Near- and Mid-Infrared Spectroscopies in Food Authentication: Coffee Varietal Identification

Gerard Downey,*,† Romain Briandet,‡ Reginald H. Wilson,‡ and E. Katherine Kemsley‡

Teagasc, The National Food Centre, Dunsinea, Castleknock, Dublin 15, Ireland, and Institute of Food Research, Norwich Research Park, Colney Lane, Norwich NR4 4UA, United Kingdom

Near- and mid-infrared spectra of a set of 56 lyophilized coffee samples were collected. The sample set comprised 29 Arabica and 27 Robusta coffees and was prepared in the laboratory. Each spectral collection was used separately to develop mathematical models for varietal authentication of the coffees using factorial discriminant analysis and partial least squares techniques. Subsequently, data from both spectral regions were combined and the chemometric approaches repeated. The relative success of the separate and combined approaches is discussed, as is the basis for the observed discriminations.

Keywords: *Food authentication; spectroscopy; near-infrared; mid-infrared*

INTRODUCTION

Food authentication is a wide-ranging issue that has come to prominence in recent years. Regulatory authorities, food processors, retailers, and consumer groups are all actively interested in ensuring that foods on retail sale are what they purport to be. Many foods have the potential to be adulterated, but those that are expensive and have production levels which vary as a natural result of fluctuations in weather and harvest conditions are particularly susceptible to this practice. Coffee is one such material, all the more so in recent times due to the increasing practice of selling coffees on the basis of their varietal and/or geographic origin.

While a number of approaches to coffee authenticity have been reported (Speer *et al.*, 1991; Frega *et al.*, 1994; Tressl *et al.*, 1982), those based on spectroscopic techniques have much to commend them due to their ease of use, speed, and relative inexpense. Previous papers have described the application of near-infrared (NIR) (Downey *et al.*, 1994; Downey and Boussion, 1996; Downey and Spengler, 1996) and mid-infrared (MIR) spectroscopy (Briandet *et al.*, 1996a,b; Kemsley *et al.*, 1995; Suchánek *et al.*, 1996) to the problem of coffee authentication; however, these research groups used different samples and investigated different authenticity issues. Given that they appear to have considerable promise for the solution of this problem and that the information in both spectral regions is analogous but not identical, it seemed appropriate to examine these spectroscopies separately and in combination on a single set of samples to enable a valid comparison of each for qualitative analysis and to determine the relative advantage, if any, of the use of combined spectra. The latter approach has been reported to be advantageous for the quantitative estimation of a limited number of forage constituents (Reeves, 1996), while NIR and MIR spectra have been analyzed for the assignment of NIR absorption bands (Barton and Himmelsbach, 1992). As far as we are aware, this is the first publication describing the use of combined NIR−MIR spectra for qualitative analysis.

* Author to whom correspondence should be addressed (telephone +353.1.805 9533; fax +353.1.805 9550; e-mail g.downey@nfc.teagasc.ie).

† Teagasc.

‡ Institute of Food Research.

MATERIALS AND METHODS

Coffee Samples. Roasted coffee beans were obtained from a number of commercial sources. Samples originated from a number of different countries and had been roasted to various degrees. No independent confirmation of their variety was made. Coffees were ground and processed to produce lyophilized powders according to a previously described process (Downey and Boussion, 1996); samples were stored in screw-capped plastic containers and NIR spectra collected within days of production. MIR spectra were collected several weeks later; in the interim, the coffee powders were stored at −20 °C. The 56 samples (29 Arabica and 27 Robusta) were divided into calibration development (20 Arabica and 18 Robusta) and evaluation (9 Arabica and 9 Robusta) sets. These sample sets were used throughout this study for the separate and combined spectral investigations.

NIR Spectroscopy. Spectra were recorded in reflectance mode using an NIRSystems 6500 instrument (NIRSystems Inc., Silver Spring, MD). Measurements were made at ambient temperature (10−20 °C) over the wavelength range 400−2498 nm at 2 nm intervals. A standard quartz cell was used; each lyophilized powder sample was scanned in duplicate (including a repack), and the mean spectrum was stored for further data processing. Instrument control and file manipulation were performed using NIRS3 software (ISI International, Port Matilda).

MIR Spectroscopy. All spectra were collected using a Monit-IR (Spectra-Tech, Applied Systems Inc.) Fourier transform infrared (FTIR) spectrometer, operating in the region 800−4000 cm^{-1}, equipped with a sealed, desiccated interferometer compartment and a deuterated triglycine sulfate detector. The spectrometer incorporates two integral sampling stations, one of which is designed for the diffuse reflectance sampling technique (DRIFT; Wilson, 1990), and comprises permanently mounted, optimized transfer optics, and sealed potassium bromide (KBr) windows to minimize the path of the infrared beam through the external atmosphere. All spectral measurements were made at nominal 8 cm^{-1} resolution, with 64 interferograms co-added before Fourier transformation. Each sample was ground for 5 min with a pestle and mortar before it was loaded into the diffuse reflectance sample cup. Taking care not to compress the sample, the surface was leveled in a predetermined direction with a spatula before analysis. Single-beam DRIFT spectra of each sample were collected and transformed to Kubelka−Munk units using a background spectrum of ground KBr.

S0021-8561(97)00337-3 CCC: $14.00

Chemometric Operations. *Spectral Transfer and Manipulation.* Both sets of spectra were exported as JCAMP-DX files and imported directly into Unscrambler (v 5.53; CAMO A/S, Trondheim, Norway) or via a specific JCAMP-DX format (Rutledge and McIntyre, 1992) into SAISIR (D. Bertrand, INRA, Nantes, France). The frequency (wavenumber) scale of the MIR spectra was converted to a wavelength scale before combination with the NIR data. No mathematical pretreatments (e.g. scatter correction) were applied to the spectra. Both spectral collections were analyzed for outlying or unusual samples by principal component analysis (PCA); none were found. After NIR and MIR spectra were merged, the combined spectra were normalized by division of readings at each wavelength in each spectrum by the standard deviation at each wavelength in the entire spectral collection. Spectra (individual and combined) for a single coffee sample are shown in Figure 1; the combined spectrum (Figure 1c) shown has been normalized for clarity.

Factorial Discriminant Analysis (FDA). Following principal component analysis of the spectral data, principal component scores for each sample were used as the basis for a FDA in SAISIR. The latter develops combinations of the principal component scores in proportions characteristic of the discrimination involved (Devaux *et al.*, 1988). Principal components 1−20 were calculated, and to guard against overfitting, discriminant models were developed on a calibration sample set and evaluated on a separate prediction sample set. (The composition of these sets is detailed below.) Gravity centers are calculated during model development and represent the mean value of discriminant scores in each cluster. To understand the molecular mechanism responsible for any observed discrimination, profiles of each discriminant function may be examined. These functions are combinations of the principal component loadings (eigenvectors) in proportions which best describe the discrimination achieved, and the relative magnitude of the function at each wavelength is in proportion to the importance of that wavelength in the discrimination process.

Partial Least Squares (PLS). PLS regression (Martens and Naes, 1989) was used for the quantitation of Robusta content in the coffee samples. Each variety was assigned a dummy variable; that is, Arabica samples were given the value 0.01 and Robusta coffees the value 1.0. This multivariate technique decomposes the spectral data matrix into a reduced number of orthogonal loadings. Unlike PCA, in which the components are developed from the spectral data only, in PLS these loadings are developed to describe the main variation in both the spectral and associated compositional data matrices (Martens and Naes, 1989). This approach is reported to produce loadings of greater relevance in the description of associated variables than PCA. Cross-validation was used to minimize the risks of overfitting the calibrations and evaluate calibration accuracy.

RESULTS AND DISCUSSION

FDA. *NIR Data.* Using the wavelength range 1100−2498 nm, three principal components were required for 100% correct identification of the calibration and evaluation sample sets; the principal components involved were numbers 4, 5, and 2 in that order, and they accounted for 0.3, 0.1, and 3.6%, respectively, of the total variance in the spectral data set. The clear separation of the Arabica and Robusta samples in the discriminant space may be seen in Figure 2, while the profile of the discriminant function responsible is shown in Figure 3. Its main features are troughs at 1673, 1917, and 2244 nm, which are close to peaks found in caffeine (Downey and Boussion, 1996). It therefore appears that variations in the content of caffeine or other alkaloids may be the basis for this observed discrimination; these results agree well with previous work on a different sample set (Downey and Boussion, 1996).

MIR Data. Using spectra in the range 5000−12 500 nm, a three principal component model was required

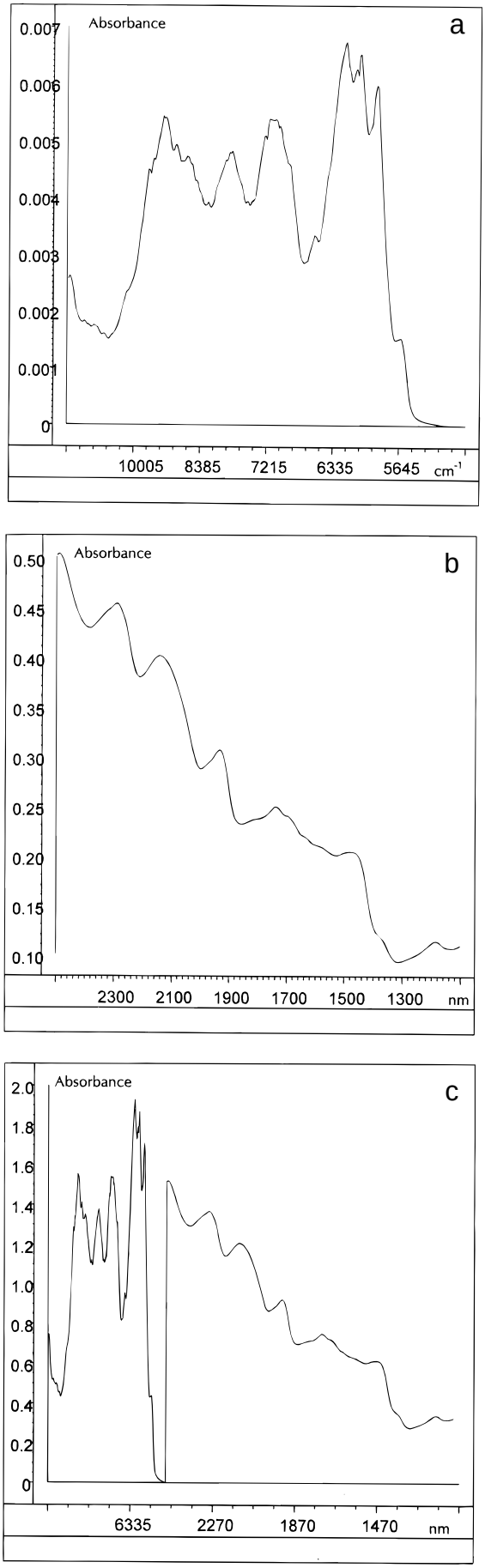

Figure 1. Reflectance spectra of lyophilized coffee: (a) MIR; (b) NIR; (c) combined and normalized (direction of NIR wavelength has been reversed to allow connection to MIR scale).

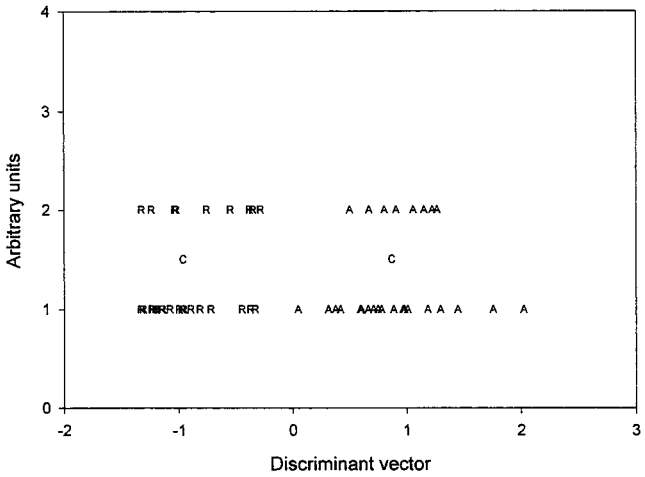

Figure 2. Discriminant scores plot for NIR spectra (A, Arabica; R, Robusta; C, cluster center). Samples have been offset vertically for greater clarity. Arbitrary unit 1 = calibration sample set; arbitrary unit 2 = evaluation sample set.

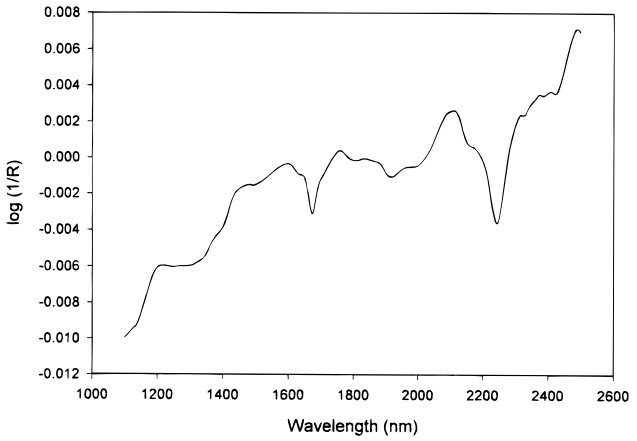

Figure 3. Discriminant profile for NIR spectra.

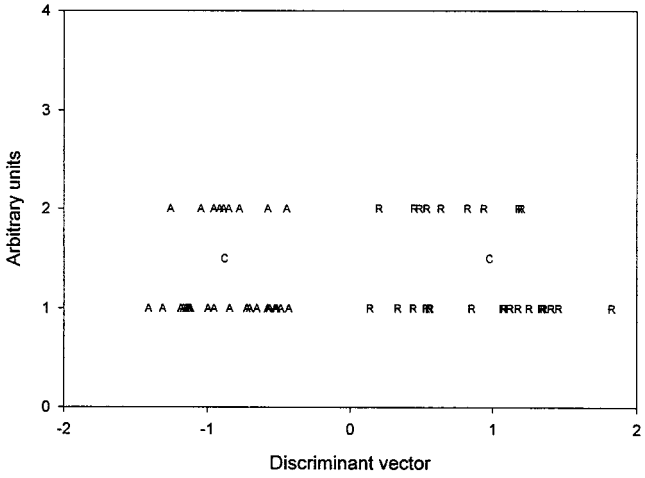

Figure 4. Discriminant scores plot for MIR spectra (A, Arabica; R, Robusta; C, cluster center). Samples have been offset vertically for greater clarity. Arbitrary unit 1 = calibration sample set; arbitrary unit 2 = evaluation sample set.

for the complete separation of Arabica and Robusta coffees. Principal components introduced were 2, 3, and 8 and described 28.2, 6.9, and 0.8%, respectively, of the variance in the spectral data set. The scores plot for all 56 samples may be seen in Figure 4, while the spectral profile of the discriminant function is shown in Figure 5. This vector is found to be similar to a spectrum of a mix of pure caffeine and chlorogenic acid. Most of the large features can be attributed to one or

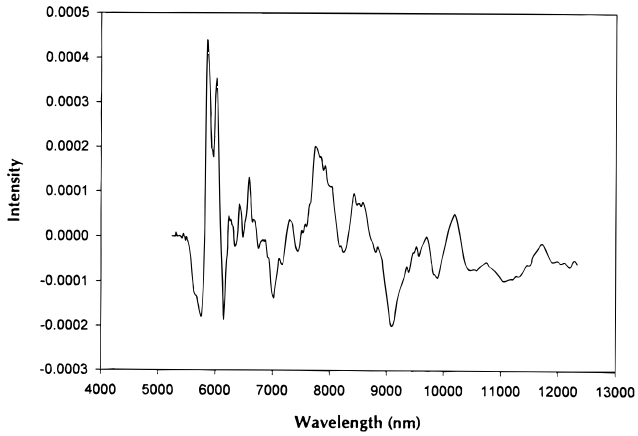

Figure 5. Discriminant profile for MIR spectra.

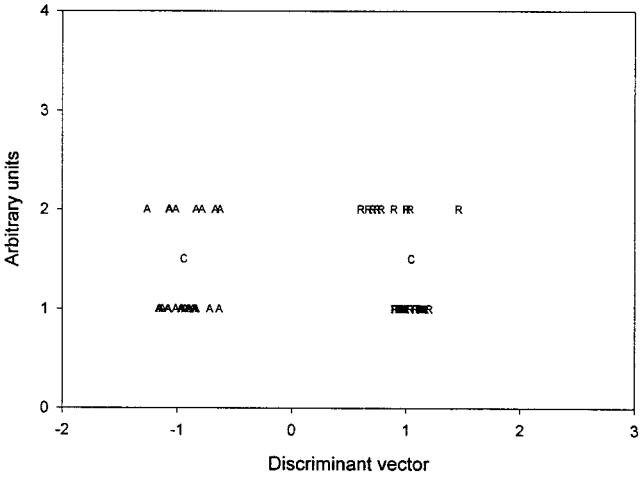

Figure 6. Discriminant scores plot for combined NIR−MIR spectra (A, Arabica; R, Robusta; C, cluster center). Samples have been offset vertically for greater clarity. Arbitrary unit 1 = calibration sample set; arbitrary unit 2 = evaluation sample set.

the other of these compounds, although the relative intensities of the bands differ somewhat. Caffeine is responsible for the large bands in the region 5700−6450 nm, while the chlorogenic acid has major bands in the region 7700−8700 nm (Briandet *et al.*, 1996a). Freeze-dried coffee contains 5.2−7.4% chlorogenic acid on a dry weight basis (dwb) and 4.5−5.1% dwb caffeine (Smith, 1985), although the relative contents of these compounds in the different species is not well-known in coffee products.

Although particle size can often affect the quality of MIR data and various treatments such as MSC or area normalization are often used to improve repeatability, we did not use such treatment here. Essentially what is important in the spectra are relative intensities rather than the absolute intensity at any given position. In principal component-based methods, one would expect the effects of intensity variation and baseline to be accounted for by one or more principal components. The approach used avoids any potential problems that may arise from overly enthusiastic preprocessing.

NIR and MIR Combined Spectra. Spectra in both spectral regions were combined and normalized before being input into the discriminant process. Using combined spectra, a single component, principal component 2, was sufficient to completely segregate the coffee varieties. The scores plot for this model is shown in Figure 6, and it is noticeable that the spread of samples around each cluster center is much less than is the case with either spectral region alone. The discriminant

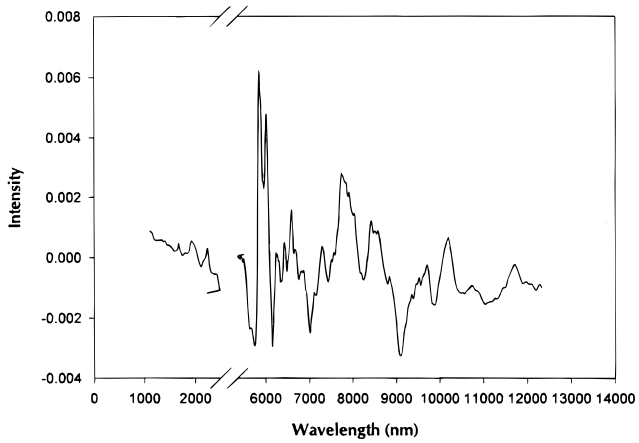

Figure 7. Discriminant profile for combined NIR−MIR spectra.

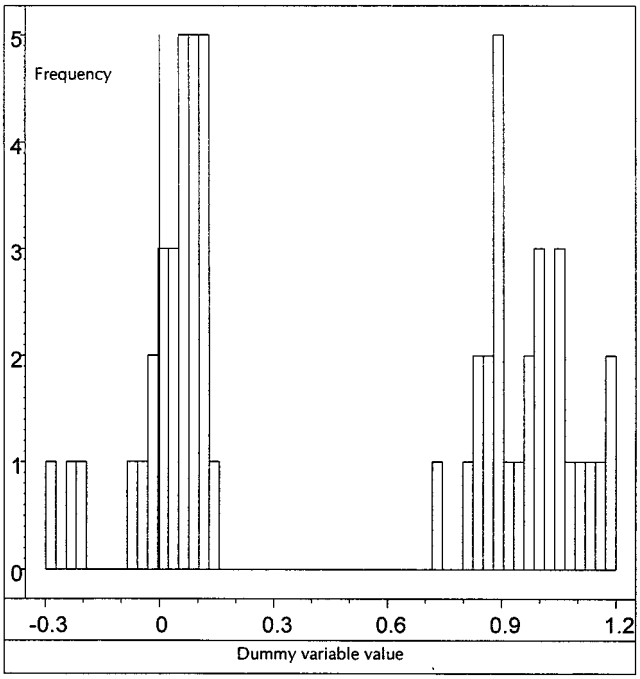

Figure 8. Histogram of coffee samples following PLS prediction, combined NIR−MIR spectra. Abscissa, "dummy" variable predicted value; ordinate, number of samples.

profile may be seen in Figure 7. In the NIR portion of this profile, major features include peaks at 1673, 1931, and 2240 nm, which are very close to those reported above for the NIR model alone. In the case of the MIR region, the major features are virtually identical to the vector shown in Figure 5 and, again, arise from caffeine and chlorogenic acid.

PLS Regression. Using NIR data recorded over the wavelength range 1100−2498 nm, a seven-loading PLS1 model was determined to be optimal for prediction of the dummy variable. (An optimal model is one that produces the first minimum in a graph of standard error of prediction versus number of components incorporated in the model.) In the case of MIR spectra, the optimal PLS regression model contained five loadings, two fewer than was the case for NIR spectra. When spectra from both spectral ranges were combined and normalized, a five-loading model was optimum. A histogram of sample composition produced by this combined spectral model is shown in Figure 8; histograms produced using only NIR or MIR spectral data produced degrees of separation similar to those of Figure 8.

Conclusions. The results reported above reveal that both NIR and MIR spectroscopies have the potential to discriminate between Arabica and Robusta lyophilized coffees. However, the combination of the two spectral regions appears to offer advantages in relation to possible model robustness, since fewer loadings or principal components were required than was the case with either technique alone. This observation is valid for both of the chemometric approaches tested (FDA and PLS) and suggests that combined spectroscopic investigations may have advantages for this and other qualitative problems. The work reported herein is only, of course, a feasibility study and requires extension to considerable numbers of commercial samples before its value may be truly estimated.

ACKNOWLEDGMENT

We thank Bewley's Ltd., Dublin, for the supply of coffee samples and for technical advice.

LITERATURE CITED

Barton, F. E., II; Himmelsbach, D. S. Two-dimensional vibrational spectroscopy II: Correlation of the absorptions of lignins in the mid- and near-infrared. *Appl. Spectrosc.* **1993**, *47* (11), 1920−1925.

Briandet, R.; Kemsley, E. K.; Wilson, R. H. Discrimination of Arabica and Robusta in Instant Coffee by Fourier Transform Infrared Spectroscopy and Chemometrics. *J. Agric. Food Chem.* **1996a**, *44*, 170−174.

Briandet, R.; Kemsley, E. K.; Wilson, R. H. Approaches to adulteration detection in instant coffees using infrared spectroscopy and chemometrics. *J. Sci. Food Agric.* **1996b**, *71*, 359−366.

Devaux, M-F.; Bertrand, D.; Robert, P.; Qannari, M. Application of multidimensional analyses to the extraction of discriminant spectral patterns from NIR spectra. *Appl. Spectrosc.* **1988**, *42* (6), 1015−1019.

Downey, G.; Boussion, J. Authentication of Coffee Bean Variety by Near-Infrared Reflectance Spectroscopy of Dried Extract. *J. Sci. Food Agric.* **1996**, *71*, 41−49.

Downey, G.; Spengler, B. Compositional analysis of coffee blends by near infrared spectroscopy. *Ir. J. Agric. Food Res.* **1996**, *35* (2), 179−188.

Downey, G.; Boussion, J.; Beauchêne, D. Authentication of whole and ground coffee beans by near infrared reflectance spectroscopy. *J. Near Infrared Spectrosc.* **1994**, *2*, 85−92.

Frega, N.; Bocci, F.; Lercker, G. High resolution gas chromatographic method for determination of Robusta coffee in commercial blends. *J. High Resolut. Chromatogr.* **1994**, *17*, 303−307.

Kemsley, E. K.; Ruault, S.; Wilson, R. H. Discrimination between coffea arabica and canephora variant robusta beans using infrared spectroscopy. *Food Chem.* **1995**, *54*, 321−326.

Martens, H.; Naes, T. *Multivariate Calibration*; Wiley: Chichester, U.K., 1989.

Reeves, J. B., III. Combining near and mid-infrared reflectance spectra: any spectral advantage? In *Near Infrared Spectroscopy: The Future Waves*; Proceedings of the 7th International Conference Spectroscopy, Montréal, Canada; Davies, A. M. C., Williams, P., Eds.; NIR Publications: Chichester, U.K., 1996; pp 14−19.

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.

Smith, A. W. Introduction. In *Coffee Chemistry*; Clarke, R. J., Macrae, R., Eds.; Elsevier Applied Science Publishers: New York, 1985.

Speer, K.; Tewis, R.; Montag, A. 16-*O*-Methylcafestol: a quality indicator for coffee. *14th Colloque, Association Scientifique International du Café (ASIC)*; ASIC: Paris, 1991.

Suchánek, M.; Filipová, H.; Volka, K.; Delgadillo, I.; Davies, A. N. Qualitative analysis of green coffee by infrared spectrometry. *Fresenius' J. Anal. Chem.* **1996**, *354*, 327−332.

Tressl, R.; Holzer, M.; Kamperschröder, H. Bildung von Aromstoffen in Rostkaffee in Abhangigkeit vom Gehalt an freien Aminosauren und reduzierenden Zuckern. In *Proceedings of the 10th Colloque, ASIC*; ASIC: Paris, 1982; pp 279−292.

Wilson, R. H. FT-IR for Food Analysis. *Trends Anal. Chem.* **1990**, *9* (4), 127−131.

Received for review April 24, 1997. Accepted August 18, 1997.[⊗] This research has been partly funded by grant aid under the Food Subprogramme of the Operational Programme for Industrial Development, which is administered by the Irish Department of Agriculture, Food and Forestry and supported by national and EU funds. We thank the concerted action FAIM (Food Authenticity: Issues and Methodologies) supported by the EU FAIR Programme.

JF970337T

[⊗] Abstract published in *Advance ACS Abstracts,* October 1, 1997.