Identification of Adulterants in Olive Oils

I.J. Wesley[a,*], F. Pacheco[b], and A.E.J. McGill[a]

[a]Seale-Hayne, Department of Agriculture and Food Studies, University of Plymouth, Devon TG12 6NQ, United Kingdom, and [b]ETGI, Escola Superior de Biotecnologia, Universidade Católica Rua Dr. António Bernardino de Almeida, 4200 Porto, Portugal

ABSTRACT: The application of discriminant analysis for identifying and quantifying adulterants in extra virgin olive oils is presented. Three adulterants were used (sunflower oil, rapeseed oil, and soybean oil) and were present in the range 5–95%. Near-infrared spectroscopy and principal components analysis were used to develop a discriminant analysis equation that could identify correctly the type of seed oil present in extra virgin olive oil in 90% of cases. Partial least squares analysis was used to develop a calibration equation that could predict the level of adulteration. Cross validation suggested that it was possible to measure the level of adulteration to an accuracy of ±0.9%. External validation of the derived calibation equation gave a standard error of performance of ±2.77%.
JAOCS 73, 515–518 (1996).

KEY WORDS: Discriminant analysis, near-infrared spectroscopy, olive oil, principal components analysis.

The issue of authenticity is becoming increasingly important in the food industry. Many products retail at a premium price on the basis of their purity and quality or their health qualities. The accidental adulteration of products during processing can lead to wastage, which will inevitably increase processing costs. It is also possible to gain financial advantage by deliberately mislabelling or adulterating products and presenting them as premium products.

Methods are required that can establish the authenticity of a product and which are quick and easy to use. Near-infrared (NIR) spectroscopy offers all of these features, as well as the opportunity to employ robust instruments, which can be used on-line and in-line to monitor production processes. In a previous paper (1), it was demonstrated that it is possible to develop a calibration for the purity of olive oil by using the NIR spectrum. The calibration predicted the purity of the oil to an accuracy of ±1.5%. Prediction of the adulterant type was also attempted. This would be of major importance in tracing the source of any adulteration. However, in that case the prediction of the adulterant type was not very accurate, with the correct adulterant being predicted only three times out of five. This poor result was attributed partially to the design of the original experiments, and it was suggested that a more balanced calibration set might improve the predictive ability of the calibration. In this paper, the results are reported of a further set of experiments that were designed to assess the potential of NIR spectroscopy as a method of accurately predicting the adulterant type in extra virgin olive oils adulterated with three different oils.

EXPERIMENTAL PROCEDURES

Sample preparation. Three adulterant types (sunflower oil, rapeseed oil, and soybean oil) and two extra virgin olive oils were used. Oil samples were obtained from bulk importers (Anglia Oils Ltd., Hull, United Kingdom and Central Edible Oils Group, Liverpool, United Kingdom) and local retail outlets. For each adulterant type, a sample set of 19 oil mixtures, covering the range 5–95% olive oil in 5% increments, was prepared. Samples were prepared on a w/w basis. Each extra-virgin olive oil was used to make half of the mixtures (selected at random) for each adulterant. A vortex mixer was used to ensure thorough mixing. In addition to the mixtures, the spectra of pure adulterants, including examples not used to make mixtures, were recorded and added to the appropriate spectral data sets, and the spectra of pure extra-virgin olive oils, including examples from additional sources, were recorded. The final sample set consisted of 23 mixtures that contained rapeseed oil, 21 mixtures with soybean oil, 23 mixtures with sunflower oil, and 6 pure extra-virgin olive oils. For the mixtures, each sample was scanned twice, giving two spectra per sample. For the pure olive oils, each sample was scanned three times giving three spectra per sample. In total, 152 spectra were obtained.

Before analysis of the spectra, the data set was divided into four files according to adulterant class (sunflower, rapeseed, soybean, or pure). From each class file, five samples were selected at random, and one spectrum of each sample was used to create a validation file of 20 spectra. In each case, the selected spectra were deleted from the class file. The remaining spectra in each class file were combined, creating a calibration file of 132 spectra. In all files, the spectra were given product codes according to adulterant category (rapeseed = 1, soybean = 2, sunflower = 3, pure extra virgin olive = 4). Examples of the pure spectra are shown in Figure 1.

NIR spectroscopy. Samples were scanned in an NIRSystems model 6500 extended-range scanning NIR spectrometer (Perstorp/NIRSystems Analytical, Maidenhead, United Kingdom). This instrument covers the NIR and visible spectral re-

*To whom correspondence should be addressed at Seale-Hayne, Department of Agriculture and Food Studies, University of Plymouth, Newton Abbot, Devon TG12 6NQ, United Kingdom.

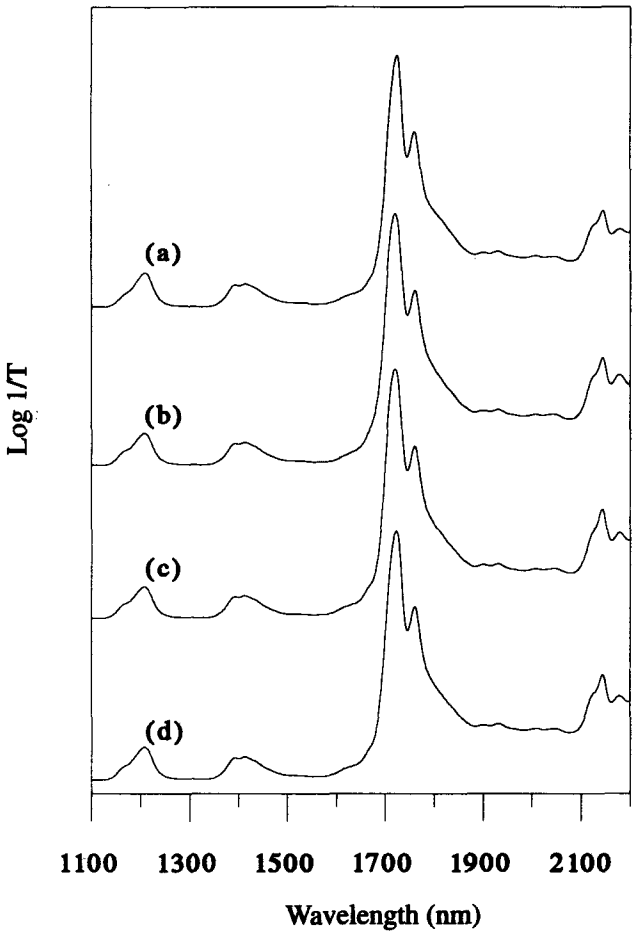

FIG. 1. Near-infrared spectra of (a) pure rapeseed oil, (b) pure soybean oil, (c) pure sunflower oil, and (d) pure extra-virgin olive oil. Spectra are offset for ease of viewing.

gion from 400–2500 nm and was configured for direct transmission measurement while utilizing a standard 1-mm pathlength quartz cuvette. Spectra were recorded as log 1/transmittance at 2 nm intervals from 400–2500 nm. The scan speed was 1.8 scans/s, and 4-point Fourier smoothing was applied. Two spectra of each sample were recorded, with the sample cell reloaded between each spectrum. Statistical analysis was carried out with ISI Systems software (Perstorp Analytical Inc., Silver Springs, MD) and Win-Discrim (Institute for Food Research, Norwich, United Kingdom).

RESULTS AND DISCUSSION

Discrimination of adulterant type. There is already a considerable amount of literature on the problem of identification of oils by spectroscopic methods. The most widely used type of parametric method for pattern recognition is discriminant analysis (2,3). Bewig *et al.* (4) achieved clear differentiation of four oils (cottonseed, canola, soybean, and peanut) by deriving a discriminant analysis equation based on four wavelengths (1704, 1802, 1816, and 2110 nm) from the second-derivative spectra. This type of simple analysis, based on selected wavelengths, holds particular attractions for practical reasons because it is comparatively inexpensive to design a system, based on filter technology, to record the required data. However, to be useful, there must be adequate discrimination between adulterant classes. Simple two-dimensional plots do not show any significant grouping of the spectra according to adulterant class. It is possible that grouping does occur in multidimensional space, but because the ISI software does not include a method of developing a discriminant analysis model based purely on wavelength measurements, it was not possible to investigate this analysis further.

An alternative method for developing a discriminant analysis model makes use of principal components analysis (PCA). PCA offers the possibility of using the data contained in the whole spectrum for the analysis, rather than selected wavelengths, by describing the variation in multidimensional data by means of a few uncorrelated variables (principal components). The principal components are linear combinations of the original spectral data that represent the maximum unexplained variation in the data. The loading plot gives some indication of which spectral region contributes to the associated principal component. In this case, all loading plots show most contribution in the 1100–1900 nm region of the spectrum, i.e., the first and second overtone region. Further than this, it is difficult to relate specific features in either the spectra or the loadings to specific oil types. Sato (5) considered the application of PCA to the NIR spectra of pure oils and suggested using the second derivative of the spectrum in the range 1600–2200 nm. Adequate discrimination of a wide range of oils was achieved by using the first two principal components. A similar analysis by Lai *et al.* (6), who used data from the mid-infrared region, suggests that the first five principal components should be used for discriminant analysis.

In this work, the Win-Discrim program was used to develop and evaluate a series of discriminant analysis models. 140 datapoints that covered the range 1100–2498 nm at 10-nm intervals were used for the analysis. The spectra are classed according to the adulterant type and the whole data set was pre-processed by using a covariance matrix to calculate the first 15 principal components of the data set. By this means, the data set is reduced to a more manageable size, while maintaining information about the class structure. Discriminant models were developed on the basis of the Mahalanobis distance (7) from the class means. Models were derived from 1–15 principal components, and the performance of each model was judged by the percentage of correct predictions of the calibration file (internal validation) and validation file (external validation). The results are presented in Figure 2. It shows that the predictive ability of the model does not improve after the addition of the 8th principal component.

Overfitting can be a problem when developing models of this type. To assess the significance of the predictions from a particular model, a random data set of the same size was created and predicted from the model. If the prediction rate of

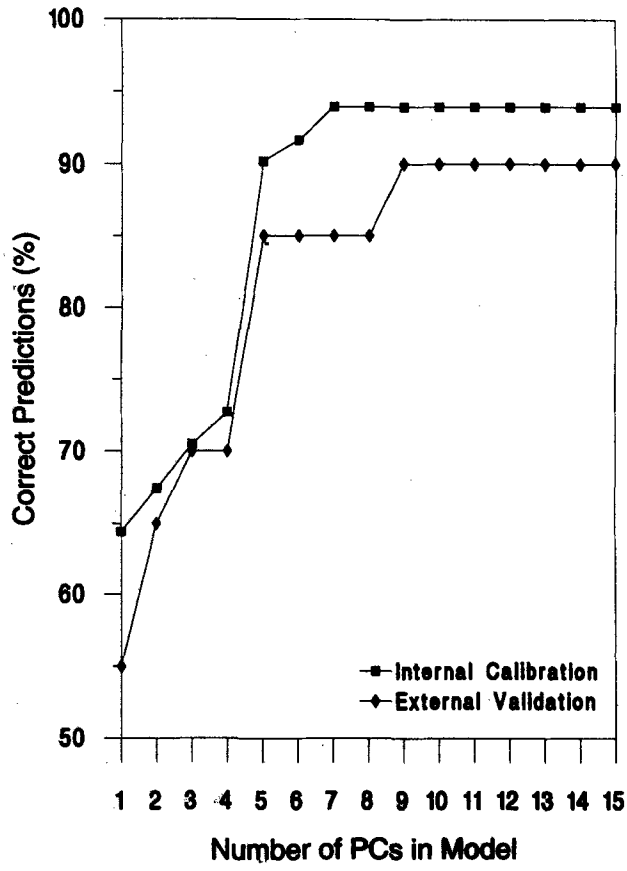

FIG. 2 Percentage of correct predictions of adulterant type against number of principal components (PCs) in model for the internal calibration and the external validation data sets.

the true data set is significantly greater than the mean of the predictions of the random data set, then the result is significant and overfitting is not a problem. A simpler test is to compare the prediction rates of the calibration and validation data sets. If the prediction rate of the validation set is greater than or equal to that of the calibration set, then overfitting is not a problem. With the 8PC model, the prediction rates were: calibration set = 93.94%, validation set = 90%, mean of random data = 39.62%, i.e., the results from the model are significant, and the model does not overfit the data.

On the basis of these results, a calibration method was developed with the ISI software. In this case, the principal components were calculated on each class of sample. The Mahalanobis distances of the unknown spectra were calculated from each class mean as before, and the spectrum was assigned on the basis of the lowest value Mahalanobis distance. This type of analysis works best when the individual classes form well- defined clusters. With this method of analysis, the prediction rate was 18/20 (90%). This was considered acceptable because the two samples that were incorrectly identified were pure oils. This class was the smallest of the four, and in this situation, it is not unusual for the prediction to be poor.

Prediction of purity. With the ISI software, an equation for purity was developed from the whole calibration file. The range 1100–2498 nm was used, with data points taken at 10-nm intervals. A first-derivative math treatment over a 12-nm gap and 4-point smoothing was used. Modified partial least squares analysis was used to develop a 10-term equation, which was used to predict the validation set. The use of the first derivative rather than no math treatment, as in the identification of adulterants, slightly improved the performance of the calibration. The results are shown in Figure 3 and given in Table 1. In this table, the number of spectra quoted is the number of spectra left after the elimination of spectral outliers by the software. Again, the samples that are poorly predicted are pure oils, which were not well represented in the calibration set. If these oils had been used to make some of the mixtures used for calibration, it is expected that they would be predicted more successfully. Removing these samples from the validation set improves the standard error of prediction to ±0.68% and the slope of the line of best fit to 1.00.

The ISI software allows both the purity prediction and adulterant prediction models to be used simultaneously. Thus, the results show that it is possible to develop an analytical system based on NIR spectroscopy and PCA that successfully predicts both the adulterant type and the purity of olive oils adulterated with seed oils.

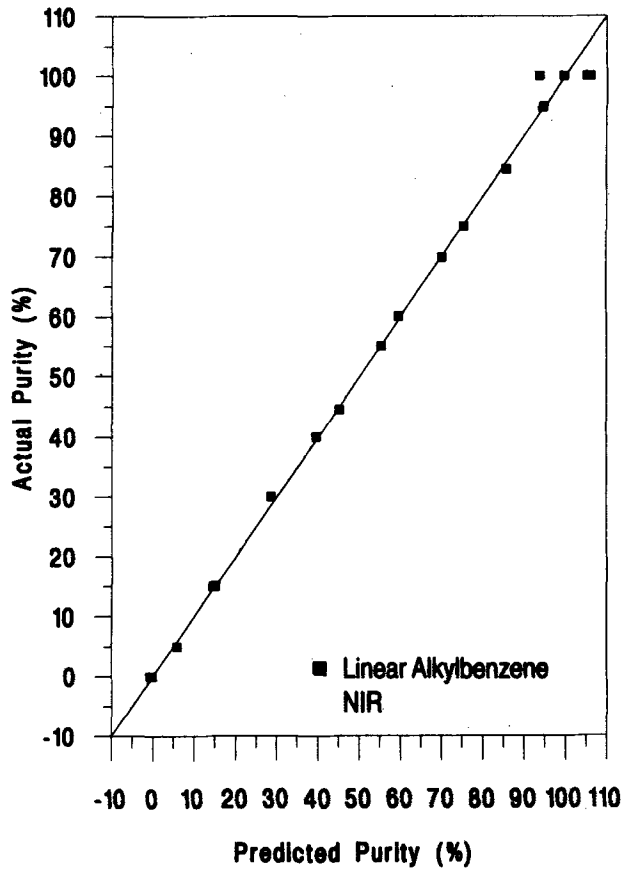

FIG. 3 Actual purity (%) against near infrared (NIR) predicted purity validation set of extra-virgin olive oils adulterated with seed oils. The line y = x is shown superimposed on the data.

TABLE 1
Calibration Statistics for Purity of Extra-Virgin Olive Oil

	Internal cross validation	External cross validation
Number of spectra	99	18
Sample range (%)	4.98–100.00	5.00–100.00
Mean of lab values (%)	49.94	65.76
Standard error of prediction (%)	0.90	2.77
r^2	1.00	0.99
Standard deviation (%)	28.34	33.79
Slope of line of best fit	1.00	0.97

ACKNOWLEDGMENTS

This work was supported by the University of Plymouth (IJW) and the European Union (FP). The assistance of E.K. Kemsley, Institute for Food Research (Norwich, United Kingdom) and Ralph Barnes, NIRSystems (Maidenhead, United Kingdom) is acknowledged.

REFERENCES

1. 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).
2. Adams, M.J., *Chemometrics in Analytical Spectroscopy*, Royal Society of Chemistry, Cambridge, 1995, pp. 123–154.
3. Brereton, R.G., *Chemometrics:Applications of Mathematics and Statistics to Laboratory Systems*, Ellis Horwood, Chichester, 1990.
4. Bewig, K.M., A.D. Clarke, C. Roberts, and N. Unklesbay, Discriminant Analysis of Vegetable Oils by Near-Infrared Reflectance Spectroscopy, *J. Am. Oil Chem. Soc. 71*:195–200 (1994).
5. Sato, T., Application of Principal-Component Analysis on Near-Infrared Spectroscopic Data of Vegetable Oils for Their Classification, *Ibid. 71*:293–298 (1994).
6. Lai, Y.W., E.K. Kemsley, and R.H. Wilson, The Potential of Fourier Transform Infrared Spectroscopy for the Authentication of Vegetable Oils, *J. Agric. Food Chem. 42*:1154–1159 (1994).
7. Mark, H.L., and D. Tunnell, Qualitative Near-Infrared Reflectance Analysis Using Mahalanobis Distances, *Analytical Chemistry 57*:1449–1456 (1985).

[Received April, 1995; accepted January 15, 1996]