Available online at www.sciencedirect.com
ScienceDirect
Food Chemistry 108 (2008) 374–383
Food Chemistry
www.elsevier.com/locate/foodchem

Analytical Methods

Geographical origin classification of olive oils by PTR-MS

Nooshin Araghipour [a,b], Jennifer Colineau [a], Alex Koot [a], Wies Akkermans [c], Jose Manuel Moreno Rojas [d], Jonathan Beauchamp [b], Armin Wisthaler [b], Tilmann D. Märk [b], Gerard Downey [e], Claude Guillou [d], Luisa Mannina [f,g], Saskia van Ruth [b,*]

[a] *RIKILT – Institute of Food Safety, Wageningen UR, P.O. Box 230, 6700 AE Wageningen, The Netherlands*
[b] *Institut für Ionenphysik und Angewandte Physik, Leopold-Franzens University, Technikerstr. 25, 6020 Innsbruck, Austria*
[c] *Biometris, Wageningen UR, P.O. Box 16, 6700 AA Wageningen, The Netherlands*
[d] *European Commission – Joint Research Centre, Institute for Health and Consumer Protection, Physical and Chemical Exposure Unit, BEVABS TP-281, Via Fermi, 2, 21020 Ispra (VA), Italy*
[e] *Ashtown Food Research Centre, Teagasc, Ashtown, Dublin 15, Ireland*
[f] *Department of S.T.A.A.M., University of Molise, I 86100 Campobasso, Italy*
[g] *Institute of Chemical Methodologies, CNR, 00016 Monterotondo Stazione, Rome, Italy*

Received 29 June 2007; received in revised form 12 September 2007; accepted 22 October 2007

Abstract

The volatile compositions of 192 olive oil samples from five different European countries were investigated by PTR-MS sample headspace analysis. The mass spectra of all samples showed many masses with high abundances, indicating the complex VOC composition of olive oil. Three different PLS-DA models were fitted to the data to classify samples into 'country', 'region' and 'district' of origin, respectively. Correct classification rates were assessed by cross-validation. The first fitted model produced an 86% success rate in classifying the samples into their country of origin. The second model, which was fitted to the Italian oils only, also demonstrated satisfactory results, with 74% of samples successfully classified into region of origin. The third model, classifying the Italian samples into district of origin, yielded a success rate of only 52%. This lower success rate might be due to either the small class set, or to genuine similarities between olive oil VOC compositions on this tight scale.

Keywords: Chemometrics; Olive oil; Origin classification; PTR-MS; PLS-DA

1. Introduction

In recent years, olive oil has steadily increased in popularity due to its associations with a healthy diet, in particular the so-called "Mediterranean diet". This trend has lead to a rise in demand of olive oil, thereby resulting in a more widespread production of this consumable. The European Union currently dominates the olive oil market, accounting for more than three-quarters of worldwide production (with Italy, Greece and Spain contributing 97% of EU production), but increasingly this is spreading to countries with generally less olive oil consumption, such as Australia and Argentina (Luchetti, 2002).

Olive oil is produced by mechanical means from the fruit of the olive tree (*Olea europaea* L.) and requires no refinement prior to consumption, thus it retains its characteristic aroma (Boskou, 1996). This aroma, and its complimentary flavour, arises from the phenolic content and the large number of volatile constituents of the oil (Morales, Rios, & Aparicio, 1997). The latter comprise a wide variety of compounds, including saturated, unsaturated, aromatic, and terpenic hydrocarbons, aldehydes, alcohols, ketones,

* Corresponding author. Fax: +31 317 417717.
E-mail address: Saskia.vanRuth@wur.nl (S. van Ruth).

0308-8146/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.foodchem.2007.10.056

esters, ethers, furans, acids, and other compounds (e.g. Aparicio & Morales, 1998; Flath, Forrey, & Guadagni, 1973).

There are many different important factors affecting the volatile composition of olive oil, such as fruit genotype, ripening, and the processing equipment used in production (Angerosa, Mostallino, Basti, & Vito, 2001; Aparicio & Morales, 1998, and references therein). The influence of climate and soil type are also crucial for olive oil volatile profiles (Kalua et al., 2007), thus the geographic origin of an oil is an important quality factor and has become more attractive to certify in recent years. In Europe this has lead to the introduction of several official regulations for agricultural products, such as protected denomination of origin (PDO), protected geographical indication (PGI), and traditional speciality guaranteed (TSG) certifications, which allow certain products to be labelled with the names of their geographical area of production (E.C., 2006).

Since the geographical classification of food plays such an important role in its quality, several methods for origin authentication have been tested for reliability. Traditional quality assessment methods of a food product rely on panels of human sensory analysts; such analyses, however, do not always offer reproducible results and are time-consuming and expensive, requiring highly trained and qualified panel testers. Furthermore, the parameters of assessment by these means are generally only suited to quality evaluations and not to the determination of geographic origin. Therefore, increasingly, there has been a push to develop alternative and complimentary analytical procedures to human sensory analysis.

In the past, there have been a number of studies using a wide variety of analytical techniques to determine geographic origin of olive oils in relation to their volatile content, with success varying according to methodologies used and variety or type of compounds measured. These include isotope ratio analyses, which exploit the chemical composition and natural isotopic ratios of certain compounds present in the samples to enable authenticity and origin verification of the sample (e.g. Angerosa et al., 1999). Analytical methods here, amongst others, include isotopic ratio mass spectrometry (IRMS) (Bréas, Sada, Reniero, Guillou, & Angerosa, 1998) and site-specific natural isotope fractionation nuclear magnetic resonance (SNIF-NMR) (Lai, Casu, Saba, Corongiu, & Dessi, 1995). Other techniques have equally been used for the geographical origin determination of food products: gas chromatography–mass spectrometry (GC–MS) (e.g. Zunin, Boggia, Salvadeo, & Evangelisti, 2005), dynamic headspace-gas chromatography (DHS-GC) (e.g. Luna, Morales, & Aparicio, 2006), solid-phase micro extraction (SPME) coupled with GC–MS (Temime, Campeol, Luigi Cioni, Daoud, & Zarrouk, 2006) or with GC-flame ionization detector (GC-FID) (Vichi, Pizzale, Conte, Buxaderas, & López-Tamames, 2003), high pressure liquid chromatography (HPLC) (e.g. Romero, Sánchez-Viñas, Gázquez, & Bagur, 2002) and electronic noses (Cosio, Ballabio, Benedetti, & Gigliotti, 2006). Spectroscopy also plays an important role in such analyses and ranges from ultra-violet (UV) to near-infrared (NIR), mid-infrared (MIR), visible and Raman spectroscopy (Reid, O'Donnell, & Downey, 2006, and references therein).

Additionally, the results from many analytical measurements rely on the strengths of the mathematical methods applied to the data for evaluation. Two popular and successful methods for the determination or distinction between geographic origins of different olive oils, for instance, are classification and influence matrix analysis (CAIMAN) and principal component analysis (PCA) (e.g. Angerosa et al., 2004; Vichi et al., 2003, and references therein).

Despite this wide array of analytical techniques, most either suffer drawbacks of being too time-consuming to be practicable for analyses of many samples, or they lack the sensitivity to provide distinguishable features between different samples. Recently, proton-transfer-reaction mass spectrometry (PTR-MS) has been used to evaluate volatiles in the headspace of extra virgin and rancid olive oils in order to detect oxidative alterations in the samples (Aprea et al., 2006). This fast and sensitive technique for volatile organic compound (VOC) detection enables direct on-line headspace analyses of complex samples to be made without the need for sample preparation, thereby potentially enabling many samples to be analysed within a short period. A multivariate statistical approach on the PTR-MS data in the aforementioned study (Aprea et al., 2006) enabled successful distinction between the two types of oil, demonstrating the capabilities of PTR-MS as being a useful tool with this respect. The aim of the present study is to evaluate PTR-MS for categorisation of olive oils according to their geographic origin using multivariate statistical methods for data evaluation.

2. Materials and methods

2.1. Olive oil samples

One hundred and ninety two Mediterranean PDO protected olive oil samples (obtained courtesy of the EU-funded TRACE project; trace.eu.org) from different olive cultivars in defined geographical areas were collected by scientists of the European Commission Joint Research Centre, Institute for Health and Consumer Protection, Physical and Chemical Exposure Unit in Ispra, Italy. These oils originated from 80 communities (cities/towns) in 45 provinces of different districts in five European countries. (Some districts in each country are divided into provinces and communities, hence the sub-divisions here). The olive oil sample distribution was as follows (and summarised in Table 1): 92 samples were from 24 provinces of nine districts in Italy; 46 were from 11 provinces in six districts of Greece; six were from four provinces of one district in Cyprus; 38 were from four provinces of one district in

Table 1
List olive oil samples, with sub-divisions for the Italian samples into districts and regions of origin

Country	Region	District	Total number of samples
Cyprus	**1**	**4**	**6**
France	**1**	**1**	**10**
Greece	**6**	**11**	**46**
Italy	**9**	**24**	**92**
	North: Veneto, Liguria	7	36
	Centre: Toscana, Umbria, Lazio, Abruzzo	9	30
	South-east: Molise, Puglia	4	18
	Sicilia: Sicilia	4	8
Spain	**1**	**4**	**38**

Spain; and 10 were from one province of one district in France. The oils were produced in the first half of 2006 and analysed in August 2006.

2.2. VOC measurements

The headspace air of the olive oil samples was analysed using PTR-MS, an analytical technique that has been described extensively elsewhere (e.g. Lindinger, Hansel, & Jordan, 1998). The advantages of PTR-MS are manifold: It is capable of on-line VOC monitoring, enabling rapid and accurate quantification of VOCs contained in a complex air matrix, such as in the headspace of food samples. Furthermore, this technique requires no sample preparation, is relatively insensitive to changes in sample humidity, and air may be used as the carrier gas, altogether allowing sample headspace air to be measured directly and immediately.

2.3. Sampling procedure

Following collection of the olive oil samples, they were stored in cold rooms at 4 °C in the absence of light. Four hours prior to analysis samples were removed from the cold storage rooms and placed in the laboratory, which was at a room temperature of approximately 20 °C.

Measurements were carried out, without any sample pre-treatment, with 5 ml aliquot of olive oil filled in a 250 ml glass vial capped with a polytetrafluoroethylene (PTFE, Teflon®) septum. Filtered outdoor air was introduced through the septum into the vial using a gas-tight syringe and exported through a second syringe. The headspace air of a sample was equilibrated for 45 min at 30 °C in a water bath and delivered directly to the inlet of the PTR-MS system. The inlet was heated to 80 °C to prevent loss of volatiles along the sampling inlet line.

The PTR-MS was operated at a standard E/N (ratio of electric field strength across the reaction chamber, E, to buffer gas number density, N, within the chamber) of 138 Td ($1\ \mathrm{Td} = 10^{-17}\ \mathrm{cm^2\ V\ molecule^{-1}}$) and measurements were made in 'mass scan' mode, whereby a complete mass spectrum in the range of 20–150 atomic mass units (amu), at a mass detection rate of $0.2\ \mathrm{s\ mass^{-1}}$, was gathered in under half a minute. The mass spectrometric data were collected in three replicates/olive oil sample with each replicate measured for 5 mass scan cycles, thereby giving an analysis time/replicate of just over 2 min.

2.4. PTR-MS data evaluation and selection

PTR-MS raw count-rate data were converted to volume mixing ratios (VMRs) for statistical analysis using a typical PTR-MS sensitivity value observed for oxygenated VOCs. Data for analysis of each oil sample were selected as follows: of the five measurement cycles (mass spectra/sample replicate), the first two cycles were discarded and the remaining three were used to provide a mean mass spectrum/replicate. Subsequently, a mean mass spectrum/sample was calculated from these replicate data (3/sample). In this manner, a dataset containing mean mass spectra/sample analysed could be compiled. These were used for subsequent multivariate analyses. Next, a mean mass spectrum for each country of origin was calculated using these individual sample mean data, which allowed for comparison of the mass spectra between the different countries.

2.5. Data analysis

In order to further explore which masses were specific for the various countries, a two-way analysis of variance (ANOVA) was carried out for each of the 130 masses measured. By applying a p value of $p = 0.05/130 = 0.00033$ and with subsequent least significant difference (LSD) tests, 45 masses came out as significant. Partial least squares discriminant analysis (PLS-DA) was carried out on the 192 samples (log transformed mean of the three mass spectra of each sample) in order to estimate a classification (into country of origin) model for the olive oil samples. This analysis was carried out using the PLS Toolbox v.4.0 for Matlab™ (Wise et al., 2006). The method of PLS-DA performs a PCA-like reduction on the independent variables (in this case, the log transformed data of the individual mass signal intensities) to obtain a maximum correlation of these with the dependent variables (here, the class membership, i.e. country, or district, etc.). The performance of the fitted model was evaluated using 10-fold cross-validation: 10% of the samples were removed at random from the complete dataset. The remaining data were then used to fit a model, which was subsequently applied to predict class membership for the removed samples. This procedure was repeated 10 times to ensure that a prediction was available for all samples.

The classification method PLS-DA is a classification method closely related to Fisher's or linear discriminant analysis (LDA). Let N be the number of independent samples (in our case, $N = 192$, each sample being the average of the three replicates), C the number of classes (in our case, the five countries) and V the number of independent variables (in this case, the log transformed data of the individual mass intensities, so $V = 128$). If $V > C$, then in LDA

there will be (at most) $C-1$ discriminant functions, meaning that $(C-1)^*V$ coefficients have to be estimated. Hence, $(C-1)^*V$ should not exceed ca. 20% of N. If the number of variables becomes too large, and especially if $(C-1)^*V$ becomes larger than N, an LDA analysis can no longer be carried out. Hence, for most mass spectrometry datasets, LDA is not a useful technique. PLS-DA offers a remedy. In PLS-DA a PCA-like reduction is performed on the space of the independent variables; the components extracted however are not the principal components (i.e. they are not the components having maximal variance) but they are components having *maximal correlation with the dependent variable* (the class membership, i.e. country, or district, etc.). The DA is then performed using these components as independent variables, and not the original variables.

Other classifications techniques are available. For SIMCA (Soft Independent Modelling of Class Analogy) generates a distinct PCA model for each class, and then combines these C models. With SIMCA, the components extracted, for each class, are the principal components, i.e. the components having the *largest variance*. These components are, in general, (very) different from the components having maximal correlation with the dependent variable. Another difference is that with SIMCA a sample might be assigned to either no class at all, or to one single class, or to more than one class, whereas with DA (both with LDA and with PLS-DA) the space of the variables (components) is divided into separate regions, one region for each class. Therefore, with DA a sample is always assigned to one and only one class. As the origin of the samples in the present dataset are known DA is an appropriate choice. Because of the large number of variables, PLS-DA was again more appropriate than LDA.

3. Results and discussion

3.1. Olive oil mass spectra

The statistical analyses in this study use the mass spectral data as 'fingerprints', i.e. the masses and their corresponding signal intensities (VMRs) in each sample mass spectrum act as a pattern for inter-comparison of the samples. All of the olive oil samples produced signals on most masses in the defined measurement range (20–150 amu), indicating the complex VOC composition of olive oil. Mean sample mass spectra for each country are displayed in Fig. 1a–e. The VOCs with higher volatility (lower mass) dominate the spectrum in terms of signal intensity, although lower masses may also result from fragmentation of larger compounds. The dominant signals are listed in Table 2, along with tentative compound assignments.

Significantly high signals on masses 57, 81, and 99 reflect the typical pattern of ions associated with hexanal, which is a compound that results during lipoxygenase (LOX) activity in the olive fruits (Morales et al., 1997; Williams, Morales, Aparicio, & Harwood, 1998). The LOX pathway is widespread in the plant kingdom and is responsible for the formation of many volatiles in olive oil, including C5 and C6 compounds. The level and activity of each enzyme involved in the LOX pathway define the concentrations of the produced volatiles. Alcohol dehydrogenase (ADH) is one of these enzymes that is responsible for the reversible reduction of aliphatic aldehydes to alcohols and is affected by environmental growth conditions (Kalua et al., 2007). Therefore, the level of C5 and C6 aldehydes and alcohols for oil samples from different regions varies (Vichi et al., 2003). The signal intensity for this compound was most prominent in the Italian oil samples, which is in keeping with other studies where the richness of C6 volatile compounds in Italian oils has also been observed (e.g. Kalua et al., 2007). In addition to the above masses, high signals were seen on other masses associated with further LOX product compounds, e.g. hexenal, hexenol and hexenyl acetate, etc. (see Table 2).

Other large signals were observed on mass 33 (and 51), which is usually associated with methanol (and hydrated methanol); however, there were no reports found in the literature of olive oil containing this alcohol, thus it is more likely that the signals on these masses arise from the fragmentation of one or more larger molecules. Additional high signals are seen on masses 43 and 61, which are likely to be acetic acid. Production of acetic acid can result from the presence of *Acetobacter* during storage of the olive fruit, especially when temperatures are relatively high (Angerosa et al., 2004). Many signals associated with aldehydes were also observed, such as mass 45 (acetaldehyde; particularly in the Spanish samples), masses 55 and 73, (butanal; particularly in the French samples), and masses 69 and 87 (pentanal; abundant in both the Greek and Spanish samples), which have previously been detected in heated or oxidized olive oils (Morales et al., 1997; Stashenko, Wong, Martínez, Mateus, & Shibamoto, 1996).

Results of the present study are in keeping with the recently published study of Cavaliere et al. (2007). In this study the selection of markers from the secondary metabolism of lipoxygenase for olive oils of various origins is described. The compounds selected included hexanal, (*E*)-2-hexenal, (*E*)-2-hexenol and (*Z*)-3-hexen-1-yl acetate.

3.2. Classifications

3.2.1. Number of components

Three models were fitted on three different datasets: on the entire 192 samples dataset, a model was fitted to classify the samples into their country of origin; and two further models were fitted on the subset of 92 Italian samples only, to classify these into their region of origin and into their district (smaller region) of origin, respectively.

In this approach, a reduction of dimensions on the individual mass signal intensities is provided in order to capture the maximum correlation with the dependent variable (country, region, or district). The number of dimensions extracted is an important parameter in such a model: mod-

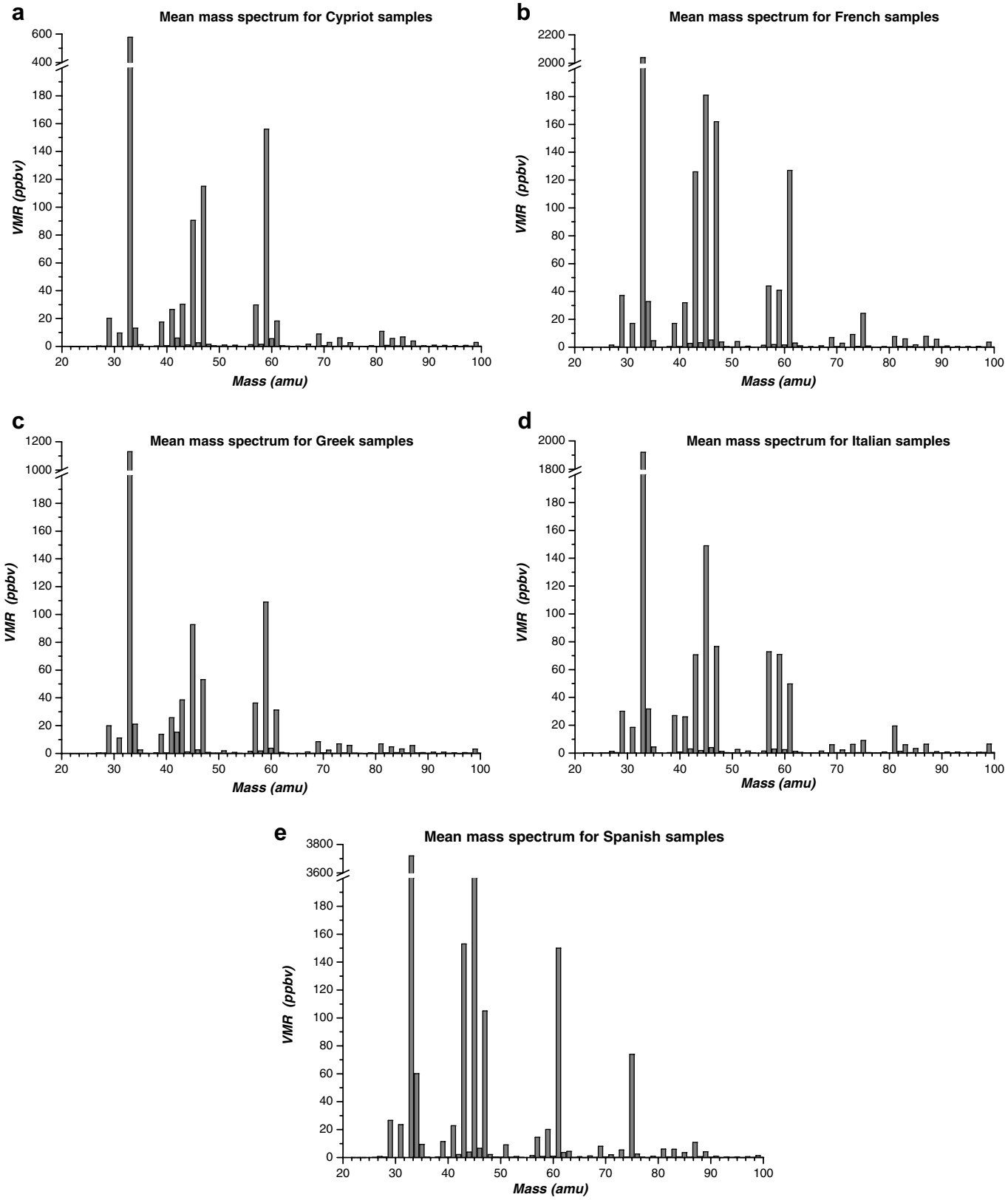

Fig. 1. (a)–(e) Mean sample mass spectra for each country (filtered for signals associated with the PTR-MS ion source: NO^+ at 30; O_2^+ at 32; major water cluster ions at 37 and 55). Note the same scale on the *y*-axis before the break (up to 200 ppbv).

els for one through six dimensions were investigated, with each model being evaluated by the cross-validation procedure described above (Section 2.5). Table 3 details the results of each PLS-DA for country classifications of all

Table 2
The predominant masses for each country with corresponding VMRs for all samples

Protonated mass (amu)	VMR (ppb)[a]					Tentative assignment[b]
	Cyprus	France	Greece	Italy	Spain	
31	10	17	11	19	**24**	
33	578	2036	1135	1918	**3716**	
35	1	5	3	4	**9**	
39	18	17	14	**27**	11	Hexenyl acetate[c] (fragment)
41	27	**32**	26	26	23	Hexanol[d,e] (fragment)
43	30	126	39	71	**153**	Acetic acid[c,f], hexanol[d,e], hexyl acetate[d,e] (fragment)
45	90	181	93	149	**246**	Acetaldehyde[g]
47	115	**162**	53	77	105	Ethanol[c]
51	1	4	2	3	**9**	
57	30	44	36	**73**	14	Hexanal[d,e,f], hexenal[e], hexanol[e] (fragment)
59	**156**	41	109	71	20	Acetone[h], propanal[f], hexenol[c,e] (fragment)
61	18	127	31	50	**150**	Acetic acid[d], hexyl acetate[d,e], methylbutyl acetate[d], ethylacetate[d], butylacetate esters[d] (fragment)
63	0	1	0	1	**4**	Dimethyl sulfide[f], acetaldehyde[g] (hydrate)
69	8	7	**9**	6	8	Pentanal[f] (fragment)
73	6	**9**	7	6	5	Butanal[g], butan-2-one[d]
75	3	24	6	9	**74**	Methylacetate[c,i]
81	11	8	7	**19**	6	Hexanal[i,j] (fragment)
83	6	**6**	5	6	6	Hexenol[d,e], hexenyl acetate[e], hexanal[c,d,e]
85	**7**	2	3	3	4	Hexanol[c,d]
87	4	8	6	6	**11**	Pentanal[f], pentanone[k]
89	0	6	1	1	**4**	Butyrate esters[i], butanoic acid[d]
99	3	4	3	**7**	1	Hexenal[d,e,f]

VMRs highlighted in bold represent the most abundant masses for the respective country, compared to the other countries. A tentative identification has also been provided, based on the literature.

[a] VMRs were calculated according to typical sensitivity value observed for oxygenated VOCs.

[b] Compounds listed are potential candidates only, not absolute identifications, and relate to those compounds reported in the literature to be present in olive oil that would produce a signal on the given mass.

[c] Angerosa et al. (2004).

[d] Kalua et al. (2007).

[e] Morales, Berry, McIntyre, and Aparicio (1998).

[f] Morales et al. (1997).

[g] Stashenko et al. (1996).

[h] Williams et al. (1998).

[i] Reiners and Grosch (1998).

[j] Kanavouras, Kiritsakis, and Hernandez (2005).

[k] Vichi et al. (2003).

Table 3
Number of correct classifications (in %) in the cross-validation for various numbers of extracted dimensions in PLS-DA[a]

	Classification criteria		
	European country	Region in Italy	District in Italy
1 Dimension	33	39	21
2 Dimensions	62	50	39
3 Dimensions	82	73	43
4 Dimensions	82	**74**	**52**
5 Dimensions	**86**	76	49
6 Dimensions	84	79	
Number of classes	5	4	9
Number of samples	192	92	92

[a] Selected number of dimensions for final model in bold.

samples, plus region and district classifications of the Italian samples. Results showed that an increase in the number of dimensions above four did not improve classification. In addition, two random permutations of the class labels were carried out, so that nonsense datasets were generated for comparison with the real model. If the respective prediction errors of the two are comparable, it can be concluded that mainly noise has been modelled. In this case, the first two dimensions, respectively, classified 27% and 19% of the olive oils correctly, which is more or less equal to the percentage expected by chance only, i.e. 20–25% for this five class case.

3.2.2. Classification into country of origin

Cross-validation of the five-dimensional model obtained from PLS-DA for all 192 samples resulted in 86% of the samples being correctly classified into their country of origin (see Table 3). High correct classification rates of the oils into four of the five countries were observed, as listed in Table 4: Cyprus, 100%; Italy, 89%; Greece, 89%, and Spain, 87%. Classification of the French samples (only 40% of samples correctly classified) was poor in this model.

The scores plot for the first three dimensions of the PLS-DA model for all samples is presented in Fig. 2.

Table 4
Classification of olive oils originating from five European countries by PLS-DA with cross-validation by their PTR-MS spectral data: absolute numbers, percentages in brackets, the correctly classified samples in bold[a]

Sample origin	Classification					Total number of samples
	Cyprus	France	Greece	Italy	Spain	
Cyprus	**6 (100%)**	0 (0%)	0 (0%)	0 (0%)	0 (0%)	6
France	0 (0%)	**4 (40%)**	2 (20%)	2 (20%)	2 (20%)	10
Greece	0 (0%)	0 (0%)	**41 (89%)**	4 (9%)	1 (2%)	46
Italy	0 (0%)	2 (2%)	5 (5%)	**82 (89%)**	3 (3%)	92
Spain	0 (0%)	2 (5%)	0 (0%)	3 (8%)	**33 (87%)**	38

[a] Number correct: 166 (86%).

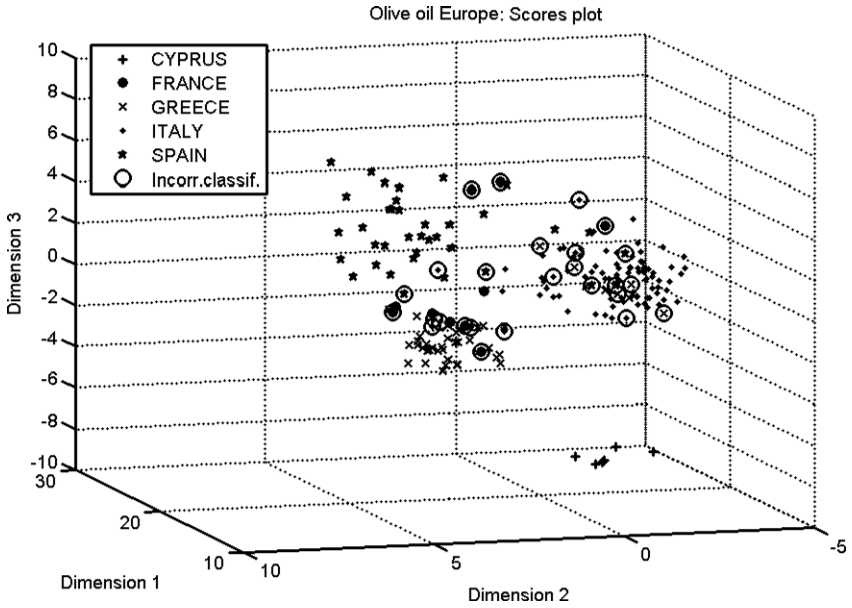

Fig. 2. Scores plot of the first three dimensions of the PLS-DA model of all 192 olive oil samples: classification into country of origin.

The separation of oils originating from Greece, Spain, Italy, and Cyprus can clearly be seen. The French olive oils display a much broader spread and, as such, a much less reliable classification, with only 40% of samples designated correctly (see Table 4). Although the French samples all originated from the same region (Paca region, Bouches-du-Rhone Province), these 10 samples came from seven different towns. In fact, the distribution in the scores plot of Fig. 2 indicates that these oils have volatile characteristics in common with oils from Greece and Spain in particular. Furthermore, the group of French samples was relatively small and an improved classification success rate may be possible using a larger number of samples.

In order to further explore which masses were specific for the various countries, a two-way ANOVA was carried out here for each mass, with subsequent LSD tests when a significant difference was observed ($p < 0.00033$). In total, 45 masses showed significant differences between countries, discriminating them into three groups, and five masses divided the countries into four groups.

3.2.3. Classification of Italian samples into region

Following successful classification of olive oil samples on the larger (European) scale for four of the five countries, a four components PLS-DA model was tested for its capabilities in classifying a more refined dataset. Since a high proportion of all olive oil samples measured were of Italian origin (92 samples in total; constituting 48% of all samples), a PLS-DA was made on a dataset of only the Italian samples. This was initially made to test classification of regions of origin in Italy, which were designated as compass regions: Veneto and Liguria – North; Toscana, Umbria, Lazio, Abruzzo – Centre; Molise and Puglia – South-east; Sicilia – Sicilia. (The district of Abruzzo was allocated as Centre, although it could equally have been designated as South-east.)

The PLS-DA four-component model successfully classified 74% of Italian samples into their regions of origin, as indicated in Table 5: Centre, 93%; Sicilia, 88%; and North, 64%; and South-east, 56%. A scores plot of first three components of Italian olive oil samples into region is shown in Fig. 3.

Table 5
Classification of olive oils originating from four regions in Italy by PLS-DA with cross-validation by their PTR-MS data: absolute numbers, percentages in brackets, and correctly classified samples in bold[a]

Sample origin	Classification				Total number of samples
	Centre	North	South-east	Sicilia	
Centre	**28 (93%)**	0 (0%)	1 (3%)	1 (3%)	30
North	4 (11%)	**23 (64%)**	7 (19%)	2 (6%)	36
South-east	0 (0%)	5 (28%)	**10 (56%)**	3 (17%)	18
Sicily	1 (13%)	0 (0%)	0 (0%)	**7 (88%)**	8

[a] Number correct: 68 (74%).

3.2.4. Classification of Italian samples into district (smaller region)

A further test of the PLS-DA four-component model was made for its classification capabilities into even smaller districts. This was again done using the Italian samples dataset, which contained olive oil originating from nine individual districts in Italy. Here, the model gave poorer classification verification, with only 52% of the samples correctly allocated to their district of origin, as indicated in Table 6: Liguria, 79%; Abruzzo, Toscana, 67%; 63%; Umbria and Sicilia, 50%; Molise, 38%; Puglia, 30%; Lazio 25%; and Veneto, 13%.

The corresponding scores plot for this analysis is presented in Fig. 4. Although the samples of each district are grouped together, these are quite spread out and there is a great deal of overlap in the clustering in which samples from two or more districts sometimes appear close together. It seems evident with these results that sample classification via this model becomes more difficult the

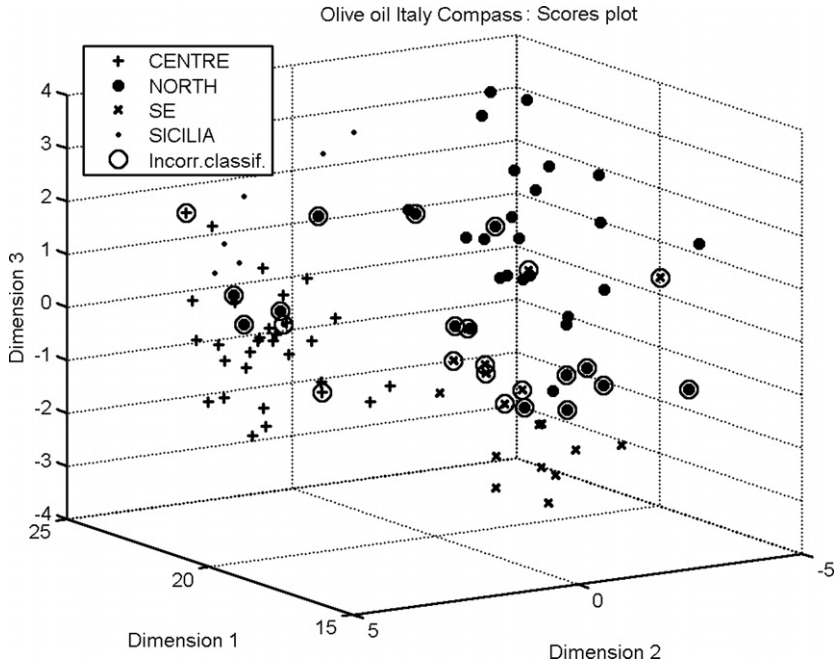

Fig. 3. Scores plot of the first three dimensions of the PLS-DA model for the 92 Italian olive oil samples: classification into region (according to points on a compass).

Table 6
Classification of olive oils originating from nine districts in Italy by PLS-DA with cross-validation by their PTR-MS spectral data: absolute numbers, percentages in brackets, the correctly classified samples in bold[a]

Sample origin	Classification									Total number of samples
	Abruzzo	Lazio	Liguria	Molise	Puglia	Sicilia	Toscana	Umbria	Veneto	
Abruzzo	**5 (63%)**	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (25%)	0 (0%)	1 (13%)	8
Lazio	2 (25%)	**2 (25%)**	0 (0%)	0 (0%)	0 (0%)	0 (0%)	3 (38%)	1 (13%)	0 (0%)	8
Liguria	0 (0%)	0 (0%)	**22 (79%)**	4 (14%)	2 (7%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	28
Molise	0 (0%)	0 (0%)	3 (38%)	**3 (38%)**	2 (25%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	8
Puglia	0 (0%)	0 (0%)	2 (20%)	4 (40%)	**3 (30%)**	0 (0%)	0 (0%)	0 (0%)	1 (10%)	10
Sicilia	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	**4 (50%)**	3 (38%)	0 (0%)	1 (13%)	8
Toscana	1 (17%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (17%)	**4 (67%)**	0 (0%)	0 (0%)	6
Umbria	2 (25%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (25%)	**4 (50%)**	0 (0%)	8
Veneto	2 (25%)	0 (0%)	0 (0%)	0 (0%)	2 (25%)	1 (13%)	2 (25%)	1 (13%)	**1 (13%)**	8

[a] Number correct: 48 (52%).

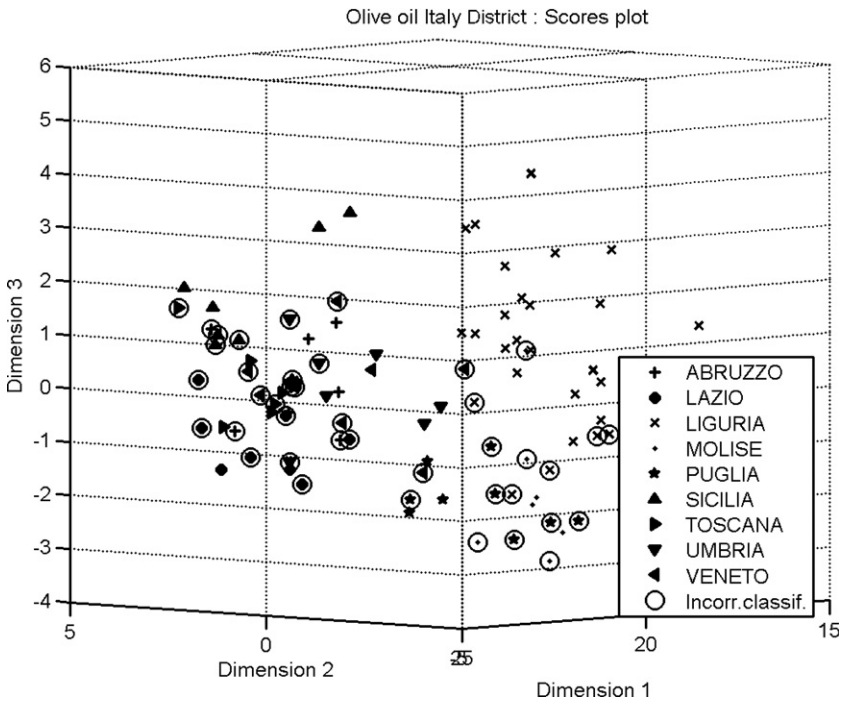

Fig. 4. Scores plot of the first three dimensions of the PLS-DA model for the 92 Italian olive oil samples: classification into district (smaller region).

smaller the district. Additionally, the number of samples per class may be of influence too. Although this analysis was carried out with at least eight samples for the smallest class, using a larger sample population may help to improve classification.

3.3. Factors influencing classification

The composition of organic compounds in a final olive oil product results from a combination of various factors such as cultivar, ripeness, processing and storage of the oil, and additionally strongly depends on climate and environmental factors. Some of the latter aspects associated with olive farming appear in the form of soil erosion, run-off to water bodies and degradation of habitats based on the interactions occurring between surface relief, soils, and water on the one hand, and the societal dynamics on the other. In terms of farming area, the altitude, slope, orientation, annual solar radiation, exposure to the prevailing winds, and fertility of the soil and environment are also important factors (Temime et al., 2006). Thus, the different VOC composition of olive oils from various regions is dependent on a complex assortment of influences. So, even if an oil is produced with the same cultivar and farming conditions and methods, the volatile nature of the final oil product could be severely altered by the fruit storage or oil production technique. Such a condition could therefore alter results of an origin classification. Nevertheless, results here show that despite all of these influential factors affecting the final olive oil product, a separation of oils according to geographic origin can be successfully made on the large scale and, to a lesser extent, on the district scale.

4. Conclusion

Mass spectral "fingerprints" were made for 192 olive oil samples from five different European countries using PTR-MS headspace analysis. Multivariate statistical analysis of these data allowed samples to be separated successfully into country of origin, using the masses as predictor variables. Further separation of the Italian samples into the smaller, region-scale (points of a compass) using this technique was also possible. On a more focussed, districtal scale, however, this method showed less strength, although this may have been attributable to the small sample size.

The advantages of this method over others are its simplicity, efficiency and reproducibility. Direct PTR-MS headspace analyses of olive oil samples were made without prior sample preparation and rapid mass spectra/sample were carried out in just over 2 min: this enabled many samples to be analysed in only a short period. Such a method could be developed to incorporate auto-sampling procedures to provide a maximum throughput of sample analyses and could provide a future screening technique for olive oil origin determination that is both fast and accurate.

Acknowledgements

The authors would like to thank the EU TRACE project and UNAPROL (the Italian consortium of olive oils) for providing the olive oil samples. Additionally, we would like

to acknowledge Ionicon Analytik GmbH, in particular Gernot Hanel, for PTR-MS support.

References

Angerosa, F., Bréas, O., Contento, S., Guillou, C., Reniero, F., & Sada, E. (1999). Application of stable isotope ratio analysis to the characterization of the geographical origin of olive oils. *Journal of Agriculture and Food Chemistry, 47*, 1013–1017.

Angerosa, F., Mostallino, R., Basti, C., & Vito, R. (2001). Influence of malaxation temperature and time on the quality of virgin olive oils. *Food Chemistry, 72*, 19–28.

Angerosa, F., Servili, R., Selvaggini, M., Taticchi, A., Esposto, S., & Montedoro, G. F. (2004). Volatile compounds in virgin olive oil: Occurrence and their relationship with the quality. *Journal of Chromatography A, 1054*, 17–31.

Aparicio, R., & Morales, M. T. (1998). Characterization of olive ripeness by green aroma compounds of virgin olive oil. *Journal of Agriculture and Food Chemistry, 46*, 1116–1122.

Aprea, E., Biasioli, F., Sani, G., Cantini, C., Märk, T. D., & Gasperi, F. (2006). Proton transfer reaction-mass spectrometry (PTR-MS) headspace analysis for rapid detection of oxidative alteration of olive oil. *Journal of Agricultural and Food Chemistry, 54*, 7635–7640.

Boskou, D. (1996). *Olive oil: Chemistry and technology*. Champaign, IL: AOCS Press (pp. 85–127).

Bréas, O., Sada, E., Reniero, F., Guillou, C., & Angerosa, F. (1998). Oxygen-18 measurement by continuous flow pyrolysis/isotope ratio mass spectrometry of vegetable oils. *Rapid Communications in Mass Spectrometry, 12*, 188–192.

Cavaliere, B., de Nino, A., Hayet, F., Lavez, A., Macchione, B., Moncef, C., et al. (2007). A metabolomic approach to the evaluation of the origin of extra virgin olive oil: A convenient statistical treatment of mass spectrometric analytical data. *Journal of Agricultural and Food Chemistry, 55*, 1454–1462.

Cosio, M. S., Ballabio, D., Benedetti, S., & Gigliotti, C. (2006). Geographical origin and authentication of extra virgin olive oils by an electronic nose in combination with artificial neural networks. *Analytica Chimica Acta, 567*, 202–210.

E.C. (2006). European Community, Regulation EC/510/2006.

Flath, R. A., Forrey, R. R., & Guadagni, D. G. (1973). Aroma components of olive oil. *Journal of Agricultural and Food Chemistry, 21*, 948–952.

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. *Food Chemistry, 100*, 273–286.

Kanavouras, A., Kiritsakis, A., & Hernandez, R. J. (2005). Comparative study on volatile analysis of extra virgin olive oil by dynamic headspace and solid phase micro-extraction. *Food Chemistry, 90*, 69–79.

Lai, A., Casu, M., Saba, G., Corongiu, F. P., & Dessi, M. (1995). A NMR investigation of the intramolecular distribution of deuterium in natural triacylglycerols. *Magnetic Resonance in Chemistry, 33*, 163–166.

Lindinger, W., Hansel, A., & Jordan, A. (1998). On the transmission function of an ion-energy and mass spectrometer. *International Journal of Mass Spectrometry and Ion Processes, 173*, 191–241.

Luchetti, F. (2002). Importance and future of olive oil in the world market – An introduction to olive oil. *European Journal of Lipid Science and Technology, 104*, 559.

Luna, G., Morales, M. T., & Aparicio, R. (2006). Characterisation of 39 varietal virgin olive oils by their volatile compositions. *Food Chemistry, 98*, 243–252.

Morales, M. T., Berry, A. J., McIntyre, P. S., & Aparicio, R. (1998). Tentative analysis of virgin olive oil aroma by supercritical fluid extraction-high-resolution gas chromatography–mass spectrometry. *Journal of Chromatography A, 819*, 267–275.

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.

Reid, L. M., O'Donnell, C. P., & Downey, G. (2006). Recent technological advances for the determination of food authenticity. *Trends in Food Science and Technology, 17*, 344–353.

Reiners, J., & Grosch, W. (1998). Odorants of virgin olive oils with different flavor profiles. *Journal of Agricultural and Food Chemistry, 46*, 2754–2763.

Romero, R., Sánchez-Viñas, M., Gázquez, D., & Bagur, M. G. (2002). Characterization of selected Spanish table wine samples according to their biogenic amine content from liquid chromatographic determination. *Journal of Agricultural and Food Chemistry, 50*, 4713–4717.

Stashenko, E. E., Wong, J. W., Martínez, J. R., Mateus, A., & Shibamoto, T. (1996). High-resolution gas chromatography with nitrogen–phosphorous detection of saturated volatile aldehydes derivatized with 2-hydrazinobenzothiazole. *Journal of Chromatography A, 752*, 209–216.

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

Vichi, S., Pizzale, L., Conte, L. S., Buxaderas, S., & López-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.

Williams, M., Morales, M. T., Aparicio, R., & Harwood, J. L. (1998). Analysis of volatiles from callus cultures of olive *Olea europaea*. *Phytochemistry, 47*, 1253–1259.

Wise, B. M., Shaver, J. M., Gallagher, N. B., Windig, W., Bro, R., & Koch, R. S. (2006). *PLS_Toolbox Version 4.0 for use with Matlab™*. Wenatchee, WA, USA: Eigenvector Research Inc.

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.