Characterization of Italian Olive Oils Based on Analytical and Nuclear Magnetic Resonance Determinations

Antonio Sacco[a,*], Maria Antonietta Brescia[a], Vitantonio Liuzzi[b], Fabiano Reniero[c], Claude Guillou[c], Stefano Ghelli[d], and Pieter van der Meer[e]

[a]Università di Bari, Dipartimento di Chimica, Campus Universitario, 70126, Bari, Italy, [b]Università di Bari, Facoltà di Agraria, Istituto di tecnologia dei Prodotti Agroalimentari, Campus Universitario, 70126, Bari, Italy, [c]Commission of the European Communities, Joint Research Centre, Institute for Health and Consumer Protection, Food Products and Consumer Goods Unit, I-21020 Ispra (VA), Italy, [d]SPIN, 42048 Rubiera (RE), Italy, and [e]Agrotechnological Research Institute (ATO-DLO), Wageningen, The Netherlands

ABSTRACT: Analytical measurements and proton nuclear magnetic resonance (^{1}H NMR) spectra of phenolic extracts were performed on a set of Italian extra-virgin olive oils from different cultivars and geographical locations of Apulia region. Multivariate statistical analysis (principal component analysis, hierarchical clustering analysis, and discriminant analysis) was applied separately to analytical and NMR data. Analytical parameters, in particular fatty acid compositions, permit the discrimination of olive variety, while ^{1}H NMR data of phenolic extracts permit a classification according to the geographical origin of the samples.

Paper no. J9505 in *JAOCS 77,* 619–625 (June 2000).

KEY WORDS: Chemometrics, extra-virgin olive oil, fatty acids composition, NMR.

Olive oil is a fundamental component of the Mediterranean area diet, and it is widely used as a condiment, cooking medium, emulsion component, and in the storage of vegetable and animal food. Olive oil cultivation originated in Asia Minor and has spread to Greece, Italy, Spain, and North Africa. Recently it also was extended in other countries with temperate climate like France.

Italy contributes about one-third of the world olive oil production, and 40% of this quantity is produced in the Apulia region. Extra-virgin olive oil is the most characteristic Apulian product, due to the favorable environmental conditions for the cultivation of olive trees. Many areas where high-quality olive oil is produced are located in this region. The characteristics of these oils are determined by a series of factors: the nature of the soil, the climate, variety of the plant, cultivation, and oil extraction techniques. European Community trade rules distinguish the identity of olive oils other agricultural products that could claim a typical quality based on geographic origin of production. These specifications protect producers of high-quality olive oils and ensure consumer awareness of product quality. Therefore, analytical techniques are needed to ensure proper classification. The denomination of protected origin (DOP) is based on the precise definition of several parameters such as: cultivar, geographical origin, agronomic practice, production technology, and organoleptic qualities. However, precise methods are needed for the assignment of a trademark to olive oils.

In recent years several attempts were made to define olive oil origin and varieties by analysis of chemical parameters. A first classification of Italian olive oils from different regions using principal component analysis (PCA) of fatty acid composition was obtained (1). Data from different analytical determinations were elaborated with a mathematical model (2) to discriminate oils coming from Andalusia according to their origin and variety. Chemometric methods on fatty acid composition data also were applied to the geographical classification of Greek oils (3). Recently a new method for determining geographic origin of Italian olive oils was developed using multivariate statistical analysis on the normalized intensities of proton nuclear magnetic resonance (^{1}H NMR) signals due to minor components (4). Discrimination of extra-virgin olive oils with respect to their region and variety was successfully realized applying PCA, principal components regression (PCR), and partial least squares (PLS) to ^{13}C NMR data (5). Another approach led to the discovery that oils of different varieties may be differentiated on the basis of the repartition of phenolic compounds, but not by differences in fatty acid composition (6). The present work shows how the statistical elaboration of analytical and NMR data may be used to authenticate the origin of olive oils in Apulia.

EXPERIMENTAL PROCEDURES

Analytical determinations. Twenty-eight virgin olive oils were supplied by the Institute of Agricultural Food Products Technologies of the University of Bari, Italy. Olive oil varieties were Coratina, Leccino, Oliarola, Olivastro, and Si-

*To whom correspondence should be addressed at Dipartimento di Chimica, Campus Universitario, Via Orabona 4, 70126 BARI, Italy.
E-mail: sacco@lgxserve.ciseca.uniba.it

mone. Olive oil samples came from three different locations (north, coastal south, and hilly hinterland) situated in a restricted area of the Italian Apulia region.

Determinations of humidity, oil acidity, peroxide value, and specific ultraviolet (UV) absorbance (K_{232} and K_{270}) were made according to official analytical methods (7). Fatty acid composition was determined by gas chromatography (GC) of methyl esters obtained after transesterification of the triacylglycerols (8), with a Fisons Instrument Model HRGC MEGA 2 equipped with a SP-2340 60 m × 0.25 mm i.d. fused-silica column coated with a 0.25 μm poly(biscyanopropylsiloxane) stationary phase. Oven temperature was programmed from 140°C for 10 min to 220°C at a rate of 1°C/min, and held at 220°C for 10 min. The flame-ionization detector temperature was 240°C; injector temperature, 220°C; splitting ratio 1:50; carrier gas, helium; and the flow rate was 20 cm/s. The individual concentrations of 14 fatty acids (palmitic, palmiteladic, palmitoleic, margaric, heptadecenoic, linolenic, stearic, oleic, *cis*-vaccenic, linoleic, arachidic, gadoleic, behenic, and lignoceric) expressed as percentage were used in this work.

NMR determinations. NMR spectra were determined on the oil phenolic compounds obtained with the following procedure: 25 g of oil was extracted with three 20-mL vol of CH_3OH. The solution was concentrated under vacuum at 35°C. The residue was resuspended in 8 mL of CH_3CN and washed three times with 15 mL hexane to eliminate the residual oil. The resulting CH_3CN solution was evaporated under vacuum at 35°C.

These oil extracts were dissolved in 0.8 mL of deuterated dimethyl sulfoxide (DMSO-d_6) (99.9% isotopic purity; Isotec Inc., Miamisburg, OH) and placed in 5-mm NMR tubes. The same bottle of deuterated solvent was used for all the measured samples. ^{1}H NMR spectra were recorded at 300 K on two high-resolution spectrometers, AMX-400 400.13 and AMX-500 500.14 MHz (Bruker Analytik GmbH, Rheinstetten, Germany), at the Wageningen NMR Centre facility. ^{1}H NMR spectra were taken with 90° pulse angle, 10 s recovery delay, and 128 scans.

A typical ^{1}H NMR spectrum of an oil phenolic extract is shown in Figure 1. ^{1}H NMR spectra of all oil extracts except two Coratina and one Oliarola samples from northern area were taken.

Chemometrics. Statistical elaboration of the data was done using STATISTICA software (StatSoft Inc., Tulsa, OK) at the Joint Research Centre. Analytical and NMR data were analyzed in two different data sets. In both cases an analysis of the matrix condition was initially performed to obtain information about collinearity among the variables. Collinearity could cause arbitrary classifications and prevent accurate matrix inversion calculations. Therefore, variables collinear to others were eliminated from the data sets.

The applied chemometric methods were PCA, hierarchical clustering analysis (HCA), and discriminant analysis (DA). From a number of original variables that are frequently correlated, PCA extracts new variables, called principal components (PC), that are uncorrelated and ordered according to their variance (9). The coefficients of the linear combinations are called loadings. The first two or three PC generally show the main structure of the data while the others are frequently connected with less useful information or noise. Therefore, starting with a set of data initially present in an *n*-dimensional space, PCA permits their representation in a space of fewer dimensions. HCA discriminates groups, called clusters, so that similarity among objects in a group is higher than similarity among objects belonging to different groups. The similarity matrix was computed by measuring the distance between two oils X_1 and X_2 using the Manhattan distance defined as:

$$\sum_{j=1}^{n} |X_1(j) - X_2(j)| \qquad [1]$$

where *n* is the number of variables. The oils were classified by using complete linkage method (10). The results are shown in a dendrogram, which may be used to detect groups of similar individuals. DA classifies an item into one of several mutually exclusive groups and it is also used to test how well two or more groups are separated (9). Its purpose is to calculate class models and boundaries, giving a rule of classification based on a set of known objects (training set). This rule can be applied to define the classification of unknown objects (test set).

RESULTS AND DISCUSSION

Variety discrimination. The concentrations of linolenic, stearic, and palmitic acids showed correlation with other fatty acid concentrations. Therefore, they were eliminated from the statistical analysis. The concentrations of the remaining fatty acids were used as input variables to perform PCA. The results are reported on the score plot in Figure 2. Separation of Coratina and Simone (positive) from Oliarola (negative) cultivars are evident on PC1, while on the second PC Simone, Olivastro and Leccino cultivars, which have similar scores on PC1, are differentiated. The loading values of the variables associated to the first two principal components, reported in Table 1, give useful information. In fact, the higher a loading of a variable on a particular principal component, the more the variable has in common with this component. Hence, the loadings can be interpreted as correlations between the variables and the components. In the present case, the PC1 loadings are high and positive for oleic and gadoleic acids and high and negative for *cis*-vaccenic and palmitoleic acids. This could reflect the relevant influence of Coratina and Simone cultivars on the content of the first two acids and of Oliarola cultivar for the second group of acids. The second PC is more correlated with the concentration of behenic acid. Considering that PCA is an unsupervised method, it is noteworthy that a separation of the oils according to the cultivar was achieved in the score plot. PCA also was applied to the entire set of analytical determinations (fatty acid composition, humidity, oil acidity, peroxide value, and specific UV absorbance). How-

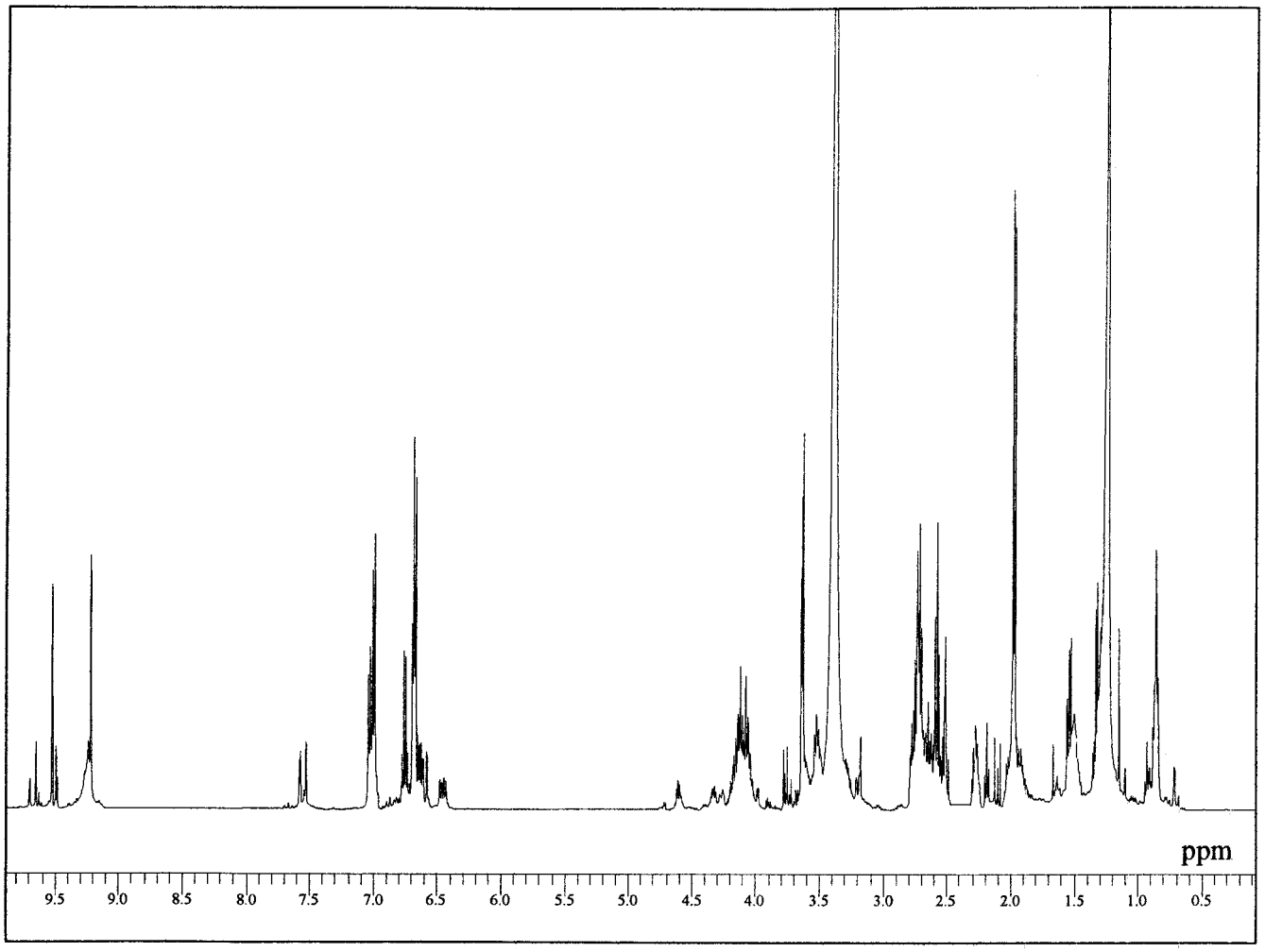

FIG. 1. ^{1}H nuclear magnetic resonance (NMR) spectrum of the olive oil phenolic extract.

ever, a poorer separation between oil varieties was obtained. It is possible to conclude that the added variables had no relevance to the discrimination considered and therefore introduced only noise in the system.

TABLE 1
Loading Values for the First Two Principal Components (PC) Obtained from Fatty Acid Composition Data

Fatty acids	PC1	PC2
Palmitic	0.427	–0.008
Palmitoleic	–0.942	0.003
Margaric	0.382	–0.663
Heptadecenoic	–0.276	–0.587
Oleic	0.948	–0.197
Cis-vaccenic	–0.819	–0.405
Linoleic	-0.680	0.513
Arachidic	0.588	0.382
Gadoleinic	0.829	0.086
Behenic	0.092	–0.733
Lignoceric	–0.216	–0.340

HCA on fatty acid composition data was performed to confirm the differentiation of the oils according to cultivars. The dendrogram shown in Figure 3 exhibits the clustering of the olive oils grouped on the basis of their cultivar. At linkage distance 7.2, it is possible to obtain five clusters in which three are prevalently composed of samples from the same cultivar, while the other two contain samples from different cultivars. This behavior is in agreement with the results obtained in the score plot (Fig. 3). DA also was performed using fatty acid composition data (Fig. 4). An excellent separation between cultivars is shown in this three-dimensional plot. The first discriminant function is weighted mostly by the concentrations of oleic and palmitoleic acids. The second and the third functions seemed to be distinguished most by the concentration of palmitoleic acid.

In order to demonstrate that this model can be safely used to classify unknown samples, it is necessary to check its reliability using a set of known samples as unknown samples to be classified. In this way it is possible to test how the model calculation depends on the actual objects in the training set. Three replicates of four samples (2 Oliarola oil, 1 Coratina

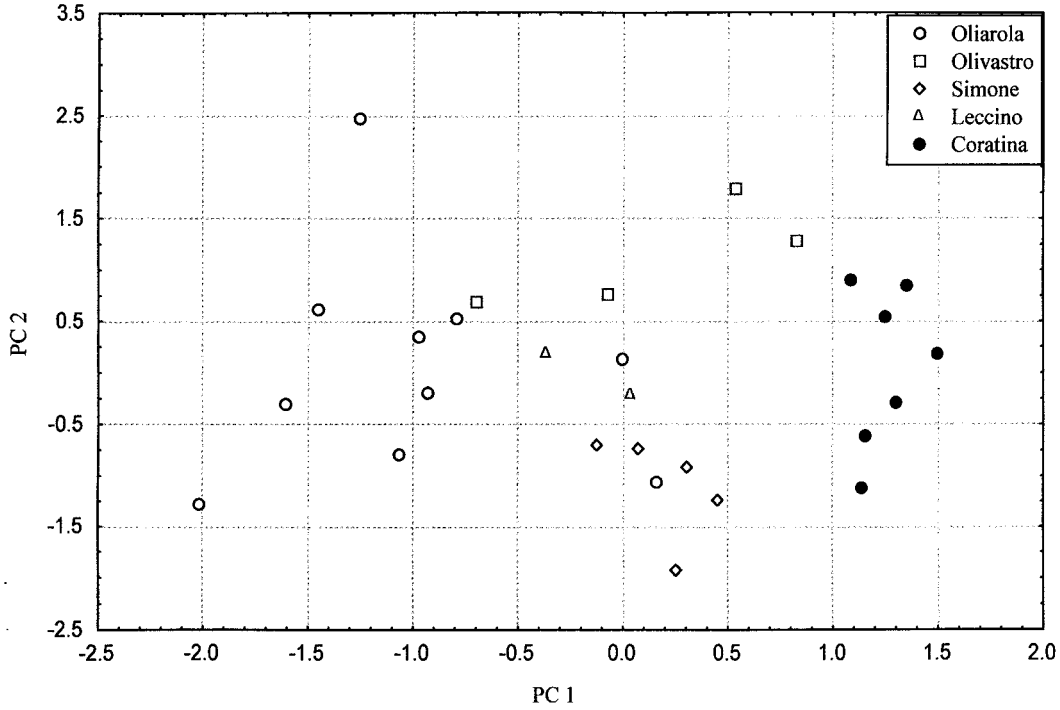

FIG. 2. Scatter plot of the scores from the first two principal components obtained using fatty acid composition data.

oil, and 1 Simone oil) have been removed from the initial population, and DA has been applied to the remaining data (training set). The excluded oils were then introduced in the system as unknowns (test set). Results showed that they have been correctly classified. Hence, the stability of the model was ascertained.

As far as ^{1}H NMR data are concerned, it has been deemed interesting to analyze spectra of olive oil phenolic compounds

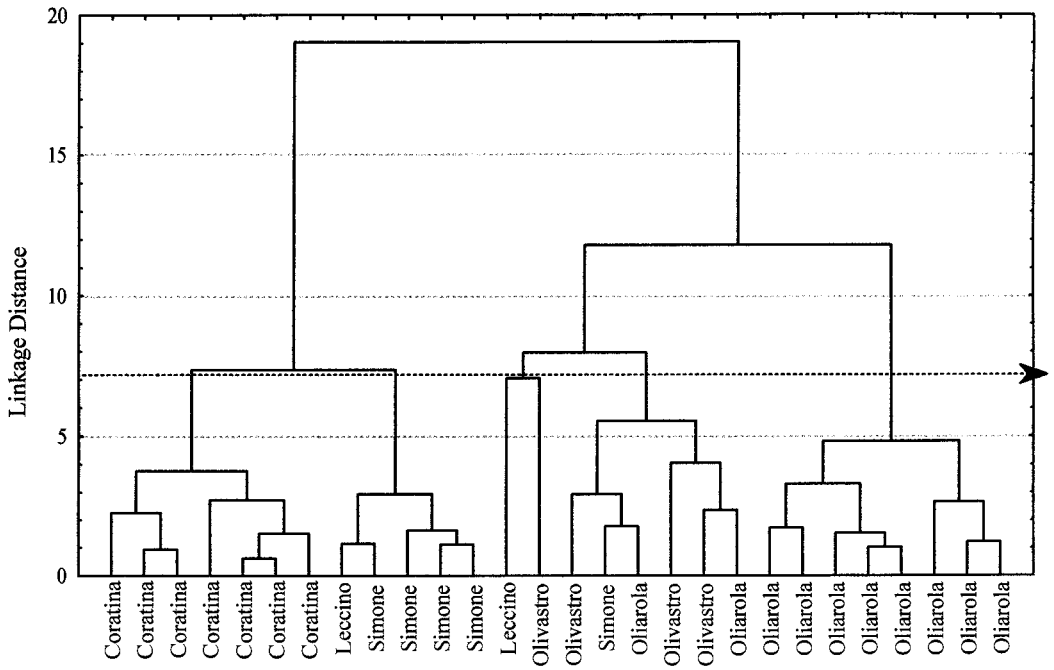

FIG. 3. Dendrogram of extra virgin olive oils obtained using fatty acid composition data.

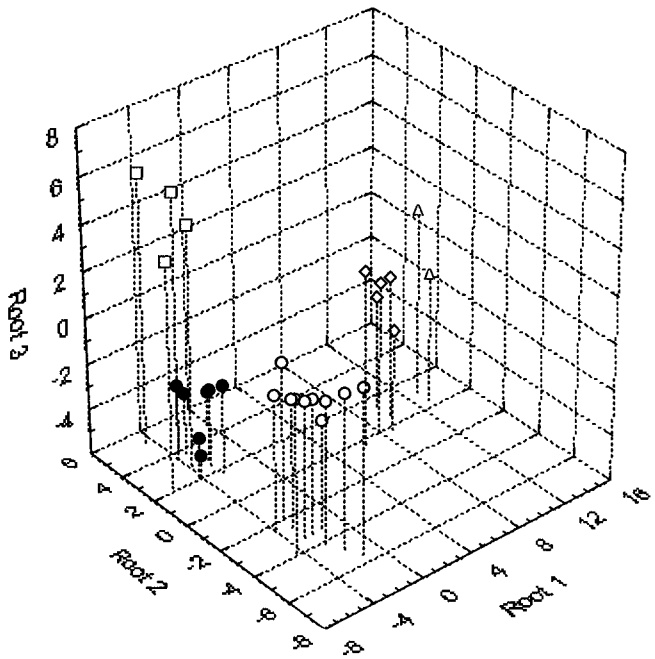

FIG. 4. Plot of the five cultivars for olive oil samples on the first three discriminant functions for fatty acid compositional data (●, Coratina; □, Olivastro; ◇, Simone; ○, Oliarola; △, Leccino).

because their content is heavily affected by the variety, origin, and maturity degree of olives (11–14). Moreover, the amount of these substances is a very important parameter for the evaluation of virgin olive oil quality since phenols are strictly related to the typical bitter taste of olive oil and to the oil's resistance to oxidation. Multivariate analysis was performed to assess the relation between phenolic compounds and olive varieties. To this purpose, ^{1}H NMR spectra of the phenolic extracts were normalized to the ^{1}H signal of the solvent peak (DMSO-d_6). Variable selection according to Shaw *et al.* (5) was applied to select the resonances having the highest power to distinguish between different oils. The selected variables were normalized intensities of peaks at 9.69, 9.64, 9.52, 9.48, 9.22 ppm (aldehydic protons); 7.57 ppm (vinylic protons); 6.98, 6.77, 6.75, 6.68, 6.64, 6.60, 6.42 ppm (aromatic protons), and 4.70 ppm. The detailed identification of these signals for the phenolic compounds in olive oils is still in progress, but it is believed that some of these signals are characteristic of secoiridoid compounds (15).

A relation between oil varieties and selected ^{1}H NMR normalized peak intensities also was investigated. A classification of the samples according to their variety with prediction ability of 100% was obtained with DA (Fig. 5). Peaks at 6.75 and 6.68 ppm contributed most to the first discriminant function, while the second discriminant function primarily was determined by peaks at 6.75 and 9.52 ppm. In this case the calculated model was not stable due to the smaller number of samples analyzed with ^{1}H NMR.

Geographical origin discrimination. As well as for varieties, a geographical origin discrimination through the performed analytical and NMR data was investigated. No correlation was found between fatty acid composition data and geographical origin of the samples. To classify olive oils according to their geographical origin, DA on the same groups of selected NMR peaks discriminated among olive varieties. The prediction ability of the obtained model was 96% (Fig. 6). The first discriminant function, which discriminates mostly between the northern area and the others, was weighted mostly by peaks at 6.75 and 9.52 ppm. The second

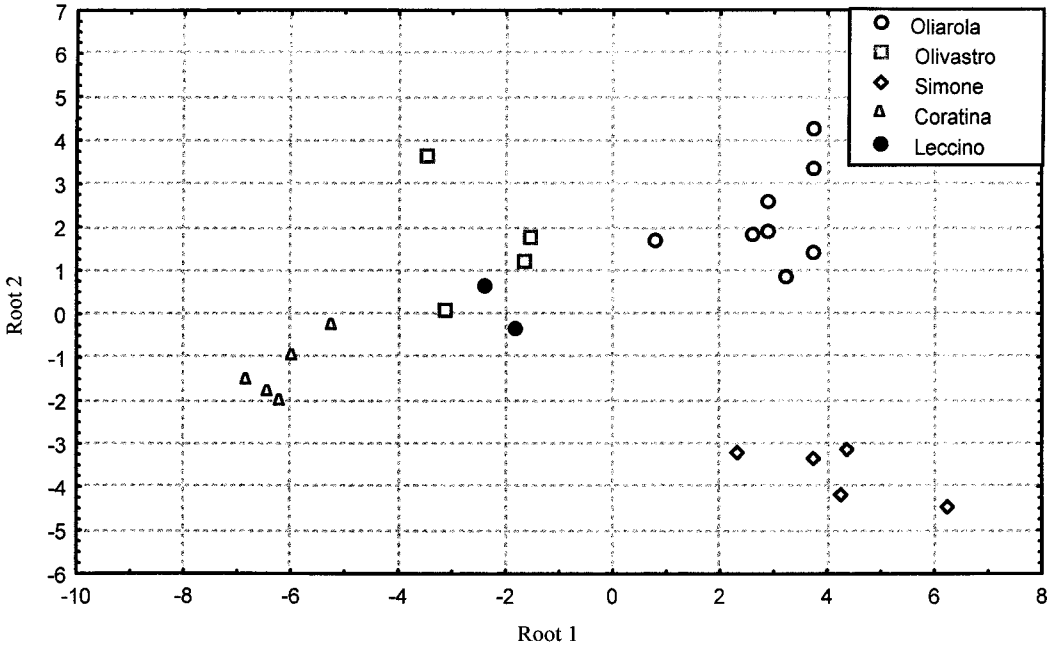

FIG. 5. Plot of the five cultivars for olive oil samples on the first two discriminant functions for NMR data. See Figure 1 for abbreviation.

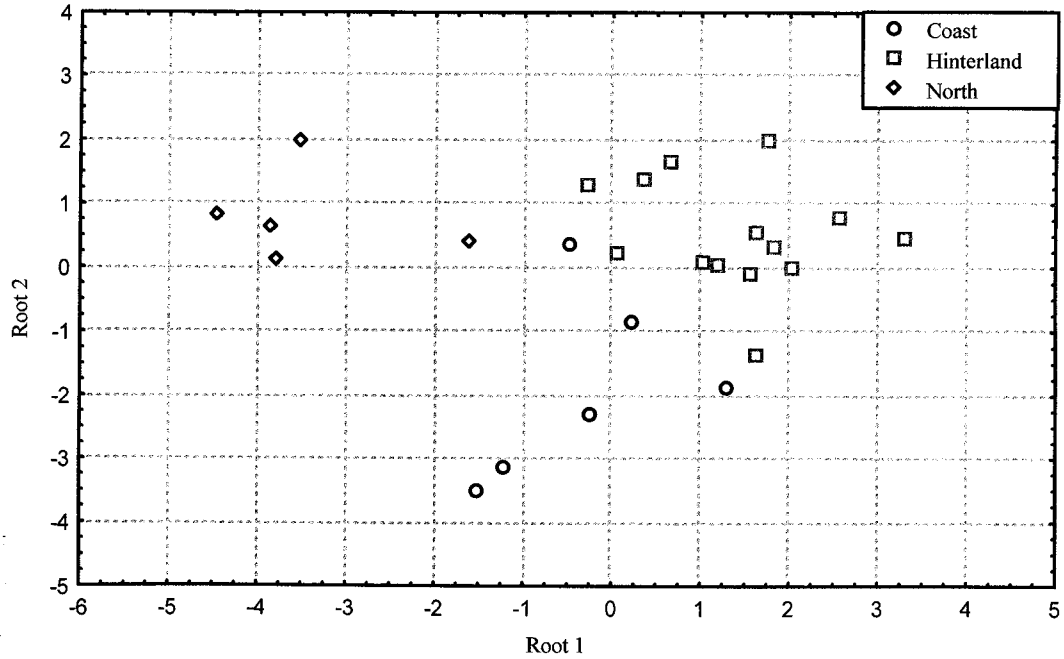

FIG. 6. Plot of the three geographical areas for olive oils on discriminant functions 1 and 2 for ^{1}H NMR data. See Figure 1 for abbreviation.

function, that distinguishes mostly between the coastal and the other geographical areas, was influenced mostly by the peak at 6.75 ppm.

The clearest discrimination was found for samples coming from the northern area using the first discriminant function. This function is marked by negative coefficients for the peak at 9.52 ppm and positive coefficients for the peak at 6.75 ppm. Thus, olive oils from the nothern area typically had lower peak intensity at 6.75 ppm and higher peak intensity at 9.52 ppm. It was not surprising that samples from the coastal and inland areas are not well separated because these two areas are only about 25 km apart. Still, the obtained result can be considered satisfactory with respect to the closeness of the production areas considered in this work. Model stability could be improved with larger data sets.

Despite the limited number of samples and the restricted geographic area considered in this study, some interesting trends were revealed. In particular, analytical parameters seemed to permit the discrimination of extra-virgin olive oils in terms of olive variety. Discrimination of the origin was not so successful. NMR data of phenolic extracts seemed to be more sensitive to the geographical origin of the samples than fatty acid composition data. The results of this study warranted extended work on an enlarged sampling region with other analytical parameters, such as sterol composition to increase the discrimination power of the method. An extension of the application also will be carried out over measurement repetitions in different harvesting years to create a data base of oil samples of Apulia region. This data base also will contain information regarding environmental conditions (such as rainfall). This information will permit the classification of the factors responsible for the variability of the qualitative characteristics of this valuable product of Apulia region.

ACKNOWLEDGMENTS

This research was supported by the European Community activity Large-Scale Facility Wageningen NMR Centre (ERBFMGECT950066).

REFERENCES

1. Forina, M., and E. Tiscornia, Pattern Recognition Methods in the Prediction of Italian Olive Oil Origin by Their Fatty Acid Content, *Ann. Chim. (Rome) 72*:143–155 (1982).
2. Aparicio, R., L. Ferreiro, A. Cert, and A. Lanzon, Caraterización de aceites de oliva vírgenes andaluces, *Grasas Aceites 41*: 23–39 (1990).
3. Tsimidou, M., and K.X. Karakostas, Geographical Classification of Greek Virgin Olive Oil by Nonparametric Multivariate Evaluation of Fatty Acid Composition, *J. Sci. Food Agric. 62*: 253–257 (1993).
4. Sacchi, R., L. Mannina, P. Fiordiponti, P. Barone, L. Paolillo, M. Patumi, and A. Segre, Characterization of Italian Extra Virgin Olive Oils Using ^{1}H NMR Spectroscopy, *J. Agric. Food Chem. 46*:3947–3951 (1998).
5. Shaw, A.D., A. di Camillo, G. Vlahov, A. Jones, G. Bianchi, J. Rowland, and D.B. Kell, Discrimination of the Variety and Region of Origin of Extra Virgin Olive Oils Using ^{13}C NMR and Multivariate Calibration with Variable Reduction, *Anal. Chim. Acta 348*:357–374 (1997).
6. Spugnoli, P., A. Parenti, D. Cardini, G. Modi, and S. Caselli, Caratterizzazione di oli extra vergini monovarietali di tre cultivar toscane, *Riv. Ital. Sostanze Grasse 75*: 227–233 (1998).
7. European Communities, Regulation 2568/91, *Off. J. Eur. Communities*, L 248 (1991).
8. Morrison, W.R., and L.H Smith, Preparation of Fatty Acid

Methyl Esters and Dimethylacetals from Lipids with Boron Fluoride-Methanol, *J. Lipid Res. 64*:373–376 (1964).
9. Massart, D.L., B.G.M. Vandeginste, S.N. Deming, Y. Michotte, and L. Kaufman, *Chemometrics: A Textbook*, Elsevier, New York, 1988, pp. 339–410.
10. Frank, I.E., and R. Todeschini, *The Data Analysis Handbook*, Elsevier Science B.V., Amsterdam, 1994, pp. 90, 153.
11. Solinas, M., L. Di Giovacchino, and A. Mascolo, The Polyphenols of Olives and Olive Oil. Note III: Influence of Temperature and Kneading Time on the Oil Polyphenol Content, *Riv. Ital. Sostanze Grasse 55*:19–23 (1978).
12. Solinas, M., HRGC Analysis of Phenolic Components in Virgin Olive Oil in Relation to the Ripening and the Variety of Olives, *Ibid. 64*:255–262 (1987).
13. Montedoro, G.F., and L. Garofolo, The Qualitative Characteristics of Virgin Olive Oils. The Influence of Variables Such as Variety, Environment, Preservation, Extraction, Conditioning of the Finished Product, *Ibid. 61*:157–168 (1984).
14. Amiot, M.T., A. Fleuriet, and J.T. Macheix, Importance and Evolution of Phenolic Compounds in Olive During Growth and Maturation, *J. Agric. Food Chem. 34*:823–826 (1986).
15. Montedoro, G.F., M. Servili, M. Baldioli, R. Selvaggini, E. Miniati, and A. Macchioni, Simple and Hydrolyzable Compounds in Virgin Olive Oil. 3. Spectroscopic Characterization of the Secoiridoid Derivatives, *Ibid. 41*:2228–2234 (1993).

[Received January 3, 2000; accepted March 20, 2000]