# Discrimination of *Arabica* and *Robusta* in Instant Coffee by Fourier Transform Infrared Spectroscopy and Chemometrics

Romain Briandet, E. Katherine Kemsley,* and Reginald H. Wilson

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

Two species of coffee bean have acquired worldwide economic importance: these are, *Coffea Arabica* and *Coffea Canephora* variant *Robusta*. *Arabica* beans are valued most highly by the trade, as they are considered to have a finer flavor than *Robusta*. In this work, Fourier transform infrared spectroscopy is explored as a rapid alternative to wet chemical methods for authentication and quantification of coffee products. Principal component analysis (PCA) is applied to spectra of freeze-dried instant coffees, acquired by DRIFT (diffuse reflection infrared Fourier transform) and ATR (attenuated total reflection) sampling techniques, and reveals clustering according to coffee species. Linear discriminant analysis of the principal component scores yields 100% correct classifications for both training and test samples. The chemical origin of the discrimination is explored through interpretation of the PCA loadings. Partial least squares regression is applied to spectra of *Arabica* and *Robusta* blends to determine the relative content of each species. Internal cross-validation gives a correlation coefficient of 0.99 and a standard error of prediction of 1.20% (w/w), illustrating the potential of the method for industrial off-line quality control analysis.

**Keywords:** *Coffee; discrimination; infrared; spectroscopy*

## INTRODUCTION

The coffee bean is obtained from the small berry-like fruit of the coffee plant, family Rubiacae, genus *coffea*. Two species have acquired worldwide economic importance: these are, *Arabica* (~90% of world coffee production) and *Canephora* variant *Robusta* (9% of production). *Arabica* beans are valued the most highly because they have a more pronounced and finer flavor than *Robusta* beans. Consumers are able to purchase coffee in many different forms, such as roasted beans, ground-roasted coffee, and instant coffee, which may be freeze-dried or agglomerated. To convert raw green beans into instant coffee, a number of processing steps are required. The beans are roasted and ground, and then the soluble solids and volatile aroma compounds are extracted. The extraction step is a highly complex process, involving mainly the physical factors of heat, water movement, and dissolution (Mabbet, 1992). Because of the greater fraction of cellulose and the introduction of hydrolysis byproduct acids and aromatics, both types of instant coffee are very different in composition from ground-roasted coffee or the soluble product obtained from brewed coffee.

In view of the higher price commanded by *Arabica* beans, it is important that the species of the various coffee products can be identified and quantified. *Arabica* and *Robusta* beans are easy to distinguish by their size, but this visual criterion is eliminated by processing. Efforts have been made to discriminate between the two species by chemical analysis. One of the most interesting methods uses high performance liquid chromatography (HPLC) to detect diterpene 16-*O*-methylcafestol, which is present only in *Robusta* coffees (Speer et al., 1991). Another method successfully differentiates the two coffee species by applying the multivariate statistical method of principal component analysis (PCA; Jolliffe, 1986) to either chromatograms obtained by static headspace capillary gas chromatography (HCGC) or by HPLC-ultraviolet (UV) analysis of the roasted ground beans, or to a combination of the results from both techniques (Bicchi et al., 1993).

The main disadvantage of wet chemical analysis is that it is time-consuming. Recent research has shown that mid-infrared (IR) spectroscopy is a simple, fast, and reliable alternative to existing methods of identification of ground roasted coffees (Kemsley et al., 1995). In the present work, IR spectroscopy is explored as a method to discriminate and quantify *Arabica* and *Robusta* freeze-dried instant coffees. Spectra are collected with a Fourier transform IR (FTIR) spectrometer, equipped with two sampling stations: a diffuse reflection (DRIFT) accessory and an overhead attenuated total reflection (ATR) accessory. DRIFT and ATR are well-established FTIR sampling techniques (Wilson and Goodfellow, 1994). DRIFT is routinely used to analyze solids, especially powdered or particulate samples, and ATR is suitable for the analysis of liquids, pastes, and semisolids. Both techniques are relatively straightforward, generally requiring the minimum of sample preparation.

Two types of quantitative analysis were carried out: discriminant and compositional analysis. Discriminant analysis was performed to differentiate between pure *Arabica* and *Robusta* coffees; data processing comprised PCA for data "compression" and linear discriminant analysis (LDA; Massart et al., 1988) for data "classification". Compositional analysis was performed by applying partial least squares (PLS) regression (Martens and Naes, 1989) to spectra of *Arabica* and *Robusta* blends.

## EXPERIMENTAL PROCEDURES

**Instrumentation.** All spectra were collected with a Monit-IR (Spectra-Tech, Applied Systems Inc.) FTIR spectrometer, operating in the region 800−4000 $cm^{-1}$, equipped with a

* Author to whom correspondence should be addressed.

0021-8561/96/1444-0170$12.00/0 

sealed, desiccated interferometer compartment and a deuterated triglycine sulphate detector. The Monit-IR instrument differs from the majority of FTIR spectrometers in that it incorporates two integral sampling stations, rather than providing for a demountable sampling accessory. One of the stations is designed for the DRIFT sampling technique, the other for overhead ATR. Both stations comprise permanently mounted, optimized transfer optics and sealed potassium bromide (KBr) windows to minimize the path of the IR beam through the external atmosphere. Consequently, spectral quality is generally very high, and there is very little spectral contamination due to carbon dioxide and water vapor.

All spectral measurements were made at nominal 8 $cm^{-1}$ resolution, with 64 interferograms co-added before Fourier transformation. To obtain DRIFT spectra, each sample was ground for 5 min with a pestle and mortar before loading into the DRIFT sample cup. Taking care not to compress the sample, the surface was levelled in a predetermined direction with a spatula before analysis. To obtain ATR spectra, each sample (pure species and blends) was dissolved in deionized water at 50 °C. The total concentration of instant coffee in each solution was 0.3 g/mL; 10 mL of solution were prepared from each sample. The ATR station was fitted with a trough-mounted zinc selenide crystal, with a 45° interface angle and nominally 11 internal reflection sites. Spectral acquisition was initiated immediately upon application of the solution to the crystal. By adopting this protocol, we ensured that settling of the small insoluble fraction onto the crystal surface produced a consistent, minor effect on all spectra.

**Discriminant Analysis.** *Samples.* Fifty-two pure samples of freeze-dried coffee were obtained at the start of this work (29 *Arabica* and 23 *Robusta*), and these were used for the DRIFT analysis. Two additional *Robusta* samples were made available for the ATR analysis. The suppliers were able to guarantee the authenticity of the samples.

*Data Acquisition.* Single-beam DRIFT spectra of each sample were collected, transformed to Kubelka-Munk units with a background spectrum of ground KBr, and truncated to 286 data points in the region 800−1900 $cm^{-1}$. To reduce the effect of irreproducible sample cup loading, a single-point baseline correction at 1900 $cm^{-1}$ was performed, followed by area normalization on the region 800−1900 $cm^{-1}$ (that is, setting the integrated area in this region equal to unity for all spectra; see Kemsley et al., 1995). Single-beam ATR spectra were also collected of the solutions of each sample. The spectra were transformed to absorbance units with an air background spectrum, and truncated to 311 data points in the region 800−2000 $cm^{-1}$. A single-point baseline correction at 2000 $cm^{-1}$ was performed, followed by area-normalization on the region 800−2000 $cm^{-1}$. Data processing was carried out with "Win-Discrim" (E. K. Kemsley, Institute of Food Research, Norwich, U.K.).

**Compositional Analysis.** *Samples.* One each of the *Arabica* and *Robusta* samples was selected at random, and 18 blends of these samples were prepared. The *Robusta* content varied in the range 0−60% (w/w).

*Data Acquisition.* Single-beam ATR spectra of the solutions of each blend were collected. The spectra were transformed to absorbance units with an air background spectrum, and truncated to 251 points in the region 930−1900 $cm^{-1}$. A single-point baseline correction at 1900 $cm^{-1}$ was performed, followed by area normalization on the region 930−1900 $cm^{-1}$. Data processing was carried out with "Unscrambler II" (Camo, Norway).

## RESULTS AND DISCUSSION

With regard to its chemical composition, coffee is undoubtedly one of the most complex of commonly encountered food commodities. Green coffee beans contain a wide range of different chemical compounds, which react and interact at all stages of coffee processing to produce a final product with an even greater diversity and complexity of structure (Clifford, 1985). The most important compounds in freeze-dried coffee are carbohydrates, minerals, caffeine, chlorogenic acid, proteins, and lipids. Most of these constituents absorb strongly in the "fingerprint" spectral region (900−1500 $cm^{-1}$), and the absorption bands frequently overlap each other. Consequently, a full assignment of the spectral features is difficult and beyond the scope of the present work.

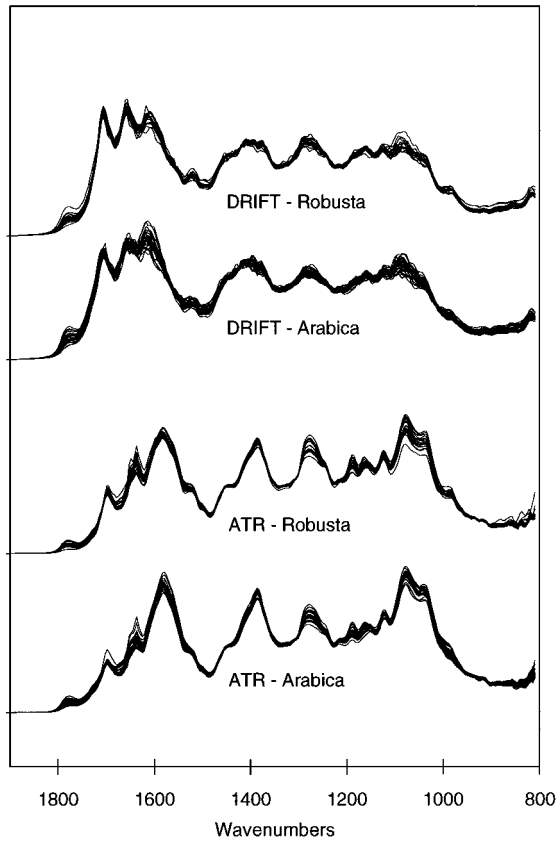


**Figure 1.** Spectra of *Arabica* and *Robusta* freeze-dried coffee samples, acquired by DRIFT and ATR sampling techniques (shown after baseline correction and area normalization; water spectrum subtracted from ATR spectra to aid clarity).

An expansion of the 800−1900-$cm^{-1}$ region of the pure *Arabica* and *Robusta* coffee spectra acquired by DRIFT and ATR is shown in Figure 1. The spectra have been baseline corrected and area normalized. In addition, a pure water spectrum has been subtracted from the ATR spectra shown here for clarity and to aid spectral interpretation, although water-subtracted data were not used in the subsequent data processing. Examination of the baseline in the region 1800−1900 $cm^{-1}$ shows that detector noise is negligible. Although *Arabica* and *Robusta* species exhibit physical differences and are known to differ in chemical composition, on inspection, the spectra of both species appear quite similar. Some small differences can be identified, particularly in the regions 1600−1800 $cm^{-1}$ and 1000−1300 $cm^{-1}$. However, these differences are inconsistent, and the spectra are so complex that multivariate statistics are needed to obtain a definitive differentiation between the two species.

The spectra from the DRIFT and ATR sampling techniques were divided into training and test sets, as shown in Table 1. PCA was applied to the two training sets. The percentage and cumulative percentage variances of the first five principal components (PCs) obtained are shown in Table 2. The test set spectra were rotated into the same PC spaces, and plots of the scores against one another were examined. The second and third PC scores for each sampling technique, which

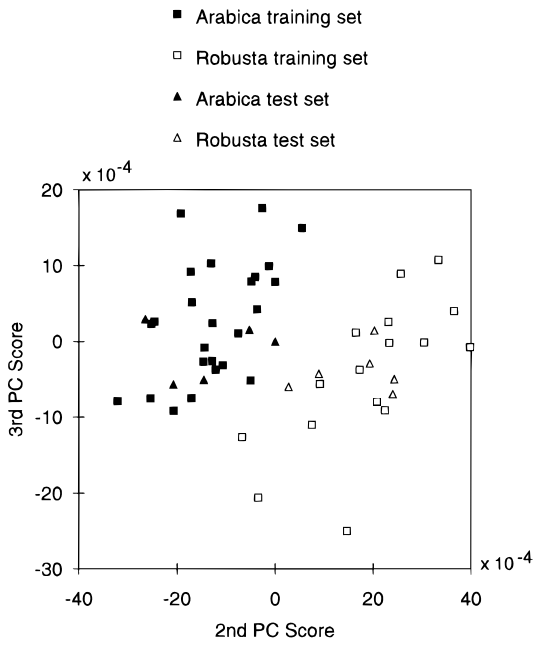


**Figure 2.** Plot of second versus third PC scores of the DRIFT training and test sets.

**Table 1. Numbers and Types of Samples in Training and Test Sets for LDA**

| | DRIFT | | ATR | |
|---|---|---|---|---|
| bean | training | test | training | test |
| *Arabica* | 25 | 5 | 25 | 4 |
| *Robusta* | 16 | 6 | 19 | 6 |

**Table 2. Percentage and Cumulative Percentage Variances for DRIFT and ATR Principal Component Analyses**

| | DRIFT | | ATR | |
|---|---|---|---|---|
| PC score | % variance | cumulative % variance | % variance | cumulative % variance |
| 1 | 40.0 | 40.0 | 95.2 | 95.2 |
| 2 | 32.3 | 72.3 | 2.8 | 98.0 |
| 3 | 7.6 | 79.9 | 1.1 | 99.1 |
| 4 | 6.2 | 86.1 | 0.6 | 99.7 |
| 5 | 3.2 | 89.3 | 0.1 | 99.8 |

revealed the most clustering with respect to coffee species, are illustrated in Figures 2 and 3, respectively. LDA was carried out for the training sets with the first three PC scores and the Mahalanobis $D^2$ metric, and the discriminant rules obtained were applied to the test sets. For both sampling methods, 100% of both training and test sets were correctly classified.

PCA loadings represent independent sources of spectral variability present in the data set. Examination of the loadings can sometimes yield insight into the chemical variability present in the samples. The first three loadings of both sampling techniques are compared in Figure 4. From inspection of the PC scores, it was found that the first loadings do not seem to play any role in the species discrimination. For the DRIFT analysis, the first loading resembles an inverted, average spectrum. This loading is interpreted as variability in the data set caused by irreproducible sampling; the area normalization pretreatment mitigates this effect to an extent, but cannot eliminate it entirely. In addition, there is a relatively large feature at 1600–1650 $cm^{-1}$; this feature may reflect differences in the residual water content of the freeze-dried samples. For the ATR analysis, the first loading also exhibits a large

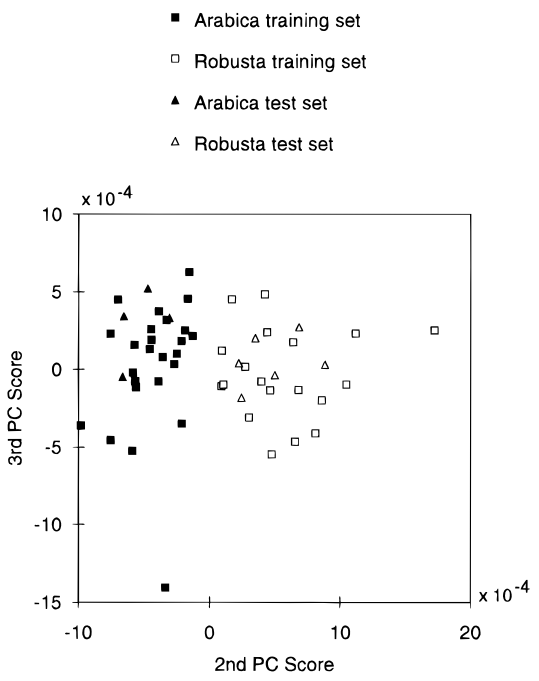


**Figure 3.** Plot of second versus third PC scores of the ATR training and test sets.

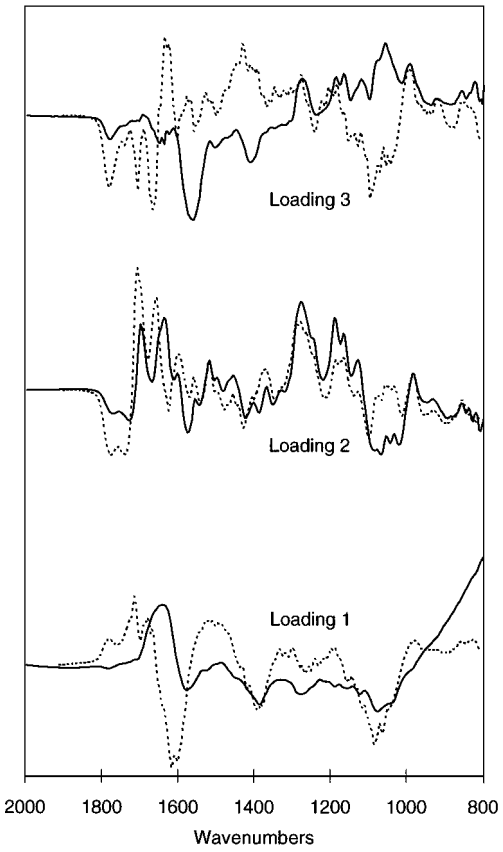


**Figure 4.** First three PC loadings of the DRIFT (- - -) and ATR (—) training sets.

positive feature in this region, along with a number of negative features across the remainder of the spectral range, which we believe reflects variability in the molar concentration of the coffee solutions. This variability may arise from differences in the initial water content of the samples; some variability may also be introduced by the settling of the small insoluble fraction (mainly starch) on the crystal surface. Because the IR spectrum of water is so strong, any variations in the water content

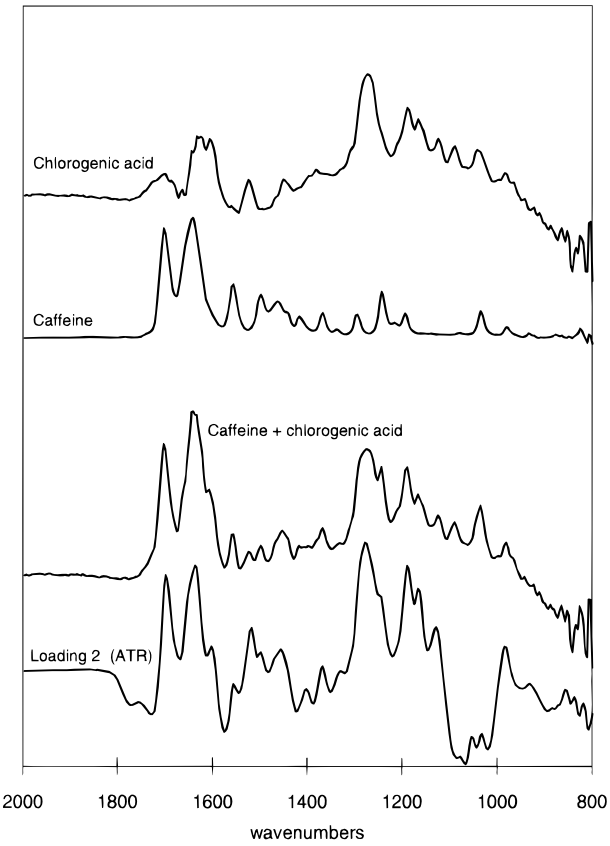


**Figure 5.** Comparison between second PC loading and the ATR spectra of pure caffeine, pure chlorogenic acid, and a mix of caffeine and chlorogenic acid.

tend to dominate, as evidenced by the large percentage variance associated with the first PC score (see Table 2).

The second PC score has the most important role in distinguishing the two species (see Figures 2 and 3). The second loading of both PCAs are quite similar across most of the region examined (see Figure 4). This similarity indicates that in both sampling techniques the same chemical differences are detected and used for the species distinction. Moreover, the second loadings are highly similar to a spectrum of a mix of pure caffeine and chlorogenic acid. The ATR spectra of caffeine, chlorogenic acid, and a mixture of the two are compared with the second PC loading obtained from the ATR data in Figure 5. Most of the large features can be attributed to one or the other of the compounds, although the relative intensities of the bands differ somewhat. Caffeine is responsible for the two large bands in the region 1550−1750 $cm^{-1}$, whereas the chlorogenic acid has major bands in the region 1150−1300 $cm^{-1}$. 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). The relative contents of these compounds in the different species is not well known in coffee products. However, *Robusta* coffee has higher caffeine and chlorogenic acid contents than *Arabica* in roasted coffee beans; the lower chlorogenic acid content of *Arabica* beans is generally associated with their superior beverage quality. These known compositional differences are consistent with our finding that the *Robusta* coffees tend to have higher scores with respect to the second loading.

The third loading contains some information that helps to discriminate between species. In both DRIFT and ATR loadings, a band occurs around 1760 $cm^{-1}$. This band may represent a differentiation on lipid

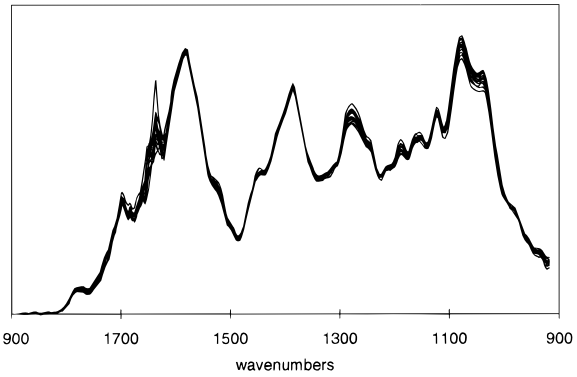


**Figure 6.** Spectra of *Arabica* and *Robusta* blends (shown after baseline correction and area normalization; water spectrum subtracted from ATR spectra to aid clarity).

content, an idea that is supported by the presence of additional features in the 1000−1400 $cm^{-1}$ region. In the DRIFT loading, the caffeine bands are again present at 1550−1750 $cm^{-1}$. In the ATR loading, the chlorogenic acid features are also present in the region 1150−1300 $cm^{-1}$. The remaining, unexplained regions of the loading may be due to carbohydrates, lipids, and probably several other constituents (for example, humic acid that accounts for 10% dwb of the freeze-dried coffee composition). A complete explanation of the loadings represents a very complex problem. However, we conclude that discrimination of the freeze-dried coffee species is based largely on the different caffeine and chlorogenic acid contents of *Arabica* and *Robusta* species.

The fingerprint region of the complete series of blend spectra is shown in Figure 6. The spectra have been baseline corrected and area-normalized, and a water absorbance spectrum was subtracted from the ATR spectra for clarity. The greatest variability in this data set occurs in the region 1600−1650 $cm^{-1}$, but this variability is likely to arise from minor differences in the water content rather than to the varying blend compositions. However, the variations in the regions 1150−1300 $cm^{-1}$ and, to a lesser extent, 1550−1750 $cm^{-1}$ support our conclusion that the chlorogenic acid and caffeine contents differ most markedly between species.

PLS regression was carried out for the *Robusta* content of the blends. Using internal cross-validation (Martens and Naes, 1989), the optimum regression model was obtained with three PLS scores. The cross-validation prediction values are plotted against the actual *Robusta* content in Figure 7. The standard error of prediction (SEP) and correlation coefficient are 1.20% (w/w) and 0.99, respectively.

## CONCLUSION

The results presented show that FTIR spectroscopy can be used to identify and quantify the *Arabica* and *Robusta* contents of freeze-dried instant coffees. DRIFT and ATR sampling accessories gave the same excellent results in discriminant analysis, with 100% of spectra in both training and independent test sets correctly classified. We believe that the discrimination is based on the different chlorogenic acid and caffeine contents of *Arabica* and *Robusta* species. The ATR sampling method is generally believed to be more accurate than DRIFT for compositional analysis, and was the technique of choice for the analysis of the coffee blends. The precision of the results obtained shows that ATR is

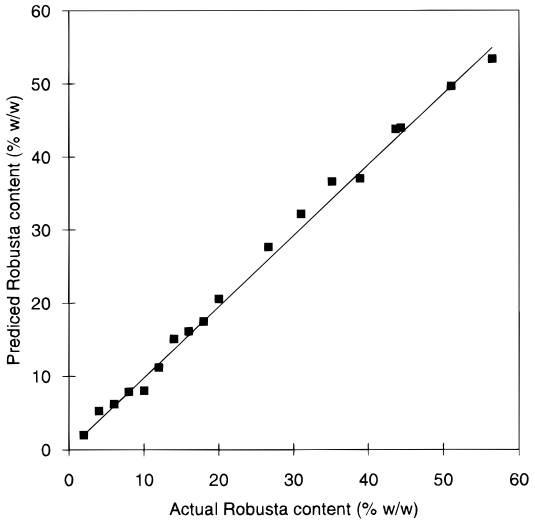


**Figure 7.** PLS model-predicted versus actual *Robusta* content in *Arabica*−*Robusta* blends.

suitable for off-line quality control of freeze-dried coffees. We anticipate that comparable results would be obtained with agglomerated instant coffees. Future work will explore whether it is possible to detect undeclared material in instant coffees.

ACKNOWLEDGMENT

We thank Mr. G. Downey, of Teagasc, National Food Center, Dublin, Ireland, for supplying the samples used in this work.

LITERATURE CITED

Bicchi, C. P.; Binello, A. E.; Legovich, M. M.; Pellegrino, G. M.; Vanni, A. C. Characterization of roasted coffee by S-HSGC and HPLC-UV and principal component analysis. *J. Agric. Food Chem.* **1993**, *41 (12)*, 2324−2328.

Clifford M. N. Chlorogenic acid. In *Coffee Chemistry*; Clarke, R. J., Macrae, R., Eds.; Elsevier Applied Science: New York, 1985.

Fuller, M. P.; Griffiths, P. R. Diffuse reflectance measurments by infrared Fourier transform spectroscopy. *Anal. Chem.* **1978**, *50 (13)*, 1906−1910.

Jolliffe, I. T. *Principal Component Analysis*; Springer-Verlag: New York, 1986.

Kemsley, E. K.; Ruault, S.; Wilson, R. H. Discrimination between coffea arabica and canephora variant robusta beans using infrared spectroscopy. *Food Chem.* **1995**, *54 (3)*, 321−326.

Mabbet, T. The secret of instant. *Coffee Cocoa Int.* **1992**, *6*, 22−23.

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

Massart, D. L.; Vandeginste, B. G. M.; Deming, S. N.; Michotte, Y.; Kaufman, L. *Data Handling in Science and Technology, Volume 2: Chemometrics, a Textbook*; Vandeginste, B. G. M., Kaufman, L., Eds.; Elsevier Science: Amsterdam, The Netherlands, 1988.

Silvetz, M. Freeze-dried coffee processing: a very basic explanation of this complicated technique. *Tea Coffee Trade J.* **1974**, *146 (1)*, 74−75.

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

Speer, K.; Tewis, R.; Montag, A. 16-O-Methylcafestol: a quality indicator for coffee; ASSIC, 14e Colloque, San Francisco, 1991.

Stoltze, A.; Masters, K. Recent developments in the manufature of instant coffee and coffee substitutes. *Food Chem.* **1979**, *4*, 31−39.

Wilson, R. H.; Goodfellow, B. G. Mid-infrared Spectroscopy. In *Spectroscopic Techniques for Food Analysis*; Wilson, R. H., Ed.; VCH: New York, 1994.

Received for review May 22, 1995. Revised manuscript received September 21, 1995. Accepted October 5, 1995.[⊗] The BBSRC and COMETT are acknowledged for financial support.

JF950305A

[⊗] Abstract published in *Advance ACS Abstracts,* December 1, 1995.