A Comparison of Mid-Infrared and Raman Spectroscopies for the Authentication of Edible Oils

N.A. Marigheto, E.K. Kemsley, M. Defernez, and R.H. Wilson*

Institute of Food Research, Norwich Laboratory, Norwich Research Park, Colney, Norwich, NR4 7UA

ABSTRACT: Robust, routine, and rapid instrumental methods for the determination of the authenticity of edible oils, and the detection of adulteration, are continually being sought. In this paper, we compare mid-infrared and Raman spectroscopies for their ability to discriminate between oils of differing botanical origin and for their ability to detect added adulterants. Furthermore, we used sufficient numbers of samples to permit a comparison of some of the chemometric methods (linear discriminant analysis, artificial neural networks) available and looked at the results obtained when the two spectroscopic datasets were combined. We show that mid-infrared spectroscopy, in combination with linear discriminant analysis, gave the best classification rates and adulteration detection levels compared to Raman or combined data.

JAOCS 75, 987–992 (1998).

KEY WORDS: Adulteration, chemometrics, edible oil, infrared spectroscopy, Raman spectroscopy.

Olive oils are classified according to purity; they can vary from extra-virgin to lampante, which is not fit for consumption. According to the International Olive Oil Council (IOOC), "virgin olive oil is the oil obtained from the fruit of the olive tree solely by mechanical or other physical means under conditions, particularly thermal conditions, that do not lead to alteration in the oil, and which has not undergone any treatment other than washing, decantation, centrifugation and filtration" (1). The adulteration of olive oils with cheaper oil not only is a commercial problem but also has health implications (2). Several methods have been applied to the adulteration problem, and previous workers have shown that spectroscopic techniques, such as infrared and Raman, have potential for such analysis. However, a problem for the potential user is which method to choose for further development and use. Quite often the choice is not clear due to different experimental designs used by different workers, limited sample numbers, and sample authenticity. Infrared and Raman spectroscopies are techniques that can easily be adapted for use by untrained personnel in laboratories or on the factory floor. A particular attraction is the ease of sample presentation. Although both Raman and infrared are forms of vibrational spectroscopy, the underlying physical processes that give rise to the two phenomena are different. As a result, bands that are strong in infrared may be weak or absent in the corresponding Raman spectrum and vice versa. For any given analysis, one may be preferred over the other.

Lai and co-workers showed that Fourier transform mid-infrared spectroscopy is able to discriminate between different vegetable oils and types of olive oil, even though extra-virgin olive oil and refined olive oil are similar chemically and spectrally (3). They have also shown that it is possible to determine, quantitatively, the level of typical adulterants in extra-virgin olive (4). The data could also be interpreted in molecular terms, so that the chemical basis of discrimination was understood. Fourier transform infrared (FTIR) spectroscopy and multidimensional analysis techniques were also used by Safar and colleagues for the characterization of edible oils, butters, and margarines, in which the lipid-rich foods were classified according to their degree of unsaturation (5). Principal component (PC) analysis of FTIR spectra was performed by Dupuy and colleagues to classify edible fats and oils with regard to their origins (6). They used two sampling methods: attenuated total reflectance (ATR) for fats and the mid-infrared optical fiber method for oils. From the interpretation of the first PC, they concluded that the basis for the discrimination between fats is the concentration of unsaturated fatty acids, and different concentrations of linoleic acid for oils (sunflower, olive, and peanut oils).

The near-infrared (NIR) region of the spectrum has also been explored for the discrimination and authentication of fats and oils. Sato used PC analysis on NIR spectroscopic data for classification of vegetable oils: soybean, corn, cottonseed, olive, rice bran, peanut, rapeseed, sesame, and coconut oils (7). Bewig and colleagues achieved differentiation between four vegetable oil types (cottonseed, peanut, soybean, and canola) and classification of unknown samples in which the second-derivative spectra of the oils were subject to discriminant analysis with Mahalanobis distances principles (8). Wesley and coworkers developed a method for predicting the level of adulteration in a set of virgin and extra-virgin olive oils adulterated with corn oil, sunflower oil, and raw olive residue by PC analysis (9). They also used NIR spectroscopy and discriminant analysis to identify and quantitate adulterants in extra-virgin olive oils (10).

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

Fourier-transform Raman spectroscopy was used by Baeten and co-workers to predict the level of adulteration in a set of virgin olive oil samples that were adulterated with soybean, corn, and raw olive oil residue oils (11). They used 24 samples: 6 genuine extra-virgin olive oils, and 6 samples adulterated with soybean, corn, or olive pomace oils at 1, 5, or 10% (w/w).

In this paper, we compare the performance of two of the spectroscopic techniques—Raman and mid-infrared—and test the value of combining the databases for the authentication of edible oils.

MATERIAL AND METHODS

Instrumental methods. Mid-infrared spectra of the samples were acquired on a Spectra-Tech Applied Systems Inc. (Stanford, CT) MonitIR FTIR spectrometer by the ATR sampling technique. The instrument was equipped with a horizontal 45° ZnSe trapezoidal ATR cell, a deuterated triglycine sulfate (DTGS) detector, and a germanium on potassium bromide substrate beamsplitter. The spectra were collected at 4 cm^{-1} resolution with co-addition of 256 interferograms. A triangular appodization function was used prior to Fourier transformation to a single-beam spectrum, which was ratioed against an air background spectrum and converted into absorbance units (12). After the acquisition of each spectrum, the oil was wiped off of the crystal with tissue paper, the crystal was cleaned with hexane, acetone, and finally distilled water. To avoid any spectral variation due to instrumental drift, and to reduce the variation from day to day to a minimum, an air background spectrum was collected before each sample spectrum.

Raman spectroscopy was carried out on a Bio-Rad FT-Raman spectrometer with a diode-pumped Topaz laser (Spectra-Physics, St. Albans, United Kingdom) operating at 1064 nm as the excitation source and a nitrogen-cooled germanium detector. Spectral acquisition conditions were: 4 cm^{-1} resolution, 256 interferograms collected before Fourier transformation, and 900 mW laser power at the sample. To compensate for the detector response profile, a spectrum of ground potassium bromide (KBr), illuminated by a white light source, was collected with compatible acquisition parameters, and all sample spectra were ratioed to this spectrum.

Materials. Our large sample collection was derived from a number of sources. Firstly, we obtained samples from local retailers. Although this source is far from ideal, it was the source used by all previous spectroscopic workers described above. However, we were able to obtain information from certain retailers about the audit trailing of selected samples, leading to increased confidence in their origin. We also obtained samples from colleagues who worked on parallel studies funded by the Ministry of Agriculture, Fisheries and Food (MAFF), which had been tested by other analytical procedures. Finally, we were able to obtain samples obtained at source *via* colleagues in a European Concerted Action ("Food Authenticity: Issues and Methodologies").

The samples of edible oils for the database were: extra-virgin olive, refined olive, sunflower, rapeseed, soybean, sesame, hazelnut, sweet almond, grapeseed, safflower, peanut, walnut, mustard, corn, palm, coconut, and palm kernel. The last three samples were solids at room temperature.

About 2 mL of each of the liquid oil samples was transferred to screw-capped fluorescence-free glass vials for the acquisition of Raman spectra. A vial was placed in the sample compartment, and the laser spot was directly focused onto it. For infrared spectral acquisition, the sample was loaded directly onto the ATR crystal. A different procedure was followed to obtain the spectra of oils that are solid at room temperature. The sample was melted at 35°C in a water bath, and the molten material was transferred to a vial for the Raman experiment or spread onto the ATR crystal for the infrared analysis and allowed to set into the solid state before acquiring the spectra. One hundred forty spectra were chosen to form a database, called 'pure oils,' in which seven different groups were present: 36 extra-virgin olive oils, 10 refined olive oils, 28 sunflower oils, 18 rapeseed oils, 9 peanut oils, 21 soybean oils, and 18 corn oils. These were further divided into a training set of 84 spectra, a tuning set of 27 spectra, and a test set of 29 spectra. The database was chosen after preliminary data processing (principal component analysis). Some oils were identified as outliers and were removed from the data set.

Samples of extra-virgin olive oils, adulterated with various olive oils and seed oils, were prepared by using oils that were purchased locally. Eleven extra-virgin olive oils were chosen to be adulterated with five olive oils; five extra-virgin olive oils were adulterated with five seed oils. The levels of adulteration were 5, 15, 25, 35, and 45% (w/w). The adulterated samples were prepared as follows: (i) the required amount of olive or seed oil was weighed into a clean and dry screw-capped glass vial, (ii) extra-virgin olive oil was added until the final weight was about 5 g, (iii) the sample was mixed by vigorously shaking, and (iv) 2 mL of this mixture was transferred to the vial used for the Raman acquisition.

In this way, 275 samples of extra-virgin olive oil, adulterated with olive oil, and 125 samples of extra-virgin oil, adulterated with seed oil, were prepared. The Raman and infrared spectra of 150 extra-virgin olive oil samples, adulterated with olive oil, and of all extra-virgin samples, adulterated with seed oil, were obtained. To evaluate different adulteration detection methods, a database called 'adulterated' was created: it contained all spectra of the adulterated samples and the 36 spectra of extra-virgin olive oils. It was further divided into a training set with 185 spectra, and tuning and test sets with 63 spectra each.

Chemometric methods. The data processing was carried out with Matlab (The Math Works Inc., Natick, MA) running on a personal computer. Macros were written in-house for carrying out the linear discriminant analysis (13,14). The artificial neural network analysis was done with NeuralDesk (Neural Computer Sciences, Southampton, UK, 1994). The spectra were area- normalized and baseline-corrected before data processing.

For the 'pure oils' database, a PC analysis by correlation matrix was performed with the training data, and the PC

scores of the training set were determined. The scores of the test set in the PC space were then determined. The PC scores were then used to: (i) perform a linear discriminant analysis (LDA) for training and test sets by using Mahalanobis distances; and (ii) carry out an artificial neural network (ANN) analysis for training and test set by using the tuning set for cross-validation.

Two analyses were carried out: one to discriminate between extra-virgin olive oils and other oils, and one to discriminate between seven oil types.

For the 'adulterated' database, a validation procedure, called "leave-one-out" or "internal cross-validation," was used. This consists of omitting one sample at a time from the data set (of *n* samples); using the remaining data as a training set; and using the sample that was removed as a test sample. This is repeated *n* times, omitting each sample in turn. Partial least squares (PLS) regression, based on the algorithm for orthogonalized PLS with one dependent variable (15) while using "internal cross-validation," was performed. To carry out the analysis with an ANN, the PC scores of the training, tuning, and test sets were first calculated as described earlier.

A two-group LDA was performed by using either PLS or a PC analysis as the reduction step. The leave-one-out procedure was used to give the number of successful classifications of the test samples.

The infrared and Raman databases were combined, and the chemometrics methods were applied to see if improved discrimination and/or adulteration detection was possible.

RESULTS AND DISCUSSION

Figures 1 and 2 show a selection of infrared and Raman spectra of extra-virgin olive oils, respectively. The percentage classification for the 'pure oils' database test set is shown in Table 1 based on two groups of samples: extra-virgin olive oils and all others, including refined olive, sunflower, rapeseed, peanut, soybean and corn oil. For the infrared data, 100% of the samples were correctly classified by using with 15 PC scores and using ANN with 12 PC scores. For the Raman data, the best prediction was 93.1% for 10 PC scores with LDA, which did not improve with the use of more PC. This result was slightly better than the classification provided by the ANN. Both infrared and Raman spectroscopic techniques worked, but in general, infrared gave more satisfactory results. This may be due to the fact that the important bands for classification are not active in Raman. It is also likely that these findings are due to the lower signal-to-noise ratio of Raman when compared to that of infrared. Combining the data sets leads to no improvement in the classification success rate when using LDA over that obtained with infrared data alone.

Table 2 shows the results of a discriminant analysis based on seven oil groups. For the infrared data with LDA, a 100% correct assignment was obtained with 10 PC scores. When analyzing the data with ANN, 20 PC values were needed to obtain 93.1% success. The fact that the ANN did not work as well as it did for the two-group analysis may be due to the number of samples. In this particular situation, there were fewer samples in each group. The results provided by the Raman data were not as good as those obtained with infrared when both LDA and ANN were used.

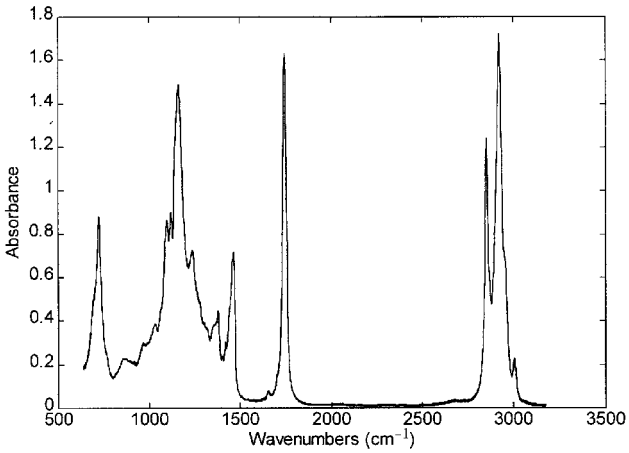

FIG. 1. Infrared spectra of typical extra-virgin olive oils.

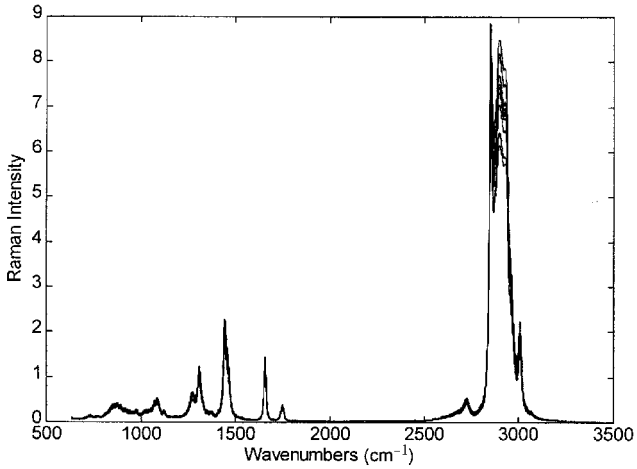

FIG. 2. FT-raman spectra of typical extra-virgin olive oils.

The results of a PLS regression with the 'adulterated' database by using "cross-validation" are shown in Table 3. These

TABLE 1
Percentage Classification Success Obtained by Different Chemometric Methods: Linear Discriminant Analysis (LDA) and Artificial Neural Networks (ANN)[a]

Number of principal components	Infrared		Raman		Combined data
	LDA	ANN	LDA	ANN	LDA
5	93	97	90	90	87
10	97	97	93	90	93
12	97	100	93	90	97
15	100	100	93	90	100
20	100	100	93	90	100

[a]Results for test set of two-group analysis.

TABLE 2
Percentage Successes of Classification Obtained by Different Chemometric Methods: LDA and ANN[a]

Number of principal components	Infrared		Raman		Combined data
	LDA	ANN	LDA	ANN	LDA
5	96	66	66	38	80
10	100	69	83	52	97
15	100	90	80	52	100
20	100	93	93	62	100

[a]Results for test set of seven-group analysis. For abbreviations see Table 1.

TABLE 3
Percentage of Correct Classified Samples Obtained When Using Partial Least Squares (PLS) Regression and ANN Applied to the 'Adulterated' Database[a]

Number of PLS factors scores	Infrared		Raman		Combined data
	PLS regression	ANN	PLS regression	ANN	ANN
5	97	27	97	95	86
8	98	84	96	94	87
10	99	89	94	92	78
12	99	68	93	95	60
15	99	75	92	95	59
20	99	60	91	95	75

[a]PC, principal component. For other abbreviation see Table 1.

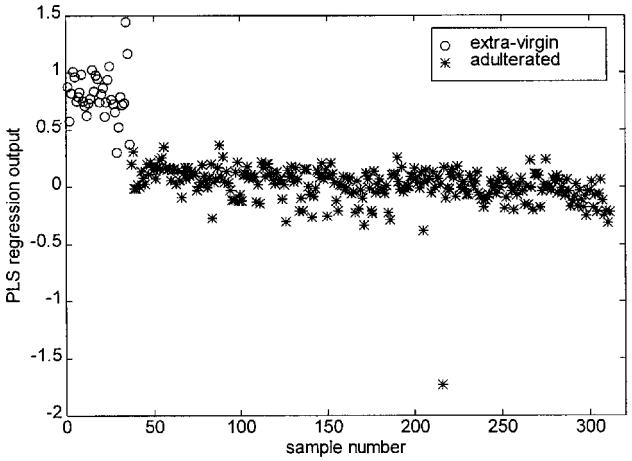

FIG. 3. Output of a partial least-squares (PLS) regression of infrared data (extra-virgin and adulterated oil samples) at 10 PLS factors.

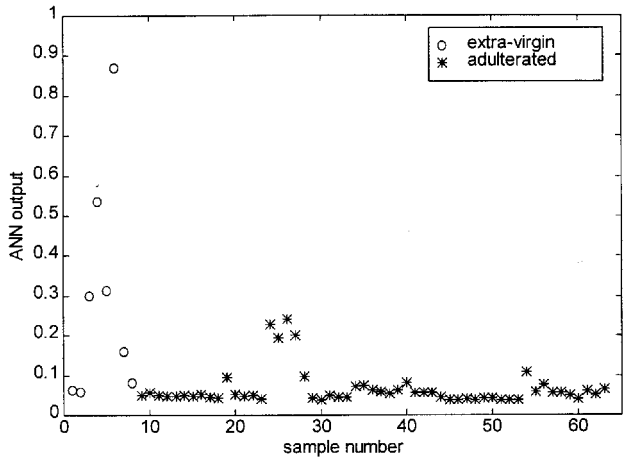

FIG. 4. Output of an artificial neural network (ANN) analysis of infrared data (extra-virgin and adulterated oil samples) with 10 principal component (PC) scores.

are compared with the output from the analysis by ANN. For the infrared data, the best prediction is obtained by using 10 PLS factors and 10 PC scores for ANN, with the PLS results being slightly better than those provided by ANN. The reason for these results may be the number of samples in each step of processing: for a PLS regression, all 311 samples in the database were used as 'training' and 'test' samples, whereas for the analysis with the ANN, the database was divided into a training set of 185 spectra, and tuning and test sets with 63 spectra each. For Raman, the best result was obtained with five PLS factors or five PC scores. However, the prediction success provided by the Raman data is lower than that provided by infrared. The combination of infrared and Raman data did not improve the level of prediction, and eight PC were necessary to give a prediction level of 87%.

The adulteration detection limit provided by PLS and ANN, when applied to the infrared data, was 5% for each of the adulterants used (the minimum adulterant level). However, for the Raman data, the detection limit was 45% when the adulterant was refined olive oil and 5% for the other adulterants. These results are given for the number of PLS factors or PC scores that provided the best prediction results. The full model plots for each of the results discussed above are shown in Figures 3 to 6.

The results of a two-group LDA, performed on the infrared data by using either PLS or a principal component analysis (PCA), as the reduction step, revealed that fewer PLS factors were required to obtain prediction success rates (99% with nine factors) when compared with the analysis based on PC scores (99% with 19 PC). As shown by Kemsley (16), the reason for this is believed to be that PLS reductions yield scores that maximize the between-groups variance. For the Raman data, the same general result was found; using PLS as a pretreatment gave a better prediction than when using PCA. However, as before, infrared data provided more correct classified samples. A summary of the results for the infrared and Raman data is presented in Figures 7 and 8, respectively.

The combination of infrared and Raman data has been proposed by spectroscopists as an approach in which the complimentary nature of the techniques should lead to improved discrimination. The idea was that different functional groups give rise to quite significantly different intensities in the infrared and Raman spectra, so that a complete vibrational spectrum is only really obtained when the two spectra are combined. We also aimed to ensure that there were sufficient samples for a valid comparison to be drawn, particularly because

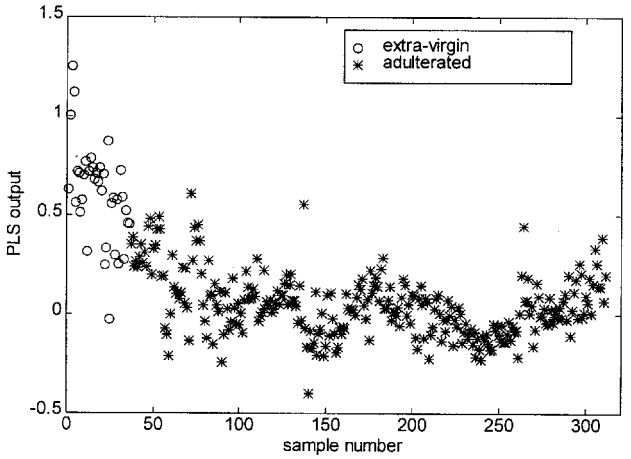

FIG. 5. Output of a PLS regression of Raman data (extra-virgin and adulterated oil samples) at 5 PLS factors. See Figure 3 for abbreviation.

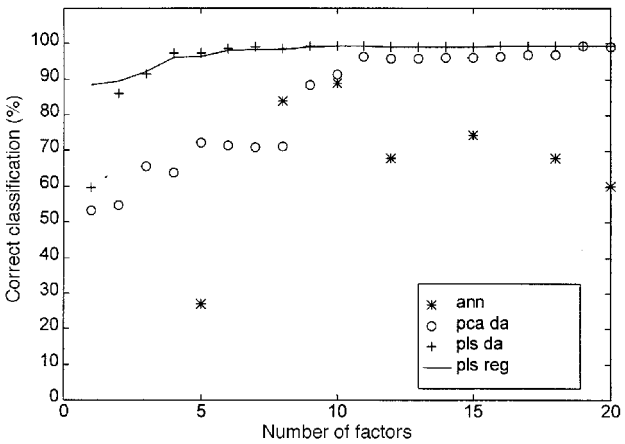

FIG. 7. Comparison between the different adulteration detection methods when applied to infrared data. Abbreviations: da, discriminant analysis; reg, regression; PCA, principal component analysis. For other abbreviations see Figures 3 and 4.

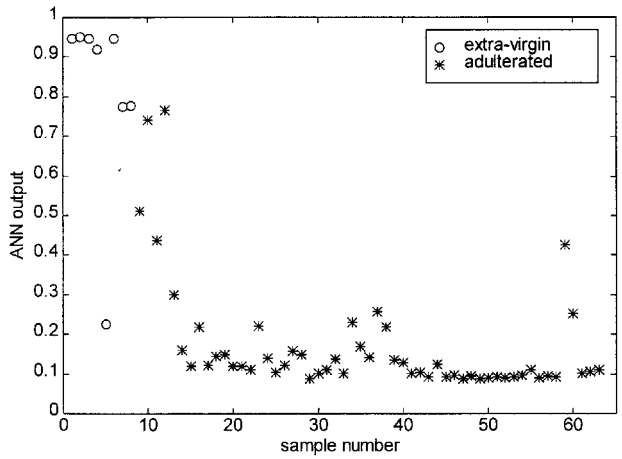

FIG. 6. Output of an ANN analysis of Raman data (extra-virgin and adulterated oil samples) with 5 PC scores. See Figure 4 for abbreviations.

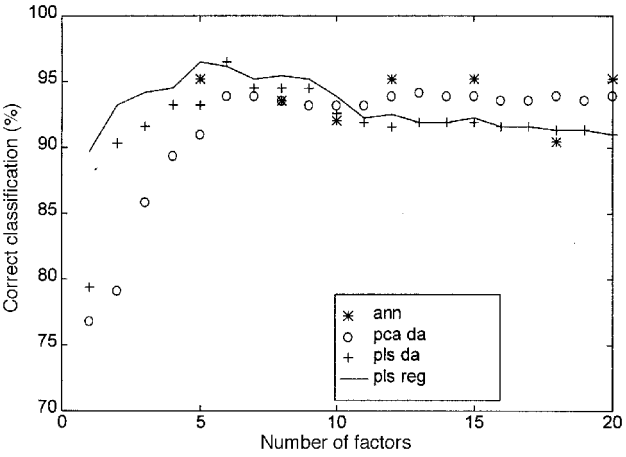

FIG. 8. Comparison between the different adulteration detection methods when applied to Raman data. For abbreviations see Figures 3, 4, and 7.

a further objective was to compare some of the multivariate statistical methods currently applied to such data.

The findings presented in this paper show that infrared is better than Raman at classifying oil samples and detecting adulteration. Furthermore, we found no additional benefit from combining the two data sets. We are not altogether surprised by this result. Although fatty acid composition is important in discriminating between most oils, the detection of refined olive oils added to extra-virgin olive oils may require the techniques to be sensitive to low levels of specific compounds. In other mixtures, the composition may be affected only to a small degree by the adulterant. The fact that Raman spectroscopy is poorer at detecting additions may result from the fact that Raman spectral quality, in terms of signal-to-noise ratio, is considerably less than that for infrared. This was most noticeable when extra-virgin olive oil was adulterated with refined olive oil.

We also found that linear discriminant analysis based on partial least-squares data reduction gave better results than artificial neural networks for classification according to botanical origin. PLS/LDA was also better at separating adulterated samples from pure samples.

We conclude that the infrared method combined with LDA based on PLS data reduction is likely to be the best technique for further investigation and development as a rapid technique.

ACKNOWLEDGMENTS

The authors thank the Ministry of Agriculture, Fisheries and Food for funding this work. They also thank Matthew Sharman of CSL Norwich and members of FAIM Concerted Action, especially Ivonne Delgadillo, for supply of samples.

REFERENCES

1. Trade Standard Applying to Olive Oil and Olive Pomace Oil, *COI/T.15/NC n° 2/REV.4,* International Olive Oil Council, Madrid, Spain, 1996.

2. Kochhar, S.P., and J.B. Rossell, The Spanish Toxic Oil Syndrome, *Nutr. Food Sci. 90*:14–15 (1984).
3. Lai, Y.W., E.K. Kemsley, and R.H. Wilson, Potential of Fourier Transform Infrared Spectroscopy for the Authentication of Vegetable Oil, *J. Agric. Food Chem. 42*:1154–1159 (1994).
4. 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).
5. Safar, M., D. Bertrand, P. Robert, M.F. Devaux, and C. Genot, Characterization of Edible Oils, Butters and Margarines by Fourier Transform Infrared Spectroscopy with Attenuated Total Reflectance, *J. Am. Oil Chem. Soc. 71*:371–377 (1994).
6. Dupuy, N., L. Duponchel, J.P. Huvenne, B. Sombret, and P. Legrand, Classification of Edible Fats and Oils by Principal Component Analysis of Fourier Transform Infrared Spectra, *Food Chem. 57*:245–251 (1996).
7. Sato, T. Application of Principal Component Analysis on Near-Infrared Spectroscopic Data of Vegetable Oils for Their Classification, *J. Am. Oil Chem. Soc. 71*:293–298 (1994).
8. Bewig, K.M., A.D. Clarke, C. Roberts, and N. Unklesbay, Discriminant Analysis of Vegetable Oils by Near-Infrared Reflectance Spectroscopy, *Ibid. 71*:195–200 (1994).
9. Wesley, I.J., R.J. Barnes and A.E.J. McGill, Measurement of Adulteration of Olive Oils by Near-Infrared Spectroscopy, *Ibid. 72*:289–292 (1995).
10. Wesley, I.J., F. Pacheco, and A.E.J. McGill, Identification of Adulterants in Olive Oils, *Ibid. 73*:515–518 (1996).
11. Baeten, V., M. Meurens, M.T. Morales, and R. Aparicio, Detection of Virgin Olive Oil Adulteration by Fourier Transform Raman Spectroscopy, *J. Agric. Food Chem. 44*:2225–2230 (1996).
12. Wilson, R.H., Fourier Transform Mid-Infrared Spectroscopy for Food Analysis, *Trends Anal. Chem. 9*:127–131 (1990).
13. Defernez, M., and E.K. Kemsley, The Use and Misuse of Chemometrics for Treating Classification Problems, *Ibid. 16*:216–221 (1997).
14. Defernez, M., and E. K. Kemsley, Establishment of Guidelines for the Application of Chemometric Methods to Food Authenticity Problems, *Report on MAFF Open Contracting*, Institute of Food Research, Norwich, United Kingdom, 1996.
15. Martens, H., and T. Naes, *Multivariate Calibration*, Wiley, Chichester, 1989, pp. 119–125.
16. Kemsley, E.K., Discriminant Analysis of High Dimensional Data: A Comparison of Principal Components Analysis (PCA) and Partial Least Squares (PLS) Data Reduction Methods, *Chem. Int. Lab. Sys. 33*:47–61.

[Received June 3, 1997; accepted April 9, 1998]