

Contents lists available at [ScienceDirect](http://ScienceDirect.com)

# Food Research International

journal homepage: [www.elsevier.com/locate/foodres](http://www.elsevier.com/locate/foodres)

## Review

# Application of infrared spectral techniques on quality and compositional attributes of coffee: An overview



Douglas Fernandes Barbin <sup>a,\*</sup>, Ana Lucia de Souza Madureira Felicio <sup>a</sup>, Da-Wen Sun <sup>b</sup>,  
Suzana Lucy Nixdorf <sup>c</sup>, Elisa Yoko Hirooka <sup>a</sup>

<sup>a</sup> Department of Food Science, Center for Agricultural Sciences, State University of Londrina, Rodovia Celso Garcia Cid, PR 445 Km 380, Campus Universitário, Londrina, PR 86055-900, Brazil

<sup>b</sup> Food Refrigeration and Computerised Food Technology (FRCFT), School of Biosystems Engineering, University College Dublin, National University of Ireland, Agriculture & Food Science Centre, Belfield, Dublin 4, Ireland

<sup>c</sup> Department of Chemistry, Center for Exact Sciences, State University of Londrina, Rodovia Celso Garcia Cid, PR 445 Km 380, Campus Universitário, Londrina, PR 86055-900, Brazil

## ARTICLE INFO

### Article history:

Received 25 November 2013  
Received in revised form 23 December 2013  
Accepted 3 January 2014  
Available online 11 January 2014

### Keywords:

Coffee  
Infrared spectroscopy  
FT-NIR  
Arabica  
Robusta  
Caffeine

## ABSTRACT

During the last two decades, near and mid-infrared spectral analyses have emerged as a reliable and promising analytical tool for objective assessment of coffee quality attributes. The literature presented in this review clearly reveals that near and mid-infrared approaches have a huge potential for gaining rapid information about the chemical composition and related properties of coffee. In addition to its ability for effectively quantifying and characterising quality attributes of some important features of coffee such as moisture, lipids and caffeine content, classification into quality grades and determination of sensory attributes, it is able to measure multiple chemical constituents simultaneously avoiding extensive sample preparation. Developing a quality evaluation system based on infrared spectral information to assess the coffee quality parameters and to ensure its authentication would bring economical benefits to the coffee industry by increasing consumer confidence in the quality of products. This paper provides an overview of the recently developed approaches and latest research carried out in near and mid-infrared spectral technology for evaluating the quality and composition of coffee and the possibility of its widespread deployment.

© 2014 Elsevier Ltd. All rights reserved.

## Contents

1. Introduction . . . . .	23
2. Spectroscopy . . . . .	24
2.1. Spectra pre-processing and multivariate data analyses . . . . .	24
3. Applications . . . . .	25
3.1. Prediction of coffee composition . . . . .	25
3.2. Authentication . . . . .	27
3.3. Classification of samples according to coffee variety and quality features . . . . .	27
3.4. Discrimination between defective and non-defective samples . . . . .	28
3.5. Prediction of sensory properties . . . . .	29
3.6. Degree of roasting . . . . .	29
3.7. Coffee residues . . . . .	30
4. Future trends . . . . .	30
5. Conclusion . . . . .	31
Acknowledgements . . . . .	31
References . . . . .	31

## 1. Introduction

Coffee is the world's most popular beverage after water, and one of the most important raw materials within the international trade for which quality is quintessential. Interest in coffee quality assessment is

\* Corresponding author. Tel.: +55 43 33715963.

E-mail address: [dfbarbin@yahoo.com.br](mailto:dfbarbin@yahoo.com.br) (D.F. Barbin).

impelled by the need to supply the consumer with a consistently high quality product at an affordable price. Indeed, quality is a major aspect for the modern coffee industry because a high quality product is the basis for success in today's particularly competitive market.

Among the several species of the genus *Coffea* identified so far, two of these varieties are economically and commercially important: *Coffea arabica* (arabica coffee) and *Coffea canephora* (robusta coffee) (Huck, Guggenbichler, & Bonn, 2005; Illy & Viani, 1996). Both coffee varieties differ mutually from a botanical perspective and in terms of quality features. *Arabica* beans are more valued by the trade because they are considered to have a finer flavour and therefore more appreciated by the consumers than *robusta*. At present, most commercially available coffee beverages are produced from *arabica* and *robusta* roasted beans or blends of these two varieties (Esteban-Díez et al., 2004a). This raises the possibility of fraudulent or accidental mislabelling, a subject of concern to food processors and regulatory authorities (Downey & Boussion, 1996). Thus, it is important that the varieties of raw beans and of various coffee products can be properly identified. A trained inspector can easily distinguish raw *arabica* and *robusta* beans by differences in size and colour. However, after roasting and milling, these visual indicators are eliminated, and identification of ground roast coffees requires alternative procedures (Kemsley, Ruault, & Wilson, 1995).

Chemical constituents of the roasted beans determine the quality of coffee as a beverage. Raw coffee beans contain a wide range of different chemical compounds, which react and interact amongst themselves at all stages of coffee roasting, resulting in greatly diverse final products (Ribeiro, Ferreira, & Salva, 2011). For instance, caffeine content which has a significant effect on the final quality of the coffee products needs to be determined fast and reliably by analytical techniques. Due to the high number of samples to be analysed, the coffee industry needs new analytical techniques providing fast and reliable data about coffee quality.

Over the past few years, a number of methods have been investigated to measure coffee quality traits in objective ways (Bicchi, Ombretta, Pellegrino, & Vanni, 1997; Farah, Monteiro, Calado, Franca, & Trugo, 2006; Franca, Mendonça, & Oliveira, 2005; Franca, Oliveira, Mendonça, & Silva, 2005; Gonzalez, Pablos, Martin, Leon-Camacho, & Valdenebro, 2001; Mendonça, Franca, & Oliveira, 2009; Mendonça, Franca, Oliveira, & Nunes, 2008; Perrone, Donangelo, & Farah, 2008). Nowadays, the coffee industry has a growing need for consistent reference methods for developing robust and accurate calibration models, which can further be used in the determination of additional samples (Esteban-Díez et al., 2004a). Spectroscopy is an increasingly growing technique due to its rapidity, simplicity, and safety, as well as its ability to measure multiple attributes simultaneously without monotonous sample preparation. More specifically, spectroscopic methods in the visible, near and mid-infrared ranges are rapid, chemical-free and low cost techniques without the need of sample processing that has been widely used to detect the chemical composition of food materials, making it suitable to be implemented on a routine basis (Jing, Cai, & Shao, 2010; Pedro & Ferreira, 2005; Santos, Sarraguça, Rangel, & Lopes, 2012; Yan-De, Yi-Bin, Xiaping, & Guisan, 2007).

This paper thus provides an overview of the application of infrared spectral technique as a reliable procedure compared to traditional analytical methods to assess coffee composition and quality parameters and classify coffee samples from different varieties and quality grades.

## 2. Spectroscopy

Near and mid-infrared spectral techniques have received considerable attention among researchers as a tool for rapid, non-destructive and simultaneous assessment of multiple attributes of organic materials without lengthy sample preparation. While near infrared (NIR) spectroscopy requires almost no sample preparations, obtaining a diffuse reflectance infrared Fourier transform (DRIFT) spectrum requires some prudence. A powdered or ground sample is placed into a stainless

steel cup of approximately 10 mm in diameter, and the sample surface is flattened. The cup is placed in a DRIFT accessory, which incorporates suitable mirrors to steer the infrared beam onto the sample, to collect the portion that is diffusely reflected from the sample surface and to direct it onto the detector to record the spectrum. For certain samples, dilution in another matrix, for example, powdered infrared grade potassium bromide (KBr), is essential to avoid optical distortion effects in the spectra (Kemsley et al., 1995). Essentially, spectroscopy provides comprehensive information of biological samples as an effect of the interaction between electromagnetic radiation and the sample material, such as reflectance, transmittance, absorbance, phosphorescence, fluorescence, and radioactive decay (Abbott, 1999; Blanco & Villarroya, 2002).

Near-infrared spectroscopy (NIRS) is based on molecular overtones and combination vibrations that can be very useful in probing bulk material with little or no sample preparation (Huck et al., 2005; Pasquini, 2003). The sensitivity of spectroscopic measurements relies on band intensities. Spectral information is repeated throughout the successive overtones and combination regions; however, the signals of bands involved become weaker by an order of magnitude. The third overtones encountered in shorter wavelengths such as the visible region (400–800 nm) have considerably weaker band intensities when compared to the second and first overtones in the NIR region (800–1800 nm) (Osborne, Fearn, & Hindle, 1993; Weyer, 1985). Although some unique molecule vibration can be properly assigned in the NIR region of the spectra, the complex composition of organic samples such as coffee, where over 800 compounds ranging from simple linear and branched structures to complex cyclic and heterocyclic structures can be encountered, makes it difficult to analyse (Illy & Viani, 1996; Lyman, Benck, Dell, Merle, & Murray-Wijelath, 2003; Wang, Jun, Bittenbender, Gautz, & Li, 2009).

The literature on chemical assignments for some of the most important NIR bands is summarized in a schematic representation of spectra-structure correlation in Fig. 1. In addition, the wavelength range in which have been reported the greatest absorbance of pure components of coffee are presented.

Applications of infrared spectral analyses have increased in food product quality analysis, and it has been widely used to predict the attributes of grains (Jing et al., 2010), tea (Chen, Zhao, Zhang, & Wang, 2006; Schulz, 2004), honey (Chen et al., 2011), fruits and vegetables (Bauer et al., 2008; Pedro & Ferreira, 2005; Xie, Ying, Ying, Yu, & Fu, 2007; Yan-De et al., 2007), beverages (Cozzolino, Smyth, & Gishen, 2003; He, Rodriguez-Saona, & Giusti, 2007; Inón, Garrigues, & Guardia, 2006), oils (Downey, McIntyre, & Davies, 2003), dairy products (Downey et al., 2005; Fagan et al., 2007; Woodcock, Fagan, O'Donnell, & Downey, 2008), eggs (Abdel-Nour, Ngadi, Prasher, & Karimi, 2011), meat (De Marchi, 2013; Liao, Fan, & Cheng, 2012), and other agricultural products, in order to substitute other commonly used destructive methods. It is widely used to quantify the composition of agricultural products because it meets the criteria of being accurate, reliable, rapid, non-destructive, and relatively inexpensive. Fourier transform infrared (FTIR) has also been investigated regarding coffee applications, although this technique is more sensitive to the heterogeneity of the individual coffee samples, because stretching and deformation vibrations are directly used and not their overtones or combination vibrations.

### 2.1. Spectra pre-processing and multivariate data analyses

Certain effects such as light scattering that occurs in diffuse reflectance NIR spectroscopy are influenced by sample presentation causing unexpected perturbations in spectra (baseline shifts, slope changes and non linearity), generating spectral changes that are not related to the studied responses and affecting the reliability of multivariate calibration models (Osborne et al., 1993). As a result, NIR signals contain information which is not only from chemical but also physical characteristics from samples and measuring conditions. Since wavelength dependency of light scatter is different from that of chemically based light

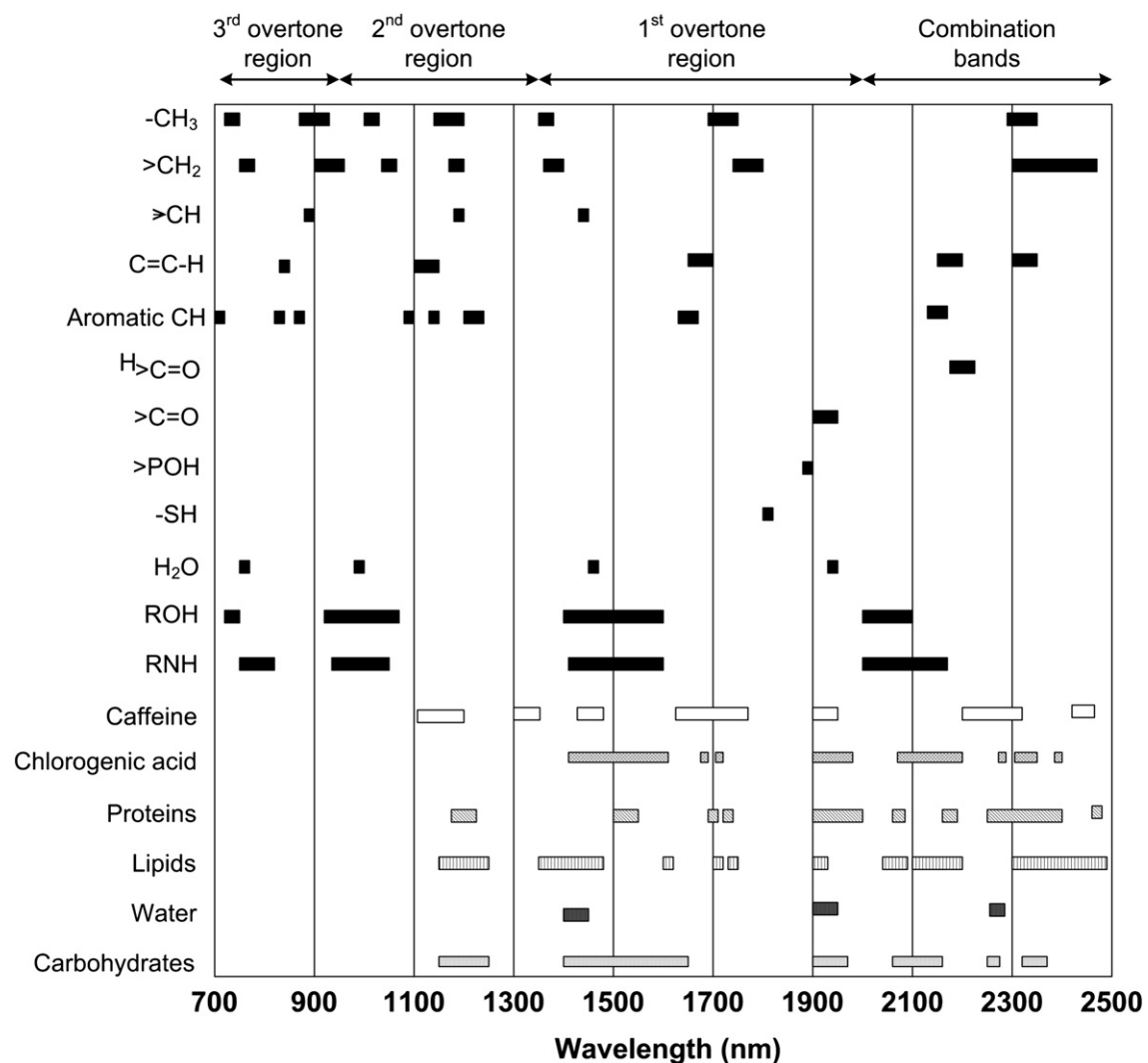


Fig. 1. Schematic representation of spectra–structure correlation and absorption regions of main components found in coffee (adapted from Osborne et al., 1993; Ribeiro et al., 2011).

absorption, scattering effects can be attenuated by some mathematical treatments, which include derivation, multiplicative scatter correction (MSC), standard normal variate (SNV) and orthogonal signal correction (OSC) (Esteban-Díez et al., 2004a). Therefore, the application of suitable pre-processing methods, aimed at correcting spectral data by minimizing the contribution of physical effects to NIR signals is a crucial step in regression model development.

Additionally, one of the characteristics inherent to spectroscopic analysis is the number of variables (several hundreds or even thousands) that exist in the NIR spectra, usually significantly larger than the number of samples. For this reason, they contain interferences coming from background, noise and overlapping bands, leading to redundancy and co-linearity in these variables. The molecular overtone and combination bands seen in the NIR range are typically very broad, leading to complex spectra; thus being a difficult task to assign particular features to specific chemical components. It brings a great challenge to build a high-quality prediction model for unknown set of samples. Thus, the application of an efficient variable selection and efficient signal de-noising and compression methods prior to the calibration step has been proved to be greatly beneficial in providing more reliable and parsimonious models (Pizarro, Esteban-Díez, & González-Sáiz, 2007; Teófilo, Martins, & Ferreira, 2009; Zhang et al., 2013).

Multivariate (multiple variables) calibration techniques such as principal components analysis (PCA), partial least squares (PLS), or

artificial neural networks (ANN) are often employed to extract the desired chemical information. Other multivariate techniques that have found applicability in spectral analyses include genetic algorithm (GA), soft independent modelling of class analogies (SIMCA), and successive projection algorithms (SPA), among others (Zhang et al., 2013). Careful development of a set of calibration samples and application of multivariate calibration techniques are essential for near-infrared analytical methods (Isaksson & Naes, 1988).

### 3. Applications

The feasibility of NIR techniques as an alternative to reference methods for coffee classification and authentication, and prediction of coffee composition have been investigated. Most recent applications of NIR spectroscopy to coffee samples are listed in Table 1. Several spectral pre-processing techniques and multivariate analyses were tested. Relevant observations and results are further presented and discussed.

#### 3.1. Prediction of coffee composition

Several works have investigated the potential application of spectral techniques to assess physical, chemical and quality parameters of coffee. Most of the investigations have focused on developing an alternative to reference methods for coffee attribute prediction and to classify coffee

**Table 1**  
Most recent applications of IR spectral analysis in coffee.

Application	Task	Wavelength range	Statistical analysis	Author(s)/year
Adulteration/ authentication	Presence of glucose, starch, chicory	800–4000 cm <sup>-1</sup> (2500–12,500 nm) FT-IR	LDA, PCA, PLS, ANN	Briandet, Kemsley, & Wilson (1996)
	Presence of barley	10,000–4000 cm <sup>-1</sup> (1000–2500 nm) FT-NIR	PLS and PLS-GA	Ebrahimi-Najafabadi et al. (2012)
	<i>Robusta</i> content in roasted coffee	1100–2500 nm	PLS	Pizarro, Esteban-Díez, & González-Sáiz (2007)
Classification/ discrimination	Authentication of Kona coffee	4000–400 cm <sup>-1</sup> (2500–25,000 nm) FT-IR	PLS and PCR	Wang et al. (2009)
	Samples from different batches and producers	800–2500 nm	PLS-DA and SIMCA	Santos, Moura, et al. (2012)
	Identifying <i>arabica</i> and <i>robusta</i>	1100–2500 nm	Potential functions methods	Esteban-Díez, González-Sáiz and Pizarro (2004a)
	Classification into <i>arabica</i> and <i>robusta</i>	1100–2500 nm	Potential functions methods	Esteban-Díez, González-Sáiz, Saenz-Gonzalez, & Pizarro (2007)
	Identifying <i>arabica</i> and <i>robusta</i>	400–2498 nm	(PCA) and FDA	Downey & Boussion (1996)
Discrimination/ prediction	Identifying <i>arabica</i> and <i>robusta</i>	400–2498 nm/4000–800 cm <sup>-1</sup> (2500–12,500 nm)	FDA and PLS	Downey, Briandet, Wilson, & Kemsley (1997)
	<i>Arabica</i> and <i>robusta</i>	800–4000 cm <sup>-1</sup> (2500–12,500 nm) FT-IR	PCA and DA	Kemsley et al. (1995)
	Defective and non-defective beans	4000–700 cm <sup>-1</sup> (2500–14,285 nm) FT-IR	PCA, clusters	Craig, Franca, & Oliveira (2012a)
	Defective and non-defective beans	4000–400 cm <sup>-1</sup> (2500–25,000 nm) FT-IR	PCA and LDA	Craig, Franca, & Oliveira (2012b)
	Defective beans	10,000–4000 cm <sup>-1</sup> (1000–2500 nm) – FT-IR	PCA and PLS	Santos, Sarraguça, Rangel, & Lopes (2012)
Composition	Caffeine content	800–4000 cm <sup>-1</sup> (2500–12,500 nm) FT-IR	Linear regression	Garrigues, Bouhsain, Garrigues, & de la Guardia (2000)
	Caffeine, theobromine and theophylline	9996–4008 cm <sup>-1</sup> (1000–2495 nm) FT-NIR	PLS	Huck et al. (2005)
	Moisture content	4500–10,000 cm <sup>-1</sup> (1000–2222 nm)	PLS	Morgano, Faria, Ferrão, Bragagnolo, & Ferreira (2008)
	Ash and lipid content	1100–2500 nm	PLS	Pizarro, Esteban-Díez, Nistal, & González-Sáiz (2004)
	Caffeine content and roasting colour	1100–2500 nm	PLS	Pizarro, Esteban-Díez, Gonzalez-Saiz, & Forina (2007)
Sensory prediction	Caffeine content	4000–10,000 cm <sup>-1</sup> (1000–2500 nm)	PLS	Zhang et al. (2013)
	Sensory attributes	4000–400 cm <sup>-1</sup> (2500–25,000 nm) FT-IR	PCA and PLS-DA	Ribeiro, Salva, & Ferreira (2010)
	Sensory attributes	1100–2500 nm	PLS	Ribeiro et al. (2011)
	Sensory properties	1100–2200 nm	PLS (and IPW-PLS)	Esteban-Díez, González-Sáiz, & Pizarro (2004c)
Degree of roasting	Prediction of roasting degree	4000–12,000 cm <sup>-1</sup> (830–2500 nm) FT-NIR	PLS	Alessandrini, Romani, Pinnavaia, & Rosa (2008)
	Effect on roasting conditions	4000–700 cm <sup>-1</sup> (2500–14,285 nm) FT-IR	–	Lyman et al. (2003)

according to physico-chemical characteristics into coffee varieties or origins.

Water content above 12.5% in green coffee causes a number of unwanted consequences like microbial growth, mycotoxin formation, altered sensorial quality of final product, and unstable production conditions among others. Several standards for reference, routine and rapid methods are already established for determination of water content in green coffee. Nevertheless, there are discussions on how effective the methods are in determining the water content. The work of Reh, Gerber, Prodolliet, and Vuataz (2006) focused on clarifying the specificity and accuracy of several available reference methods for determination of water content in green coffee, demonstrating that only ISO 1446 exclusively measures water. For all drying oven based methods it was observed that degradation of the product contributes to the overall weight loss. Near-infrared spectroscopy and colour measurement were applied to study the different drying methods, establishing the degree of degradation and the completeness of the drying process. The drying process was straightforwardly monitored due to the specificity of the 1940 nm wavelength band for water molecules by comparing the weight loss results with the residual NIRS absorption left after drying. It was observed that neither reference methods ISO 1446 nor ISO 6673 allows complete drying of green coffee, since ISO 1446 has proved to let residual water content at the end of the drying process (e.g., 2 weeks). This residual value was very sensitive to the particle size, due to the very low water diffusion in dehydrated green coffee material. All methods discussed (ISO 1447; ISO 6673; ISO 1446) were shown to be highly correlated with each other with R<sup>2</sup> values above 0.99.

On another study, NIR spectroscopy was investigated to measure the moisture content in raw coffee. The original NIR spectra were pre-processed through different transformations and mathematical pre-treatments, such as the Kubelka–Munk transformation; multiplicative signal correction (MSC); spline smoothing and movable average. The

most important bands were mainly related to the vibration of O–H and C–H bonds between 4650 and 4900, 5004 and 5062 cm<sup>-1</sup>. The regression model allowed for the determination of moisture content of raw coffee samples with a standard error of validation of 0.298 g · 100 g<sup>-1</sup> and correlation coefficient (r) of 0.818 for validation, with an average relative error of 4.1% (Morgano et al., 2008).

Caffeine content is an important component of coffee and thus has been the target of several investigations. Ground *arabica* coffee samples within a wide range of roasted levels were analysed by NIR in order to obtain a reduced-spectrum model for quantitative determination of caffeine content. PLS regression was used to construct calibration models based on chemometric analyses of the NIR spectral data and caffeine contents determined by the high performance liquid chromatography with ultraviolet (HPLC–UV) detection method as the reference values. Comparing the respective quality of the different models studied, the application of a second derivative pre-treatment and stability competitive adaptive reweighted sampling (SCARS) variable selection provided a notably improved regression model, with a root mean square error of prediction (RMSEP) of 0.378 mg/g, mean relative error of 1.98% and mean relative standard deviation (RSD) of 1.71%. Thus, the results revealed the feasibility of NIR spectroscopy for at-line application to predict the caffeine content of unknown roasted coffee samples, thanks to the short analysis time of a few seconds and non-destructive advantages of NIRS (Zhang et al., 2013).

Regarding caffeine content, a fast procedure was proposed for the FTIR determination of this component in roasted coffee samples with reduced amount of organic solvent used. The method involves extracting the caffeine with CHCl<sub>3</sub>, and measuring absorbance at 1659 cm<sup>-1</sup> using a baseline established between 1900 and 830 cm<sup>-1</sup>. The procedure provides a drastic reduction of the organic solvent consumed by the reference chromatographic UV–spectrometric determination, with a limit of detection of the order of 3 mg L<sup>-1</sup> of caffeine. The accuracy of the FTIR procedure was higher than 94% (Garrigues et al., 2000).

Quantitative calibration models were developed using diffuse reflectance NIR spectra of liquid coffee beverages as predictors collected by a FT-NIR spectrometer. Determination coefficients ( $R^2$ ) for the calibration models of instant coffee, plant fat and sugar, were 98.97%, 99.94% and 99.18% respectively; with root mean square errors of cross validation (RMSECV) of 2.12, 0.72 and 2.01 respectively, indicating that NIR spectroscopy can be applied for the simultaneous determination of the three main ingredients in liquid coffee beverage (Wang, Min, Duan, Xiong, & Li, 2012).

An analytical method based on NIRS for the quantification of alkaloids (caffeine, theobromine and theophylline) was investigated in roasted coffee after discrimination of the rough green beans into *arabica* and *robusta*. This validated method was compared to the most commonly used liquid chromatography (LC) connected to UV and mass spectrometric (MS) detection. Among the most important spectral features, it was noticed that the vibration of the 2nd overtone of the carbonyl group ( $5352\text{ cm}^{-1}$ ), C–H stretch and C–H deformation vibration, and the  $\text{CH}_2$  overtone ( $5742\text{ cm}^{-1}$ ), was caused by ingredients such as lipids, volatile and non-volatile acids, chlorogenic acids, alkaloids and some aroma compounds. Theobromine has one methyl group less than caffeine and showed differences compared to caffeine at  $7353\text{ cm}^{-1}$  (C–H str. + C–H def.),  $5865\text{ cm}^{-1}$  (C–H str., first overtone) and  $4383\text{ cm}^{-1}$  (C–H str. + C–H def.). Analysis of 83 liquid coffee extracts resulted in correlation coefficients of 0.86 and 0.85 in a concentration range between 0.10 and  $4.13\text{ g}\cdot 100\text{ g}^{-1}$ , with values for standard error of prediction (SEP) of 0.07 and  $0.10\text{ g}\cdot 100\text{ g}^{-1}$  for caffeine and theobromine, respectively. The robustness and reproducibility of the NIRS model for the determination of caffeine and theobromine showed that it can be used to predict the content of these components in liquid extracts, with the possibility for the coffee producing industry to replace the LC method usually applied to determine these components in the routine analysis (Huck, Maurer, Popp, Basener, & Bonn, 1999). However, due to the lower limit of detection of the NIRS-method, the analysis of theophylline by LC method has to be favoured due to the sensitivity of the method (Huck et al., 2005).

### 3.2. Authentication

Food authentication is a major issue that has become increasingly important in recent years. The need to guarantee the actual origin of a product and to determine whether it has been mixed with adulterants is a major challenge to the food industry, regulatory authorities and consumers, which are all interested in authenticating raw materials and food products in order to satisfy food quality and safety requirements (Pizarro, Esteban-Diez, & González-Sáiz, 2007).

The excellent quality of Kona coffee, the variety of “Kona typica” grown in Kona-Island (Hawaii, U.S.A.), carries a unique stamp and is among the most distinctive coffee products in the world. Fourier transform infrared (FTIR) spectroscopy integrated with an attenuated total reflectance (ATR) accessory and multivariate analysis were used to develop an effective analytical tool to discriminate pure Kona coffee and its adulterants and to quantify the percentage of Kona coffee in Kona coffee blends. A calibration model was successfully validated using nine blend sets of 100% Kona coffee mixture and its adulterant, 100% non-Kona coffee mixture. Distinct peak variations of ground and brewed coffee blends in the spectral region were observed between  $800$  and  $1900\text{ cm}^{-1}$ . A PLS-2nd derivative calibration model based on brewed Kona coffee with mean centering data processing showed the best predictability with the lowest standard error of calibration value of 0.81 and the highest  $R^2$  value of 0.99 (Wang et al., 2009).

Regarding the identification of different coffee blends and authentication issues, an investigation based on NIR spectral information has been carried out for the identification and quantification of the fraudulent addition of barley in roasted and ground coffee samples. Nine types of commercial coffee samples, including pure *arabica*, *robusta* and mixtures of them were chosen to extensively explore the variability of

coffee present on the market. Furthermore, four types of barley samples, with different roasting degrees, were blended to coffee in concentrations between 2 and 20% weight per weight of barley, with the lowest limit being surely lower than the sensorial limit of detection, and resolution step (2%) lower than human sensorial capability for distinguishing between close quantities. A genetic algorithm (GA) was applied in order to obtain simplified models, taking into account only informative regions of the spectral profiles. Very low prediction errors were achieved on an external test set, revealing that NIRS can find applications to quantify adulterants in coffee powder. However, it must be emphasized that sample collection and analysis should be performed through a number of years, in order to account for a wider range of variability factors and to obtain models for a large-scale applicability (Ebrahimi-Najafabadi et al., 2012). The promising results obtained can be considered in future applications to quantify different blends of varieties in order to identify fraudulent mixtures.

Among other types of coffee adulteration, the most commonly encountered are adulteration with coffee substitutes such as chicory, malt, figs, cereals, caramel, starch, maltodextrins or glucose, as well as roasted or even unroasted coffee husks; mixing of two species (addition of cheaper *robusta* to pure *arabica* coffee); or mixing of expensive coffee beans from one growing region with cheap beans grown in another region. Most used methods for detection of coffee adulteration are generally based on chemical parameters (Franca, Oliveira, Mendonça, & Silva, 2005; Gonzalez et al., 2001; Mendonça et al., 2009). Infrared spectroscopy has been explored as an alternative to existing methods for the detection of undeclared material in instant coffee. Two different FTIR sampling methods were tested: diffuse reflectance, and attenuated total reflectance. PCA and linear discriminant analysis (LDA) were applied to distinguish between pure samples and samples adulterated with glucose, starch or chicory in the range  $20$ – $100\text{ g kg}^{-1}$ , achieving a 98% successful classification rate. Simultaneous PLS regression was carried out for the content determination of three added carbohydrates (xylose, glucose and fructose). An artificial neural network was used to classify pure and adulterated instant coffee based on spectral information, attaining 100% of correctly classified samples. The results presented show that the different statistical approaches yield complementary information. The PCA gives a visual representation of the position of samples in a low-dimensional space, and the LDA uses the relative positions of samples in this space to form a useful classification rule. The ANN analysis offered an improvement over the classification results obtained by LDA, while the PLS regressions for the added xylose, glucose and fructose contents show that FTIR spectroscopy has the required sensitivity for determining the carbohydrate profile of samples (Briand et al., 1996).

### 3.3. Classification of samples according to coffee variety and quality features

Generally, *arabica* coffee beans are considered superior to *robusta* and are therefore the more expensive of the two. As a result, the identification of adulterations and mislabelling is very important regarding the consumer protection (Huck et al., 2005). The potential of NIR reflectance spectroscopy for discriminating between *arabica* and *robusta* coffees and characterization of blends between these two varieties has been investigated with relative success. Downey et al. (1997) compared the near- and mid-infrared spectra of a set of 56 lyophilized coffee samples to discriminate between *arabica* and *robusta* lyophilized coffees. Separate mathematical models were used for distinct wavelength ranges used as predictors separately and combined. Varietal authentication of the coffees was tested using factorial discriminant analysis (FDA) and PLS. The combination of the two wavelength ranges required fewer loadings or principal components for both FDA and PLS approaches tested than with either spectral range alone, suggesting that a broader spectral range may have advantages for this particular type of assessment.

On another study, dried beverages were produced by both lyophilisation and air-drying under vacuum on a glass-fibre filter paper, and spectral information in the wavelength range of 1100–2498 nm was used for classification into coffee varieties. Identification of spectral features from bonding structures observed at 1672, 2158, 2250, 2332 and 2416 nm was related with features in caffeine, implying the difference in caffeine between dried *arabica* and *robusta* coffee beverages. In addition, peaks related to water at 1454 and 1930 nm were observed influencing the analyses. Correct classification rates of 87% and 95% were achieved on lyophilised and vacuum-dried beverage samples respectively, supporting its application for authentication tasks. It was observed that the basis for this discrimination appears to involve caffeine and/or other alkaloids, thus indicating that the quantification of these components is a relevant aspect for coffee analyses (Downey & Bousson, 1996).

In an attempt to improve the classification models constructed using original roasted coffee NIR spectra, orthogonal signal correction (OSC) and direct orthogonal signal correction (DOSC) methods were applied on the raw spectra. The proposed method aimed for removing information not related to an explicitly chosen chemical variable that could provide discrimination between *arabica* and *robusta* coffees. Caffeine content was preferred as this specific descriptor from among other chemical parameters (chlorogenic acid and total acidity), since it showed a notable discrepancy between varieties. The results obtained before and after transforming the spectra were analysed and compared in order to evaluate the effect of orthogonal corrections. It was demonstrated that the application of an orthogonal signal correction method induced a substantial improvement of the classification models in comparison with the model developed on the basis of original spectra, even though classification models developed from unprocessed NIR spectra provided satisfactory classification results. Moreover, the results obtained using the methodologies proposed were as good as those obtained with the classification model constructed using a chemical descriptor in terms of both classification and prediction abilities (Esteban-Diez et al., 2004a). On another study, FTIR spectroscopy in the mid-infrared region (800–4000  $\text{cm}^{-1}$ ) combined with the diffuse reflectance infrared Fourier transform spectroscopy (DRIFTS) sampling technique and PCA revealed clear grouping of the spectra for discrimination of *arabica* and *robusta* ground roast coffees. Classical discriminant analysis (DA) based on principal components scores yielded 100% successful discrimination. However, this work did not report the identification of blends of different concentrations from the two varieties (Kemsley et al., 1995).

Nevertheless, although results have shown the feasibility of using original NIR spectra of coffee samples to develop classification models with good discrimination abilities between pure coffee varieties, it would be interesting to avoid potential errors when working with different coffee blends. Quantifying the *robusta* variety content of roasted coffee samples, as a means for controlling and avoiding coffee adulteration, is a very important issue taking into account the great variability of the final sale price depending on coffee varietal origin (Pizarro, Esteban-Diez, & González-Sáiz, 2007).

Esteban-Diez et al., 2007 proposed an innovative approach capable of discriminating between *arabica* and *robusta* pure coffee samples and blends of these two varieties. The results reported have demonstrated that the combination of near-infrared spectroscopy with the OSC method and with a class-modelling technique, such as potential functions method, yielded better results than using the raw spectra. Therefore, this approach can be used as an improved strategy not only for discriminating between *arabica* and *robusta* pure coffee varieties, but also for differentiating between blends of the two species, even when separate categories of blends are defined depending on the actual *robusta* content in samples. The applied strategy did not imply any additional reference analysis, since the response variable was precisely the percentage of *robusta* coffee in each sample already known for calibration. Therefore, the applied classification strategy only relied on NIR

measurements, proving its reliability and effectiveness to assess the genuineness of coffee samples.

Orthogonal wavelet correction (OWAVEC) pre-processing method combined with PLS regression was applied in order to simultaneously operate two crucial pre-processing steps in multivariate calibration: signal correction and data compression. Several pre-processing methods (mean centering, first derivative, OSC and DOSC) were additionally applied in order to find calibration models with the best predictive ability and to evaluate the performance of the OWAVEC method, comparing the respective quality of the different regression models. However, the success of this methodology was highly dependent on signal pre-processing methods applied to minimize the spectral variation not due to the parameter of interest but due to variation in experimental or sample conditions. The regression models developed for predicting *robusta* content of blends, after pre-processing first derivative NIR spectra by OWAVEC, were accurate (0.79% RMSEP) and considerably better compared to the models obtained from raw data (in terms of both reliability and parsimony) and improved results were also obtained with respect to the other orthogonal signal correction methods tested. The high quality calibration models constructed suggest that the proposed methodology would be a suitable tool for the quantitative detection of *robusta* addition to *arabica* coffees. However, the authors state that although it was a feasibility study and further research should be carried out before its value for authentication purposes. The data set used in this work was designed to cover, as much as possible, the great natural variability inherent in commercially available coffees by considering different roasting conditions and degrees, and varied geographical origins. However, it is important to emphasize that the dynamic nature of the coffee market would require a more exhaustive collection of calibration samples to serve as the basis for developing a predictive model for use in the routine quality control of coffee samples (Pizarro, Esteban-Diez, & González-Sáiz, 2007).

Analogous to the situation found in calibration, a classification model constructed from spectra measured on one instrument may not be valid for prediction of class from spectra measured on a second instrument. Thus, the transfer of multivariate classification models between different near-infrared spectrometers was investigated for the discrimination of whole, green *arabica* and *robusta* coffee beans. Orthogonal signal correction trained on a vector of discrete class identities was found to perform well in the pre-processing of data allowing the transfer of a classification model developed on data acquired on one instrument to be used on another instrument, resulting in robust models for the discrimination of green coffee beans on both spectrometers with misclassification errors in the range of 5–10% (Myles, Zimmerman, & Brown, 2006).

Other studies have reported the ability of NIR spectroscopy in discriminating coffees from different qualities. Santos, Moura, et al. (2012) described the use of near infrared spectroscopy (NIRS) and chemometric techniques for the classification of coffee samples from different lots and producers acquired in supermarkets and roasting industries in Brazil. In general, it was found that NIR spectroscopy, coupled with either SIMCA or partial least squares discriminant analysis (PLS-DA) multivariate models, can be a useful tool to differentiate roasted coffee grains.

#### 3.4. Discrimination between defective and non-defective samples

Characterisation of coffee quality based on bean quality assessment is associated with the relative amount of defective beans among non-defective ones. It is therefore important to develop a methodology that enables a fast assessment of coffee grade and capable of becoming an analytical tool to regulate coffee quality. A methodology for quality assessment of green coffee based on near infrared spectroscopy (NIRS) was proposed for fast assessment of *arabica* and *robusta* coffee varieties from different geographical locations and identification of the presence of defective beans in a batch. PLS regression relating the NIR

spectrum to the mass fraction of defective and non-defective beans presented relative errors around 5% for four coffee provenances investigated (two *robusta*, Indonesia and Vietnam, and two *arabica*, Colombia and Nicaragua). The results obtained are an indication that although the defects could not be recognized individually it could be possible to quantify the non-defective quality against the remaining qualities, showing that NIRS enables a fast, simple and quantitative evaluation of green coffee following general international guidelines of coffee classification (Santos, Sarraguça, Rangel, & Lopes, 2012).

The feasibility of employing DRIFTS for discrimination between defective and non-defective coffees after roasting and grinding has shown promising results. PCA applied to the spectra and of the first derivatives of the spectra provided separation of the samples into four groups: non-defective, black, dark sour and light sour, with immature beans scattered among the sour samples. Classification models based on LDA ranged from 95 to 100% (Craig et al., 2012b). On another study, Craig et al. (2012a) evaluated the potential of FTIR spectroscopy combined with PCA and cluster analysis to discriminate defective and non-defective beans. Results have shown the separation of the samples into two major groups: non-defective/light sour and black/dark sour/immature, indicating that non-defective and defective (black, immature and sour) coffee samples could be separated into distinct groups based on transmittance or reflectance spectra. It is indicated that FTIR presents potential for the development of a fast and reliable analytical methodology for discrimination between defective and non-defective coffee, although the authors suggest that further studies should be conducted employing larger sets of samples in order to develop more robust predictive models. It is emphasized, however, that all classes of compounds are evenly distributed throughout the bean, including its surface. Therefore, when the beans are ground, the chemical makeup of the bean surface will remain embedded in the sample and thus will contribute to the makeup of the FTIR spectra.

### 3.5. Prediction of sensory properties

Further studies were carried out to establish a relationship between the sensory attributes of the beverage and the chemical components of the coffee beans. Chemometric analyses based on coffee beverage sensory data and NIR spectra of *arabica* roasted coffee samples aimed to predict the scores of acidity, bitterness, flavour, cleanliness, body and overall quality of coffee beverage determined by sensory panellists. PLS were used to establish prediction models and the ordered predictor selection (OPS) algorithm was applied as a primary step to select the wavelengths mostly related to each sensory attribute in order to select only significant regions for the regression models. All models presented good correlation ( $r_{cv}$ ) values for the respective attribute predicted thus being useful for associating specific spectral regions to the spectra of the pure compounds. It was observed that the regions of the spectrum defined as important for sensory quality were closely related to the NIR spectra of pure caffeine, trigonelline, 5-caffeoylquinic acid, cellulose, coffee lipids, sucrose and casein. The NIR analyses confirmed the relationship between the sensory characteristics of the beverage and the chemical composition of the roasted grain, demonstrating that the lipids and proteins in the roasted bean were closely related to the attribute of body in the coffee beverage, caffeine and chlorogenic acids to bitterness and chlorogenic acid to acidity, and the flavour, cleanliness and overall quality were related to the caffeine, trigonelline, chlorogenic acid, polysaccharides, sucrose and protein present in the roasted coffee beans (Ribeiro et al., 2011).

Application of NIR spectroscopy for predicting sensory attributes from an 'espresso' quality assurance (perceived acidity, mouthfeel (body), bitterness and aftertaste) was studied. Wavelength selection was performed applying iterative predictor weighting-PLS (IPW-PLS) (Forina, Casolino, & Pizarro, 1999) in order to take into account only significant and characteristic spectral features to improve the quality of the final regression models constructed. The objective was to test the

hypothesis that a low number of NIR absorption bands corresponding to certain compounds strongly related to a specific attribute can predict it with similar or even improved accuracy and reliability as using the whole NIR wavelength range. Specific scores obtained for each espresso sample in the sensory analysis were the direct consequence of the particular chemical composition of the roasted coffee used as raw material. Thus, separate PLS regression models were constructed based on the corresponding NIR roasted coffee spectra in order to model each sensory property studied. The results provided by the calibration models were comparable in terms of accuracy to the evaluations provided by a trained sensory panel, proving the feasibility of using such methodology in on-line or routine applications (Esteban-Díez et al., 2004c).

### 3.6. Degree of roasting

Roasting colour and quality parameters have been targeted as main attributes investigated by NIR spectroscopy (Esteban-Díez, González-Sáiz, & Pizarro, 2004b). For instance, FTIR-ATR was used to study the effect of dark, medium and light roasting conditions on coffee brewed from Guatemala Antigua *arabica* coffee beans. The 1800–1680  $\text{cm}^{-1}$  carbonyl region of the spectra provided a 'flavour-print' which appears to be consistent with the taste and aroma perceived by sensory panellists. The composition of some carbonyl compounds (vinyl esters/lactones, esters, aldehydes, ketones, and acids) in the brewed coffee appears to be dependent upon the heating rate of the green beans to the onset of the first and second cracks. It was observed that when the rate of heating to the onset of the first and second cracks was kept constant, the types of carbonyl compounds formed were similar, varying only in their concentration. This difference in concentration is apparently due to the additional heating of the coffee bean beyond the second crack. When the heating rate to the onset of the first and second cracks was varied, both the types and concentration of the carbonyl compounds formed during roasting were affected. Thus, heating rates of green coffee beans to the onset of the first and second cracks are important determinants of the basic taste and aroma of brewed coffee (Lyman et al., 2003).

The potential of mid-infrared diffuse reflectance spectroscopy with Fourier transform was investigated for discrimination of commercial coffee samples. A clear discrimination of decaffeinated from medium and dark roasted coffees was observed by application of PCA. A PLS-DA model correctly classified 100% of the external validation and prediction samples according to their roasting degree. Chemometric analysis 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 (Ribeiro et al., 2010).

The most common mathematical treatments applied to correct systematic variations unrelated to the studied responses are derivation SNV and MSC. Other mathematical treatments, such as OSC and DOSC have been developed in order to minimize the variability unrelated to the response in spectral data. It has been confirmed that the pre-processing methods most usually applied (derivation, MSC, SNV) can reduce the variation unrelated to the modelled response, and that, in some cases, they lead to better calibration models compared to those obtained from the original spectra. However, in certain applications, it was shown that none of these pre-treatments can remove completely all the systematic variability. By contrast, OSC and DOSC have proven their relative effectiveness to correct coffee spectra, both for the quantification of ash content and for the determination of total lipids. These transformations removed at least a portion of information unrelated to the response, leading to significantly improved and simpler calibration models that only require a few components to model the data and providing predictive abilities much higher than other methods. Compared to models obtained with the original data and the data corrected by derivation, SNV and MSC, OSC and DOSC treatments gave PLS calibration models with improved prediction abilities (4.9 and 3.3% RMSEP with corrected data versus 7.1 and 8.3% RMSEP with

original data, respectively). Therefore, both OSC and DOSC could be used to develop reliable regression models based on filtered data for the specific applications of this work (Pizarro et al., 2004).

The relationship between some coffee roasting variables (weight loss, density and moisture) with near infrared (NIR) spectra of original raw (green) and differently roasted coffee samples was investigated in order to predict coffee roasting degree. Using PLS regression, a prediction of the three modelled roasting responses was performed. Most significant spectral differences were observed around 6600–7000, 5400–5900, 5300–5000, 4800–4500 and 4400–4100  $\text{cm}^{-1}$ . Such spectral variations were likely related to changes in chemical composition occurring during roasting, like the formation or degradation of certain compounds responsible for particular absorptions that strongly affect the final spectral profile. For instance, water band intensity at around 6896 (first overtone of O–H stretching) and 5154  $\text{cm}^{-1}$  (combination band of O–H stretching and O–H deformation) decreased gradually during roasting. Obtained data allowed constructing robust and reliable models for the prediction of roasting variables of unknown roasted coffee samples, since measured and predicted values showed high correlation coefficients ( $r$  from 0.92 to 0.98). Results provided by calibration models were comparable in terms of accuracy to the conventional analyses, revealing the feasibility for on-line or routine applications of NIR spectra regarding roasting process (Alessandrini et al., 2008).

Another study focused on four of the most important constituents of espresso and roasted coffee that are essential from a quality assurance standpoint: total acidity, caffeine content, chlorogenic acids and roasted bean colour. Similarly to the trend observed by Alessandrini et al. (2008), it was observed that water band intensity at 1450 (first overtone of O–H stretching) and at 1940 nm (combination band of O–H stretching and O–H deformation) decreased gradually when coffees were roasted darker, since during roasting water reduction is proportional to the degree of roasting. Other significant spectral differences appeared around 1100–1250, 1450–1700, 1780–1850, 1920–2000 and 2050–2200 nm, allowing the correlation of spectral variations with certain changes in chemical composition occurring during roasting, i.e., with the formation or degradation of certain compounds such as degradation of chlorogenic acids, changes in carbohydrates and amino acids. A comprehensive list of wavelength assignments for each attribute predicted is proposed by the authors. Thanks to the conjunction of pre-processing methods and multivariate calibration with NIR spectroscopy, separate regression models were developed and could be used for the prediction of the referred quality parameters (Esteban-Diez et al., 2004c).

The feasibility of applying variable selection techniques to extract a minimum number (maximum parsimony) of informative predictors from NIR spectra to measure in real time roasting colour and caffeine content was investigated to predict quality parameters of unknown coffee samples. Several variable selection techniques were tested including iterative predictor weighting (IPW) (Forina et al., 1999), interactive stepwise elimination (ISE) (Boggia, Forina, Fossa, & Mosti, 1997), and uninformative variable elimination (UVE) (Centner et al., 1996); and approaches based on ordinary least-squares (OLS) regression, such as genetic algorithms coupled with OLS regression (GA-OLS) (Leardi, Boggia, & Terrile, 1992; Lucasius & Kateman, 1991), and stepwise orthogonalization of predictors (SELECT) (Forina, Lanteri, Casale, & Cerrato-Olivero, 2007). Selected wavelengths related to O–H, C–H and N–H stretching second and first overtones mainly assigned to chlorogenic acids, carbohydrates, amino acids and water were observed for modelling roasting colour and caffeine content. The application of the stepwise orthogonalization of predictors provided notably improved regression models, with root-mean-square errors of the residuals in external prediction (RMSEP) equal to 3.68 and 1.46% for roasting colour and caffeine content, respectively. The use in industrial scale of the calibration models proposed has the potential of reducing analytical time, efforts, and costs of assessing these roasted coffee quality parameters, allowing for online monitoring of the coffee roasting

process and for controlling coffee quality and possible quality changes (Pizarro, Esteban-Diez, Gonzalez-Saiz, & Forina, 2007).

### 3.7. Coffee residues

Spent coffee grounds are a significant by-product which contains high levels of bioactive compounds, such as chlorogenic acid and flavonoids that have recognized antioxidant properties. Fourier-transform near infrared (FT-NIR) spectroscopy was used to determine the antioxidant capacity, total phenolic and total flavonoid contents of several spent coffee grounds samples. FT-NIR spectra were related to total flavonoid and total phenolic contents and the antioxidant capacity using PLS modelling. Results were promising, with an  $R^2$  for an independent test set of 0.93, 0.96, 0.95 and 0.95 for antioxidant capacity of spent coffee grounds, antioxidant capacity of spent coffee grounds ethanolic extracts, total flavonoid and total phenolic contents, respectively. It was confirmed that NIR spectroscopy has possible applications for routine assessment of these parameters and is a viable and advantageous alternative to chemical procedures involving laborious extractions (Páscoa et al., 2013).

On another study, the potential of near-infrared spectroscopy (NIR) and multivariate analysis to replace reference methods in the characterization of some constituents of coffee and banana residues (leaves and husks) was investigated. The evaluated parameters were Klason lignin, acid soluble lignin, total lignin, extractives, moisture, ash and acid insoluble residue contents of 102 coffee residues samples. The precision of the NIR methodology performed well compared to the reference method for almost all the parameters, except for moisture. The models for all the analytes exhibited  $R^2 > 0.80$ , with the exception of ash and acid soluble lignin content, which were predicted poorly ( $R^2 < 0.80$ ). Thus, NIR spectroscopy proved to be useful for chemical prediction of coffee residues, representing a faster and more economical alternative to the standard methodologies (Rambo, Amorim, & Ferreira, 2013).

## 4. Future trends

The results of previous research works presented in this review confirmed that IR spectral techniques are well suited for classification, authentication and prediction of essential properties in coffee samples.

Both FTIR and NIR spectroscopy have been successfully exploited. It can work in tandem or even replace standard methods in situations where classification, sorting, or identification of coffee varieties is required. NIRS calibrations allow performing coffee sample characterisation in a short time comparatively to classical methods to discriminate the coffee varieties, which are laborious and require lengthy sample preparation. Considering the high spectral resolution used in recent studies, deeper research is required to extract the useful information and reject the voluminous data that do not contribute to the application.

As a result, it is expected that the spectroscopic techniques will overcome other time-demanding methods in the near future for more complex applications in monitoring coffee quality. Also, the abundant chemical features laid in the spectra will enable implementing some applications such as authentication of superior coffee quality; detection of adulteration and defective grains and discrimination between different coffee qualities. The accurate classification of different quality grades will be very important for correct pricing and categorization of coffee which provide some economical benefits for producers by increasing consumer confidence in the supplied products.

The samples used in most of the studies reported were designed to cover as best as possible the large natural intrinsic variability of commercially available coffees by considering different blends, roasting conditions and geographical origins. However, transfer of the technique out of the laboratory will require its extension to commercial samples. The dynamic nature and particular needs of the coffee market would demand a continuous and more exhaustive collection of calibration samples that should be incessantly added to develop classification and



prediction models using NIR spectroscopic techniques (Downey & Boussson, 1996; Esteban-Díez et al., 2007). On the other hand, it must be emphasized that the application of NIR spectroscopy would not imply any additional effort regarding further analytical determinations in future samples to be characterised. The measurement of reference chemical parameters is only required for samples used to build the calibration dataset. Once calibration NIR spectra have been established, the pre-processing steps and prediction models obtained can be directly applied to future sample spectra without the need for measuring any reference values (Esteban-Díez et al., 2004a). In this way, from a practical application standpoint, the classification and prediction methodologies proposed only rely on NIR measurements.

IR spectroscopy techniques may find applications in the coffee industry for simple product inspection, prediction of sensory features or for quantification and prediction of sample composition. The non-destructive and multivariate characteristics of NIR spectroscopy techniques provide an interesting tool for quality monitoring changing the measurements of coffee quality from the laboratories to the processing lines. The use of FTIR and/or NIR spectroscopy for on/inline monitoring of quality in coffee industry is a challenge for the future.

## 5. Conclusion

This review has covered some of the recent applications of infrared spectroscopy in coffee quality assessment. In aid of greater regulatory inspection and consistent with the coffee industry's goal of providing superior product quality, both FTIR and NIR spectroscopy have shown the potential to enhance the knowledge and understanding of product parameters along the production process. The main advantage is that it is a chemical-free assessment method where sample preparation time is critically reduced or even eliminated, thus being a useful reference for traditional methods. The wide application of this system would seem to offer a number of potential advantages, including reduced labour costs, decreased errors during sample preparation or subjective judgement, and the creation of product data in a real time for documentation, traceability, and labelling. However, it is necessary for researchers to further transfer the knowledge from laboratory scale to industrial lines, so that the coffee industry can realistically benefit from the possibility of performing this non-destructive technique without additional laborious and time-consuming chemical analyses, enabling early sorting of produce and thereby improving quality management. Considering the recent advances in multivariate modelling and the analytical requirements of the most recent concepts of quality, it is anticipated that NIR techniques may progressively become a routine method for coffee quality and safety assessment and process monitoring and control.

## Acknowledgements

The authors gratefully acknowledge the financial support from the Coordination for the Improvement of Higher Education Personnel (CAPES) strategic research initiative under the Brazilian Ministry of Education, project number 23038.019085/2009-14.

## References

- Abbott, J. A., 1999. Quality measurement of fruits and vegetables. *Postharvest Biology and Technology*, 15(3), 207–225.
- Abdel-Nour, N., Ngadi, M., Prasher, S., & Karimi, Y., 2011. Prediction of egg freshness and albumen quality using visible/near infrared spectroscopy. *Food and Bioprocess Technology*, 4(5), 731–736.
- Alessandrini, L., Romani, S., Pinnavaia, G., & Rosa, M.D., 2008. Near infrared spectroscopy: An analytical tool to predict coffee roasting degree. *Analytica Chimica Acta*, 625, 95–102.
- Bauer, R., Nieuwoudt, H., Bauer, F. F., Kossmann, J., Koch, K. R., & Esbensen, K. H., 2008. FTIR spectroscopy for grape and wine analysis. *Analytical Chemistry*, 80, 1371e1379.
- Bicchi, C. P., Ombretta, M. P., Pellegrino, G., & Vanni, A.C., 1997. Characterization of roasted coffee and coffee beverages by solid phase microextraction-gas chromatography and principal component analysis. *Journal of Agricultural and Food Chemistry*, 45(1997), 4680–4686.
- Blanco, M., & Villarroya, I., 2002. *Trends in analytical chemistry*, 21(4), 240–250.
- Boggia, R., Forina, M., Fossa, P., & Mosti, L., 1997. Chemometrics study and validation strategies in the structure–activity relationships of a new class of cardiotoxic agents. *Quantitative Structure–Activity Relationships*, 16, 201–206.
- Briand, R., Kemsley, E. K., & Wilson, R. H., 1996. Approaches to adulteration detection in instant coffees using infrared spectroscopy and chemometrics. *Journal of the Science of Food and Agriculture*, 71, 359–366.
- Centner, V., Massart, D. L., de Noord, O. E., De Jong, S., Vandeginste, B.M., & Sterna, C., 1996. Elimination of uninformative variables for multivariate calibration. *Analytical Chemistry*, 68, 3851–3858.
- Chen, L., Xue, X., Ye, Z., Zhou, J., Chen, F., & Zhao, J., 2011. Determination of Chinese honey adulterated with high fructose corn syrup by near infrared spectroscopy. *Food Chemistry*, 128(4), 1110–1114.
- 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. *Analytica Chimica Acta*, 572, 77–84.
- Cozzolino, D., Smyth, H. E., & 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. *Journal of Agricultural and Food Chemistry*, 51, 7703–7708.
- Craig, A. P., Franca, A. S., & Oliveira, L. S., 2012a. Evaluation of the potential of FTIR and chemometrics for separation between defective and non-defective coffees. *Food Chemistry*, 132(2012), 1368–1374.
- Craig, A. P., Franca, A. S., & Oliveira, L. S., 2012b. Discrimination between defective and non-defective roasted coffees by diffuse reflectance infrared Fourier transform spectroscopy. *LWT – Food Science and Technology*, 47, 505–511.
- De Marchi, M., 2013. On-line prediction of beef quality traits using near infrared spectroscopy. *Meat Science*, 94(4), 455–460.
- Downey, G., & Boussson, J., 1996. Authentication of coffee bean variety by near-infrared reflectance spectroscopy of dried extract. *Journal of the Science of Food and Agriculture*, 71, 41–49.
- Downey, G., Briand, R., Wilson, R. H., & Kemsley, E. K., 1997. Near- and mid-infrared spectroscopies in food authentication: Coffee varietal identification. *Journal of Agricultural and Food Chemistry*, 45, 4357–4361.
- Downey, G., McIntyre, P., & 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. *Applied Spectroscopy*, 57, 158–163.
- Downey, G., Sheehan, E., Delahunty, C., O'Callaghan, D., Guinee, T., & Howard, V., 2005. June–September. Prediction of maturity and sensory attributes of Cheddar cheese using near-infrared spectroscopy. *International Dairy Journal*, 15(6–9), 701–709. <http://dx.doi.org/10.1016/j.idairyj.2004.06.013> (ISSN 0958–6946).
- Ebrahimi-Najafabadi, H., Leardi, R., Oliveri, P., Casolino, M. C., Jalali-Heravi, M., & Lanteri, S., 2012. Detection of addition of barley to coffee using near infrared spectroscopy and chemometric techniques. *Talanta*, 99, 175–179.
- Esteban-Díez, I., González-Sáiz, J. M., & Pizarro, C., 2004a. An evaluation of orthogonal signal correction methods for the characterisation of *arabica* and *robusta* coffee varieties by NIRS. *Analytica Chimica Acta*, 514, 57–67.
- Esteban-Díez, I., González-Sáiz, J. M., & Pizarro, C., 2004b. Prediction of roasting colour and other quality parameters of roasted coffee samples by near infrared spectroscopy. A feasibility study. *Journal of Near Infrared Spectroscopy*, 12(5), 287–297.
- Esteban-Díez, I., González-Sáiz, J. M., & Pizarro, C. I., 2004c. Prediction of sensory properties of espresso from roasted coffee samples by near-infrared spectroscopy. *Analytica Chimica Acta*, 525, 171–182.
- Esteban-Díez, I., González-Sáiz, J. M., Saenz-Gonzalez, J. M., & Pizarro, C. I., 2007. Coffee varietal differentiation based on near infrared spectroscopy. *Talanta*, 71, 221–229.
- Fagan, C. C., O'Donnell, C. P., O'Callaghan, D. J., Downey, G., Sheehan, E. M., Delahunty, C. M., et al., 2007. Application of mid-infrared spectroscopy to the prediction of maturity and sensory texture attributes of cheddar cheese. *Journal of Food Science*, 72, E130–E137.
- Farah, A., Monteiro, M. C., Calado, V., Franca, A. S., & Trugo, L. C., 2006. Correlation between cup quality and chemical attributes of Brazilian coffee. *Food Chemistry*, 98, 373–380.
- Forina, M., Casolino, C., & Pizarro, C., 1999. Iterative predictor weighting (IPW) PLS: a technique for the elimination of useless predictors in regression problems. *Journal of Chemometrics*, 13, 165–184.
- Forina, M., Lanteri, S., Casale, M., & Cerrato-Oliveros, M. C., 2007. Stepwise orthogonalization of predictors in classification and regression techniques: An “old” technique revisited. *Chemometrics and Intelligent Laboratory Systems*, 87, 252–261.
- Franca, A. S., Mendonça, J. C. F., & Oliveira, S. D., 2005. Composition of green and roasted coffees of different cup qualities. *LWT – Food Science and Technology*, 38, 709–715.
- Franca, A. S., Oliveira, L. S., Mendonça, J. C. F., & Silva, X. A., 2005. Physical and chemical attributes of defective crude and roasted coffee beans. *Food Chemistry*, 90, 89–94.
- Garrigues, J. M., Bouhsain, Z., Garrigues, S., & de la Guardia, M., 2000. Fourier transform infrared determination of caffeine in roasted coffee samples. *Journal of Analytical Chemistry*, 366(3), 319–322.
- Gonzalez, A. G., Pablos, F., Martin, M. J., Leon-Camacho, M., & Valdenebro, M. S., 2001. HPLC analysis of tocopherols and triglycerides in coffee and their use as authentication parameters. *Food Chemistry*, 73(2001), 93–101.
- He, J., Rodriguez-Saona, L. E., & Giusti, M. M., 2007. Mid infrared spectroscopy for juice authentication – Rapid differentiation of commercial juices. *Journal of Agricultural and Food Chemistry*, 55, 4443–4452.
- Huck, C. W., Guggenbichler, W., & Bonn, G. K., 2005. Analysis of caffeine, theobromine and theophylline in coffee by near infrared spectroscopy (NIRS) compared to high-performance liquid chromatography (HPLC) coupled to mass spectrometry. *Analytica Chimica Acta*, 538, 195–203.

- Huck, C. W., Maurer, R., Popp, M., Basener, N., & Bonn, G. K., 1999. Quantitative Fourier transform near infrared reflectance spectroscopy (NIRS) compared to high performance liquid chromatography (HPLC) of a flavone in *Flos Primulae veris* extractives. *Pharmaceutical and Pharmacological Letters*, 9, 26–29.
- Illy, A., & Viani, R., 1996. *Espresso coffee: The chemistry of quality*. London: Academic Press.
- 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. *Analytica Chimica Acta*, 571, 167–174.
- Isaksson, T., & Naes, T., 1988. The effect of multiplicative scatter correction (MSC) and linearity improvement in NIR spectroscopy. *Applied Spectroscopy*, 42, 1273–1284.
- ISO International Standard, 1978. *Green coffee – Determination of moisture (routine method)*. ISO 1447:1978.
- ISO International Standard, 1983. *Green coffee – Determination of loss in mass at 105 °C*. ISO 6673:1983.
- ISO International Standard, 2001. *Green coffee – Determination of moisture (basic reference method)*. ISO 1446:2001.
- Jing, M., Cai, W., & Shao, X., 2010. Quantitative determination of the components in corn and tobacco samples by using near-infrared spectroscopy and multiblock partial least squares. *Analytical Letters*, 43(12), 1910–1921.
- Kemsley, E. K., Ruault, S., & Wilson, R. H., 1995. Discrimination between *Coffea arabica* and *Coffea canephora* variant *robusta* beans using infrared spectroscopy. *Food Chemistry*, 54, 321–326.
- Leardi, R., Boggia, R., & Terrile, M., 1992. Genetic algorithms as a strategy for feature selection. *Journal of Chemometrics*, 6, 267–282.
- Liao, Y., Fan, Y., & Cheng, F., 2012. On-line prediction of pH values in fresh pork using visible/near-infrared spectroscopy with wavelet de-noising and variable selection methods. *Journal of Food Engineering*, 109(4), 668–675.
- Lucasius, C. B., & Kateman, G., 1991. Genetic algorithms for large-scale optimization in chemometrics: An application. *Trends in Analytical Chemistry*, 10, 254–261.
- Lyman, D. J., Benck, R., Dell, S., Merle, S., & Murray-Wijelath, J., 2003. FTIR-ATR analysis of brewed coffee: Effect of roasting conditions. *Journal of Agricultural and Food Chemistry*, 51, 3268–3272.
- Mendonça, J. C. F., Franca, A. S., & Oliveira, L. S., 2009. Physical characterization of non-defective and defective Arabica and Robusta coffees before and after roasting. *Journal of Food Engineering*, 92, 474–479.
- Mendonça, J. C. F., Franca, A. S., Oliveira, L. S., & Nunes, M., 2008. Chemical characterization of non-defective and defective green Arabica and Robusta coffees by Electrospray Ionization–Mass Spectrometry (ESI–MS). *Food Chemistry*, 111, 490–497.
- Morgano, M.A., Faria, C. G., Ferrão, M. F., Bragagnolo, N., & Ferreira, M. M. C., 2008. Determinação de umidade em café cru usando espectroscopia NIR e regressão multivariada [Determination of moisture in raw coffee by near infra-red reflectance spectroscopy and multivariate regression]. *Ciência e Tecnologia de Alimentos*, 28(1), 12–17.
- Myles, A. J., Zimmerman, T. A., & Brown, S. D., 2006. Transfer of multivariate classification models between laboratory and process near-infrared spectrometers for the discrimination of green Arabica and Robusta coffee beans. *Applied Spectroscopy*, 60(10), 246A–272A (1097–1216, 1198–1203(6)).
- Osborne, B. G., Fearn, T., & Hindle, P. H., 1993. *Practical NIR spectroscopy with applications in food and beverage analysis* (2nd ed.). Harlow, UK: Longman Scientific and Technical.
- Páscoa, R. N. M. J., Magalhães, L. M., & Lopes, J. A., 2013. FT-NIR spectroscopy as a tool for valorization of spent coffee grounds: Application to assessment of antioxidant properties. *Food Research International*, 51(2013), 579–586.
- Pasquini, C., 2003. Near Infrared Spectroscopy: Fundamentals, practical aspects and analytical applications. *Journal of Brazilian Chemical Society*, 14(2), 198–219.
- Pedro, A.M. K., & Ferreira, M. M. C., 2005. Nondestructive determination of solids and carotenoids in tomato products by near-infrared spectroscopy and multivariate calibration. *Analytical Chemistry*, 77, 2505–2511.
- Perrone, D., Donangelo, C. M., & Farah, A., 2008. Fast simultaneous analysis of caffeine, trigonelline, nicotinic acid and sucrose in coffee by liquid chromatography–mass spectrometry. *Food Chemistry*, 110, 1030–1035.
- Pizarro, C., Esteban-Diez, I., & 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. *Analytica Chimica Acta*, 585, 266–276.
- Pizarro, C., Esteban-Diez, I., Gonzalez-Saiz, J. -M., & Forina, M., 2007. Use of near-infrared spectroscopy and feature selection techniques for predicting the caffeine content and roasting color in roasted coffees. *Journal of Agricultural and Food Chemistry*, 55, 7477–7488.
- Pizarro, C., Esteban-Diez, I., Nistal, A. -J., & González-Sáiz, J. -M., 2004. Influence of data pre-processing on the quantitative determination of the ash content and lipids in roasted coffee by near infrared spectroscopy. *Analytica Chimica Acta*, 509, 217–227.
- Rambo, M. K. D., Amorim, E. P., & Ferreira, M. M. C., 2013. Potential of visible-near infrared spectroscopy combined with chemometrics for analysis of some constituents of coffee and banana residues. *Analytica Chimica Acta*, 775, 41–49.
- Reh, C. T., Gerber, A., Prodoliet, J., & Vuataz, G., 2006. Water content determination in green coffee – Method comparison to study specificity and accuracy. *Food Chemistry*, 96, 423–430.
- Ribeiro, J. S., Ferreira, M. M. C., & Salva, T. J. G., 2011. Chemometric models for the quantitative descriptive sensory analysis of Arabica coffee beverages using near infrared spectroscopy. *Talanta*, 83(2011), 1352–1358.
- Ribeiro, J. S., Salva, T. J., & Ferreira, M. M., 2010. Chemometric studies for quality control of processed Brazilian coffees using DRIFTS. *Journal of Food Quality*, 33, 212–227.
- Santos, K. M., Moura, M. F. V., Azevedo, F. G., Lima, K. M. G., Raimundo, I. M., Jr., & Pasquini, C., 2012. Classification of Brazilian coffee using near-infrared spectroscopy and multivariate calibration. *Analytical Letters*, 45(7), 774–781.
- Santos, J. R., Sarraguça, M. C., Rangel, A. O. S. S., & Lopes, J. A., 2012. Evaluation of green coffee beans quality using near infrared spectroscopy: A quantitative approach. *Food Chemistry*, 135, 1828–1835.
- Schulz, H., 2004. Application in analysis of coffee, tea, cocoa, tobacco, spices, medicinal and aromatic plants, and related products. In C. Roberts, J. Workman, & J. Reeves (Eds.), *Near-infrared spectroscopy in agriculture. Agronomy Monograph, No. 44*. (pp. 345–376). Madison, USA: American Society of Agronomy/Crop Science of America/Soil Science Society of America.
- 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. *Journal of Chemometrics*, 23, 32–41.
- Wang, J., Jun, S., Bittenbender, H. C., Gautz, L., & Li, Q. X., 2009. Fourier transform infrared spectroscopy for Kona coffee authentication. *Journal of Food Science*, 74(5), C385–C391.
- Wang, D., Min, S. -G., Duan, J., Xiong, Y. -M., & Li, Q. -Q., 2012. Simultaneous determination of instant coffee, plant fat and sugar content in liquid coffee beverage by diffuse reflectance near-infrared spectroscopy. *Guang Pu Xue Yu Guang Pu Fen Xi/Spectroscopy and Spectral Analysis*, 32(4), 982–984.
- Weyer, L. G., 1985. Near-infrared spectroscopy of organic substances. *Applied Spectroscopy Reviews*, 21, 1–43.
- Woodcock, T., Fagan, C. C., O'Donnell, C. P., & Downey, G., 2008. Application of near and mid-infrared spectroscopy to determine cheese quality and authenticity. *Food & Bioprocess Technology*, 1(2), 117–129. <http://dx.doi.org/10.1007/s11947-007-0033-y>.
- Xie, L., Ying, Y., Ying, T., Yu, H. And, & Fu, X., 2007. Discrimination of transgenic tomatoes based on visible/near-infrared spectra. *Analytica Chimica Acta*, 584, 379–384.
- Yan-De, L., Yi-Bin, Y., Xiaping, F. And, & Guisan, L., 2007. Experiments on predicting sugar content in apples by FT-NIR technique. *Journal of Food Engineering*, 80, 986–989.
- Zhang, X., Li, W., Yin, B., Chen, W., Kelly, D. P., Wang, X., et al., 2013. Improvement of near infrared spectroscopic (NIRS) analysis of caffeine in roasted Arabica coffee by variable selection method of stability competitive adaptive reweighted sampling (SCARS). *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 114, 350–356.