JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

FTIR Spectroscopy and Multivariate Analysis Can Distinguish the Geographic Origin of Extra Virgin Olive Oils

HENRI S. TAPP, MARIANNE DEFERNEZ, AND E. KATHERINE KEMSLEY*

Institute of Food Research, Norwich Research Park, Colney Lane, Norwich NR4 7UA, United Kingdom

This work investigates whether Fourier transform infrared spectroscopy (FTIR), in combination with multivariate analysis, can distinguish extra virgin olive oils from different producing countries. Duplicate spectra were collected from 60 oils from four European countries. Two approaches to data analysis were used as follows: first, the "whole spectrum" method of partial least squares (PLS) followed by distance-based linear discriminant analysis (LDA) applied to the PLS scores, and second, a genetic algorithm (GA) for variate selection from the raw data, followed by LDA applied to the selected subset. The PLS−LDA approach produced a cross-validation success rate of 96%, whereas the GA−LDA approach achieved a 100% cross-validation success rate, from subsets comprising only eight variates. Neither the selected variate nor the whole spectrum approach was able to offer insight into the origin of the discrimination in biochemical terms. However, FTIR analysis is rapid, and this work shows that it has the required discriminatory power to potentially offer a "black box" method of screening oils to verify their country of origin.

KEYWORDS: Spectroscopy; infrared; olive oil; country of origin; multivariate analysis

INTRODUCTION

Olive oil is produced from the fruit of the *Olea europaea* tree, which is cultivated extensively across the Mediterranean basin. It is known that the composition of olive oils varies with geographic origin (*1*, *2*), due to a number of different factors: regional differences in climate, soil, and agricultural practice, as well as the subvariety that is grown—different cultivars tend to be grown in different countries. However, it is also known that the major and minor components of olive oil are each influenced differently by these factors; hence, the overall relationship between origin and composition is subtle and may not remain consistent across more than one harvest.

Verifying the declared origin, or determining the origin of an unidentified olive oil, is, therefore, a challenging problem. There have been many studies attempting to elucidate the link between composition and geographic origin using a variety of analytical techniques, in general coupled with multivariate analysis. Chromatographic techniques (*3*) (gas chromatography, GC; with mass spectrometry, GC-MS; and high-performance liquid chromatography, HPLC) have been extensively used to quantify fatty acids (*4*−*6*), as well as minor compound classes such as sterols (*7*), phenolics (*8*, *9*), and hydrocarbons (*10*, *11*). Stefanoudaki et al. (*12*) used HPLC to obtain triglyceride composition from two cultivars, each grown in more than one geographical location; multivariate analysis separated the data according to cultivar and, within each cultivar, showed grouping according to region of origin. Bucci et al. (*13*) were also able to discriminate between cultivars using fatty acid compositional data. However, Cinquanta et al. (*14*) used HPLC to analyze phenolic compounds in a selection of cultivars and suggested that ripeness, soil, and climate had a greater influence than cultivar on the phenolic content.

In recent years, a number of studies have used high-field nuclear magnetic resonance (NMR) with the aim of distinguishing oils of different origin. Vlahov et al. (*15*) used carbon-13 NMR spectra to differentiate between monovarietal olive oils from different regions of Italy; the discrimination was attributed to differences in fatty acid composition. Similar differentiation was obtained by Mannina et al. (*16*), using hydrogen-1 NMR.

Another molecular spectroscopy technique, Fourier transform infrared (FTIR), has been shown to be useful for the analysis of olive oils. It has been used to detect adulteration with other vegetable oils (*17*−*20*) and to quantify free fatty acids (*21*). However, the compositional variances of interest in these applications are comparatively large, often manifest in the raw spectral data. There is no precedent for the use of FTIR spectroscopy to discriminate between oils on the basis of their country of origin. The main purpose of the work presented in this paper is to ascertain whether FTIR, in combination with multivariate analysis, has the sensitivity needed to detect such small systematic differences in composition.

The major advantage of FTIR over NMR or the chromatographic approaches is that no complex sample preparation (extraction, separation, derivatization) is required. The FTIR analysis described here is relatively fast: the complete protocol

* To whom correspondence should be addressed. E-mail: kate.kemsley@bbsrc.ac.uk.

10.1021/jf030232s CCC: $25.00
Published on Web 09/12/2003

Table 1. Numbers of Extra Virgin Olive Oils from Each Country of Each Origin, along with Cultivar Information

group designation	country of origin	no. of samples	variety or varieties
1	Greece	10	Koroneiki
2	Italy	17	Coratina (× 2)
			Coratina–Ogliarola blend (× 7)
			Coratina–Ogliarola–Garganica blend
			Leccino–Ogliarola–Garganica blend
			Frantoio–Ogliarola–Garganica blend
			Ogliarola–Garganica blend
			Ogliarola–Salentina blend (× 4)
3	Portugal	8	Cobrancosa
			Galega
			Cobrancosa–Galega blend
			Cobrancosa–Picual–Blanqueta blend
			Picual
			Arbequina
			Cordovil–Galega–Verdeal blend
			Macanilha–Carrasquenha–Almendralejo blend
4	Spain	25	Hojiblanca (× 14)
			Arbequina (× 2)
			Cornicabra
			unspecified blend (× 8)
	total:	60	

(comprising acquisition of background and sample spectra, cleaning the sampling accessory, and acquisition of spectra to verify cleanliness) takes around 20 min per sample. Provided it has the required discriminatory power, FTIR potentially offers a means of rapid screening of olive oils for verifying their country of origin.

MATERIALS AND METHODS

Samples. Sixty authenticated samples of virgin olive oils, originating from four European producing countries, were obtained from the International Olive Oil Council in Madrid. **Table 1** details the numbers of samples from each country, along with some additional information that was available on the cultivars used.

Instrumentation. All spectra were collected on a Spectra-Tech Applied Systems Inc. (Shelton, CT) MonitIR FTIR spectrometer system, equipped with a sealed and desiccated interferometer, a room temperature deuterated triglycine sulfate detector, and an attenuated total reflectance (ATR) accessory, built into one of two dedicated sampling stations. The ATR crystal was a nominal 11-reflection zinc selenide crystal mounted in a trough plate. The crystal geometry was 45° parallelogram with mirrored angled faces.

Spectral Acquisition. Data were acquired in two discrete periods, each of around 2 weeks duration. In each period, an absorbance spectrum was collected of each sample, with the order in which samples were presented to the spectrometer randomized with respect to the country of origin. Concatenating the data from both acquisition periods gave a data set of 120 absorbance spectra in total (i.e., duplicate acquisitions from each of the 60 samples). Because of the random orders in which samples were analyzed during each period, the elapsed time between duplicate acquisitions was different for each sample, ranging between 1 and 24 days (mean interval = 15 days). Before and between spectral acquisitions, samples were stored in the dark at ambient temperature. The spectrometer was sited in an air-conditioned room (21 °C), and samples were allowed to equilibrate to this temperature immediately before analysis.

To obtain each absorbance spectrum, a background single-beam spectrum of the clean, dry ATR crystal was first collected. The spectral range was 800−4000 cm^{-1}, the nominal resolution was 4 cm^{-1}, and 256 interferograms were coadded before Fourier transformation using triangular apodization. Immediately following collection of each background, approximately 5 mL of sample was applied to the ATR crystal using a transfer pipet, ensuring that no air bubbles were trapped on the crystal surface. A single-beam spectrum of the sample was collected with the same acquisition conditions and converted to an absorbance spectrum using the background.

The ATR plate was cleaned in situ by scrubbing with 0.2% Triton-X 100 solution followed by ethanol and allowed to dry. Cleanliness was verified by collecting an absorbance spectrum of the crystal, using the most recently collected background as a reference. The cleaning procedure was repeated as necessary until this absorbance spectrum was found to contain substantially only noise, at which point the crystal was deemed to be clean and ready to acquire a new background to be used for the subsequent sample.

Chemometric Analysis. All data analysis was carried out using Matlab (The Mathworks Inc., Cambridge, U.K.). All absorbance spectra were truncated to 570 data points in the "fingerprint" region, 799−1897 cm^{-1}, to give a data matrix of dimensions [120 × 570] for analysis.

A selection of multivariate modeling methods was applied to the data, with the aim of discriminating between the four different countries of origin. The first of these comprised the "whole spectrum" method of partial least squares (PLS) followed by distance-based linear discriminant analysis (LDA) using the Mahalanobis metric (PLS−LDA, as described in Al-Jowder et al. (*22*)). The PLS regression step, onto a set of dummy variates encoded to represent the four groups of samples, compresses the original data into a smaller number of "scores", while tailoring them to the subsequent LDA step. A range of subsets comprising different numbers of PLS scores are then passed to the LDA, to obtain a classification model. To ensure a realistic impression of model performance, internal cross-validation was used throughout, using a cross-validation segment size of two, corresponding to the duplicate spectra from each sample (for any given model dimensionality, the routine was repeated 120/2 = 60 times, with each pair of spectra acting as the test items in turn; this approach to cross-validation has been used throughout the work in this paper). Classification success rates were evaluated for the test items only. The optimum model dimensionality was identified as that with the maximum classification success rate by cross-validation. This model was approximated graphically by a similarly cross-validated canonical variate analysis (CVA, *23*) applied to PLS scores of the optimal subset size.

The second approach to data analysis comprised variate selection, to identify a low-dimensional subset of raw variates to pass to the cross-validated LDA routine. Variate selection algorithms have long been used as precursors to multivariate analysis of spectroscopic data, in particular before the advent of whole spectrum chemometric methods. A difficulty is that the total number of variates d to choose from is very large: $d = 570$ in the present work; hence, the number of possible subsets of size r is astronomical (with $r = 8$, say, ${}^{d}C_r = d!/(r!(d - r)!) \sim 2.6 \times 10^{17}$). However, in recent years, attention has turned back to variate selection approaches, and some studies have shown that they offer better performance than whole spectrum methods (*24*, *25*). Genetic algorithms (GAs) have been found to be effective at selecting variables from high-dimensional data sets (*26*−*28*). In the present work, we have written a GA to search for a small subset of variates to pass to cross-validated LDA. The criterion for termination of GA evolution was either attainment of a 100% classification success rate or no further improvement in success rate for 10 generations. The complete routine was carried out repeatedly, because the random nature of the initial "chromosomes" (variate identifiers), as well as the possibility of convergence on local optima, can affect the outcome of any single GA evolution. The best solution from each execution of the GA was retained. The frequency of occurrence of each variate calculated across all retained solutions was interpreted as an indication of its usefulness as a discriminator.

RESULTS AND DISCUSSION

All 120 spectra are shown in **Figure 1a**. Spectral quality is high; instrumental (detector) noise is not visible on the baseline, and the amount of baseline shift does not indicate baseline correction to be essential. However, with regard to differentiating between sample types, it is immediately clear that little can be gained from examining the raw spectral data, which appear

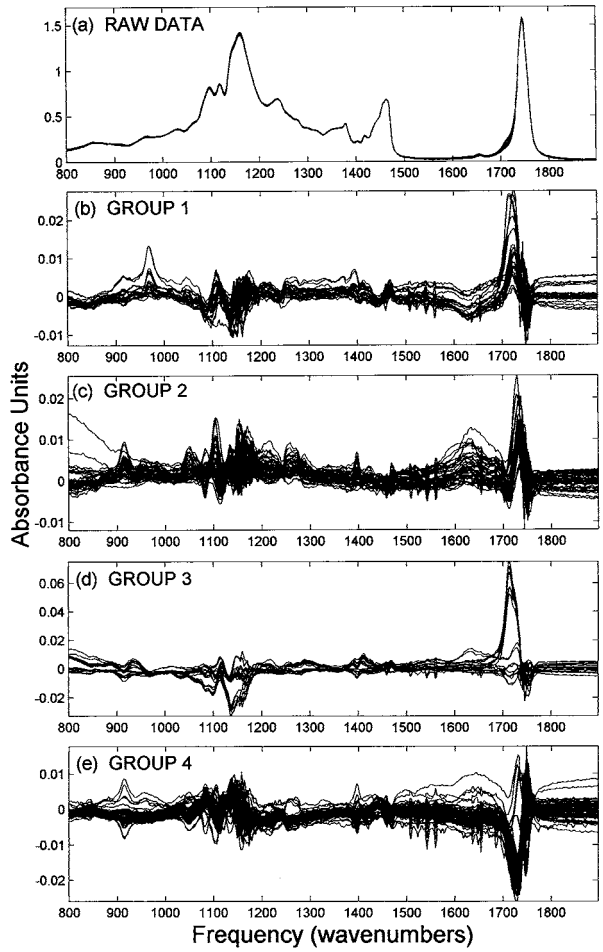

Figure 1. (**a**) Raw spectral data; (**b**–**e**) centered data set, shown separately for each of the four groups.

virtually indistinguishable. The spectra are dominated by absorptions arising from triglycerides, which form the major component of olive (and other vegetable) oils. In the presence of these strong absorptions, it is very difficult to see more subtle spectral contributions arising from, for example, differences in the fatty acid composition or from the nonglyceridic minor components.

Figure 1b−**e** shows the "centered" data (obtained by subtracting the mean of all 120 spectra from each individual spectrum) grouped according to sample origin. It is immediately clear that there are systematic spectral features present, indicating that FTIR is indeed sensitive to compositional differences between extra virgin olive oils. Instrumental noise is present across the whole frequency range; however, its magnitude (which can be estimated from the region of substantially no absorbance, 1800−1896 cm^{-1}) is in comparison small. Note the large difference between the absorbance unit scales for the raw and centered data (**Figure 1a** cf. **Figure 1b**−**e**). The most prominent features in the centered data are ∼4% of the size of the largest bands in the raw spectra. The multivariate modeling to follow will investigate whether the information present in these features is sufficient to distinguish systematically between the groups of centered data.

The centered data are essentially difference spectra. Some of the features may arise from variations in the relative abundance of individual absorbing species, but shifts in the position and width of peaks will also manifest as peaks or troughs. Moreover, in such complex samples, the total number of absorption bands in the raw spectra is very large, and most overlap one another. It is therefore difficult to assign features in **Figure 1b**−**e** to individual chemical components. This difficulty is compounded by the sparse and incomplete literature on spectral assignment of edible oils. The majority of the features are probably caused by compositional differences in fatty acids, as they are present in relatively large amounts and their proportions are affected by sample origin and cultivar (*6*, *12*, *13*). The main fatty acids in olive oils are oleic, palmitic, and linoleic; these compounds exhibit discernibly different infrared spectra (see, for example, the SDBSWeb library at http://www.aist.go.jp/RIODB/SDBS/).

Figure 2 shows an expansion of the 900−1150 cm^{-1} region in the centered data belonging to group 3 (Portuguese). These data arise from duplicate spectra of eight different samples. The figure shows that pairs of duplicate spectra are very similar, irrespective of the time interval between acquisitions—duplicates collected, for example, 4 days apart are no better matched than those separated by 4 weeks. This suggests that neither instrumental drift nor changes in the sample over the time scales concerned have had a serious negative impact upon the measurement repeatability. We believe, however, that to obtain this degree of repeatability, careful and painstaking spectral acquisition is essential, in view of the very small spectral differences that are of interest here. Preliminary and parallel studies (to be published elsewhere) using different instrumenta-

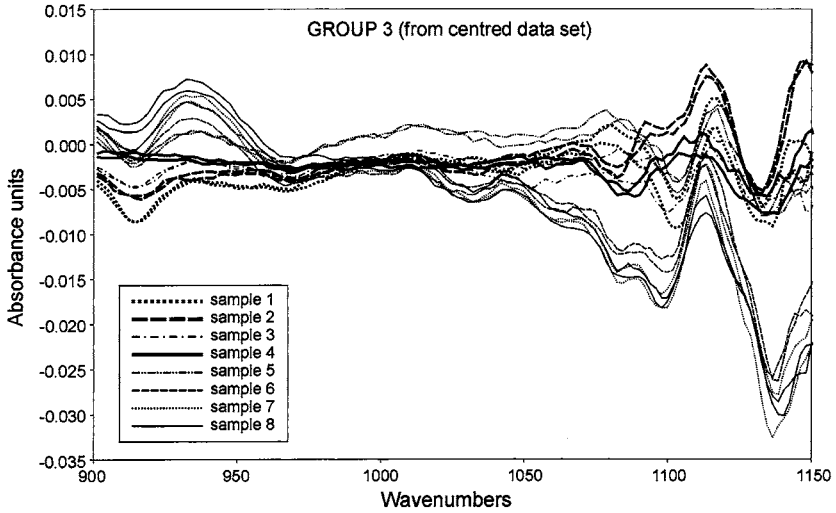

Figure 2. Expansion of the 900–1150 cm^{-1} spectral region for the centered data in group 3.

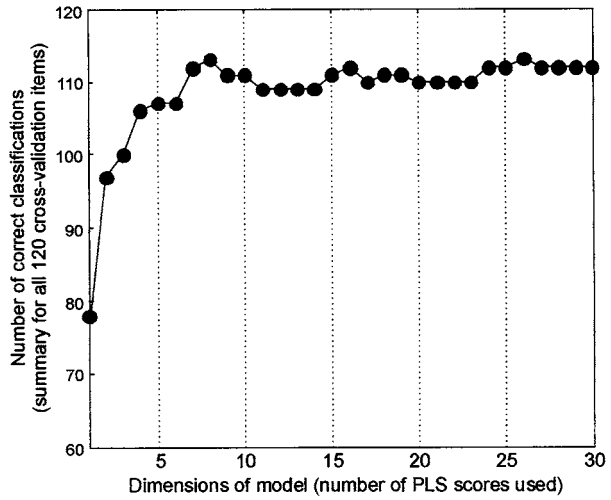

Figure 3. Number of classification successes (summed over the test segments) vs the number of PLS factors used in the PLS–LDA method.

tion (spectrometer, crystal material) and protocols (cleaning regime) have shown that experimental and instrumental conditions impact substantially both spectral quality and repeatability.

Multivariate analyses were carried out to determine whether the different groups of data could be systematically distinguished, beginning with the PLS−LDA procedure. The classification success rate (summarized across the test items) is shown vs the number of PLS factors used in **Figure 3**. There is a clear first optimum at eight PLS factors, at which the cross-validation success rate is ∼96% (115 out of 120 samples correctly classified). This is an encouraging performance for a method aiming to offer a rapid screen (rather than a diagnostic test). **Figure 4a** graphically represents the classification model, showing the first vs second CV scores from a cross-validated PLS−CVA using eight PLS factors. It is interesting to note that in this model, the Spanish and Portuguese groups are closest together. This may be a reflection of their neighboring geographical locations and consequent similarity in climate or other environmental conditions or of greater genetic similarity between the groups (some cultivars are represented in both). We note that for olive oils, distinguishing the country of origin simultaneously amounts to distinguishing between groups of cultivars, due to the regional differences in the varieties that are cultivated.

In whole spectrum methods, attempts are often made to interpret the "loadings": the vectors used in the various linear transformations, which can sometimes offer information on the nature of the biochemical factors that are varying in the samples. Understanding the biochemical basis of any discrimination would be interesting, although not essential for an effective screening method—it could be argued that "black box" techniques are harder to defraud. In the PLS−CVA procedure, successive linear transformations relate the data in a matrix **X** through to the canonical variate scores, **T**. Considering the vector of first CV scores $\mathbf{t}_1$, we can write:

$$\mathbf{t}_1 = \mathbf{Z}_r\,\mathbf{q}_1 = \mathbf{X}\,\mathbf{P}_r\,\mathbf{q}_1 = \mathbf{X}\,\mathbf{w}_1$$

in which $\mathbf{Z}_r$ is an $(n \times r)$ matrix of PLS scores, $\mathbf{q}_1$ is the first CV loading vector, $\mathbf{P}_r$ is a $(d \times r)$ matrix of PLS loadings, $\mathbf{w}_1$ is the vector of weights that relate $\mathbf{t}_1$ directly to **X**, and $d = 570$, $n = 120$, and in the optimum model, $r = 8$. From **Figure 4a**, the projections of the data onto $\mathbf{w}_1$ effectively discriminate group 2 (Italian) from other oils. **Figure 4b** compares the vector $\mathbf{w}_1$ with the means of the four groups of data calculated from the centered data shown in **Figure 1b−e**. Unfortunately, there

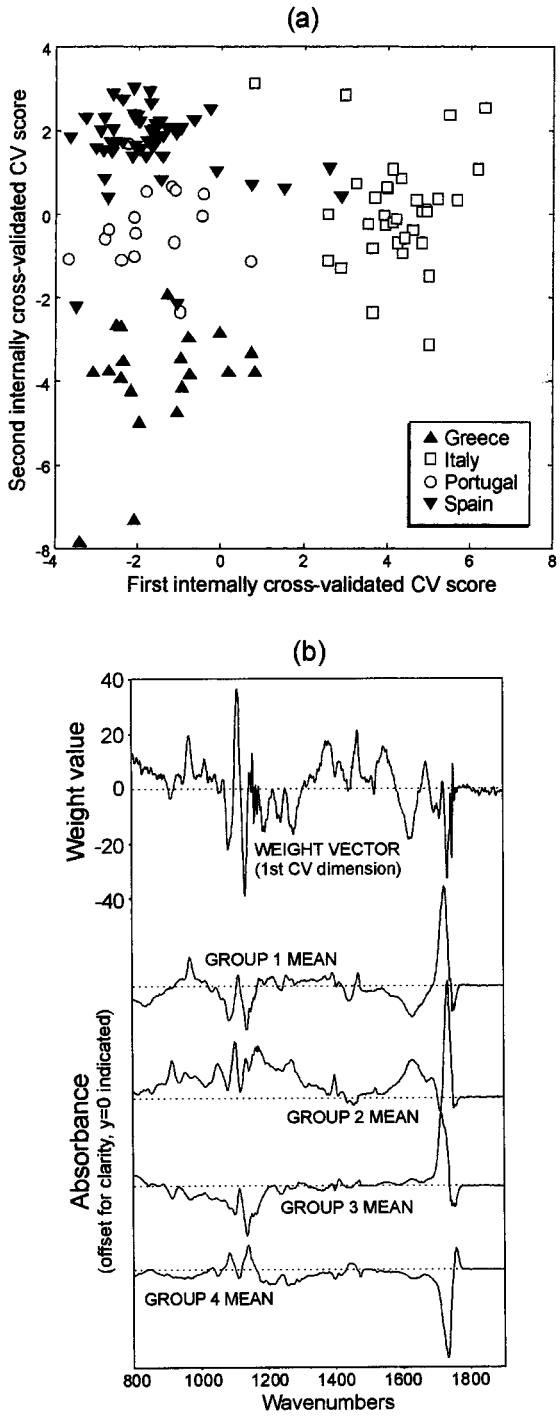

Figure 4. (**a**) First vs second CV scores from the PLS–CVA method (based on eight PLS factors). (**b**) First composite weight vector $\mathbf{w}_1$ (relating first CV dimension to raw data) shown above the mean of each group (calculated from centered data set).

is no simple interpretation of the weight vector in terms of feature assignments, although it appears to reflect several of the features present in the group means.

The variate selection approach was implemented. The GA aimed to identify a subset of raw variates that produce maximum cross-validation success rate in LDA. A subset size of eight was elected, due to the finding from the PLS−LDA analysis that eight factors were needed for an optimal model. Because the initial chromosomes are random and the nature of the problem suggests that there may be many local optima, which hinder convergence, the GA was executed 1000 times. The

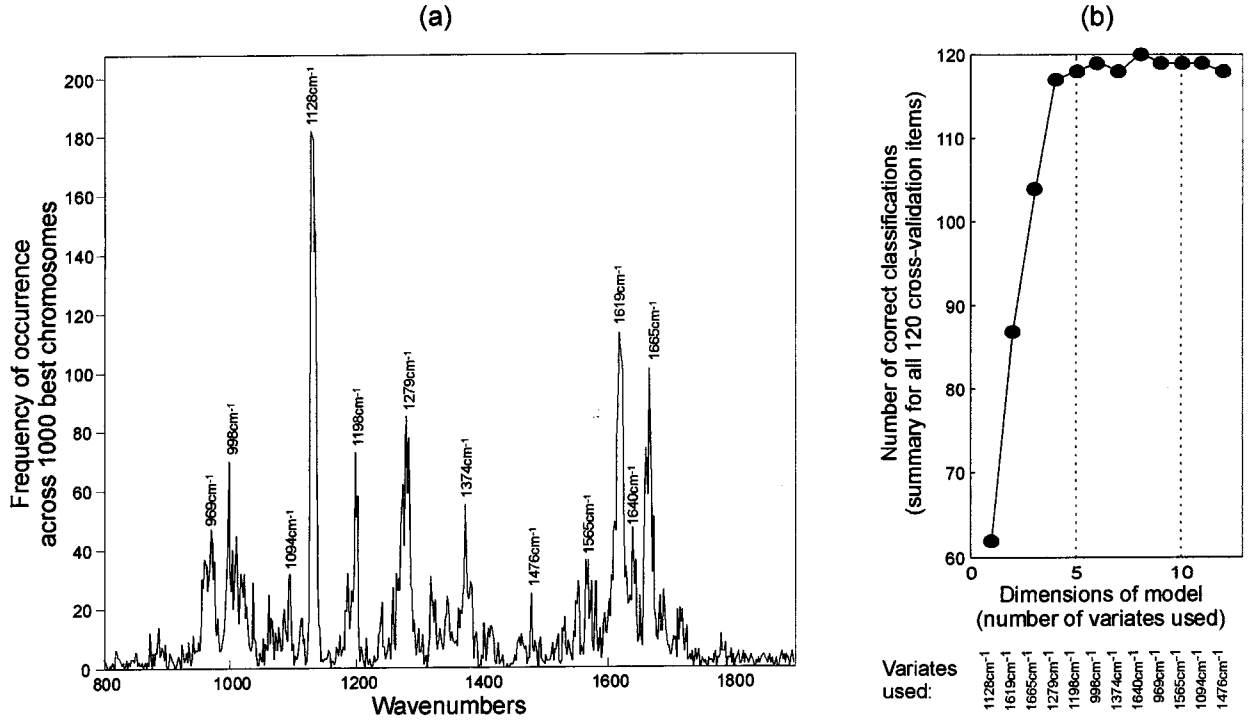

Figure 5. (**a**) Histogram showing frequency of occurrence of each variate within the 1000 GA solutions. (**b**) Number of classification successes (summed over the test segments) vs number of raw variates used. The spectral frequencies of the variates used are also given.

maximum cross-validation success rate obtained was 100%—all 120 samples correctly classified. Furthermore, different subsets of variates were equally able to produce this success rate: of the 1000 executions, 100% was obtained on 94 occasions, from 84 unique subsets. The large majority of the remaining executions converged on local optima yielding 118 or 119 successes. It is possible that further unique subsets yielding a 100% success rate could be identified with more repetitions of the GA or if the GA was refined to improve the search efficiency.

Figure 5a shows a histogram depicting the frequency of occurrence of each variate across all 1000 solutions. The histogram can be considered as a "pseudo-loading", with spectral-like features. Inspection of the solutions showed that the vast majority comprised representative variates from across several of these bands (rather than several from within one band). The largest features are peak-picked and labeled with the corresponding spectral frequency; we believe that these represent successively the most useful variates. Subsets of from 1 to all 12 of these variates were passed to the cross-validated LDA routine and ranked in order of their occurrence frequency, largest to smallest. The success rate as a function of the number of variates used is shown in **Figure 5b**. From comparison with **Figure 3**, it is clear that the selected variate approach has performed far better than PLS−LDA. A 100% success rate was achieved from a subset of eight variates and a ~98% success rate (117 correct assignments) using just four variates—both exceeding that achieved by PLS−LDA. This is a very substantial reduction in complexity as compared with the whole spectrum method. In principle, parsimonious models are desirable because they are more likely to be robust and transferable, for example, between spectrometers; this premise will be tested in work currently being planned.

Figure 6 compares the frequency positions of the eight highest-ranked variates (as identified on the histogram) with the mean of each group, again calculated from the centered data. Once more, it is difficult to relate the individual variates to

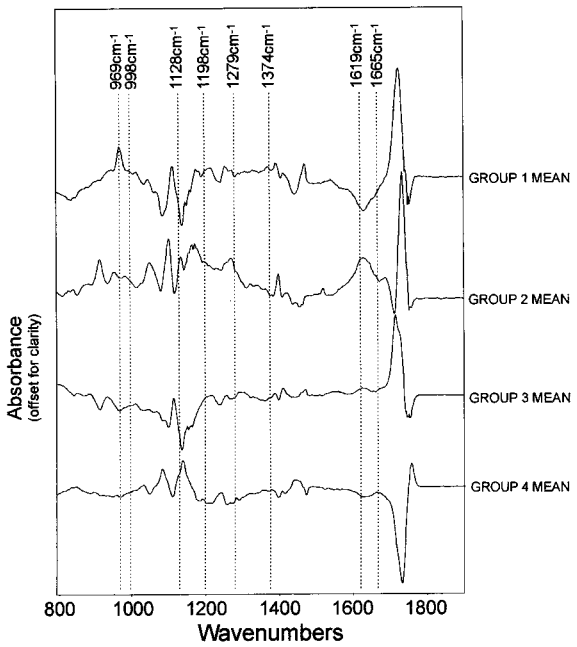

Figure 6. Subset of eight variates producing a 100% success rate, indicated by the vertical dashed lines, superposed upon the mean of each group (calculated from centered data set).

features present in the centered data, with the exception of the 969 cm^{-1} variate, which corresponds to a peak in the group 1 data (Greek) that is absent from the remaining groups. It seems, then, that neither the GA−LDA nor the PLS−LDA approach is able to provide significant insight, in biochemical terms, into the basis of the discriminatory models. The GA−LDA method, however, was much better at distinguishing between the groups of data. In conclusion, infrared spectroscopy combined with multivariate analysis has the potential to act as a black box screening technique, to verify the country of origin of extra virgin olive oils.

ACKNOWLEDGMENT

We thank Fausto Luchetti at the International Olive Oil Council for providing samples.

LITERATURE CITED

(1) Fedeli, E. Lipids of olives. In *Progress of Chemistry of Fats and Other Lipids*; Ralph, E., Holman, T., Eds.; Pergamon Press: Paris, France, 1977; pp 57−74.
(2) Kiritsakis, A. K. Flavor components of olive oil−A review. *J. Am. Oil Chem. Soc.* **1998**, *75*, 673−681.
(3) Aparicio, R.; Aparicio-Ruiz, R. Authentication of vegetable oils by chromatographic techniques. *J. Chromatogr. A* **2000**, *881*, 93−104.
(4) Tsimidou, M.; Karakostas, K. X. Geographical classification of Greek virgin olive oil by nonparametric multivariate evaluation of fatty-acid composition. *J. Sci. Food Agric.* **1993**, *62*, 253−257.
(5) Spangenberg, C. E.; Macko, S. A.; Hunziker, J. Characterization of olive oil by carbon isotope analysis of individual fatty acids: Implications for authentication. *J. Agric. Food Chem.* **1998**, *46*, 4179−4184.
(6) Stefanoudaki, E.; Kotsifaki, F.; Koutsaftakis, A. Classification of virgin olive oils of the two major Cretan cultivars based on their fatty acid composition. *J. Am. Oil Chem. Soc.* **1999**, *76*, 623−626.
(7) Alberghina, G.; Caruso, L.; Fisichella, S.; Musumarra, G. Geographical classification of Sicilian olive oils in terms of sterols and fatty-acids content. *J. Sci. Food Agric.* **1991**, *56*, 445−455.
(8) Tasioula-Margari, M.; Okogeri, O. Isolation and characterization of virgin olive oil phenolic compounds by HPLC/UV and GC-MS. *J. Food Sci.* **2001**, *66*, 530−534.
(9) Garcia, A.; Brenes, M.; Romero, C.; Garcia, P.; Garrido, A. Study of phenolic compounds in virgin olive oils of the Picual variety. *Eur. Food Res. Technol.* **2002**, *215*, 407−412.
(10) Webster, L.; Simpson, P.; Shanks, A. M.; Moffat, C. F. The authentication of olive oil on the basis of hydrocarbon concentration and composition. *Analyst* **1999**, *125*, 97−104.
(11) Bortolomeazzi, R.; Berno, P.; Pizzale, L.; Conte, L. S. Sesquiterpene, alkene, and alkane hydrocarbons in virgin olive oils of different varieties and geographical origins. *J. Agric. Food Chem.* **2001**, *49*, 3278−3283.
(12) Stefanoudaki, E.; Kotsifaki, F.; Koutsaftakis, A. The potential of HPLC triglyceride profiles for the classification of Cretan olive oils. *Food Chem.* **1997**, *60*, 425−432.
(13) Bucci, R.; Magri, A. D.; Magri, A. L.; Marini, D.; Marini, F. Chemical authentication of extra virgin olive oil varieties by supervised chemometric procedures. *J. Agric. Food Chem.* **2002**, *50*, 413−418.
(14) Cinquanta, L.; Esti, M.; LaNotte, E. Evolution of phenolic compounds virgin olive oil during storage. *J. Am. Oil Chem. Soc.* **1997**, *74*, 1259−1264.
(15) Vlahov, G.; Schiavone, C.; Simone, N. Quantitative C-13 NMR method using the DEPT pulse sequence for the determination of the geographical origin (DOP) of olive oils. *Magn. Reson. Chem.* **2001**, *39*, 689−695.
(16) Mannina, L.; Patumi, M.; Proietti, N.; Bassi, D.; Segre, A. L. Geographical characterization of Italian extra virgin olive oils using high-field H-1 NMR spectroscopy. *J. Agric. Food Chem.* **2001**, *49*, 2687−2696.
(17) Lai, Y. W.; Kemsley, E. K.; Wilson, R. H. quantitative-analysis of potential adulterants of extra virgin olive oil using infrared-spectroscopy. *Food Chem.* **1995**, *53*, 95−98.
(18) Guillen, M. D.; Cabo, N. Infrared spectroscopy in the study of edible oils and fats. *J. Sci. Food Agric.* **1997**, *75*, 1−11.
(19) Yang, H.; Irudayaraj, J. Comparison of near-infrared, Fourier transform-infrared, and Fourier transform-Raman methods for determining olive pomace oil adulteration in extra virgin olive oil. *J. Am. Oil Chem. Soc.* **2001**, *78*, 889−895.
(20) Ozen, B. F.; Mauer, L. J. Detection of hazelnut oil adulteration using FT-IR spectroscopy. *J. Agric. Food Chem.* **2002**, *50*, 3898−3901.
(21) Bertran, E.; Blanco, M.; Coello, J.; Iturriaga, H.; Maspoch, S.; Montoliu, I. Determination of olive oil free fatty acid by Fourier transform infrared spectroscopy. *J. Am. Oil Chem. Soc.* **1999**, *76*, 611−616.
(22) Al Jowder, O.; Kemsley, E. K.; Wilson, R. H. Detection of adulteration in cooked meat products by mid-infrared spectroscopy. *J. Agric. Food Chem.* **2002**, *50*, 1325−1329.
(23) Krzanowski, W. J. *Principles of Multivariate Analysis*; Clarendon Press: Oxford, U.K., 1993; pp 291−306.
(24) Walmsley, A. D. Improved variable selection procedure for multivariate linear regression. *Anal. Chim. Acta* **1997**, *354*, 225−232.
(25) Shaw, A. D.; diCamillo, 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 C-13 NMR and multivariate calibration with variable reduction. *Anal. Chim. Acta* **1997**, *348*, 357−374.
(26) Leardi, R.; Seasholtz, M. B.; Pell, R. J. Variable selection for multivariate calibration using a genetic algorithm: prediction of additive concentrations in polymer films from Fourier transform-infrared spectral data. *Anal. Chim. Acta* **2002**, *461*, 189−200.
(27) Costa, P. A. D.; Poppi, R. J. Application of genetic algorithms in the variable selection in mid infrared spectroscopy. Simultaneous determination of glucose, maltose and fructose. *Quim. Nova* **2002**, *25*, 46−52.
(28) Goicoechea, H. C.; Olivieri, A. C. Wavelength selection for multivariate calibration using a genetic algorithm: A novel initialization strategy. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1146−1153.

Received for review March 31, 2003. Revised manuscript received July 11, 2003. Accepted July 16, 2003. This work was funded by the Biotechnology and Biological Sciences Research Council's CSG grant.

JF030232S