



Detection of refined olive oil adulteration with refined hazelnut oil by employing NMR spectroscopy and multivariate statistical analysis

Alexia Agiomyrghianaki^a, Panos. V. Petrakis^b, Photis Dais^{a,*}

^a NMR Laboratory, Department of Chemistry, University of Crete, P.O. Box 2208, Voutes Campus, 710 03 Heraklion, Crete, Greece

^b National Agricultural Research Foundation, Institute of Mediterranean Forest Research, Laboratory of Entomology, Terma Alkmanos, Ilissia, 115 28, Athens, Greece

ARTICLE INFO

Article history:

Received 3 November 2009

Accepted 5 November 2009

Available online 12 November 2009

Keywords:

NMR spectroscopy

Refined olive oil

Refined hazelnut oil

Adulteration

Canonical discriminant analysis

Classification binary tree

ABSTRACT

NMR spectroscopy was employed for the detection of adulteration of refined olive oil with refined hazelnut oil. Fatty acids and iodine number were determined by ¹H NMR, whereas ³¹P NMR was used for the quantification of minor compounds including phenolic compounds, diacylglycerols, sterols, and free fatty acids (free acidity). Classification of the refined oils based on their fatty acids content and the concentration of their minor compounds was achieved by using the forward stepwise canonical discriminant analysis (CDA) and the classification binary trees (CBTs). Both methods provided good discrimination between the refined hazelnut and olive oils. Different admixtures of refined olive oils with refined hazelnut oils were prepared and analyzed by ¹H NMR and ³¹P NMR spectroscopy. Subsequent application of CDA to the NMR data allowed the detection of the presence of refined hazelnut oils in refined olive oils at percentages higher than 5%. Application of the non-linear classification method of the binary trees offered better possibilities of measuring adulteration of the refined olive oils at a lower limit of detection than that obtained by the CDA method.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Numerous attempts have been made in the past to detect virgin olive oil adulteration with seed oils and particularly with hazelnut oil [1–23]. Although no official analytical method has been adopted by the European Commission and other international institutions for the detection of hazelnut oil, several methods using chromatographic and spectroscopic techniques have been proposed to distinguish extra virgin or virgin olive oil from hazelnut oil. Several of these methodologies detect virgin olive oil adulteration with crude hazelnut oil on the basis of chemical compounds that are present in one oil but not in the other oil. For instance, filbertone, which was considered as a hazelnut oil marker [1], had been proposed to detect olive oil fraud [1–4], whereas other researchers suggested to determining chlorophylls [5], tocopherols [6], volatile compounds [7,8] and phenolic compounds [9,10]. Detection of virgin olive oil with refined hazelnut oil (RHZO) presents increasing difficulties because the filbertone marker and other minor components in hazelnut oil are removed upon refinement [3,4], and in addition the triacylglycerol profile of both oils is very similar. Chromatography has been employed to distinguish RHZO in virgin olive oil on the basis of triacylglycerol [11,12] and/or sterol

[11–13] profiles. Several spectroscopic techniques in combination with multivariate statistical methods were successful in classifying edible oils and detect the presence of RHZO with little or great success [14–21]. The real problem, however, is the detection of RHZO in refined olive oil (ROO), since their minor components, and in particular their glyceridic composition, were considered to be very similar. Recent applications of FT MIR and FT-Raman and fluorescence spectroscopy [22,23] allowed detection of ROO adulteration with RHZO at concentrations >8%.

NMR spectroscopy is one of the most promising spectroscopic techniques for the analysis of complex systems, such as food matrices. It has been proven to be very successful in the analysis of olive oil and in combination with multivariate statistical methods was able to classify edible oils and detect olive oil adulteration with seed oils and olive oils of inferior quality [15,16,20,24–26]. It made possible the detection of the presence of refined hazelnut oil in virgin olive oil at concentrations 5–10% [14–16,20]. ¹H and ¹³C NMR spectroscopy shows a number of advantages relative to other analytical techniques; it is fast (less than 5 min are required to record an one-dimensional ¹H NMR spectrum of an olive oil sample in chloroform-d); it needs no calibration with internal standards or separation of the various components prior to the analysis; it shows remarkable selectivity, inasmuch it identifies unknown compounds at a molecular level; it gives a wealth of information in a single experiment; it is quantitative with excellent repeatability and reproducibility. Disadvantages of NMR spectroscopy may be considered its low sensitivity compared with chromatographic

* Corresponding author. Tel.: +30 281054037; fax: +30 2810545001.

E-mail addresses: ppetrakis@fria.gr (Panos.V. Petrakis), dais@chemistry.uoc.gr (P. Dais).

and mass spectrometry techniques and the high cost of the analysis, since the NMR spectrometers are quite expensive. Nevertheless, sensitivity problems have been overcome lately by using strong magnetic fields produced by super-conductive selenoids and properly designed probes, especially cryogenic probes, whereas the screening of a large number of samples in a short period of time may compensate the high cost of the analysis. The same advantages shares the NMR spectroscopy of the phosphorus-31 nucleus, although derivatization with phosphorus reagent is needed prior to the analysis, since apart from phospholipids, compounds in olive oil do not contain phosphorus atoms in their molecules. This method is based on the replacement of the acidic hydrogens of the hydroxyl and carboxyl groups of olive oil constituents with the phosphorus reagent 2-chloro-4,4,5,5-tetramethylphospholane and the use of ^{31}P NMR spectroscopy to identify the phosphitylated compounds [27]. This facile magnetic resonance method supplements ^1H and ^{13}C NMR techniques, especially in cases where severe overlapped signals in ^1H NMR spectra or long relaxation times of the insensitive ^{13}C nuclei render the analysis a difficult task.

The aim of the present study was to assess the potential of ^1H and ^{31}P NMR spectroscopy to detect ROO adulteration with RHZO at low concentration. This goal will be achieved upon application to the NMR data two statistical methods, namely, forward stepwise canonical discriminant analysis (CDA) and classification binary trees (CBTs).

2. Experimental

2.1. Refined oils

A collection of 20 samples of refined olive oil were supplied by the olive oil company MINERVA (Athens, Greece). The refined oils were obtained upon refinement of lampante olive oils originated from various locations of Greece. Twenty commercial samples of RHZO were purchased from Henry Lamotte (Bremen, Germany). To detect ROO adulteration, admixtures were prepared by mixing different samples of ROO with different samples of RHZO. The percentage of RHZO in ROO of the adulterated refined olive oils (AOO) varied from 1% to 95% (1%, 2%, 3%, 4%, 5%, 7%, 10%, 12%, 15%, 17%, 20%, 25%, 30%, 50%, 60%, 70%, 80%, 90%, 95%: w/w). The ROO and RHZO samples were selected randomly for the formation of each one of the 19 admixtures. All oil samples and their admixtures were analyzed twice.

2.2. Reagents

Pinacol, triethylamine, phosphorus trichloride, protonated solvents (reagent or analytical grade), and deuterated solvents used in the present study were purchased from Sigma–Aldrich (Athens-Greece). The derivatizing phosphorous reagent 2-chloro-4,4,5,5-tetramethyldioxaphospholane used in ^{31}P NMR spectroscopy was synthesized from pinacol and phosphorus trichloride following the method described in the literature [28]. However, to increase the yield of the reaction, we utilized hexane solvent instead of benzene and pyridine instead of triethylamine suggested in the original method. This modification resulted in ~45% yield of the product against 19% obtained with the original method.

2.3. Extraction of minor polar components from refined oil

Phenolic compounds were extracted following the method developed by Montedoro et al. [29] using 35 g of olive oil and a mixture of methanol–water (80:20, v/v). The polyphenol extracts were used immediately for sample preparation and ^{31}P NMR measurements.

2.4. Sample preparation for ^{31}P NMR spectral analysis

A stock solution was prepared by dissolving 0.6 mg of chromium acetylacetonate, $\text{Cr}(\text{acac})_3$ (0.165 μM) and 13.5 mg cyclohexanol (13.47 mM) in 10 ml of a mixture of pyridine and CDCl_3 solvents (1.6:1.0, volume ratio) and protected from moisture with 5A molecular sieves. Cyclohexanol was used as an internal standard for quantification purposes. 100–150 mg of each oil was placed in a 5 mm NMR tube. The required volumes of the stock solution (0.4 mL) and the reagent **1** (15 μL) were added. The reaction mixture was left to react for about 15 min at room temperature. Upon completion of the reaction, the solution was used to obtain the ^{31}P NMR spectra.

2.5. NMR experiments

All NMR experiments were conducted on a Bruker AMX500 spectrometer operating at 500.1 and 202.2 MHz for proton and phosphorus-31 nuclei, respectively, at $26 \pm 1^\circ\text{C}$.

One-dimensional ^{13}P NMR spectra were recorded by employing the inverse gated decoupling technique in order to suppress NOE effects. Typical spectral parameters for quantitative studies using reagent **1** were: 90° pulse width 12.5 μs , sweep width 55 kHz, relaxation delay 25 s, memory size 32 K. To ensure quantitative spectra, the magnitude of the relaxation delay adopted was more than five times the relaxation time ($T_1 = 4.6$ s) of the phosphitylated cyclohexanol; 32 transients were accumulated for each spectrum. For all FIDs, line broadening of 1 Hz was applied and drift correction was performed prior to Fourier transform. Polynomial fourth-order base-line correction was performed before integration.

One-dimensional high-resolution ^1H NMR spectra were acquired with the following acquisition parameters: time domain 32 K; 90° pulse width 9.3 μs ; spectral width 12 ppm; relaxation delay 2 s. 16 scans and 8 dummy scans were accumulated. Base-line correction was performed carefully by applying a polynomial fourth-order function in order to achieve a quantitative evaluation of all signals of interest. The spectra were acquired without spinning the NMR tube in order to avoid artificial signals, such as spinning side bands of the first or higher order.

2.6. Statistical methods

Classification by stepwise forward CDA is used to achieve the most discriminative variables for the arrangement of samples in a space of reduced dimensionality in a way that maximizes the distances between the *a priori* formed groups and the independence of the axes of the configuration [30,31]. The *a priori* groups are ROO, RHZO, and the adulterated refined olive oils (AOO). The chemical compounds contributing most to the discrimination of preformed groups are shown by means of the *F-ratio* as a criterion for inclusion or removal of the compound in the forward stepwise CDA mode. Wilks' λ and the *F* approximation were used to check the significance and estimate the importance of each compound in CDA analysis. The significance of the representation of the original configuration of the samples by the newly formed discriminant axes was expressed by the last value of Wilks' λ . The performance of the method was measured by the percentage of total variance explained by all significant discriminant axes. In a sense, this is a measure of non-linearity and complex interrelations among variables, since CDA is a linear technique. The classificatory efficiency of CDA can be seen in the resulting classification tables which illustrate the group relations of the samples. The ability of the oil samples to be predicted by the rest is another property of CDA, although the percentage of the correctly classified oil samples is a property of the original data rather than an intrinsic feature

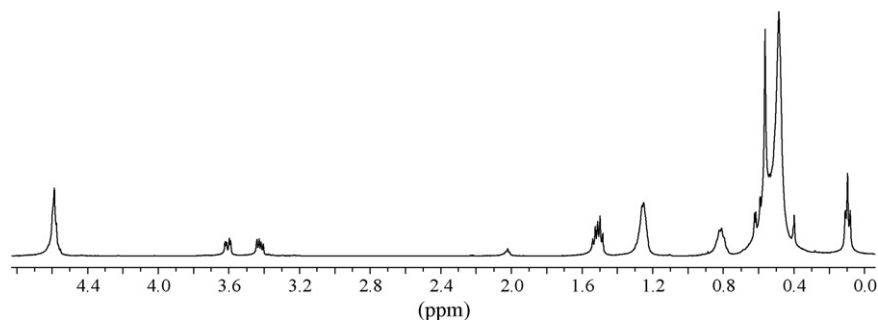


Fig. 1. 500 MHz ^1H NMR spectrum of a ROO sample, from which fatty acids and iodine value can be determined.

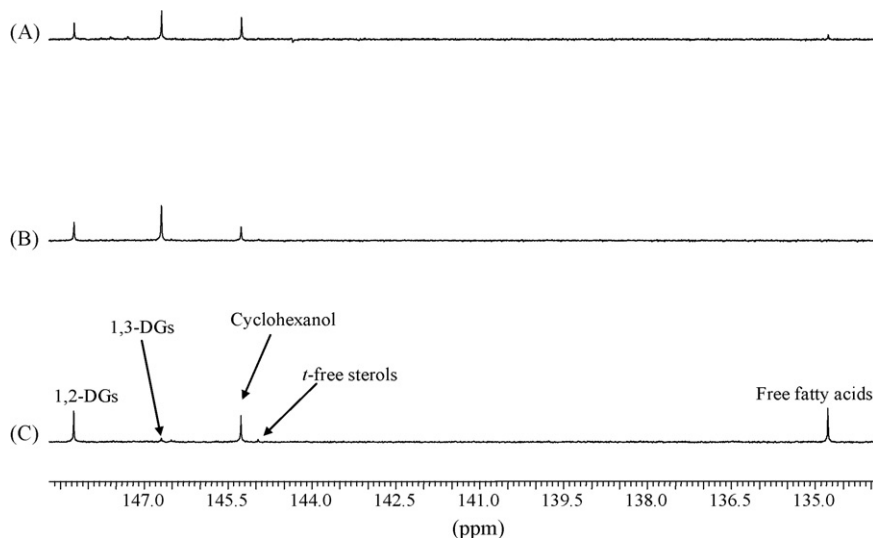


Fig. 2. 200.2 MHz ^{31}P NMR spectra of (A) RHZO, (B) ROO, and (C) virgin olive oil samples from which diacylglycerols, total free sterols and free acidity can be determined.

of the algorithm. Classification is achieved by means of the Mahalanobis distance of each oil sample from the group centroid in order to estimate the probability of misclassification for each one of the predefined groups. Oil samples, which were not engaged in the calculation of classification functions of CDA, are classified in the same

way. This feature of CDA reveals its ability to classify oil samples with known compositional constitution, but unknown group affiliation [32]. In other words, CDA predicts class affiliation by using a linear combination of all compounds engaged in the construction of classification equations.

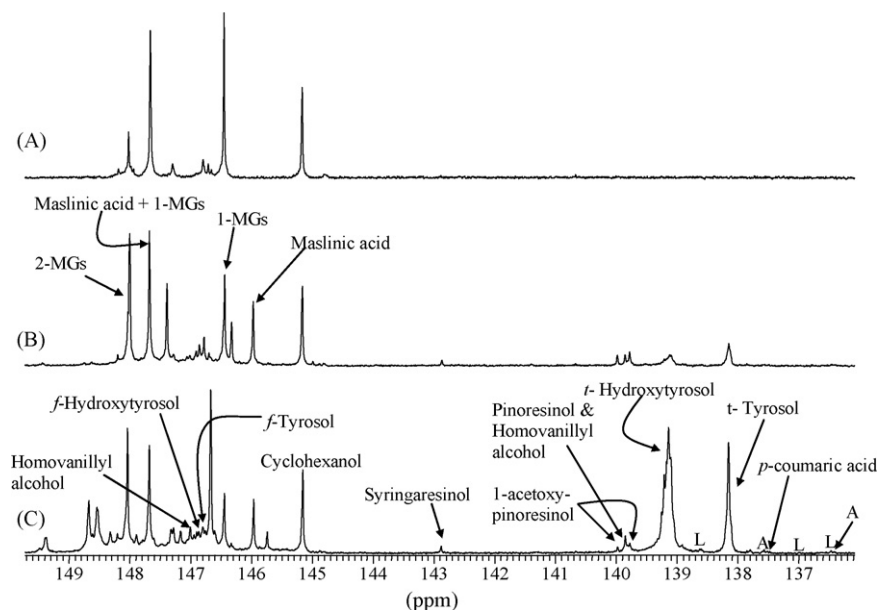


Fig. 3. 200.2 MHz ^{31}P NMR spectra of the polar fraction of (A) RHZO, (B) ROO, and (C) virgin olive oil samples from which phenolic compounds can be determined.

Table 1

Percentage content of the compounds in each group used in CDA and CBT analysis.

	Compound	RHZO		ROO		Adulterated oil samples (AOO)					
		Mean	Stdv	Mean	Stdv	n1		n2		n3	
						Mean	Stdv	Mean	Stdv	Mean	Stdv
1	Syringaresinol	0.00	0.00	0.16	0.23	0.12	0.04	0.05	0.06	0.04	0.06
2*	1-Acetoxy-pinoreosinol [#]	0.00	0.00	0.51	0.71	0.11	0.05	0.10	0.09	0.01	0.01
3	<i>t</i> -Hydroxytyrosol [#]	0.00	0.00	9.33	25.14	0.00	0.00	0.00	0.00	0.00	0.00
4	<i>t</i> -Tyrosol	0.00	0.00	6.28	14.96	0.09	0.08	0.14	0.00	0.01	0.01
5	<i>p</i> -Coumaric acid	0.00	0.00	0.04	0.08	0.04	0.02	0.03	0.03	0.01	0.03
6	<i>f</i> -Tyrosol	0.00	0.00	0.89	1.87	0.60	0.72	0.42	0.31	0.15	0.21
7*	<i>f</i> -Hydroxytyrosol	0.00	0.00	0.03	0.10	0.00	0.00	0.00	0.00	0.00	0.00
8*	Homovanillyl alcohol	0.00	0.00	0.19	0.58	0.00	0.00	0.14	0.25	0.00	0.00
9*	Luteolin [#]	0.00	0.00	0.07	0.09	0.05	0.03	0.07	0.04	0.03	0.05
10	Apigenin [#]	0.00	0.00	0.05	0.06	0.04	0.04	0.04	0.02	0.03	0.04
11	Pinoreosinol [#]	0.00	0.00	0.72	0.78	0.80	0.12	0.30	0.27	0.06	0.09
12	<i>t</i> -Phenols	0.00	0.00	18.28	41.71	1.85	0.88	1.30	0.33	0.34	0.44
13	1,2-DGs [#]	2.03	0.40	3.45	1.10	5.08	0.33	5.06	0.33	3.54	0.19
14*	1,3-DGs	0.81	0.22	1.73	0.35	2.37	0.15	2.39	0.15	1.67	0.14
15	<i>t</i> -DGs	2.85	0.61	5.18	1.40	7.45	0.48	7.45	0.46	5.21	0.32
16*	<i>D</i> (=1,2-DGs/ <i>t</i> -DGs)	0.28	0.03	0.35	0.09	0.32	0.00	0.32	0.01	0.32	0.01
17*	Free acidity	0.07	0.05	0.09	0.19	0.02	0.01	0.07	0.04	0.11	0.03
18*	Sterols [#]	0.03	0.02	0.06	0.02	0.04	0.02	0.05	0.03	0.03	0.02
19	Linolenic acid [#]	0.09	0.06	0.31	0.11	0.37	0.06	0.35	0.06	0.19	0.07
20*	Linoleic acid	10.21	2.27	7.79	1.70	6.59	0.43	8.54	0.89	10.99	0.82
21*	Oleic acid	81.70	4.81	76.21	3.02	75.23	0.63	73.27	0.89	73.47	1.26
22	SFA	7.96	2.87	15.69	2.25	17.80	0.31	17.85	0.14	15.34	1.61
23*	IV [#]	88.23	1.68	79.66	2.24	77.42	0.77	80.02	0.51	84.50	1.82

f = stands for free, *t* = stands for total, DGs = diacylglycerols, *t*-DGs = total DGs, SFA = saturated fatty acids, IV = iodine value, stdv = stands for standard deviation.

* Index compounds that are important in CDA.

Index compounds that are important in CBT.

In contrast to CDA, CBTs is a statistical method capable to handle non-linearity and any possible interactions of compositional parameters, in order to produce a set of simple rules in the form of inequalities, which identify existing samples and with this information find the group affiliation of any new sample [33]; the latter is essential in any adulteration detection system. The construction of a CBT comprises the split of the original set of samples into two parts on the basis of a criterion involving one or two compounds in an algebraic expression. All chemical compounds involved in the construction of a CBT comprise the best diagnostic variable set for the groups already defined prior to the analysis. The groups

at the beginning of a CBT construction are RHZO, ROO, and *n1*, *n2*, *n3* groups of AOO with small (1–10%), medium (12–30%) and high (50–95%) RHZO content, respectively. The original AOO group is split into three subgroups because: (a) in this way the method is capable to reveal connections of the adulterated samples with the original groups of ROO and RHZO, and (b) this is the minimum number of groups that could manifest the “intermediacy principle”, that is *n3* will be closer (clustered together) to RHZO, *n1* to ROO, while *n2* samples will be between them. It should be noted that CDA does not need this split because the final arrangement is expected to depict the picture of intermediacy [26]. The tree grows according to splits that produce maximally informative and ‘pure’, according to an impurity function. The reduction of error in the classification is monitored by means of a loss function. Several loss functions have been proposed [33]; in this study the splitting method has been made through the information theory compatible *Gini-Simpson* index in the form of “towing” a term coined by Breiman et al. [33] on the basis of the better reduction in error they achieved.

All statistical analyses used routines and packages in R system [34] and employed the packages “rpart” [35], and “vegan” [32]. Some algorithms and graphics were applied via SYSTAT (SYSTAT, 1999).

3. Results and discussion

3.1. Data analysis

Fig. 1 shows a typical ¹H NMR spectrum of a ROO sample from which the fatty acid profile and the iodine number can be calculated [16]. This spectrum shows similar characteristics with the proton spectra of RHZO samples (not shown), although it differs in certain signal intensities. Fig. 2 depicts the ³¹P NMR spectra of the phosphitylated refined oils; diacylglycerols, free sterols, and free acidity can be determined from this spectrum [27]. The same figure shows the corresponding spectrum of a virgin olive oil sample. The

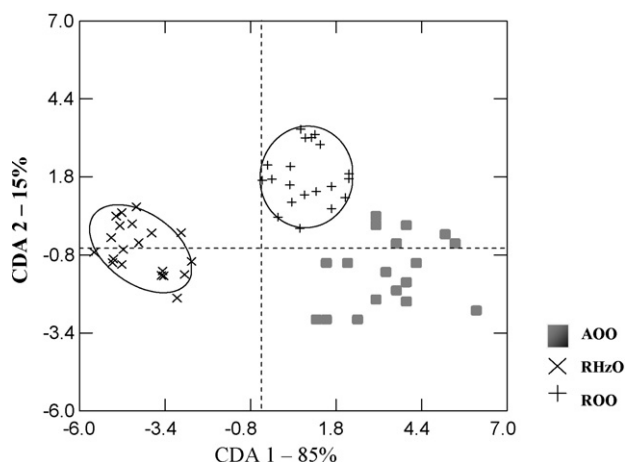


Fig. 4. CDA scattergram of refined olive oil (ROO) (+), refined hazelnut oil (RHZO) (x) samples, and adulterated ROO with RHZO. The latter admixtures denoted as AOO and shown as squares (■) were not taken into account in the construction of the scattergram. The two discriminant axes account for by all variation in the original data set (CDA 1 = 85% and CDA 2 = 15%). Ninety-five percent ellipses are drawn around the ROO and RHZO group centroids. Samples lying outside the group ellipses are correctly classified together with the AOO samples. The crosswire indicates the origin (0, 0) of the scattergram.

Table 2

Canonical discriminant functions of all important variables standardized by within their variances.

No	Chemical compound ^a	Important also in CBT ^b	Canonical function coefficient 1	Canonical function coefficient 2
1	1-Acetoxypinoresinol	E	0.654	0.749
2	<i>f</i> -Hydroxytyrosol		0.359	−0.597
3	Homovanillyl alcohol		−0.945	0.270
4	Luteolin	F	0.985	−0.247
5	1,2-DGs		1.076	−0.332
6	D (=1,2-DGs/ <i>t</i> -DGs)		−0.601	0.408
7	Free acidity		1.195	−1.037
8	Sterols	D	0.191	0.494
9	Linoleic acid		−5.997	3.494
10	Oleic acid		−5.228	2.737
11	IV	A	4.094	−2.938

^a See footnote in Table 1 for the meaning of symbols.^b Symbols at split in CBT (see Table 4).

higher acidity of the virgin olive oil sample as compared to those of ROO and RHZO is reflected on the much stronger signal of the free fatty acids at δ 134.8. Fig. 3 compares the ^{31}P NMR spectra of the polar part of the refined oils recorded after phosphorylation of the hydroxyl groups of their minor compounds. This methodology was used for the detection and quantification of polyphenols in olive oil [36]. Also, Fig. 3 illustrates the ^{31}P NMR spectrum of a virgin olive oil sample. The most characteristic feature of the ^{31}P NMR spectra in Fig. 3 is the complete absence of phenolic compounds in RHZO. Previous studies [37,38] have shown that refining process reduces significantly or even removes completely [39,40] the content of phenolic compounds of olive oil. The mean losses of total phenols of olive oil (virgin and lampante) amount to 75–85%, respectively. This is shown clearly upon comparison of the ^{31}P NMR spectra of a virgin olive oil sample and its refined counterpart (Fig. 3). Hazelnut oil, on the other hand, contains much smaller amounts of phenolic compounds than olive oil [41], and hence the polyphenol content is expected to be practically absent after refinement. The absence of phenolic compounds in RHZO serves as a means to differentiate between RHZO and ROO (see below).

Table 1 lists the mean values of phenolic compounds, fatty acids, diacylglycerols, free sterols, iodine values, and free acidity for both refined oils. The compositional parameters for each oil, including the adulterated oils, are provided as supporting material. The mean values in Table 1 shows that there are discernible differences in fatty acid profiles, especially for linolenic acid and saturated fatty acids. However, this interesting difference alone cannot detect the adulteration of ROO with RHZO at low percentage [41]. Both refined oils are characterized by a higher content of 1,3-DGs than 1,2-DGs, contrary with what is observed for the virgin olive oil samples. This is nicely reflected in Fig. 2 on the respective signal intensities between the refined oils and virgin olive oil. Refinement accelerates the isomerisation of 1,2-DGs to 1,3-DGs, which reaches rapidly its equilibrium state and the ratio D (=1,2-DGs/Total DGs) assumes its equilibrium value (~ 0.33) [42]. These parameters may not be suitable to discriminate between the refined oils. However, measurable differences were observed in the mean values of free sterols and iodine values (Table 1). In particular, ROO samples are characterized by higher sterols content and lower iodine values than the corresponding values of the RHZO samples. The amounts of phenolic compounds as determined by NMR are within the range of phenolic content determined by HPLC [37,38].

3.2. Chemometric approach for the detection of adulteration of refined olive oil

The entire variation of compounds in the original data set within, ROO, RHZO, and AOO groups (Table 1) was fully explained by CDA (Wilks' $\lambda = 0.031$, $F_{\text{approx}} = 19.46$, $df_1 = 22$, $df_2 = 92$, $P < 10^{-4}$). The eigenvalues are $\lambda_1 = 10.35$, and $\lambda_2 = 1.82$ estimating the rate of the

explained variation by the two discriminant axes, i.e. 85% and 15%, respectively. The discriminant scores of the oil samples can be seen in Fig. 4. Both axes have a chemical meaning since their canonical correlations with the original important compounds is high ($r_1 = 0.96$ and $r_2 = 0.80$). The important compounds are variously weighted on the two axes as shown by the standardized coefficients of the canonical functions in Table 2. Linoleic and oleic acids are highly weighted in the negative part of CDA-1 while the iodine value in the positive part. The same compounds are highly weighted with reversed signs on CDA-2. For the linoleic and oleic acids this pattern was expected since they are the products of the same biogenetic pathway [43]. The opposite sign of the iodine coefficients are explained on the basis of the chemical treatment of the oil for the production of ROO and RHZO. The dotted lines in Fig. 4 at the origin (0, 0) of the CDA diagram leave intact only the ROO group. RHZO and the adulterated samples are located at the two ends of CDA-1 but are cut by CDA-2. This indicates that no axis alone is enough to separate adulterated oils from the groups of RHZO and ROO. However, the classification efficacy of CDA is considered to be perfect as shown by the classification matrix in Table 3 (panel a). The percentage classification for the three groups of oils was 100%. The Jackknifed classification (cross-validation) in Table 3 (panel b) indicate that classification was correct for 95% of the samples, and only 5% of the adulterated olive samples cannot be predicted by the rest. Importantly, the adulterated oils are not found in the space between the two groups which comprise the source (RHZO) and the adulteration target (ROO) as a putative intermediate product of mixing [14,16]. The violation of the “intermediacy principle” can be explained on the basis of the non-biological chemical profile of ROO.

Classification of the NMR data into the ‘leaves’ of the binary trees is achieved by starting from an original configuration of five groups. The first two groups are the refined hazelnut oil (RHZO group) and the refined olive oil (ROO group) and in the next three groups are the

Table 3

Classification matrix of the CDA analysis of the NMR data in Table 1.

	Predicted group affiliation			
	RHZO	ROO	AOO	% correct classifications
[a] Actual groups of samples				
RHZO	19	0	0	100
ROO	0	20	0	100
AOO	0	0	20	100
Total classifications	19	20	20	100
[b] Actual groups of samples				
RHZO	17	0	2	89
ROO	0	20	0	100
AOO	1	0	19	95
Total classifications	18	20	21	95

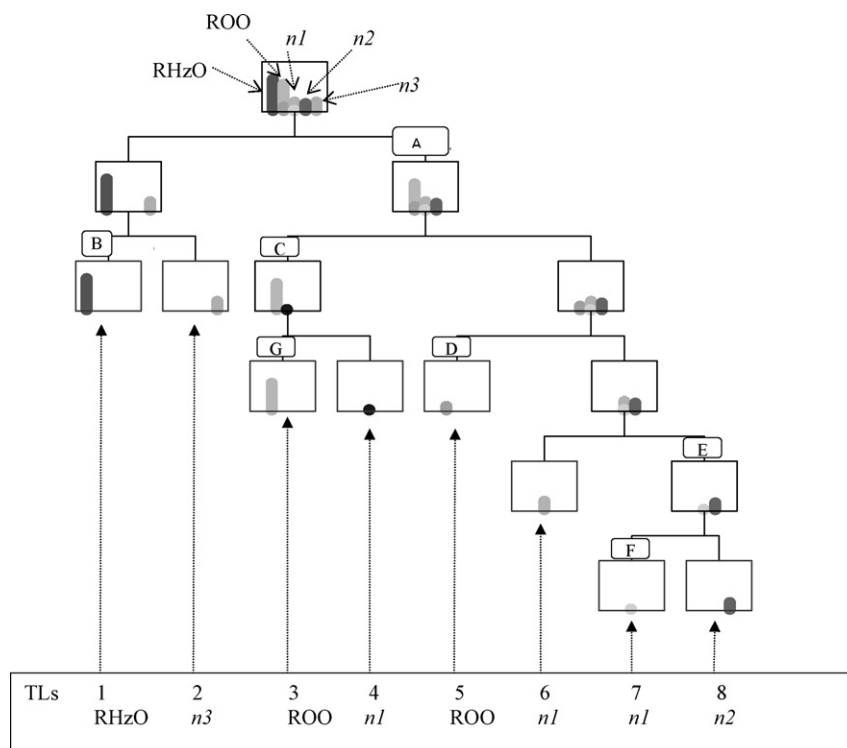


Fig. 5. CBT classification treegram (mobile) of all oil samples. The splitting criteria are those given in Table 4. The original groups are in the root node and in this order they are shown in the entire tree as circles stacked in proportion to the size of each group. The terminal leaves (TLs) at the bottom box correspond to a single group. Eight leaves – indicated by arrows – are required for a complete classification of the samples, since a specific group can form more than one TL.

adulteration levels—i.e. small (1–10%, *n1* group), medium (12–35%, *n2* group) (large 50–95%, *n3* group); ideally five site groups are produced. In the employed model the group identifier is the dependent variable, while the independent variables are the chemical compounds derived by NMR spectroscopy (Table 1). In the construction of the mobile (Fig. 5), the overall reduction in error was 100% and eight terminal leaves were found, each one corresponding to one group (Table 4). It is interesting to note that three compounds, namely *t*-hydroxytyrosol, pinoreosin, and luteolin occur twice in the splitting criteria. The growth of the tree shows that RHZO samples form the first terminal leaf (TL1) early in the process and in a second step the ROO group (TL3 & TL5) (Fig. 5). The adulterated oils spread at different terminal leaves on the basis of different compounds. On the basis of iodine value and pinoreosin (symbol A at split), oils of the *n3* group are clustered together with hazelnut oils but later they separate on the basis of 1,3-DGs and *t*-hydroxytyrosol (symbol B) (Fig. 5). The same pattern is presented with oils of the ROO group. However, here the oil samples at TL5 are separated from adulterated oil samples at another branch of the tree in accordance with the similarity of the refined hazelnut and olive oils in minor constituents. Overall the adulterated oil samples are distributed as follows: *n1* is found at TL4, TL6, and TL7; *n2* is found at TL8 while the group *n3* at TL2 as a result of the proximity to hazelnut oils. Even-

tually, all splitting criteria are used to separate the adulterated oils which form as many terminal leaves as their RHZO dictates. This means that *n1* is found at three terminal leaves while *n2* and *n3* at one leaf (Table 4 and Fig. 5).

With regard to the significant compounds, CDA uses a larger number of compounds (11 in Table 2) than CBT (9 compounds in seven linear combinations or splitting criteria in Table 4). The fatty acids oleic, linoleic and linolenic are of particular importance. With some exceptions, they were found to be significant in the geographical classification of Greek virgin olive oils either by CDA or CBT [31]. In this study, CDA estimates the first two acids as important while CBT the last one. It seems that the selection is justified in terms of the lipoxygenases pathway (LOX pathway) [43]. A classification method needs only one acid since both linoleic and linolenic acids are the substrates for LOX. In the same line of argumentation oleic is converted to linoleic acid in a bioreaction catalyzed by enzymes encoded by two widespread Fatty Acid Desaturases (FAD) genes. In CDA both oleic and linoleic are significant compounds (Table 2) and have very similar canonical correlations. The pure classification CBT technique does not need both acids. Instead, linoleic acid is enough for the entire classification to be described in CBT (Table 4). This is in agreement with the fact that the refinement process does not disturb the relative proportions of fatty acids.

Table 4
Significant compounds and linear combinations of two compounds expressed as improvement of CBT classification.

Symbol at split	Compounds ^a	Criterion	Fraction of predicted samples	Significance of compounds
A	IV-Pinoreosin	<83.24	0.367	0.367
B	1,3-DGs + <i>t</i> -hydroxytyrosol	<2.96	0.580	0.213
C	Luteolin- <i>t</i> -hydroxytyrosol	<0.03	0.738	0.158
D	Linoleic acid-Sterols	<0.15	0.822	0.084
E	Pinoreosin-1-acetoxypinoreosin	<0.59	0.918	0.096
F	Luteolin	<0.04	0.956	0.038
G	Apigenin	<0.19	1.000	0.044

^a See footnote in Table 1 for the meaning of symbols.

For both methods, diacylglycerols appear to be significant compounds. 1,2-DGs is important for CDA, but with a different behavior of canonical coefficient than the ratio *D*. In CBT method 1,3-DGs are found significant. These compounds have lower variation within oil samples and vary substantially among the original groups. It is strange that *t*-DGs are not important in both methods presumably because of the similarity that exists in the lipid biosynthetic pathway that leads from diacylglycerols to triacylglycerols [44].

Phenols occur in ROO group though at reduced amounts in comparison to virgin olive oil but they are absent from RH₂O as a result of chemical treatment of refinement, in spite the fact that phenols have been reported from hazel tree leaves and several species of hazel family (Angiospermae: Coryllaceae). Almost every split in the CBT tree (Fig. 5) contains one or two phenols. Exception is the split D which is the difference between linolenic acid and sterols. Amongst phenols, *t*-hydroxytyrosol, luteolin, and 1-acetoxypinoresinol are important in both CDA and CBT analyses, whereas homovanillic alcohol, apigenin and pinoresinol are important in CDA and CBT classification, respectively. Syringaresinol, *t*-tyrosol and *f*-tyrosol, *p*-coumaric acid and total phenols are not significant compounds. They have high standard deviation within groups (Table 1). Possibly, the large variation of these phenolic compounds amongst the various samples caused their exclusion from the important compound set in both classifications together with the fact that they are compounds of the same phenol biosynthetic path.

Phytosterols are also important compounds in both CDA and CBT analyses. In nature they are distributed to many species in small quantities less than 1% and usually much less. They are important nutrients for all eucaryotes; insects incapable in synthesizing them acquire the necessary amounts from endosymbiotic and in some case exobiotic bacteria from plant hosts, which supply sterols and in some cases vitamins. In CBT sterols appear in the splitting criterion that separates ROO oils from the adulterated ones in groups *n1* and *n2*.

4. Conclusions

¹H and ¹³P NMR spectroscopy in conjunction with multivariate statistical analyses (CDA and CBTs) has been used with respect to its ability to discriminate between ROO and RH₂O and to detect ROO fraud with RH₂O. Overall, both statistical analyses showed perfect classification that is all oil samples were grouped correctly in the original group. Both analyses used different significant compounds, although in the same class and a more or less equal number of them. CBTs gave a more detailed and biologically sound hierarchical splitting while it started from more groups than CDA. This is the result of (a) non-predefined strategy, (b) its capability in treating non-linear and/or discontinuous data structures, and (c) its orientation towards hierarchical classification unlike CDA which keeps a balance between classification and ordination in a fixed number of groups [45]. RH₂O was the source group of this adulteration and ROO was the group which was adulterated, and these groups cannot be altered in CDA. The reported CBTs classification is superior to other methods combining linear algorithms as principal components and simple non-linear techniques (e.g. artificial neural networks) that need a scarce knowledge of the engaged compounds and for this are unable to substantiate their biological basis.

Acknowledgments

We are indebted to the program Regional Pole of Innovation for Value Creation funded by the administration of the Region of Crete and the European Union.

References

- [1] G.P. Blanch, M. Caja, M.L. Riuz del Castillo, J. Agric. Food Chem. 46 (1998) 5153–5157.
- [2] M.T. Morales, G. Luna, R. Aparicio, Food Chem. 91 (2004) 293–301.
- [3] G. Flores, M.L. del Castillo, M. Herraiz, G.P. Blanch, Food Chem. 97 (2006) 742–749.
- [4] G. Flores, M.L. del Castillo, G.P. Blanch, M. Herraiz, Food Chem. 97 (2006) 336–342.
- [5] A. Sayago, M.T. Morales, R. Aparicio, Eur. Food Res. Technol. 218 (2004) 480–483.
- [6] C. Mariani, G. Bellan, G. Morchio, A. Pellegrino, Riv. Ital. Sostanze Grasse 76 (1999) 59–67.
- [7] F. Peña, S. Cárdenas, M. Gallego, M. Velcárcel, J. Chromatogr. A 1074 (2005) 215–221.
- [8] S. Milder-Szkudlarz, H.H. Jeleń, Food Chem. 110 (2008) 751–761.
- [9] M.H. Gordon, C. Covell, N. Kirsch, J. Am. Oil. Chem. Soc. 78 (2001) 621–624.
- [10] D. Zabarar, M.H. Gordon, Food Chem. 84 (2004) 475–483.
- [11] S. Vichi, L. Pizzale, E. Toffano, R. Bartolomeazzi, L. Conte, J. AOAC Int. 84 (2001) 1534–1541.
- [12] J. Parcerisa, I. Casals, J. Boatella, R. Codony, M. Rafecas, J. Chromatogr. A 881 (2000) 149–158.
- [13] L. Cercasi, M.T. Rodriguez-Estrada, G. Lercker, J. Chromatogr. A 985 (2003) 211–220.
- [14] C. Faulh, F. Reniero, C. Guillou, Magn. Reson. Chem. 38 (2000) 436–443.
- [15] L. Mannina, M. Patumi, P. Fiordiponti, M.C. Emanuele, A.L. Segre, Ital. J. Food Sci. 11 (1999) 139–149.
- [16] G. Vigli, A. Philippidis, A. Spyros, P. Dais, J. Agric. Food Chem. 51 (2003) 5715–5722.
- [17] D. Šmejkalova, A. Piccolo, Food Chem. 118 (2010) 153–158.
- [18] B.F. Ozen, L.J. Mauer, J. Agric. Food Chem. 50 (2002) 3898–3901.
- [19] E.C. López-Diez, G. Bianchi, R. Goodacre, J. Agric. Food Chem. 51 (2003) 6145–6150.
- [20] L. Diego, G. González, L. Mannina, M. D' Imperio, A.L. Segre, R. Aparicio, Eur. Food Res. Technol. 219 (2004) 545–548.
- [21] M.J. Lerma-García, G. Ramis-Ramos, J.M. Herrero-Martínez, E.F. Simó-Alfonso, Food Chem. 118 (2010) 78–83.
- [22] V. Baeten, J.A.F. Pierna, P. Dardienne, M. Meurens, D.L. García-González, R. Aparicio-Ruiz, J. Agric. Food Chem. 53 (2005) 6201–6206.
- [23] A. Sayago, D.L. García-González, M.T. Morales, R. Aparicio, J. Agric. Food Chem. 55 (2007) 2068–2071.
- [24] T. Mavromoustakos, M. Zervou, G. Bonas, A. Kolocouris, P. Petrakis, J. Am. Oil Chem. Soc. 77 (2000) 405–411.
- [25] J. Vlahov, Magn. Reson. Chem. 35 (1997) S8–S12.
- [26] G. Fragaki, A. Spyros, G. Siragakis, E. Salivaras, P. Dais, J. Agric. Food Chem. 53 (2005) 2810–2816.
- [27] A. Spyros, P. Dais, J. Agric. Food Chem. 48 (2000) 802–805.
- [28] A. Zwierzak, Can. J. Chem. 45 (1967) 2501–2512.
- [29] G. Montedoro, M. Servili, M. Baldioli, R. Selvaggini, E. Miniati, J. Agric. Food Chem. 40 (1992) 1571–1576.
- [30] R.A. Johnson, D.M. Wichern, Applied Multivariate Statistical Analysis, Prentice Hall, New Jersey, USA, 1998, p. 820.
- [31] P.V. Petrakis, A. Agiomyrgianaki, S. Christophoridou, A. Spyros, P. Dais, J. Agric. Food Chem. 14 (2008) 3200–3207.
- [32] J. Oksanen, R. Kindt, P. Legendre, R. O'Hara, G.L. Simpson, M.H.H. Stevens, H. Wagner, Vegan: Community Ecology Package, R package version 1.13-2. <http://vegan.r-forge.r-project.org/> (accessed 17.05.2008).
- [33] L. Breiman, J. Friedman, R. Olshen, C. Stone, Classification Regression Trees, Wadsworth, California, USA, 1984.
- [34] R Development Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2008.
- [35] T.M. Therneau, B. Atkinson, rpart: Recursive Partitioning, R package version 3.1-41, R port by Brian Ripley, 2008.
- [36] S. Christophoridou, P. Dais, J. Agric. Food Chem. 54 (2006) 656–664.
- [37] R.W. Owen, W. Mier, A. Giacosa, W.E. Hull, B. Spiegelhalter, H. Bartsch, Food Chem. Toxicol. 38 (2000) 647–659.
- [38] C. Nergiz, Int. J. Food Sci. Technol. 28 (1993) 461–464.
- [39] N. Cortesi, A. Ponzini, E. Fedeli, Riv. Ital. Sostanze Grasse 58 (1981) 108–114.
- [40] A. García, M.V. Ruiz-Méndez, C. Romero, M. Brenes, J. Am. Oil Chem. Soc. 83 (2006) 159–164.
- [41] S. Arranz, R. Cert, J. Perez-Jiménez, A. Cert, F. Saura-Calixto, Food Chem. 110 (2008) 985–990.
- [42] A. Spyros, A. Philippidis, P. Dais, J. Agric. Food Chem. 52 (2004) 157–162.
- [43] D.F. Hildebrand, Physiol. Plant 76 (1989) 249–253.
- [44] R.E. Minto, B.J. Blacklock, Progr. Lipid Res. 47 (2008) 233–306.
- [45] R.A. Pimentel, in: J.T. Sorensen, R. Footitt (Eds.), An Introduction to Ordination, Principal Component Analysis and Discriminant Analysis, Elsevier, Amsterdam, 1992, pp. 11–28.