Use of ^{13}C Nuclear Magnetic Resonance Distortionless Enhancement by Polarization Transfer Pulse Sequence and Multivariate Analysis to Discriminate Olive Oil Cultivars

Giovanna Vlahov[a,]*, Adrian D. Shaw[b], and Douglas B. Kell[b]

[a]Istituto Sperimentale per la Elaiotecnica, 65013 Città S. Angelo, Pescara, Italy, and [b]Institute of Biological Sciences, University of Wales, Aberystwyth, Ceredigion SY23 3DD, Wales, United Kingdom

ABSTRACT: Distortionless enhancement by polarization transfer (DEPT) pulse sequence was used to set up a quantitative high-resolution ^{13}C nuclear magnetic resonance (NMR) method to discriminate olive oils by cultivars and geographical origin. DEPT pulse sequence enhances the intensity of NMR signals from nuclei of low magnetogyric ratio. The nuclear spin polarization is transferred from spins with large Boltzmann population differences (usually protons) to nuclear species characterized by low Boltzmann factors, e.g., ^{13}C. The signal enhancement of ^{13}C spectra ensures the accuracy of resonance integration, which is a major task when the resonance intensities of different spectra must be compared. The resonances of triglyceride acyl chains $C_{n:0}$, $C_{18:1}$, $C_{18:2}$, and $C_{18:3}$, were also assigned. Multivariate analysis was carried out on the 35 carbon signals obtained. By using variable reduction techniques, coupled with standard statistical methods—partial least squares and principal components analysis—it was largely possible to separate the samples according to their variety and region of origin. With one problem variety removed, 100% prediction of the three remaining varieties was achieved. Similarly, by using the three regions with greatest representation in the data, all but one of a test set of 34 samples were correctly predicted. Thus, the composition of olive oils from different cultivars and of different geographical origin were compared and successfully studied by multivariate analysis. These considerations in conjunction with the structural elucidations of triglyceride molecules demonstrated that ^{13}C NMR is among the most powerful techniques yet described for analysis of olive oils.

Paper no. J8756 in *JAOCS 76,* 1223–1231 (October 1999).

KEY WORDS: Multivariate analysis, ^{13}C NMR DEPT, olive oil.

Spectroscopic methods of analysis—vibrational, infrared (IR), and Raman spectroscopy, including mass spectrometry—coupled with chemometrics (including methods of supervised and unsupervised pattern recognition) are becoming accepted methods of detection of adulteration and of classification by variety and geographic origin of natural triacylglycerol mixtures such as olive oils (1–10). The methods are based on the use of spectroscopic data, e.g., the intensities of IR or Raman bands, which are interpreted by employing chemometrics to extract relevant information for oil classification (authenticity, cultivar, and geographic origin).

Quantitative ^{13}C nuclear magnetic resonance (NMR) spectroscopy has been widely used in analyzing natural triglyceride mixtures. It has proven useful in determining acyl chain compositions and their distribution among the *sn*-1(3) and *sn*-2 glycerol positions in different edible oils (11), and the technique was tentatively employed in the stereospecific analysis of olive oil triacylglycerides (12). Conventional ^{13}C NMR quantitative spectroscopy coupled with statistical methods, including principal components analysis (PCA) and partial least squares (PLS), to interpret spectral data has been used in the classification of extra-virgin olive oil by cultivars and by geographical origins (13,14).

The precautions adopted for running conventional ^{13}C NMR spectra demanding accurate intensity comparisons include the use of relaxation delays between pulses that are long enough to recover all the magnetization before the next pulse is applied, in order to avoid signal saturation. Furthermore, since the intensities of ^{13}C resonances can be distorted by nuclear Overhauser effect (NOE), which is observed during proton decoupling, the spectra were measured by NOE suppressed by using an inverse-gated sequence (15).

These precautions, along with the poor sensitivity of the ^{13}C nucleus owing to its low natural isotopic abundance (1.1%) and low gyromagnetic ratio, reduce substantially the signal-to-noise ratio of ^{13}C spectra and integration accuracy.

The polarization transfer pulse sequences INEPT (insensitive nuclei enhanced by polarization transfer) and DEPT (distortionless enhancement by polarization transfer) are widely used in ^{13}C spectroscopy to detect carbon multiplicity (16). However, their property of transferring the magnetization created originally on the abundant ^{1}H spins to the dilute ^{13}C spins with which they are coupled brings with it the powerful advantage of sensitivity enhancement of ^{13}C nuclei. As a result, the signal intensities are enhanced by a factor proportional to γ_H/γ_C, where γ_H and γ_C are the gyromagnetic ratios of proton and ^{13}C nuclei, respectively. Moreover, the experiment repetition rate in the polarization transfer experiments

*To whom correspondence should be addressed at Istituto Sperimentale per la Elaiotecnica, Contrada Fonte Umano 65013 Città S. Angelo, Pescara, Italy. E-mail: g. vlahov@collaboratore.unich.it

is determined by the longitudinal relaxation times (T1) of ^{1}H nuclei, which are considerably shorter than those of ^{13}C nuclei. Consequently, the pulse delay, which has to be five times the longest T1 in order to avoid signal saturation, becomes significantly shorter, thus lowering spectrum acquisition time.

INEPT pulse sequence gives signal enhancements equivalent to those of the DEPT sequence. Nevertheless, DEPT has been the experiment of choice to optimize quantitative methodology. Because it contains fewer pulses, along with a lower dependence of signal intensities on single-bond C–H couplings, error sources are reduced, thus providing reliable quantitative results (17) .

The aim of this research was to improve quantitative ^{13}C NMR methodology by using the DEPT pulse sequence to ensure signal-to-noise ratios high enough to enable integration accuracy in a shorter time than that used in conventional ^{13}C NMR methodology, i.e., 100 min (14). The methodology was also used to check the actual reliability of ^{13}C NMR and chemometrics in discriminating olive oils by cultivars and geographical origin.

EXPERIMENTAL PROCEDURES

Olive oil sample. Olive fruit samples of Leccino (n = 28), Moraiolo (n = 8), and I-77 (n = 8) cultivars were sampled in different Italian regions; the Dritta (n = 12) cultivar was sampled only in the Abruzzo region. Crushing and pressing of samples in the micro oil mill of the Institute within 24 h harvesting provided the samples with which we worked. The oil samples were stored at 255 K until they were analyzed by NMR spectroscopy. They were prepared for ^{13}C NMR analysis by diluting 200 mg of olive oil with deuterated chloroform (0.5 mL $CDCl_3$). Two replicate measurements were made on all the olive oil samples.

NMR spectroscopy. Spectra were measured using a 300 MHz Bruker spectrometer (Karlsruhe, Germany) equipped with a 5-mm proton/carbon dual probe at 298 K. The DEPT spectra were obtained by using the pulse sequence "DEPT polarization transfer from ^{1}H to X nuclei for refocussed decoupled spectra" as encoded by Bruker:

^{1}H:	D1	- 90 -	D2	-180 -	D2	- P0 -	D2	- BB	DECOUPLE
X (^{13}C):				90		180		FID	

Free induction decays (FID) were acquired at a 3.3-s acquisition time, using 64k points and a spectral width encompassing the olefinic, glyceridic, and aliphatic carbon resonances. The carbonyl carbons are not detected by DEPT pulse sequence since polarization transfer to quaternary carbons is inefficient, the long-range C–H couplings being small compared with $^{1}J_{C–H}$ couplings (15). The spectrum, which was the result of 256 scans, required a total accumulation time of 50 min. The experiment time was therefore reduced by 50% as compared with the time needed to run 256 scans by conventional ^{13}C NMR quantitative methodology that used inverse-gated sequence.

The FID were zero-filled to 128k points, and resolution was enhanced before Fourier transformation. The Gaussian multiplication function was applied by entering the parameters –0.12 Lorentzian narrowing and 0.08 Gaussian broadening, which provided the desired improvement in resolution and sensitivity.

P0 proton pulse, depending on the sought multiplicity selection, was set at 45 degrees to give positive methine, methylene, and methyl carbon signals.

The D1 relaxation delay was optimized, on the basis of T1 longitudinal relaxation times of proton nuclei, to 8 s in order to avoid saturation of the ^{1}H population.

The delay for polarization transfer D2, defined by 1/2 $J_{C–H}$, was optimized to $J_{C–H}$ = 144 Hz, which was obtained by averaging C–H couplings experimentally determined by INEPT pulse sequence.

T1 longitudinal relaxation times of ^{1}H nuclei were measured by means of the inversion–recovery pulse sequence.

Data analysis. PCA was performed using the program SCAN (Minitab Inc., State College, PA) on a PC under Microsoft MS-DOS 6.2 (Redmond, WA). Further data analysis was performed using Microsoft Excel 5 macros written specifically for the purpose of carrying out variable selection, PLS, principal components regression (PCR), PCA (18) and Neural Nets (14). The macros made use of statistical software written in C++, and were run on a PC under Microsoft Windows NT 3.51.

RESULTS AND DISCUSSION

High-resolution ^{13}C spectrum assignment. The aim of this research was to measure the intensities of ^{13}C resonances of triglyceride molecules, which constitute the major portion (98%) of olive oil matrix. As far as the fatty acid profile of triglyceride is concerned, olive oil is rated a "high oleic acid" oil (55.0–83.0%) but containing minor components such as palmitic (7.5–20.0%), stearic (0.5–5.0%), linoleic (3.5–21.0%), and linolenic (≤0.9%) acids as well (19). ^{13}C signals of saturated (no chain differentiation on the basis of chain length is achievable by NMR), oleyl, linoleyl, and linolenyl chains were assigned by using a soybean oil spectrum because this oil, which differs from olive oil in quantitative chain composition rather than in qualitative acid profile, enabled the detection of linolenyl chain. Thus, the DEPT spectrum of the triglyceride fraction of soybean oil provided evidence that ^{13}C resonances were grouped into three well-defined spectral regions: 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 chemical shifts of ^{13}C resonances for the major acyl chains, i.e., saturated ($C_{n:0}$), oleyl ($C_{18:1}$), linoleyl ($C_{18:2}$) and linolenyl ($C_{18:3}$) chains, are reported in Table 1. Double-bond carbons of oleyl, linoleyl, and linolenyl chains were assigned according to chemical shifts previously reported (11).

Double-bond carbon resonances, which were resolved both according to the degree of chain unsaturation and to the chain position on glycerol backbone, evidenced a remarkable

TABLE 1
^{13}C Nuclear Magnetic Resonance Chemical Shifts of Acyl Chains of Soybean Oil Triglycerides

Resonance	Acyl chain carbon[a,b]	Chemical shift (ppm)	Resonance	Acyl chain carbon	Chemical shift (ppm)
			R 24	12 O α + β	29.76
R 1	16 Ln α + β	131.84	R 25	Unresolved peaks	29.68[c]
R 2	9 Ln α	130.13	R 26	Unknown	29.66
R 3	13 L α + β; 9 Ln β	130.10	R 27	Unresolved peaks	29.61[c]
R 4	10 O β	129.95	R 28	14 O α + β	29.53
R 5	10 O α	129.94	R 29	6 S α	29.47
R 6	9 L α	129.90	R 30	15 S α	29.37
R 7	9 L β	129.88	R 31	15 L α + β	29.34
R 8	9 O α	129.65	R 32	13 + 15 O α + β	29.32
R 9	9 O β	129.62	R 33	5 S α	29.27
R 10	13 Ln α + β	128.22	R 34	5 O + L + Ln β	29.18
R 11	12 Ln α	128.18	R 35	5 O + L + Ln α	29.16
R 12	12 Ln β	128.17	R 36	Unresolved peaks	29.10[c]
R 13	10 L β	128.06	R 37	4 O + L + Ln α	29.06
R 14	10 L α	128.04	R 38	4 O + L + Ln β	29.03
R 15	12 L α	127.88	R 39	8 + 11 O α + β	27.17[c]
R 16	12 L β	127.87		8 + 14 L α +β	
R 17	10 Ln β	127.74		8 Ln α +β	
R 18	10 Ln α	127.73	R 40	11 L + Ln α + β	25.56[c]
R 19	15 Ln α + β	127.10		14 Ln α + β	25.51
R 20	β-Glycerol	68.89	R 41	3 O + L + Ln β	24.85
R 21	α-(α′) Glycerol	62.05		3 S + O + L + Ln α	24.82
R 22	2 O + L + Ln β	34.14	R 42	17 S α	22.68
	2 S α	34.00		17 O α + β	22.67
	2 O + L + Ln α	33.98		17 L α + β	22.57
R 23	16 S α	31.93		17 Ln α + β	20.54
	16 O α + β	31.91	R 43	18 Ln α + β	14.25
	16 L α + β	31.52		18 S α 18 O α + β	14.08
				18 L α + β	14.04

[a]1(3)- and 2-positions of glycerol are designated by the Greek symbols α and β, respectively.
[b]Labeling of acyl chains: S, saturated; O, oleyl; L, linoleyl; Ln, linolenyl chain.
[c]Chemical shifts not definitely assigned; they represent the middle frequency of the multiplets.

pattern. The logarithm of shift differences of double-bond carbons in monoenoic ($C_{18:1}$) and nonconjugated polyenoic ($C_{18:2}$, $C_{18:3}$) glycerides were linearly correlated with the number of C–C bonds separating them from the ester group. This relationship was explained by a σ-inductive mechanism of transmission of dipole effects of the ester moiety and multiple bonds through saturated C–C bonds of polymethylene chains (20,21). The chemical shifts of double-bond carbons that were predicted by the σ-inductive theory were in good agreement with the shift values experimentally determined.

The resonances of unsaturated carbons of oleyl, linoleyl, and linolenyl chains were baseline-resolved, except for C-9 of the linolenyl chain and C-13 of the linoleyl chain, which overlapped. Moreover, the C-9 and C-10 signals of all chains and the C-12 resonances of linoleyl and linolenyl chains were split according to the chain 1(3)- and 2-position on the glycerol backbone, whereas no splittings were detected for C-13 of linoleyl and linolenyl chains, and for C-15 and C-16 of linolenyl chain. In particular, in 2-position chains, the carbon of a double-bond pair closer to carbonyl carbon C1 was shifted to a lower frequency compared with that of 1(3)-chains, whereas the double bond carbon further from C1 resonated at a higher frequency than that of 1(3)-chains. Moreover, for a double-bond carbon, the shift differences between the resonances of 1(3)- and 2-position chains become smaller moving away from the ester group toward the methyl end. Thus, the shift separations between C-9 lines at 1(3)- and 2-glycerol positions for oleyl and linoleyl chains were +0.03 and +0.02 ppm, respectively, whereas those of C-10 lines of oleyl, linoleyl, and linolenyl chains, were lower and opposite in sign, i.e., –0.01, –0.02, and –0.01 ppm, respectively. The shift differences of C-12 resonances at the 1(3)- and 2-positions for linoleyl and linolenyl chains were +0.01 ppm, but no shift difference was detected for C-13, C-15, and C-16 of the linolenyl chain.

The appearance of 68.89 and 62.05 ppm signals, which were assigned to C-2 and C-1(3) of glycerol in triglyceride molecules, respectively, confirmed that mono- and diglyceride species were absent in the oils examined, as was expected for high-quality olive oils (22).

C-2 and C-3 signals were resolved according to the chain position on the glycerol backbone whereas no chain differentiation was attainable. C-2 of oleyl, linoleyl, and linolenyl chains at 2- and 1,3-positions resonated as single peaks at 34.14 and 33.98 ppm, respectively. C-2 of saturated fatty acids in 1,3-positions appeared at 34.00 ppm. No signal was

detected for these chains at the 2-glycerol position, of which the maximal acceptable level for high-quality olive oils is ≤1.5% (19).

C-3 resonances of 2-position chains were shifted to a higher frequency (24.85 ppm) from 1,3-chains (24.82 ppm), thus confirming the pattern of C-2 resonances. The assignments of C-2 and C-3 resonances were based on the T1 longitudinal relaxation times, which were found to be slightly shorter for 2-chains, than for 1,3-chains because of their lower mobility and as a consequence a more efficient relaxation process (23).

C-16, C-17, and C-18 carbons appeared as three sets of resonances centered at 31.73, 22.63, and 14.06 ppm, respectively. Within each group of signals, saturated, oleyl, and linoleyl chains resonated from higher to lower frequency, in that order.

The carbons C-16 to C-18 and C-4 to C-15 were assigned by means of T1 longitudinal relaxation times of ^{13}C nuclei which, in fatty acids, were found to increase regularly toward the methyl end of alkyl chains (24).

C-17 and C-18 of linolenyl chain were shifted to higher frequencies from the corresponding signals of the other chains, at 20.54 and 14.26, respectively; they were assigned according to literature data (25). C-17 of the linolenyl chain, in spite of being allylic to the C-15,16 double bond with a shift predicted close to ≈27 ppm, showed a resonance frequency (20.54 ppm) that seemed to be highly influenced by the chain-end methyl group according to chemical shift data for linear alkanes (26).

The chemical shifts of allylic carbons C-8 and C-11 in the oleyl chain, C-8 and C-14 in the linoleyl chain, and C-8 in the linolenyl chain centered at 27.17 ppm, confirmed the *cis* configuration of C-9,10 and C-12,13 double bonds . The low frequency shift by ≈2 ppm (≈27 ppm) from the central methylene shifts (≈29 ppm, unperturbed shift value) which is registered in a *cis*-chain, was produced by a strong γ-gauche interaction occurring between the two allylic carbons (27). The bis-allylic carbons C-11 of linoleyl and linolenyl chains and C-14 of linolenyl chain were further shifted at lower frequency, 25.56 and 25.51 ppm (25), respectively, because a bis-allylic carbon has two γ-gauche interactions with two other carbons which are both allylic to the double bonds.

Quantitative ^{13}C NMR spectra. The success of ^{13}C NMR spectroscopy coupled with statistical methods of multivariate analysis in discriminating olive oils by cultivars and geographical origin is determined by ^{13}C spectra, which have to be quantitative, and by the technique of signal integration.

Quantitative ^{13}C spectra require that all the resonance intensities be rigorously comparable in all the olive oil samples. This means measuring spectra where no signal saturation is allowed to occur, by using relaxation delay of the order of five times the longest T1 of ^{1}H nuclei (15). To fulfill this requirement, we determined ^{1}H T1 longitudinal relaxation times of acyl chains of olive oil triglycerides by the inversion-recovery pulse sequence. They are reported in Table 2. The ^{1}H nuclei relaxed increasingly slowly going from the central glycerol group to the methyl ends of the fatty acids as the mobility became higher and higher from the glycerol skeleton and proceeding along the methylene chain (28).

TABLE 2
T1 Longitudinal Relaxation Times of ^{1}H Nuclei of Acyl Chains of Olive Oil Triglycerides

^{1}H resonances	Chemical shift (ppm)	T1 (s)
Glyceryl residue		
1 (3)	4.30	0.42
1′ (3′)	4.15	0.44
2	5.26	0.79
Acyl residues		
2	2.30	0.53
3	1.57	0.73
4 → 7	1.25	0.86
8	2.00	1.06
9, 10	5.34	1.26
11	2.76	1.30
18	0.87	1.51

Polarization transfer from protons to ^{13}C depends on the D2 delay which is based on the value of C–H coupling constant according to the relationship D2 = 1/2 J_{C-H}. The ^{13}C-^{1}H coupling constants of carbons, which were representative of sp^2 and sp^3 hybridization, were experimentally determined on an olive oil sample by means of INEPT pulse sequence, without refocusing and decoupling. The spectrum, except for the crowded region of the methylene envelope (20–40 ppm), showed the methyl group resonating as a quartet according to the −1:−1:1:1 intensity pattern (J_{C-H} = 124 Hz), the glycerol C-2 as a doublet with an antiphase pair of lines (J_{C-H} = 149 Hz), and the glycerol C-1(3) as a triplet with −1:0:1 structure (J_{C-H} = 149 Hz).

The =C–H coupling of C-9 was determined by the antiphase pair of lines of the corresponding doublet (J_{C-H} = 152 Hz). The $^{1}J_{C-H}$ coupling values are reported in Table 3. Since the resonance intensities of all the ^{13}C nuclei have to be used for the quantitative purposes of this research, and considering that the $^{1}J_{C-H}$ values ranged from 124 to 152 Hz, the D2 delay was optimized by adopting the compromise of setting J_{C-H} = 144 Hz, that is, the average coupling constant C–H.

The intensities, i.e., areas, of ^{13}C resonances of olive oil spectra were determined by integration using the software provided with the spectrometer. Any slope or roll in the base-

TABLE 3
^{13}C–^{1}H Coupling Constants of Acyl Chain Carbons of Olive Oil Triglycerides

^{13}C resonances	Chemical shift (ppm)	$J_{^{13}C-^{1}H}$ (Hz)
Glyceryl residue		
1, 3-	62.05	149 (*t*)
2 -	68.89	149 (*d*)
Acyl residue		
9	129.63	152 (*d*)
18	14.06	124 (*q*)

line was removed prior to integration since integrals are extremely sensitive to small changes in the baseline. The phase, which represents the actual difficulty in achieving integration accuracy, was carefully set to the absorption-mode signal.

The accuracy of the integral increases with the number of data points defining the peak (16); therefore, the number of data points used during the Fourier transformation was doubled by the technique of zero-fill. This improved the resonance definition by reducing the frequency spacing between the points, i.e., digital resolution, from 0.3 to 0.15 Hz. The ^{13}C resonances of triglyceride acyl chains have linewidths of approximately 1.0 Hz, so 0.15 Hz/point appeared to be a suitable digitization to define the lines properly and ensure integration accuracy.

Whenever a group of resonances was poorly resolved and included peaks with low signal-to-noise ratios, the resonances were integrated as a whole rather than measuring the areas of single lines in order to minimize integration errors. This strategy was adopted for the resonances labeled R3, R22, R23, R25, R27, R36, R39–R43, whose components are reported in detail in Table 1. In considering that in all olive oil samples the linolenyl chain was not detected under the signal-to-noise ratio values adopted, 35 intensities of ^{13}C nuclei were measured resonating at the same chemical shifts in all the spectra examined.

Multivariate statistical analysis. The optimal number of variables for the formation of statistical models is often not the total number of variables available (14,29–35) because many variables contain noise, whereas others are collinear and essentially contain the same information. The principle of parsimony (36) in particular bears this out in statistical theory. This principle states that if a simpler model is equally as good as a more complex model for a given set of data (training set), then it will tend to be better when applied to a previously unseen set of data (test set). For this reason, it is important to find the best variables for prediction before proceeding.

PLS was carried out using the macros (described in more detail in Ref. 14), which encode targets as 1 and 0. For example, a simple PLS2 target prediction (Y) matrix may look as shown in Scheme 1. A prediction is taken to be for a particular target type (variety in this case) if it is the highest number in the row, so that an output matrix may look like Scheme 2.

		Encoding		
		Dritta (Dr)	Leccino (Le)	Moraiolo (Mo)
	Dr1	1	0	0
	Dr2	1	0	0
Sample	Dr3	1	0	0
Index	Le1	0	1	0
	Le2	0	1	0
	Mo1	0	0	1
	Mo2	0	0	1

SCHEME 1

		Prediction			
		Dritta	Leccino	Moraiolo	Identified as
	Dr1	0.623	0.42	0.012	Dritta
Sample	Dr2	0.978	0.014	0.001	Dritta
Index	Le1	0.024	0.859	0.001	Leccino
	Le2	0.287	0.325	0.337	Moraiolo

SCHEME 2

Test and training sets were chosen in such a way that the training set encompassed the test set as far as possible. That is, extreme objects were placed in the training set. Half of the objects (or half + 1, where the number of total objects is odd) for each class (variety or region) were placed in the training set. This was achieved by using a program written by Dr. Alun Jones, called MultiPlex, an in-house extension of the Duplex method (37). For a PLS2 prediction, using weighted characteristicity *w* variable selection (14) (essentially the ratio of the inner variance to the outer variance of the object classes—in this case varieties—see Ref. 38), the best prediction is obtained with just 15 of the variables, using 8 PLS factors (Fig. 1).

Figure 1 was produced by using results from the test set only. The test set was previously unseen to the model, except that it was used, along with the training set, in determining the optimal number of factors for prediction.

A three-dimensional PCA scores plot at this point (Fig. 2) showed that all the oils of variety Dritta cluster very tightly. The oils of variety I-77 were well distinguished in the lower right, whereas the Leccino oils were spread around, and bisected with a line of Moraiolo oils. Our previous studies indicated that the variety Leccino is more difficult to predict, a fact associated with the effect of age in this particular variety. It is the most rustic of Italian varieties, and cannot self-pollinate (39). It is also widespread throughout Italy; consequently, it is likely that there is considerable genetic variation.

A similar pattern can be seen in the PLS2 scores plots for the 15 best variables (Fig. 3). Closer analysis of these PLS2 results yields a confusion matrix (Scheme 3). This shows a total of 8 incorrectly identified oils from a test set of 56 oils, or a correct prediction of 85.7%.

Examination of the scores plots suggested that linear separation of the varieties would be possible if the samples of va-

		Identified as			
	Confusion	Dritta	I-77	Leccino	Moraiolo
	Dritta	10	0	2	0
Actual	I-77	0	7	1	0
	Leccino	1	0	27	0
	Moraiolo	1	0	3	4

SCHEME 3

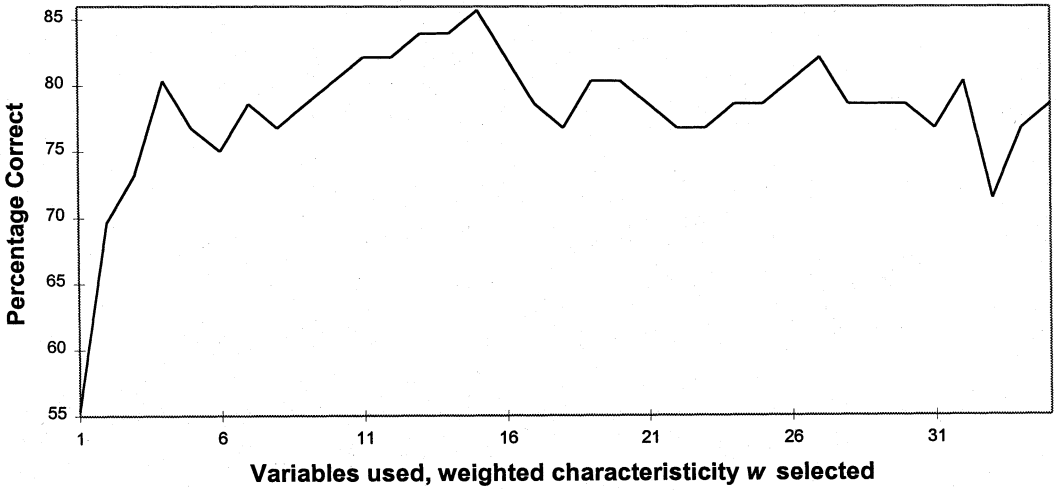

FIG. 1. Partial least squares 2 (PLS2) prediction of variety of olive oil. Percentage correct for test set only for models from the best 1 to all 35 variables. Graph shows that optimal prediction is achieved using the best 15 variables.

riety Leccino were removed from the data, and this was indeed found to be the case. The 15 best variables without this variety provided a 100% prediction for all three varieties. The scores plot (18) is shown in Figure 4.

For any kind of discrimination, it is necessary to have a reasonable representation from each object type in order to form a good model. In this investigation, where regional discrimination is involved, this means using the largest three regions (i.e., Toscana, Abruzzo, and Puglia). These three regions are geographically separated, and so the oils from these regions would be expected to be well-differentiated. Among these three regions, there are two I-77 representatives, and be-

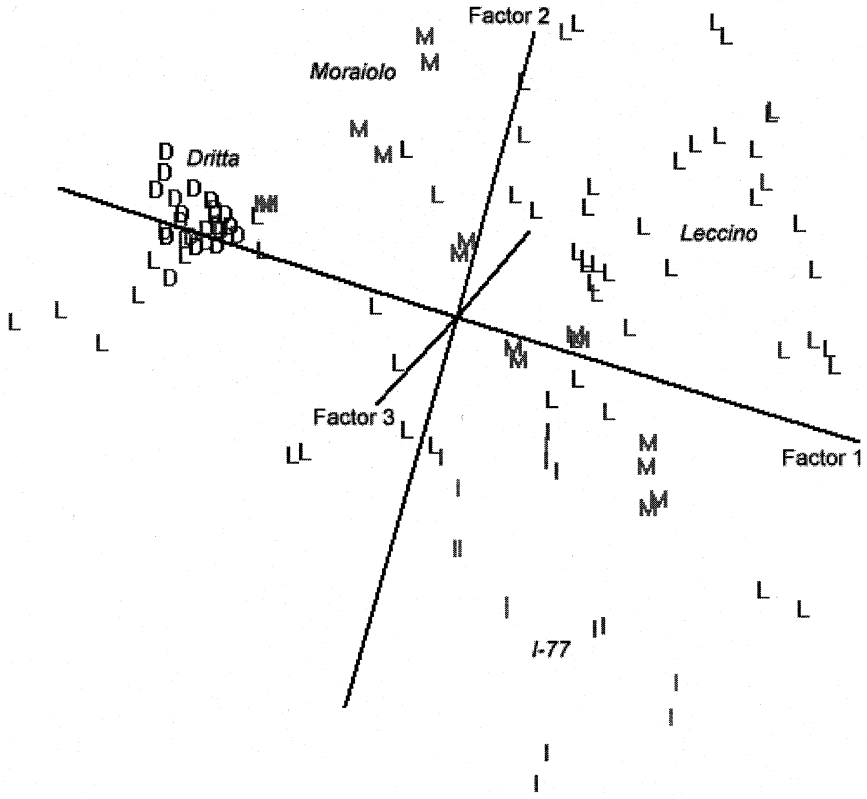

FIG. 2. Principal components analysis three-dimensional scores plot of olive varieties. Fifteen best variables, as selected by characteristicity *w* weighted in favor of the larger varieties. Graph shows elements of factors 1, 2, and 3. Dritta (D) variety clusters very tightly, I-77 (I) is distinguishable, and Leccino (L) is diffuse. Moraiolo (M) oils intersect the Leccino oils.

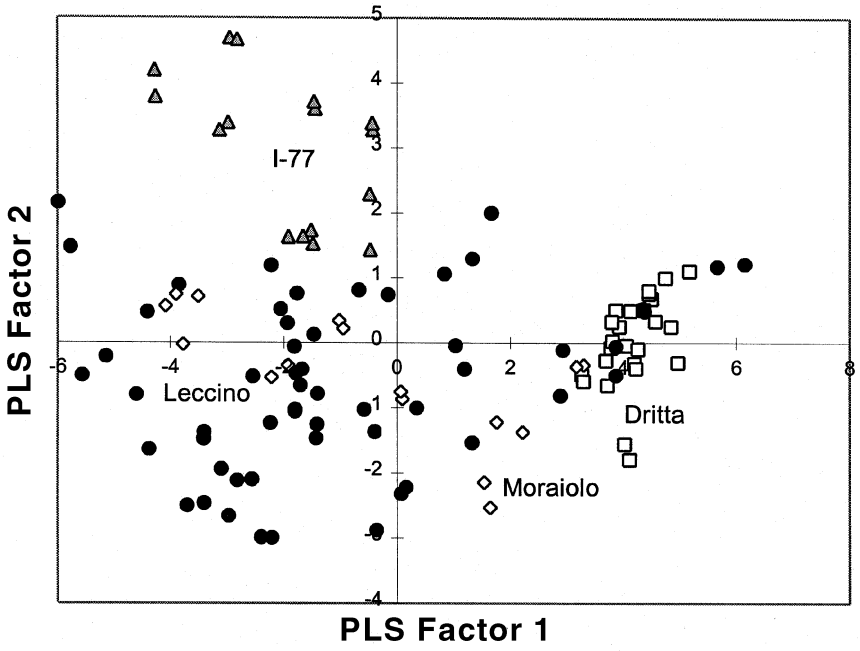

FIG. 3. PLS2 scores plot of olive varieties. Fifteen best variables, as selected by characteristicity *w* weighted in favor of the larger varieties. For abbreviation see Figure 1.

tween 8 and 12 Leccino representatives from each region. The variety Moraiolo is found only in Tuscany, and Dritta only in Abruzzo in these data.

The results show that, like the ^{13}C NMR data presented before (14), olive oils from different regions are not so easy to predict as varieties. Figure 5 shows that a PLS2 model with these three regions is able to achieve an all-but-one prediction.

The models at the 31 and 32 best variables both selected five factors as being optimal. In both instances the wrong sample was an Abruzzo which was falsely identified as a Toscana. The training set also showed two samples incorrectly identified; one, another Abruzzo identified as a Toscana, the other a Puglia identified as an Abruzzo. In the scores chart in Figure 6 these three samples are identified by an arrow. The two Abruzzo oils incorrectly identified were the only two from that region of variety I-77 (the remainder

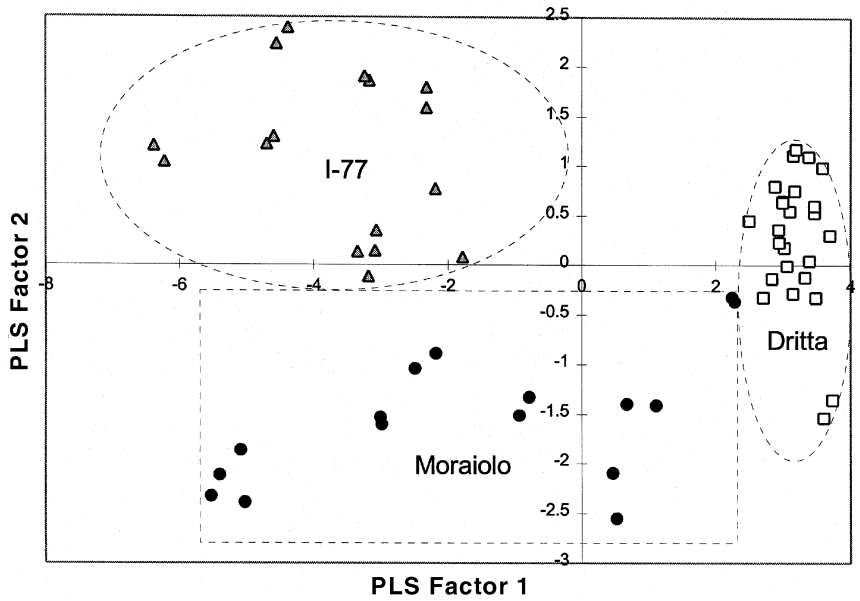

FIG. 4. PLS2 scores plot of olive varieties Moraiolo, Dritta, and I-77 only. Fifteen best variables, as selected by characteristicity *w* weighted in favor of the larger varieties. For abbreviation see Figure 1.

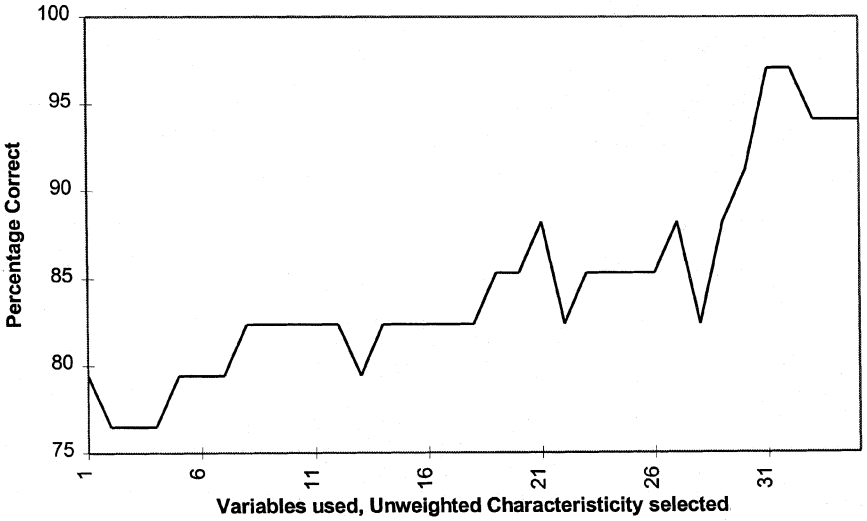

FIG. 5. PLS2 prediction of region of origin of olive oil. Percentage correct for test set only for models from the best 1 variable to all 35. Graph shows that optimal prediction is achieved with the best 31 or 32 variables. For abbreviation see Figure 1.

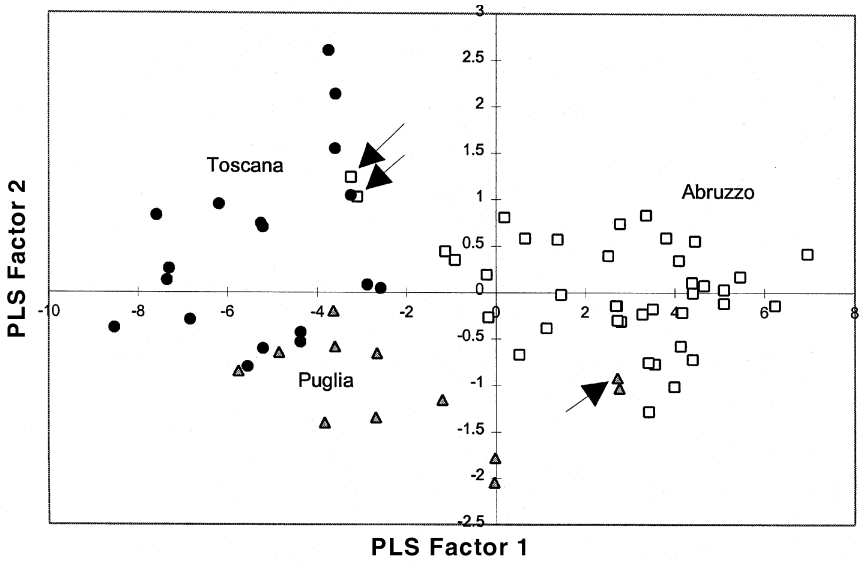

FIG. 6. PLS2 scores plot of olive oil regions of origin, Abruzzo, Puglia, and Tuscany only. Thirty-one best variables, as selected by characteristicity *w* unweighted. Three samples incorrectly identified are indicated with arrows. For abbreviation see Figure 1.

being all Dritta and Leccino oils). The incorrect Puglia oil was a Leccino, for which there is no easy explanation beyond the known variability of that variety.

These results compare very favorably with our previous ^{13}C NMR results, despite the fact that DEPT NMR is unable to give signals from the carbonyl region, which were found to be very important in our previous work (13,14). These findings confirm that Fourier transform NMR can be used as a quantitative technique. The repeatability of the methodology, i.e., within-run (short term) precision, and the reproducibility, i.e., between-run (very long term) precision, enable rigorous quantitative analyses of olive oils.

ACKNOWLEDGMENTS

Dr. Alun Jones (c/o Adrian Shaw, Institute of Biological Sciences, University of Wales-Aberystwyth), was the author of the C++ statistical software. This work was supported by research grants from Ispettorato Centrale Repressione Frodi (ICRF), Ministero per le Politiche Agricole (MIPA), Italy.

REFERENCES

1. Goodacre, R., D.B. Kell, and G. Bianchi, Rapid Assessment of the Adulteration of Virgin Olive Oils by Other Seed Oils Using Pyrolysis Mass Spectrometry and Artificial Neural Networks Developments in the Detection of Adulteration of Olive Oil, *J. Sci. Food Agric. 63*:297–307 (1993).

2. Lai, Y.W., E.K. Kemsley, and R.H. Wilson, Potential of Fourier Transform Infrared Spectroscopy for the Authentication of Vegetable Oils, *J. Agric. Food Chem. 42*:1154–1159 (1994).
3. Li-Chan, E., Developments in the Detection of Adulteration of Olive Oil, *Trends Food Sci. Tech. 5*:3–11 (1994).
4. Van de Voort, F.R., FTIR Spectroscopy in Edible Oil Analysis, *INFORM 5*:1038–1042 (1994).
5. Zamora, R., J.L. Navarro, and F.J. Hidalgo, Identification and Classification of Olive Oils by High-Resolution C-13 Nuclear Magnetic Resonance, *J. Am. Oil Chem. Soc. 71*:361–364 (1994).
6. Lai, Y.W., E.K. Kemsley, and R.H. Wilson, Quantitative Analysis of Potential Adulterants of Extra Virgin Olive Oil Using Infrared Spectroscopy, *Food Chem. 53*:95–98 (1995).
7. Wesley, I.J., R.J. Barnes, and A.E.J. McGill, Measurement of Adulteration of Olive Oils by Near-Infrared Spectroscopy, *J. Am. Oil Chem. Soc. 72*:289–292 (1995).
8. Sacchi, R., M. Patumi, G. Fontanazza, P. Barone, P. Fiordiponti, L. Mannina, E. Rossi, and A.L. Segre, A High-Field ^{1}H Nuclear Magnetic Resonance Study of the Minor Components in Virgin Olive Oils, *Ibid. 73*:747–758 (1996).
9. Segre, A.L., L. Mannina, P. Barone, and R. Sacchi, Quality and Geographical Origin of Virgin Olive Oil as Determined by High-Field ^{1}H-NMR, *Bruker Report 143*:27–28 (1996).
10. Wesley, I.J., F. Pachero, and A.E.J. McGill, Identification of Adulterants in Olive Oils, *J. Am. Oil Chem. Soc. 73*:515–518 (1996).
11. Wollenberg, K.F., Quantitative High-Resolution C-13 Nuclear Magnetic Resonance of the Olefinic and Carbonyl Carbons of Edible Vegetable Oils, *Ibid. 67*:487–494 (1990).
12. Vlahov, G., The Structure of Triglycerides of Monovarietal Olive Oils: A ^{13}C-NMR Comparative Study, *Fett/Lipid 98*:203–205 (1996).
13. Shaw, A.D., A. Di Camillo, G. Vlahov, A. Jones, G. Bianchi, J. Rowland, and D.B. Kell, Discrimination of Different Olive Oils using ^{13}C NMR and Variable Reduction, in *Proceedings of the Conference on Food Authenticity '96*, Norwich, 1996.
14. 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).
15. Freeman, R., *A Handbook of Nuclear Magnetic Resonance*, Longman Group, Oxford, 1988.
16. Derome, A.E., *Modern NMR Techniques for Chemistry Research*, Pergamon Press Limited, Oxford, 1991.
17. Breitmaier, E., and W. Voelter, *Carbon-13 NMR Spectroscopy,* VCH, Weinheim, 1989.
18. Martens, H., and T. Næs, *Multivariate Calibration*, Wiley & Sons Ltd., Chichester, 1989.
19. IOOC, *International Trade Standards Applying to Olive Oils and Olive-Residue Oils (COI/T. 15/NC n2/Rev. 6),* International Olive Oil Council, 1997.
20. Bianchi, G., O.W. Howarth, C.J. Samuel, and G. Vlahov, Long-Range Sigma-Inductive Interactions Through Saturated C–C Bonds in Polymethylene Chains, *J. Chem. Soc., Perkin Trans. 2*:1427–1432 (1995).
21. Howarth, O.W., C.J. Samuel, and G. Vlahov, The Sigma-Inductive Effects of C=C and C–C Bonds—Predictability of NMR Shifts at sp^2 Carbon in Nonconjugated Polyenoic Acids, Esters and Glycerides, *Ibid. 2*:2307–2310 (1995).
22. Vlahov, G., Improved Quantitative C-13 Nuclear Magnetic Resonance Criteria for Determination of Grades of Virgin Olive Oils. The Normal Ranges for Diglycerides in Olive Oil, *J. Am. Oil Chem. Soc. 73*:1201–1203 (1996).
23. Howarth, O.W., and G. Vlahov, ^{13}C Nuclear Magnetic Resonance Study of Cyclopropenoid Triacylglycerols, *Chem. Phys. Lipids 81*:81–85 (1996).
24. Bengsch, E., B. Perly, C. Deleuze, and A. Valero, A General Rule for the Assignment of the C-13 NMR Peaks in Fatty-Acid Chains, *J. Magn. Res. 68*:1–13 (1986).
25. Gunstone, F.D., M.R. Pollard, C.M. Scrimgeour, and H.S. Vedanayagam, Fatty Acids: Part 50. ^{13}C Nuclear Magnetic Resonance Studies of Olefinic Fatty Acids and Esters, *Chem. Phys. Lipids 18*:115–129 (1977).
26. Grant, D.M., and E.G. Paul, Carbon-13 Magnetic Resonance. II. Chemical Shift Data for the Alkanes, *J. Am. Chem. Soc. 86*:2984–2990 (1964).
27. Batchelor, J.G., R.J. Cushley, and J.H. Prestegard, Carbon-13 Fourier Transform Nuclear Magnetic Resonance. VIII. Role of Steric and Electric Field Effects in Fatty Acid Spectra, *J. Org. Chem. 39*:1698–1705 (1974).
28. Breitmaier, E., K.H. Spohn, and S. Berger, ^{13}C Spin-Lattice Relaxation Times and the Mobility of Organic Molecules in Solution, *Angew. Chem. Internat. Ed. 14*:144–159 (1975).
29. Seasholtz, M.B., and B. Kowalski, The Parsimony Principle Applied to Multivariate Calibration, *Anal. Chim. Acta 277*:165–177 (1993).
30. Chattergee, S.K., and S.K. Samanta, An Alternative Approach to Variable Selection for Prediction, *Commun. Stat. Theory Methods 23*:2157–2174 (1994).
31. Heikka, R., P. Minkkinen, and V.M. Taavitsainen, Comparison of Variable Selection and Regression Methods in Multivariate Calibration of a Process Analyzer, *Proc. Cont. Qual. 6*:47–54 (1994).
32. Lindgren, F., P. Geladi, A. Berglund, M. Sjostrom, and S. Wold, Interactive Variable Selection (IVS) for PLS2. Chemical Application, *J. Chemom. 9*:331–342 (1995).
33. Miller, C.E., The Use of Chemometric Techniques in Process Analytical Method Development and Operation, *Chem. Int. Lab. Sys. 30*:11–22 (1995).
34. Kubinyi, H., Evolutionary Variable Selection in Regression and PLS Analyses, *J. Chemom. 10*:119–133 (1996).
35. Broadhurst, D., R. Goodacre, A. Jones, J.J. Rowland, and D.B. Kell, Genetic Algorithms as a Method for Variable Selection in Multiple Linear Regression and Partial Least Squares Regression, with Applications to Pyrolysis Mass Spectrometry, *Anal. Chim. Acta 348*:71–86 (1997).
36. Seasholtz, M.B., and B. Kowalski, The Parsimony Principle Applied to Multivariate Calibration, *Ibid. 277*:165–177 (1993).
37. Snee, R.D., Validation of Regression Models: Methods and Examples, *Technometrics 19*:415–428 (1977).
38. Eshuis, W., P.G. Kistemaker, and H.L.C. Meuzelaar, Some Numerical Aspects of Reproducibility and Specificity, in *Analytical Pyrolysis*, edited by C.E.R. Jones and C.A. Cramers, Elsevier, Amsterdam, 1977.
39. Boschelle, O., A. Giomo, L. Conte, and G. Lercker, Caratterizzazione della Cultivar di Olivo del Golfo di Trieste Mediante Metodi Chemiometrici Applicati ai Dati Chimica-Fisici, *Riv. Ital. Sostanze Grasse 71*:57–65 (1994).

[Received January 12, 1998; accepted May 25, 1999]