# Determination of Geographical Origin of Olive Oils Using [13]C Nuclear Magnetic Resonance Spectroscopy. I – Classification of Olive Oils of the Puglia Region with Denomination of Protected Origin

Giovanna Vlahov,* Paolo Del Re, and Nicola Simone

Istituto Sperimentale per la Elaiotecnica, Contrada Fonte Umano,
65013 Città S. Angelo (Pescara), Italy

[13]C nuclear magnetic resonance spectroscopy was used to classify olive oils from the three production areas of the Puglia region labeled with the "denomination of protected origin" (DPO) Terra di Bari, Colline di Brindisi, and Dauno. High resolution ^{13}C spectra of 173 olive oil samples were measured, and the intensity data of triacylglycerol resonances were processed by using linear discriminant analysis, which was carried out stepwise for variable selection. The olive oil samples from the DPOs Colline di Brindisi and Terra di Bari were 90% correctly classified, whereas only 74% of "Dauno" DPO oils were classified in the true group. The performance of the discriminant model was verified by applying the cross-validation procedure based on the "leave one out" formalism. The discriminant model was evaluated against a blind test set of olive oils from the three DPO areas. All the oils used for the purpose were correctly assigned to their respective groups, with the exception of the Dauno oil samples based on the Coratina cv. They were misclassified as Terra di Bari oils because of a common monovarietal composition.

KEYWORDS: ^{13}C NMR; LDA Classification; Puglia Region DOP olive oils

## INTRODUCTION

Conventional quantitative ^{13}C NMR was applied jointly with multivariate analysis, to discriminate olive oils by cultivar and geographical origin (*1*). ^{13}C NMR methodology, which applies the distortionless enhancement by polarization transfer (DEPT) pulse sequence, has been successively used to detect protonated carbons at higher sensitivity resulting from spin population transfer from the abundant ^{1}H spins to the dilute ^{13}C spins (*2*). The ^{13}C NMR DEPT methodology, it has been optimized for high field (500 MHz) spectra, was used to discriminate olive oils sampled in the Italian areas labeled with the "denomination of protected origin" (DPO). The principal component analysis (PCA) was carried out to explore the multivariate structure of ^{13}C NMR data (they represent the resonance intensities of ^{13}C nuclei of triglycerides). Olive oils were grouped according to different DPO areas. The main factor for the oil discrimination appeared to be the cultivar more than the geographical origin (*3*).

This result has provided further evidence about the reliability of the ^{13}C NMR DEPT methodology for ensuring that resonance intensity measurements are accurate and precise. This minimizes the oil differences due to measurement errors and emphasizes the oil variability based on different cultivars and/or geographical areas of production.

The problem of monitoring the geographical origin of olive oils is related to the calculation of statistical models for olive oil classification by their production areas. These models should enable the determination of the production area of an unknown oil sample.

Standard statistical classification techniques (linear discriminant analysis and bayesian analysis) (*4*) and a neural network approach (*5*) have been used to classify olive oils from different Italian regions by their fatty acid composition. The prediction ability of the two methods was about 95% (prediction ability of Kohonen neural network was even 100% for certain classes).

Considering the more advanced structural information provided by ^{13}C NMR on the positional distribution of fatty acids in triacylglycerols, olive oil classification was tried by using ^{13}C resonances of triacylglycerols. In particular, linear discriminant analysis (LDA) was applied to the ^{13}C NMR data of olive oils sampled in the DPO areas of the Puglia region, with the purpose of calculating the classification model of the oil productions labeled as Terra di Bari, Colline di Brindisi, and Dauno DPOs.

The goal was to maintain the quality of DPO olive oil productions of the Puglia region (quality being perceived in terms of geographical origin) by adopting ^{13}C NMR as an objective physicochemical method.

## MATERIALS AND METHODS

**Materials.** The olive oils were sampled during the olive oil campaigns from 2000 to 2002 in the three DPO areas of the Puglia Region. The total numbers of olive oils sampled from 2000 to 2002

* Corresponding author. Tel.: +39 (0) 85 95212. Fax: +39 (0) 85 959518. E-mail: vlahov@unich.it.

Puglia DPO Olive Oil Classification

*J. Agric. Food Chem.,* Vol. 51, No. 19, 2003 **5613**

for each DPO area were the following: 99 oil samples from "Terra di Bari" DPO corresponding to the Bari province (predominant cv Coratina), 31 oil samples from "Colline di Brindisi" DPO corresponding to the Brindisi province, (predominant cv Ogliarola), 43 oil samples from "Dauno" DPO corresponding to the Foggia province (cvs Ogliarola Garganica, Peranzana, Coratina).

**$^{13}$C NMR Spectroscopy.** Two hundred milligrams of oil sample were dissolved in 0.5 mL of deuterated chloroform (CDCl$_3$, Sigma Aldrich, Milano, Italia) for NMR analysis. The spectra were recorded on a $^{\text{UNITY}}$ INOVA Narrow Bore 500 MHz NMR Spectrometer equipped with a UNIX-based Sun Microsystems workstation (Varian NMR Instruments, Palo Alto, California).

$^{13}$C spectra were recorded at 25°C by using the DEPT pulse sequence under the constraints (the proton pulse was set at 45 degrees) that all the protonated carbons CH, CH$_2$, and CH$_3$ of the acyl chains were detected as positive resonances. The polarization transfer delay calculated by the coupling constant (1/ 2 $J_{\text{C−H}}$) was optimized to $J_{\text{C−H}}$ = 144 Hz, the $J_{\text{C−H}}$ value was obtained by averaging C−H scalar couplings experimentally determined by the insensitive nuclei enhanced by polarization transfer (INEPT) pulse sequence (*2*). The spectra were measured under proton decoupling by using the Waltz-16 broadband decoupling.

$^{13}$C spectra were acquired with 128K data points, an acquisition time of 2.9 s, and a delay between pulses of 10 s, which was set at four times the value of the longest spin−lattice relaxation time ($T_1$) value for the proton connected to the carbon nuclei to avoid signal saturation. Signal averaging was carried out for 128 transients requiring a total acquisition time of 28 min.

The free induction decays (FIDs) were zero-filled to make the number of data points increase as a power of $2^2$, they were multiplied by a resolution enhancement function to improve resolution and sensitivity of the carbon resonances and finally Fourier transformed. The intensities (i.e., areas) of $^{13}$C resonances of olive oil spectra were determined by integration using the software provided with the NMR spectrometer.

**Data Analysis.** Multivariate analysis of variance (MANOVA) was applied for pairwise and multiple comparison of the means of the oil groups under study by using Minitab Statistical package (MINITAB Release 13, Minitab Ltd., Coventry, United Kingdom).

The data were analyzed by the multivariate method which applies LDA. The Fisher approach to the discrimination problem was used by looking for linear combinations (discriminant functions) of the original variables ($^{13}$C resonance intensities). The linear combination $Z$ of the $X$ original $p$ variables

$$Z = a_1 X_1 + a_2 X_2 + \cdots + a_p X_p$$

can separate the groups if the $Z$ mean value changes considerably from group to group while the $Z$ value within a group remains fairly constant. Hence, the coefficients $a_1, a_2 \cdots a_p$ have to be chosen so as to define the $Z$ linear combination for which the $F$ ratio of the variation between and within groups (for a one-way analysis of variance on $Z$ values), is as large as possible (*6*).

Linear discriminant analysis was carried out stepwise for variable selection by means of Statgraphics Statistical package (STATGRAPHICS Plus 19.0, Statistical Graphics Corporation, Englewood Cliffs, N. J.). The backward elimination procedure was followed for stepwise variable selection (*6*). The procedure removes, one at a time from the discriminant funtions calculated with all the variables, those variables that contribute the least to the olive oil discrimination in correspondence of the lowest $F$ statistic values.

The cross-validation routine that was available on Minitab software was used to check the error rate in misclassifying observations. The procedure works by omitting each observation one at a time, recalculating the classification function by using the remaining data and then classifying the omitted observation.

## RESULTS AND DISCUSSION

**Preselection of Variables for Discriminant Analysis.** The $^{13}$C NMR DEPT spectrum of an olive oil sample reveals the

triacylglycerol resonances. They are grouped into three spectral regions for a total of 49 resonances, which were fully assigned (*3*): unsaturated carbons ranging from 132.1 to 126.8 ppm, glycerol carbons ranging from 69.1 to 61.6 ppm, and aliphatic carbons ranging from 34.5 to 13.9 ppm. The carbonyl carbons resonating in the range from 173.2 to 172.6 ppm are not detected by DEPT pulse sequence (*2*). Assuming that all the olive oil samples differ only in relative amounts of long chain acids of triacylglycerols, and that their absolute resonance intensities are the quantitative parameters for the correlations of oils by geographical areas, the multivariate method using LDA was applied to $^{13}$C NMR data of olive oil samples produced in the three DPO areas of the Puglia region.

Linear discriminant analysis was carried out to explore the possibility of grouping olive oil samples corresponding to the three DPO areas of the Puglia region. The low sampling rates available during two campaigns, for Dauno and Colline di Brindisi DPO oils (they were of 10 and 5 oil samples, respectively) did not enable the evaluation of data variability due to different years of production. Therefore, the olive oils from each DPO area sampled during three years were considered as a whole under the assumption that they could still enable the classification of the three DPO oils.

Because the data of the DPO areas of Colline di Brindisi and Dauno were insufficient to calculate the discriminant functions (only 31 and 43 olive oil samples, respectively, were available for 49 variables), a subset of the original variables was selected by chemical evaluation of resonances and their relationships with the discriminant problem.

Fatty acids of triacylglycerols are the most abundant tracer substances of olive oil cultivars and of the latitude, both cultivar and latitude factors defining the "unique character of a particular oil-producing area". Considering that latitude determines the temperature ranges of the growing areas of olive trees and that the unsaturation degree of fatty acids was strongly dependent on growing temperatures (*7*), the resonances of unsaturated carbons of oleate and linoleate chains and of saturated chain carbons 3, 5, 6, 15, and $\omega$3 ($\omega$3 is the third carbon from the CH$_3$ chain end) were selected. The carbons resonating at 29.72, 29.68, and 29.64 ppm, because of severe overlapping of the mid chain methylene carbons, were tentatively assigned to saturated chains by using integral values (**Table 1**).

The resonances of glycerol backbone carbons of allylic (8, 11) and methyl chain end (12−16) carbons of oleate chain were also selected. Moreover, the chemical relevance of the determination by $^{13}$C NMR of the acyl chain positions on the glycerol backbone ($\alpha$- and $\beta$-positions), which appeared to be an important factor in olive oil discrimination (*3*), suggested to consider the resonances of oleate and linoleate chains (overlapping resonances) at $\alpha(\alpha')$-positions (carbons 2,3,5) and at $\beta$-positions (carbons 2,4,5), in the subset of variables (**Table 1**).

**Linear Discriminant Analysis.** The subset of 33 variables was used to carry out linear discriminant analysis. Eight variables were found to be significant predictors of the three olive oil groups by using the backward elimination procedure for stepwise variable selection. **Table 2** shows the selected variables along with their standardized coefficients.

Two tests of significance were employed in conjunction with discriminant function analysis. The likelihood ratio test based on the Wilks' $\Lambda$ criterion (*8*) gave clear evidence that the overall differences among the means of the three oil groups were statistically significant beyond the 0.001 level. The Hotelling's

**5614** *J. Agric. Food Chem.,* Vol. 51, No. 19, 2003

Vlahov et al.

**Table 1.** ¹³C NMR Spectrum Resonances for Classification of DPO Olive Oils of the Puglia Region (125.7 MHz, CDCl₃)

| peak | chemical shift ppm | acyl chain carbons | peak | chemical shift ppm | acyl chain carbons |
|---|---|---|---|---|---|
| | olefinic | | | O (L) β-pos | |
| 1 | 130.15 | C-13 L [a]α,β[b] | 22 | 34.16 | C-2 |
| 2 | 129.96 | C-10 O β | 23 | 29.20 | C-5 |
| 3 | 129.94 | C-10 O α | 24 | 29.05 | C-4 |
| 4 | 129.91 | C-9 L α | | | |
| 5 | 129.89 | C-9 L β | | | |
| 6 | 129.65 | C-9 O α | | O (L) α-pos | |
| 7 | 129.63 | C-9 O β | 25 | 33.99 | C-2 |
| 8 | 128.07 | C-10 L β | 26 | 29.18 | C-5 |
| 9 | 128.05 | C-10 L α | 27 | 24.84 | C-3 |
| 10 | 127.89 | C-12 L α | | | |
| 11 | 127.88 | C-12 L β | | | |
| | | | | O α (β) pos. | |
| | glycerol | | 28 | 31.92 | C-16 |
| 12 | 68.90 | Gl β | 29 | 29.77 | C-12 |
| 13 | 62.06 | Gl α | 30 | 29.54 | C-14 |
| | | | 31 | 29.34 | C-15,13 |
| | saturated | | 32 | 27.22 | C-11 |
| 14 | 31.94 | ω3[c] | 33 | 27.16 | C-8 |
| 15 | 29.72 | na[d] | | | |
| 16 | 29.68 | na[d] | | | |
| 17 | 29.64 | na[d] | | | |
| 18 | 29.49 | C-6 | | | |
| 19 | 29.38 | C-15 | | | |
| 20 | 29.28 | C-5 | | | |
| 21 | 24.86 | C-3 | | | |

[a] The fatty acids are saturated (S), *cis*-9-octadecenoic acid (O) and *cic,cis*-9,12-octadecadienoic acid (L). [b] The 1(3)- and 2-glycerol positions are indicated by the Greek letters α and β, respectively. [c] ω3 is the third carbon from the CH₃ end. [d] The assignments are only tentative (see text under Preselection of Variables for Discriminant Analysis).
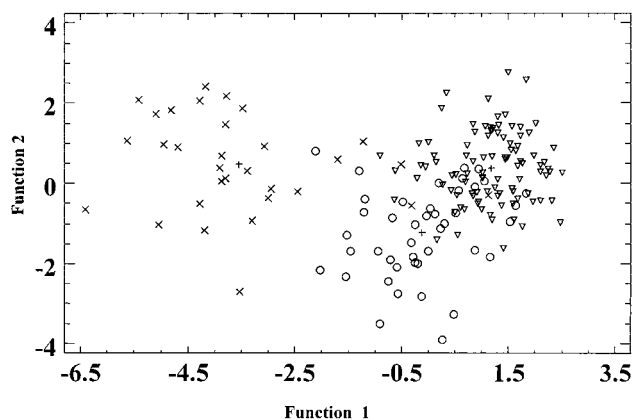
**Table 2.** Linear Discriminant Analysis of ¹³C NMR Data of Olive Oils from the Three DOP Areas of Puglia Region: Standardized Coefficients of Discriminant Functions 1 and 2

| variables | discriminant function 1 | discriminant function 2 |
|---|---|---|
| 4 C−9 L α | 1.06 | −1.11 |
| 5 C−9 L β | −2.33 | 0.71 |
| 14 ω3 S | 0.89 | 4.52 |
| 19 C−15 S | 0.16 | −3.39 |
| 21 C−3 S | −0.66 | −2.26 |
| 26 C−5 O,L α | 2.03 | 1.84 |
| 30 C−14 O α, β | 1.07 | −5.34 |
| 32 C−11 O α, β | −3.19 | 2.68 |

$T^2$ test confirmed that there were significant differences between any pair of oil groups ($P = 0.001$).

Two discriminant functions were calculated, this being the minimum of the number of selected variables (8) and the number of groups (3) minus 1. The significance of the discriminant functions was tested by using the Wilks' Λ criterion. Both discriminant functions were found to be statistically significant beyond the 0.001 level.

**Figure 1** shows the discriminant score plot of the olive oil samples on the discriminant function 1 and 2 along with the centroid values (they represent the average values of each discriminant function in correspondence of the oil groups). The first discriminant dimension separated the oil group of Colline di Brindisi DPO from the oils of Terra di Bari and Dauno groups, the Colline di Brindisi group mean being far from the means of groups Terra di Bari and Dauno. In terms of the second dimension, the oil group of Dauno DPO was not separated by a considerable distance from the groups Colline di Brindisi and



**Figure 1.** Plot of 173 olive oils from the DPO areas of Puglia Region against their values for two discriminant functions: olive oils from Terra di Bari (▽; 1.2, 0.4), Colline di Brindisi (x; −3.5, 0.5), and Dauno (○; −0.1, −1.2) DPO areas. The centroid (+) of each olive oil group was also indicated in parentheses.

**Table 3.** Linear Discriminant Analysis (LDA) of ¹³C NMR Data of Olive Oils from the Three DPO Areas of Puglia Region, Terra di Bari, Colline di Brindisi, and Dauno: Classification Table

| | | assigned group | | | |
|---|---|---|---|---|---|
| true group | group size | Terra di Bari | Colline di Brindisi | Dauno | correct classification % |
| Terra di Bari | 99 | 89[a] | 0 | 10 | 89.9 |
| | | 87[b] | 0 | 12 | 87.9 |
| Collinedi Brindisi | 31 | 2[a] | 28 | 1 | 90.3 |
| | | 3[b] | 27 | 1 | 87.1 |
| Dauno | 43 | 10[a] | 1 | 32 | 74.4 |
| | | 10[b] | 2 | 31 | 72.1 |

[a] Group assignments without cross-validation. [b] Group assignments with cross-validation.

Terra di Bari, the latter two were very close together, thus being practically indistinguishable.

Among the 173 olive oils used to calculate the two discriminant functions, 149 (86.1%) were correctly classified in the true group. The leave-one-out cross-validation was applied to check the percentage of correctly classified oils, as the percentage could be overestimated, considering that the oil samples being classified were also used to build the classification functions.

The summary of classification of the three oil groups without and with cross-validation is reported in **Table 3**. The results obtained by applying the cross-validation routine (the proportions of correct classification were found to be slightly lower) evidenced that 87.9 and 87.1% of the olive oils from the DPO areas Terra di Bari and Colline di Brindisi, respectively, were correctly classified, whereas only 72.1% of the DPO Dauno oils were assigned in the true group.

A closer examination of misclassified observations confirmed that the classification of olive oils from the three DPO areas of the Puglia region appeared to be highly dependent on the cultivar factor. There was evidence that Dauno olive oils (10 samples) were classified as Terra di Bari because of a common cultivar composition based on Coratina variety, whereas Terra di Bari olive oils (10 samples) were classified as Dauno because of prevailing Ogliarola variety in both oil sets.

The standardized discriminant weights, which are reported in **Table 2**, showed that dimension 1 (discriminant function 1) was most highly weighed in the negative direction with the unsaturated chain carbons C-11 O α, β; that is, the oils with

**Table 4.** Acyl Chain Composition of Olive Oil Samples from the DPOs Terra di Bari, Colline di Brindisi and Dauno

| DPO | saturated[a] % | oleate % | linoleate % |
|---|---|---|---|
| Colline Brindisi | $21.6 \pm 0.7$[b] | $67.5 \pm 0.9$ | $10.9 \pm 0.3$ |
| Terra Bari | $17.4 \pm 0.3$ | $74.6 \pm 0.5$ | $8.0 \pm 0.2$ |
| Dauno | $20.2 \pm 0.6$ | $70.2 \pm 0.9$ | $9.6 \pm 0.5$ |

[a] Variables 6–7 (C-9 O $\alpha$–$\beta$), 8–9 (C-10 L $\alpha$–$\beta$), 18 (C-6 S), were used to calculate the acyl chain composition of the three DPOs. [b] Mean $\pm$ 95% confidence limits.

higher oleate chain content (Terra di Bari and Dauno DPOs) scored high on dimension 1, in good agreement with the acyl chain compositions reported in **Table 4**.

However, carbons C-14 O $\alpha$, $\beta$ contributed to the second discriminant function with a weight of $-5.34$ compared to 4.52 for $\omega$3 carbon of saturated chains; that is, the oils with the highest saturated chain content (i.e., the Colline di Brindisi DPO oils) scored high on the second dimension. This result was confirmed by the data of **Table 4**, whereas the oils of the Terra di Bari DPO with the lowest saturated chain concentration (17.4%), scored as high as the Collina di Brindisi oils on the second dimension.

Furthermore, the first dimension seemed most highly weighed in the positive direction, with the unsaturated chain carbons at glycerol $\alpha$- positions, but the lack of resolution of C-14 and C-11 oleate resonances at $\alpha$- and $\beta$- positions could not further confirm this evidence.

Finally, to evaluate the prediction ability of the discriminant model, it was subjected to a blind test using oil samples from the three classes that were not used to calculate the discriminant model. The results evidenced that the prediction ability of the discriminant model was an even 100% (0 errors) for the oil classes Terra di Bari (10 oil samples based on Coratina cv), Colline di Brindisi (5 oil samples based on Ogliarola cv), and for the Dauno oil class (6 oil samples based on Peranzana and Ogliarola cultivars). However, the DPO Dauno oils based on Coratina cv (6 oil samples) were assigned to the group Terra di Bari, with which they shared the same predominant cultivar (Coratina).

## LITERATURE CITED

(1) Shaw, A. D.; Di Camillo, A.; Vlahov, G.; Jones, A.; Bianchi, G.; Rowland, J.; Kell, D. B. 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* **1997**, *348*, 357−374.

(2) Vlahov, G.; Shaw, A. D.; Kell, D. B. Use of ^{13}C nuclear magnetic resonance distortionless enhancement by polarization transfer pulse sequence and multivariate analysis to discriminate olive oil cultivars. *J. Am. Oil Chem. Soc.* **1999**, *76*, 1223−1231.

(3) Vlahov, G.; Schiavone, C.; Simone, N. Quantitative ^{13}C NMR method using DEPT pulse sequence for the determination of the geographical origin (DOP) of olive oils. *Magn. Reson. Chem.* **2001**, *39*, 689−695.

(4) Forina, M.; Triscornia, E. Pattern recognition methods in the prediction of Italian olive oil origin by their fatty acid content. *Annali Chim.* **1982**, *72*, 143−155.

(5) Zupan, J.; Novic, M.; Li, X.; Gasteiger, J. Classification of multicomponent analytical data of olive oils using different neural networks. *Anal. Chim. Acta* **1994**, *292*, 219−234.

(6) Flury, B.; Riedwyl, H. In *Multivariate statistics. A Practical Approach.* Chapman & Hall: London, UK, 1988, pp 68−71.

(7) Lotti, G.; Izzo, R.; Riu, R. Influenza del clima sulla composizione acidica e sterolica degli oli di oliva. *Riv. Soc. Ital. Scienza Alimentazione* **1982**, *2*, 115−126.

(8) Sadocchi, S. In *Manuale di Analisi Statistica Multivariata.* Angeli, F., Ed.; Milano, Italy, 1993, pp 180−185.