunisia and Portugal), in addition to a group of oils from other countries. To this aim, we collected a total of 1217 virgin olive oil samples with different categories (EVOO and VOO) according to chemical and sensorial analysis, and from three different olive oil crops (2016–2017, 2017–2018, 2018–2019). The profile of Volatile Organic Compounds (VOCs) of all the samples was then characterized by a method recently optimized and validated by our group (Fortini, Migliorini, Cherubini, Cecchi, & Calamai, 2017): this method is based on a widely used technique (HS-SPME-GC-MS) and on a new approach for the quantification of 73 VOCs. This approach, using up to 11 internal standard for area normalization, allows for quantitation of VOCs with a higher accuracy than the typical methods only using one internal standard and in a wider range of calibration (Fortini et al., 2017). Finally, this method has never been applied for authentication of the geographical origin of virgin olive oils. At the same time, the composition of FAs of a reduced group of samples was analyzed and all the collected data were subjected to suitable statistical analysis in order to propose the model.

2. Materials and methods

2.1. Chemicals and standard solutions

All chemicals and standards of analytical reagent grade were from Sigma-Aldrich (Steinheim, Germany). Volatile standards used for preparing solutions for external calibration curves (ExtStd) in a refined olive oil free from VOCs were from Sigma-Aldrich (Steinheim, Germany): purity was 50.0% for *Z*-3-hexenal, 90.0% for *E,E*-hepta-2,4-dienal and *E,E*-nona-2,4-dienal, 85.0% for *E,E*-deca-2,4-dienal and ≥95.0% for all other standards. Internal standard solution was prepared by weighing acetic acid-2,2,2-d_3 (≥99.0%), 6-chloro-2-hexanone (≥97.0%), butanol-d_{10} (≥99.0%), 4-methyl-2-pentanol (≥98.0%), 3-octanone (≥98.0%), ethyl acetate-d_8 (≥99.0%), 3,4-dimethylphenol (≥98.0%), toluene-d_8 (≥99.6%) and trimethylacetaldehyde (≥96.0%) (all from Sigma-Aldrich, Steinheim, Germany) in refined olive oil (IntStd). Six diluted solutions constituted by the same amount of IstStd and different amounts of ExtStd were then prepared for quantification purpose according to previous works (Fortini et al., 2017). These solutions were stored in the dark at −20 °C until analyses.

FAME mix CRM18917 - C14–C22 referenced material was purchased from Supelco, Sigma-Aldrich (Darmstadt, Germany). Inert gasses (He and N_2 99.999% purity) were supplied by SOL gas company.

2.2. Samples

A total of 1217 virgin olive oil samples each from different geographical origins were collected from the Carapelli laboratory

(Carapelli S.p.A., Tavarnelle Val di Pesa, Florence, Italy) during 2016/17, 2017/18 and 2018/19. Samples distribution was almost representative of the worldwide virgin olive oil production: in particular, samples were from Spain (code "S", 340 samples), Italy ("I", 408), Greece ("G", 246), Tunisia ("T", 85), Portugal ("P", 98) and other ("O", 40), with samples labelled as "O" that were from Peru (3), Morocco (8), Australia (24), Albania (1), California (2), and Argentina (2). Samples were classified as EVOO (600 samples) or VOO (617 samples) after chemical and sensorial analysis, according to the methods described in the next paragraph.

Furthermore, a set of 80 samples constituted by binary mixtures of virgin olive oils of different geographical origins belonging to Italy, Spain, Tunisia, Greece and Portugal was collected; these samples were prepared in the Carapelli laboratory starting from oils with a known origin.

2.3. Chemical and sensorial analysis for oil classification

Legal quality indices used for classification of samples were determined according to European Regulations (European Economic Community, 1991): the chemical ones consisted of free acidity, peroxide value and spectrophotometric indices, while for the sensorial one, the Panel Test was carried out by a team acknowledged by the Italian Ministry of Agricultural Policies (MIPAAF), according to EU Reg. 2568/1991. The obtained results used for confirming samples' classification are not reported.

2.4. Fatty acid composition

The composition of fatty acids (FAs) of a subset of 180 samples (G = 36, I = 36, P = 36, S = 36, O = 36 (2 from Peru, 16 from Australia, 18 from Tunisia)) was analyzed according to the International Olive Council official method (IOC/T.20/Doc No. 33/Rev.1) after slight modifications. Briefly, trans-esterification was applied to the samples for preparation of fatty acid methyl esters (FAMEs): 0.02 g of samples were dissolved in 4 mL of heptane in the presence of 0.4 mL of 2.0 N methanolic potassium hydroxide and vigorously shaken. The obtained FAMEs mix was analyzed by GC with an Agilent Technologies (7890B) chromatograph (Palo Alto, CA, USA) equipped with a capillary GC column BPX70, 60 m × 0.25 mm i.d., 0.25 μm f.t., from SGE Analytical Science (Ringwood Victoria, Australia) and a FID detector. Injection volume, 1 μL; carrier gas, nitrogen at 1 mL/min; injector and detector temperature, 250 °C. The initial oven temperature was kept at 180 °C for 12 min, raised to 220 °C, with 10 °C/min gradient, and it was maintained for 5 min; after that, the temperature reached 250 °C, with 12 °C/min gradient, and finally maintained for 5 min. Ten FAs, namely C16:0, C16:1, C18:0, C18:1, C18:2, C18:3, C20:0, C20:1, C22:0 and C24:0, in addition to C18:1, C18:2 and C18:3 *trans* FAs, were identified based on retention time and comparison with C14–C22 referenced material FAME mix CRM18917 from Supelco, Sigma-Aldrich (Darmstadt, Germany). For each sample, the composition of FAs was evaluated in terms of peak area percentage.

2.5. HS-SPME-GC-MS analysis of volatile organic compounds

Volatile fraction of all samples was analyzed by the validated HS-SPME-GC-MS method previously described (Fortini et al., 2017). Briefly, approx. 4.3000 g of sample and approx. 0.1000 g of the internal standard mix solution were exactly weighed into a 20 mL screw cap vials. A 1-cm SPME fiber 50/30 μm DVB/CAR/PDMS by Agilent (Palo Alto, CA, USA) was exposed under orbital shaking at 400 rpm and 45 °C for 20 min in the vial headspace, after sample equilibration for 5 min at 45 °C. VOCs were then desorbed for 1.7 min in the injection port of a 6890N GC system equipped with a 5975-model MS detector, (Agilent, Palo Alto, CA, USA) and separated using a HP-Innowax capillary column (50 m × 0.2 mm i.d., 0.4 μm film thickness), then the fiber was conditioned for 20 min at 260 °C. Initial oven temperature was 40 °C for 2 min; it was then raised to 156 °C with 4 °C/min, and then to 260 °C with 10 °C/min. Carrier gas was helium at 1.2 mL/min. The temperature of transfer line and ion source were 250 °C and 230 °C, respectively. Mass detector conditions: scan mode, 30–350 Th mass range, 1500 Th/s, IE energy 70 eV.

The 73 VOCs were identified by comparison with mass spectra and retention times of authentic standards; since 2-methylbutanol and 3-methylbutanol co-elute, they were considered together and, for each sample, the output was given by 72 quantitative results. For each VOC, the more suitable internal standard was selected for area normalization before quantitation. Each VOC was quantified using a six point linear least squares calibration line in which the area ratio was plotted versus the amount ratio. Each calibration line was built using the relative pure standard. The method was validated as previously described (Fortini et al., 2017).

2.6. Data analysis

Two data matrices (namely $M_{ij}1$ and $M_{ij}2$) were built, both having the following form:

$$M_{ij} = (S_i, C_j)$$

$M_{ij}1$ was built with the data-set also including data from composition of FAs: in this matrix S_i is the ith of the 180 samples and C_j is the jth of the 84 variables (12 FAs and 72 VOCs). $M_{ij}2$ was built with the data-set only including data related to the VOCs profile: in this matrix S_i is the ith of the 1217 samples and C_j is the jth of the 72 variables (namely, the 72 VOCs). Each value in the matrices is the amount (in % w/w for FAs and in mg/kg for the VOCs) of the considered sample.

Starting from these two matrices, three approaches for classification of samples according to their geographical origin have been developed and proposed. All the three approaches were based on the use of Linear Discriminant Analysis (LDA) as pattern recognition technique. It is based on searching for the discriminant functions able to achieve maximum separation among the samples' categories through maximization of between-class variance and minimization of within-class variance. These new discriminant functions are a linear combination of the original variables, and are named canonical variables (Pizarro et al., 2011). Wilks' Lambda test was used to confirm the significance of the discriminant functions in discriminating between the different origins. During the development of the approaches, we kept into account that, when LDA is applied as pattern recognition technique, the number of variables should not exceed the number of samples (Pizarro et al., 2011). Diverse strategies for reducing the number of variables were applied, as described in the following paragraphs in which the approaches are presented: these strategies mainly used two statistical tools, namely Principal Component Analysis (PCA) and one-way Analysis of Variance ANOVA.

PCA was used on $M_{ij}1$ as non-supervised technique with the main goal of reducing the complexity of data, transforming the original variables in new ones, called Principal Components (PCs), to be used in the following LDA, using the scores as the quantitative values. These PCs are orthogonal to each other, meaning that each PC is uncorrelated with the other ones. The first 20 PCs were selected, based on eigenvalues higher than 1 (total explained variance 78.0%). One-way ANOVA and *F*-test were run on $M_{ij}2$ for assessing which variables are able to discriminate between oils from different origins; Fisher Least Significant Difference (LSD) test was then used for comparing the averaged values and for assessing which origins are differentiated by that variable, at level of significance of 0.05.

The stability of all the proposed models was internally validated by the leave-one-out cross-validation procedure. Furthermore, for the models involving a number of samples big enough, an external validation was performed using a test-set of samples not used to construct

Table 1
Results obtained for each geographical origin after leave-one-out cross-validation using data from both fatty acid composition and volatile profile. In each line, there is the number (in the brackets) and the percentage of Virgin Olive Oil samples of that origin assigned by the model to the origin in the columns. The figures in bold corresponds to the correctly assigned samples.

Origin	Spain	Portugal	Italy	Greece	Other	Total	Classified
Spain	76.5% (26)	20.6% (7)	2.9% (1)	–	–	36	94.4% (34)
Portugal	22.2% (8)	77.8% (28)	–	–	–	36	100.0% (36)
Italy	–	2.9% (1)	91.4% (32)	–	5.7% (2)	36	97.2% (35)
Greece	–	–	–	100.0% (35)	–	36	97.2% (35)
Other	–	5.7% (2)	8.6% (3)	–	85.7% (30)	36	97.2% (35)

the model. Finally, one of the three models was furtherly validated using a set of 80 samples constituted by binary mixtures of virgin olive oils of different geographical origins, in order to test its capability of recognize the origin of both the oils present in the mixture.

All statistical processing of data were performed using OriginPro 2018 (OriginLab Corporation, Northampton, MA 01060 USA http://www.originlab.com).

3. Results and discussion

The final objective of this paper was to propose a reliable chemometric approach for the authentication of geographical origin of virgin olive oils from the main worldwide producing countries. The method we want to propose must be as easily applicable as possible in testing laboratories, with sustainable timing and costs.

To this aim, we selected two type of chemical analysis (composition of FAs and volatile profile) able to give lots of information, and suitable statistical tools for treating the data, for analyzing up to 1217 commercial virgin olive oil samples from several origins. We obtained quantitative data for a total of 84 parameters (72 VOCs and 12 FAs) for each of the analyzed samples. Starting from this huge dataset and applying suitable statistical tools, we proposed three approaches for authentication of the geographical origin of virgin olive oils, as described in the following paragraphs.

Because many are the papers in the literature that describe the volatile profile and the FAs profile of virgin olive oils, this aspect is out of the aims of this work and not discussed. In the next paragraphs, the obtained data will only be used for building models for assessing geographical origin of olive oil and to briefly discuss on what VOCs differentiated the most between the diverse origins.

3.1. Authentication of geographical origin according to VOCs and FAs

To our knowledge, none of the works in the literature proposed a model for authentication of the geographical origin of VOOs including samples from the first five worldwide producing countries simultaneously, and based on quantitative data on VOCs and FAs of a high number of commercial virgin olive oils from three consecutive olive oil crops. For these reasons, the first model we built for authentication of geographical origin involved composition of FAs together with the profile of the VOCs. The composition of FAs has been used in the past for classification of olive oils according to their origin, with sometimes satisfactory and sometimes less satisfactory results (Di Bella et al., 2007; Kosma et al., 2017; Longobardi et al., 2012). These studies revealed a wide variability of FA composition as a function of geographical origin.

The dataset (matrix $M_{ij}1$) was constituted by 5 classes (namely, Spain, Portugal, Italy, Greece, Other) each with the same number of samples (36). All the variables in the matrix $M_{ij}1$ were considered with no selection, so that none of the available information was a priori excluded. By this way, we applied a PCA-LDA approach for building a model that likely gives the best predictive results with the available data. Since LDA can be used only with matrices having a number of samples higher than the number of variables for each class (Pouliarekou et al., 2011), we initially applied PCA on the 84 variables (12 FAs and 72 VOCs) in order to select a reduced number of PCs for using them in the following LDA. We selected the first 20 PCs, as described in the Material and Methods section. Consequently, LDA was run using a new matrix ($M_{ij}1b$, 180 × 20) in which S_i is the ith of the 180 samples and C_j is the jth of the 20 selected scores from the PCs. The predictive capability of the model was internally validated by leave-one-out cross-validation procedure. The model was built to be able to classify samples in the categories Spain (S), Portugal (P), Italy (I), Greece (G) and Other (O) and the samples were allocated in one of the 5 categories only if the post classification probability (CP th %) was higher than 50% for that category. This threshold value was chosen as a compromise to have the highest number of samples classified by the model with no risk of having samples with more than one classification.

Table 1 shows the results obtained after LDA and leave-one-out cross-validation. Overall, 175 out of the 180 samples (97.2%) were classified by the model, and 151 out of the 175 classified samples (86.3%) were correctly allocated. Noteworthy, 91.4% of samples from Italy and 100% of samples from Greece were correctly classified, while the percentage of correctly classified Spanish and Portuguese samples was slightly lower. Regarding the wrongly classified samples of these two origins, it emerges that all the Portuguese samples were classified as Spanish and vice-versa. These findings can be explained considering that the territories of the two countries (Spain and Portugal) both belong to the Iberian peninsula, and are not so clearly distinguishable with this approach using all the information initially available with no selection.

The obtained results were very satisfactory and even higher than those reported in the literature by other authors working on the quantitative data of VOCs and FAs, which obtained 81.1% of correct classification working on samples from different geographical area of Greece (Kosma et al., 2017).

3.2. Authentication of geographical origin only based on VOCs

The results obtained using the PCA-LDA approach working with both FAs and VOCs quantitative data gave satisfactory results. In order to propose a cheaper and more easily applicable model, to gain qualitative information, to try to better distinguish also between Spain and Portugal and to extend the groups of classes also adding Tunisia, we built a new model only based on VOCs of a huge number of samples (1217).

The new dataset (matrices $M_{ij}2$) was thus constituted by all samples from the 6 classes reported in paragraph 2.2 (Spain, Portugal, Italy, Greece, Tunisia and Other, this latter category with samples from countries in Europe, South America, North America, Africa and Oceania).

Data in the matrix $M_{ij}2$ were initially subjected to one-way analysis of variance (ANOVA) and averaged values were compared by LSD in order to assess which couples of classes are distinguished by each VOC. In Table 2, the 72 quantified VOCs are listed according to decreasing F-ratio and the significance of each VOC in discriminating between each possible couple of origins is reported in the next 15 columns.

As we can see, 64 out of the 72 VOCs have a ratio F/F-crit greater

Table 2
List of the 72 quantified VOCs (1st column) sorted by decreasing Fisher *F*-ratio (2nd column). The following 15 columns report all the possible binary combinations of the 6 geographical origins (S, Spain; T, Tunisia; P, Portugal; I, Italy; G, Greece; O, Other) of VOO samples: value 1 indicates that the difference of means for the VOC in that row is significant at the 0.05 level for the two origins, while value 0 indicates that the difference is not significant. The first 25 VOCs, used for building the model, are in bold in the table; the name of the molecules originated by LOX pathway is followed by †

n°	VOCs	F/F-crit.	S-T	P-T	P–S	I-T	I–S	I–P	G-T	G-S	G-P	G-I	O-T	O–S	O–P	O–I	O-G
1	Hexyl acetate †	106.7	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1
2	*E*-2-hexenal †	82.5	1	1	1	1	1	0	1	1	1	1	0	1	1	1	1
3	*Z*-3-hexenyl acetate †	68.9	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1
4	Pentan-2-ol	65.5	1	1	1	0	1	1	0	1	1	0	1	1	1	1	1
5	Octan-1-ol	54.1	1	1	0	1	0	0	1	1	1	1	0	1	1	1	1
6	Oct-1-en-3-ol	50.1	1	1	0	1	0	0	1	1	1	1	1	1	1	1	0
7	Heptan-1-ol	46.3	1	1	0	1	0	0	0	1	1	1	0	1	1	1	1
8	Hexan-1-ol †	45.6	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1
9	Nonanal	43.0	0	1	1	0	0	0	1	1	1	1	0	0	0	0	1
10	Phenyl ethanol	42.6	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0
11	6-methylhept-5-en-2-one	41.8	0	0	1	0	1	0	1	1	1	1	0	0	0	0	1
12	*E*-2-hexenol †	41.0	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1
13	*Z*-3-hexenol †	35.7	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1
14	Pent-1-en-3-ol †	35.6	1	1	1	1	1	1	1	1	0	1	0	1	0	1	1
15	Octanal	30.9	1	1	0	1	0	1	1	1	1	1	1	1	1	1	0
16	*E*-2-decenal	30.5	0	0	0	0	1	0	1	1	1	1	0	0	0	0	1
17	Ethyl Acetate	26.7	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1
18	Heptan-2-ol	25.8	1	0	1	1	1	1	1	1	1	1	0	1	0	1	1
19	Oct-1-en-3-one	25.1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1
20	Methyl Acetate	24.4	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1
21	Ethanol	22.7	1	0	1	0	1	0	0	1	1	1	1	1	0	0	1
22	Acetic acid	21.7	1	1	1	1	1	1	0	1	1	1	1	0	1	0	1
23	Heptanal	21.2	1	1	0	1	0	0	1	1	1	1	0	1	1	1	1
24	Octane	20.8	1	0	1	0	1	0	1	1	1	1	1	0	1	1	0
25	Hexanal †	20.1	1	0	1	0	1	0	1	1	1	1	0	1	1	1	1
26	Propanoic acid	17.8	1	1	0	1	0	1	1	0	1	0	1	1	1	1	1
27	2-methyl + 3-methyl-1-butanol	17.1	1	1	1	0	1	1	0	1	1	0	1	0	0	1	1
28	2-methylbutanal	16.0	1	1	1	1	0	1	1	0	1	1	1	0	1	0	0
29	*E,E*-Hepta-2,4-dienal	15.4	0	0	1	1	1	1	0	0	0	1	0	0	0	1	0
30	Isobutanol	14.4	1	1	0	0	1	1	1	0	0	1	0	1	1	0	1
31	Guaiacol	14.4	0	1	1	0	0	1	0	0	1	0	1	1	1	1	0
32	Isovaleraldehyde	14.2	1	0	1	1	0	1	1	1	1	1	1	0	1	0	0
33	Octan-2-one	14.2	1	1	0	1	0	0	0	1	1	1	1	0	1	1	1
34	Methanol	13.6	1	1	1	1	1	0	0	1	1	1	1	1	0	1	0
35	Decanal	12.8	0	0	0	1	1	1	0	0	0	1	0	0	0	1	0
36	*Z*-3-hexenal †	12.6	0	1	1	1	1	0	0	0	1	1	0	0	1	1	0
37	Nonan-1-ol	12.6	0	0	0	0	0	0	1	1	1	1	0	0	0	0	1
38	Benzaldehyde	12.5	0	0	0	0	0	0	1	1	1	1	0	0	0	0	1
39	*E,E*-Deca-2,4-dienal	12.5	1	1	1	1	0	1	1	0	1	0	1	0	1	0	0
40	4-Ethyl phenol	12.1	0	1	1	1	1	0	1	1	0	0	0	1	1	1	1
41	Butanoic acid	11.4	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1
42	*E*-2-octenal	11.2	1	1	1	1	1	0	1	0	1	1	1	0	1	1	0
43	Ethyl butanoate	10.0	1	0	1	0	1	0	1	1	0	0	1	0	1	1	1
44	Pent-1-en-3-one †	9.9	0	0	0	1	1	1	0	1	1	1	0	0	0	1	1
45	*E*-2-pentenal †	9.4	0	0	1	1	1	1	1	1	1	0	0	1	0	0	0
46	Heptan-2-one	9.2	1	1	0	1	0	0	1	1	1	1	1	0	0	0	0
47	Pentan-3-one	8.4	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1
48	Ethyl propanoate	8.0	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1
49	*E*-2-pentenol †	6.3	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0
50	Limonene	6.3	1	1	0	1	1	0	0	1	1	1	0	0	0	0	1
51	Butan-2-ol	6.1	1	0	1	0	1	0	0	1	0	0	1	1	1	1	1
52	*E,E*-Hexa-2,4-dienal	5.7	0	1	1	1	1	0	0	0	1	1	0	0	1	1	0
53	*E*-2-heptenal	5.4	1	1	0	1	0	0	1	1	0	1	1	1	1	1	1
54	Propanol	3.8	1	0	0	0	1	1	1	0	0	1	1	0	0	1	0
55	*E*-3-hexenol	3.6	0	1	1	0	0	1	1	0	1	0	1	1	0	1	0
56	*E*-2-nonenal	3.3	1	1	0	1	0	0	0	1	1	1	0	0	0	0	0
57	*Z*-2-hexenol	3.2	0	0	0	1	1	1	0	0	0	1	0	0	0	0	0
58	4-Ethylguaiacol	2.4	0	1	1	0	0	1	0	1	0	0	1	1	1	1	1
59	Pentanoic acid	2.3	0	0	0	0	0	0	0	1	0	1	1	1	1	1	1
60	Valeraldehyde	1.9	0	0	1	0	1	0	0	0	1	1	1	1	1	1	1
61	Phenol	1.8	1	0	1	1	0	1	1	0	1	0	0	0	0	0	0
62	Hexanoic acid	1.1	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0
63	Heptane	1.0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0
64	Pentanol	1.0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
65	Butyl acetate	0.9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
66	Butan-2-one	0.9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
67	*E,E*-Nona-2,4-dienal	0.8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
68	Methyl propanoate	0.7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
69	*E*-2-hexenyl acetate	0.6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
70	*Z*-2-pentenol †	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
71	Octan-2-ol	0.4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
72	Nonan-2-one	0.2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 3
Composition of training and test sets, according to the geographical origin of Virgin Olive Oil samples.

Origin	Training-set	Test-set
Tunisia	74	11
Spain	308	32
Portugal	74	24
Italy	308	100
Greece	206	40
Other	30	10
Total	1000	217

than 1, thus being able to differentiate at least one couple of origin. The only VOCs not capable of differentiating between the origins were butyl acetate, butan-2-one, *E,E*-nona-2,4-dienal, methyl propanoate, *E*-2-hexenyl acetate, *Z*-2-pentenol, octan-2-ol and nonan-2-one (Table 2).

It also immediately emerged that the first three VOCs belong to the molecules originated by the LOX pathway and that 7 out of the 13 molecules from the LOX pathway are ranked in the first 15 positions in Table 2. Starting from this information, and in an attempt to propose an approach using a very low number of quantified VOCs, we built a model only using the 13 molecules originated from the LOX pathway, namely hexyl acetate, *E*-2-hexenal, *Z*-3-hexenyl acetate, hexan-1-ol, *E*-2-hexenol, *Z*-3-hexenol, pent-1-en-3-ol, hexanal, *Z*-3-hexenal, pent-1-en-3-one, *E*-2-pentenal, *E*-2-pentenol and *Z*-2-pentenol.

Given the availability of a higher number of data, the entire data-set has been randomly split in two subsets: a training-set (1000 samples) for building and internally validate the model, and a test-set (217 samples) to be used for the external validation of the model. The composition of the two subsets is shown in Table 3.

For building the model, LDA was thus run using a new matrix (M_{ij}2b, 1000 × 13) in which S_i is the *i*th of the 1000 samples in the training-set and C_j is the *j*th of the 13 selected VOCs. As for the model built using both VOCs and FAs, the predictive capability of the model was internally validated by leave-one-out cross-validation procedure. The model was built to be able to classify samples in the categories Spain (S), Portugal (P), Italy (I), Greece (G), Tunisia (T) and Other (O) and again the samples were allocated in one of the 6 categories only if the post classification probability (CP th %) was higher than 50% for that category. The prediction capability of this model was furtherly assessed using the test-set as set of independent samples for the external validation. The obtained results from both internal and external validation are shown in Table 4. Overall, the model was able to classify 855 out of the 1000 samples (85.5%) of the training set and 188 out of the 217 samples (86.6%) of the test set. This means that this model has a lower capability of classifying the samples according to their origin with respect to the one built with both FAs and VOCs (Table 1). Moreover, also the percentage of correctly classified samples was lower: indeed, 706 out of the 855 classified samples (82.6%) of the training-set and 145 out of the 188 classified samples (77.1%) of the test-set were correctly classified, that is lower than the 86.2% obtained with the first model. In particular, the model was not able to correctly classify a sufficient % of samples from Italy and Portugal, suggesting that the molecules from the LOX pathway alone are not able to well distinguish these two classes from the others.

In order to improve the model capability in classification of samples according to their geographical origin, we built a third model, based on the same approach of the second model and using the same subsets of samples (Table 3), but selecting a new group of VOCs according to their F/F-crit ratio values, which had to be higher than 20. By this way, the first 25 VOCs in Table 2 were selected and the new model was built running LDA on the new matrix (M_{ij}2c, 1000 × 25) in which S_i is the *i*th of the 1000 samples in the training-set and C_j is the *j*th of the 25 selected VOCs. From a chemical point of view, 11 out of the 25 selected VOCs were alcohols, while the others were aldehydes (6), esters (4), ketones (2), carboxylic acids (1) and hydrocarbons (1), suggesting that alcohols are the chemical class more able to differentiating between the different origins, followed by aldehydes and esters. Again, the model was built to be able to classify samples in the categories Spain (S), Portugal (P), Italy (I), Greece (G), Tunisia (T) and Other (O) according to post classification probability (CP th %) higher than 50% for that category and its predictive capability was validated both internally, by leave-one-out cross-validation procedure, and externally, using the test-set as set of independent samples. Fig. 1 shows the score of each sample on the plane of the first two canonical variables (74% of the variance explained), while the obtained results are shown in Table 5. Overall, the improved model was able to classify 914 out of the 1000 samples (91.4%) of the training set and 205 out of the 217 samples (94.5%) of the test set. The capability of this model of classifying the samples according to their origin is definitely higher than that only based on the molecules from LOX pathway and only slightly lower than the one built with both FAs and VOCs. Also the percentage of correctly classified samples was strongly increased with respect to the second model: 806 out of the 914 classified samples (88.1%) of the training-set and 179 out of the 205 classified samples (87.3%) of the test-set were correctly

Table 4
Prediction obtained for each geographical origin using data from the 13 molecules originated from the LOX pathway. For each origin, two types of validation obtained by LDA are reported: the internal one was based on leave-one-out cross-validation run on the training-set; the external one was obtained applying the model on the external test-set. In each line, we reported the percentage and the number (in the brackets) of Virgin Olive Oil samples of that origin assigned by the model to the different origins. The figures in bold corresponds to the correctly assigned samples. The last two columns report the total number of samples belonging to the conditions of each line, and the percentage and the number (in the brackets) of samples classified by the model according to the criteria described in the text.

Origin	Validation	**Tunisia**	**Spain**	**Portugal**	**Italy**	**Greece**	**Other**	Total	Classified
Tunisia	Internal	**87.1% (61)**	**-**	–	10.0% (7)	1.4% (1)	1.4% (1)	74	94.6% (70)
	External	**100% (11)**	**-**	–	–	–	–	11	100% (11)
Spain	Internal	2.7% (7)	**86.7% (229)**	8.0% (21)	2.7% (7)	–	–	308	85.7% (264)
	External	–	**96.6% (28)**	–	3.5% (1)	–	–	32	90.6% (29)
Portugal	Internal	1.8% (1)	12.5% (7)	**75.0% (42)**	1.8% (1)	–	8.9% (5)	74	75.7% (56)
	External	–	18.2% (4)	**50.0% (11)**	–	13.6% (3)	18.2% (4)	24	91.7% (22)
Italy	Internal	11.5% (30)	3.1% (8)	3.4% (9)	**72.5% (190)**	0.8% (2)	8.8% (23)	308	85.1% (262)
	External	13.3% (11)	2.4% (2)	2.4% (2)	**66.3% (55)**	3.6% (3)	12.1% (10)	100	83.0% (83)
Greece	Internal	4.5% (8)	1.1% (2)	–	–	**93.3% (167)**	1.1% (2)	206	86.9% (179)
	External	2.9% (1)	2.9% (1)	–	–	**94.3% (33)**	–	40	87.5% (35)
Other	Internal	12.5% (3)	8.3% (2)	4.2% (1)	4.2% (1)	–	**70.8% (17)**	30	80.0% (24)
	External	–	–	12.5% (1)	–	–	**87.5% (7)**	10	80.0% (8)

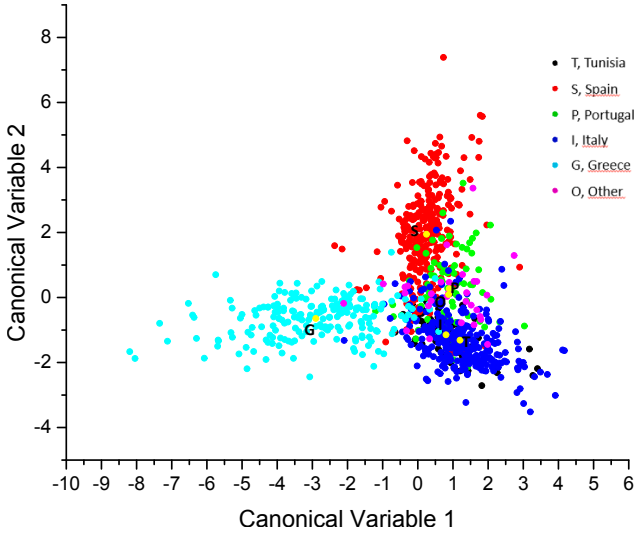

Fig. 1. Score of each Virgin Olive Oil sample on the plane of the first two canonical variables (74% of the variance explained). In yellow, the mean position of all the scores related to a single geographical origin.

classified, that is even higher than the 86.2% obtained with the first model, built using both FAs and VOCs (Table 6). In particular, this model is able to correctly classify samples from all the classes of geographical origin in percentages of at least 86%, reaching percentages higher than 97% for Greek and Tunisian samples when it was used in the external test-set. In this model, Portuguese samples are the ones worst classified, mainly because a certain percentage of them are classified as Spanish, likely because both the two countries belong to the Iberic peninsula, as discussed above.

All the prediction results obtained with the three proposed models are compared in Table 6. The first PCA-LDA model built using both FAs and VOCs gave the best results in terms of percentage of classified samples (97.2%), but the best results in terms of percentage of correctly classified samples (88.1%) were obtained using the third model, built using the 25 VOCs selected by highest F-ratio values, with an only slightly lower percentage of classified samples (94.5% in external validation). It is worth noting that external prediction of this latter model gave better results in terms of classified samples and comparable results in terms of correctly classified samples, suggesting a strong robustness of the proposed model.

Table 6
Comparison of the prediction results obtained with the three proposed models in terms of percentage of Virgin Olive Oil samples classified according to the proposed criteria and of percentage of samples correctly classified according to their geographical origin. Values are expressed in percentage. In the brackets, the prediction results obtained by external validation.

	Model	Classified samples (%)	Correctly classified samples (%)
1	**PCA-LDA (FAs + VOCs)**	97.2	86.3
2	**LDA (VOCs from LOX)**	85.5 (86.6)	82.6 (77.1)
3	**ANOVA-LDA (25 VOCs)**	91.4 (94.5)	88.1 (87.3)

For these reasons, for authentication of the geographical origin of commercial virgin olive oil samples, we propose the third model using LDA based on the quantitative data about the first 25 VOCs in Table 2 (in bold), selected after applying one-way ANOVA.

3.3. Application of the proposed model to binary mixtures of VOOs from different geographical origins

The prediction capability of the third model was further validated applying it for predict the geographical origin of 80 virgin olive oils constituted by binary mixtures of oils from Tunisia, Spain, Portugal, Italy and Greece. This approach is intended as preliminary, in that applied only to a set of available samples that are not homogenous in terms of representativeness of the five origins, with a prevalence of mixtures Spain/Portugal. The results are presented in Table 7 in which, for each sample, the real composition and the predicted composition by the model are indicated. Only for 5 out of the 80 samples (6.3%) the predicted main origin resulted in disagreement, for 11 (13.8%) it was confused between Spain and Portugal, and for 64 (80%) it was in agreement.

These preliminary results confirm again the reliability of the third model in authentication of the geographical origin of virgin olive oil. Further studies are desirable for confirming the usefulness of the proposed approach in authentication of the geographical origin of mixtures of virgin olive oils of different provenances, also aiming at proposing a reliable chemometric tool for discovering frauds related to the geographical origin of EVOOs.

Table 5
Prediction obtained for each geographical origin using data from the first 25 VOCs selected according to the highest *F*-ratio values (Table 2). For each origin two types of validation obtained by LDA are reported: the internal one was based on leave-one-out cross-validation run on the training-set; the external one was obtained applying the model on the external test-set. In each line, we reported the percentage and the number (in the brackets) of Virgin Olive Oil samples of that origin assigned by the model to the different origins. The figures in bold corresponds to the correctly assigned samples. The last two columns report the total number of samples belonging to the conditions of each line and the percentage and the number (in the brackets) of samples classified by the model according to the criteria described in the text.

Origin	Validation	Tunisia	Spain	Portugal	Italy	Greece	Other	Total	Classified
Tunisia	Internal	**91.2% (62)**	-	1.5% (1)	5.9% (4)	–	1.5% (1)	74	91.9% (68)
	External	**100% (11)**	-	–	–	–	–	11	100% (11)
Spain	Internal	2.6% (7)	**89.1% (245)**	3.3% (9)	5.1% (14)	–	–	308	89.3% (275)
	External	–	**90.3% (28)**	3.2% (1)	6.5% (2)	–	–	32	96.9% (31)
Portugal	Internal	3.2% (2)	12.7% (8)	**77.8% (49)**	1.6% (1)	–	4.8% (3)	74	85.1% (63)
	External	8.3% (2)	16.7% (4)	**66.7% (16)**	–	4.2% (1)	4.2% (1)	24	100% (24)
Italy	Internal	3.5% (10)	2.1% (6)	2.1% (6)	**86.9% (251)**	0.7% (2)	4.8% (14)	308	93.8% (289)
	External	2.1% (2)	2.1% (2)	3.2% (3)	**86.2% (81)**	2.1% (2)	4.3% (4)	100	94.0% (94)
Greece	Internal	0.5% (1)	1.0% (2)	–	3.6% (7)	**94.3% (182)**	0.5% (1)	206	93.7% (193)
	External	–	–	2.8% (1)	–	**97.2% (35)**	–	40	90.0% (36)
Other	Internal	15.4% (4)	7.7% (2)	7.7% (2)	3.9% (1)	–	**65.4% (17)**	30	86.7% (26)
	External	–	–	11.1% (1)	–	–	**88.9% (8)**	10	90.0% (9)

Table 7
Evaluation of the composition of binary mixtures of VOOs from different geographical origins by the third model. In the column "result", "ok" indicates that the predicted main origin is in agreement with the real composition, "no" indicates that it is not in agreement and "sp" indicates that the predicted main origin is Spain while the real main origin is Portugal or vice versa.

n°	Composition					**Result**	**Prediction**					
	S%	P%	T%	I%	G%		**S%**	**P%**	**T%**	**I%**	**G%**	**O%**
M-1	90		10			ok	**99**	**1**				
M-2	90		10			ok	**99**					**1**
M-3	90	10				ok	**93**	**6**				**1**
M-4	90	10				**no**	**34**	**27**				**39**
M-5	90	10				ok	**97**	**3**				
M-6	89		11			ok	**97**	**3**				
M-7	89		11			ok	**96**	**2**				**2**
M-8	89		11			ok	**94**	**4**		**1**		**1**
M-9	89	11				ok	**98**			**1**		**1**
M-10	88		12			ok	**99**	**1**				
M-11	88		12			ok	**98**	**2**				
M-12	88				12	sp		**100**				
M-13	85		15			ok	**99**	**1**				
M-14	85		15			ok	**91**	**2**	**2**	**3**		**2**
M-15	85	15				ok	**98**	**2**				
M-16	83	17				**no**	**8**	**2**				**90**
M-17	81		19			ok	**97**	**1**		**1**		**1**
M-18	81		19			ok	**100**					
M-19	81	19				ok	**88**	**6**				**6**
M-20	80		20			ok	**96**	**2**		**1**		**1**
M-21	80		20			ok	**89**	**8**		**2**		**1**
M-22	80	20				ok	**92**	**3**				**5**
M-23	80	20				ok	**81**	**10**				**9**
M-24	80	20				ok	**98**	**2**				
M-25	80	20				ok	**96**	**2**		**1**		**1**
M-26	80	20				ok	**98**	**2**				
M-27	80	20				sp		**100**				
M-28	80	20				ok	**93**	**7**				
M-29	79		21			ok	**66**	**1**	**10**	**13**	**2**	**8**
M-30	79		21			ok	**82**	**9**	**1**	**5**		**3**
M-31	79		21			ok	**80**	**2**	**2**	**9**		**7**
M-32	78	22				ok	**95**	**4**		**1**		
M-33	77	23				ok	**96**	**3**		**1**		
M-34	75	25				ok	**85**	**8**				**7**
M-35	75	25				ok	**97**					**3**
M-36	75	25				ok	**97**	**1**		**1**		**1**
M-37	75	25				ok	**98**	**2**				
M-38	73	27				ok	**86**	**9**				**5**
M-39	71	29				ok	**95**	**4**				**1**
M-40	71	29				ok	**95**	**4**		**1**		
M-41	71	29				ok	**98**	**1**				**1**
M-42	70	30				ok	**57**	**40**				**3**
M-43	70	30				ok	**64**	**35**				**1**
M-44	70	30				ok	**100**					
M-45	70	30				ok	**93**	**5**		**1**		**1**
M-46	70	30				ok	**90**	**3**		**5**	**1**	**1**
M-47	70	30				ok	**94**	**6**				
M-48	70	30				ok	**98**	**1**				**1**
M-49	69	31				sp	**17**	**80**		**1**		**2**
M-50	69				31	sp	**87**	**11**				**2**
M-51	67	33				ok	**97**	**2**		**1**		
M-52	65	35				ok	**94**	**4**		**1**		**1**
M-53	64	36				ok	**78**	**20**				**2**
M-54	64	36				ok	**86**	**4**				**10**
M-55	64	36				sp	**13**	**87**				
M-56	65	35				ok	**72**	**26**		**1**		**1**
M-57	61	39				ok	**86**	**13**				**1**
M-58	61	39				ok	**85**	**14**				**1**
M-59	60	40				ok	**66**	**34**				
M-60	60	40				ok	**90**	**3**	**1**	**4**		**2**
M-61	59	41				ok	**93**	**7**				
M-62	59	41				ok	**91**	**8**				**1**
M-63	59	41				ok	**97**	**2**				**1**
M-64	55	45				ok	**85**	**11**				**4**
M-65	55	45				ok	**74**	**15**		**1**		**10**
M-66	55	45				ok	**46**	**45**		**6**		**3**
M-67	50	50				sp	**79**	**19**		**1**		**1**
M-68	50				50	**no**	**99**	**1**				
M-69	31	69				sp	**96**	**3**				**1**
M-70	31	69				sp	**63**	**35**		**2**		

(*continued on next page*)

Table 7 (*continued*)

n°	Composition					Result	Prediction					
	S%	P%	T%	I%	G%		S%	P%	T%	I%	G%	O%
M-71	31	69				sp	**89**	**11**				
M-72	25	75				sp	**74**	**23**		**2**		**1**
M-73	24	76				sp	**90**	**10**				
M-74		75		25		ok		**97**		**3**		
M-75		55		45		**no**	**4**	**9**		**56**		**31**
M-76		55		45		**no**	**5**	**7**		**13**		**75**
M-77		55		45		ok	**5**	**5**		**46**		**44**
M-78		24			76	ok					**100**	
M-79		24			76	ok		**6**		**1**	**93**	
M-80		24			76	ok					**100**	

4. Conclusions

A reliable chemometric approach for authentication of the geographical origin of virgin olive oils from the main worldwide producing countries has been proposed in this manuscript. The model was built using quantitative data collected analyzing the volatile fraction of 1217 virgin olive oils by a HS-SPME-GC-MS quantitation method never applied for this purpose: the model applies Linear Discriminant Analysis only on 25 VOCs selected using one-way ANOVA and allows classifying oil samples according to the categories of Italy, Spain, Tunisia, Portugal, Greece and Other.

The main novelties with respect to the current literature are the use of a dataset based on more than 1200 commercial virgin olive oil samples from three consecutive olive oil crops, the application of a recently validated HS-SPME-GC-MS method, the simultaneous authentication of the geographical origin of virgin olive oils from the first five worldwide producing countries, and the application of the proposed chemometric approach for the identification of more than one origin in binary mixtures of VOOs.

The proposed model showed a very good predictive capability, with 87.3% of correctly classified samples during external validation, and percentages even higher than 97.1% for some specific origins. It showed a good predictive capability also when it was applied on binary mixtures of oils from different origins. Some attempts to improve the model should be tried searching for further volatile molecules able to differentiate between the different origins, enlarging the number of discriminated origins and developing further statistical models.

The approach proposed in this manuscript represents a very useful tool for the olive oil sector, easily applicable in testing laboratories for the quality control of virgin olive oils thus helping in protecting consumers and producers from frauds.

CRediT authorship contribution statement

Lorenzo Cecchi: Formal analysis, Conceptualization, Methodology, Investigation, Writing - original draft. **Marzia Migliorini:** Formal analysis, Conceptualization, Investigation. **Elisa Giambanelli:** Formal analysis, Investigation. **Adolfo Rossetti:** Resources, Supervision. **Anna Cane:** Resources, Project administration. **Nadia Mulinacci:** Supervision, Conceptualization, Writing - review & editing. **Fabrizio Melani:** Conceptualization, Data curation, Methodology, Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research is part of the FOODOLEAPLUS project (financial support POR-FSE 2014-2020, Regione Toscana, Italy).

Abbreviations

ANOVA	Analysis of Variance
CAD	Charged Aerosol Detector
EVOO	extra virgin olive oil
FAs	fatty acids
FAMEs	fatty acid methyl esters
FID	Flame Ionization Detector
HCA	Hierarchical Cluster Analysis
HPLC	High Performance Liquid Chromatogrpahy
HS-SPME-GC-MS	head space-solid phase micro extraction-gas chromatograph-mass detector
IRMS	isotope ratio analysis
LIBS	Laser Induced Breakdown Spectroscopy
LOX	lipoxygenase
LSD	Least Significant Difference
LVOO	lampante virgin olive oil
MS	Mass Detector
NMR	nuclear magnetic resonance
PCA	Principal Component Analysis
PLS-DA	Partial Least Squares Discriminant Analysis
SIFT-MS	Selected Ion Flow Tube Mass Spectrometry
VOC	Volatile Organic Compound
VOO	virgin olive oil

References

Angerosa, F. (2002). Influence of volatile compounds on virgin olive oil quality evaluated by analytical approaches and sensor panels. *European Journal of Lipid Science and Technology, 104*, 639–660. https://doi.org/10.1002/1438-9312(200210)104:9/10<639::AID-EJLT639>3.0.CO;2-U.

Angerosa, F., Lanza, B., D'Alessandro, N., Marsilio, V., & Cumitini, S. (1999). Olive oil off-odour compounds produced by *aspergillus* and *penicillum*. *Acta Horticulturae, 474*, 695–699. https://doi.org/10.17660/ActaHortic.1999.474.144.

Angerosa, F., Mostallino, R., Basti, C., & Vito, R. (2001). Influence of malaxation temperature and time on the quality of virgin olive oil. *Food Chemistry, 72*, 19–28. https://doi.org/10.1016/S0308-8146(00)00194-1.

Aparicio, R., Rocha, S. M., Delgadillo, I., & Morales, M. T. (2000). Detection of rancid defect in virgin olive oil by the electronic nose. *Journal of Agricultural and Food Chemistry, 48*, 853–860. https://doi.org/10.1021/jf9814087.

Bajoub, A., Medina-Rodriguez, S., Ajal, E. A., Cuadros-Rodriguez, L., Monasterio, R. P., Vercammen, J., et al. (2018). A metabolic fingerprinting approach based on selected ion flow tube mass spectrometry (SIFT-MS) and chemometrics: A reliable tool for mediterranean origin-labeled olive oil authentication. *Food Research International, 106*, 233–242. https://doi.org/10.1016/j.foodres.2017.12.027.

Ben Mansour, A., Chtourou, F., Khbou, W., Flamini, G., & Bouaziz, M. (2017). Phenolic and volatile compounds of Neb Jmel olive oil cultivar according to their geographical origin using chemometrics. *European Food Research and Technology, 243*, 403–418. https://doi.org/10.1007/s00217-016-2754-5.

Ben Mohamed, M., Rocchetti, G., Montesano, D., Ben Ali, S., Guasmi, F., Grati-Kamoun,

N., et al. (2018). Discrimination of Tunisian and Italian extra-virgin olive oils according to their phenolic and sterolic fingerprints. *Food Research International, 106*, 920–927. https://doi.org/10.1016/j.foodres.2018.02.010.

Berlioz, B., Cordella, C., Cavalli, J. F., Lizzani-Cuvelier, L., Loiseau, A. M., & Fernandez, X. (2006). Comparison of the amounts of volatile compounds in French protected designation of origin virgin olive oils. *Journal of Agricultural and Food Chemistry, 54*, 10092–10101. https://doi.org/10.1021/jf061796+.

Cajka, T., Riddellova, K., Klimankova, E., Cerna, M., Pudil, F., & Hajslova, J. (2010). Traceability of olive oil based on volatiles pattern and multivariate analysis. *Food Chemistry, 121*, 282–289. https://doi.org/10.1016/j.foodchem.2009.12.011.

Camin, F., Larcher, R., Nicolini, G., Bontempo, L., Bertoldi, D., et al. (2010). Isotopic and elemental data for tracing the origin of European olive oils. *Journal of Agricultural and Food Chemistry, 58*(1), 570–577. https://doi.org/10.1021/jf902814s.

Campestre, C., Angelini, G., Gasbarri, C., & Angerosa, F. (2017). The compounds responsible for the sensory profile in monovarietal virgin olive oils. *Molecules, 22*, 1833–1860. https://doi.org/10.3390/molecules22111833.

Cecchi, L., Migliorini, M., Giambanelli, E., Rossetti, A., Cane, A., Melani, F., et al. (2019). Headspace solid-phase microextraction-gas chromatography-mass spectrometry quantification of the volatile profile of more than 1200 virgin olive oils for supporting the Panel test in their classification: Comparison of different chemometric approaches. *Journal of Agricultural and Food Chemistry, 67*, 9112–9120. https://doi.org/10.1021/acs.jafc.9b03346.

Choe, E., & Min, D. B. (2006). Mechanism and factors for edible oil oxidation. *Comprehensive Reviews in Food Science and Food Safety, 5*, 169–186. https://doi.org/10.1111/j.1541-4337.2006.00009.x.

Dais, P., & Hatzakis, E. (2013). Quality assessment and authentication of virgin olive oil by NMR spectroscopy: A critical review. *Analytica Chimica Acta, 765*, 1–27. https://doi.org/10.1016/j.aca.2012.12.003.

Di Bella, G., Maisano, R., La Pera, L., Lo Turco, V., Salvo, F., & Dugo, G. (2007). Statistical characterization of Sicilian olive oils from the Peloritana and Maghrebian zones according to the fatty acid profile. *Journal of Agricultural and Food Chemistry, 55*, 6568–6574. https://doi.org/10.1021/jf070523r.

European Community (2006). Council Regulation (EC) No. 510/2006 of 20 March 2006 on the protection of geographical indications and designations of origin for agricultural products and foodstuffs. *Off. J. Eur. Commun. L93*, 12–25.

European Economic Community (1991). Commission Regulation (EEC) No. 2568/1991 of 11 July 1991 on the characteristic of olive oil and olive-residue oil and on the relevant methods of analysis. *Off. J. Eur. Commun. L248*, 1–83.

Fortini, M., Migliorini, M., Cherubini, C., Cecchi, L., & Calamai, L. (2017). Multiple internal standard normalization for improving HS-SPME-GC-MS quantitation in virgin olive oil volatile organic compounds (VOO-VOCs) profile. *Talanta, 165*, 641–652. https://doi.org/10.1016/j.talanta.2016.12.082.

Fregapane, G., & Salvador, M. D. (2019). Chemical and sensory characteristics of extra virgin olive oils produced in central Iberian peninsula under the protected designation of origin quality scheme. *European Journal of Lipid Science and Technology, 121*, 1800134. https://doi.org/10.1002/ejlt.201800134.

Garcia, R., Martins, N., & Cabrita, N. J. (2013). Putative markers of adulteration of extra virgin olive oil with refined olive oil: Prospects and limitations. *Food Research International, 54*, 2039–2044. https://doi.org/10.1016/j.foodres.2013.05.008.

Gazeli, O., Bellou, E., Stefas, D., & Couris, S. (2020). Laser-based classification of olive oils assisted by machine learning. *Food Chemistry, 302*, 125329. https://doi.org/10.1016/j.foodchem.2019.125329.

Ghisoni, S., Lucini, L., Angilletta, F., Rocchetti, G., Farinelli, D., Tombesi, S., et al. (2019). Discrimination of extra-virgin olive oils from different cultivars and geographical origins by untargeted metabolomics. *Food Research International, 121*, 746–753. https://doi.org/10.1016/j.foodres.2018.12.052.

Gumus, Z. P., Celenk, V. U., Tekin, S., Yurdakul, O., & Ertas, H. (2017). Determination of trace elements and stable carbon isotope ratios in virgin olive oils from Western Turkey to authenticate geographical origin with a chemometric approach. *European Food Research and Technology, 243*, 1719–1727. https://doi.org/10.1007/s00217-017-2876-4.

Ingallina, C., Cerreto, A., Mannina, L., Circi, S., Vista, S., Capitani, D., et al. (2019). Extra-virgin olive oils from nine Italian regions: An ^{1}H-NMR-chemometric characterization. *Metabolites, 9*, 65. https://doi.org/10.3390/metabo9040065.

IOC/T.20/Doc No 33/Rev1. (2017). *Official method of analysis. Determination of fatty acid methyl esters by gas chromatography*. International Olive Council. Available online: http://www.internationaloliveoil.org/estaticos/view/224-testing-methods (accessed on September 28th, 2019) .

Kalua, C. M., Allen, M. S., Bedgood, D. R., Bishop, A. G., Prenzler, P. D., & Robards, K. (2007). Olive oil volatile compounds, flavor development and quality: A critical review. *Food Chemistry, 100*, 273–286. https://doi.org/10.1016/j.foodchem.2005.09.059.

Kosma, I., Vatavali, K., Kontakos, S., Kontominas, M., Kiritsakis, A., & Badeka, A. (2017). Geographical differentiation of Greek extra virgin olive oil from late-harvested Koroneiki cultivar fruits. *Journal of the American Oil Chemists Society, 94*, 1373–1384. https://doi.org/10.1007/s11746-017-3036-5.

Longobardi, F., Ventrella, A., Casiello, G., Sacco, D., Tasioula-Margar, M., Kiritsakis, A. K., et al. (2012). Characterization of the geographical origin of Western Greek virgin olive oils based on instrumental and multivariate statistical analysis. *Food Chemistry, 133*, 169–175. https://doi.org/10.1016/j.foodchem.2011.09.130.

Lukic, I., Carlin, S., Horvat, I., & Vrhovsek (2019). Combined targeted and untargeted profiling of volatile aroma compounds with comprehensive two-dimensional gas chromatography for differentiation of virgin olive oils according to variety and geographical origin. *Food Chemistry, 270*, 403–414. https://doi.org/10.1016/j.foodchem.2018.07.133.

Mallamace, D., Vasi, S., Corsaro, C., Naccari, C., Clodoveo, M. L., Dugo, G., et al. (2017). Calorimetric analysis points out the physical-chemistry of organic olive oils and reveals the geographic origin. *Physica A, 486*, 925–932. https://doi.org/10.1016/j.physa.2017.06.015.

Melucci, D., Bendini, A., Tesini, F., Barbieri, S., Zappi, A., Vichi, S., et al. (2016). Rapid direct analysis to discriminate geographical origin of extra virgin olive oils by flash gas chromatography electronic nose and chemometrics. *Food Chemistry, 204*, 263–273. https://doi.org/10.1016/j.foodchem.2016.02.131.

Morales, M. T., Luna, R., & Aparicio, R. (2005). Comparative study of virgin olive oil sensory defects. *Food Chemistry, 91*, 293–301. https://doi.org/10.1016/j.foodchem.2004.06.011.

Morales, M. T., Rios, J. J., & Aparicio, R. (1997). Changes in the volatile composition of virgin olive oil during oxidation: Flavors and off-flavors. *Journal of Agricultural and Food Chemistry, 45*, 2666–2673. https://doi.org/10.1021/jf960585+.

Ouni, Y., Flamini, G., Issaoui, M., Nabil, B. Y., Cioni, P. L., Hammami, M., et al. (2011). Volatile compound and compositional quality of virgin olive oil from oueslati variety: Influence of geographical origin. *Food Chemistry, 124*, 1770–1776. https://doi.org/10.1016/j.foodchem.2010.08.023.

Persuric, Z., Saftic, L., Masek, T., & Pavelic, S. K. (2018). Comparison of triacylglycerol analysis by MALDI-TOD/MS, fatty acid analysis by GC-MS and non-selective analysis by NIRS in combination with chemometrics for determination of extra virgin olive oil geographical origin. A case study. *LWT – Food Science and Technology, 95*, 326–332. https://doi.org/10.1016/j.lwt.2018.04.072.

Pizarro, C., Rodriguez-Tecedor, S., Perez-del-Notario, N., & Gonzalez-Saiz, J. M. (2011). Recognition of volatile compound as markers in geographical discrimination of Spanish extra virgin olive oils by chemometric analysis of non-specific chromatography volatile profiles. *Journal of Chromatography A, 1218*, 518–523. https://doi.org/10.1016/j.chroma.2010.11.045.

Portarena, S., Baldacchini, C., & Brugnoli, E. (2017). Geographical discrimination of extra-virgin olive oils from the Italian coasts by combining stable isotope data and carotenoid content within a multivariate analysis. *Food Chemistry, 215*, 1–6. https://doi.org/10.1016/j.foodchem.2016.07.135.

Pouliarekou, E., Badeka, A., Tasioula-Margari, M., Kontakos, S., Longobardi, F., & Kontominas, M. G. (2011). Characterization and classification of Western Greek olive oil according to cultivar and geographical origin based on volatile compounds. *Journal of Chromatography A, 1218*, 7534–7542. https://doi.org/10.1016/j.chroma.2011.07.081.

Quintanilla-Casas, B., Bertin, S., Leik, K., Bustamante, J., Guardiola, F., Valli, E., et al. (2020). Profiling versus fingerprinting analysis of sesquiterpene hydrocarbons for the geographical authentication of extra virgin olive oils. *Food Chemistry, 307*, 125556. https://doi.org/10.1016/j.foodchem.2019.125556.

Sayago, A., Gonzalez-Dominguez, R., Beltran, R., & Fernandez-Recamales, A. (2018). Combination of complementary data mining methods for geographical characterization of extra virgin olive oils based on mineral composition. *Food Chemistry, 261*, 42–50. https://doi.org/10.1016/j.foodchem.2018.04.019.

Souayah, F., Rodrigues, N., Veloso, A. C. A., Dias, L. G., Pereira, J. A., Oueslati, S., et al. (2017). Discrimination of olive oil by cultivar, geographical origin and quality using potentiometric electronic tongue fingerprints. *Journal of the American Oil Chemists Society, 94*, 1417–1429. https://doi.org/10.1007/s11746-017-3051-6.

Techer, I., Medini, S., Janin, M., & Arregui, M. (2017). Impact of agricultural practice on the Sr isotopic composition of food products: Application to discriminate the geographic origin of olives and olive oil. *Applied Geochemistry, 82*, 1–14. https://doi.org/10.1016/j.apgeochem.2017.05.010.

Trapani, S., Breschi, C., Cecchi, L., Guerrini, L., Mulinacci, N., Parenti, A., et al. (2017). Indirect indices of oxidative damage to phenolic compounds for the implementation of olive paste malaxation optimization charts. *Journal of Food Engineering, 207*, 24–34. https://doi.org/10.1016/j.jfoodeng.2017.03.012.

Union, E. (2012). Regulation (EU) No. 1151/2012 of the european parliament and of the council of 21 November 2012 on quality schemes for agricultural products and foodstuffs. *Off. J. Eur. Commun. L343*, 1–29.

Vera, D. N., Jimenez-Carvelo, A. M., Cuadros-Rodriguez, L., Ruisanchez, I., & Pilar Callao, M. (2019). Authentication of the geographical origin of extra-virgin olive oil of the Arbequina cultivar by chromatographic fingerprinting and chemometrics. *Talanta, 203*, 194–202. https://doi.org/10.1016/j.talanta.2019.05.064.

Vichi, S., Pizzale, L., Conte, L. S., Buxaderas, S., & Lopez-Tamames, E. (2003). Solid-phase microextraction in the analysis of virgin olive oil volatile fraction: Characterization of virgin olive oils from two distinct geographical areas of northern Italy. *Journal of Agricultural and Food Chemistry, 51*, 6572–6577. https://doi.org/10.1021/jf030269c.

Zunin, P., Boggia, R., Salvadeo, P., & Evangelisti, F. (2005). Geographical traceability of West Liguria extravirgin olive oils by the analysis of volatile terpenoid hydrocarbons. *Journal of Chromatography A, 1089*(1–2), 243–249. https://doi.org/10.1016/j.chroma.2005.07.005 In this issue.