Food Chemistry 204 (2016) 263–273

Contents lists available at ScienceDirect

Food Chemistry

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

Rapid direct analysis to discriminate geographic origin of extra virgin olive oils by flash gas chromatography electronic nose and chemometrics

Dora Melucci [a], Alessandra Bendini [b,*], Federica Tesini [b], Sara Barbieri [b], Alessandro Zappi [a], Stefania Vichi [c], Lanfranco Conte [d], Tullia Gallina Toschi [b]

[a] Department of Chemistry Ciamician, University of Bologna, Via Selmi, 2, 40126 Bologna, Italy
[b] Department of Agricultural and Food Sciences (DiSTAL), University of Bologna, P.zza Goidanich 60, 47521 Cesena, Italy
[c] Department of Food Science and Nutrition, University of Barcelona, Food and Nutrition Torribera Campus, Av. Prat de la Riba, 171,, S.ta Coloma de Gramenet, Spain
[d] Department of Food Science, University of Udine, Via Sondrio 2/a, 33100 Udine, Italy

ARTICLE INFO

Article history:
Received 7 September 2015
Received in revised form 11 January 2016
Accepted 21 February 2016
Available online 23 February 2016

Keywords:
Extra virgin olive oil
Geographic origin
FGC E-nose
SPME/GC–MS
Headspace volatile compounds
Non-target analysis
PCA
PLS-DA

ABSTRACT

At present, the geographical origin of extra virgin olive oils can be ensured by documented traceability, although chemical analysis may add information that is useful for possible confirmation. This preliminary study investigated the effectiveness of flash gas chromatography electronic nose and multivariate data analysis to perform rapid screening of commercial extra virgin olive oils characterized by a different geographical origin declared in the label. A comparison with solid phase micro extraction coupled to gas chromatography mass spectrometry was also performed. The new method is suitable to verify the geographic origin of extra virgin olive oils based on principal components analysis and discriminant analysis applied to the volatile profile of the headspace as a fingerprint. The selected variables were suitable in discriminating between "100% Italian" and "non-100% Italian" oils. Partial least squares discriminant analysis also allowed prediction of the degree of membership of unknown samples to the classes examined.

1. Introduction

In an increasingly globalized world, certification of food quality is one of the most important goals for scientists in the agri-food sector. Consumer demand of traceability and authenticity of food products is also increasing, and the international agencies dealing with food quality have recently published specific guidelines in this regard (FAO, 2003). Extra virgin olive oil (EVOO) is a typical Mediterranean food product characterized by a multi-millenary tradition that arouses great appreciation among consumers. Within the Mediterranean basin, Italy is a key producer of olive oil. The vast economic interests may give rise to illegal activities aimed to increase profit, such as a false declaration of geographic origin, thus falsifying traceability and, consequently, authenticity of the product. The European Union (EU) has recently concluded a decennial *iter* to establish regulations about olive oil with the aim of regulating production and commercialization of this important product. Regulation EU No. 1019/02 defined how to correctly pack and label oils, and the last Commission Implementing Regulation, 2013 EU No. 1335/13 made it obligatory to indicate the geographic origin on the label. In EU Regulation No. 29/12 (European Commission Implementing Regulation, 2012), it is reported that in order to ensure that consumers are not misled and the olive oil market is not distraught, information concerning the geographic area in which olives are harvested and olive oil is obtained should be stated on the packaging or labels. For greater clarification, the document also defines that simple provisions as 'blend of olive oils of European Union origin' or 'blend of olive oils not of European Union origin' or 'blend of olive oils of European Union origin and not of European Union origin' should be stated for labeling of origin.

The mandatory necessity of certifying the geographical origin makes it highly desirable to assess origin not only by documentation of verification, but also by rapid analytical methods. In this regard, it is necessary to apply high performance instrumental analytical methods, and the large number of variables imposes the use of chemometrics, whose outputs provide useful and easy-to-visualize information extracted from data while simultaneously

* Corresponding author.
E-mail addresses: dora.melucci@unibo.it (D. Melucci), alessandra.bendini@unibo.it (A. Bendini), federica.tesini@unibo.it (F. Tesini), sara.barbieri@unibo.it (S. Barbieri), alessandro.zappi12@gmail.com (A. Zappi), stefaniavichi@ub.edu (S. Vichi), lanfranco.conte@uniud.it (L. Conte), tullia.gallinatoschi@unibo.it (T. Gallina Toschi).

http://dx.doi.org/10.1016/j.foodchem.2016.02.131
0308-8146/

discarding useless information (analytical noise and redundant information).

There is an urgent need to extend the representativeness of a database established on chromatographic, spectroscopic, and spectrometric compositional data profiles to clearly identify the most promising techniques in order to confirm the geographic origin of EVOOs and verify the conformity of label-declared geographic origin, as well as to provide one or more harmonized methods for sharing markers that are useful to check the product's conformity to specific standards (e.g., geographic origin). All the factors identified by compositional analysis of EVOOs are important. Mass spectrometry together with various spectrometric and chromatographic analytical techniques have been applied to determine the chemical composition, and many of these instrumental analytical techniques have been used in tandem with chemometrics (Gouvinhas, De Almeida, Carvalho, Machado, & Barros, 2015; Azizian et al., 2015; Mendes et al., 2015; Sinelli et al., 2010; Diraman & Dibeklioğlu, 2009). In this context, adulteration of EVOOs has been studied by liquid chromatography (HPLC), gas chromatography (GC), and linear discriminant analysis (LDA) using fatty acids (FA) and triacylglycerols (TGs) as markers (Ollivier, Artaud, Pinatel, Durbec, & Guerere, 2006; Jabeur et al., 2014). HPLC-mass spectrometry (MS) and LDA allowed determination of the phenolic profile for discrimination of geographical origin (Taamalli, Arráez Román, Zarrouk, Segura-Carretero, & Fernán dez-Gutiérrez, 2012). In particular, specific volatile compounds or their classes (e.g., terpenoid compounds) have been used to discriminate EVOO samples according to geographic origin (Ben Temime, Campeol, Cioni, Daoud, & Zarrouk, 2006; Cecchi & Alfei, 2013; Vichi, Pizzale, Conte, Buxaderas, & Lopez-Tamames,, 2003; Zunin, Boggia, Salvadeo, & Evangelisti, 2005). Many EVOOs have also been classified according to their geographic origin using the combination of FA and/or TG profiles with other compounds such as sterols, polyphenols, and volatiles using conventional and new analytical approaches, as recently reviewed (Gallina Toschi, Bendini, Lozano-Sanchez, Segura-Carretero, & Conte, 2013; García-González, Luna, Morales, & Aparicio, 2009). Several publications have described the use of volatile-species distribution as a fingerprint to assess traceability, authentication, and non-degradation based on head-space sampling and GC in tandem with several chemometric tools: analysis of variance (ANOVA) and correlation analysis (Cecchi & Alfei, 2013); principal components analysis (PCA) (Cimato et al., 2006); LDA (Pouliarekou et al., 2011); PCA and hierarchical clustering analysis (HCA) (Procida, Giomo, Cichelli, & Conte, 2005).

Among the chemical species in EVOO, many volatiles have been related with specific sensory characteristics (Aparicio, Morales, & Alonso, 1996; Cerretani, Salvador, Bendini, & Fregapane, 2008). Over the last decade, "e-sensing" technologies have undergone important developments from a technical and commercial point of view, and electronic noses have been designed to mimic the human sense of olfaction in order to detect and recognize flavors and off-flavors in different food matrices (Peng, Tian, Chen, Li, & Gao, 2015). Moreover, the electronic nose results have been successfully correlated to those obtained with other techniques (sensory, GC, and GC–MS) (Mildner-Szkudlarz & Jelen, 2008; Lerma-García et al., 2010).

In a traditional multivariate approach, the variables are concentrations of several compounds: this means that the scientist chooses beforehand which chemical species are relevant; in contrast, when tools like PCA or partial least squares discriminant analysis (PLS-DA) are applied to *full chromatograms*, there is no risk to discard species with retention times not corresponding to chemical species already known to influence EVOO quality. The advantages of such an approach have recently been described (Melucci et al., 2013).

The aim of this study was to analyze the headspace profile of commercial EVOOs with different geographic origin using electronic nose with ultra fast gas chromatography (FGC E-nose), which is able to perform the separation on two short columns of different polarities working in parallel and detect analytes with a flame ionization detector (FID). The FGC E-nose was used to discriminate between products labeled as "100% Italian EVOO" and "non-100% Italian" coming from other countries in the EU, and in particular Spain and Greece. PCA, LDA, and HCA were applied as exploratory tools. Data processing was initially applied to datasets made from peak areas at retention times corresponding to significant species; in this case, a comparison between the non-target analysis performed by FGC E-nose and SPME/GC–MS achieved two purposes: (i) to demonstrate that the discriminating power of FGC E-nose was comparable with SPME/GC–MS; (ii) to assign FGC E-nose retention times to specific volatile compounds. In a second step, the full chromatograms, obtained on two different sets of samples analyzed in two different laboratories, were processed by applying PLS-DA as a chemometric tool.

2. Materials and methods

2.1. Samples

The two sets of samples named Set A and Set B were formed by 27 and 251 EVOOs, respectively, and were collected from COOP Italia before distribution by the supermarket chain (COOP Italia is a consortium that acts as a central retailer and is one of the most important supermarket chains in Italy; it also carries out marketing activities and performs quality control). Set A was composed of 5 PDO (Protected Designation of Origin) and PGI (Protected Geographical Indication) Italian samples, 13 samples declared as produced and processed exclusively in Italy (100% Italian, I code), and 9 samples produced in countries which are members of the European Union (Mixtures, M code). All samples in Set A were collected during the 2012–2013 harvest period. Set B included 132 samples labeled as 100% Italian (I) and 119 samples labeled as non-100% Italian (M) EVOOs collected during the 2013–2014 harvest period. Even if the actual identity of the samples was confidential, all the olive oils were bottled (in dark or transparent glass bottles) in Italy. Moreover, samples considered as 100% Italian were assumed to be as declared, according to specific quality control checks, and based on chemometric control with single-class PCA models and Hotelling analysis for outliers elimination applied to confirm the geographic class. All samples were stored at 10 °C in darkness before analysis.

2.2. Sensory Evaluation

A IOC panel test method was carried out on samples in Set A by a group of 8 selected trained assessors, all members of the Professional Committee DiSTAL. Sample evaluation was performed according to the official procedure (Reg. (EC) 640/2008). Moreover, the presence of green notes and other positive attributes were evaluated with reference to the list of descriptors for PDO EVOOs developed and agreed by the International Olive Oil Council, 2005 (IOOC/T.20/Doc. No. 22, 2005).

2.3. FGC E-Nose

The same type of FGC E-nose Heracles II (AlphaMos, Toulouse, France) was used for both sets of samples but in two different laboratories (Set A was analyzed in Toulouse, Set B in the laboratory of COOP Italia in Bologna, Italy). The Heracles II was equipped with two columns working in parallel mode: a non-polar column

(MXT5: 5% diphenyl, 95% methylpolysiloxane, 10 m length and 180 μm diameter) and a slightly polar column (MXT1701: 14% cyanopropylphenyl, 86% methylpolysiloxane, 10 m length and 180 μm diameter). A single comprehensive chromatogram was created by joining the chromatograms obtained with the two columns; such an approach may help in preventing/reducing incorrect identifications due to overlapping of chromatograms obtained with two different columns, and represents a useful tool for improved identification. An aliquot of each sample (2 g ± 1%) was placed in a 20 mL vial and sealed with a magnetic plug. The vial was placed in the Heracles' auto-sampler, which placed it in a shaker oven where it remained for 20 min at 50 °C, shaken at 500 rpm. Next, a syringe pierced the silicone septum of the magnetic plug and sampled 5 ml of the head space. Prior to the chromatographic separation, the 5-ml headspace aliquot was adsorbed on a CARBOWAX trap maintained at 40 °C for 65 s while the carrier gas (H_2) flowed through it in order to concentrate the analytes and to remove excess air and moisture. Subsequently, desorption was obtained by increasing the temperature of the trap up to 240 °C in 93 s and the sample was injected. The thermal program started at 40 °C (held for 2 s) and increased up to 270 °C at 3 °C s^{-1}; the final temperature was held for 21 s. The total separation time was 100 s. At the end of each column, a FID detector was placed and the acquired signal was digitalized every 0.01 s. For calibration, an alkane solution (from *n*-hexane to *n*-hexadecane) was used to convert retention time in Kovats indices and identify the volatile compounds using specific software (AromaChemBase). Samples were analyzed in triplicate or quadruplicate for both Set A and Set B.

2.4. SPME/GC–MS

The headspace composition was investigated by SPME coupled to GC separation and MS detection. This same analysis was performed in two different laboratories: samples in Set A were analyzed at the University of Bologna (Italy), whereas the laboratory of the University of Barcelona (Spain) performed analysis on Set B. The same kind of instrument, a gas chromatograph Agilent 6890 N Network and a quadrupolar mass-selective spectrometry Agilent 5973 Network detector (Agilent Technologies, Palo Alto, CA, USA), provided with a split–splitless injection port and helium as the carrier gas (linear velocity of 17 cm s^{-1}) was used. Slight differences in analytical conditions were applied.

For analysis of Set A: SPME was carried out by weighing 1.5 g of sample, spiked with 4-methyl-2-pentanone (internal standard dissolved in refined sunflower oil) to a concentration of 10 mg kg^{-1} in a 10 mL vial fitted with a silicone septum. The vial was placed in a water bath at 40 °C and maintaining the oil sample under magnetic stirring for 2 min (conditioning) and then a DVB/CAR/PDMS fiber (50/30 μm, 2 cm long from Supelco Ltd., Bellefonte, PA) was exposed for 30 min in the headspace of the sample. After exposition, the fiber was retracted into the needle and immediately desorbed for 3 min in the injection port of a gas chromatograph (250 °C). Compounds were separated on a ZB-WAX column 30 m, 0.25 mm ID, 1.00 μm film thickness (Chemtek Analytic, Bologna, Italy). Column temperature was held at 40 °C for 10 min and increased to 200 °C at 3 °C min^{-1}. The ion source and transfer line were at 180 °C and 230 °C, respectively. Electron impact mass spectra were recorded at 70 eV ionization energy in the 20–250 amu mass range, 2 scans s^{-1}.

For analysis of Set B, SPME extraction was performed according to Vichi et al. (2003) and differed from the method applied for Set A only for use of a different internal standard, 4-methyl-2-pentanol (Sigma–Aldrich, St. Louis, MO). The fiber was then desorbed at 260 °C in the gas chromatograph injection port for 5 min. Separation of compounds was performed on two columns with distinct polarity: Supelcowax-10 and Equity-5 (both 30 m × 0.25 mm I.D., 0.25 μm film thickness), both purchased from Supelco (Supelco Ltd., Bellefonte, PA, USA). The column temperature was held at 40 °C for 5 min and increased to 200 °C at 4 °C min^{-1}. The injector temperature was 260 °C, and the transfer line temperature was 280 °C. Electron impact mass spectra were recorded at 70 eV ionization energy in the 30–300 amu mass range, 2 scans s^{-1}.

Identification of volatile compounds was mainly carried out by a comparison of mass spectral data with information from the National Institute of Standards and Technology (NIST) library (2005 version) and checked with pure standards. Linear retention indexes were also calculated and compared with those available in the literature. Relative amounts of volatile compounds were expressed as mg of internal standard per kg of oil, applying a response factor of 1. All determinations were carried out in triplicate or duplicate for Set A and Set B, respectively.

2.5. Software

The FGC E-nose data processing was carried out with Alphasoft V12.44 and AroChembase software. XLSTAT version 2011.1.03 software (Addinsoft, USA) was used to elaborate ANOVA and PCA on Set A. Preliminary PCA on Set B and PLS-DA were performed using The Unscrambler version 9.8 (CAMO, Norway).

2.6. Chemometrics

In this work, a first explorative step was carried out using peak areas that were automatically calculated by the software that controls each instrument. All data based on peak area were preprocessed by autoscaling.

Principal component analysis is a well-known chemometric procedure which rotates the original space to another one whose versors are the principal components (PCs) oriented along directions containing the maximum explained variance (EV) and mutually orthogonal. Score and loadings plots are obtained, allowing for easy visualization of samples and variables and verification of their role in the analytical problem. Hotelling analysis, applied to PCA scores, calculates the covariance ellipsoid corresponding to 95% confidence level (and visually draws it on the scores plot); therefore, samples falling outside of the ellipsoid are those in the multivariate Gaussian tails and may be considered outliers and discarded from further analyses. Linear discriminant analysis is a multivariate classification tool which rotates the original space, but unlike PCA its aim is to maximize separation between classes, minimizing at the same time distances between objects in the same class; in this way, new objects may be projected onto this new scores space and assigned to one of the classes of the training set. HCA may also be applied to identify eventual sub-classes by calculating multidimensional Euclidean distances between objects and grouping those closest to each other. In the present investigation, it was highly expected that various sub-categories may be included in the very broad category "non-100% Italian" (M, for example mixtures from Spain, Greece, Italy).

Once the preliminary exploration by PCA, HCA, and LDA was completed, the work was extended by creating models, or equations involving experimental variables. A very useful response variable is the degree of belonging of objects to the possible classes involved in the analytical problem. The main interest was in quantifying the degree of belonging to class I (y_I) and the degree of belonging to class M (y_M). Few tens of objects are available while up to thousands of variables (digitized signal) are generated by a FGC E-nose chromatogram. Thus, the only adequate modeling tool is PLS regression (in particular, PLS-DA), which exploits PCs and maximizes both EV and correlation between regressors (the variables, that is the chromatographic signals at various retention

times) and the response (degree of belonging, y). The choice of using full chromatograms has important advantages: (i) no pre-selection of significant retention times is needed, thus by-passing the non-target character of FID signals; when no pre-selection is done, the risk of discarding useful information is avoided; (ii) errors related to incorrect integration in peak-area calculation are avoided. Of course, some disadvantages must also be considered when using whole chromatograms as predictors: a number of correlated variables much higher than the number of objects may lead to overfitting, which provides modeling noise instead of useful information. However, chemometric modeling tools offer reliable methods for controlling these problems to obtain good performance of PLS-regression, based on objective measures: in particular, root mean square error (RMSE) and correlation coefficients. Predictive ability (also for LDA) was evaluated by the well-known cross-validation (CV) procedure (Brereton, 2007).

3. Results and discussion

3.1. Explorative analysis of sample Sets A and B

The exploration of Set A was considered as a preliminary step in the method development as it was the first to be analyzed and consequently taken into account to better define a chemometric approach for discriminating such a large number of olive oil samples subsequently studied. This first set of 27 samples was very useful for exploring Set B in depth and in establishing the method.

3.2. PCA from SPME/GC–MS peak areas of Set A

According to the sensory analysis performed by IOC panel test method, the 27 samples of Set A were classified as EVOO (8 samples) and VOO (19 samples); for EVOOs, the intensity of fruity was light (4 samples) and medium (4 samples), and the presence of secondary notes (olfactory and gustatory sensations) of almond, tomato, and grass was also found. The VOOs showed several sensory defects, although "fusty-muddy" (off-flavor of oils from olives stored in large amounts for many days before processing, or of oils left in contact with the sediment for a long period of time, both leading to anaerobic fermentation) was the most common. Other sensory defects found in VOO samples were rancid and winey-vinegary.

The volatile compounds identified and quantified in the headspace of the analyzed samples by SPME/GC–MS are reported in Fig. 1, which shows an overlap between SPME/GC–MS traces relative to the profiles in volatiles molecules for M15 (mixture, non-100% Italian), I13 (100% Italian), and I23 (Italian PDO) samples. It is interesting to note that the non-100% Italian sample (M15) showed a high content of C_6 lipoxygenase (LOX) esters (hexylacetate and (*Z*)-3-hexenylacetate), which contribute to the positive sensory notes of "sweet", "fruity", and "banana-like" (Kalua et al., 2007) and, on the other hand, a tendency towards a lower content in (*E*)-2-hexenal and (*E*)-2-hexenol, both positively correlated with green sensory attributes such as "freshly cut grass", "bitter almond", and "leaves" (Angerosa, 2002; Morales, Luna, & Aparicio, 2005). Moreover, a larger peak of a compound tentatively identified as dodecene could be observed (see also Fig. 2). Generally, samples I13 and I23, respectively, 100% Italian and Italian PDO, were characterized by a major richness in compounds derived from the secondary pathway of LOX (i.e., C_5 molecules and pentene dimers).

Volatile data obtained from SPME/GC–MS were elaborated by PCA to compare the profile of volatile compounds (Fig. 2). A selection of the most discriminant volatile compounds obtained by ANOVA was performed to improve separation among samples. The first two components explained 81% of total variance (48% for the first latent variable and 33% for the second). Considering the locations of products on the PCA scores plot, it is possible to point out that the non-100% Italian samples (M) were grouped in a cluster located in the quadrant of negative values of PC1 and positive values of PC2, whereas Italian samples (100% Italian and Italian PDO/PGI, I) were concentrated mainly between the two quadrants corresponding to negative values of PC2. The different direction/location of vectors (PCA loadings) shows which molecules were involved in the aroma variations among samples, according to the previous explanation. This statistical elaboration allowed to discriminate the samples according to their different geographic origin (non-100% Italian vs. Italian), but not in terms of sensory quality: in fact, each cluster contains both VOOs and EVOOs. The application of FGC E-nose on the set of samples allowed hypothetical identification of 25 different compounds based on Kovats retention indices and the AroChembase software equipped with a library built on the scientific literature to display the associated sensory features.

3.3. PLS-DA from FGC E-nose full chromatograms of Set A

Fig. 2 clearly demonstrated that the discriminating power of the volatile profile with respect to geographic origin can be identified: this preliminary result encouraged further chemometric exploration. Once the discrimination potential of PCA based on Set A, the same set was used to explore the potentials of the other key chemometric tool chosen, namely PLS-DA. In order to make this check independent of the analytical procedure (modality of introduction of volatiles in the GC column) and of the nature of chemometric variables (peak areas or full chromatograms), thus reinforcing eventual confirmation of the intrinsic discriminating power of the volatile profile, the PLS-DA was applied to full chromatograms obtained by FGC E-nose analysis of Set A. To reduce the calculation complexity, one retention time every 10 was selected: hence, the number of variables was reduced from 20,000 to 2000. For the sake of succinctness, the PLS-DA model is not reported herein, but its good performance may be summarized as follows: (i) the scores-plot is analogous to the one shown in Fig. 2 (I samples on negative PC2 values and M samples with negative PC1 and positive PC2); (ii) high total EV (96.9% in the first 2 PCs) was obtained; the plot of predicted vs. experimental responses showed low RMSE (0.071) and $RMSE_{CV}$ (0.15) with high correlation ($R^2 = 0.980$; $R^2_{CV} = 0.908$).

Following the demonstration that the volatiles profile is intrinsically related to geographic origin (independently of whether the volatiles are identified in the GC column by E-nose or SPME, and independently of choosing variable peak areas or full chromatograms), in depth analysis of the large training set (Set B) was initiated.

3.4. PCA models based on FGC E-nose peak areas of Set B

Considering Set B, the training set to create chemometric models and the unknown set to apply models must be extracted from all 251 EVOO samples that were analyzed in quadruplicate by FGC E-nose. Each replicate corresponds to a row of the data set (object), and thus 251 samples gave 1004 objects. In this first step of multivariate analysis of Set B, the variables are the peak areas. Choosing the training set is a delicate step, because the fidelity of the characteristics declared about the samples is crucial to the model's performance. In order to obtain a very reliable and consistent training set, the following rationale was used. A PCA model was created from the 100% Italian samples, and Hotelling analysis was performed. Only objects far inside the Hotelling ellipse were chosen; 224 objects were thus selected. The same was done with

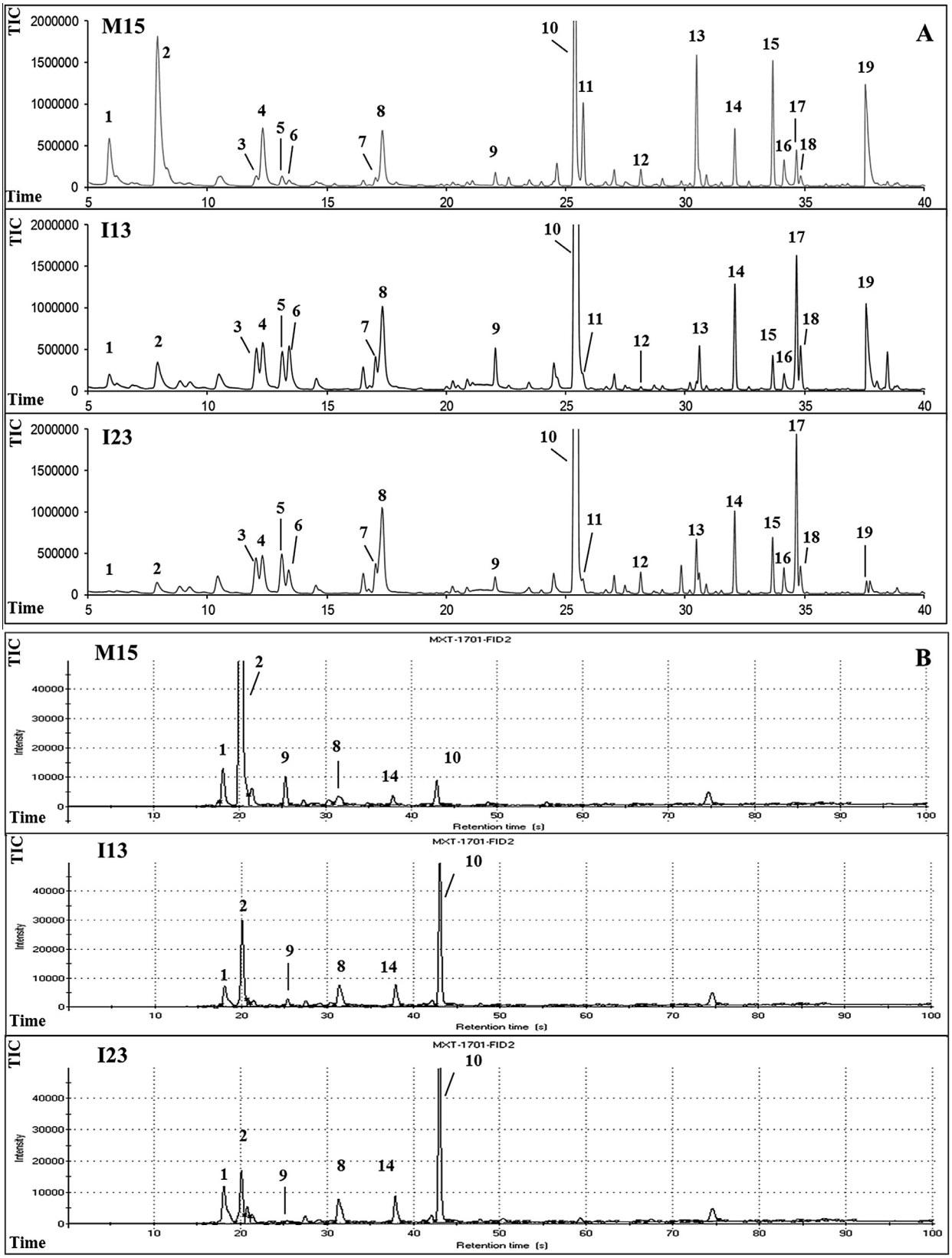

Fig. 1. (A) Overlapping of volatile GC traces obtained by SPME/GC–MS analysis (Set A). Samples: M15 (non-100% Italian), I13 and I23 (100% Italian). Peaks are reported in order of elution: 1: ethyl acetate; 2: ethanol; 3: 3 ethyl-1,5-octadiene (I); 4: IS; 5: 3 ethyl-1,5-octadiene (II); 6: 1-penten-3-one; 7: 4,8-dimethyl-1,7-nonadiene; 8: hexanal; 9: 1-penten-3-ol; 10: (*E*)-2-hexenal; 11: 1-dodecene; 12: hexylacetate; 13: (*Z*)-3-hexenylacetate; 14: hexanol; 15: (*Z*)-3-hexenol; 16: nonanal; 17: (*E*)-2-hexenol; 18: (*E,E*)-2,4-hexadienal; 19: acetic acid. (B) Overlapping of sensors (volatiles) as detected by FGC E-Nose (Set A). Samples and peak numbers according to the (A).

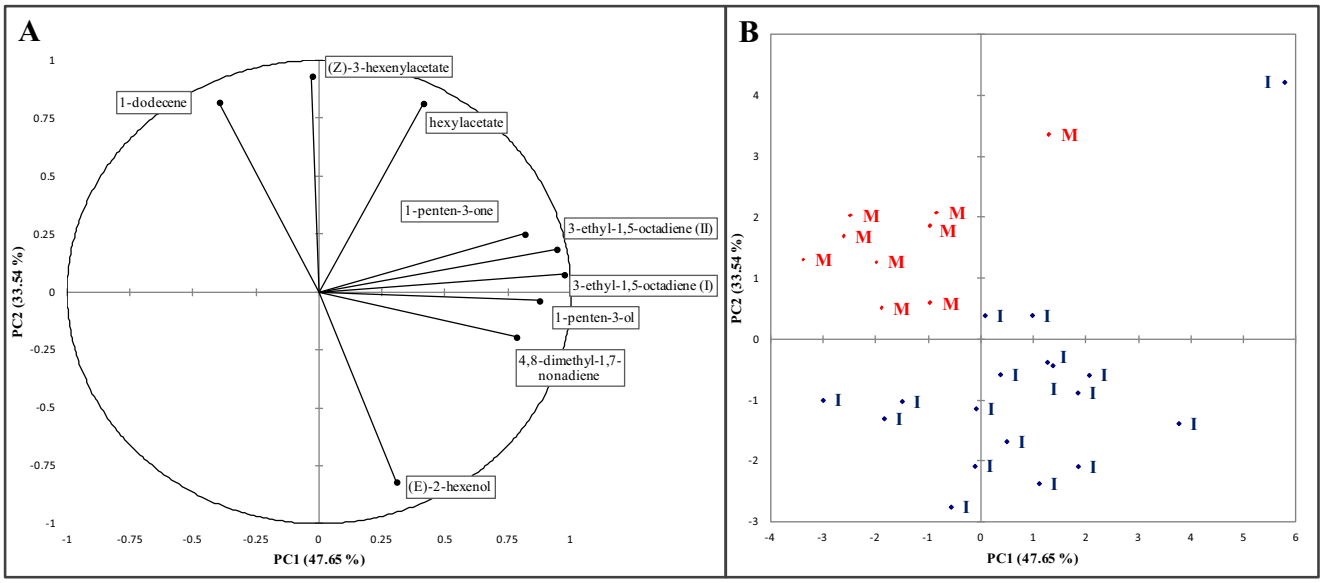

Fig. 2. (A) PCA loadings obtained using the selected variables on SPME/GC–MS data (Set A). (B) PCA score plot obtained using the selected variables on SPME/GC–MS data (Set A).

the M samples, and 269 objects were selected. Therefore, the training set was formed of 493 objects. To verify the suitableness of samples, a LDA scores plot (not reported) was created, and separation between classes was excellent (94.2% correct assignments in cross validation). This is not an obvious result: based on FGC E-nose areas, all Italian samples formed a homogeneous PCA cluster, and all M samples constituted another homogeneous PCA cluster, but LDA showed that these two clusters are separated, thus demonstrating the discriminating ability of FGC E-nose variables and hence of the volatiles profile. This preliminary exploration allowed identification of the variables that were related to high discriminating power. In order to explore eventual subgroups in M category (very wide in this case), a HCA was performed. In fact, 3 clusters were observed in the M category, termed M1, M2, and M3 (dendrogram not shown).

The PCA analysis of these 493 selected objects obtained the results reported in Fig. 3. The resulting PCA model showed good performance since 81.3% EV was obtained in calibration mode with only 6 PCs of 20 original variables. It can be seen that the centroid of the I-cluster is far distant from the centroid of the M-cluster. This is another important proof of the suitability of the volatiles profile (here represented with FGC E-nose variables) to discriminate the geographic origin with respect to 100% Italian and non-100% Italian EVOOs. However, several M2

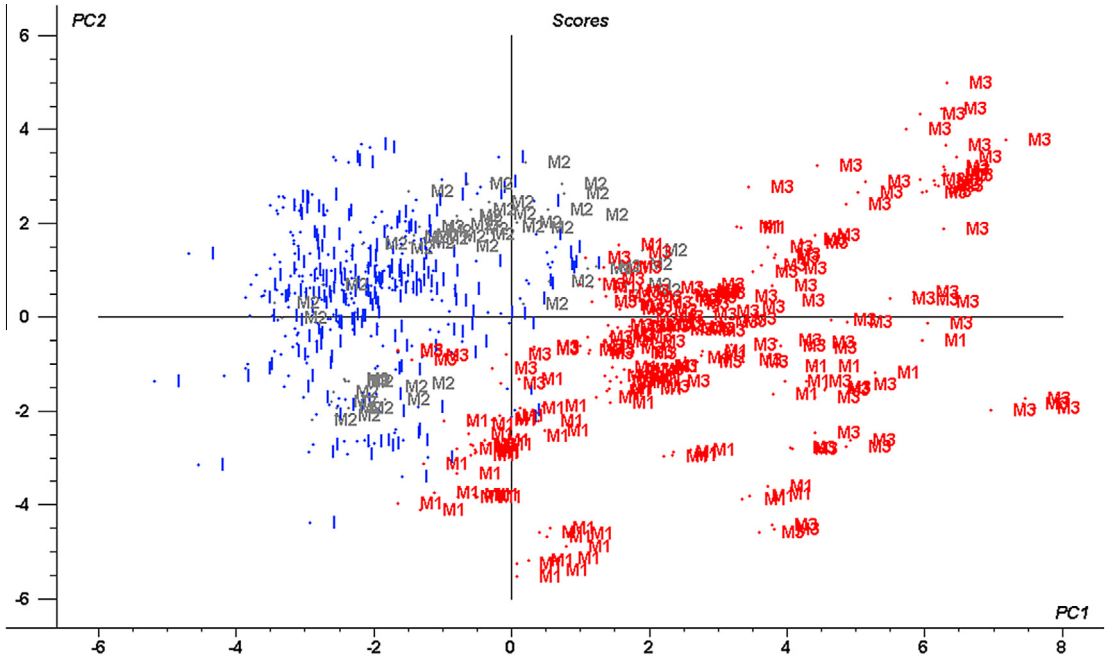

Fig. 3. Scores plot of FGC E-nose peak areas of 493-objects dataset selected by Hotelling (Set B). M1, M2, M3: clusters identified by HCA. EV = 39% along PC1, 18% along PC2. 95% EV is obtained with 11 PCs in calibration and 15 PCs in cross validation mode.

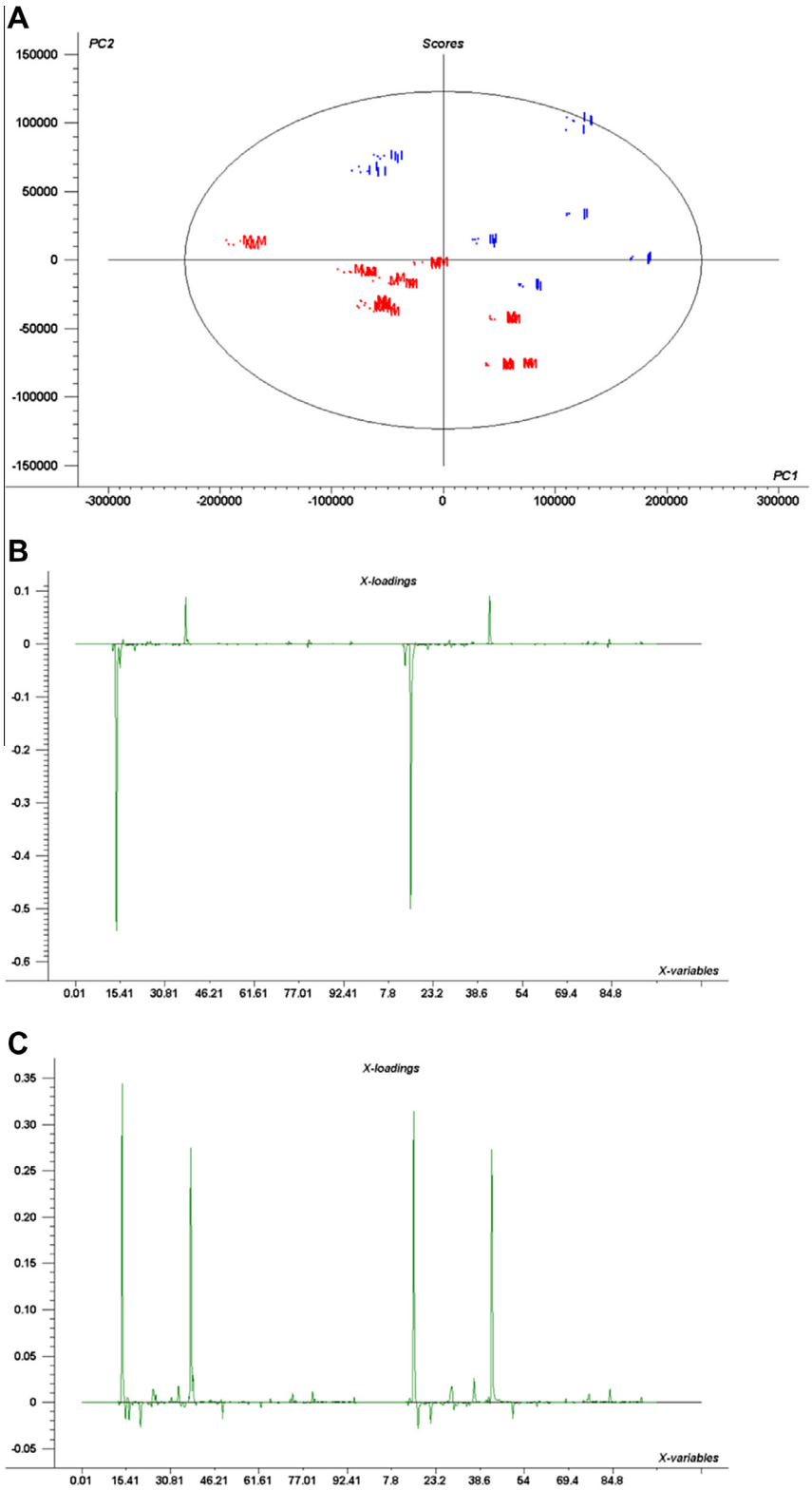

Fig. 4. PLS-DA from FGC E-nose full chromatograms of 439-objects sub-dataset (Set B). (A) Scores plot, PC1-EV: 87%, PC2-EV: 7%. (B) PC1-loadings plot. (C) PC2-loadings plot.

samples in the scores plot in Fig. 3 are near the I centroid; this is not surprising, since a sample classified as "non-100% Italian" may contain a fraction of Italian EVOO. The samples with the highest distance from the centroid were from four suppliers who declared that they were from EU countries, but did not contain Italian oil. In order to avoid doubts related to the geographical origin of samples in the training set, in the subsequent discussion a sub-dataset was created where the M2 samples were discarded (439 objects remained), and M1 and M3 were joined again in a unique M class.

3.5. PLS-DA from FGC E-nose full chromatograms of Set B

To check the opportunity of using full chromatograms as prediction variables, PLS-DA was performed on the 439 object subdataset of Set B. This procedure is almost identical to the one used in Section 3.2, although in this case there is a much higher number of objects. The outputs relevant to the PLS-DA model are reported in Fig. 4A–C. A well-defined separation between Italian and non-100% Italian classes is obtained. Comparison between the scores plot in Fig. 4A and loadings relevant to PC1 (Fig. 4B) allows determination of which FGC E-nose retention times discriminate objects with positive PC1 scores with respect to objects with negative PC1 scores; the analogous comparison allows to study the FGC E-nose discriminating retention times along PC2. The figures of merit related to the PLS-DA response plot (calculated vs. experimental degree of belonging to Italian class) were as follows: a very low RMSE was obtained for both descriptive and predictive ability (0.203 and 0.207, respectively); very few PCs contained over 99% of variance: for each chromatogram, 2000 signals at several retention times were acquired and only 2 PCs contained an high level of information (PC1-EV: 87%, PC2-EV: 7%; total EV = 94%). Both in calibration and in validation, the slopes of response plot were very high (0.834 and 0.833, respectively) and the offsets were close to the ideal null value. Determination coefficients were also high (0.835 and 0.839). This is a very strong result, because models created in Sections 3.3 and 3.5 were obtained by two different laboratories working in a completely independent manner, and using two different sample sets from different harvest periods analyzed with different instruments and experimental conditions.

The good PLS-DA model obtained was applied to M2 samples that were used as unknowns to be predicted. In all cases, a relative standard deviation (RSD) of about 20% degree of belonging (y_I or y_M) was obtained. Predicted values for y_I or y_M that were higher than 70% were considered to correspond to "full" *I* or *M* character, respectively; values resulting lower than 30% were assumed to indicate non-belonging. The result of prediction was the following: 6 ME2 samples of 51 (11.8%) were predicted as "non-100% Italian"; 19 samples (37.3%) were predicted as "100% Italian"; the remaining 26 ME2 samples were predicted as partially "100% Italian" and partially "non-100% Italian".

3.6. PCA models based on SPME/GC–MS peak areas of Set B

In order to compare FCG E-nose results with a well known technique such as SPME/GC–MS, a new dataset was created on the basis of the PCA shown in Section 3.4, according to the following criteria. Samples for which all the replicates gave points that were very close to the I-centroid were selected as "surely Italian samples". Samples for which all the replicates give points very close to the M-centroid were selected as "surely non-100% Italian" samples. In this way, 7 I samples and 9 M samples were extracted, and the I–M sub-dataset was obtained. The scores plot obtained from I–M dataset is reported in Fig. 5, where the Hotelling ellipse is seen.

The I–M dataset extracted from Set B was processed by SPME/GC–MS, and careful and detailed analysis of mass spectra was performed to identify molecules corresponding to significant chromatographic peaks. It must be pointed out that neither the gas chromatographic conditions nor the headspace conditions respectively employed for SPME/GC–MS and FGC E-nose were identical. Moreover, correlation analysis between SPME/GC–MS and FGC E-nose chromatograms may show eventual correspondences between species identified in SPME/GC–MS and FGC E-nose retention times. This could help in bypassing the non-target character of FGC E-nose analyses.

Since SPME/GC–MS analyses were performed in two replicates (Set B), the 7+9 samples corresponded to 14+18 objects. The species identified by SPME/GC–MS analysis were the following: 1-hexanol; 1-octanol; 1-octen-3-ol; 1-penten-3-ol; 1-penten-3-one; 2,4-decadienal; 2,4-hexadienal; 2-butenal; 2-heptanone; 2-methylbutanal; 2-methylbutanol; 2-octanol; 3,4-diethyl 1,5-hexadiene; 3,4-diethyl meso-1,5-hexadiene; 3,5-octadien-2-one; 3,7-decadiene; 3-ethyl 1,5-octadiene; 3-methylbutanal; 3-methylbutanol; 3-pentanone; acetic acid; acetone; α-copaene; α-murolene; α-pinene; benzeneethanol (2-Phenylethanol); benzenemethanol; citronellal; decanal; decane; dimethylnonadienal; (*E*,*E*)-α-farnesene; (*E*)-2-heptenal; (*E*)-2-hexenal; (*E*)-2-hexenol; (*E*)-2-pentenal; (*E*)-2-pentenol; (*E*)-β-ocimene; ethanol; ethyl acetate; formic acid; heptanal; hexanal; hexane; hexylacetate; isoamylacetate; isoamylalcohol; limonene; methanol; methylacetate; methyloctane; murolene; nonanal; octanal; octane; pentanal; propanal; (*Z*)-2-pentenol; (*Z*)-3-hexenal; (*Z*)-3-hexenol;

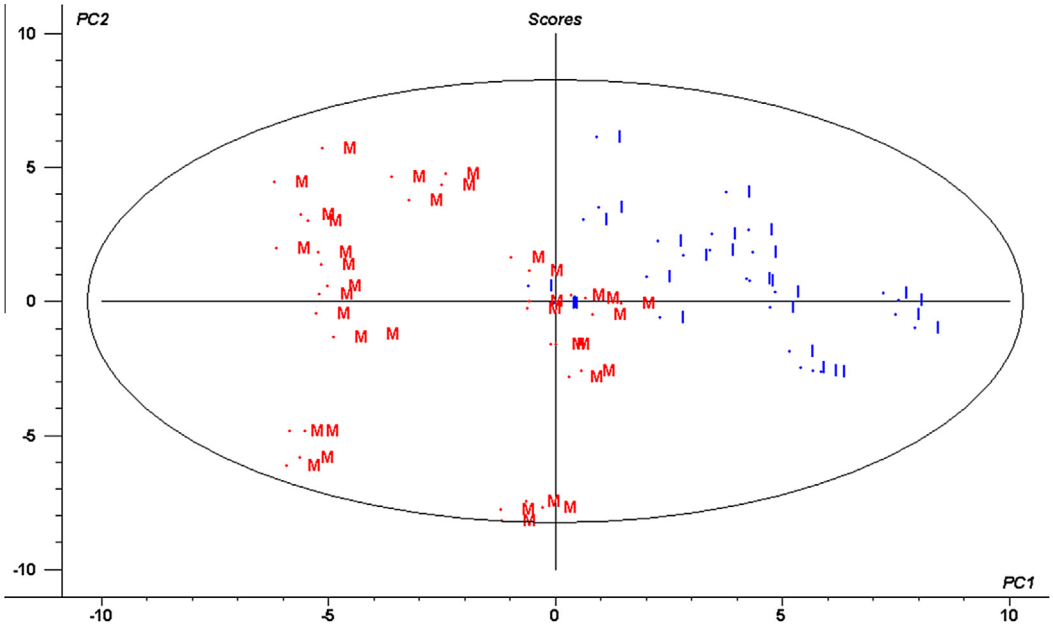

Fig. 5. Scores plot of FGC E-nose peak areas of I–M dataset (Set B). PC1-EV: 25%, PC2-EV: 16%.

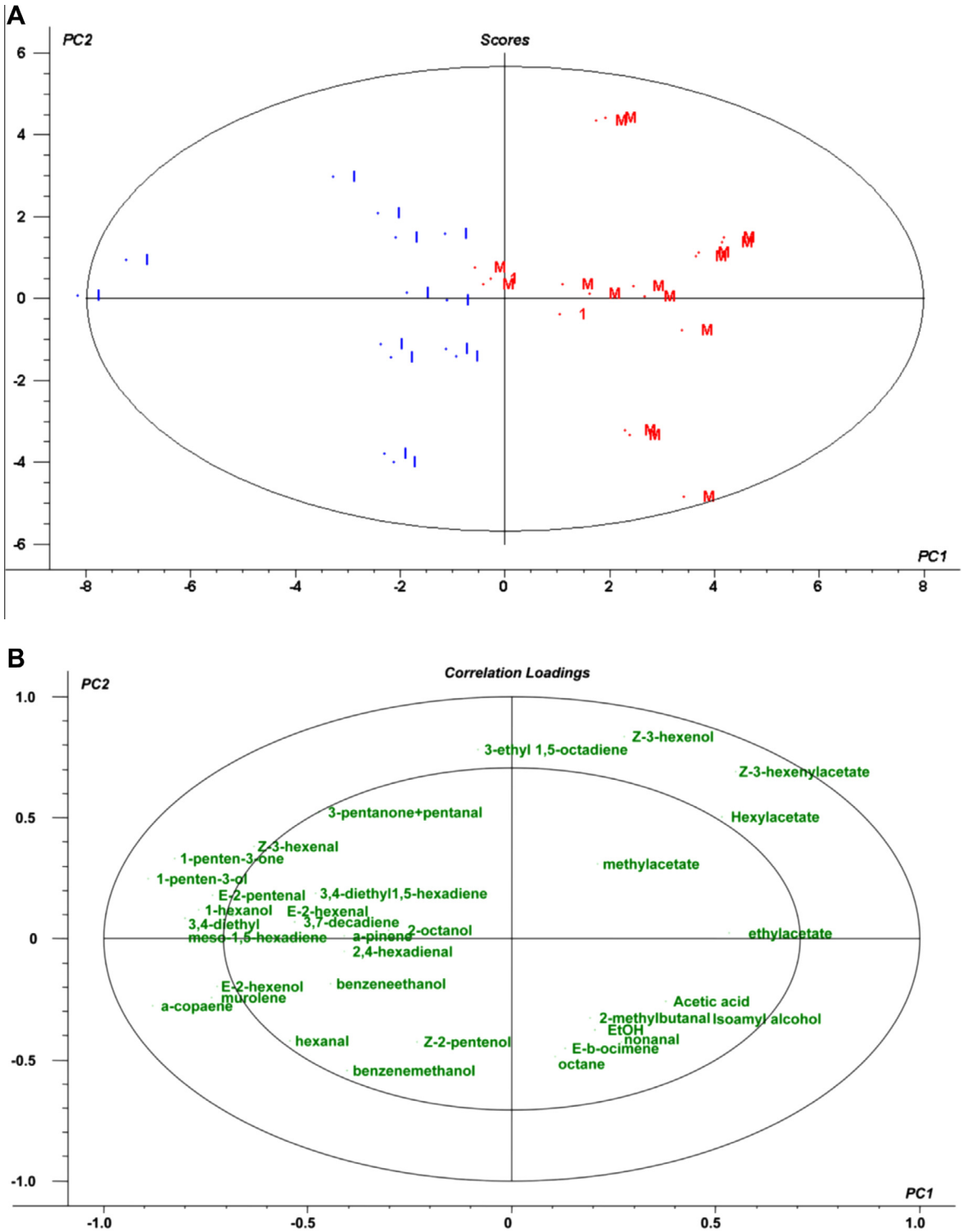

Fig. 6. (A) Scores plot of SPME/GC–MS peak areas of I–M dataset (Set B). PC1-EV: 28%, PC2-EV: 14%. (B) Correlation loadings plot of SPME/GC–MS peak areas of I–M dataset (Set B).

(*Z*)-3-hexenylacetate; (*Z*)-4,8-dimethylnonatriene. Each of these species was a "variable" in a dataset created by putting the I–M samples on lines and the SPME/GC–MS peak-area values (in total ion current, TIC) in the corresponding columns. There were 62 species detected, although some were detected by both the more-polar and by less-polar columns. Hence, there were 89 variables in the SPME/GC–MS dataset, which was more than the number of species detected. The PCA model obtained by the I–M SPME/GC–MS dataset is reported in Fig. 6A, where the Hotelling ellipse is seen. An excellent separation was observed between I and M clusters, thus confirming that headspace GC may discriminate the Italian quality of EVOOs. The corresponding correlation loadings plot (Fig. 6B) showed which species are especially important in discriminating samples: the molecules in the zone between the internal and the external ellipses are the most

important variables; molecules with absolute values of loadings higher than 0.3 may be considered significantly relevant. It is interesting to observe that molecules relevant to a separation along PC1, namely with respect to the separation between I and M, are due to primary or secondary metabolic compounds of the LOX pathway and terpenes. This has a chemical–biological basis, since molecules derived from these enzymatic activities are known to be influenced by the cultivar and geographic origin. Comparison between Figs. 5 and 6A shows that the FGC E-nose peak areas and SPME/GC–MS peak areas yield a very similar PCA model: this confirms that headspace GC data (independently of how volatiles are brought into the GC column, i.e., FGC E-nose or SPME/GC–MS) are suitable for discriminating between 100% Italian and non-100% Italian samples, and that FGC E-nose performance in this discrimination is not significantly different with respect to SPME/GC–MS. It must be pointed out that the extraction of the training set samples from the initial samples was performed based on data pre-processing on objects obtained by FGC E-nose; the fact that these objects gave good results even with SPME/GC–MS data demonstrates that the initial choice was not a tautology: MS data are completely independent from FID data. The comparison between the scores plot and the correlation loadings plot, respectively reported in Fig. 6A and B, shows that I samples are characterized by negative PC1 scores and M samples are characterized by positive PC1 scores; this suggests that molecules identified by MS spectra and characterized by negative PC1 loadings and positive PC1 loadings may be related to I and M samples, respectively.

3.7. Correlation between FGC E-nose and SPME/GC–MS data of Set B

In order to study the correlation between FID variables and MS variables, a dataset in which lines corresponded to the I–M samples discussed in Sections 3.4 and 3.6 was created; all columns relevant to FGC E-nose peak areas and SPME/GC–MS peak areas relevant to the more polar column are reported. The correlation matrix for the FID-peaks and MS-peaks was calculated, and correlation coefficients with significant or considerable values for highly discriminating FGC E-nose peaks (see Section 3.4) were observed. For instance, correlation coefficients higher than 0.8 were observed for ethanol, methylacetate, ethylacetate, 1-penten-3-one, 1-penten-3-ol, (*E*)-1-hexenal, 1-hexanol, and (*E*)-2-hexenol. This analysis shows that accurate study may lead to identification of FGC E-nose peaks, thus bypassing the shortcomings of this technique: it is a non-target analysis; when a significant signal is not linkable to a chemical characteristics, the chemometric results are less strong. It must be underlined that high correlation between retention time and a molecule does not imply that the molecule is an important variable; the present correlation analysis simply has an identification purpose. Importance of variables is determined by loadings: the important molecules are those lying in the outside elliptical ring shown in correlation loadings plot (Fig. 6B). Complete identification of FGC E-nose signal is beyond the scope of the present work, which aims to demonstrate that FGC E-nose based chemometric models are not less reliable that those obtained with SPME/GC–MS data.

4. Conclusions

This study demonstrates that FGC E-nose is suitable for checking geographical traceability of EVOO, even using non-target chromatographic signals of the volatile fraction as variables for multivariate analysis. As a consequence, the feasibility of comparing the geographic origin of standard EVOOs to the origin of an unknown EVOO using FGC E-nose chromatograms as a fingerprint has been assessed. A PLS-DA model, able to discriminate between oils labeled as "100% Italian" (I) and oils labeled as EU oils mixture, considered as "non-100% Italian" (M), was created. This means that when a good, reliable training set coming from a certain production year is available, it is possible to verify, through direct and rapid analysis, whether unknown samples belong to the same statistical population as the training set. Moreover, it is possible to quantify the degree of belonging of unknown samples to the category "100% Italian". The performance of geographic discrimination of FGC E-nose was comparable with SPME/GC–MS, and the results obtained by the two techniques on the same dataset were not significantly different. Comparison between FGC E-nose and SPME/GC–MS signals allowed for eventual correlations between some FGC E-nose retention times and particular molecules identified by their MS spectra in SPME/GC–MS analysis.

Both approaches utilized to analyze volatile compounds were able to discriminate samples with different geographical origin (M vs. I), but each offers specific advantages and limitations: SPME/GC–MS provided more reliable diagnostic information on the identity of compounds thanks to the study of the specific ion fragment profile and the possibility to consult the library of mass spectra, but a lengthy time for analysis and for data processing is required. FGC E-nose was a very fast analytical tool (only 100 s of acquisition time and virtually no need for solvents), discriminating samples with a higher explained variance and allowed for comprehensive data processing with automatic identification of molecules. These results highlight the potential of FGC E-nose for rapid control of the compliance of information on geographic origin declared in the label. This analytical approach seems particularly interesting for food providers, commercial suppliers, and retailers that intend to avoid media scandals of this sector thanks to a more efficient protection and promotion of the integrity of the olive oil image. The main effort concerns the possibility to build, season by season (even by each distributor) an internal or shared and representative data base to be used to screen and control, year after year, EVOOs labeled with a specific origin.

Acknowledgements

The authors wish to thank Fernando Gottardi and Martino Barbanera of COOP Italia for providing the samples, guiding students in carrying out part of the experimental work, and in sharing their expert knowledge about quality control of 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.

Aparicio, R., Morales, M. T., & Alonso, M. V. (1996). Relationship between volatile compounds and sensory attributes of olive oils by the sensory wheel. *Journal of the American Oil Chemists' Society, 73*(10), 1253–1264.

Azizian, H., Mossoba, M. M., Fardin-, A. R., Delmonte, P., Karunathilaka, S. R., & Kramer, J. K. G. (2015). Novel, rapid identification, and quantification of adulterants in extra virgin olive oil using near-infrared spectroscopy and chemometrics. *Lipids, 50*, 705–718.

Ben Temime, S., Campeol, E., Cioni, P. L., Daoud, D., & Zarrouk, M. (2006). Volatile compounds from Chétoui olive oil and variations induced by growing area. *Food Chemistry, 99*, 315–325.

Brereton, R. G. (2007). *Applied chemometrics for scientists*.Chichester, England: John Wiley & Sons.

Cecchi, T., & Alfei, B. (2013). Volatile profiles of Italian monovarietal extra virgin olive oils via HS-SPME-GC–MS: Newly identified compounds, flavors molecular markers, and terpenic profile. *Food Chemistry, 141*(3), 2025–2035.

Cerretani, L., Salvador, M. D., Bendini, A., & Fregapane, G. (2008). Relationship between sensory evaluation performed by Italian and Spanish official panels and volatile and phenolic profiles of virgin olive oils. *Chemosensory Perception, 1*(4), 258–267.

Cimato, A., Dello Monaco, D., Distante, C., Epifani, M., Siciliano, P., Taurino, A. M., ... Sani, G. (2006). Analysis of single-cultivar extra virgin olive oils by means of an Electronic Nose and HS-SPME/GC/MS methods. *Sensors and Actuators B Chemical, 114*(2), 674–680.

Commission Implementing Regulation (EU) No. 1335/2013 of 13 December 2013, amending Implementing Regulation (EU) No. 29/2012 on marketing standards for olive oil, Official Journal of the European Union, L 335/14.

Diraman, H., & Dibeklioğlu, H. (2009). Characterization of Turkish virgin olive oils produced from early harvest olives. *JAOCS Journal of the American Oil Chemists' Society, 86*(7), 663–674.

European Commission Implementing Regulation (2012). EU No 29/2012 13 January 2012. Marketing standards for olive oil.

European Community (2008). Commission Regulation No. 640/2008. Amending regulation No. 2568/91/EEC. *Official Journal of the European Community, L178*, 11–16.

Food and Agriculture Organization of the United Nations World Health Organization (2003). Traceability implementation in developing countries, its possibilities and its constraints. A few case studies. 1–159.

Gallina Toschi, T., Bendini, A., Lozano-Sanchez, J., Segura-Carretero, A., & Conte, L. (2013). Misdescription of edible oils: Flowcharts of analytical choices in a forensic view. *European Journal of Lipid Science and Technology, 115*, 1205–1223.

García-González, D. L., Luna, G., Morales, M. T., & Aparicio, R. (2009). Stepwise geographical traceability of virgin olive oils by chemical profiles using artificial neural network models. *European Journal of Lipid Science and Technology, 111*, 1003–1013.

Gouvinhas, I., De Almeida, J. M. M. M., Carvalho, T., Machado, N., & Barros, A. I. R. N. A. (2015). Discrimination and characterisation of extra virgin olive oils from three cultivars in different maturation stages using Fourier transform infrared spectroscopy in tandem with chemometrics. *Food Chemistry, 174*, 226–232.

International Olive Oil Council, Selection of the characteristic descriptors of the designation of origin (2005). IOOC/T.20/Doc. No. 22.

Jabeur, H., Zribi, A., Makni, J., Rebai, A., Abdelhedi, R., & Bouaziz, M. (2014). Detection of chemlali extra-virgin olive oil adulteration mixed with soybean oil, corn oil, and sunflower oil by using GC and HPLC. *Journal of Agricultural and Food Chemistry, 62*(21), 4893–4904.

Kalua, C. M., Allen, M. S., Bedgood, D. R., Jr, Bishop, A. G., Prenzler, P. D., & Robards, K. (2007). Olive oil volatile compounds, flavour development and quality: A critical review. *Food Chemistry, 100*, 273–286.

Lerma-García, M. J., Cerretani, L., Cevoli, C., Simó-Alfonso, E. F., Bendini, A., & Gallina Toschi, T. (2010). Use of electronic nose to determine defect percentage in oils. Comparison with sensory panel results. *Sensor and Actuators B Chemical, 147*, 283–289.

Melucci, D., Fedi, S., Locatelli, M., Locatelli, C., Montalbani, S., & Cappelletti, M. (2013). Application of pyrolysis-gas chromatography-mass spectrometry and multivariate analysis to study bacteria and fungi in biofilms used for bioremediation. *Current Drug Targets, 14*, 1023–1033.

Mendes, T. O., da Rocha, R. A., Porto, B. L. S., de Oliveira, M. A. L., Dos Anjos, V. D. C., & Bell, M. J. V. (2015). Quantification of extra-virgin olive oil adulteration with soybean oil: A comparative study of NIR, MIR, and Raman Spectroscopy associated with chemometric approaches. *Food Analytical Methods, 8*, 2339–2346.

Mildner-Szkudlarz, S., & Jelen, H. H. (2008). The potential of different techniques for volatile compounds analysis coupled with PCA for the detection of the adulteration of olive oil with hazelnut oil. *Food Chemistry, 110*, 751–761.

Morales, M. T., Luna, G., & Aparicio, R. (2005). Comparative study of virgin olive oil sensory defects. *Food Chemistry, 91*, 293–301.

Ollivier, D., Artaud, J., Pinatel, C., Durbec, J. P., & Guerere, M. (2006). Differentiation of French virgin olive oil RDOs by sensory characteristics, fatty acid and triacylglycerol compositions and chemometrics. *Food Chemistry, 97*, 382–393.

Peng, Q., Tian, R., Chen, F., Li, B., & Gao, H. (2015). Discrimination of producing area of Chinese Tongshan kaoliang spirit using electronic nose sensing characteristics combined with the chemometrics methods. *Food Chemistry, 178*, 301–305.

Pouliarekou, E., Badeka, A., Tasioula-Margari, M., Kontakos, S., Longobardi, F., & Kontominas, M. G. (2011). Characterization and classification of Western Greek olive oils according to cultivar and geographical origin based on volatile compounds. *Journal of Chromatography A, 1218*(42), 7534–7542.

Procida, G., Giomo, A., Cichelli, A., & Conte, L. S. (2005). Study of volatile compounds of defective virgin olive oils and sensory evaluation: A chemometric approach. *Journal of the Science of Food and Agriculture, 85*(13), 2175–2183.

Sinelli, N., Casale, M., Di Egidio, V., Oliveri, P., Bassi, D., Tura, D., & Casiraghi, E. (2010). Varietal discrimination of extra virgin olive oils by near and mid infrared spectroscopy. *Food Research International, 43*(8), 2126–2131.

Taamalli, A., Arráez Román, D., Zarrouk, M., Segura-Carretero, A., & Fernández-Gutiérrez, A. (2012). Classification of 'Chemlali' accessions according to the geographical area using chemometric methods of phenolic profiles analysed by HPLC-ESI-TOF-MS. *Food Chemistry, 132*(1), 561–566.

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.

Zunin, P., Boggia, R., Salvadeo, P., & Evangelisti, F. (2005). Geographical traceability of West Liguria extra virgin olive oils by the analysis of volatile terpenoid hydrocarbons. *Journal of Chromatography A, 1089*, 243–249.