Giorgio Bianchi[a], Lucia Giansante[a], Adrian Shaw[b], Douglas B. Kell[b]

[a] Istituto Sperimentale per la Elaiotecnica, Contrada Fonte Umano, Città Sant'Angelo, Pescara, Italy
[b] Institute of Biological Sciences, University of Wales, Aberystwyth, United Kingdom

Chemometric criteria for the characterisation of Italian Protected Denomination of Origin (DOP) olive oils from their metabolic profiles

This study was conducted and carried out as a consequence of the European directives n°2081/92 and n°2037/93 in which regulations for the protection of denomination of origin of food commodities were established. The research work was intended to investigate if monovarietal oils may be differentiated by their basic chemical composition. A positive outcome of this pilot study would hopefully permit extension to the characterisation of Italian DOP (Protected Denomination of Origin) olive oils.

In the olive, long-chain alcohols, triterpenes and fatty acids, formed in distinct biosynthetic compartments, provide characteristic compositional data of an olive cultivar. The three classes of compounds were qualitatively and quantitatively determined by GC in six Italian olive cultivars, *Coratina* and *Provenzale* from Puglia, *Frantoio* and *Moraiolo* from Toscana, *Bosana* from Sardegna and *Dritta* from Abruzzo. Basic statistics and multivariate methods were first applied to the GC data of each class of compounds and then to the complete set of data with the aim of obtaining a clear discrimination of the cultivars.

When Principal Components Analysis (PCA) was applied to the whole set of data consisting of alcohols, triterpenes and acids, the PCA revealed that the compounds (variables) that gave the better class distinction were: cycloartenol for *Coratina*, acids C20:0, C17:0, C18:0 for *Dritta*, citrostadienol for *Frantoio* and β-sitosterol for *Moraiolo*. *Bosana* and *Provenzale* correlated with erythrodiol and uvaol. A correct assignment of each oil sample to its monovarietal group was obtained.

Keywords: Monovarietal olive oils, chemometrics, olive metabolites.

1 Introduction

The geographical origin and the specific olive cultivar from which oil is obtained are the major official characterising elements of Italian DOP (Protected Denomination of Origin) olive oils. DOP olive oils, on the market since the olive harvest of 1998-1999, are considered the best among Extra Virgin Olive Oils because of their expected authenticity and specified organoleptic characteristics. Because of their high commercial value there is a great temptation for fraud by marketing non-authentic or adulterated DOP oils. Moreover, the European directives n° 2037/93 and n° 2081/92 stress the need to regulate and protect the denomination of origin of food commodities [1, 2].

Both consumers and DOP oil producers are concerned about the present lack of suitable, reliable analytical methods to determine the authenticity of this most prized commodity. The official EU analytical methods are useless as a means to protect DOP oil [3], since they merely allow for the determination of the grade of an olive oil, and whether or not it is adulterated with lower grade seed oil.

Since the early nineties, many research institutions devoted to the study of vegetable oils, especially olive oil, have explored and applied various methods potentially capable of characterising DOP oils for their authenticity; almost all these studies were based on the application of chemometrics [4, 5]. Chemometric techniques are especially suitable for handling the large amounts of data produced by modern analytical methods, such as nuclear magnetic resonance, infrared and *Raman* spectroscopies, mass and pyrolysis mass spectrometry, gas and liquid chromatographies [6]. Chemometrics proved to be a powerful tool for answering complex questions such as what is the geographical origin of a wine, and it is generally believed to be applicable to other food products [7, 8].

Commonly, the data analysed were general chemical composition, chemico-physical parameters and also agronomical and environmental data. The most common chemometric methods applied to the study were ex-

Correspondence: *Giorgio Bianchi*, Istituto Sperimentale per la Elaiotecnica, Contrada Fonte Umano, 65013 Città Sant'Angelo, Pescara, Italy. Phone: +39-085-95294, Fax: +39-085-959518; e-mail: elaiotec@unich.it

 0931-5985/2001/0303-0141 $17.50+.50/0

ploratory methods of pattern recognition, such as Principal Components Analysis (PCA), Hierarchical Cluster Analysis (HCA), and classification methods, such as Soft Independent Modelling of Class Analogy (SIMCA), Partial Least Square (PLS), Linear Discriminant Analysis (LDA) and Artificial Neural Networks (ANNs) [9, 10]. Additionally, a specific expert system SEXIA (a Spanish acronym for the Expert System to Identify Oils), has been designed to achieve the characterisation of European virgin olive oils on the basis of fifty-five chemical compounds [11, 12]. For supervised learning methods such as PLS and ANNs, the data-splitting program Multiplex has been used to sort the pyrolysis and NMR spectra of olive oils before analysis/prediction [13, 14].

Parametric and non-parametric methods have been applied to the complex chemical composition of oils from various Mediterranean countries, obtaining a satisfactory classification of oil samples according to their geographical origin [10, 15–17]. In further chemometric work, the composition of fatty acids has been used to compare the ability for discrimination of two different artificial neural network methods on oils of different origin.

Besides fatty acids, sterolic fraction composition has also been used in chemometric studies aimed at the characterisation of the geographic origin of extra virgin olive oil [18]. Fatty acid and sterol composition have been shown to have a significant power to discriminate even oils produced within close geographical proximity, even harvested at different maturity stages [19–21]. It has been found that sterols, triterpenic alcohols and hydrocarbons undergo changes in relation to orography (climate, altitude, distance fron sea etc.): by means of discriminant and regression analysis procedures the oils were classified into various groups, such as valley and mountain oils, respectively [22, 23]. In addition, it has been reported that chemical variables such as Δ^7-stigmasterol, cycloartenol, stigmasterol and other minor polar compounds are not affected significantly over a period of years [24, 25].

In this paper we report on the results obtained in an effort to identify monovarietal oils (also called elemental oils) obtained from some common Italian olive cultivars. The major aim of this study was to explore and evaluate further the potential and reliability of chemometric studies on the compositional variables of oil with the eventual aim of correctly characterising Italian DOP oils. To this end it was decided to use as characteristic variables the same analytical parameters as those of the EC official methods of analysis, namely (i) fatty acids, (ii) tetracyclic as well as pentacyclic triterpenoids and (iii) long-chain aliphatic alcohols.

The study consisted of a prior descriptive analysis, comparing cultivars for every variable, followed by bivariate and multivariate analysis of the three classes of compounds (i), (ii) and (iii) one at a time and of the complete set of data (i) + (ii) + (iii), all together in a run.

The three classes (i), (ii) and (iii) are biosynthesised according to independent and genetically controlled pathways [26, 27]. Consequently, the genetic makeup of an olive cultivar determines the biosynthetic reaction sequences leading to the lipid classes of compounds comprising the oil. As a consequence of the genetic makeup, a given product of the pathway considered may be present in variable concentrations or even absent in extreme cases. There is also the possibility that an intermediate in a reaction step may accumulate at the expense of either the precursor or the final product [28]. The rational basis for this assumption lies in the fact that in a cultivar there may be a complete or partial block on a particular step in a metabolic pathway through a modulation of the action of a specific enzyme.

2 Materials and methods

2.1 Origin of olives and production of the oil

The data set included numerous samples of monovarietal oils obtained from the following olive cultivars grown in various Italian regions: Abruzzo, *cv Dritta*, 17 samples; Puglia, *cv Coratina*, 12 samples; *cv Provenzale*, 6 samples; Sardegna, *cv Bosana*, 9 samples; Toscana, *cv Frantoio*, 12 samples; *cv Moraiolo*, 7 samples. Olives were harvested in the period November – December of the 1996 season. The numerous representatives of the cultivars examined were collected at well-separated sites covering a large area of each geographical region under examination. Olive batches of 3-5 kg were crushed in a small hammer metal mill at the Pescara Institute. The olive paste underwent malaxation for 20 min at 28 °C. The oil was separated by a vertical centrifuge with gravitational discharge through the base. The yields were in a wide range, 6-10%.

2.2 Determination of unsaponifiables

Long-chain fatty alcohols and polycyclic triterpenes were obtained in a one-step reaction. The determination of the composition of the unsaponifiable fraction, without prior TLC separation, into classes of compounds was performed by gas chromatography with a polar column of high thermal stability as reported in the literature [29, 30]. In the gas chromatographic analysis of long-chain fatty alcohols and cyclic triterpenes only the major components were identified and quantitatively evaluated (Fig. 1).

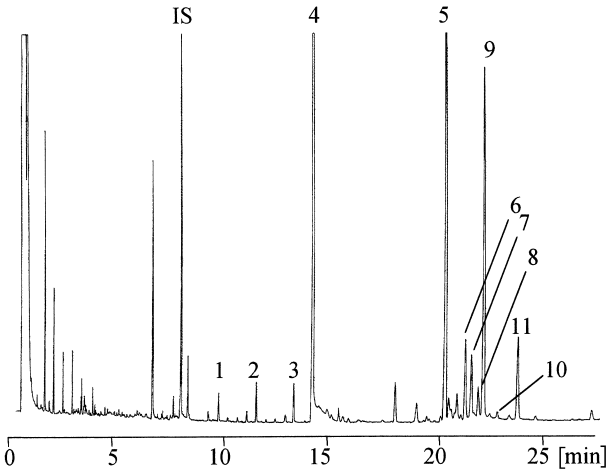

Fig. 1. Gas chromatographic trace of the unsaponifiable fraction of extra virgin olive oil. IS – internal standard (eicosanol); 1 – docosanol; 2 – tetracosanol; 3 – hexacosanol; 4 – squalene; 5 – β-sitosterol; 6 – Δ^5-avenasterol; 7 – cycloartenol; 8 – erythrodiol; 9 – 24-methylencycloartanol; 10 – uvaol; 11 – citrostadienol.

Saponification was performed on 1.25 g oil samples to which a known amount of 1-eicosanol, 0.1 g of standard dissolved in 200 ml of $CHCl_3$ (*Sigma Aldrich*, Deisenhofen, Germany), was added as internal standard in 25 ml of an ethanolic KOH solution. The reaction flask, fitted with a reflux condenser, was kept at a gentle simmer, for 15–20 min, until the solution was clear. A further 5 ml of ethanol were added to the solution, and shaken to homogenise. The recovery of the unsaponifiable matter was performed by solid-phase extraction (SPE) with aluminium oxide neutral cartridges (*Supelco* Inc., Bellefonte, USA). A 0.5 ml aliquot of sample solution was pipetted onto the column head, and allowed to pass completely into the column; the unsaponifiable matter was eluted first with 0.5 ml of ethanol then with 6 ml of diethyl ether. The solvent of the eluted solutions was removed by rotary evaporation. The samples obtained following this procedure, free from any organic potassium salt, were derivatised immediately for GC analysis as trimethylsilyl (TMS) derivatives. The silylating reagent was a 9:3:1 mixture of pyridine, hexamethyldisilazane and chlorotrimethylsilane. The derivatization was carried out by reacting 100 µl of silylating reagent with 10 mg of the dried sample for 30 min at room temperature. The reaction solution was evaporated to dryness under a N_2 stream, then dissolved in 1 ml iso-octane. The qualitative and quantitative composition of the whole unsaponifiable fraction was determined in a HRGC Model 5160 Mega gas chromatograph (*Carlo Erba*, Milano, Italy). Samples (0.1 µl) were injected onto the column. The instrument was equipped with a fused silica capillary column (25 m x 0.32 mm x 0.10 µm), with a 50% phenyl/50% methylpolysiloxane stationary phase CP-TAP-CB (*Chrompack* Inc., Bridgewater, NJ, USA). The carrier gas was hydrogen, inlet pressure 70 kPa at a flow rate of 2.3 ml/min. The temperature of the flame-ionisation detector (FID) was 330 °C. The oven temperature was programmed from 70 to 160 °C at 30 °C/min, 0 min at 160 °C, from 160 to 240 °C at 8 °C/min, 2 min at 240 °C, from 240 to 300 °C at 3 °C/min, 10 min at 300 °C. Peak identification was carried out by comparison of the relative retention time with those reported in the literature [29, 30] and with retention time of standards (*Sigma Aldrich*, Deisenhofen).

The long-chain alcohols docosanol (C22), tetracosanol (C24), hexacosanol (C26) and the polycyclic triterpenes β-sitosterol, Δ^5-avenasterol, cycloartenol, 24-methylencycloartenol, erythrodiol, uvaol and citrostadienol were expressed as area percentage of the two classes of the compounds long-chain alcohols and triterpenes, respectively.

2.3 Determination of fatty acid methyl esters

Oil fatty acid analysis was carried out on methyl esters prepared according to the EU method [3]. Gas chromatographic analysis and peak area estimation were carried out using a HRGC gas chromatograph (*Carlo Erba*, Milano, Italy) equipped with a fused silica capillary column SP 2380 (60 m x 0.32 mm x 0.20 µm, *Supelco* Inc., Bellefonte, USA) with hydrogen as the carrier gas, inlet pressure of 50 kPa; samples were injected onto the column. The temperature program was from 120 to 165 °C at 30 °C/min, 25 min at 165 °C, from 165 to 200 °C at 5 °C/min, 15 min at 200 °C, from 200 to 230 °C at 5 °C/min, 10 min at 230 °C, FID temperature at 230 °C. The identification of peaks was performed by comparing the peak retention time with those of methyl ester standards.

The peaks for all fatty acids methyl esters were well resolved, with the exception of that of vaccenic acid (11c – 18:1) which was only partially separated from that of oleic acid under the chromatographic conditions used. Ten fatty acids, palmitic C16:0, palmitoleic C16:1, heptadecanoic C17:0, heptadecenoic C17:1, stearic C18:0, oleic C18:1, linoleic C18:2, arachidic C20:0, linolenic C18:3 and eicosenoic C20:1, expressed as percentage of fatty acid methyl ester fraction, were used in this work.

2.4 Data processing

General descriptive statistics were processed in Microsoft® Excel 97, while for multivariate statistics the S-Plus® 2000 Professional statistical package (MathSoft) was used.

3 Results and discussion

The raw data generated from the three classes of compounds were interpreted by using summarised descriptive statistics and arranged in Tab. 1. One way Analysis of Variance (ANOVA) was applied to test which variables contributed most to differentiate olive oil cultivars. Subsequently, bivariate statistics were applied, thus determining the effect that a given variable exerts on another by regression analysis. Principal Components Analysis was applied sequentially on the three classes of compounds one at a time and on all variables as a whole.

3.1 Compositional trends of fatty acids, alcohols and polycyclic triterpenes of monovarietal oils

Tab. 1 shows the class relevant compositional data in percentage derived for each monovarietal oil sample. Examination of the data in Tab. 1 reveals large differences and clear similarities among the six monovarietal oils. The extent of the variations observed depends on the class of compounds considered.

Fatty acids: All the samples show chemical compositions that are well in the average literature value ranges [12]. The data set reveals rather broad ranges for C18:1, C18:2 and C16:0 in that order. The most evident variations in the fatty acids are observed for C18:1, *Bosana* oils showing the largest range, about 67-75, whilst *Coratina* and *Moraiolo* oils vary least, about 79-81. The latter two varieties of oil are quite similar on the basis of this acid. It is noteworthy that *Bosana* oil is characterised by the lowest percentage of C18:1, and contains the largest percentage of C18:2, C16:0 and C16:1. For a given variety, significant variations in C18:2 and C16:0 were found; the remaining acids, however, are present in comparatively constant percentages.

Triterpenes: Tab. 1 shows the large variability in composition of olive oil triterpenes. Although the data reported in the table are within the reported literature ranges [12], some features are worth noting: e.g., β-sitosterol is the major sterol except in the case of Dritta, which is characterised by an unusually high content of 24-methylencycloartenol. The latter component, although close to cycloartenol in the biosynthetic pathway, appears to correlate with its homologue to a limited extent only. A closer examination of the triterpene data reveals other associations and differentiations. It can be seen that β-sitosterol brings about three groups with two associations (i) *Coratina-Frantoio-Moraiolo*, (ii) *Bosana-Provenzale* and (iii) the *Dritta* oils alone. With Δ^5-avenasterol three groups are observed: (i) Bosana-Frantoio-Moraiolo-Provenzale, (ii) *Dritta*, (iii) *Coratina*.

Long-chain fatty alcohols: The three C22, C24, C26 alcohols show a compositional pattern in fair agreement with their biosynthetic scheme that has been suggested to proceed through an elongation-reduction in which C2 units are incorporated into the precursor, e.g. the C18 chain, that in succession leads to the homologues C22–C26 [26, 27]. All six cultivars studied show typical genetic control on the formation of these substances. The three homologues are consistently present at a considerable percentage. A detailed numerical examination of the data reveals that all the three homologues vary to a great extent; in most cases C22 and C26 vary more than twofold in their own ranges. Such large variations, not found in the other two classes of compounds, might be due to the fact that alcohols are derived from the epicuticular waxes present on the cuticle of the olive fruit. The latter substances are known to be particularly affected by environmental stress, and are consequently more susceptible to variations in chemical composition.

A common general observation can be drawn from the examination of the whole set of compositional data: variations observed within monovarietal oil groups are smaller than variations between oils from different cultivars. Since almost all fatty acids, alcohols and triterpenoids bring about different associations or differentiations, it seems likely that suitable multivariate methods may be more appropriate in characterising the oil varieties.

3.2 Chemometrics

Early in this research work, we sought to find simple criteria based on oil components which would allow us to distinguish each oil variety from the others. First, an exploratory study was carried out for each variable individually to test the differences between cultivars, with one-way ANOVA. The significance was estimated comparing $F_{observed}$ with F_{critic} for *(g-1)* and *(N-g)* degrees of freedom, where $g = 6$ is the number of groups and $N = 63$ is the total number of samples. Tab. 1 shows F_{values} for each variable; it can be seen that all variables show an $F_{observed}$ bigger than F_{critic}; this means that all variables show a highly significant difference between cultivars.

The second criterion considered was to compare quantitatively the homologues within each class of compound, that is, the substances which proved to be biosynthetically correlated. [26, 27]. The data shown in Tab. 1 prove the strict biosynthetic relationship among the three homologous alcohols. In fact, considering the extremes of the homologous chain C22–C24–C26, one observes how an increase of the final homologous alcohol C26 corresponds to a low percentage value of the starter C22 (*Coratina*) and *vice versa* (*Provenzale*). The homologous alcohols C22, C24 and C26 and the pair of fatty acids, palmitic

Tab. 1. Summary of compositional data and their Analysis of Variance.

Classes of compounds (variables)	Bosana			Coratina			Dritta			Frantoio			Moraiolo			Provenzale			ANOVA**
	mean*	min	max	mean*	min	max	mean*	min	max	mean*	min	max	mean*	min	max	mean*	min	max	$F_{observed}$
Acids																			
Myristic	***0.01***	0.00	0.01	***0.00***	0.00	0.01	***0.01***	0.00	0.05	***0.00***	0.00	0.01	***0.00***	0.00	0.01	***0.00***	0.00	0.01	**4.27**
Palmitic	***12.41***	11.20	14.30	***9.11***	8.36	9.61	***11.31***	9.92	12.21	***10.08***	9.45	11.64	***10.67***	10.24	11.24	***10.85***	10.46	11.57	**33.17**
Palmitoleic	***0.96***	0.77	1.39	***0.37***	0.31	0.41	***0.87***	0.72	1.12	***0.64***	0.53	1.02	***0.60***	0.53	0.75	***0.72***	0.61	0.88	**35.23**
Heptadecanoic	***0.04***	0.03	0.05	***0.04***	0.04	0.05	***0.05***	0.04	0.06	***0.04***	0.04	0.05	***0.04***	0.03	0.04	***0.04***	0.03	0.04	**13.9**
Heptadecenoic	***0.09***	0.06	0.12	***0.06***	0.04	0.08	***0.08***	0.07	0.10	***0.10***	0.08	0.13	***0.08***	0.06	0.09	***0.09***	0.06	0.13	**10.76**
Stearic	***2.01***	1.77	2.26	***2.09***	1.79	2.46	***2.48***	2.14	2.90	***1.62***	1.47	1.94	***1.66***	1.57	1.75	***2.01***	1.74	2.36	**38.46**
Oleic	***72.32***	67.09	75.67	***80.79***	79.80	81.74	***77.08***	73.64	79.13	***80.70***	77.31	82.23	***80.08***	79.26	81.47	***75.99***	72.97	78.12	**36.53**
Linoleic	***10.49***	7.58	14.25	***5.90***	4.93	7.74	***6.75***	4.93	8.86	***5.38***	4.64	7.45	***5.50***	4.55	6.32	***8.84***	7.23	11.05	**24.77**
Linolenic	***0.83***	0.71	1.01	***0.71***	0.63	0.83	***0.58***	0.50	0.64	***0.62***	0.50	0.74	***0.67***	0.58	0.88	***0.65***	0.54	0.81	**13.78**
Arachidic	***0.36***	0.34	0.41	***0.37***	0.31	0.43	***0.38***	0.32	0.45	***0.31***	0.28	0.33	***0.29***	0.27	0.32	***0.35***	0.32	0.37	**18.5**
Eicosenoic	***0.33***	0.29	0.37	***0.45***	0.37	0.48	***0.26***	0.22	0.30	***0.36***	0.29	0.39	***0.32***	0.28	0.39	***0.33***	0.29	0.38	**59.38**
Behenic	***0.12***	0.09	0.15	***0.10***	0.08	0.13	***0.10***	0.07	0.13	***0.10***	0.09	0.12	***0.10***	0.08	0.12	***0.12***	0.10	0.15	**5.19**
Lignoceric	***0.03***	0.00	0.05	***0.00***	0.00	0.01	***0.03***	0.00	0.09	***0.03***	0.00	0.05	***0.01***	0.00	0.04	***0.00***	0.00	0.00	**11.27**
Alcohols																			
Docosanol	***23.08***	16.73	30.01	***7.65***	2.26	17.93	***11.00***	6.54	23.35	***14.69***	6.03	25.50	***16.05***	11.70	20.01	***25.65***	21.82	33.11	**17.88**
Tetracosanol	***38.48***	31.26	43.10	***23.20***	18.95	32.24	***26.23***	19.22	35.24	***29.42***	20.28	40.23	***33.72***	24.91	39.68	***37.52***	34.80	39.49	**16.14**
Hexacosanol	***38.45***	26.89	52.00	***69.15***	49.83	78.79	***62.77***	47.46	73.42	***55.88***	34.27	73.69	***50.22***	42.18	63.39	***36.84***	28.06	43.21	**21.75**
Triterpenes																			
β-Sitosterol	***42.69***	33.00	59.60	***50.49***	31.05	62.06	***33.47***	26.67	62.08	***52.51***	42.89	62.69	***52.80***	44.68	60.40	***43.21***	31.10	52.68	**11.58**
Δ^5-Avenasterol	***8.83***	6.91	11.01	***4.43***	3.39	6.38	***12.00***	5.98	14.22	***9.04***	7.24	11.76	***7.55***	5.43	9.23	***7.33***	4.67	8.50	**12.48**
Cycloartenol	***11.15***	5.89	16.13	***28.08***	20.79	39.86	***7.12***	3.91	10.80	***4.00***	2.36	9.73	***9.21***	7.62	12.22	***8.64***	6.24	12.10	**66.26**
24-Methylen-Cycloartanol	***23.41***	12.70	29.88	***10.31***	6.27	21.42	***37.47***	15.24	45.84	***22.45***	13.89	34.55	***17.90***	10.93	24.30	***29.68***	23.93	40.87	**30.93**
Citrostadienol	***8.29***	7.18	9.89	***3.74***	2.43	4.67	***6.17***	4.90	8.02	***9.01***	6.70	10.57	***6.80***	5.86	8.11	***6.72***	4.42	7.77	**27.13**
Erythrodiol	***4.90***	3.25	6.06	***2.68***	1.41	4.51	***3.39***	2.38	5.13	***2.56***	1.16	5.56	***4.83***	2.65	7.65	***3.76***	2.55	4.95	**8.69**
Uvaol	***0.73***	0.45	1.34	***0.28***	0.10	0.84	***0.37***	0.18	1.02	***0.42***	0.16	1.10	***0.91***	0.47	1.25	***0.66***	0.20	1.28	**5.69**

* Mean values in percentage within each of the three classes of compounds, alcohols, acids and triterpenes

** *Analysis of Variance*: degree of freedom (g-1) = 5 and (N-g) = 57, probability < 0.01, F_{critic} = 3.36

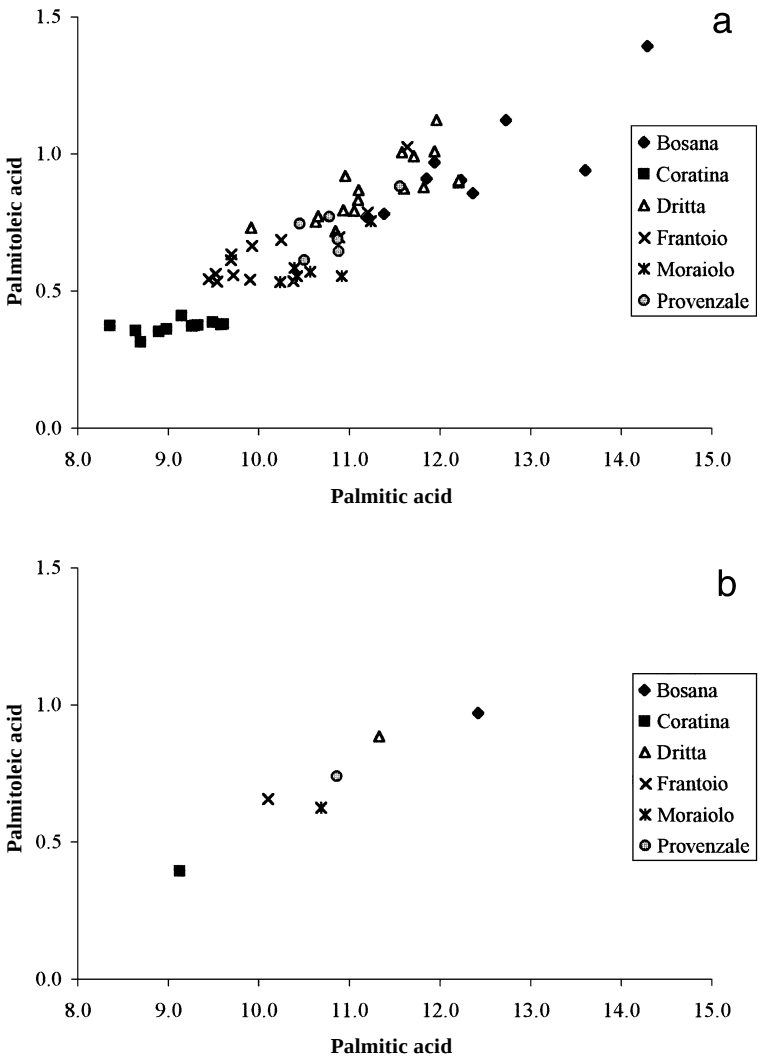

Fig. 2. a) Two-dimensional scatter plots of the correlation between palmitoleic and palmitic acids content. In Fig. 2a all samples are reported, whereas **b)** shows the correlation of the mean values of the two acids, calculated for the six groups of monovarietal oils.

(C16:0) and palmitoleic (C16:1), respectively, are the most evident examples of this.

The good correlation explained by the pair of fatty acids, typical for each of the six cultivars examined, besides illustrating the soundness of the biosynthetic pathways, indicate that this approach is a simple means for a tentative discrimination among the monovarietal oils.

Thus, scatter plots shown in Figs. 2a and 2b reveal the ability of defined oil components to differentiate monovarietal oils: the greater the ability, the more separated the points representing the oils.

Similar results have been obtained from other pairs of substances. Consequently, since almost all oil components (variables) contribute to a greater or lesser extent to the differentiation of oil varieties, multivariate methods were expected to be the appropriate tools to distinguish between them.

As seen and discussed above, the oil variables correlate: therefore investigation of all the variables simultaneously is necessary to reveal all their hidden relationships, and Principal Components Analysis (PCA) is considered a suitable method which can be applied to the matrix of original data.

PCA is a technique to visualise data and find the real dimension of a data set. The measured *p* parameters describe and set out each sample (object) in a *(p-1)*-dimensional space (*p* variables). PCA generates a set of new orthogonal variables (axes), as linear combinations of the original variables, so that the maximal amount of variance contained in the data is concentrated in the first principal component. The second principal component, orthogonal to the first, contains the next largest possible variation; the third next, orthogonal to both the first two and so on. Therefore, by selecting only the most important principal components the original matrix may be

greatly simplified without substantial loss of information. The PCA results are graphically displayed using two plots. In the first, the samples' scores are plotted to show the relationship between the samples; in the second, the loadings of the original measured variables on successive principal components are plotted to aid the interpretation of components in terms of the original variables [31]. The score and loading plots can be interpreted together, because objects with high scores for a specific PC also have high values for the variables with high loadings and low values for those exhibiting low loadings. A generally significant feature of the PCA method is that the results obtained are not independent of the units in which the variables are measured: all variables should be measured on the same scale. Since the three classes of compounds (i, ii, iii) are measured in different scales, the PCA was applied to the correlation matrices. This is equivalent to the use of the variance-covariance matrices of standardised variables by subtracting mean value and dividing by the standard deviation, i.e. by normalising to unit variance [32].

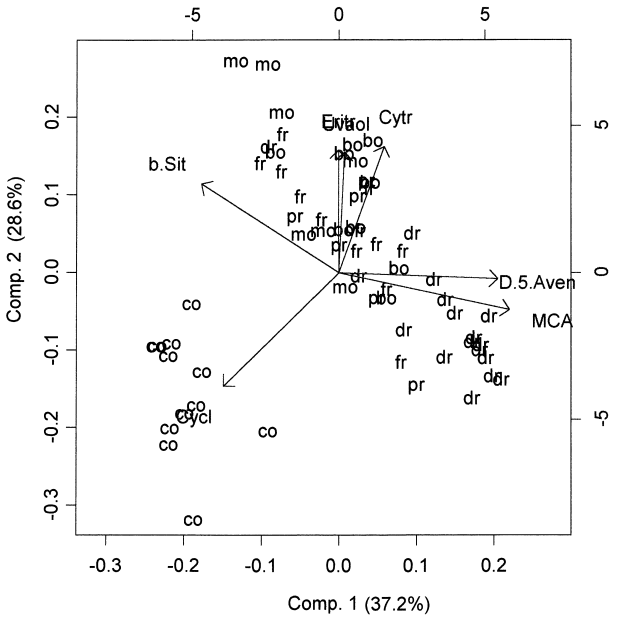

Fig. 3. Triterpenes: PC1 versus PC2, loadings and scores (see text).

A first PCA was applied to the long-chain alcohols C22, C24, C26 of all the oil samples. Two PCs explaining the total variance of the set of the three variables failed in any oil classification and will not be commented on further. In a second PCA analysing the triterpenes, the first three PCs explained 85.9% of variance. The first PC accounted for 37.2% of total variance and was highly correlated with 24-methylencycloartenol, Δ^5-avenasterol and β-sitosterol in that order. The second PC (28.6% of variance) was highly correlated to the two pairs citrostadienol-cycloartenol, and erythrodiol-uvaol. These were the best parameters to describe the oil samples, and the biplot score loadings revealed that the cluster of *Coratina* was due to its high content of cycloartenol, whilst oils of *Dritta* were correlated with their content of 24-methylen cycloartenol and Δ^5-avenasterol (Fig. 3). Furthermore, Fig. 4 – showing the plane defined by the first and third PC – evidenced how citrostadienol is highly correlated with *Frantoio* oils.

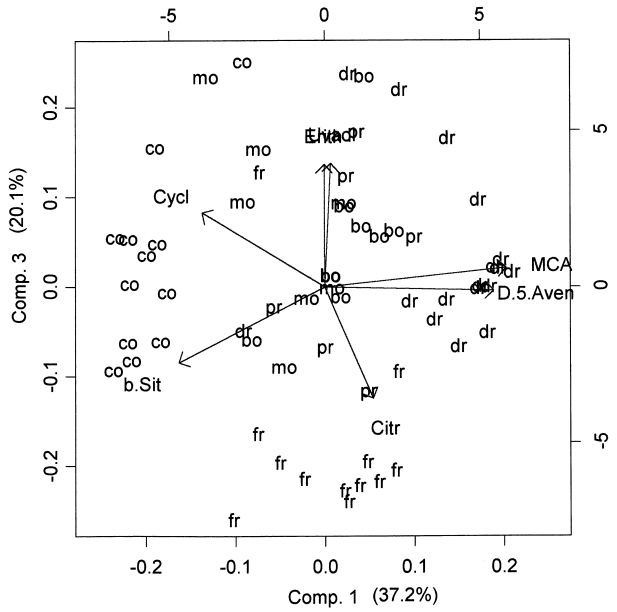

Fig. 4. Triterpenes: PC1 versus PC3, loadings and scores (see text).

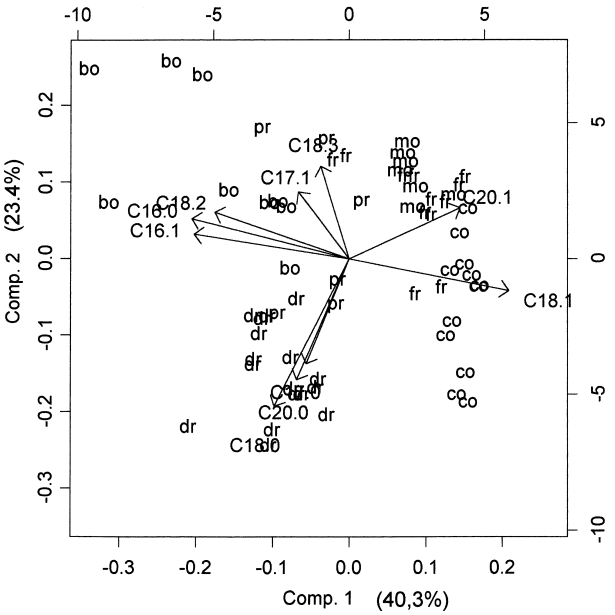

Fig. 5. Fatty acids: PC1 versus PC2, loadings and scores (see text).

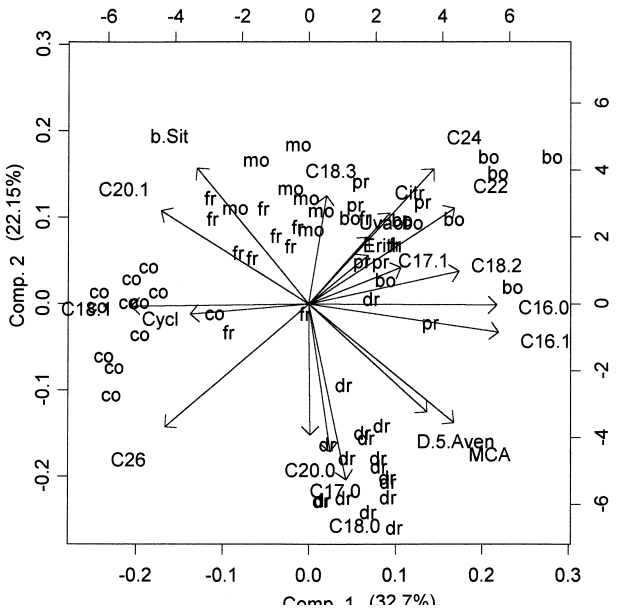

Fig. 6. Long-chain alcohols, triterpenes, fatty acids: PC1 versus PC2, loadings and scores (see text).

When PCA was applied to the fatty acids data, the percentage of variance explained by the first two PCs was 63.7% (Fig. 5). Inspection of Fig. 5 reveals only partially distinguished sample groups of the oils examined. In the loadings and scores biplot shown in Fig. 5 *Coratina* oils appear correlated with C18:1 whilst *Frantoio* and *Moraiolo* oils were correlated with C20:1, both variables being explained by PC1; the second PC2 correlated fairly well

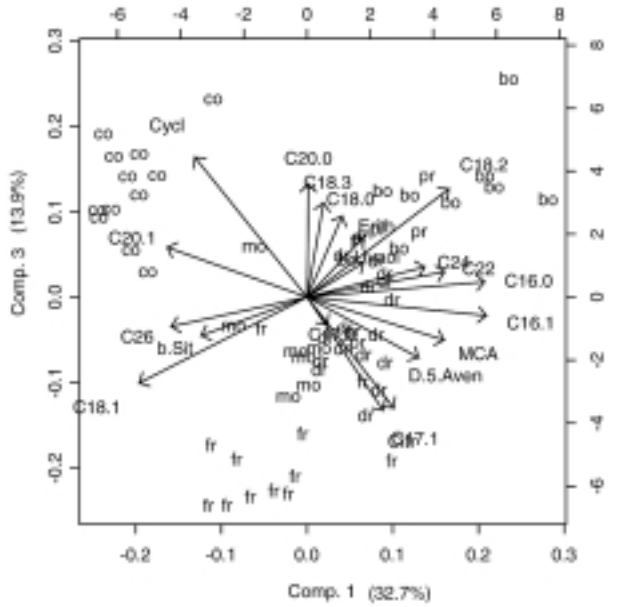

Fig. 7. Long-chain alcohols, triterpenes, fatty acids: PC1 versus PC3, loadings and scores (see text).

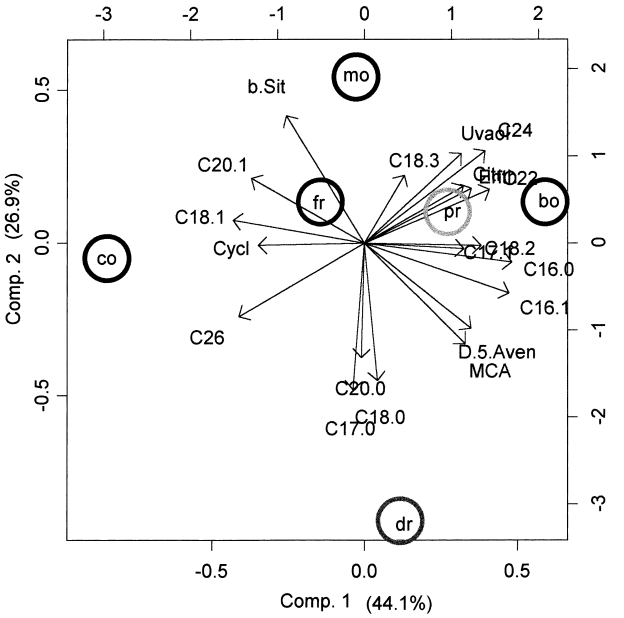

Fig. 8. Long-chain alcohols, triterpenes, fatty acids: PC1 versus PC2, loadings and scores (see text).

with *Dritta* oil samples, explaining most of the variance of the saturated acids C17:0, C18:0, C20:0.

Following the above, independent study of the three subsets of variables, we analysed all the variables at the same time, obtaining more satisfactory results in the differentiation of the six monovarietal oils.

The first three PCs explained 32.7, 22.1 and 13.9%, respectively, of the total variance present in all the original variables. Fig. 6 shows the PCA projection of all oil samples from the different olive oil cultivars and the loadings of the described variables. The PC1–PC2 and PC1–PC3 plots permitted an excellent differentiation of the six groups of oils. Figs. 6 and 7 present the scores and the loadings of the PC1–PC2 and PC1–PC3, respectively, in their corresponding subspaces, explaining up to 68.7% of the initial variance. PC1 (32.7% of the variance) is mainly associated with C18:1 and cycloartenol on one side and with C18:2 and C16:0, C16:1 on the other (the two groups of variables having opposite signs in the plots).

In a second PCA analysis in which the mean values (for a given cultivar) of all oil sample variables were used, the first three PC explained 44.1, 26.9 and 17.1% of the total variance. As is apparent from the relative position of the scores in the plot (Fig. 8), the use of mean values giving all oils an equal weight in the data analysis increased the accuracy of the PCA, revealing clearer differences between the oil varieties. The position of oils from *Dritta*, *Moraiolo* and *Coratina* are the most distinct, whilst *Provenzale* and *Bosana* oils appear to have several characteristics in common and are therefore found closer in

the PC subspace. The score and loading biplots demonstrate the actual correlation between variables and oils.

For instance, the cluster of *Coratina* oils correlates with oleic acid (C18:1) and cycloartenol; *Dritta* correlates with the three saturated acids C17:0, C18:0, C20:0 and with Δ^5-avenasterol and 24-methylenyicloartenol. Other clearly distinct clusters are *Bosana* and *Frantoio* correlating with linoleic acid (C18:2), β-sitosterol and eicosenoic acid (C20:1), respectively.

In conclusion, although we have found in previous works [14, 33-35] that supervised methods of multivariate data analysis (such as Partial Least Squares analysis and Artificial Neural Networks) are much more discriminating in this field and are necessary for the assessment of olive oil adulteration, the unsupervised learning methods described herein were sufficient to obtain a correct assignment of each olive oil sample to its monovarietal group.

Acknowledgement

Part of the paper was presented at the 23rd World Congress and Exhibition of the International Society for Fat Research (ISF) 3-7 October 1999, Brighton, England.

References

[1] European Communities (EC) (1992) Official Journal of the Commission of the European Communities, Regulation No 2081/92, July 14, 1992.

[2] European Communities (EC) (1993) Official Journal of the Commission of the European Communities, Regulation No 2037/93, July 27, 1993.

[3] European Communities (EC) (1991) Official Journal of the Commission of the European Communities, Regulation No 2568/91 L 248, September 5, 1991.

[4] *R. Aparicio:* Characterization: Mathematical Procedures for Chemical Analysis. In: Handbook of olive oil: Analysis and properties. Eds. *J. Harwood, R. Aparicio*, Aspen Publisher, Inc., Gaithersburg, Maryland (USA) 2000, pp. 285–312.

[5] *R. Aparicio:* Authentication. In: Handbook of olive oil: Analysis and properties. Eds. J. Harwood, R. Aparicio, Aspen Publisher, Inc., Gaithersburg, Maryland (USA) 2000, pp. 491–513.

[6] *M. Defernez, E. K. Kemsley:* The use and misuse of chemometrics for treating classification problems. Trends Anal. Chem. **16** (1997) 216–221.

[7] *S. Wold, M. Sjöström:* Chemometrics, present and future success. Chem. Intel. Lab. Systems **44** (1998) 3–14.

[8] *A. Kaufmann:* Multivariate statistics as a classification tool in the food laboratory. J. AOAC Int. **80** (1997) 665–675.

[9] *P. Kaufmann:* Chemometrics in lipid analysis. In: Advances in Lipid Methodology-One. Ed. *W. W. Christie*, The Oil Press, Ayr (UK) 1992, pp. 149–180.

[10] *J. Zupan, M. Novic, X. Li, J. Gasteiger:* Classification of multicomponent analytical data of olive oils using different neural networks. Anal. Chim. Acta **292** (1994) 219–234.

[11] *R. Aparicio, V. Alonso:* Characterization of virgin olive oils by SEXIA expert system. Prog. Lipid Res. **33** (1994) 29–38.

[12] *A. Kiritsakis, W. W. Christie:* Analysis of edible oils. In: Handbook of olive oil: Analysis and properties. Eds. *J. Harwood, R. Aparicio*, Aspen Publisher, Inc., Gaithersburg, Maryland (USA) 2000, pp. 129–158.

[13] *A. Jones, A. D. Shaw, G. J. Salter, G. Bianchi, D. B. Kell:* The exploitation of chemometric methods in the analysis of spectroscopic data: application to olive oils. In: Lipid Analysis of Oils and Fats. Ed. *R. J. Hamilton*, Chapman and Hall, London (UK) 1998, pp. 317–376.

[14] *G. J. Salter, M. Lazzari, L. Giansante, A. Jones, G. Surricchio, D. B. Kell, G. Bianchi:* Determination of the geographical origin of Italian extra virgin olive oil using pyrolysis mass spectrometry and artificial neural networks. J. Anal. Appl. Pyrolysis **40-41** (1997) 159–170.

[15] *O. Eddib, G. Nickless:* Elucidation of olive oil classification by chemometrics. Analyst **112** (1987) 391–395.

[16] *M. Tsimidou, K. Karakostas:* Geographical classification of Greek virgin olive oil by non-parametric multivariate evaluation of fatty acid composition. J. Sci. Food Agric. **62** (1993) 253–257.

[17] *M. Alonso, R. Aparicio:* Characterization of European virgin olive oils using fatty acids. Grasas y Aceites **44** (1993) 18–24.

[18] *R. Leardi, V. Paganuzzi:* Caratterizzazione dell'origine di oli di oliva extravergini mediante metodi chemiometrici applicati alla frazione sterolica. Riv. Ital. Sost. Grasse **74** (1987) 131–136.

[19] *E. Stefanoudaki, F. Kotsifaki, A. Koutsaftakis:* Classification of virgin olive oils of the two major Cretan cultivars based on their fatty acid composition. J. Am. Oil Chem. Soc. **76** (1999) 623–626.

[20] *G. Alberghina, L. Caruso, S. Fisichella, G. Musumarra:* Geographical classification of Sicilian olive oils in terms of sterols and fatty acids content. J. Sci. Food Agric. **56** (1991) 445–455.

[21] *C. M. Lanza, C. Russo, F. Tomaselli:* Relationship between geographical origin and fatty acid composition of extra-virgin olive oils produced in three areas of eastern Sicily. Ital. J. Food Sci. **4** (1998) 359–366.

[22] *L. Ferreiro, R. Aparicio:* Influencia de la altitud en la composicion quimica de los aceites de oliva virgenes de Andalucia. Ecuaciones matematicas de clasificacion. Grasas y Aceites **43** (1992) 149–156.

[23] *C. Armanino, R. Leardi, S. Lanteri, G. Modi:* Chemometric analysis of Tuscan olive oils. Chem. Intel. Lab. Systems **5** (1989) 343–354.

[24] *S. Alessandri, A. Cimato, G. Modi, A. Mattei, A. Crescenzi, S. Caselli, S. Tracchi:* Univariate models to classify Tuscan virgin olive oils by zone. Riv. Ital. Sost. Grasse **74** (1997) 155–164.

[25] *S. Alessandri, A. Cimato, G. Modi, A. Crescenzi, S. Caselli, S. Tracchi:* Multivariate models to classify Tuscan virgin olive oils by zone. Grasas y Aceites **50** (1999) 369–378.

[26] *T. W. Goodwin, E. I. Mercer:* Introduction to plant biochemistry. 2th Ed., Pergamon Press Ltd., Oxford (UK) 1983, pp. 273–327

[27] *P. von Wettstein-Knowles:* Biosynthesis and genetics of waxes. In: Waxes: chemistry, molecular biology and func-

tions. Ed. *R. J. Hamilton,* The Oil Press Ltd., Dundee (UK) 1995, pp. 91–130.

[28] *G. Bianchi:* Chemical genetics of plant surface lipid biosynthesis. Gazz. Chim. Ital. **117** (1987) 707–716.

[29] *N. Frega, F. Bocci, G. Lercker:* Direct gas chromatographic analysis of the unsaponifiable fraction of different oils with a polar capillary column. J. Am. Oil Chem. Soc. **69** (1992) 447–450.

[30] *N. Frega, F. Bocci, G. Lercker:* Composizione lipidica della drupa di olivo di due cultivars della zona del Chianti in funzione della maturazione. Nota II: frazione dell'insaponificabile. Riv. Ital. Sost. Grasse **79** (1992) 77–81.

[31] *J. Piggot, K. Sharman:* Methods of multivariate dimensionality reduction. In: *Statistical Procedures in Food Research*. Ed. *J. R. Piggott*, Elsevier Applied Science, London (UK) 1986, pp. 181–232.

[32] *D. L. Massart, B. G. M. Vandeginste, S. N. Deming, Y. Michotte, L. Kaufman:* Chemometrics: a textbook. Eds. *B. G. M. Vandeginste, L. Kaufman* Elsevier, Amsterdam (The Netherlands), 1988.

[33] *R. Goodacre, D. B. Kell, G. Bianchi:* Neural networks and olive oil. Nature **359** (1992) 594.

[34] *R. Goodacre, D. B. Kell, G. Bianchi:* Rapid assessment of the adulteration of virgin olive oils by other seed oils using pyrolysis mass spectrometry and artificial neural networks. J. Sci. Food Agric. **63** (1993) 297–307.

[35] *A. D. Shaw, A. Di Camillo, G. Vlahov, D. B. Kell, J. Rowland, G. Bianchi:* Discrimination of the variety and region of origin of extra virgin olive oil using ^{13}CNMR and multivariate calibration with variable reduction. Anal. Chim. Acta **348** (1997) 357–374.

[Received: June 19, 2000; accepted: January 11, 2001]