Detecting and Quantifying Sunflower Oil Adulteration in Extra Virgin Olive Oils from the Eastern Mediterranean by Visible and Near-Infrared Spectroscopy

GERARD DOWNEY,*,[†] PETER MCINTYRE,[‡] AND ANTONY N. DAVIES[‡,§]

Teagasc, The National Food Centre, Ashtown, Dublin 15, Ireland, School of Applied Sciences, University of Glamorgan, Pontypridd, Mid-Glamorgan CF37 1DL, U.K., and Creon Lab Control AG, Europallee 27, D-50226 Frechen, Germany

One hundred and thirty-eight oil samples have been analyzed by visible and near-infrared transflectance spectroscopy. These comprised 46 pure extra virgin olive oils and the same oils adulterated with 1% (w/w) and 5% (w/w) sunflower oil. A number of multivariate mathematical approaches were investigated to detect and quantify the sunflower oil adulterant. These included hierarchical cluster analysis, soft independent modeling of class analogy (SIMCA method), and partial least squares regression (PLS). A number of wavelength ranges and data pretreatments were explored. The accuracy of these mathematical models was compared, and the most successful models were identified. Complete classification accuracy was achieved using 1st derivative spectral data in the 400−2498 nm range. Prediction of adulterant content was possible with a standard error equal to 0.8% using 1st derivative data between 1100 and 2498 nm. Spectral features and chemical literature were studied to isolate the structural basis for these models.

KEYWORDS: Olive oil; adulteration; classification; quantification; sunflower oil; SIMCA; partial least squares regression

INTRODUCTION

In 1994−95, olive oil represented approximately 3% of the world oils and fats market with ∼75% of global production taking place in Mediterranean countries (*1*). Olive oils (from *Olea Europa sativa*) are marketed according to the process used for their extraction (*2*). Virgin olive oils (extra virgin, virgin, ordinary virgin, and lampante virgin) are produced using only cold pressing techniques. They are the most sought after on account of their organoleptic and nutritional properties (*1*) and are therefore the most expensive grades. For this latter reason, they are a potential target for adulteration or mislabeling. The main adulteration issue involves addition of other cheaper oils (*3−5*), such as sunflower oil; this is similar in composition to olive oil and is often cultivated on the same farms. A significant labeling issue involves false claims concerning the geographic origin of an olive oil. A report on the efficacy of visible and near-infrared spectroscopy in addressing the latter issue is in preparation (Downey et al., unpublished data).

The main strategy for detecting adulteration in olive oils has been to compare the chemical compositions of suspect samples with the identity characteristics of olive oil grades promulgated by the European Commission (*6, 7*). This task is complex, as a result of the natural variation inherent in olive oil samples arising from differences in geographical origin, effects of weather during growth and harvesting, and so forth. Analytical methods investigated include determination of fatty acid profiles by gas liquid chromatography (*4*), high-pressure liquid chromatography (*8, 9*) measurement of slip points and iodine values (*10*), stable carbon isotope ratio analysis (*11*), and pyrolysis mass spectrometry (*12*). These techniques are time-consuming, expensive, and generally destructive of the sample material. Spectroscopic techniques have advantages in terms of speed and expense per test. Raman (*13*), ultraviolet (*14*), mid-infrared (*15−18*), and near-infrared (*19−24*) have all been studied for the quantification of adulterants.

Near-infrared spectroscopy has also demonstrated a capability for discrimination between sets of similar biological materials such as wheat flour (*25*), milk powders (*26, 27*), and coffee varieties (*28*). This paper discusses the potential of NIR technology (a) for discriminating between authentic extra virgin olive oils and the same oils adulterated by the addition of sunflower oil and (b) for quantifying the level of sunflower oil adulterant present. These questions were addressed by the application of soft independent modeling of class analogy (SIMCA) and partial least squares regression (PLS1) techniques.

MATERIALS AND METHODS

Samples. Olive oil and sunflower oil (not high oleic) samples were collected as part of a European Union supported research project (*29*);

* To whom correspondence should be addressed. Telephone: +353.1. 8059500. Fax: +353.1.8059550. E-mail: gdowney@nfc.teagasc.ie.
† Teagasc.
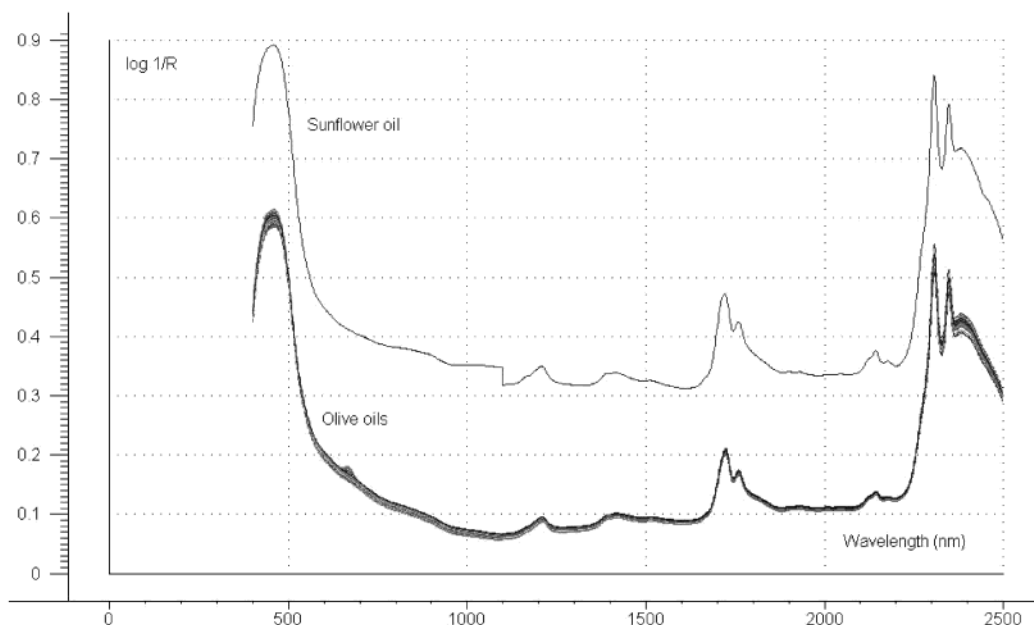‡ University of Glamorgan.
§ Creon Lab Control.

Sunflower Oil Adulteration in Extra Virgin Olive Oils

J. Agric. Food Chem., Vol. 50, No. 20, 2002 **5521**

Figure 1. Visible and near-infrared transflectance spectra of authentic and adulterated extra virgin olive oil samples ($n = 138$) plus authentic sunflower oil (offset for clarity).

they were collected at random from farm lots at oil mills in the three eastern Mediterranean regions by Elais S.A. (Peireus, Greece) and stored at −18 °C for 3 months. A total of 138 samples were available in this work. These comprised 46 authentic extra virgin olive oils from Greece and these same 46 samples adulterated by the addition of 1% (w/w) and 5% (w/w) authentic sunflower oil. Samples were thawed at room temperature and stored for a further 3 months at 18−24 °C in screw-capped glass vials in the dark prior to spectral acquisition.

Near-Infrared Spectroscopy. Visible and near-infrared transflectance spectra (400−2498 nm) were recorded on a NIRSystems 6500 scanning monochromator (FOSS NIRSystems, Silver Spring, MD) fitted with a sample transport accessory. Samples (0.5 mL approximately) were placed on the inside of the quartz window of a 0.1 mm path length camlock cell (part number IH-0355-1); this accessory is fitted with a circular quartz window and a gold-plated backing plate. The backing plate is constructed with flanges around its rim to produce the exact path length specified. Gaps in this flange permit the egress of excess sample to the rear of the plate. Instrument control and initial spectral manipulation were performed with WinISI II software (v1.04a; Infrasoft International, Port Matilda, MD). Oil spectra were recorded at an ambient temperature of between 18 and 24 °C. Between samples, the sample cell components were washed with detergent in warm water and rinsed with tepid tap water and then with distilled water at room temperature. Cells were dried using paper tissue. Oil samples were scanned in duplicate, with the mean spectra being used in all subsequent calculations.

Chemometrics. Raw spectral files were exported from WinISI in JCAMP.DX format (*30*) as described in ref *31* and then imported into Pirouette (v. 3.02; Infometrix Inc., Woodinville, WA) and The Unscrambler (v 7.6; CAMO A/S, Oslo, Norway) for data analysis. Spectral data were examined for unusual or outlying samples by principal component analysis (The Unscrambler). Preliminary analysis of the data set for natural groupings was performed by hierarchical cluster analysis in Pirouette. Classification was by single-category SIMCA (soft independent modeling of class analogy) also using Pirouette; full cross-validation was used to develop models. Samples were assigned to either a calibration or prediction set on the basis of their positions in the main spectral file; that is, even-numbered samples were used for model development, and odd-numbered samples were used for model evaluation. Unless otherwise stated, all classification results are at the 95% confidence level. Quantification of sunflower oil adulteration levels was by partial least squares regression (PLS1) in The Unscrambler; no outliers were removed during calibration. Full

cross-validation was used to develop and evaluate these regression models; model accuracy was assessed using the root-mean-square error of prediction.

Models were developed using three wavelength ranges: 400−2498 nm, 400−750 nm (visible), and 1100−2498 nm (near-infrared). The data pretreatments examined were none and 1st and 2nd derivatives (*32*); these were calculated using the Savitzky−Golay method and used two data points on either side of the measurement wavelength. In all cases, only optimal models, that is those using the number of components or loadings which produced the first local minimum in the prediction error versus number of loadings plot, are discussed in this paper.

RESULTS AND DISCUSSION

Spectra. The visible and near-infrared spectra of all oil samples are shown in **Figure 1**. As a quality control measure, the repeatabilities of these spectral recordings were measured by calculating the root-mean-square of the difference between replicates; mean values for segment 1 (400−1100 nm) and segment 2 (1100−2498 nm) were 2500×10^{-6} absorbance units. These values are similar to the within-day repeatability reported previously (*24*), although these latter authors used subsets of the spectral range 1100−2498 nm in their measurements.

Absorption maxima are clearly evident at 450, 672, 1210, 1722, 1760, 2310, 2346, and 2386 nm together with smaller absorption bands seen at 2124 and 2144 nm. Bands around 1200 nm arise from 2nd overtones of C−H stretching vibrations (*24*) while those at 1722 and 1760 nm are attributed to the first overtone of C−H stretching vibrations of methyl, methylene, and ethylene groups. Oleic acid has been reported to absorb at 1725 nm (*20*) while saturates and trans-unsaturated triglycerides have absorption maxima at 1725 and 1760 nm (*33*). Absorbances at 2310, 2346, and 2386 nm arise from combination bands arising from C−H stretching vibrations and other vibrational modes. Smaller peaks at 2124 and 2144 nm have been ascribed to the presence of cis double bonds (*34*).

A spectrum of pure sunflower oil is also shown in **Figure 1**; this spectrum is offset by +0.3 absorbance units for clarity. It is readily apparent that no significant difference exists between
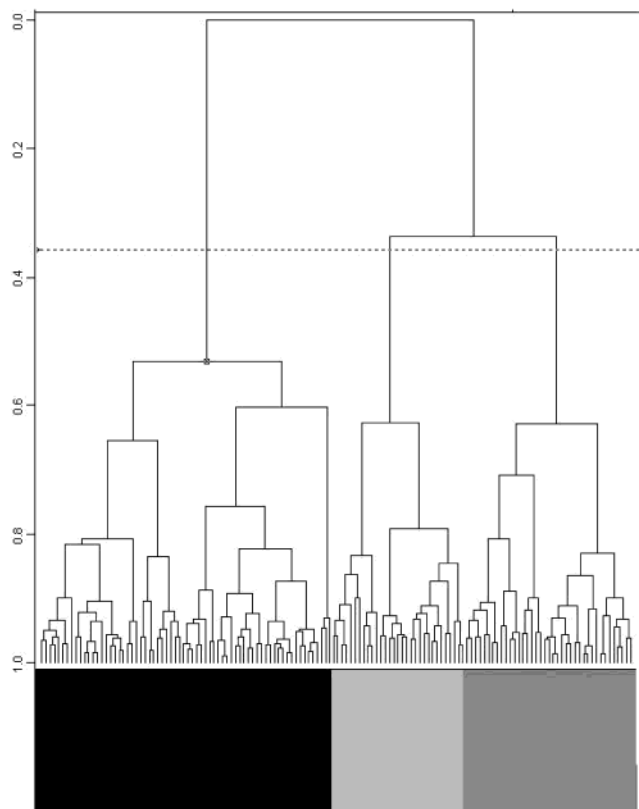
Figure 2. Dendrogram of entire oil data set showing three natural groupings at a similarity value of 0.365 (400–2498 nm wavelength range; raw data; incremental linkage method).

Table 1. Summary SIMCA Classification Results for the Entire Oil Sample Set (*n* = 115)[a]

wavelength range (nm)	data type	no. of PCs in optimal model	samples correctly classified	false negatives	false positives
400–2498	raw	5	112	0	3
	1st der[b]	5	115	0	0
	2nd der[c]	10	97	0	18
400–750	raw	3	113	0	2
	1st der[b]	3	80	0	35
	2nd der[c]	9	66	0	49
1100–2498	raw	2	97	3	15
	1st der[b]	5	112	3	0
	2nd der[c]	4	112	3	0

[a] Model developed using 50% (*n* = 23) of unadulterated olive oils. [b] 1st derivative. [c] 2nd derivative.

it and the olive oil spectra that can be detected by the naked eye. Olive and sunflower oils differ in their composition principally in their content of linoleic and oleic acids. In the case of olive oils, typical figures for these are 12.3 and 66.3%, respectively; for sunflower oil, mean values of 66.2 and 25.1%, respectively, have been quoted (*35*). Spectral differences between these oils have been reported in the mid-infrared region around 912 cm^{-1}.

Examination of 2-dimensional scores plots following a principal component analysis on data in the three wavelength ranges did not reveal any unusual or outlying samples.

Hierarchical Cluster Analysis. The presence of natural groupings in the spectral collection was evident when data in the wavelength ranges 400−2498 nm, 400−750 nm, or 1100−2498 nm were studied. Best results were obtained using an incremental linkage clustering method (*36*) which uses a sums of squares approach in calculating intercluster distances. This method minimizes dispersion within groups and favors the formation of small clusters of approximately equal size. Experience has shown that the centroidal methods, incremental in particular, are more adept at finding clusters when groups overlap, contain roughly equal numbers of similar samples, and are shaped as ellipsoids rather than long strands (S. Ramos, Infometrix, personal communication). Using raw data in the 400−2498 nm wavelength range, a dendrogram revealing three clusters at a similarity value of 0.365 is shown in **Figure 2**. Examination of the samples contained in each grouping shows the leftmost and middle groups to comprise olive oil adulterated with 5 and 1% w/w sunflower oil, respectively; some overlap occurs between these two groups. The group on the extreme right comprises pure extra virgin olive oils and some oils

adulterated at the 1% level. Detection of such natural groupings suggests that discrimination between at least some of them may be possible.

Classification. Forty-six (46) of the samples comprised unadulterated extra virgin olive oils. A principal component model to describe this oil type was developed using half of these samples, that is every alternate such sample in the spectral file. Models were developed using raw, 1st derivative, and 2nd derivative spectral data. These models were then deployed to classify all of the oil samples (*n* = 138). Summary data describing these models and their performance are shown in **Table 1**. With the exception of the 1st and 2nd derivative spectra for the 1100−2498 nm range, all of the authentic olive oils were correctly identified. False positive results (i.e. when an adulterated sample is classified as authentic) represent a serious classification error. The use of derivatives of visible spectral data produced a very high level of such errors. Overall, the best model derived was that produced by a 1st derivative treatment of spectral data over the entire wavelength range; this gave a 100% correct classification at the 99% probability level. Models based on 1st and 2nd derivatives of spectral data in the range 1100−2498 nm produced models of similar accuracy—only 1 false negative and no false positives.

Repetition of this classification study using all of the unadulterated olive oils for SIMCA model development did not change the overall pattern of results. The main variation was a reduction in the number of false positive classifications for each model. The number of principal components required to describe the various SIMCA models exhibited a variation of ±1 when using the complete (*n* = 46) or attenuated (*n* = 23) sample calibration set.

Quantification. Quantification of the sunflower oil contents of the adulterated oil samples was performed using PLS1 regression. As in the qualification study described above, raw and derivatized spectral data were investigated separately. A summary of the results is shown in **Table 2**. These show that the visible wavelength range produced the least accurate prediction equations, irrespective of data pretreatment. Overall, the best prediction results were obtained using either raw spectral data in the 400−2498 nm range or 1st or 2nd derivative data in the 1100−2498 nm range; these models produced a root-mean-square error of prediction approximately equal to 0.8% and a correlation coefficient equal to 0.93. Detailed study of these regression models revealed that the use of 1st derivative data in the wavelength range 1100−2498 nm may prove to be the best choice. The optimum number of loadings to be used was clearer with this data set, and an obvious segregation of the unadulterated versus adulterated samples was apparent in the
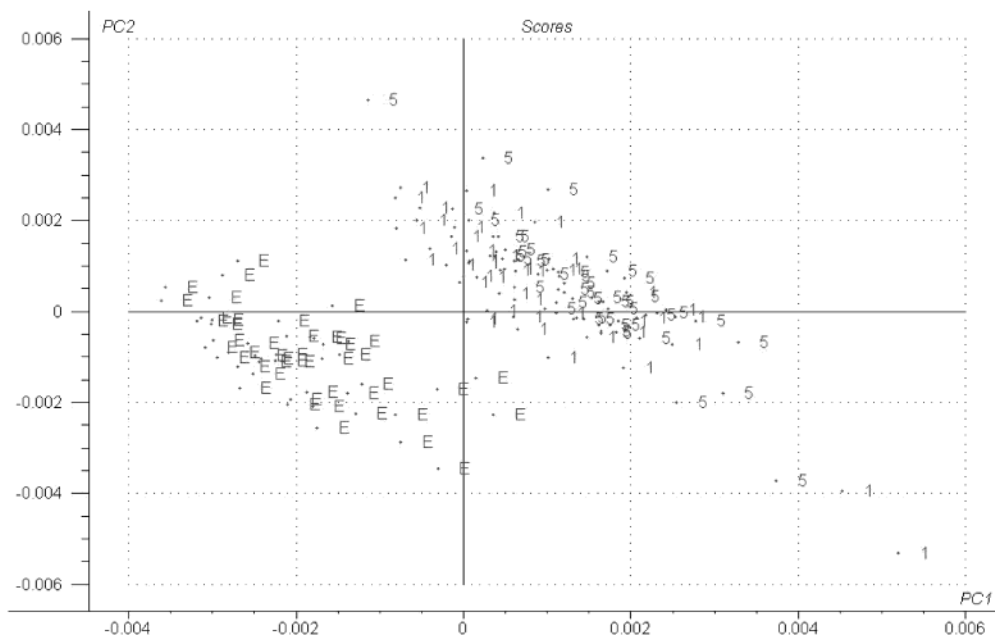
Figure 3. Sample scores plot on PLS loadings 1 and 2 (1st derivative spectral data; 1100–2498 nm; E = pure olive oils; 1 = 1% w/w sunflower adulteration; 5 = 5% w/w sunflower adulteration).
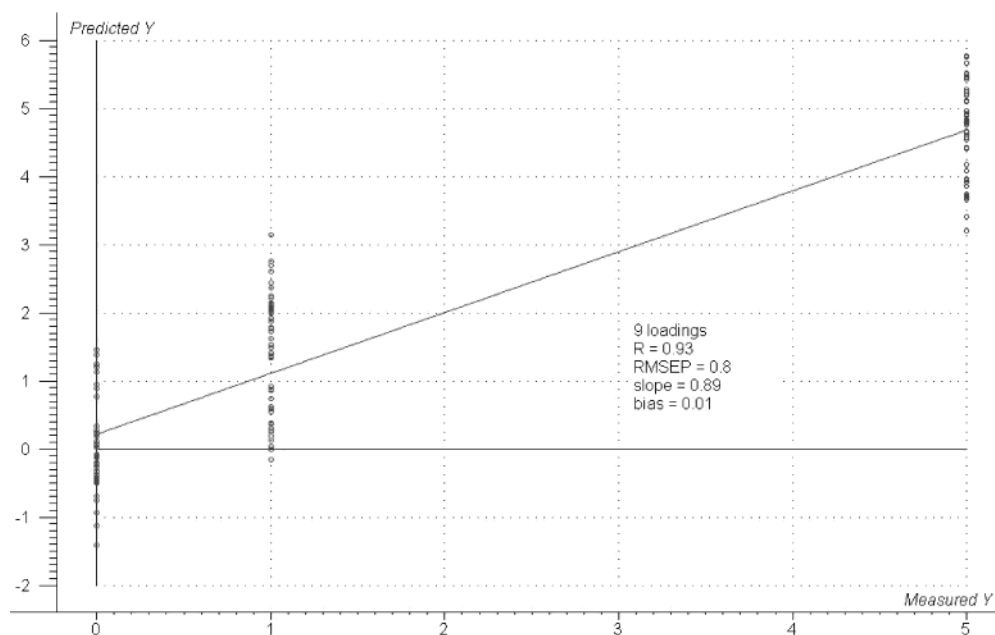

Figure 4. PLS regression of predicted vs actual sunflower oil content in authentic and adulterated extra virgin olive oil samples (n = 115; 1100–2498 nm wavelength range; 1st derivative pretreatment).

multivariate space defined by the first two PLS loadings (**Figure 3**). A graphical display of the regression produced using this model is shown in **Figure 4**. Given that the range in sunflower oil content of this data set is equal to 5.0%, the range/error ratio (RER) for this optimal calibration is calculated to be 16. This indicates that it is suitable for practical use. The limit of detection of sunflower oil adulteration is equal to 1.6% using this model.

PLS loadings 1 and 2 and 3 and 4 are shown in **Figures 5** and **6**, respectively. Loadings 1 and 2 account for 61 and 34% of the variance in spectral data, respectively, together with 39% and 11% of the Y data (sunflower oil content), respectively. These loadings mainly utilize spectral data in the regions around 1200, 1400, 1700, and 2300 nm. Given the general similarity

Table 2. Prediction of Sunflower Oil Content in Adulterated Olive Oils[a]

wavelength range (nm)	data type	no. of loadings in optimal model	RMSEP[b]	R[c]
400–2498	raw	8	0.79	0.93
	1st der	7	0.96	0.90
	2nd der	4	1.24	0.82
400–750	raw	2	1.42	0.75
	1st der	4	1.06	0.87
	2nd der	2	1.45	0.74
1100–2498	raw	8	0.9	0.91
	1st der	9	0.8	0.93
	2nd der	7	0.78	0.93

[a] Partial least squares regression with full cross-validation. [b] Root-mean-square error of prediction. [c] Correlation coefficient.

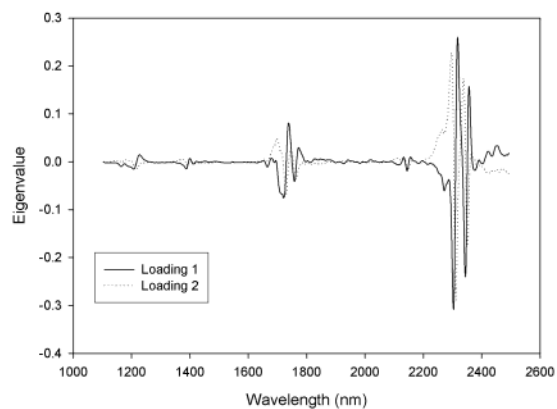**5524** *J. Agric. Food Chem.,* Vol. 50, No. 20, 2002

Downey et al.

Figure 5. PLS loadings 1 and 2 (sunflower oil content prediction; 1100–2498 nm wavelength range; 1st derivative data pretreatment).
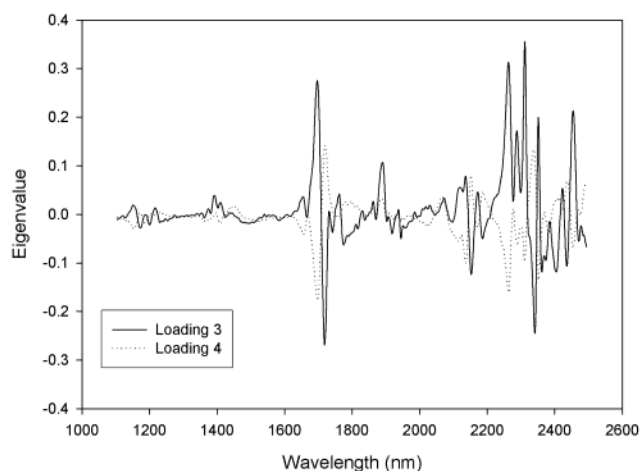

Figure 6. PLS loadings 3 and 4 (sunflower oil content prediction; 1100–2498 nm wavelength range; 1st derivative data pretreatment).

of sunflower and olive oils in chemical composition, it is difficult to relate any of the important wavelengths in these loadings plots to specific oil components, although oleic acid, found in greater amounts in olive oils, is reported (20) to exhibit an absorption peak around 1725 nm.

CONCLUSIONS

Using a limited number of authentic and adulterated oils, it has been demonstrated that the multivariate classification method SIMCA (soft independent modeling of class analogy) can successfully discriminate between authentic extra virgin olive oils and the same oils adulterated with sunflower oil at levels as low as 1% (w/w). It should be stressed that this work has only involved a single sunflower oil sample. Natural variations in chemical composition between samples of this oil will occur, for example, between varieties, harvest, geographic locations, and so forth. Therefore, the results reported in this work require extension to a greater number of sunflower and olive oil samples before they can be completely endorsed. The greatest classification accuracy was achieved using the 1st derivative of spectral data in the wavelength range 1100−2498 nm. Using a confidence level of 1%, 100% correct classification was achieved in both calibration and prediction sample sets. Once adulteration has been detected, quantification of the sunflower oil adulterant was achieved by partial least squares regression with a cross-validation prediction error equal to 0.8%. This level of accuracy is suitable for industrial use.

LITERATURE CITED

(1) Aparicio, R.; McIntyre, P. Oils and Fats. In *Food Authenticity: Issues and Methodologies*; Lees, M., Ed.; Eurofins Scientific: France, 1998; pp 231−269.

(2) International Olive Oil Council COI/T.15 no.2, Madrid, Spain, 1996.

(3) Firestone, D.; Summers, J. L.; Reina, R. J.; Adams, W. S. Detection of adulterated and misbranded olive oil products. *J. Am. Oil Chem. Soc.* **1985**, 62, 1558−1562.

(4) Firestone, D.; Carson, K. L.; Reina, R. J. Update on control of olive oil adulteration and misbranding in the United States. *J. Am. Oil Chem. Soc.* **1988**, 65, 788−792.

(5) Salivaras, M.; McCurdy, A. R. Detection of olive oil adulteration with canola oil from triacyglycerol analysis by reversed-phase high-performance liquid chromatography. *J. Am. Oil Chem. Soc.* **1992**, 69, 935−938.

(6) European Union. *Official Journal of the Commission of the European Communities*; Regulation No. 2632/94, L280, 1996.

(7) European Union. *Official Journal of the Commission of the European Communities*; Regulation No. 656/95, 1995.

(8) Cortesi, N. High performance liquid chromatography (HPLC) in the analysis of olive oil. *Olivae* **1993**, 45, 40−42.

(9) Mariani, C.; Fedeli, E. Gas chromatography in olive oil analysis. *Olivae* **1993**, 45, 34−39.

(10) Bartlett, J. G.; Mahon, J. H. Identification of oils and the detection of oil adulteration by differential IR spectroscopy. *J.−Assoc. Off. Anal. Chem.* **1958**, 41, 450−459.

(11) Rossell, J. B. Stable carbon isotope ratios in establishing maize oil purity. *Fat Sci. Technol.* **1994**, 96, 304−308.

(12) Goodacre, R.; Kell, D. B.; Bianchi, G. Rapid assessment of the adulteration of virgin olive oils by other seed oils using pyrolysis mass spectrometry and artificial neural networks. *J. Sci. Food Agric.* **1993**, 63, 297−307.

(13) Davies, A. N.; McIntyre, P.; Morgan, E. Study of the use of molecular spectroscopy for the authentication of extra virgin olive oils. Part 1: Fourier transform Raman spectroscopy. *Appl. Spectrosc.* **2000**, 54, 1864−1867.

(14) Calapaj, R.; Chiricosta, S.; Saija, G. Evaluation of gas chromatographic and spectrophotometric analytical results to check the presence of seed oils in olive oil samples. *Riv. Ital. Sostanze Grasse* **1993**, 70, 585−594.

(15) Lai, Y. W.; Kelmsley, E. K.; Wilson, R. H. Quantitative analysis of potential adulterants of extra virgin olive oil using IR spectroscopy. *Food Chem.* **1995**, 53, 95−98.

(16) Dupuy, N.; Duponchel, L.; Huvenne, J. P.; Sombret, P.; Legrand, P. Classification of edible fats and oils by principal component analysis of Fourier transform infrared spectra. *Food Chem.* **1996**, 57, 245−251.

(17) Guillen, M. D.; Cabo, N. Usefulness of the frequencies of some Fourier transform infrared spectroscopic bands for evaluating the composition of edible oil mixtures. *Fett/Lipid* **1999**, 101, 71−76.

(18) Küpper, L.; Heise, H. M.; Lampen, P.; Davies, A. N.; McIntyre, P. Authentication and quantification of extra virgin olive oils by attenuated total reflectance infrared spectroscopy using silver halide fiber probes and partial least-squares calibration. *Appl. Spectrosc.* **2001**, 55, 563−570.

(19) Wesley, I. J.; Barnes, R. J.; McGill, A. E. J. Measurement of adulteration of olive oils by near-infrared spectroscopy. *J. Am. Oil Chem. Soc.* **1995**, 72, 289−292.

(20) Cho, R. K.; Iwamoto, M. *The purity identification of sesame oil by near-infrared reflectance spectroscopy*; Proceedings of the 2nd International NIRS Conference, Tsukuba, Japan; Iwamoto, M., Kawano, S., Eds.; Korin Publishing Company: Tokyo, Japan, 1990; pp 142−149.

(21) Bewig, K. M.; Clarke, A. D.; Roberts, C.; Unklesbay, N. Discriminant analysis of vegetable oils by near-infrared reflectance spectroscopy. *J. Am Oil Chem. Soc.* **1994**, *71*, 195−200.

(22) Sato, T. Application of principal component analysis on near-infrared spectroscopic data of vegetable oils for their classification. *J. Am. Oil Chem. Soc.* **1994**, *71*, 293−298.

(23) Wesley, I. J.; Pacheco, F.; McGill, A. E. J. Identification of adulterants in olive oil. *J. Am. Oil Chem. Soc.* **1996**, *73*, 515−518.

(24) Hourant, P.; Baeten, V.; Morales, M. T.; Meurens, M.; Aparicio, R. Oil and fat classification by selected bands of near-infrared spectroscopy. *Appl. Spectrosc.* **2000**, *54*, 1168−1174.

(25) Downey, G.; Spengler, B. Compositional analysis of coffee blends by near-infrared spectroscopy. *Ir. J. Agric. Food Res.* **1996**, *35*, 179−188.

(26) Downey, G.; Robert, P.; Bertrand, D.; Kelly, P. Classification of commercial skim milk powders according to heat treatment using factorial discriminant analysis of near infrared reflectance spectra. *Appl. Spectrosc.* **1990**, *44*, 150−155.

(27) Maraboli, A.; Cattaneo, T. M. P.; Giangiacomo, R. Detection of vegetable proteins from soy, pea and wheat isolates in milk powder by near-infrared spectroscopy. *J. Near Infrared Spectrosc.* **2002**, *10*, 63−69.

(28) Rannou, H.; Downey, G. Discrimination of raw pork, chicken and turkey meat by spectroscopy in the visible, near- and mid-infrared ranges. *Anal. Commun.* **1997**, *34*, 401−404.

(29) EU Contract No. AIR-2-CT 92-1224. Responsible Scientific Officer Dr. Xabier Goenaga, Directorate-General XII, Brussels, Belgium, 1992.

(30) McDonald, R. S.; Wilks, P. A. A standard form for exchange of infrared spectra in computer readable form. *Appl. Spectrosc.* **1988**, *42*, 151−162.

(31) Rutledge, D. N.; McIntyre, P. A proposed European implementation of the JCAMP. DX data file transfer format. *Chemom. Intell. Lab. Syst.* **1992**, *16*, 95−101.

(32) Hruschka, R. Data Analysis: Wavelength selection. In *Near Infrared Technology in the Agricultural and Food Industries*; Williams, P., Norris, K., Eds.; American Association of Cereal Chemists: St. Paul, MN, 1987; p 47.

(33) Sato, T.; Kawano, S.; Iwamoto, M. Near-IR spectral patterns of fatty acid analysis from fats and oils. *J. Am. Oil Chem. Soc.* **1991**, *68*, 827−833.

(34) Holman, R. T.; Edmondson, P. G. Near-Infrared spectra of fatty acids and some related substances. *Anal. Chem.* **1956**, *28*, 1533−1538.

(35) Neumüller, O.-A., Ed. *Römpps Chemie-Lexikon*, 8th ed.; Franckh'sche Verlagshandlung: Stuttgart, 1985.

(36) Ward, J. J., Jr. Hierarchical grouping to optimise an objective function. *J. Am. Stat. Assoc.* **1963**, *58*, 236−244.