Food Chemistry **25** (1987) 227–239

Authentication of Virgin Olive Oils Using Principal Component Analysis of Triglyceride and Fatty Acid Profiles: Part 1—Classification of Greek Olive Oils

Maria Tsimidou,* Robert Macrae

Department of Food Science, University of Reading, Whiteknights, Reading RG6 2AP, Great Britain

&

Ian Wilson

Department of Applied Statistics, University of Reading, Whiteknights, Reading RG6 2AN, Great Britain

ABSTRACT

Forty-five authentic samples of Greek virgin olive oils, collected from various locations over two seasons, were analysed for fatty acid composition by GLC and for triglyceride composition by HPLC. The data obtained were analysed statistically using Principal Component Analysis. The olive oils could be separated into distinct groups using either the fatty acid or triglyceride data, although the latter provided a higher level of discrimination. It was concluded that the method showed considerable potential for both the characterisation of olive oils and the detection of adulteration.

INTRODUCTION

Multivariate statistics have been used extensively in the characterisation of oils in recent years (Forina & Armanino, 1982; Forina & Tiscornia, 1982; Forina *et al.*, 1983*a,b*; Piepponen *et al.*, 1983; Derde *et al.*, 1984; Gaydou *et al.*, 1984) but not in establishing the authenticity of edible oils (Van Niekerk & Burger, 1985). In most of these applications fatty acid compositional data have been used as the analytical parameters, whilst a similar approach using triglyceride profiles has not been studied. This, in part, may be attributed to the greater difficulty in obtaining reliable quantitative data for

* Present address: State Chemical Laboratory, Poutetsi 19, Ioannina 45110, Greece.

Food Chemistry 0308-8146/87/$03·50

the latter. However, with recent developments in HPLC techniques, including the introduction of novel detection systems (Robinson *et al.*, 1985), these data are now more readily available. Furthermore, the triglyceride compositional data are more likely to be characteristic of a given type of oil or fat as they will contain structural information; for example, the position of the fatty acid residues on the glycerol backbone, which information is lost on transesterification necessary for Fatty Acid analysis by GLC.

Very sophisticated multivariate statistical programs are now available for processing data such as those produced by instrumental chromatographic techniques. In the present work only those techniques that would allow an interpretation of the results directly in terms of the original data were considered. Principal Component Analysis (PCA), stepwise linear regression, discriminant analysis and canonical correlation were used. PCA was found to give adequate discrimination and was adopted throughout. This has the further advantage that it is a technique which may be readily assimilated even by those workers without formal training in statistics. PCA was applied to both fatty acid and triglyceride compositional data for a wide range of authentic olive oils and the discriminating power of the two data sets was compared. In a second stage of the work the same methodology was used to study the adulteration of virgin olive oil with various vegetable oils.

MATERIALS AND METHODS

Samples

Forty-five Greek virgin olive oils from 2 years' harvest, of different varieties and from different origins, were supplied by the Greek Quality Control Authorities (State Chemical Laboratory, Tsoha 16, Athens, Greece). Table 1 shows the characteristics of these samples.

Preparation of fatty acid methyl esters and their determination

Oil (0·02 g), toluene (1 ml) and sodium methoxide (0·5 N, 2 ml) were mixed and left overnight at room temperature. The mixture was then acidified with glacial acetic acid (0·1 ml), saturated sodium chloride was added and the methyl esters extracted with *n*-heptane (2 × 5 ml), after vigorous shaking with the aid of an electrical shaker. The organic layer was separated, dried over anhydrous sodium sulphate (2 g), decanted and filtered. The residue was washed with *n*-heptane (2 × 2 ml) and the washings combined with the organic layer. The solvent was removed by evaporation under reduced pressure and the residue (methyl esters) dissolved in *n*-heptane. These were then analysed by GLC using a Pye Unicam chromatograph (Model 204) on a

TABLE 1
Characteristics of Greek Virgin Olive Oils

Set 1: 1981/1982 Harvest			*Set 2: 1982/1983 Harvest*		
Sample	*Origin*	*Variety*	*Sample*	*Origin*	*Variety*
A	Sterea Hellas (C)	AMFISSA	A′	Crete (S)	TSOUNATI
B	Sterea Hellas (C)	AMFISSA	B′	Crete (S)	TSOUNATI
C	Sterea Hellas (C)	AMFISSA	C′	Peloponnese (SW)	KORONIA
D	Sterea Hellas (C)	AMFISSA	D′	Peloponnese (SW)	KORONIA
E	Sterea Hellas (C)	AMFISSA	E′	Peloponnese (SW)	KORONIA
F	Sterea Hellas (C)	AMFISSA	F′	Peloponnese (SW)	KORONIA
G	Sterea Hellas (C)	PATPA	G′	Peloponnese (SW)	KORONIA
H	Sterea Hellas (C)	AMFISSA	H′	Peloponnese (SW)	LIANOELIA
I	Sterea Hellas (C)	LADOELIA	I′	Peloponnese (SW)	LIANOELIA
J	Sterea Hellas (C)	AMFISSA	K′	Peloponnese (SW)	LIANOELIA
K	Sterea Hellas (C)	AMFISSA	L′	Peloponnese (SW)	LIANOELIA
L	Corfu (NW)	LIANOELIA	M′	Peloponnese (SW)	LIANOELIA
M	Corfu (NW)	LIANOELIA	N′	Peloponnese (SW)	LIANOELIA
N	Evros (NE)	MESOKARPOS	O′	Sterea Hellas (C)	AMFISSA
O	Sterea Hellas (C)	MESOKARPOS	P′	Sterea Hellas (C)	AMFISSA
P	Peloponnese (SW)	KORONIA	Q′	Sterea Hellas (C)	PATRA
Q	Peloponnese (SW)	KORONIA	S′	Sterea Hellas (C)	PATRA
R	Peloponnese (SW)	KORONIA	T′	Sterea Hellas (C)	AMFISSA
S	Peloponnese (SW)	KORONIA	U′	Sterea Hellas (C)	AMFISSA
T	Peloponnese (SW)	KORONIA	V′	Crete (C)	TSOUNATI
U	Peloponnese (SW)	KORONIA	W′	Crete (C)	THROUMPA
			X′	Crete (C)	KORONIA
			Y′	Crete (C)	TSOUNATI
			Z′	Crete (C)	THROUMPA

N, North; S, South; E, East; W, West; C, Central.
Further details of these samples are available from the authors.

glass column (2 m × 3 mm ID) coated with 15% DEGS on Chromosorb WAW DMCS (80–100 mesh) at 180°C and carrier gas flow 75–80 ml/min. Triplicates were prepared for each sample and each was chromatographed once. The reproducibility of the method was evaluated by the analysis of seven replicates of the same sample (% peak area, +/−CV%: palmitic, 11·25+/−0·22; palmitoleic, 0·70+/−4·44; stearic, 2·59+/−1·57; oleic, 78·44+/−0·18; linoleic, 7·01+/−1·82).

Triglyceride determination

Oil (5 μl of a 2·5% w/v solution) was injected via a Rheodyne injection valve (20 μl loop) onto a reversed-phase Spherisorb ODS-2, 5 μm (250 × 4·6 mm ID) analytical column with a mobile phase of acetone/acetonitrile (65:35, v/v) at a flow rate of 1 ml/min. Detection was achieved with an Applied Chromatography Systems Mass Detector (Model No. 750/14) using the following settings: evaporation temperature., 50°C; air pressure, 1·52 bar:

photomultiplier, $\times 2$; sensitivity, $\times 2$. Each sample was chromatographed three times.

Statistical analysis

PCA is a linear transformation of a set of original data $(x_1, \ldots, x_p)$ to a set of uncorrelated components $(y_1, \ldots, y_p)$ in such a way that only a few of the resulting variables account for the majority of the variability observed in the original data (Mardia *et al.*, 1979; Marriot, 1976). Therefore, a reduction in dimensionality is achieved with minimal loss of information. This reduction is significant in the evaluation of large data sets, containing many interrelated variables generated by instrumental analytical techniques, e.g. HPLC, GLC. The arithmetic value of each principal component is given by the equation:

$$(\mathrm{PC}) = a_1 \frac{(x_1 - \bar{x}_1)}{sd_1} + \ldots$$

where x_1 are measurements of the original variables; $\bar{x}_1$ are mean values for the corresponding variables, sd_1 are standard deviations for the corresponding variables, a_1 are loadings of the linear transformation.

Visualisation of the results of PCA is usually achieved by plotting pairs of the first few PCs. A number of coefficients equal to the number of the original variables is assigned to each PC. Each of these coefficients indicates the significance of the participation of the particular variable to the PC in question. Although there is no hypothesis tested in this analysis, PCA may be used to classify objects in terms of predefined groups (or classes), such as variety, origin, etc. It is also possible to define factors that classify the objects into distinct groups, to assign an unknown object to the most similar class or to detect outliers. Such interpretations have been made in the past but they are not based on preconceived ideas. Any given statistical method has inherent limitations and it is recommended that, where possible, complementary techniques should be used together to avoid misinterpretation of data (Bauden, 1984). The PCA program, and other statistical programs, used in the present study was part of the SAS library (SAS Institute Inc., Cary, N. Carolina, 1982).

RESULTS AND DISCUSSION

Determination of fatty acid composition

The Fatty Acids of the virgin olive oils were determined using conventional GLC techniques, preceded by methyl ester formation. The latter was carried

out under alkaline conditions using sodium methoxide at room temperature. The extended reaction time required (*c.* 16 h) was not considered a disadvantage, as a large number of samples should be prepared the previous day and allowed to stand overnight prior to GLC analysis. The reproducibility of the transesterification was checked by suitable replication and the results were compared with transesterification carried out at elevated temperature.

Determination of triglyceride composition

The triglyceride compositions of the same oils were determined by non-aqueous reversed-phase HPLC in conjunction with the mass detector. The nature of the injection solvent used with a mobile phase of acetone/acetonitrile proved to be critical in terms of resolution, and THF was used for this purpose throughout this study (Tsimidou & Macrae, 1985). The mass detector had been previously evaluated in detail (Robinson *et al.*, 1985), from which it was clear that more reliable quantitative data could be obtained under isocratic conditions. Thus, as an adequate and characteristic triglyceride profile could be achieved under these conditions, the additional uncertainty of using gradient elution was avoided. A complete range of the necessary triglycerides was not available for calibration purposes and the data were therefore treated as percentage areas. Some idea of the likely differences between the percentage areas and the percentage composition may be obtained from the previously reported evaluation of the detector (Robinson *et al.*, 1985).

Classification of Greek virgin olive oils—fatty acid composition

Five fatty acids predominated in the olive oil samples studied; namely, palmitic, palmitoleic, stearic, oleic and linoleic acids. No myristic acid was detected under the experimental conditions used, whereas traces of linolenic and arachidic acids were apparent in most samples. The ranges found were in agreement with those recommended by the *Codex Alimentarius* (FAO/WHO, 1969) and also those values reported in other studies on Greek olive oils (Gracian, 1968; International Oil Council, 1979), so that the sample sets were considered as representative of authentic Greek virgin olive oils, although the sets were statistically rather small. Visual examination of the data, or indeed attempts to use simple correlations between components, did not lead to any meaningful discrimination between the samples. PCA, on the other hand, revealed similarities among certain samples with some distinct groupings. Table 2 shows the SAS–PCA output for the evaluation of the 135 observations (45 triplicates) for the FAME data. As the data are recorded as % areas only four variables are used; (*JJ*) palmitic, (*KK*) stearic, (*LL*) oleic

TABLE 2
Standard SAS–PCA Output for the FA Composition Data for the Authentic Olive Oil Samples

Principal component analysis

135 observations
4 variables

Simple statistics

	JJ	*KK*	*LL*	*MM*
Mean	1 100·119	264·000 0	7 815·496	756·866 7
Standard deviation	140·660	35·261 1	254·401	167·719 7

Correlations

	JJ	*KK*	*LL*	*MM*
JJ	1·000 0	−0·203 5	−0·693 0	0·152 2
KK	−0·203 5	1·000 0	0·190 7	−0·282 8
LL	−0·693 0	0·190 7	1·000 0	−0·723 1
MM	0·152 2	−0·282 8	−0·723 1	1·000 0

	Eigenvalue	Difference	Proportion	Cumulative
PRIN 1	2·202 686	1·283 406	0·550 672	0·550 672
PRIN 2	0·919 281	0·102 364	0·229 820	0·780 492
PRIN 3	0·816 916	0·755 799	0·204 229	0·984 721
PRIN 4	0·061 117		0·015 279	1·000 00

Eigenvectors

	PRIN 1	PRIN 2	PRIN 3	PRIN 4
JJ	0·484 177	0·488 548	−0·554 058	0·468 950
KK	−0·304 363	0·789 072	0·522 290	0·109 276
LL	−0·637 581	−0·237 739	−0·161 901	0·714 674
MM	0·516 165	−0·286 647	0·627 711	0·507 333

Principal component, PC or PRIN.

TABLE 3
Eigenvectors of PCA of the HPLC TG data

Variables	Eigenvectors PRIN 1	PRIN 2	PRIN 3	TGs[a]
A	0·314	0·401	0·331	LOL, LPL
B	0·312	0·533	0·254	LOO
C	0·357	0·461	−0·350	PLO
D	0·387	−0·365	0·405	OOO
E	0·454	−0·258	−0·279	POO
F	0·396	−0·154	−0·580	POP, PPP
G	0·401	−0·343	0·351	SOO

[a] Expected triglycerides from Dong & Dicesare (1983).
L, Linoleic; O, Oleic; P, Palmitic; S, Stearic acids.

and (*MM*) linoleic acids. The eigenvalue (Table 2) for each PC (or the Cumulative) is a measure of the participation of each PC (or combination of PCs) in the explanation of the total variability of the original data (Marriot, 1976). The first PC accounted for 55%, and the first two cumulatively for 78%, so that the third PC was also included. Plots of the first three PCs, taken in pairs, did not seem to be influenced by the harvest year. However, geographical origin and variety grouped the oils, sometimes quite clearly, as shown in Figs 1 and 2, respectively. A strong correlation was observed

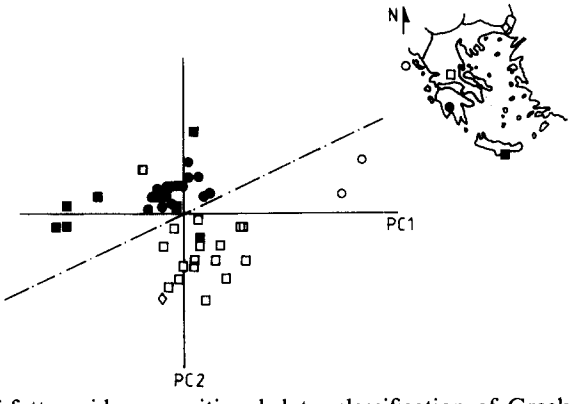

Fig. 1. PCA of fatty acid compositional data, classification of Greek olive oils on the basis of geographical origin. Corfu, ○; Sterea Hellas, □; Crete, ■; Peloponnese, ●; Evros, ◇.

between latitude and a direction on the plane of the first two PCs, separating oils from the North and South. The two samples not correctly classified may be due to the result of other factors (e.g. variety) as explained below. The ability to classify oils on this basis may be of direct relevance to trade control, as olive oils are often marketed with reference to their origin (e.g. Cretan oils) both internally and internationally.

All the plots of the first three PCs seemed to be affected by variety. However, as origin contributed to the values of PC1 the plot of the PC2 versus PC3 was more suitable for studying the grouping of the oils on the basis of variety. The main drawback with such a classification lies in the confusion found in the literature (Siggelakis, 1982) on the names of the olive tree cultivars (local or even misused names) which limits the discriminating power of any statistical treatment. Thus, LIANOELIA proved to be another trivial name for KORONIA, whereas the name 'LADOELIA' is either referred to as KORONIA or used as a collective trivial name. This could explain why Sample I, although originating from Sterea Hellas, was grouped together with oils from Peloponesse.

Eigenvector is another term for the PC coefficients and a number of

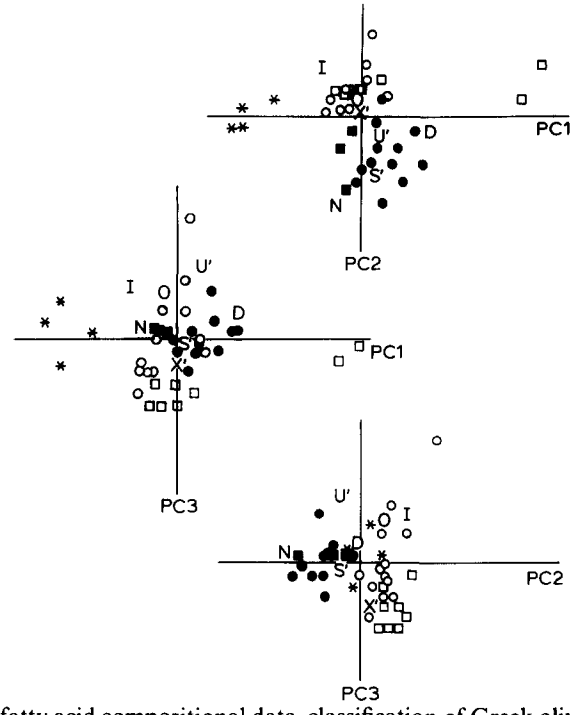

Fig. 2. PCA of fatty acid compositional data, classification of Greek olive oils on the basis of variety. LIANOELIA, □; AMFISSA, ●; KORONIA, ○; TSOUNATI, *; PATRA, ■; UNIDENTIFIED VARIETIES I(LADOELIA), O, U', D, S', N, X'.

loadings are assigned to each PC which represent the participation of a particular variable to the PC in question. The loadings shown in Table 2 suggest that oleic acid and linoleic acids contribute more to PC1, whereas the overall saturation dominates PC2. Figures 3 and 4 show the influence of linoleic acid and palmitic acid on the distribution of oils across PC1 and PC2, respectively. Indeed, regression of linoleic acid versus PC1 and palmitic

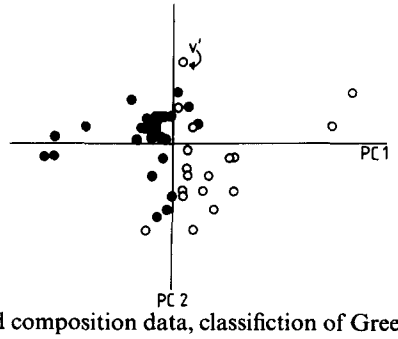

Fig. 3. PCA of fatty acid composition data, classifiction of Greek olive oils on the basis of linoleic acid content, expressed as % area; >7·6, ●; <7·6, ○. (7·6 is mean linoleic acid content).

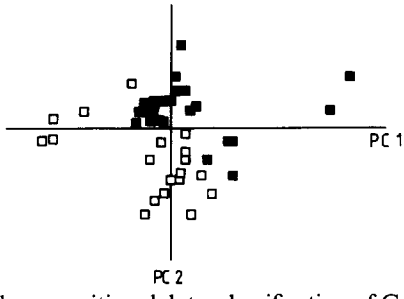

Fig. 4. PCA of fatty acid compositional data, classification of Greek olive oils on the basis of palmitic acid content, expressed as % area; >11·0, ■; <11·0, □. (11·0 is mean palmitic acid content).

acid versus PC2 gave quite high correlation coefficients, 0·778 and 0·627, respectively. If the samples are examined with respect to their linoleic acid contents, they may be assigned to one of two main subgroups of oils, one with more, and the other with less, than the mean value for the sample set. Similarly, two subgroups can be found with regard to palmitic acid levels. From this point of view the Cretan oil (V') is no longer an outlier as it is grouped with oils of similar linoleic acid content and not according to its origin; compare Figs 1 and 3.

PCA of fatty acid compositional data may therefore be used to classify oils with respect to their geographical origin, variety and linoleic (or palmitic) acid content. Projections on different principal component planes may contribute to a better understanding of the classification rules and of the reasons for apparent 'outliers'.

Classification of Greek virgin olive oils—triglyceride composition

The triglyceride profiles of the virgin olive oils studied were all found to be qualitatively similar. A typical chromatogram under the conditions used is shown in Fig. 5, in which the peaks A–G represent the triglyceride components used in the subsequent PCA. Some of these peaks are doubtless complex. A complete range of the necessary triglyceride standards was also not available and so the data were treated as percentage areas, expressed on a unit weight basis. Nonetheless, evaluation of these uncalibrated data was found to be very promising and produced a similar classification to that already reported for the fatty acid data. Table 3 gives the eigenvectors for the first three PCs used in the statistical evaluation, explaining cumulatively 53%, 75% and 90%, respectively. The PCA of the triglyceride profiles was superior with respect to the year of harvest. Oils from the two harvest years were differentiated across PC1, a result which is not explained but which was verified for oils of the same variety or origin (Fig. 6). This influence is

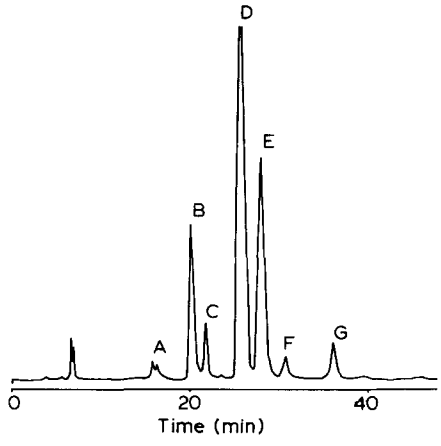

Fig. 5. Typical HPLC triglyceride profile of olive oil, A–G major triglyceride groups used in PCA.

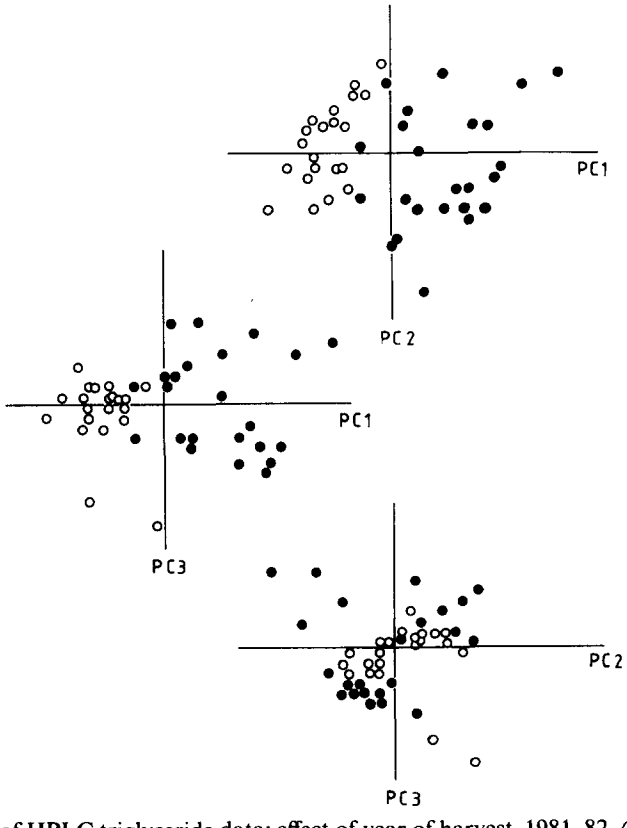

Fig. 6. PCA of HPLC triglyceride data; effect of year of harvest, 1981–82, ○; 1982–83, ●.

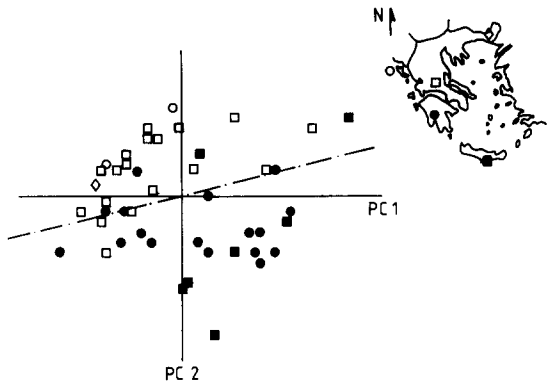

Fig. 7. PCA of HPLC triglyceride data; effect of geographical origin, Corfu, ○; Sterea Hellas, □; Crete, ■; Peloponnese, ●; Evros, ◇.

removed from the plane of PC2 and PC3 and may be related to the 'size' of the measurements (Marriot, 1976), as reported in cases where all the eigenvectors of PC1 are of similar size and are all positive.

PC2 was related to latitude (Fig. 7) and the same outliers were observed as with the fatty acid data. Variety affected the dispersion of the samples across PC3 and the plane of PC2 and PC3 provided adequate discrimination of the oils on this basis (Fig. 8). Stepwise linear regression showed that the three variables *B*, *D* and *E* out of the seven (*A*–*G*) explained the majority of the linoleic acid content ($r^2 = 0{\cdot}928$), a fact which may be useful in reducing the amount of data that needs to be obtained, especially if calibrated data are to be sought. The high correlation might be expected since the chromatographic conditions used would result in the elution of most unsaturated species. This is illustrated in Fig. 9 where the oils are arranged across PC2 in order of increasing linoleic acid content.

HPLC-triglyceride profiles not only provide detailed 'fingerprints' of the

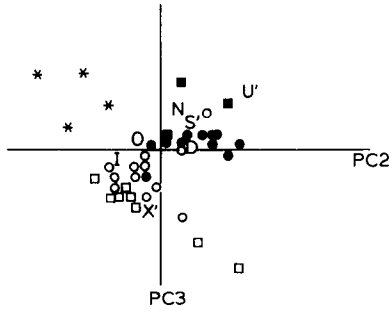

Fig. 8. PCA of HPLC triglyceride data; effect of variety, LIANOELIA, □; AMFISSA, ●; KORONIA, ○; TSOUNATI, ∗; PATRA, ■.

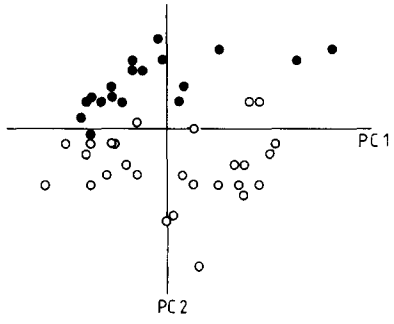

Fig. 9. PCA of HPLC triglyceride data; effect of unsaturation, expressed as linoleic acid content of the oils; % area; >7·6, ●; 7·6, ○ (7·6 is mean linoleic acid content).

oils studied but the peak area data may be used to classify olive oils according to the year of harvest, variety and geographical origin and can also provide an indication of the level of linoleic acid present.

CONCLUSION

PCA was found to be a promising technique for the classification of virgin olive oils using either fatty acid or triglyceride compositional data, the latter providing somewhat better discrimination. The classification is not based on any hypothesis so that the dispersion of the objects (samples) cannot be tested for its significance. However, this dispersion is obtained by a simple projection on the planes of the PCs and therefore is real. Such simple statistical treatments may well find application in quality control laboratories.

REFERENCES

Bauden, D. (1984). Disclose: An integrated set of multivariate display procedures for chemical and pharmaceutical data, *Analytica Chimica Acta*, **158**, 363–8.

Derde, M. P., Coomans, D. & Massart, D. L. (1984). SIMCA demonstrated with characterisation and classification of Italian olive oils, *Journal of Association of Analytical Chemists*, **67**, 721–7.

Dong, M. W. & Dicesare, J. L. (1983). Improved separation of natural triglycerides by liquid chromatography using columns packed with 3 μm particles, *Journal American Oil Chemist's Society*, **60**, 788–91.

FAO/WHO (1969). Joint food standards programme, CAC/RS 33-1969, vigin and refined olive-residue oil.

Forina, M. & Armanino, C. (1982). Eigenvector projection and simplified non-linear mapping of fatty acid content of Italian olive oils, *Annali di Chimica*, **72**, 127–41.

Forina, M. & Tiscornia, E. (1982). Pattern recognition methods in the prediction of Italian olive oil origin by their fatty acid content, *Annali di Chimica*, **72**, 143–55.

Forina, M., Armanino, C., Lanteri, S. & Tiscornia, E. (1983*a*). Classification of olive oils from their fatty acid composition. In: *Food research and data analysis* (Martens, H. & Russuwurm, H. (Eds)), Applied Sci. Publ. Ltd, London, 189–214.

Forina, M., Armanino, C., Lanteri, S., Calcogno, C. & Tiscornia, E. (1983*b*). Volutazione delle caratteristiche chimiche dell'olio di oliva in funzione dell'annata di produzione mediante metodi di classificozione multivariate, *Rivista Italiana delle Sostanze Grasse*, **LX**, 607–13.

Gaydou, E. M., Iatrides, M. C., Tamayanarivo, R. & Peiffer, G. (1984). Pattern recognition analysis of fatty acids. Application to beef fat tissue characterisation, *Journal of Agriculture and Food Chemistry*, **32**, 651–5.

Gracian, J. T. (1968). The chemistry and analysis of olive oil. In: *Analysis and characterisation of oils, fats and fat products*, Vol. 2 (Bookenoogen, H. A.), 317–591.

International Olive Oil Council (1979). File No. 14.

Mardia, K. U., Kent, J. T. & Bibby, J. M. (1979). Principal component analysis. In: *Multivariate analysis*, Academic Press, London, 213–54.

Marriot, F. H. C. (1976). *The interpretation of multiple observations*, Academic Press, London.

Piepponen, S., Vainionpaa, A. & Roos, A. (1983). Classification of some food oils. In: *Food research and data analysis.* (Martens, H. & Russuwurm, H. (Eds)), Applied Sci. Publ. Ltd, London, 426.

Robinson, J. L., Tsimidou, M. & Macrae, R. (1985). Evaluation of the mass detector for quantitative detection of triglycerides and fatty acid methyl esters, *Journal of Chromatography*, **324**, 35–51.

Siggelakis, H. M. (1982). *Olive tree, Greek and foreign varieties.* (Elaiourgiki, S. Y. N. P. E. (Ed), Athens.

Tsimidou, M. & Macrae, R. (1985). Reversed-phase chromatography of triglycerides—Theoretical and practical aspects of the influence of injection solvents, *J. Chromatographic Science*, **23**, 155–60.

Van Niekerk, P. J. & Burger, A. (1985). The estimation of the composition of edible oil mixtures. *Journal American Oil Chemists Society*, **62**, 531–8.