

# CHEMOMETRIC STUDIES FOR QUALITY CONTROL OF PROCESSED BRAZILIAN COFFEES USING DRIFTS

JULIANO S. RIBEIRO<sup>1,2</sup>, TEREZINHA J. SALVA<sup>2</sup> and  
MÁRCIA M.C. FERREIRA<sup>1,3</sup>

<sup>1</sup>*Theoretical and Applied Chemometrics Laboratory  
Chemistry Institute  
University of Campinas  
P.O. Box 6154, 13083-970 Campinas, SP, Brazil*

<sup>2</sup>*Centro de Análise e Pesquisa Tecnológica do Agronegócio do Café  
“Alcides Carvalho”  
Instituto Agronômico de Campinas  
Campinas, SP, Brazil*

Received for Publication March 15, 2008  
Accepted for Publication December 16, 2008

## ABSTRACT

*In this work, the potential of mid-infrared diffuse reflectance spectroscopy with Fourier transform for discrimination of 29 commercial Brazilian coffee samples with different industrial processing, i.e., caffeine extraction and roasting degree, was evaluated. The statistical treatments applied to pre-treated spectral data were principal component analysis and partial least squares – discriminant analysis (PLS-DA). The ordered predictors selection method was used for variable selection. The chemometric analyses of the mid-infra-red spectra allowed inferring on the lower carbohydrate, caffeine and chlorogenic acid concentration as well as on the higher water content in the decaffeinated coffee. The technique also allowed speculation on the higher lipid and lower water content in the dark roasted coffee compared with traditional roasted coffee. A clear discrimination of decaffeinated from medium and dark roasted coffees was observed in PC1. PLS-DA was used for the discrimination between medium and dark roasted coffees. A model with one latent variable correctly classified 100% of the external validation and prediction samples according to their roasting degree.*

<sup>3</sup> Corresponding author. TEL: +55-19-3521-3102; FAX: +55-19-3521-3023; EMAIL: marcia@iqm.unicamp.br

## PRACTICAL APPLICATIONS

Diffuse reflectance mid infrared spectroscopy (DRIFTS), principal component analysis and partial least squares were successfully applied to discriminate decaffeinated coffees from nondecaffeinated coffees and to discriminate roasted coffees by their roasting degree. This study have shown that DRIFTS coupled with chemometrics consists in a simple and straightforward analytical method for monitoring the roasted coffee authenticity, and the results could help in developing an alternative and inexpensive method for quality control of coffee products.

## INTRODUCTION

The annual world coffee market is worth about 60 billion dollars (Brando 2007). Production and trade systems, as well as processing conditions and chemical composition of the beans, contribute to the final price of the coffee. Traditionally, strategies used to certify chemical characteristics of coffee have relied on wet chemistry, but these methods are destructive, time-consuming and sometimes expensive. As demand for rapid and cheaper controls is growing, wet chemistry methodologies are being replaced by dry methods (Rubayiza and Meurens 2005).

Recent research has shown that near-infrared (NIR) and mid-infrared (MIR) spectroscopy are useful in the food industry (Chen *et al.* 2006; Small 2006; Cen and He 2007; Esteban-Díez *et al.* 2007). Such methodologies, when coupled with chemometric data analysis techniques, may be regarded as rapid and reliable means of process control, as well as for product quality and safety certifications (Pedro and Ferreira 2005; Li *et al.* 2006; Xie *et al.* 2007). Contrary to the great number of studies carried out by NIR diffuse reflectance spectroscopy to characterize food products (Downey *et al.* 2003; Andrés and Bona 2005; Yan-de *et al.* 2007) and coffee varieties (Kemsley *et al.* 1995; Briandet *et al.* 1996; Esteban-Díez *et al.* 2004; Rubayiza and Meurens 2005; Pizarro *et al.* 2007), similar studies dealing with MIR diffuse reflectance spectroscopy are scarce (Reeves and Zapt 1998; Wilson and Tapp 1999; Bertelli *et al.* 2007).

MIR has potentially some advantages over NIR for the analysis of food, and recent interest has increased its use in areas where NIR spectroscopy has been successfully used (Fagan *et al.* 2007; He *et al.* 2007). Spectral features in the NIR region are due to molecular absorptions of overtones and combinations of fundamental vibrational bands from the MIR region, limiting its use for classification purposes. Band assignments are not easy to interpret because one single band can result from a combination of severely overlapped

vibrations. The weak and broad bands are the other shortcomings of this method. On the other hand, absorbances in MIR region are richer of qualitative (molecular structures) information that can be easier distinguished and interpreted (Olinger and Griffiths 1993). Therefore, Fourier-transform MIR spectroscopy is an appealing technique to be used for classification purposes when qualitative characterization and interpretation is important (Bertelli *et al.* 2007).

The majority of MIR applications for food quality in the literature comprise analysis of liquid foods, such as wines, fruit juices and beers (Cozzolino *et al.* 2003; Reid *et al.* 2005; Inón *et al.* 2006). More recently (Ferrão *et al.* 2003; Karoui *et al.* 2006; Fagan *et al.* 2007), solid samples like cheese and coffee have also been analyzed by diffuse reflectance infrared Fourier transformation spectroscopy (DRIFTS).

DRIFTS can be used to detect compositional differences between food samples on the basis of vibrations of several chemical groups at specific wavelengths in the MIR region of the spectrum. The results from DRIFTS can provide information regarding the chemical fine structure of a food sample, allowing screening of food products based on qualitative attributes (Reid *et al.* 2005).

The color and composition of coffee beans are changed during the roasting process. A combination of temperature and time determines the intensity of heat applied to the product and the degree of modification in the grains. The higher the intensity of the roasting process the higher the chemical degradation. Release of large amounts of carbon dioxide and formation of several hundreds of substances associated with the taste and aroma of the coffee occur during the roasting. A heavy roasting process implies dark brown bitter beans and the lack of typical coffee aroma, whereas a light roasting process may be insufficient for the completion of all pyrolytic reactions, resulting in light brown coffee with underdeveloped organoleptic characteristics (Buffo and Cardelli-Freire 2004).

When studying the roasting of arabica coffee, Toci *et al.* (2006) found that the process degraded a considerable amount of sacarose (50% in light roast) and clorogenic acids. About proteins, the reduction was around 10% in light degrees of roasting, increasing to 20% in dark roasted beans. The concentration of trigoneline was also reduced with the degree of roasting (Toci *et al.* 2006).

Besides roasting, decaffeination is another process commonly applied to green coffee beans. The influence of the decaffeination process on the bean composition and on the final product quality depends on the extraction method. Although coffee decaffeination processes may be done using supercritical fluid (CO<sub>2</sub>), most of the decaffeination methods employ organic solvents like dichloromethane and ethyl acetate (Clarke and Vitzthum 2001).

Decaffeination with dichloromethane causes modifications in chemical composition of the beans. Reduction of the total glycides, represented mainly by a considerable loss of sucrose (20 and 60% for *C. canephora* e and *C. arabica*, respectively), has been observed. It has been shown that the process may also induce the extraction of 16 % of the clorogenic acids in *C. arabica* and 11 % in *C. canephora* (Toci *et al.* 2006).

In the present work, the feasibility of mid-infrared diffuse reflectance spectroscopy coupled with chemometric analysis to differentiate commercial coffees was investigated. Principal component analysis (PCA) (Wold *et al.* 1987) and partial least squares-discriminant analysis (PLS-DA) (Barker and Rayens 2003) were applied to discriminate decaffeinated coffees and to classify roasted coffees by the roasting degree.

## MATERIALS AND METHODS

### Coffee Samples and Data Acquisition

Twenty-nine samples of commercial Brazilian coffees were purchased from local stores. Twenty-one of them were used to build the calibration model (calibration set). From these, seven were decaffeinated and light roasted (DC), seven were traditional or medium roasted (TR) and seven were dark roasted (DR). All samples were from the same brand from different production batches. The eight remaining samples, five medium roasted (TR) and three dark roasted, were used for prediction (prediction set).

### DRIFT Spectra Acquisition

Each sample was ground for 1 min with mortar and pestle before loading into the static sample cup with five cavities (0.5 cm of surface diameter and 0.20 cm<sup>3</sup> of internal volume). Sample compression inside the cup was avoided, and the surface was leveled with a spatula. All the spectra were collected using a Nicolet 520 FT-IR Spectrometer equipped with a diffuse reflectance accessory model Jasco DR81, using 256 scans in the 4,000–400/cm range, with 4/cm of nominal resolution and 64 interferograms co-added before Fourier Transformation. KBr was used for the background spectrum. Three different portion of a coffee sample were placed into each cavity of the instrumental cup and the spectrum recorded.

### Data Analyses

The original spectral profiles were organized into a matrix format  $\mathbf{X}$  ( $I \times J$ ), where each replicate was considered as one sample. Data analysis was

carried out in Matlab 6.5 software (The MathWorks, Co., Natick, MA) using the computational package PLS\_Toolbox (Eigenvector Research, Inc. – PLS\_Toolbox version 3.02.) (Wise *et al.* 2004).

PCA was the method used for exploratory data analysis (Wold *et al.* 1987). In this method, a  $J$ -dimensional matrix  $\mathbf{X}$  is projected into a lower,  $K$ -dimensional space through a decomposition into scores  $\mathbf{S}$  ( $I \times K$ ) and loadings  $\mathbf{L}$  ( $J \times K$ ) matrices whose product models the systematic variation in the data, and into a residual matrix  $\mathbf{R}$  ( $I \times J$ ), which, in the ideal case, contains only the measurement errors. The plot of two columns from the scores matrix  $\mathbf{S}$  against each other gives a two-dimensional projection of original data set, whereas the plot from columns of the loadings matrix  $\mathbf{L}$  displays the correlation among the variables (Beebe *et al.* 1997).

Whilst PLS is used as a calibration method, PLS-DA is a classification method wherein a model is built between the matrix  $\mathbf{X}$  and the matrix of known classes  $\mathbf{Y}$  (Barker and Rayens 2003). In PLS-DA each class is described by a column in  $\mathbf{Y}$ . A value 1 or 0 is assigned to each class variable, depending on the class to which an object belongs. The model, of course, will not predict either 1 or 0 exactly, so a threshold (predicted probability) is determined, above which a sample is considered to be in the class and below which a sample does not belong to that class. Basically, a normal distribution is fitted to the predicted values from the PLS-DA calibration model and then used to calculate the probability (based on Bayes' Theorem) of a given prediction sample value (Xing *et al.* 2007). Using the estimated distribution for each class, thresholds are calculated and the prediction samples that are located between the range of the estimated distributions, are the samples predicted as false positives and false negatives.

The performance of the final PLS-DA model was evaluated in terms of the root mean square error of cross validation (RMSECV) and by the correlation coefficient of cross-validation ( $r_{cv}$ ). A leave-one-sample-out cross-validation was performed. In this procedure one spectrum was excluded at a time, the model was built and the estimated class of the excluded sample was used to calculate RMSECV. Once the model had been internally and externally validated, it was tested for the prediction of eight new samples.

Two pre-treatments were applied to the original data matrix in this study. First, a Savitzky-Golay smoothing (Savitzky and Golay 1964), with a window size of 10 points, was used to reduce random variations (experimental noise) and then, the systematic variations were reduced by a multiplicative signal correction (MSC) (Isaksson and Naes 1988).

A visual inspection of the spectra indicated that the decaffeinated samples could be discriminated from the others and so, a visual selection was carried out prior to the exploratory analysis using PCA. The ordered predictors selection (OPS) method (Teófilo *et al.* 2009) was applied to find the best spectral

regions for construction of the PLS-DA model. In this method, an intuitive vector is formed by a combination of vectors, such as regression and correlation vectors, among others. Based on the absolute values of the informative vector elements, the importance of each response (instrumental) variable to the model is defined. In the next step, the response variables are reordered into decreasing order according to their importance. Finally, the ordered variables are tested using increments over a previously defined window. The RMSECV and correlation coefficient of cross-validation ( $r_{cv}$ ) values are stored for each analyzed window. The best set of variables is indicated by the lowest RMSECV and highest  $r_{cv}$ .

## RESULTS AND DISCUSSION

### PCA – Discrimination of Decaffeinated Coffees

The original spectra from calibration set were organized into a matrix format ( $63 \times 1,856$ ) and pre-treated as described in Materials and Methods section. Even after pre-treatment, the spectra still contained regions with high signal variations of instrumental origin (2,380–2,280 and 500–400/cm), thus these regions were excluded from data analysis, resulting in a data matrix ( $63 \times 1,772$ ). PCA was applied to this data matrix and in PC2 two well distinguished clusters were revealed (Figure not shown), one of them consisting only of DC samples while TR and DR were completely mixed into the other.

Figure 1A shows an example of two overlapped spectra; one resulted from the mean of traditional and dark roasted samples and other from the mean of decaffeinated coffee. It was visible that several regions could distinguish DC from the others. In order to get more insight into which spectral regions could be responsible for this discrimination, a visual variable selection was applied to the pretreated spectra. From 1,772 pre-treated variables, 269 variables were selected and organized in 9 regions, as indicated with numbers in Fig. 1A. It can be seen in this figure that the regions selected coincide with those of largest spectral differences. Figure 1B shows the expanded region indicated in Fig. 1A for the purpose of better visualization.

PCA was applied to the nine regions selected. From the scores plot in Fig. 2A, a distinct visual clustering appeared between decaffeinated and the other two classes when the data were displayed with respect to the first two principal components. Decaffeinated coffees were located on the right side of PC1, which describes 66.57% of the original information, well separated from medium and dark roasted samples on the right side of PC1.

The loadings plot of PC1 (Fig. 2B) indicates that regions 1, 16 and 17 have positive values and significantly contribute to the discrimination of

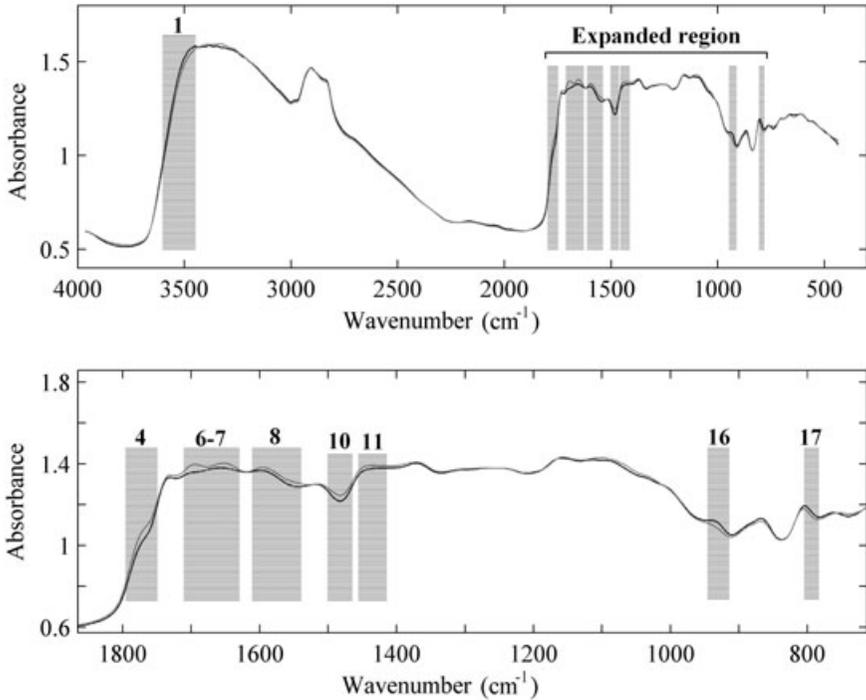


FIG. 1. MEAN MID-INFRARED SPECTRA OF DECAFFEINATED (—), AND TRADITIONAL AND DARK ROASTED (—) COFFEE SAMPLES

The regions visually selected for discrimination of both types of coffee are indicated in (A) and the expanded region for better visualization is indicated in (B).

decaffeinated coffee samples from the others. On the other hand, medium and dark roasted coffee samples have negative scores (Fig. 2A) because they are characterized by higher absorbance in the other six regions which have negative loadings values, as shown in Fig. 2B.

According to Kemsley *et al.* (1995), carbohydrates are responsible for the absorbance in the regions 10 and 11, named as “finger print” regions. The negative loadings of these two regions (Fig. 2B) indicated that carbohydrate concentrations in decaffeinated samples, although less roasted, are lower than in the non-decaffeinated coffees, confirming previous findings that the caffeine extraction process also reduces the sugar concentration in the beans (Toci *et al.* 2006).

The spectral region 1 from 3,600 to 3,450/cm is well known to represent O-H bond stretching (Silverstein and Webster 1998) and could characterize water and carbohydrates. Since the experiments indicated that the carbohy-

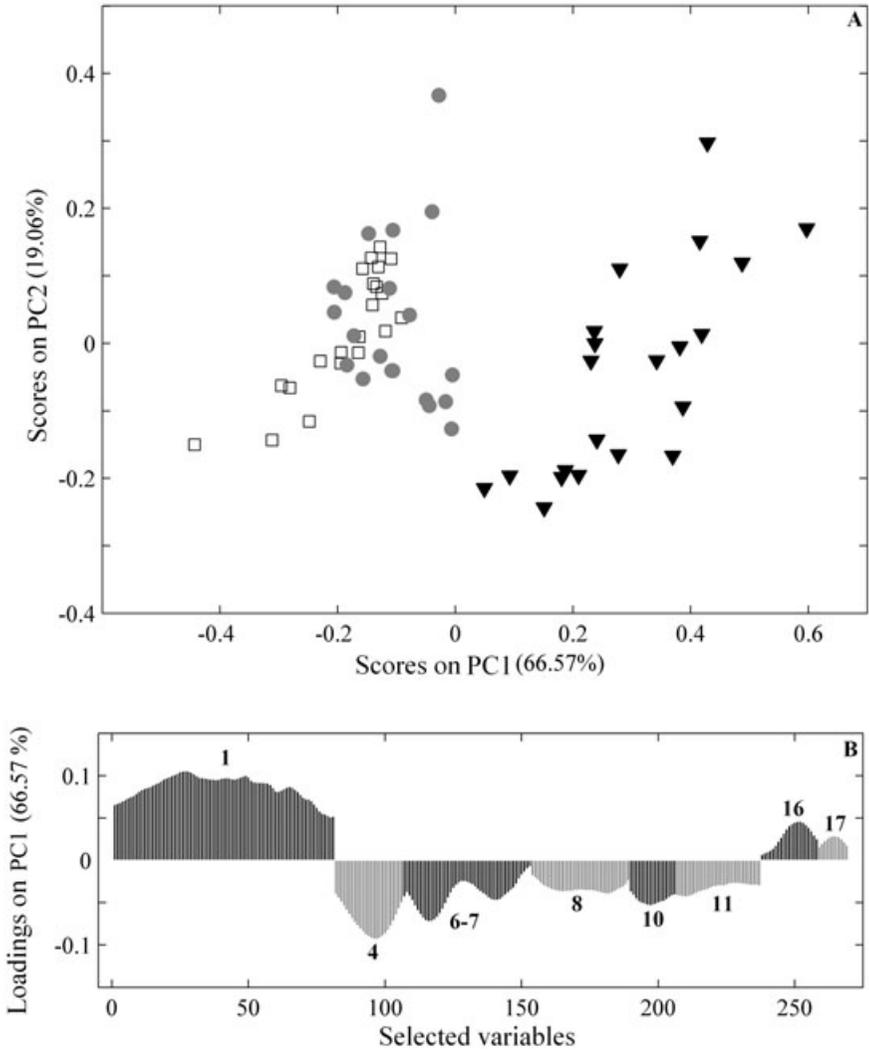


FIG. 2. PC1 × PC2 SCORES PLOT (A)

(▼) Decaffeinated samples, (□) dark roasted and (●) medium roasted. PC1 loadings plot indicating the nine regions visually selected for PCA analysis (B).

drate concentration was reduced by the overall process (decaffeination and light roasting) the stretching of region 1 (Fig. 1) should be explained mainly by the higher water content. These results indicated that decaffeinated and light roasted samples had more intrinsic water than traditional and dark roasted coffees.

The results shown in loadings of PC1 (Fig. 2B) indicate lower absorbance in regions 4 (1,795–1,750/cm), 6 (1,710–1,680/cm) and 7 (1,665–1,630/cm) for decaffeinated samples. These regions exhibit characteristic bands arising from carbonyl C=O bond stretching (Silverstein and Webster 1998). Proteins, free amino acids, lipids, chlorogenic acids and alkaloids (caffeine and trigonelline) could be responsible for the absorbances in these regions. According to Fabian *et al.* (1992), wavenumbers ranging from 1,600 to 1,700/cm are highly related to chlorogenic acids and caffeine concentration in coffees. So, the smaller intensities in regions 4, 6 and 7 demonstrate that decaffeinated samples present lower concentrations of caffeine and chlorogenic acids, in agreement with the findings of Toci *et al.* (2006) that the decaffeination process, besides caffeine, also extracts chlorogenic acids.

Table 1 indicates the wavenumbers of stretching and deformation vibrations with the respective vibrational modes that occur in the 9 regions visually selected to discriminate the coffee samples and also shows other regions that will be relevant for the next discriminant study.

TABLE 1.  
CHARACTERISTIC INFRARED ABSORPTION FREQUENCIES AND ATTRIBUTIONS IN  
REGIONS SELECTED FOR PC

Visual* and OPS† regions selected	Wavenumber range‡ (cm <sup>-1</sup> )	Vibrational modes
1*†	3,600–3,450	O-H str.
2†	3,370–3,350	N-H str., NH <sub>3</sub> <sup>+</sup> str., NH <sub>2</sub> str.
3†	3,000–2,820	C-H str.
4*†	1,795–1,750	C=O str.
5†	1,750–1,720	C=O str.
6*†	1,710–1,680	C=O str.
7*†	1,665–1,630	C=C str., C=N str., C=O str.
8*	1,610–1,540	NO <sub>2</sub> str., NH <sub>3</sub> <sup>+</sup> def., N=N str., C=O str.
9†	1,570–1,500	NH <sub>3</sub> <sup>+</sup> def., N=O str., C=C str., N-H def.
10*	1,500–1,465	C=C str., N=O str., NH def.
11*†	1,455–1,415	C-H def., S=O str.
12†	1,400–1,380	O-H str., C-H def., CH <sub>3</sub> def.
13†	1,325–1,295	S=O str., C-H def., NO <sub>2</sub> str.
14†	1,275–1,260	O-NO <sub>2</sub> vib., ROR str.
15†	1,240–1,170	S=O str., C-N vib.
16*	945–915	C-H def.
17*	804–784	C-H def.

\* Variables selected visually.

† Variables selected by OPS.

‡ Values of wavenumber extracted from Parikh 1974; Silverstein and Webster 1998.  
def., deformation; str., stretching; vib, vibration.

## Partial Least Squares-Discriminant Analysis

As the TR and DR coffee samples were mixed in the scores plot of previous analysis (Fig. 2A) and well distinguished from DC, partial least squares – discriminant analysis (PLS-DA) was applied to discriminate/classify the samples according to their roasting degree. Thus, the decaffeinated coffees were excluded from the original data matrix. The calibration set (14 samples = 42 replicates) was randomly split into the training set consisting of 10 samples (30 replicates) and external validation set of four samples (twelve replicate spectra). For prediction, three dark and five medium roasted samples were used (24 replicates).

To construct this model, it was important to identify those regions that could be chemically related to the roasting degree. The important variables of the pre-treated spectral profiles that could be involved in the discrimination of types of coffee samples were selected by the OPS method (Teófilo *et al.* 2009) and the original variables were reduced from 1,772 to 208. Figure 3A shows an example of two overlapped spectra, one from traditional and one from dark roasted coffee. Figure 3B,C show the expanded regions indicated in Fig. 3A.

When using only one latent variable, the statistical parameters indicated a low root mean square error of cross validation (RMSECV = 0.171) and a high correlation coefficient ( $r_{cv} = 0.938$ ). Measured values ( $y = 1$ ) correspond to the traditional class while  $y = 0$  is used for dark roasted samples. The calculated thresholds (see materials and methods) were  $1 \pm 0.584$  and  $0 \pm 0.416$  for traditional and dark roasted, respectively. All samples from the external validation and prediction sets were correctly classified.

Figure 4A shows the scores for latent variable 1 (LV1) plotted against LV2. In this plot the samples from calibration, and prediction sets were clustered in the subspace defined by the first two components of the PLS-DA model. LV1 described 43.81 % of total variance and revealed good separation of the medium from dark roasted coffees.

The loadings plot of LV1 (Fig. 4B) indicated that regions 1, 2, 6, 7, 9 and 12 with positive values significantly contributed to the discrimination of traditional coffee samples (higher absorbance) from dark roasted samples. The other seven regions, important to dark roasted coffees, have negative values.

As mentioned previously in PCA analysis, water is one of the main reasons for the spectral intensities in the region between 3,600 and 3,450/cm (region 1), confirming that dark roasting coffees contain lower amounts of water.

Lipid degradation during the coffee roasting is very low (Toci *et al.* 2006). Since the dry mass of the grain decreases, the higher the roasting degree, higher is the lipid concentration (w/w) in the roasted coffee bean.

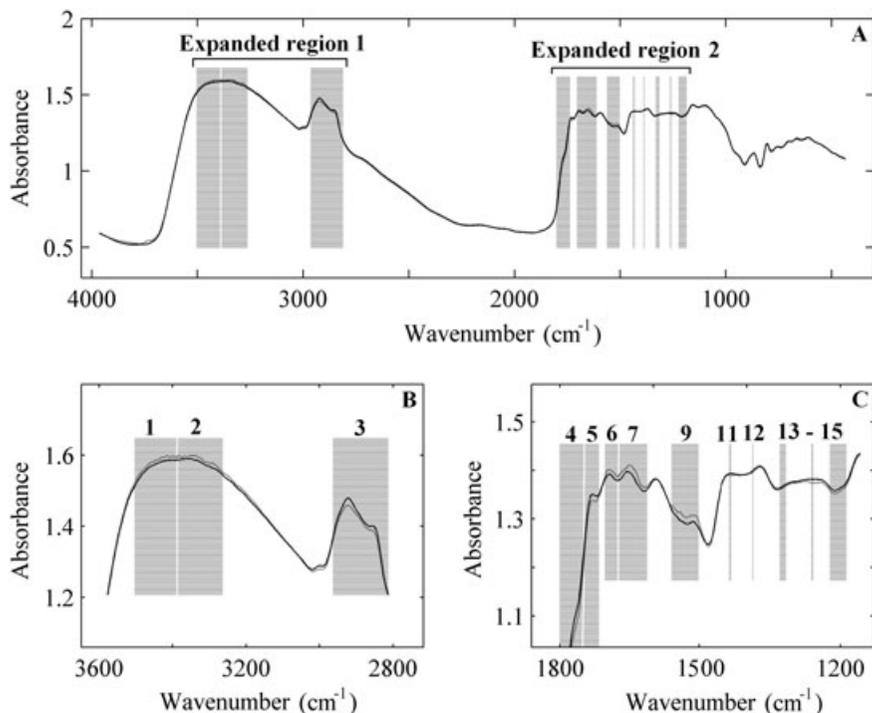


FIG. 3. MID-INFRARED SPECTRA REGIONS SELECTED BY THE OPS METHOD FOR THE CONSTRUCTION OF THE PLS-DA MODEL

Traditional (—) and dark roasted (---) coffee samples (A); expanded region 1 (B) and expanded region 2 (C).

Thus, region 3 in Fig. 3, with negative loadings (corresponding to C-H str. in Table 1) could be explained by an increase in the lipid content.

The degradation of the main classes of compounds that constitute the green coffee beans (chlorogenic acids, amino acids, alkaloids and carbohydrates) generally into volatile and non volatile compounds, during the roasting process (Toci *et al.* 2006), would be responsible for the differences in the expanded region C (Fig. 3C) but, due to the highly complex composition of the coffee matrix, it was not possible to infer about compounds absorbing in this spectral regions selected by the algorithm.

Table 1 also shows the wavenumbers of stretching and deformation vibrations with the respective vibrational modes in the regions selected by OPS to discriminate medium from dark roasted samples.

PLS-DA analysis has shown that it is possible to discriminate medium and dark roasted coffees by performing a variable selection. The discrimination was masked when all useful and nonwavelength were take into account.

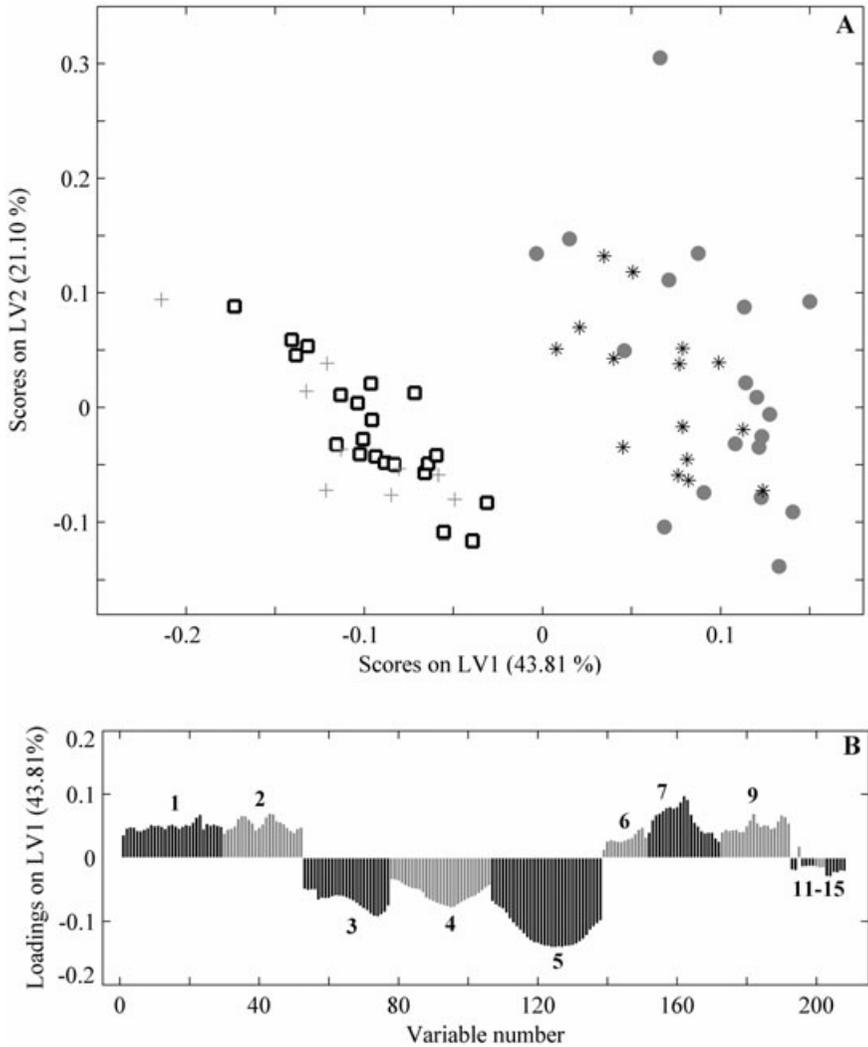


FIG. 4. LV1  $\times$  LV2 SCORES PLOT FROM PLS-DA MODEL

Dark roasted ( $\square$ ) and medium roasted ( $\bullet$ ) coffee samples of the calibration set. Dark roasted (+) and medium roasted (\*) samples of the prediction set (A). Lv1 loadings plot shows the 10 regions selected by ops method (B).

## CONCLUSIONS

The results reported in this work revealed that diffuse reflectance MIR spectroscopy coupled with chemometric tools had enabled the maximization

of chemical information on processed coffees. The chemometric analyses of the MIR spectra allowed inferring on the lower carbohydrate, caffeine and chlorogenic acid concentration as well as on the higher moisture in the decaffeinated coffee compared with traditional and dark roasted coffees. The technique also allowed speculation on the higher lipid and lower moisture content in the dark roasted coffee. Moreover, the chemometric analysis allowed visually locate the wavenumbers for the caffeine and chlorogenic acids in coffee. The exploratory data analysis (PCA) showed the high potential of the technique to discriminate decaffeinated from non decaffeinated coffees. The PLS-DA regression model, using only one latent variable, correctly classified all training and prediction coffee samples according to their roasting degree.

### ACKNOWLEDGMENTS

This work was supported by grants from CAPES and FAPESP.

### REFERENCES

- ANDRÉS, J.M. and BONA, M.T. 2005. Analysis of coal by diffuse reflectance near-infrared spectroscopy. *Anal. Chim. Acta* *535*, 123–132.
- BARKER, M. and RAYENS, W. 2003. Partial least squares for discrimination. *J. Chemometr.* *17*(3), 166–173.
- BEEBE, K.R., PELL, R.J. and SEASHOLTZ, M.B. 1997. *Chemometrics – A Practical Guide*, John Wiley & Sons, New York, NY.
- BERTELLI, D., PLESSI, M., SABATINI, A.G., LOLLI, M. and GRILLENZONI, F. 2007. Classification of Italian honeys by mid-infrared diffuse reflectance spectroscopy (DRIFTS). *Food Chem.* *101*, 1565–1570.
- BRANDO, C.H.J. 2007. Tendências futuras de mercado e demanda de pesquisa: Um ponto de vista. In *C. Cafés de Qualidade: Aspectos Tecnológicos, Científicos e Comerciais* (T.J.G. Salva, O.G. Filho, R.A. Thomaziello and L. Fazuoli, eds.) pp. 352–355, Instituto Agronômico de Campinas, Campinas, Brazil.
- BRIANDET, R., KEMSLEY, E.K. and WILSON, R.H. 1996. Discrimination of Arabica and Robusta in instant coffee by Fourier transform infrared spectroscopy and Chemometrics. *J. Agric. Food Chem.* *44*, 170–174.
- BUFFO, R.A. and CARDELLI-FREIRE, C. 2004. Coffee flavour: An overview. *Flavour Fragrance J.* *19*, 99–104.
- CLARKE, R.J. and VITZTHUM, O.G. 2001. *Coffee: Recent Developments*, World Agriculture Series, Blackwell Science Ltd., Ames, IA.

- CEN, H. and HE, Y. 2007. Theory and application of near infrared reflectance spectroscopy in determination of food quality. *Trends Food Sci. Technol.* 18, 72–83.
- CHEN, Q., ZHAO, J., ZHANG, H. and WANG, X. 2006. Feasibility study on qualitative and quantitative analysis in tea by near infrared spectroscopy with multivariate calibration. *Anal. Chim. Acta* 572, 77–84.
- COZZOLINO, D., SMYTH, H.E. and GISHEN, M. 2003. Feasibility study on the use of visible and near-infrared spectroscopy together with chemometrics to discriminate between commercial white wines of different varietal origins. *J. Agric. Food Chem.* 51, 7703–7708.
- DOWNEY, G., MCINTYRE, P. and DAVIES, A.N. 2003. Geographic classification of extra virgin olive oils from the eastern Mediterranean by chemometric analysis of visible and near-infrared spectroscopy data. *Appl. Spectrosc.* 57, 158–163.
- ESTEBAN-DÍEZ, I., GONZÁLEZ-SÁIZ, J.M. and PIZARRO, C. 2004. Prediction of sensory properties of espresso from roasted coffee samples by near-infrared spectroscopy. *Anal. Chim. Acta* 525, 171–182.
- ESTEBAN-DÍEZ, I., GONZÁLEZ-SÁIZ, J.M., SÁENZ-GONZÁLEZ, C. and PIZARRO, C. 2007. Coffee varietal differentiation based on near infrared spectroscopy. *Talanta* 71, 221–229.
- FABIAN, Z., SALGO, A. and IZVEKOV, V.P. 1992. Determination of actual coffee content in instant coffee mixtures. In *Near infra-red Spectroscopy Bridging the Gap between Data Analysis and NIR Applications* (K.I. Hildrum, T. Isaksson, T. Naes and A. Tandberg, eds.) p. 357, Ellis Horwood Limited, Chichester, England.
- FAGAN, C.C., O'DONNELL, C.P., O'CALLAGHAN, D.J., DOWNEY, G., SHEEHAN, E.M., DELAHUNTY, C.M., EVERARD, C., GUINEE, T.P. and HOWARD, V. 2007. Application of mid-infrared spectroscopy to the prediction of maturity and sensory texture attributes of cheddar cheese. *J. Food Sci.* 72, E130–E137.
- FERRÃO, M.F., FURTADO, J.C., NEUMANN, L.G., KONZEN, P.H.A., MORGANO, M.A., BRAGAGNOLO, N. and FERREIRA, M.M.C. 2003. Técnica não destrutiva de análise de tanino em café empregando espectroscopia no infravermelho e algoritmo genético. *Tecno-lóg* 7, 9–26.
- HE, J., RODRIGUEZ-SAONA, L.E. and GIUSTI, M.M. 2007. Mid infrared spectroscopy for juice authentication-rapid differentiation of comercial juices. *J. Agric. Food Chem.* 55, 4443–4452.
- INÓN, F.A., GARRIGUES, S. and GUARDIA, M. 2006. Combination of mid- and near-infrared spectroscopy for the determination of the quality properties of beers. *Anal. Chim. Acta* 571, 167–174.

- ISAKSSON, T. and NAES, T. 1988. The effect of multiplicative scatter correction (MSC) and linearity improvement in NIR spectroscopy. *Appl. Spectrosc.* *42*, 1273–1284.
- KAROUI, R., MOUAZEN, A.M., DUFOUR, E., PILLINEL, L., SCHALLER, E., PICQUE, D., DE BAERDEMAEKER, J. and BOSSET, J.O. 2006. A comparison and joint use of NIR and MIR spectroscopic methods for the determination of some parameters in European emmental cheese. *Eur. Food Res. Technol.* *223*, 44–50.
- KEMSLEY, E.K., RUAULT, S. and WILSON, R.H. 1995. Discrimination between *Coffea arabica* and *Coffea canephora* variant robusta beans using infrared spectroscopy. *Anal. Nutr. Clin. Met. Sect.* *54*, 321–326.
- LI, W., HUANG, D., LIU, L. and JIANG, B. 2006. The application of near infrared spectroscopy in the quality control analysis of glass/phenolic resin prepreg. *J. Mater. Sci.* *41*, 7183–7189.
- OLINGER, J.M. and GRIFFITHS, P.R. 1993. Effects of sample dilution and particle size/morphology on diffuse reflectance spectra of carbohydrate systems in the near- and mid-infrared. Part I: Single analytes. *Appl. Spectrosc.* *47*, 687–694.
- PARIKH, V.M. 1974. *Absorption Spectroscopy of Organic Molecules*, Addison-Wesley, London, UK.
- PEDRO, A.M.K. and FERREIRA, M.M.C. 2005. Nondestructive determination of solids and carotenoids in tomato products by near-infrared spectroscopy and multivariate calibration. *Anal. Chem.* *77*, 2505–2511.
- PIZARRO, C., ESTEBAN-DÍEZ, I. and GONZÁLEZ-SÁIZ, J.M. 2007. Mixture resolution according to the percentage of robusta variety in order to detect adulteration in roasted coffee by near infrared spectroscopy. *Anal. Chim. Acta* *585*, 266–276.
- REEVES, J.B. and ZAPT, C.M. 1998. Mid-infrared diffuse reflectance spectroscopy for discriminant analysis of food ingredients. *J. Agric. Food Chem.* *46*, 3614–3622.
- REID, L.M., WOODCOCK, T., O'DONNELL, C.P., KELLY, J.D. and DOWNEY, G. 2005. Differentiation of apple juice samples on basis of heat treatment and variety using chemometrics analysis of MIR and NIR data. *Food Res. Int.* *38*, 1109–1115.
- RUBAYIZA, A.B. and MEURENS, M. 2005. Chemical discrimination of arabica and robusta coffees by Fourier transform Raman spectroscopy. *J. Agric. Food Chem.* *53*, 4654–4659.
- SAVITZKY, A. and GOLAY, M. 1964. Smoothing + differentiation of data by simplified least squares procedures. *Anal. Chem.* *36*, 1627–1629.
- SILVERSTEIN, R.M. and WEBSTER, F.X. 1998. *Spectrometric Identification of Organic Compounds*, 6th Ed., Wiley, New York, NY.

- SMALL, G.W. 2006. Chemometrics and near-infrared spectroscopy: Avoiding the pitfalls. *Trends Anal. Chem.* 25, 1057–1066.
- TEÓFILO, R.F., MARTINS, J.P.A. and FERREIRA, M.M.C. 2009. Sorting variables by using informative vectors as a strategy for featuring selection in multivariate regression. *J. Chemom.* 23, 32–41.
- TOCI, A., FARAH, A. and TRUGO, L.C. 2006. Effect of decaffeination using dichloromethane on the chemical composition of arabica and robusta raw and roasted coffees. *Quím. Nova* 29, 965–971.
- WILSON, R.H. and TAPP, H.S. 1999. Mid-infrared spectroscopy for food analysis: Recent new applications and relevant developments in sample presentation methods. *Trends Anal. Chem.* 18, 85–92.
- WISE, B.M., GALLAGHER, N.B., BRO, R., SHAVER, J.M., WINDIG, W. and KOCH, R.S. 2004. *PLS\_Toolbox 3.5, for Use with MatlabTM*, Version 3.5, Eigenvector Research, Inc., Wenatchee, WA.
- WOLD, S., ESBENSEN, K. and GELADI, P. 1987. Principal component analysis. *Chem. Intell. Lab. Syst.* 2, 37–52.
- XIE, L., YING, Y., YING, T., YU, H. and FU, X. 2007. Discrimination of transgenic tomatoes based on visible/near-infrared spectra. *Anal. Chim. Acta* 584, 379–384.
- XING, J., SAEYS, W. and DE BAERDEMAEKER, J. 2007. Combination of chemometric tools and image processing for bruise detection on apples. *Comput. Electr. Agric.* 56, 1–13.
- YAN-DE, L., YI-BIN, Y., XIAPING, F. and GUIBAN, L. 2007. Experiments on predicting sugar content in apples by FT-NIR technique. *J. Food Eng.* 80, 986–989.