Control of the extractable content of bioactive compounds in coffee beans by near infrared hyperspectral imaging.

Julio Nogales-Bueno[a,b], Berta Baca-Bocanegra[a], Laura Romero-Molina[a], Alicia Martínez-López[a], Ana Elisa Rato[c], Francisco José Heredia[a], José Miguel Hernández-Hierro[a], María Luisa Escudero-Gilete[a]*, María Lourdes González-Miret[a].

Affiliations

a) *Food Colour and Quality Laboratory, Área de Nutrición y Bromatología, Facultad de Farmacia, Universidad de Sevilla, 41012, Sevilla, Spain*
b) *Department of Animal Production, University of Cordoba, Campus of Rabanales, Córdoba, 14071, Spain*
c) *MED-Mediterranean Institute for Agriculture, Environment and Development, Instituto de Investigação e Formação Avançada, Universidade de Évora, Pólo da Mitra, Ap. 94, 7006-554 Évora, Portugal;*

*** Corresponding author**: María Luisa Escudero-Gilete

Phone: +34 955420938

E-mail: gilete@us.es

Abstract

Control of coffee quality has a great importance for being one of the most important raw materials within the international trade. The extractable composition of coffee has been studied in recent decades and the use of non-destructive methodologies is being continuously promoted. In this study, near infrared hyperspectral imaging has been applied to develop non-destructive methods for the control of extractable contents of caffeine, chlorogenic acid, total phenolics and melanoidins in coffee beans.

Extractable contents and trends obtained among the different coffee types analysed are similar to those obtained previously in other studies. Moreover, modified partial least square (MPLS) regressions produced prediction models with standard errors of prediction in external validation of 12.01%, 15.61% and 17.61% for caffeine, chlorogenic acid and total phenolics, respectively. Therefore, results obtained for these three parameters indicate that NIR spectroscopy has a great potential for their control in coffee beans.

Keywords: coffee; near infrared hyperspectral imaging; caffeine; chlorogenic acid; total phenolics.

1. Introduction

Coffee is one of the most popular beverages being consumed in whole world. The coffee beans found on the market belong to two different species of the genus *Coffea*: *Coffea arabica* L. (*arabica* coffee) and *Coffea canephora* (Pierre) ex Frohner (*robusta* coffee). In the economic and public health areas, the control of coffee quality has a great importance as it is one of the most important raw materials within the international trade (Barbin, Felicio, Sun, Nixdorf, & Hirooka, 2014). Coffee beans composition and, more importantly, their extractable composition, i.e., the compounds that can pass into water in the infusion process, has been continuously monitored in recent decades (Borrelli, Visconti, Mennella, Anese, & Fogliano, 2002; Delgado-Andrade & Morales, 2005; Hečimović, Belščak-Cvitanović, Horžić, & Komes, 2011; Huck, Guggenbichler, & Bonn, 2005; Illy & Viani, 1995; Karpinska, Świsłocka, & Lewandowski, 2017).

Among these compounds, caffeine is perhaps the most popular. Differences in caffeine content among coffee samples of different varieties, seasons, roasting or grinding processes, etc. have been reported (Bell, Wetzel, & Grand, 1996). This alkaloid has well-known pharmacological effects on several body systems (central nervous system, heart, gastrointestinal system, renal system, respiratory system, peripheral and central vasculature) (Bessada, Alves, & Oliveira, 2018; Jahrami et al., 2020; Leonard, Watson, & Mohs, 1987). However, other properties of caffeine, such as stimulant effects, decreased sleepiness and increased attention, make coffee positively valued by consumers (Clarke & Macrae, 1988). In addition, other families present in coffee (e.g., phenolics, chlorogenic acids, or melanoidins) have been reported as compounds with recognized health benefits. Phenolic compounds and their main representatives in coffee, chlorogenic acids, have been widely reported for their antioxidant, anti-inflammatory and anticancer activities (Meng, Cao, Feng, Peng, & Hu, 2013; Sato et al., 2011; Upadhyay & Mohan Rao, 2013). In the same way, coffee melanoidins have demonstrated antioxidant and antimicrobial activities (Borrelli et al., 2002; Rufian-Henares & de la Cueva, 2009; Steinhart, Luger, & Piost, 2001).

Therefore, it is well-known the importance of the extractable composition of roasted coffee. The need to continuously improve existing methods and develop new ones to control this composition is clearly demonstrated. Apart from traditional methods based on chromatography, spectrophotometry, mass spectrometry, etc., several spectroscopic methods have been proposed for the control of caffeine in coffee beans (Huck et al., 2005; Pizarro, Esteban-Díez, González-Sáiz, & Forina, 2007; X. Zhang et al., 2013), even by the application of hyperspectral imaging (C. Zhang, Jiang, Liu, & He, 2017). However, only one study has been found in which chlorogenic acid is determined in coffee beans by spectroscopic methods (Liang, Lu, Hu, & Kitts, 2016) and none for phenolic compounds or melanoidins.

Although coffee bean size is not adequate for chemical imaging, hyperspectral imaging can provide some other advantages such as collecting spectral information of a high amount of coffee beans in a short period of time or implementation of on-line application in an easy way. Moreover, hyperspectral imaging is being successfully applied for the control of other quality parameters in coffee beans such as roast degree, variety or coffee brand (Bona et al., 2017; Chu, Yu, Zhao, & He, 2018; Nansen, Singh, Mian, Allison, & Simmons, 2016; C. Zhang, Liu, & He, 2018). Consequently, the main aim of this study is to apply near infrared hyperspectral imaging to the development of methods for the control of extractable contents of caffeine, chlorogenic acid, total phenolics and melanoidins in coffee beans. To the best of our knowledge, this is the first time that the aforementioned aim has been carried out.

2. Material and methods

2.1. *Samples*

A number of coffee brands present in the Spanish market were selected (Carrefour®, Marcilla®, Bonka®, Catunambú®, Eroski®, Supersol®, Hacendado®, Camelo Barco®, Día®). They were purchased twice, in November 2013 and April 2015. In order to correctly represent the Spanish coffee market, packages with differences in roasting type (natural, torrefacto or blend) and caffeine content (normal and decaffeinated) were purchased. Torrefacto is obtained by adding sugar to the coffee beans during the last step of the roasting process. A total of 144 samples

were collected from 36 different packages. Each sample contained approximately 7 g of coffee beans. Then, they were stored in sealed plastic bags in a dry place at room temperature until spectroscopic analysis.

2.2. *Near infrared hyperspectral imaging*

Spectroscopic analysis was carried out in a hyperspectral system. Samples were extracted from the plastic bags and individually placed on a polyethylene plastic tray. Coffee beans were evenly distributed over the tray to maximize their exposed surface. Then, the tray was placed under the hyperspectral device to image acquisition. Three images were acquired for each sample. Coffee beans were randomly repositioned before each image acquisition.

Hyperspectral images were acquired following the procedure described in detail in Hernández-Hierro, Nogales-Bueno, Rodríguez-Pulido, and Heredia (2013). Briefly, hyperspectral device (Infaimon S.L., Barcelona, Spain) comprised a Xenics® XEVA-USB InGaAs camera (Xenics Infrared Solutions, Inc., Leuven, Belgium) with a spatial resolution of 320 × 256 pixels. It covers the spectral range between 900 and 1700 nm with a spectral resolution of 3.25 nm thanks to the spectrograph Specim ImSpector N17E Enhanced (Spectral Imaging Ltd., Oulu, Finland). Images were calibrated by means of a two-point calibration. Then, regions of interest (ROIs) were identified by lineal discriminant model, prior developed, using the reflectance values from three wavelengths (1207, 1386 and 1501 nm). In that way, pixels into the images were classified as plastic tray pixels or coffee pixels and only the spectra of coffee pixels were saved. Due to reduced efficiency of the sensor in the extremes of its spectral range, only the effective wavelength 950–1650 nm regions were used in data analysis.

2.3. *Chemicals*

Methanol was supplied by J.T. Baker® (Phillipsburg, NJ, USA) and di-sodium hydrogen phosphate by Scharlab (Barcelona, Spain). Caffeine and chlorogenic acid reference standards were provided by Sigma-Aldrich® (Saint Louis, MO, USA), while gallic acid reference standard

was provided by Analytical Carlo Elba®. Folin–Ciocalteu reagent was supplied by Merk® (Darmstadt, Germany) and sodium carbonate by Panreac® (Barcelona, Spain).

2.4. *Reference parameters*

After hyperspectral image acquisition, each 7-gram coffee sample was ground in a IKA® A11 basic mill. Then coffee beverage was prepared in a domestic coffee maker with 70 mL of distilled water. The resulting drink was made up to 100 mL and an aliquot was kept for further analysis.

2.4.1. *Caffeine and chlorogenic acid*

Caffeine and chlorogenic acid contents in coffee beverages were determined by means of a chromatographic method. One millilitre of sample was made up to 50 mL. An aliquot of this solution was filtered (0.45 µm) into chromatographic vials and 20 µL were injected in duplicated. Chromatographic analyses were performed on a Hewlett-Packard 1200 series HPLC equipped with an autosampler, a quaternary HPLC pump, a column heater, a diode array detector (DAD), and a data treatment station. A Kromasil C18 column (4.6 mm × 150 mm, 5 µm particle size) thermostated at 25 °C was used. An isocratic flow of 1 mL/min was used with a unique solvent (0.025 mol/L di-sodium hydrogen phosphate, pH 3.0, buffer in 45% methanol). The preferred detection wavelengths were 270 nm for caffeine and 325 nm for chlorogenic acid. They both were identified and quantified according to their spectroscopic and chromatographic features by comparing with reference standards. Elution times were respectively for chlorogenic acid and caffeine 1.80 and 2.49 min. Results were expressed as mg of analyte per g of coffee beans. The standard error was generally around 10%, so the error and degree of accuracy of the reference method was considered appropriate to use these data as reference values.

2.4.2. *Total phenolics*

Total phenolics were determined following the Folin–Ciocalteu method (Singleton & Rossi, 1965). One millilitre of sample was made up to 50 mL. Then, 100 µL of this extract was mixed

142 with 1.0 mL of sodium carbonate (20 g/100 mL), 520 µL of Folin–Ciocalteu reagent and made
143 up to 10 mL with ultrapure water. After 2 hours at room temperature, Folin–Ciocalteu analysis
144 was performed on an Agilent 8453 UV–Vis spectrophotometer (Palo Alto, USA), equipped with
145 diode array detection (DAD), measuring absorbance at 765 nm. The extract volumes were
146 modified appropriately for the samples that needed it. For quantification, results were expressed
147 as mg of gallic acid equivalents per g of coffee beans.

148 2.4.3.*Melanoidins*

149 Melanoidins were determined following the method described in Pérez-Hernández, Chávez-
150 Quiroz, Medina-Juárez, and Gámez Meza (2012). A standard calibration curve was performed
151 at 420 nm, which is the wavelength absorbed by melanoidins. Melanoidins structure is
152 indeterminate yet and, therefore, there is not available a reference standard in the market. The
153 lack of this standard was supplied using an extract of roasted coffee as the source of
154 melanoidins. A torrefacto coffee was chosen because it was expected to be very rich in these
155 compounds. This coffee extract was considered as stock solution and it was diluted afterwards
156 10 times. After reading the absorbance of every dilution, the calibration curve was constructed
157 by plotting absorbance values as a function of the melanoidins concentrations. For that, a
158 specific extinction coefficient of 1.1289 $Lg^{-1}cm^{-1}$ was applied (Tagliazucchi, Verzelloni, &
159 Conte, 2010). Then, samples were diluted (1:50 for torrefacto and 1:25 for the remaining
160 samples) and melanoidins were spectrophotometrically determined at 420 nm. Results were
161 expressed as mg of melanoidins per g of coffee beans.

162 2.5. *Data analysis*

163 *2.5.1.Analysis of variance*

164 Significant differences between different types of coffee (natural, torrefacto, blend and
165 decaffeinated) and between the different sets of samples (calibration and validation) were
166 evaluated by one-way analysis of the variance (ANOVA). A Tukey *post hoc* test was run to
167 confirm where the differences occurred between groups. Extractable contents of caffeine,

chlorogenic acid, total phenolics and melanoidins were used as independent variables. ANOVAs were developed with Statistica v.8.0 software (StatSoft Inc., OK, USA, 2007).

2.5.2.Principal component analysis

Coffee spectra from each different package were randomly divided into calibration and validation sets. Calibration set was analysed by means of a principal component analysis (PCA). This analysis was applied in order to provide information about the latent structure of spectral matrix (spectral outliers, sample distribution and possible differences between sample classes). Moreover, this method is also an important source of knowledge with which to create cross-validation groups used in the calibration process (Brereton, 2003; Shenk & Westerhaus, 1995). PCA was developed with Win ISI® (v1.50) (Infrasoft International, LLC, Port. Matilda, PA, USA).

2.5.3.Modified partial least square regressions

After spectral outliers were identified and removed from the spectral matrix, reference parameters (extractable contents of caffeine, chlorogenic acid, total phenolics and melanoidins in coffee samples) were assigned to their respective spectrum. Modified partial least square (MPLS) regression were applied to the calibration set in order to obtain prediction models for controlling the reference parameters following the procedure described in detail in (Nogales-Bueno, Hernández-Hierro, Rodríguez-Pulido, & Heredia, 2014). Different spectral pre-treatments were tested in order to remove or reduce scattering effects produced for the different textures and sample sizes (Dhanoa, Lister, & Barnes, 1995; Geladi, MacDougall, & Martens, 1985). The best model for each reference parameter was saved. Pre-treatments were multiplicative scattering correction (MSC), detrend, standard normal variate (SNV), different derivatives and none pre-treatments. Chemical outliers were identified and removed by the evaluation of the T-statistic and setting the threshold in 2.5 units. To do so, residual error was obtained by comparison between the predicted values and the chemical

193 values. Finally, the standard error of cross-validation (SECV) was evaluated and
194 expressed as percentage.
195 Then, the goodness of each MPLS model was tested. To this end, models were applied
196 to the samples allocated into the validation set and the results were compared to the
197 chemical values previously determined. In this way, a standard error of prediction (SEP)
198 in external validation was obtained for each reference parameter. MPLS models were
199 developed and tested with Win ISI® (v1.50) (Infrasoft International, LLC, Port. Matilda,
200 PA, USA).

201 **3. Results and discussion**

202 *3.1. Extractable contents of caffeine, chlorogenic acid, total phenolics and melanoidins in*
203 *coffee samples*

204 Table 1 shows mean and standard error of the mean for the reference parameters measured. Data
205 for all samples, different types of samples (natural, torrefacto, blend and decaffeinated) and
206 different sets of samples (calibration and validation) are shown. The results obtained for the
207 whole group of samples are similar to those described in previous studies (Belguidoum, Amira-
208 Guebailia, Boulmokh, & Houache, 2014; Lopes et al., 2016; Lopez-Galilea, de Pena, & Cid,
209 2008; Ludwig, Bravo, De Peña, & Cid, 2013).
210 Furthermore, when different types of coffee are taken into account, the ANOVA result
211 obviously shows that decaffeinated coffee has significantly lower caffeine values than other
212 types. However, it also shows other interesting results: torrefacto coffee has a significantly
213 lower amount of chlorogenic acid and significantly higher amounts of total phenolics and
214 melanoidins than natural coffee. These results were also reported (López-Galilea, Andueza,
215 Leonardo, Paz de Peña, & Cid, 2006; Lopez-Galilea et al., 2008; Ludwig et al., 2013). Although
216 torrefacto coffee is often considered a poor quality coffee, these results indicate that, overall, it
217 has better health qualities, such as increased its potential antioxidant activity.

When looking for significant differences between the calibration and validation sets, these two groups were found to be homogeneous. This is good evidence that the random selection procedure performed generates two new sets of samples that correctly represent all data variability.

3.2. Spectral information

The spectra of coffee samples in absorbance units were extracted from the hyperspectral images. The mean spectra of each type of coffee are shown in Figure 1a. It can be seen that the shape of the spectrum is very similar for the four types of coffee. The most remarkable feature is that the entire torrefacto coffee spectrum shows a higher absorbance. This characteristic may be due to the roasted sugar coating this coffee, which may alter the way these beans scatter light. Calibration and validation sets were randomly created by selecting, respectively, 75% and 25% of the acquired spectra. Initially, the structure of the calibration set was evaluated by a PCA. Prior to the development of this unsupervised pattern recognition method, a SNV (2,5,5,1) pre-treatment was applied to the calibration spectra. The numbers in parentheses indicate the number of the derivative, the gap over which the derivative is calculated and two different smooths, respectively. The spectral variability explained was 97% using 12 principal components. Mahalanobis distances (H) from the mean spectrum of the entire sample set were calculated in this hyperspace. Samples were ranked in order of their H values and the $H > 3$ criterion was applied. Two decaffeinated samples were identified as H-outliers and they were removed from the calibration set. Figure 1b shows the projection of the samples on the plane defined by the first and second principal component. It is possible to observe a partial separation between samples, particularly between torrefacto samples and the rest when the samples are represented according to the type of coffee.

3.3. *MPLS regression methods*

Raw spectra allocated into the calibration set were used to develop MPLS regression methods for the prediction of the reference parameters. Different pre-treatments were applied and the

model that obtained the best result for each parameter was saved. Table 2 shows most important data related to these models: The pre-treatment applied, the number of terms or factors created by the MPLS algorithm, the number of samples retained and their mean and standard deviation and, finally, a number of statistical descriptors. Among these descriptors, the coefficient of determination (RSQ) and the ratio of performance to deviation (RPD) indicate how good the fit between chemical and spectral data is. The RSQ values obtained were good for caffeine and chlorogenic acid and acceptable for total phenolics. However, the RSQ value obtained for melanoidins indicated that the fit between this parameter and the NIR spectrum of the samples was not correct. This lack of fit can be due to the high difference between the melanoidins extracted in torrefacto and the remaining samples (Table 1). These melanoidins values were not homogeneous enough for the correct development of PLS regressions. These results are confirmed by the RPD values. Only for caffeine and chlorogenic acid, the obtained RPD suggest, respectively, a good and an acceptable fit of the model.

Furthermore, the goodness of the MPLS models were also assessed according to the SECV values. These calibration errors confirmed the problems for predicting melanoidins. However, the errors for caffeine, chlorogenic acid and total phenolics were quite good. This fact was reinforced when models were validated using the validation set of samples. SEP errors in external validation were obtained. When expressed as percentages with respect to the mean, SEP ranged from 12.01 to 17.61%. Taking into account the Bias and Slope of the models, the more promising values are those obtained for caffeine and chorogenic acid. Therefore, results obtained for these parameters indicate that NIR spectroscopy has a great potential for controlling caffeine and chlorogenic acid in coffee beans. The errors obtained for total phenolics are also interesting, although its low values of RPD and Slope in external validation indicate that more samples might be needed for obtaining a better fit. Similar errors were found for caffeine in (Huck et al., 2005; Pizarro et al., 2007; C. Zhang et al., 2017; X. Zhang et al., 2013) and for chlorogenic acid in (Liang et al., 2016). No similar studies have been found in coffee for the control of total phenolics, although they have been found in other matrices with errors of the same order of magnitude (Baca-Bocanegra et al., 2019; Baca-Bocanegra, Nogales-Bueno,

Heredia, & Hernández-Hierro, 2018; Baca-Bocanegra, Nogales-Bueno, Hernández-Hierro, & Heredia, 2018; Nogales-Bueno et al., 2014; Schulz, Engelhardt, Wegent, Drews, & Lapczynski, 1999).

PLS factors of the developed models were extracted and studied (Figure 2). It can be seen that the region around 1400 nm has a great importance in the prediction of caffeine, chlorogenic acid and total phenolics. Moreover, the region around 1200 nm also shows important contributions to the model loadings, mainly for the control of chlorogenic acid and total phenolics. These regions have been linked to combination bands of the –OH functional group, and also to C−H aromatic second overtones and C−H third overtones (Barbin et al., 2014; Hernández-Hierro et al., 2013; Siesler, Ozaky, Kawata, & Heise, 2002). These vibrational modes can be attributed to the predicted components and also to components linked to their extraction from the coffee.

Therefore, results obtained are promising. Developed methods allow predicting caffeine, chlorogenic acid and total phenolics in coffee beans in a non-polluting and non-destructive way and with acceptable prediction errors. The standard errors for the reference methods applied are generally around 5-10%. In consequence, the developed spectroscopic methods have to be considered preliminary. They need to be implemented with more data coming from other coffee brands, types, dates etc.

4. Conclusion

In this study, the extractable content of caffeine, chlorogenic acid, total phenolics and melanoidins has been determined from a series of samples of natural, torrefacto, blend and decaffeinated coffee beans. Trends obtained among the different coffee types analysed are similar to those obtained previously in other studies: overall, torrefacto coffee shows significantly higher amounts of compounds that are traditionally related to health profits. This result increases the interest in having subjected to additional studies this type of coffee, typical of the Spanish market.

The analysis of the spectral information extracted from the different types of coffee has shown some trends among the different samples. Mainly, torrefacto samples have slightly different

spectral characteristics than other types of coffee. This is surely due to the differences produced on the surface of the torrefacto coffee beans by their characteristic roasting procedure.

When spectral and chemical data are jointly employed for the development of prediction methods in coffee beans, interesting results were obtained. These methods showed a good fit for the prediction of caffeine and chlorogenic acid and moderate for total phenolics. The errors obtained indicate that NIR hyperspectral imaging has a great potential for the prediction of these parameters. However, additional studies involving more coffee brands, types, locations and dates should be added to these models to develop more reliable and robust methods and to improve the results.

5. Abbreviations used

ANOVA, analysis of variance; H, Mahalanobis distance; MPLS, modified partial least square; MSC, multiplicative scattering correction; NIR, near infrared; PC, principal component; PCA, principal component analysis; ROI, regions of interest; RSQ, coefficient of determination; SECV, standard error of cross validation; SEP, standard error of prediction; SNV, standard normal variate.

6. Acknowledgements

The authors thank the technical staff of Biology Service (Servicios Generales de Investigación (SGI), Universidad de Sevilla).

7. Fundings

This study was supported by Universidad de Sevilla [grant numbers VPPI-II.2, VPPI-II.4], by FEDER and National funds through the Programa Operacional Regional ALENTEJO 2020 [ALT20-03-0145- FEDER-000005 – Eficiencia da tecnologia NIR para avaliação da maturação e qualidade de frutos] and the Spanish MINECO [Juan de la Cierva contract FJC2018-037967-I].

8. References

Baca-Bocanegra, B., Nogales-Bueno, J., García-Estévez, I., Escribano-Bailón, M. T., Hernández-Hierro, J. M., & Heredia, F. J. (2019). Screening of Wine Extractable Total Phenolic and Ellagitannin Contents in Revalorized Cooperage By-products: Evaluation by Micro-NIRS Technology. *Food and Bioprocess Technology*, *12*(3), 477-485. doi: 10.1007/s11947-018-2225-z

Baca-Bocanegra, B., Nogales-Bueno, J., Heredia, F. J., & Hernández-Hierro, J. M. (2018). Estimation of Total Phenols, Flavanols and Extractability of Phenolic Compounds in Grape Seeds Using Vibrational Spectroscopy and Chemometric Tools. *Sensors*, *18*(8), 2426.

Baca-Bocanegra, B., Nogales-Bueno, J., Hernández-Hierro, J. M., & Heredia, F. J. (2018). Evaluation of extractable polyphenols released to wine from cooperage byproduct by near infrared hyperspectral imaging. *Food Chemistry*, *244*, 206-212. doi: https://doi.org/10.1016/j.foodchem.2017.10.027

Barbin, D. F., Felicio, A. L. d. S. M., Sun, D.-W., Nixdorf, S. L., & Hirooka, E. Y. (2014). Application of infrared spectral techniques on quality and compositional attributes of coffee: An overview. *Food research international*, *61*, 23-32.

Belguidoum, K., Amira-Guebailia, H., Boulmokh, Y., & Houache, O. (2014). HPLC coupled to UV–vis detection for quantitative determination of phenolic compounds and caffeine in different brands of coffee in the Algerian market. *Journal of the Taiwan Institute of Chemical Engineers*, *45*(4), 1314-1320. doi: https://doi.org/10.1016/j.jtice.2014.03.014

Bell, L. N., Wetzel, C. R., & Grand, A. N. (1996). Caffeine content in coffee as influenced by grinding and brewing techniques. *Food Research International*, *29*(8), 785-789. doi: https://doi.org/10.1016/S0963-9969(97)00002-1

Bessada, S. M. F., Alves, R. C., & Oliveira, M. B. P. P. (2018). Caffeine-based food supplements and beverages: Trends of consumption for performance purposes and

349 safety concerns. *Food Research International, 109*, 310-319. doi:
350 https://doi.org/10.1016/j.foodres.2018.04.050
351 Bona, E., Marquetti, I., Link, J. V., Makimori, G. Y. F., da Costa Arca, V., Guimarães Lemes,
352 A. L., . . . Poppi, R. J. (2017). Support vector machines in tandem with infrared
353 spectroscopy for geographical classification of green arabica coffee. *LWT - Food*
354 *Science and Technology, 76*, 330-336. doi: https://doi.org/10.1016/j.lwt.2016.04.048
355 Borrelli, R. C., Visconti, A., Mennella, C., Anese, M., & Fogliano, V. (2002). Chemical
356 characterization and antioxidant properties of coffee melanoidins. *J Agric Food Chem*,
357 *50*(22), 6527-6533.
358 Brereton, R. G. (2003). *Chemometrics : data analysis for the laboratory and chemical plant*.
359 Chichester, West Sussex, England: J. Wiley.
360 Chu, B., Yu, K., Zhao, Y., & He, Y. (2018). Development of Noninvasive Classification
361 Methods for Different Roasting Degrees of Coffee Beans Using Hyperspectral Imaging.
362 *Sensors, 18*(4). doi: 10.3390/s18041259
363 Clarke, R. J., & Macrae, R. (1988). *Coffee: physiology* (Vol. 3): Springer Science & Business
364 Media.
365 Delgado-Andrade, C., & Morales, F. J. (2005). Unraveling the contribution of melanoidins to
366 the antioxidant activity of coffee brews. *Journal of Agricultural and Food Chemistry*,
367 *53*(5), 1403-1407.
368 Dhanoa, M. S., Lister, S. J., & Barnes, R. J. (1995). On the Scales Associated with Near-
369 Infrared Reflectance Difference Spectra. *Applied Spectroscopy*, *49*(6), 765-772.
370 Geladi, P., MacDougall, D., & Martens, H. (1985). Linearization and Scatter-Correction for
371 Near-Infrared Reflectance Spectra of Meat. *Applied Spectroscopy*, *39*(3), 491-500.
372 Hečimović, I., Belščak-Cvitanović, A., Horžić, D., & Komes, D. (2011). Comparative study of
373 polyphenols and caffeine in different coffee varieties affected by the degree of roasting.
374 *Food Chemistry*, *129*(3), 991-1000. doi:
375 https://doi.org/10.1016/j.foodchem.2011.05.059

Hernández-Hierro, J. M., Nogales-Bueno, J., Rodríguez-Pulido, F. J., & Heredia, F. J. (2013). Feasibility study on the use of near-infrared hyperspectral imaging for the screening of anthocyanins in intact grapes during ripening. *Journal of Agricultural and Food Chemistry, 61*(41), 9804-9809.

Huck, C., Guggenbichler, W., & Bonn, G. (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*(1), 195-203.

Illy, A., & Viani, R. (1995). *Espresso Coffee: The Chemistry of Quality*: Academic Press.

Jahrami, H., Al-Mutarid, M., Penson, P. E., Al-Islam Faris, M. e., Saif, Z., & Hammad, L. (2020). Intake of Caffeine and Its Association with Physical and Mental Health Status among University Students in Bahrain. *Foods, 9*(4), 473.

Karpinska, J., Świsłocka, R., & Lewandowski, W. (2017). A mystery of a cup of coffee; an insight look by chemist. *BioFactors*.

Leonard, T. K., Watson, R. R., & Mohs, M. E. (1987). The effects of caffeine on various body systems: a review. *Journal of the American Dietetic Association, 87*(8), 1048-1053.

Liang, N., Lu, X., Hu, Y., & Kitts, D. D. (2016). Application of Attenuated Total Reflectance-Fourier Transformed Infrared (ATR-FTIR) Spectroscopy To Determine the Chlorogenic Acid Isomer Profile and Antioxidant Capacity of Coffee Beans. *Journal of Agricultural and Food Chemistry, 64*(3), 681-689. doi: 10.1021/acs.jafc.5b05682

Lopes, G. R., Ferreira, A. S., Pinto, M., Passos, C. P., Coelho, E., Rodrigues, C., . . . Coimbra, M. A. (2016). Carbohydrate content, dietary fibre and melanoidins: Composition of espresso from single-dose coffee capsules. *Food Research International, 89*, 989-996. doi: https://doi.org/10.1016/j.foodres.2016.01.018

López-Galilea, I., Andueza, S., Leonardo, I. d., Paz de Peña, M., & Cid, C. (2006). Influence of torrefacto roast on antioxidant and pro-oxidant activity of coffee. *Food Chemistry, 94*(1), 75-80. doi: https://doi.org/10.1016/j.foodchem.2004.10.052

403 Lopez-Galilea, I., de Pena, M. P., & Cid, C. (2008). Application of multivariate analysis to
404 investigate potential antioxidants in conventional and torrefacto roasted coffee.
405 *European Food Research and Technology, 227*(1), 141-149. doi: 10.1007/s00217-007-
406 0703-z
407 Ludwig, I. A., Bravo, J., De Peña, M. P., & Cid, C. (2013). Effect of sugar addition (torrefacto)
408 during roasting process on antioxidant capacity and phenolics of coffee. *LWT - Food*
409 *Science and Technology, 51*(2), 553-559. doi: https://doi.org/10.1016/j.lwt.2012.12.010
410 Meng, S., Cao, J., Feng, Q., Peng, J., & Hu, Y. (2013). Roles of chlorogenic acid on regulating
411 glucose and lipids metabolism: a review. *Evidence-Based Complementary and*
412 *Alternative Medicine, 2013*.
413 Nansen, C., Singh, K., Mian, A., Allison, B. J., & Simmons, C. W. (2016). Using hyperspectral
414 imaging to characterize consistency of coffee brands and their respective roasting
415 classes. *Journal of Food Engineering, 190*, 34-39. doi: 10.1016/j.jfoodeng.2016.06.010
416 Nogales-Bueno, J., Hernández-Hierro, J. M., Rodríguez-Pulido, F. J., & Heredia, F. J. (2014).
417 Determination of technological maturity of grapes and total phenolic compounds of
418 grape skins in red and white cultivars during ripening by near infrared hyperspectral
419 image: A preliminary approach. *Food Chemistry, 152*, 586-591.
420 Pérez-Hernández, L. M., Chávez-Quiroz, K., Medina-Juárez, L. Á., & Gámez Meza, N. (2012).
421 Phenolic Characterization, Melanoidins, and Antioxidant Activity of Some Commercial
422 Coffees from Coffea arabica and Coffea canephora. *Journal of the Mexican Chemical*
423 *Society, 56*, 430-435.
424 Pizarro, C., Esteban-Díez, I., González-Sáiz, J.-M., & Forina, M. (2007). Use of near-infrared
425 spectroscopy and feature selection techniques for predicting the caffeine content and
426 roasting color in roasted coffees. *Journal of Agricultural and Food Chemistry, 55*(18),
427 7477-7488.
428 Rufian-Henares, J. A., & de la Cueva, S. P. (2009). Antimicrobial Activity of Coffee
429 Melanoidin: A Study of Their Metal-Chelating Properties. *Journal of Agricultural and*
430 *Food Chemistry, 57*(2), 432-438.

Sato, Y., Itagaki, S., Kurokawa, T., Ogura, J., Kobayashi, M., Hirano, T., . . . Iseki, K. (2011). In vitro and in vivo antioxidant properties of chlorogenic acid and caffeic acid. *International Journal of Pharmaceutics, 403*(1), 136-138.

Schulz, H., Engelhardt, U. H., Wegent, A., Drews, H. H., & Lapczynski, S. (1999). Application of near-infrared reflectance spectroscopy to the simultaneous prediction of alkaloids and phenolic substances in green tea leaves. *Journal of Agricultural and Food Chemistry, 47*(12), 5064-5067.

Shenk, J. S., & Westerhaus, M. O. (1995). *Routine Operation,Calibration, Development and Network System Management Manual*. Silver Spring, Maryland: NIRSystems.

Siesler, H. W., Ozaky, Y., Kawata, S., & Heise, H. M. (2002). *Near infrared spectroscopy: principles, instruments, applications*. Weinheim, Germany: Wiley-VCH.

Singleton, V. L., & Rossi, J. A. (1965). Colorimetry of Total Phenolics with Phosphomolybdic-Phosphotungstic Acid Reagents. *American Journal of Enology and Viticulture*(16), 144-158.

Steinhart, H., Luger, A., & Piost, J. (2001). *Antioxidative effect of coffee melanoidins.* Paper presented at the Proceedings of the 19th Colloquium ASIC, Trieste.

Tagliazucchi, D., Verzelloni, E., & Conte, A. (2010). Effect of dietary melanoidins on lipid peroxidation during simulated gastric digestion: their possible role in the prevention of oxidative damage. *J Agric Food Chem, 58*(4), 2513-2519. doi: 10.1021/jf903701h

Upadhyay, R., & Mohan Rao, L. J. (2013). An outlook on chlorogenic acids—occurrence, chemistry, technology, and biological activities. *Critical Reviews in Food Science and Nutrition, 53*(9), 968-984.

Zhang, C., Jiang, H., Liu, F., & He, Y. (2017). Application of Near-Infrared Hyperspectral Imaging with Variable Selection Methods to Determine and Visualize Caffeine Content of Coffee Beans. *Food and Bioprocess Technology, 10*(1), 213-221. doi: 10.1007/s11947-016-1809-8

Zhang, C., Liu, F., & He, Y. (2018). Identification of coffee bean varieties using hyperspectral imaging: influence of preprocessing methods and pixel-wise spectra analysis. *Scientific Reports, 8*. doi: 10.1038/s41598-018-20270-y

Zhang, X., Li, W., Yin, B., Chen, W., Kelly, D. P., Wang, X., . . . Du, Y. (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.

Figure captions

Figure 1: a) Average NIR spectra of natural, torrefacto, decaffeinated and blend coffees. b) Score plot of the first two principal components after PCA performed on NIR spectra recorded from coffee beans. The individual data points have been colour coded according to coffee type.

Figure 2: Loading plots of the MPLS models for caffeine (a), chlorogenic acid (b) and total phenolics (c).

Table 1

Table 1. Extractable contents of caffeine, chlorogenic acid, total phenolics and melanoidins.

Reference parameter	All samples (N=144)	Coffee type (N[a])				Set (N[a])	
		Natural (64)	Torrefacto (36)	Blend (16)	Decaffeinated (28)	Calibration (108)	Validation (36)
Caffeine	14.71 ± 0.58	17.90 ± 0.22[a]	18.29 ± 0.36[a]	17.92 ± 0.65[a]	0.96 ± 0.03[b]	14.71 ± 0.68[A]	14.70 ± 1.18[A]
Chlorogenic acid	7.23 ± 0.14	7.37 ± 0.22[a]	6.46 ± 0.31[b]	8.42 ± 0.37[a]	7.25 ± 0.19[ab]	7.24 ± 0.17[A]	7.20 ± 0.27[A]
Total phenolics	35.93 ± 0.59	34.50 ± 0.77[a]	38.63 ± 0.99[b]	36.53 ± 2.65[ab]	35.42 ± 1.39[ab]	35.91 ± 0.70[A]	36.02 ± 1.07[A]
Melanoidins	83.58± 2.90	70.26 ± 2.08[a]	115.54 ± 8.43[b]	79.14 ± 4.70[a]	75.49 ± 3.68[a]	83.47 ± 3.20[A]	83.94 ± 6.59[A]

[a]N: Number of sample of the correspondent group; Results are expressed as mg of analyte per g of coffee bean. Means ± standard errors of means are presented for different sample groups. For each reference parameter, different letters in the same row indicate statistical differences (Tukey test, α=0.05). Lowercase and uppercase letters correspond to different ANOVAs.

Table 2. Main statistical descriptors for the MPLS models developed for coffee beans in the NIR zone close to 950-1650 nm for extractable caffeine, chlorogenic acid, total phenolics and melanoidins.

Reference Parameters	Spectral pretreatments	PLS factors	N[a]	Mean	SD[b]	SEC[c]	RSQ[d]	SECV[e]	RPD[f]	$Bias_c$[g]	$Slope_c$[h]	SEP[i]	SEP(%)[i]	$Bias_v$[g]	$Slope_v$[h]
Caffeine[j]	Detrend 1,5,5,1	9	103	14.95	6.91	1.10	0.97	1.29	5.36	-0.05	0.99	1.80	12.01	-0.30	0.99
Chlorogenic acid[k]	SNV 1,5,5,1	5	101	7.18	1.71	0.74	0.81	0.91	1.88	-0.04	0.95	1.12	15.61	-0.19	0.75
Total phenolics[l]	SNV + detrend 2,5,5,1	7	101	35.67	6.06	3.94	0.58	4.63	1.31	0.06	0.96	6.28	17.61	0.03	0.51
Melanoidins[m]	MSC 2,15,15,1	4	100	77.93	24.90	20.81	0.30	21.49	1.16	-0.47	0.92	25.20	32.33	-3.79	0.75

[a]N: number of samples (calibration set); [b]SD: standard deviation; [c]SEC: standard error of calibration; [d]RSQ: coefficient of determination (calibration set); [e]SECV: standard error of cross-validation; [f]RPD: ratio of performance to deviation; [g]Bias: bias of the regression in calibration (c) and external validation (v); [h]Slope: slope of the regression in calibration (c) and external validation (v); [i]SEP: standard error of prediction in the external validation (also expressed as percentages with respect to the mean); [j]Caffeine: expressed in mg per g of coffee bean; [k]Chlorogenic acid: expressed in mg per g of coffee bean; [l]Total phenolics: expressed as mg of gallic acid equivalents per g of coffee beans; [m]Melanoidins: expressed as mg per g of coffee beans.

Figure 1

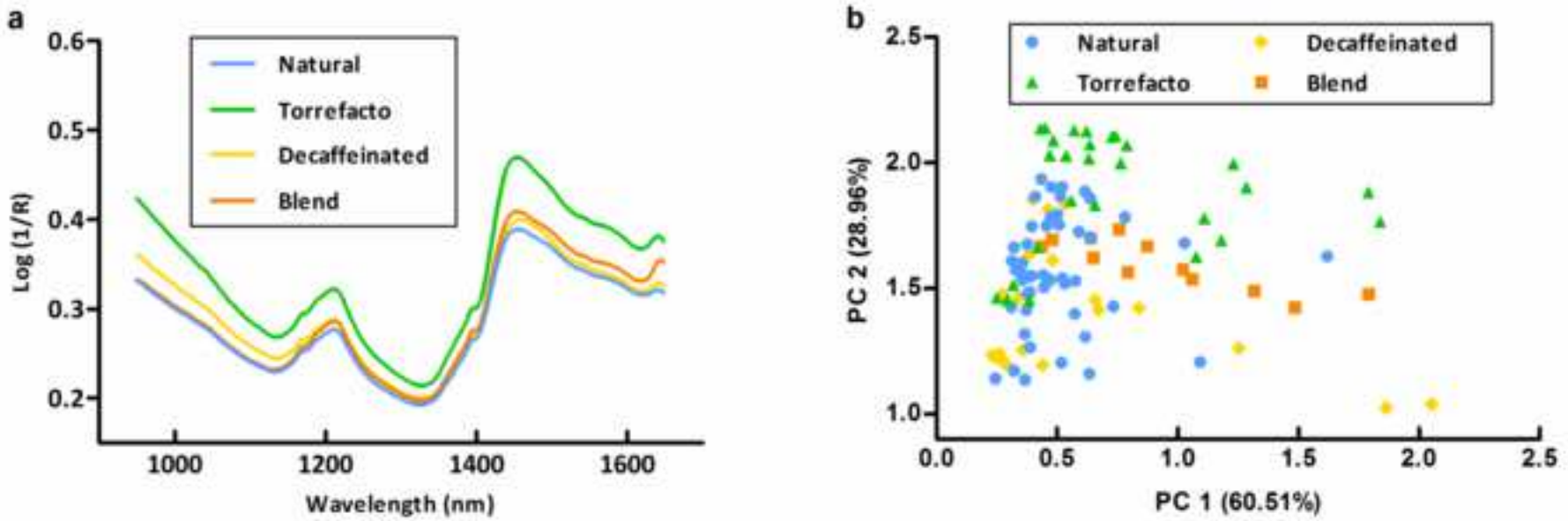

Figure 2

Click here to access/download;Figure;Nogales-Bueno_Cafe_Figure2.tif