



Analytical Methods

Traceability of olive oil based on volatiles pattern and multivariate analysis

Tomas Cajka, Katerina Riddellova, Eva Klimankova, Monika Cerna, Frantisek Pudil, Jana Hajslova *

Institute of Chemical Technology, Prague, Faculty of Food and Biochemical Technology, Department of Food Chemistry and Analysis, Technicka 3, 166 28 Prague 6, Czech Republic

ARTICLE INFO

Article history:

Received 7 July 2009

Received in revised form 5 October 2009

Accepted 1 December 2009

Keywords:

Olive oil
Traceability
Authenticity
Head-space solid-phase microextraction
Comprehensive two-dimensional gas chromatography
Mass spectrometry
Multivariate analysis

ABSTRACT

An automated head-space solid-phase microextraction (HS-SPME)-based sampling procedure, coupled to gas chromatography–ion trap mass spectrometry (GC–ITMS), was developed and employed for fast characterisation of olive oil volatiles. In total, 914 samples were collected, over three production seasons, in north-western Italy–Liguria ($n = 210$) and other regions—in addition to the rest of Italy, Spain, France, Greece, Cyprus, and Turkey ($n = 704$) with the aim to distinguish, based on analytical (profiling) data, the olive oils labelled as “Ligurian” (protected denomination of origin region, PDO) from all the others (“non-Ligurian”). For the chemometric analysis, linear discriminant analysis (LDA) and artificial neural networks with multilayer perceptrons (ANN-MLP) were tested. Employing LDA, somewhat lower recognition (81.4%) and prediction (61.7%) abilities were obtained. The classification model was significantly improved using ANN-MLP. Under these conditions, the recognition (90.1%) and prediction (81.1%) abilities were achieved. The diagnostic value of the data obtained by one-dimensional GC–ITMS were compared with those generated by two-dimensional gas chromatography–time-of-flight mass spectrometry (GC \times GC–TOFMS), allowing a comprehensive analysis of olive oil volatiles.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

Olive oil, obtained from the fruit of olive trees (*Olea europaea* L.), represents a very popular food commodity, not only for its delicious flavour, but also for its health-promoting potential provided by polyunsaturated fatty acids and other compounds, such as phenolics (hydroxytyrosol, tyrosol, and oleuropein), squalene and oleic acid (monounsaturated fatty acid) (Waterman & Lockwood, 2007). Virgin olive oil is recovered by mechanical (pressure) or other physical means, applied under mild thermal conditions, avoiding alterations in its nature. The only allowed crude oil treatments include washing, decanting, centrifuging, and filtering (Codex Standard for Olive Oils and Olive Pomace Oils, 2003). According to the International Olive Council trade standard, the best brand, extra virgin olive oil, should meet specified physico-chemical characteristics, including low free acids content (<0.8 g/100 g, expressed as oleic acid) (Trade standard applying to olive oils and olive-pomace oils). The higher price of extra virgin olive oil compared to virgin olive oil is due to a limited production rate and strict requirements for its high quality parameters (Commission Regulations (EC) No. 1019/2002). One of the factors important for the consumers' choice is the geographic area of olive oil production. Up to now, the EU has specified 85 protected denomination of origin (PDO) regions for olive oil: 7 French, 15 Greek, 37 Italian (including “Riviera Ligure”), 6 Portuguese, 1 Slovenian, 19 Spanish

(DOOR – Database of Origin, 2009). Unfortunately, economic fraud, such as false claims of geographical origin on product labels, cannot be fully avoided. To protect the market from fraudulent practices and false label claims, a wide range of analytical strategies has been developed to confirm olive oil authenticity (only most recent papers are quoted here to document the state-of-the-art in this field). One of the oldest approaches is represented by the analysis of fatty acid methylesters and phytosterols by gas chromatography, employing a flame ionisation detector (GC–FID) (Angerosa, Campestre, & Giansante, 2006). Also, spectroscopic techniques, such as nuclear magnetic resonance (NMR), Fourier transform infrared (FT-IR) spectroscopy, fluorescence spectroscopy, near-infrared (NIR) spectroscopy, and/or mass spectrometry, have been widely used for fast olive oil fingerprinting (Dupuy et al., 2005; Hennessy, Downey & O'Donnell, 2009; Rezzi et al., 2005; Woodcock, Downey & O'Donnell, 2008). Besides analysis of the non-volatile fraction of olive oil, an examination of the complex profile of volatiles might be considered as a strategy for olive oil authentication (Angerosa et al., 2004). The use of direct head-space–mass spectrometry (HS–MS), yielding a fingerprint of the electron ionisation fragments, was reported in one of the published studies (Lorenzo, Pavon, Laespada, Pinto, & Cordero, 2002). For isolation of volatiles, dynamic head-space—involving their stripping from olive oil, trapping on a suitable adsorbent, and subsequent thermal desorption, or elution with a solvent—used to be “classic” sample preparation procedures prior to their gas chromatographic separation (Angerosa et al., 2004). Currently, a more preferred approach, used for characterisation of olive oil head-space volatiles, is their

* Corresponding author. Tel.: +420 220 443 185; fax: +420 220 443 184.
E-mail address: jana.hajslova@vscht.cz (J. Hajslova).

pre-concentration by solid-phase microextraction (SPME), followed by GC–MS. (Angerosa et al., 2004; Cavalli, Fernandez, Lizzani-Cuvelier, & Loiseau, 2004; Lorenzo et al., 2002; Mildner-Szkudlarz, Jelen, Zawirska-Wojtasiak, & Wasowicz, 2003; Ribeiro, Costa Freitas, & Gomes da Silva, 2008; Vichi et al., 2003). It should be noted, that only relatively small numbers of samples were examined in these studies; moreover, none of them explored the analytical challenge to identify geographical origin within a large sample set involving several harvest years.

If large volumes of data provided by the above fingerprint techniques are to be processed, smart chemometric tools are required to fully utilise this information. In most cases, principal component analysis (PCA) is used for a preliminary inspection of the data structure. In the next step, various classification methods such as linear discriminant analysis (LDA), partial least-squares discriminant analysis (PLS-DA), soft independent modelling of class analogy (SIMCA), or artificial neural networks (ANNs) are the most commonly considered options (Berrueta, Alonso-Salces, & Heberger, 2007; Bhadesia, 1999; Lebart, Morineau, & Warwick, 1984; Naes, Isaksson, Fearn, & Davies, 2002; Wold, 1976).

The current study has been conducted within the EU-funded TRACE project (www.trace.eu.org), aiming at development of cost-effective traceability methods and systems to provide consumers with added confidence in the authenticity of food in the European market. In addition to the spectroscopic methods mentioned above, we recorded fingerprints of olive oil volatiles by an HS-SPME-based procedure coupled to a GC–ion trap MS (ITMS), the objective being to provide an alternative screening tool for geographic origin traceability. Advanced chemometric strategies represented by LDA and ANN were employed for interpretation of the acquired data set. The feasibility of this approach for authentication purposes was demonstrated by an exercise in which olive oils labelled as “Ligurian” (protected denomination of origin region) were to be distinguished from other Mediterranean olive oil samples.

2. Materials and methods

2.1. Chemicals and materials

The SPME fibres tested were: (i) 100 μm polydimethylsiloxane (PDMS), (ii) 65 μm polydimethylsiloxane/divinylbenzene (PDMS/DVB), (iii) 65 μm Carbowax/divinylbenzene (CW/DVB), and (iv) 50/30 μm divinylbenzene/Carboxen/polydimethylsiloxane (DVB/CAR/PDMS). All of them were supplied by Supelco (Bellefonte, PA, USA). Prior to use, all fibres were conditioned, following the manufacturer's recommendations.

For GC–ITMS experiments an HP-INNOWax, polyethylene glycol (Agilent, Palo Alto, CA, USA) column; 30 m \times 0.25 mm I.D., 0.25 μm film thickness, was employed.

The system used for GC \times GC experiments comprised an HP-INNOWax, polyethylene glycol (Agilent, Palo Alto, CA, USA) primary column; 30 m \times 0.25 mm I.D., 0.25 μm film thickness, coupled via a column connector (Agilent, Palo Alto, CA, USA) to a BPX-50, 50% phenyl polysilphenylenesiloxane, second column; 1.25 m \times 0.1 mm I.D., 0.1 μm film thickness (*configuration A*). For further comparison, a GC configuration consisting of a DB-5 ms, 5% phenyl polysilphenylenesiloxane (Agilent, Palo Alto, CA, USA) primary column; 30 m \times 0.25 mm I.D., 0.25 μm film thickness, coupled via a column connector to a Supelcowax 10, polyethylene glycol (Supelco, Bellefonte, PA, USA) second column, 1.25 m \times 0.1 mm I.D., 0.1 μm film thickness (*configuration B*), was employed.

A mixture of *n*-alkanes (C₈–C₂₀) dissolved in *n*-hexane, which was employed for retention index determinations, was supplied

by Supelco (Bellefonte, PA, USA). The calculation was done for components eluting between *n*-octane and *n*-eicosane.

2.2. Analysed samples

In total, 914 extra virgin olive oil samples were collected within the framework of the EU TRACE project (www.trace.eu.org) over a period of three harvests—2005–2007, see Table 1. The samples were stored in a refrigerator at 4 °C between delivery and analysis (in maximum 4 weeks).

2.3. HS-SPME procedure

Prior to analysis, the sample was equilibrated to laboratory temperature (30 min), followed by placing 2 g aliquot in a 10 ml glass vial for SPME, which was then sealed using a magnetic cap with a PTFE/silicon septum.

The optimised conditions of HS-SPME procedure were as follows: SPME fibre: 50/30 μm DVB/CAR/PDMS; incubation time: 5 min; incubation and extraction temperature: 40 °C; agitator speed: 500 rpm; extraction time: 15 min; desorption temperature: 250 °C; desorption time: 60 s (splitless).

2.4. GC analysis

2.4.1. GC–ITMS

2.4.1.1. General. A system consisting of a Trace GC 2000 gas chromatograph equipped with a PTV injector (Thermo Quest, San Jose, CA, USA), a CombiPal autosampler for automated SPME (CTC Analytics, Zwingen, Switzerland), and a Polaris Q ion trap mass spectrometer (Finnigan, San Jose, CA, USA) was used.

2.4.1.2. GC separation. Oven temperature programme: 45 °C (1 min), 5 °C/min to 170 °C, 10 °C/min to 260 °C (1 min); carrier gas: helium (purity 99.9999%); column flow: 1 ml/min; inlet temperature: 250 °C.

2.4.1.3. ITMS detection. Electron ionisation mode (70 eV); ion source temperature: 200 °C; mass range (segment scan): *m/z* 35–70, 71–110, 111–160, 161–220, 221–320; acquisition rate: 1.3 scan/s; detector voltage: on the bases of autotune set-up.

2.4.1.4. Data processing. XCALIBUR (Finnigan, USA) software (v. 1.2.2) was used for instrument control, data acquisition, and data processing. Identification of compounds was based on a NIST 1.7 mass spectra library search.

2.4.2. GC \times GC–TOFMS

2.4.2.1. General. A Pegasus 4D system, consisting of an Agilent 6890N gas chromatograph equipped with a split/splitless injector (Agilent Technologies, Palo Alto, CA, USA), an MPS2 autosampler for automated SPME (Gerstel, Mülheim an der Ruhr, Germany), and a Pegasus III high-speed time-of-flight mass spectrometer (Leco Corp., St. Joseph, MI, USA), was used. Inside the GC oven a cryogenic modulator (N₂ jets-hot air jets technology) and a secondary oven (Leco Corp., St. Joseph, MI, USA) were mounted. Resistively heated air was used as a medium for hot jets, while cold jets were supplied by gaseous nitrogen cooled by liquid nitrogen.

2.4.2.2. GC \times GC separation—configuration A. Primary oven temperature programme: 45 °C (1 min), 5 °C/min to 170 °C, 40 °C/min to 260 °C (1.75 min); secondary oven temperature: +5 °C above the primary oven temperature; modulator offset: +20 °C above the primary oven temperature; modulation period: 3 s (hot pulse 0.6 s); carrier gas: helium (purity 99.9999%); column flow: 1.3 ml/min; inlet temperature: 250 °C.

Table 1
Overview of number and geographical origin of analysed olive oil samples within TRACE project.

Sampling year	Geographical origin							Total
	Liguria	Italy	Spain	France	Greece	Cyprus	Turkey	
2005	63	164	42	9	25	0	14	317
2006	79	173	38	10	46	6	0	352
2007	68	116	34	20	7	0	0	245
All years	210	453	114	39	78	6	14	914

2.4.2.3. GC × GC separation—configuration B. Primary oven temperature programme: 45 °C (1 min), 5 °C/min to 170 °C, 40 °C/min to 250 °C (2 min); secondary oven temperature: +15 °C above the primary oven temperature; modulator offset: +30 °C above the primary oven temperature; modulation period: 3 s (hot pulse 0.6 s); carrier gas: helium (purity 99.9999%); column flow: 1.3 ml/min; inlet temperature: 250 °C.

2.4.2.4. TOFMS detection. Electron ionisation mode (70 eV); ion source temperature: 220 °C; mass range: m/z 25–300; acquisition rate: 150 spectra/s; detector voltage: –1750 V.

2.4.2.5. Data processing. ChromaTOF (LECO Corp.) software (v. 2.31) was used for instrument control, data acquisition, and data processing. Identification of compounds was based on a NIST 2005 mass spectra library search. The confirmation of tentatively identified compounds was performed by comparing calculated linear retention indices with those available in NIST 2008.

2.5. Chemometric analysis

Chemometric analysis included principal component analysis and formation of an artificial neural network model, employing the software package STATISTICA “Neural Networks” (v. 6, 2003, StatSoft, Inc., Tulsa, OK, USA, www.statsoft.com). For linear discriminant analysis, the software statistiXL (v. 1.8, 2008, statistiXL, Broadway–Nedlands, Australia, www.statistiXL.com) was used.

3. Results and discussion

3.1. General

In the first phase of this study, the HS-SPME–GC–ITMS procedure was optimised for obtaining (as far as possible) a broad profile of volatile compounds released from olive oil samples. In the follow-up part, this method was employed for an examination of a large set of olive oils, which were presumed to differ in profiles of volatiles, depending on their geographical origin. The feasibility of employing this approach for the traceability purpose is presented in paragraphs below.

3.2. Optimisation of HS-SPME method

All the key parameters that may affect the SPME performance, such as a type of fibre coating, incubation/extraction time, and sample temperature, were subjects of optimisation carried out within our study. Among the tested fibres (see Section 2.1), the complexity (i.e. number of extracted compounds) of recorded olive oil profiles decreased in order: DVB/CAR/PDMS > PDMS/DVB > CW/DVB > PDMS. For this reason, the fibre with combined sorbents, 50/30 µm DVB/CAR/PDMS, was used in subsequent experiments.

Extraction temperatures of 30, 40, 50, 60, and 70 °C with a 15 min extraction period (+5 min incubation time at selected temperature) were employed to test DVB/CAR/PDMS fibre sorption efficiency. For most of the volatiles selected, an increase in extrac-

tion temperature up to 70 °C led to a growth in signal intensity, supposedly due to their facilitated transfer into the head-space. However, due to a risk of undesirable formation of volatiles (at elevated sampling temperature) from precursors contained in samples, 40 °C was selected as a compromise in the final method. In the next step, incubation times of 5, 10, 20, and 30 min were tested at 40 °C, followed by a 15 min extraction. No significant differences were observed for incubation times ranging between 5 and 20 min; after 30 min, some drop of signal intensity occurred. With regard to these results, a short incubation time, only 5 min, was used in the final method.

The last parameter we tuned was extraction time. Growth of the detector signal was observed for sampling times of 5, 15, 30, 50, and 60 min at 40 °C (with a 5 min incubation period). To avoid GC column and detector overflow, an extraction time of 15 min was used in subsequent experiments. This relatively short extraction time also allowed a reasonable sample throughput.

Relatively good repeatability of peak areas (RSDs ≤ 20%) of the optimised HS-SPME–GC–ITMS procedure was obtained within a series of 10 consecutive analyses for all 44 selected volatile compounds.

3.3. Characterisation of olive oil volatiles and selection of markers

For the chemometric analysis, several potential markers (volatiles) were selected after careful inspection of the overlaid GC profiles of the analysed olive oil samples (Fig. 1). The key selection criterion was a distinct difference in intensity of a particular peak among examined samples. The list of alcohols, aldehydes, ketones, esters, carboxylic acids, and hydrocarbons selected as markers, together with their retention indices, is shown in Table 2. It should be noted that the identification of some volatiles in the GC–ITMS system was only tentative since spectral match was not strong enough; thus, to confirm the identity of these compounds, comprehensive two-dimensional gas chromatography, coupled to time-of-flight mass spectrometry (GC × GC–TOFMS), was considered as an option.

3.4. Improved identification of olive oil volatiles using GC × GC–TOFMS

The decision to exploit the potential of GC × GC–TOFMS for obtaining more reliable, unambiguous identification of marker olive oil volatiles was motivated by challenging features of this technology (not provided by the conventional, one-dimensional GC–ITMS system used for profiling purposes within this study): (i) higher peak capacity resulting from the combination of two different GC columns with independent separation mechanisms, (ii) an increase of S/N ratio (due to re-focusing of the analyte in the modulator), and hence, not only improved detectability, but also better spectral quality, enabled by improved separation of chemical noise in the GC × GC system (significantly higher peak capacity), (iii) formation of structured chromatograms thanks to complementary separation mechanisms occurring on both columns, and (iv) the ability to deconvolute overlapping peaks unresolved even under

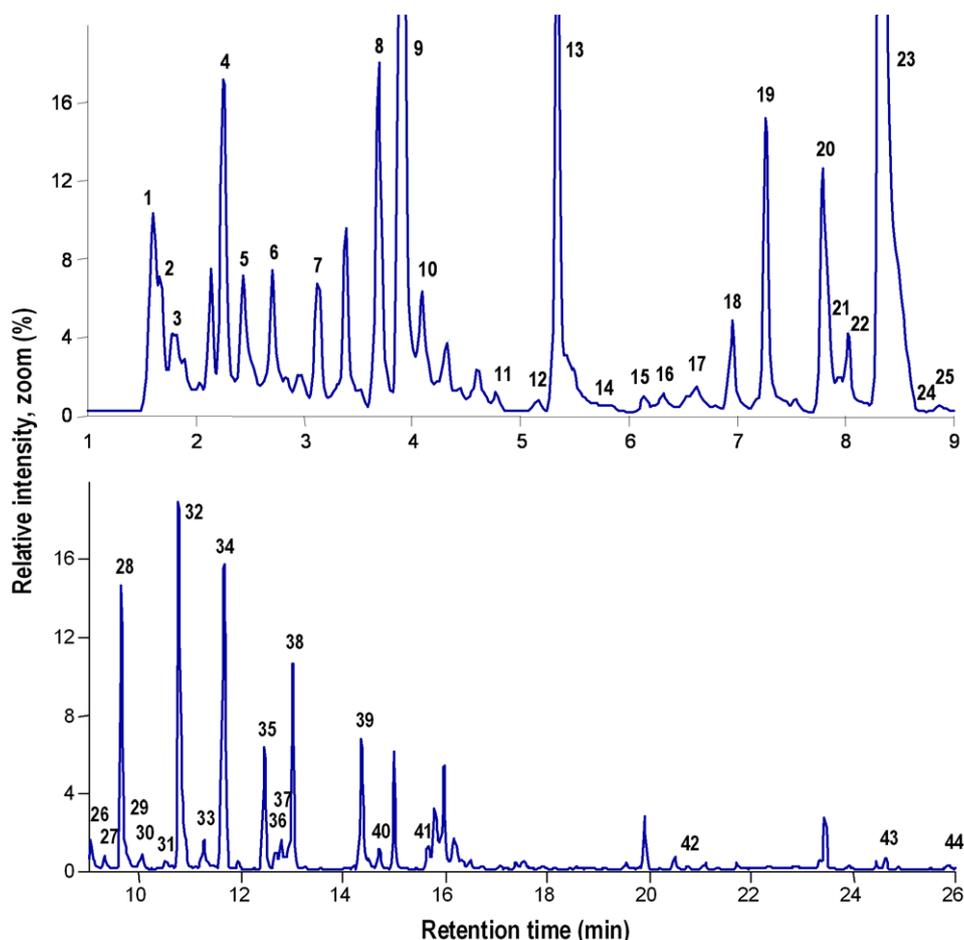


Fig. 1. HS-SPME–GC–ITMS chromatogram of olive oil volatiles with marked marker compounds; for their description see Table 1.

conditions of GC \times GC (Cajka, Hajslova, & Mastovska, 2008; Dalluge, Beens, & Brinkman, 2003). For the separation of volatiles of olive oil samples, two GC \times GC column set-ups were tested. In the first instance, a “standard” combination, consisting of a non-polar narrow-bore (5% phenyl polysilphenylenesiloxane, DB-5 ms) and a polar microbore column (polyethylene glycol, Supelcowax 10), was used. In addition to this “orthogonal” separation approach, “reversed-type” column combination was examined, employing a polar narrow-bore column (polyethylene glycol, HP-INNOWax) and a medium-polar microbore column (50% phenyl polysilphenylenesiloxane, BPX-50). As shown in Fig. 2, these approaches provided substantially differing grouping of compounds. With the non-polar \times polar column combination, the 2D separation plane is relatively well occupied. However, compared to the polar \times medium-polar column combination, higher temperature offset between primary and secondary oven was required (5 vs. 15 °C) to avoid wrap-around of some peaks. It should be noted that the use of a relatively unstable stationary phase in the second dimension limits the setting of the upper temperature. Hence, the elution of less volatile compounds can be rather problematic (band broadening due to higher retention times). In addition, highly polar analytes (e.g. formic and acetic acid), strongly retained in the non-polar \times polar column set-up, may cause co-elutions such as the broad wrap-around peaks with peaks eluting in their own modulation cycle. In so far as the identification of volatiles in the GC \times GC–TOFMS system was performed on the basis of criteria such as (i) similarity of measured mass spectra to NIST library records (>800), (ii) signal-to-noise (S/N) ratio (>100), and (iii) linear retention indices (± 20 index units), the number of iden-

tified volatiles was higher in the case of the non-polar \times polar column combination (170) than in the polar \times medium-polar column combination (119). The lower number of identified compounds in the latter case can be attributed to a greater column bleed of the polar, relatively long column used in the first dimension, and thus, worse quality of mass spectra, but also lower peak capacity of this system. In any case, this number of reported compounds was still fairly high compared to 1D-GC–ITMS (50) for the same olive oil sample. In other words, GC \times GC–TOFMS offered a remarkable potential for both chromatographic and spectrometric separation of olive oil volatiles and, consequently, minimisation of the risk of incorrect identification. However, taking into account the large volume of data generated by GC \times GC–TOFMS and their rather demanding handling, the conventional approach, employing 1D-GC–MS, was preferred for the analysis of 914 samples.

3.5. Chemometric analysis

3.5.1. General

Prior to the chemometric analysis, employing PCA (unsupervised pattern recognition technique) and LDA/ANN (supervised pattern recognition techniques), the raw data (914 samples, each characterised by 44 markers – variables) acquired by GC–ITMS analyses in the form of absolute peak intensities (total ion current, TIC) at respective retention time windows were pre-processed using a “constant row sum” approach; i.e. each variable was divided by the sum of all variables for each sample. This procedure transformed all data into a uniform range of variability (Berrueta et al., 2007). (Note: The other scaling methods such as “range trans-

Table 2
Overview of olive oil volatiles (selected markers) employed for chemometric analysis.

No.	Compound	RI	RI _{NIST}	Identification
1	Penta-1,3-diene	689	728	b
2	3-Methylhexane	700	680	b
3	Propanal	804	790	b
4	Acetone	841	846	b
5	Prop-2-enal	864	869	b
6	Ethyl acetate	896	891	b
7	Ethanol	943	940	a,b
8	Pentanal	988	982	b
9	2-Methylpentan-3-one	1007	1000	b
10	3-Ethyl-octa-1,5-diene (isomer I)	1033	1027	b
11	Ethyl 2-methylbutanoate	1062	1051	b
12	3-Ethyl-octa-1,5-diene (isomer II)	1080	n/a	b
13	Hexanal	1094	1099	a,b
14	Pent-2-enal (isomer I)	1119	1124	b
15	Ethylbenzene	1139	1134	b
16	Pent-2-enal (isomer II)	1145	1144	b
17	1,4-Xylene	1153	1150	a,b
18	Pent-1-en-3-ol	1170	1165	a,b
19	Heptan-2-one	1191	1187	b
20	Dodecane	1201	1200	b
21	Hex-2-yn-1-ol	1213	n/a	b
22	3-Methylbutan-1-ol	1215	1211	b
23	Hex-2-enal	1230	1225	a,b
24	1-Ethyl-2-methylbenzene	1234	1231	b
25	Unknown (<i>m/z</i> 41, 55, 70, 83, 97, 111)	1244	n/a	c
26	3,7-Dimethylocta-1,3,6-triene	1257	1255	b
27	Unknown (<i>m/z</i> 41, 53, 67, 74, 124)	1263	n/a	c
28	Hexyl acetate	1276	1265	a,b
29	1,2,4-Trimethylbenzene	1290	1294	b
30	Octanal	1298	1291	b
31	Unknown (<i>m/z</i> 41, 53, 69, 79, 91, 107, 122, 135, 150)	1311	n/a	c
32	Pent-2-en-1-ol	1329	1318	a,b
33	Unknown (<i>m/z</i> 43, 55, 87, 101, 114)	1342	n/a	c
34	Hexan-1-ol	1360	1356	a,b
35	Hex-3-en-1-ol	1372	1378	a,b
36	Nonanal	1401	1390	b
37	Hexa-2,4-dienal	1411	1409	b
38	Hex-2-en-1-ol	1415	1417	a,b
39	Acetic acid	1492	1480	a,b
40	Copaene	1498	1493	b
41	Octa-3,5-dien-2-one	1531	1536	b
42	Farnesene	1755	1753	a,b
43	2-Phenylethanol	1927	1937	a,b
44	Nonanoic acid	2138	2123	a,b

n/a, not available.

RI, retention index.

RI_{NIST}, retention index available from NIST 2008 mass library.

^a GC–ITMS.

^b GC × GC–TOFMS.

^c Compound not identified under the conditions of GC–ITMS and GC × GC–TOFMS (the most characteristic ions given).

formation” and “logarithmic transformation” were also tested, but the results obtained by these classification models were less satisfactory.)

3.5.2. Principal component analysis

Principal component analysis (PCA), applied in the first stage of the data processing, represents one of the most frequently used chemometric tools. One of its attractive features is the possibility to project (in a relatively easy way) particular data from a higher to a lower dimensional space and then reconstruct them without any preliminary assumptions about their distribution (Stanimirova, Daszykowski, & Walczak, 2007). In the preliminary analysis of the olive oil data, PCA was performed to investigate any potentially existing clustering of samples on the basis of geographical origin and the year of harvest. The first principal compound (PC) accounted for 16.7% of variance, while the second PC contributed 9.5%. The most important PCs (with eigenvalues >1), contributed to 65% of total variance. As Fig. 3 shows, using this approach, the olive oils were divided into groups described as “Ligurian” and

“non-Ligurian” and, at the same time, the year of harvest. The formed clusters indicated a large variation in the profiles of olive oil volatiles between the three harvest years, possibly caused by different weather conditions during the growing season. To overcome this limitation, we decided to form LDA and ANN models that integrate the data from all three harvests instead of separate ones that led to worse performance (when model of one year was applied to other years). For the LDA and ANN, the data were randomly split to obtain an equal number of samples in a calibration (training) set ($n = 280$, 1:1 ratio of Ligurian and non-Ligurian samples) and the remaining samples ($n = 634$) were used as test set.

3.5.3. Linear discriminant analysis

Linear discriminant analysis (LDA) is probably the most frequently used supervised pattern recognition method. In principle, LDA determines linear discriminant functions, which maximise the ratio of between-class variance and minimise the ratio of within-class variance. It should be noted that, whereas PCA selects a direction that retains maximal structure among the data in a lower

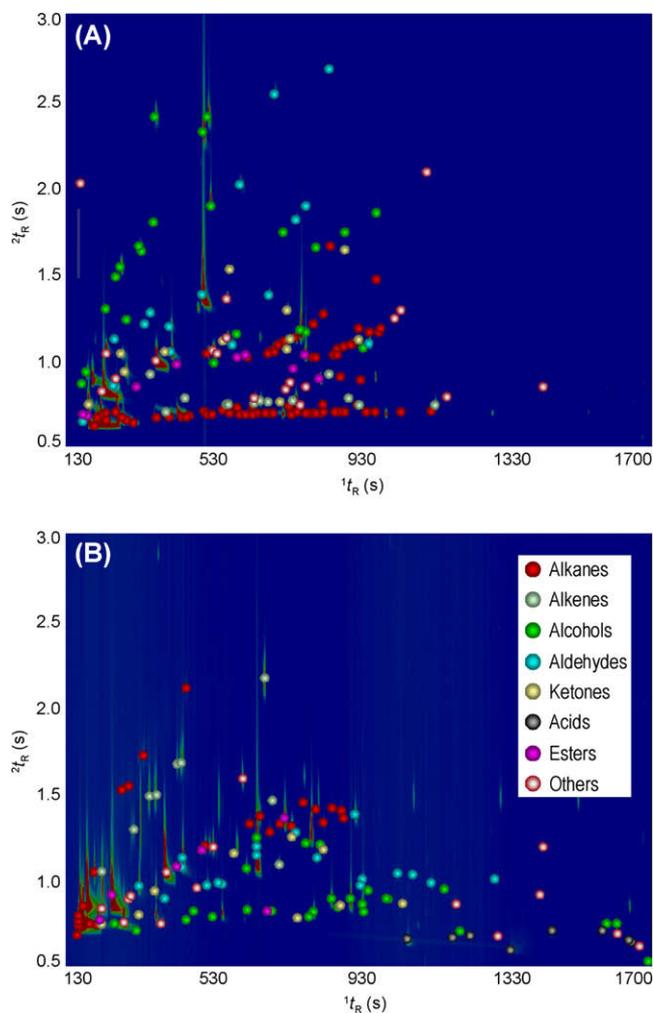


Fig. 2. Separation of olive oil volatiles in different GC \times GC systems consisting of the following capillary columns: (A) DB-5 ms \times Supelcowax 10 (orthogonal column configuration), (B) HP-INNOWax \times BPX-50 (reversed-type column configuration).

dimension, LDA selects a direction that achieves maximum separation among the given classes (Berrueta et al., 2007). As Table 3 shows, the recognition and prediction abilities of this model were somewhat low, 81.4% and 61.7%, respectively. (Note: Recognition ability represents percentage of the samples in the training set successfully classified, whereas prediction ability is the percentage of the samples in the test set correctly classified by using the model developed during the training step, Berrueta et al., 2007.) The analysis of the misclassified samples showed that the LDA model presented a relatively high sensitivity (89.9%) but a quite low selectivity (58.2%); thus, the model is able to identify most of Ligurian samples; however, its ability to classify the non-Ligurian samples is small. This led us to the conclusion that the relationship between predictor variables (independents, inputs) and predicted variables (dependents, outputs) is very complex and requires an “advanced” chemometric tool. Under these conditions, we decided to employ ANN, a challenging tool applicable for non-linear modelling (Yi, Gerardo, Lee, & Lee, 2006).

3.5.4. Artificial neural networks

The most common neural network (ANN) approach to regression-type problems is multilayer perceptrons (MLP) (Olszewski, Ryniecki, & Boniecki, 2008; Siripatrawan & Harte, 2007). An ANN-MLP, using the back propagation, was employed to predict the geographical origin of olive oil samples based on the pattern

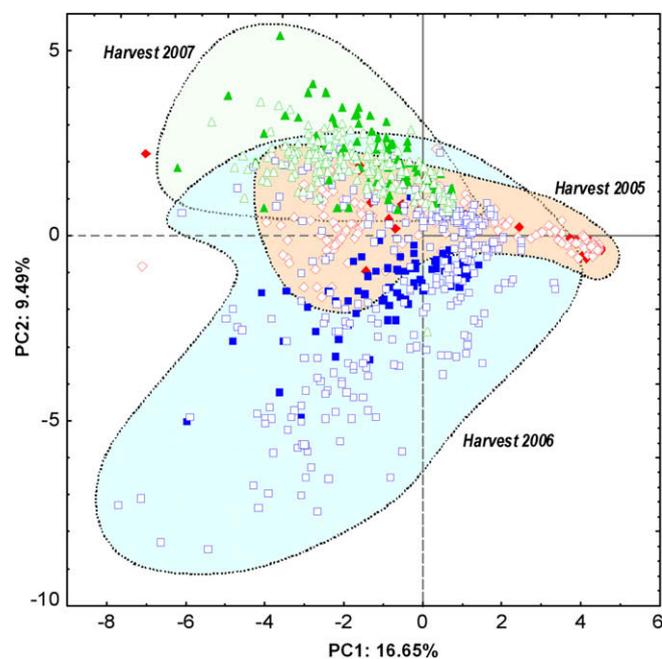


Fig. 3. PCA clustering: (A) year 2005 (\blacklozenge Liguria, \diamond non-Liguria); year 2006 (\blacksquare Liguria, \square non-Liguria); year 2007 (\blacktriangle Liguria, \triangle non-Liguria).

Table 3

Overall summary of performance of tested chemometric models employed in this study.

	Recognition ability [%]		Prediction ability [%]
	Training subset	Selection subset	Test subset
ANN-MLP model 44:44–25–1:1	94.1	84.0	81.1
LDA model 44 variables	81.4		61.7

of their volatiles. In the first step, the calibration data set was randomly divided by the software into two subsets: (i) *training subset* (2/3 of calibration data set, $n = 186$), which is used to accomplish the network model training; (ii) *selection subset* (1/3 of calibration data set, $n = 94$) for checking the network within the training process to avoid network overtraining. The *test subset* (i.e. remaining data, $n = 634$) represents the tool to assess the quality of the generated model. *Intelligent Problem Solver* was employed for the analysis of data. The search for an appropriate ANN model was restricted only to MLP networks. In total, 50 networks were tested, of which the best 10 were retained. The network architecture created for the olive oil data matrix included: (i) an input layer consisting of 44 neurons (marker compounds), (ii) one hidden layer with 25 neurons, and (iii) an output layer represented by one neuron providing origin classification. The ANN was trained using selected parameters from the data sets, followed by the validation using an independent data set to estimate the olive oil origin (Ligurian vs. non-Ligurian). The training started with different initial random weights, and was optimised during the process. Typically, the learning process continues by epoch-by-epoch (through single complete training processes) until the synaptic weights and bias level of the network are stabilised (Siripatrawan & Harte, 2007). In this study, the network was trained by a back propagation algorithm (100 epochs), followed by a conjugate gradient algorithm (20 epochs). Finally, a network with the smallest error (misclassification of sample origin in particular case) was selected. As Table 3 shows, the recognition ability, consisting of training and selection subsets, was 90.1% (94.1% for training subset and 84.0% for selec-

tion subset), and the prediction ability was 81.1%, and thus, substantially improved as compared to the LDA model (61.7%). Using the ANN-MLP model, the misclassified samples showed quite high sensitivity and selectivity (84.1% and 80.7%, respectively). In other TRACE studies, employed for examination of the same olive oil sample set and FR-IR spectroscopy, followed by partial least-squares discriminant analysis (PLS-DA), the best model was characterised by a sensitivity of 86% and corresponding selectivity of 78% (Hennessy, Downey, & O'Donnell, 2009), thus, comparable results were obtained. Somewhat better sensitivity and selectivity (92.8% and 81.5%, respectively) were obtained by the NIR spectroscopy technique with PLS-DA (Woodcock, Downey, & O'Donnell, 2008). In another TRACE study, assessing only samples obtained from two harvest years by FR-IR spectroscopy, the classification and regression trees (CART) showed a relatively high selectivity (81.7%) but a quite low sensitivity (34.8%); improvements of selectivity (58.5%) and sensitivity (93.8%) were observed when support vector machines (SVM) with Euclidian distance-based Pearson VII Universal Kernel (PUK) were applied (Caetano et al., 2007).

4. Conclusions

The results obtained within this traceability exercise, aiming to distinguish Ligurian olive oil samples from other Mediterranean regions (non-Ligurian), can be summarised as follows:

- (i) Among various approaches applicable for (non-target) profiling of olive oil characteristics for the identification of geographical origin, fingerprinting by employing SPME sampling of head-space volatiles, followed by GC separation and MS detection, represents an effective analytical option, since the whole process can be fully automated. Moreover, the identity of fingerprint components can be obtained, based on mass spectral data (which is not the case for IR-based methods). From the two analytical strategies tested, GC × GC–TOFMS provided a more comprehensive fingerprint of olive oil volatiles (and better spectral quality). Nevertheless, the number of diagnostic sample markers (44 peaks) delivered by conventional one-dimensional GC–ITMS was sufficient for chemometric analysis.
- (ii) PCA showed relatively large inter-annual variability in sample composition from particular regions, there was some overlap of the three harvests (three years) data. The model developed for a single year data set was not fully applicable for other sampling years. A more reliable approach appeared to be a model that consisted of the three-year sampling data. Employing this strategy, the recognition and prediction abilities obtained for ANN-MLP (non-linear modelling tool) were 90.1% and 81.1%, respectively, while somewhat lower abilities were achieved for LDA (81.4% and 61.7%, respectively).
- (iii) Considering transition of this technique into an industrial setting, comprehensive databases should be established in order to take into account the variability of factors. In addition to weather conditions during the growing season, possible fluctuation of storage conditions and the periods in the time scale “harvest → processing → sample analysis” may also be reflected in the volatiles profile. The developed technique is not stand-alone and for 100% accuracy, samples that are misclassified by the SPME–GC–ITMS technique would require further confirmatory analysis using a complementary approach such as FT-IR spectroscopy; a 20% possibility of a model rejecting a Ligurian sample or falsely accepting a non-Ligurian sample means that this technique may be used as a screening tool.

Disclaimer

The information reported reflects the authors' views; the European Commission is not liable for any use of the information contained therein. Mention of brand or firm names in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the Institute of Chemical Technology, Prague.

Acknowledgements

The financial support by the European Commission through the 6th Framework Programme (Contract No. FP6-FOOD–2004-006942 – TRACE) and the Ministry of Education, Youth and Sports of the Czech Republic (MSM 6046137305) is gratefully acknowledged.

References

- Angerosa, F., Campestre, C., & Giansante, L. (2006). Analysis and authentication. In D. Boskou (Ed.), *Olive oil: Chemistry and technology* (pp. 113–154). Champaign, Illinois: AOCS Publishing.
- Angerosa, F., Servili, M., Selvaggini, R., Taticchi, A., Esposito, 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.
- Berrueta, L. A., Alonso-Salces, R. M., & Heberger, K. (2007). Supervised pattern recognition in food analysis. *Journal of Chromatography A*, 1158, 196–214.
- Bhadeshia, H. K. D. H. (1999). Neural networks in materials science. *ISIJ International*, 39, 966–979.
- Caetano, S., Ustun, B., Hennessy, S., Smeyers-Verbeke, J., Meissen, W., Downey, G., et al. (2007). Geographical classification of olive oils by the application of CART and SVM to their FT-IR. *Journal of Chemometrics*, 21, 324–334.
- Cajka, T., Hajslova, J., & Mastovska, K. (2008). Mass spectrometry and hyphenated instruments in food analysis. In S. Otle (Ed.), *Handbook of food analysis instruments* (pp. 197–228). Boca Raton: CRC Press.
- Cavalli, J.-F., Fernandez, X., Lizzani-Cuvellier, L., & Loiseau, A.-M. (2004). Characterization of volatile compounds of French and Spanish virgin olive oils by HS-SPME: Identification of quality-freshness markers. *Food Chemistry*, 88, 151–157.
- Codex Standard for Olive Oils and Olive Pomace Oils (2003). Codex Standard 33–1981 [Rev. 2-2003].
- Dalluge, J., Beens, J., & Brinkman, U. A. T. (2003). Comprehensive two-dimensional gas chromatography: A powerful and versatile analytical tool. *Journal of Chromatography A*, 1000, 69–108.
- DOOR – Database of Origin and Registration. <<http://ec.europa.eu/agriculture/quality/door/>> Accessed 07.07.09.
- Dupuy, N., Le Dreau, Y., Ollivier, D., Artaud, J., Pinatel, C., & Kister, J. (2005). Origin of French virgin olive oil registered designation of origins predicted by chemometric analysis of synchronous excitation-emission fluorescence spectra. *Journal of Agricultural Food Chemistry*, 53, 9631–9638.
- European Commission. Commission Regulations (EC) No. 1019/2002 of 13 June 2002 on the marketing standards for olive oil (2002). *Official Journal of the European Communities*, L155, 27–31.
- Hennessy, S., Downey, G., & O'Donnell, C. P. (2009). Confirmation of food origin claims by Fourier transform infrared spectroscopy and chemometrics: Extra virgin olive oil from Liguria. *Journal of Agricultural Food Chemistry*, 57, 1735–1741.
- Trade standard applying to olive oils and olive-pomace oils, International Olive Council, COI/T.15/NC No. 3/Rev. 3, 24 November 2008.
- Lebart, L., Morineau, A., & Warwick, K. M. (1984). *Multivariate descriptive analysis correspondence analysis and related techniques for large matrices*. New York: Wiley.
- Lorenzo, J. M., Pavon, J. L. P., Laespada, M. E. F., Pinto, C. G., & Cordero, B. M. (2002). Detection of adulterants in olive oil by headspace-mass spectrometry. *Journal of Chromatography A*, 945, 221–230.
- Mildner-Szkudlarz, S., Jelen, H. H., Zawirska-Wojtasiak, R., & Wasowicz, E. (2003). Application of headspace – Solid phase microextraction and multivariate analysis for plant oils differentiation. *Food Chemistry*, 83, 515–522.
- Naes, T., Isaksson, T., Fearn, T., & Davies, T. (2002). *A user-friendly guide to multivariate calibration and classification*. Chichester: NIR Publications.
- Olszewski, T., Ryniecki, A., & Boniecki, P. (2008). Neural network development for automatic identification of the endpoint of drying barley in bulk. *Journal of Research and Applications in Agricultural Engineering*, 53, 26–31.
- Rezzi, S., Axelson, D. E., Heberger, K., Reniero, F., Mariani, C., & Guillou, C. (2005). Classification of olive oils using high throughput flow ¹H NMR fingerprinting with principal component analysis, linear discriminant analysis and probabilistic neural networks. *Analytica Chimica Acta*, 552, 13–24.
- Ribeiro, L. H., Costa Freitas, A. M., & Gomes da Silva, M. D. R. (2008). The use of headspace solid phase microextraction for the characterization of volatile compounds in olive oil matrices. *Talanta*, 77, 110–117.

- Siripatrawan, U., & Harte, B. R. (2007). Solid phase microextraction/gas chromatography/mass spectrometry integrated with chemometrics for detection of *Salmonella typhimurium* contamination in a packaged fresh vegetable. *Analytica Chimica Acta*, *581*, 63–70.
- Stanimirova, I., Daszykowski, M., & Walczak, B. (2007). Dealing with missing values and outliers in principal component analysis. *Talanta*, *72*, 172–178.
- Vichi, S., Castellote, A. I., Pizzale, L., Conte, L. S., Buxaderas, S., & Lopez-Tamames, E. (2003). Analysis of virgin olive oil volatile compounds by headspace solid-phase microextraction coupled to gas chromatography with mass spectrometric and flame ionization detection. *Journal of Chromatography A*, *983*, 19–33.
- Waterman, E., & Lockwood, B. (2007). Active components and clinical applications of olive oil. *Alternative Medicine Review*, *12*, 331–342.
- Wold, S. (1976). Pattern-recognition by means of disjoint principal components models. *Pattern Recognition*, *8*, 127–139.
- Woodcock, T., Downey, G., & O'Donnell, C. P. (2008). Confirmation of declared provenance of European extra virgin olive oil samples by NIR spectroscopy. *Journal of Agricultural Food Chemistry*, *56*, 11520–11525.
- Yi, S., Gerardo, B. D., Lee, Y.-S., & Lee, J. (2006). Intelligent information search mechanism using filtering and NFC based on multi-agents in the distributed environment. In M. Gavrilova et al. (Eds.), *Computational science and its applications—ICCSA 2006* (pp. 144–153). Heidelberg: Springer.