

# Mid-infrared spectroscopy and chemometric tools applied to quantification of adulteration in Arabica coffee by addition of roasted barley

W. Terouzi, A. Oussama

**Abstract**— Attenuated Total Reflectance-Fourier Transform Mid Infrared Spectroscopy (ATR-FTMIR), combined with chemometric methods, has been used to quantify the roasted barley content in a binary mixture with Arabica coffee samples, as a means for controlling and avoiding coffee adulteration, which is a very important issue taking into account the great variability of the final sale price depending on coffee varietal origin. Blends of Arabica coffee with different percentages of roasted barley were measured using ATR-FTIR spectroscopy. Spectral and reference data were firstly analyzed by principal component analysis (PCA). Partial least square regression (PLSR) was used to establish calibration model. Excellent correlation between ATR-FTIR analysis and studied coffee blends was obtained  $R^2 = 0.99$ ; with Root Mean Square Errors of Prediction  $< 0.34$ , Limit of Detection 1.012%, and Relative Prediction Errors as low as 2.054. This result demonstrated the feasibility of ATR-FTIR spectroscopy combined with chemometrics to quantify successfully binary mixtures of Arabica coffee in the 0–45 % weight ratio range of roasted barley with a reliable, rapid and inexpensive tool without the need for sample preparation.

**IndexTerms**—Adulteration, Arabica coffee, chemometrics, infrared spectroscopy

## I. INTRODUCTION

Food authenticity is a major issue worldwide. It has been the target of government authorities, and presents a huge importance for consumers, food processors, and industries, in order to satisfy food quality and safety requirements [1], [2]. Coffee is one of the main food products commercialized in the world. Assurance of quality of roasted coffees has attracted widespread attention for controlling and preventing coffee adulteration, also given the great difference in the final sale price [3]. The most common coffee frauds include addition of roasted and ground vegetable products of lower commercial value, as maize, barley, rye, wheat, etc. [4], [5]. A simple visual inspection is not an appropriate method for differentiating between the genuine coffee samples and the fraudulent ones. As a consequence, many analytical methods using different techniques have been developed for coffee analysis, including chromatographic analysis [6], UV-Vis

spectroscopy [7] and nuclear magnetic resonance [8]. But, these techniques require more time to prepare samples, have high costs, and generate too much residues. Recently, several studies have described the effective application of near infrared spectroscopy coupled with chemometrics to address the problem of coffee authentication. This technique is often used to discriminate Arabica coffee by geographical and genotype origin [9] or blends of Arabica and Robusta grains [10], [11], [12] and to detect adulteration or defects in coffees [13], [14]. However, the authenticity of coffee by MIR spectroscopy combined with chemometrics has not been reported so far, even though mid-infrared is a region used for quantitative and qualitative analysis of several products.

The current work presents an application of ATR-FTMIR spectroscopy coupled with chemometric methods for quantification analysis of the fraudulent addition of roasted barley in Arabica coffee. This application was considered to develop improved and reliable regression model (PLSR) which could later be used as a quick and accurate analysis tool for quantifying the actual percentage of roasted barley in the binary blend with Arabica coffee.

## II. MATERIALS AND METHODS

### A. Samples preparation

In this study, to prepare the adulterated coffee samples we used:

- One Kilogram of Pure roasted coffee beans of Arabica variety from Ethiopia was purchased in a local supermarket grinded with an electric coffee grinder and preserved at 17°C until preparation of blends.
- Good quality crude barley was obtained from local market: Two kilograms of Moroccan barley picked up in Beni-Mellal, were roasted at 220°C for 3 hours. The roasted barley was grinded with an electric coffee grinder and preserved at 17°C until preparation of blends.

Samples were prepared by mixing Arabica coffee (A) with roasted Barley (B). Samples with a final mass of 10 g were prepared in different percentages in the 0–45 % weight ratio range of roasted Barley. All the samples were stored in a dry and dark location at ambient temperature (25°C) until analysis.

The final data base consists of 45 samples, containing spectroscopic and compositional information of the analyzed mixtures. Among which 30 samples (calibration set) were randomly selected for establishing principal component analysis (PCA) and partial least square regression (PLSR) models. Other 15 samples were used to test the applicability of the regression model (prediction set).

W. TEROUZI, Laboratory of Spectro- Chemometrics and environment, Faculty of Science and Technology of Beni Mellal, University of Sultan Moulay slimane, 21000- Beni Mellal, Morocco

A. OUSSAMA, Laboratory of Spectro- Chemometrics and environment, Faculty of Science and Technology of Beni Mellal, University of Sultan Moulay slimane, 21000- Beni Mellal, Morocco

# Mid-infrared spectroscopy and chemometric tools applied to quantification of adulteration in Arabica coffee by addition of roasted barley

## B. ATR-FTIR analysis

ATR-FTIR spectra were obtained using a Vector 22 Bruker FTIR Spectrophotometer equipped with an attenuated total reflectance accessory (ATR single reflexion, Diamond, incident angle 45°, Pike Miracle, Pike Technologies, Madison, USA) with DTGS detector, Globar (MIR) Source and KBr Germanium separator, with a resolution of 4 cm<sup>-1</sup> at 80 scans. Spectra were scanned in the absorbance mode from 4000 to 600cm<sup>-1</sup> and the data were handled with OPUS logiciel. About 1g of each binary blend powder samples of roasted Barley and Arabica coffee were directly placed, without preparation on an Attenuated Total Reflectance cell provided with a diamond crystal. Analyses were carried out at room temperature (25°C). The background was collected before every sample was measured. Between spectra, the ATR plate was cleaned in situ by scrubbing with ethanol solution, enabling to dry the ATR.

## C. Data pre-processing procedures

In this study, a series of pre-processing elaborations were tested on the spectral data prior to the multivariate calibration. In fact, several pre-processing methods were applied before calibration development in order to find regression model with as high a predictive power as possible. The Savitzky–Golay [15] and Norris gap [16] algorithms were tested for data derivatisation. Standard normal variate (SNV) and multiple scatter correction (MSC) [17] were also tested. For data pre-treatment giving best result is the derivative function. In all PCA and PLSR models, second derivative through the Gap algorithm has been applied as preprocessing technique with uncentered data, in order to correct the spectrum by separating overlapping peaks and to enhance spectral differences.

## D. Chemometric methods

### • Principal Component Analysis (PCA)

Principal component analysis (PCA) is an unsupervised technique commonly used for quantification, characterization and classification of data. It is based on variance, transforms the original measurement variables into new uncorrelated variables called principal components [18], [19]. It maps samples through scores and variables by the loadings in a new space defined by the principal components. The PCs are a simple linear combination of original variables. The scores vectors describe the relationship between the samples and allow checking if they are similar or dissimilar, typical or outlier. It provides a reduction in data set dimensionality and allows linear combinations of the original independent variables that are used to explain the maximum of data set variance [20].

### • Partial least squares regression (PLSR)

Partial least squares regression (PLSR) [21] is popular and the most commonly used multivariate calibration chemometrics methods. It is able to resolve overlapping spectral responses [22]. It assumes a linear relationship between the measured sample parameters (for example, concentration or content) and the experimentally measured spectra.

PLSR attempts to maximize the covariance between X and y data blocks as it searches for the factor subspace most congruent to both data blocks. A new matrix of weights

(reflecting the covariance structure between the X and y) is calculated and provided rich factor interpretation information [23].

In this study, the collected ATR-FTMIR spectra will be used as the X matrix, and the roasted barley compositions of the different samples will be used as the Y vector.

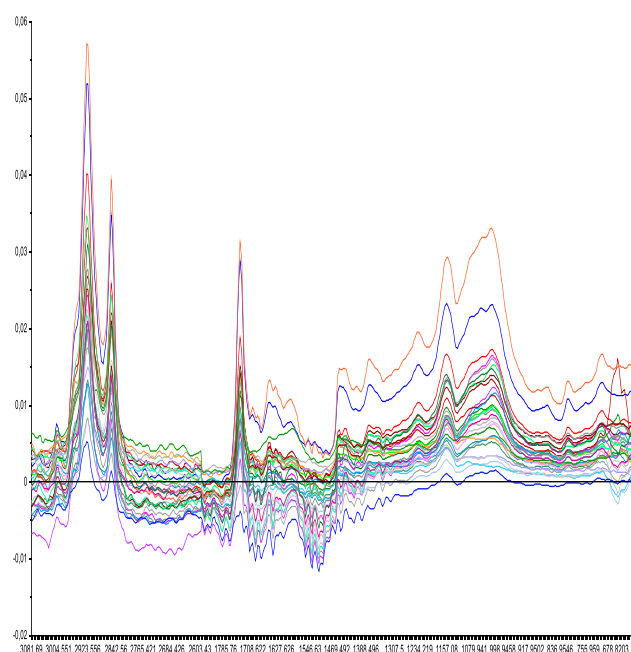
### • Software

The pre-treatment procedures and all chemometric models were performed by using the Unscrambler X software version 10.2 from Computer Aided Modelling (CAMO, Trondheim, Norway).

## III. RESULTS AND DISCUSSION

### A. Data acquisition

ATR-FTIR spectra of 45 samples of the studied binary mixtures were recorded and divided in two sets: a calibration set of 30 samples and an external validation set of 15 samples. One spectrum is the average of 80 scans of the same sample of adulterated coffee. The average spectra of all considered samples in calibration set are presented in Fig.1.



**Fig.1. ATR-FTIR spectra of the binary mixture (Arabica coffee – roasted Barley: A-B) samples of calibration set in the 0–45 % weight ratio range**

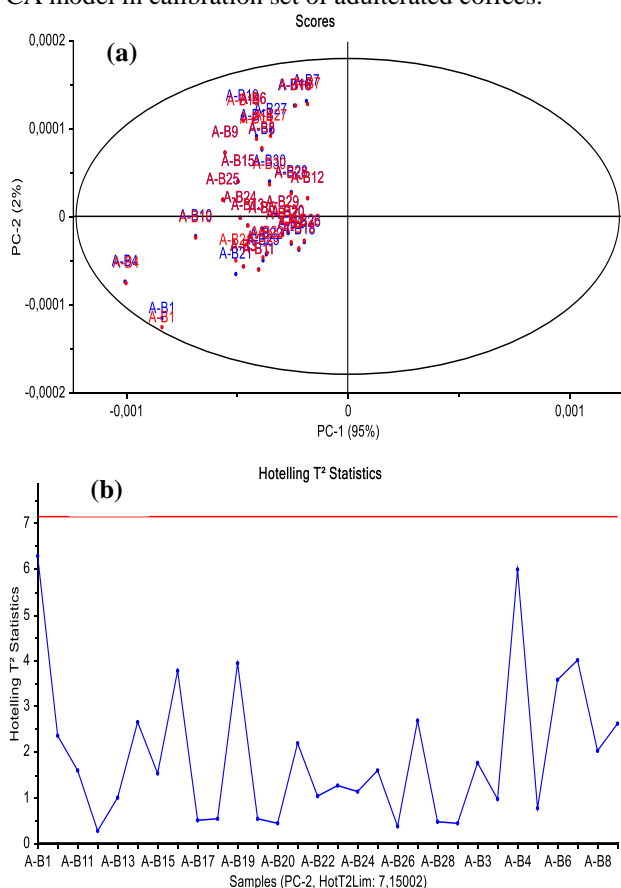
Fig.1 shows the ATR-FTIR spectra of the studied coffee blends at frequency region of 4000–600 cm<sup>-1</sup>. The obtained spectra are dominated by typical bands of holocellulosic materials in the 1200 - 900 cm<sup>-1</sup> region [17]. In the 4000-1200 cm<sup>-1</sup> region, All the spectra were dominated by two peaks at 2860 and 2850cm<sup>-1</sup>, due to bands arising from asymmetrical and symmetrical stretching vibrations of methylene (–CH<sub>2</sub>) groups. The peak at 3008 cm<sup>-1</sup> could be assigned to the functional group (trans =C–H stretch), and strong bands between 1750 and 1650cm<sup>-1</sup> arising from the stretching vibration of the ester carbonyl functional groups of the triglycerides. The band of the aromatic ring stretch of the lignin appears at 1604 cm<sup>-1</sup>. The bands from 1480 to 1450cm<sup>-1</sup> could be assigned to the bending vibrations of the –CH<sub>2</sub> and –CH<sub>3</sub> aliphatic groups [24], [25].

According to **Fig.1**, the MIR spectra obtained for calibration set of the studied coffee blends to be similar. Since the native FTIR spectra did not furnish enough information to build a reliable prediction model, a preliminary treatment of the data seemed necessary to extract better analytical information. At the same time, a data pre-treatment was considered useful to minimize instrumental problems as baseline fluctuation or noise. Different mathematical elaborations were so explored to handle the spectral data. Derivative elaboration showed the most interesting result. In particular, a significant enrichment in the data variance was reached when the starting data were transformed in derivative signals by means of Gap derivative algorithm. Different mathematical parameters in the derivative procedure were tested and results were optimized when the following parameters were selected: 2nd order, gap size 17; with uncentered data.

### B. Multivariate analysis

#### • PCA modeling

Principal component analysis was carried out to detect the presence of any spectral outliers in the spectral data, prior to develop a prediction model using PLS regression. Many studies indicate that PCA is a useful tool for the identification of spectral outliers in the absorbance spectra of the samples and can be employed to increase the quality of the prediction-model [26]. **Fig.2** shows the score plot obtained by PCA model in calibration set of adulterated coffees.



**Fig.2. PC1 / PC2 Score plot by PCA analysis on the calibration set of binary mixtures (Arabica coffee-roasted Barley) samples: (a) Hotelling T<sup>2</sup> ellipse; (b) Hotelling T<sup>2</sup> statistics**

**Fig.2.(a)** is the Hotelling T<sup>2</sup> ellipse plot. This 95% confidence ellipse can be included in scores plots and reveals potential

outliers, lying outside the ellipse. **Fig.2.(b)** is the Hotelling T<sup>2</sup> statistics plot. It is an alternative to plotting sample leverages. The plot displays the Hotelling T<sup>2</sup> statistic for each sample as a line plot. The associated critical limit (with a default p-value of 5%) is displayed as a red line.

The Hotelling T<sup>2</sup> statistic has a linear relationship to the leverage for a given sample. Its critical limit is based on an F-test. Use it to identify outliers or detect situations where a process is operating outside normal conditions.

According to **Fig.2 ((a) and (b))** of PCA score plot, there's no « outliers » in the data set. Then, the prediction model (PLSR) was building with all samples of calibration set.

#### • PLSR modeling

In general, the modeling consists of two steps: (1) calibration, where data characteristics (Calibration and internal validation samples) are investigated to find a model for their behavior; and (2) External validation, where data that did not participate in the calibration step (external validation samples) are used to evaluate the model adequacy and capability.

The quantification of roasted Barley in adulterated coffees was carried out using PLS algorithm. The PLSR model is built by considering the all spectra range 4000–600 cm<sup>-1</sup> with X as variable and the Y variables is associated to the different percentages of the roasted Barley. The PLSR model was evaluated using coefficient of determination (R<sup>2</sup>) in calibration, root-mean-square error of calibration (RMSEC) and cross validation (RMSECV).

The performance of the PLSR models on the independent validation set is assessed using R<sup>2</sup>, RMSEP and the residual prediction deviation (RPD). Here, the criteria of classifying RPD values [27] is adopted as follows: an RPD value below 1.5 indicates that the calibration is not usable; an RPD value between 1.5 and 2.0 indicates the possibility of differentiating between high and low values; an RPD value between 2.0 and 2.5 makes possible approximate quantitative predictions. For RPD value between 2.5 and 3.0 and beyond 3.0, the prediction is classified as good and excellent, respectively. Generally, a good model should have high values of R<sup>2</sup> and RPD, and low values of RMSEC, RMSECV and RMSEP.

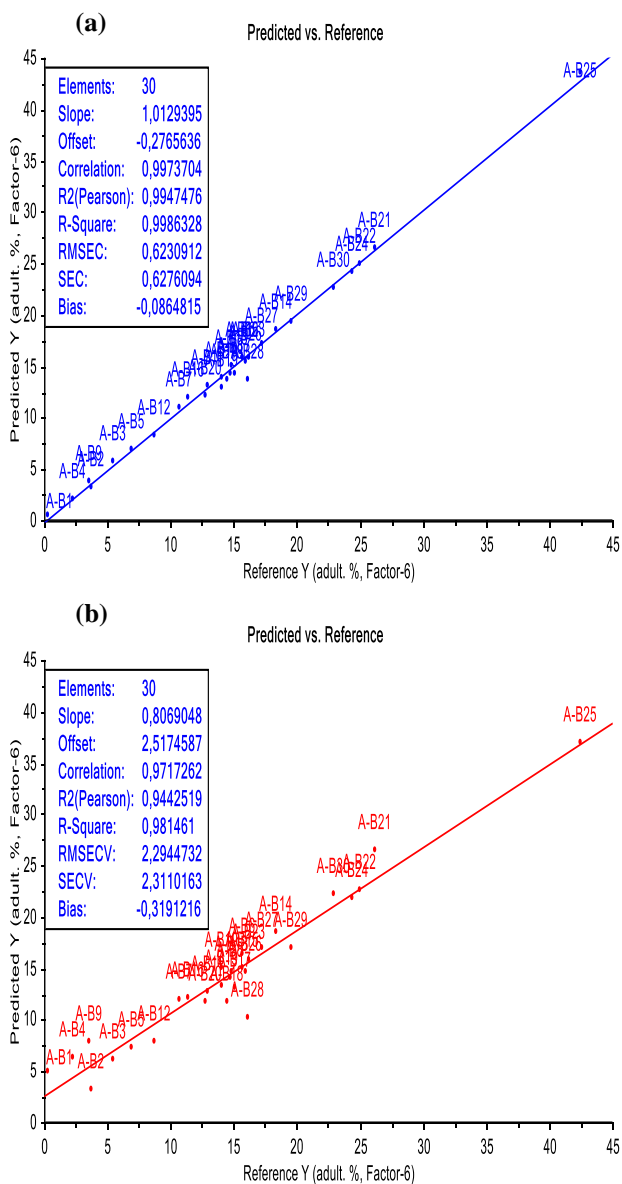
**Fig.3** shows the PLSR model which correlates the « actual » and « predicted » values of roasted Barley percentages obtained from ATR-FTMIR spectra. The term « actual » refers to the known percentage of roasted Barley. The « predicted » refers to a value calculated by the PLSR model using spectral data. The difference between the actual and the predicted percentage is relatively small with coefficient of determination (R<sup>2</sup>) values 0.99 with calibration set (**Fig.3.(a)**) and 0.9814 with internal validation (**Fig.3.(b)**). The low value RMSEC (< 0.624) indicates the good performance of PLSR model [28].

Additionally, validity of the model was checked by running several diagnostics including R<sup>2</sup>, root mean standard error of calibration (RMSEC) and root mean standard error of cross validation (RMSECV). Root mean square error of cross-validation (RMSECV), recovery percentage and coefficient of determination (R<sup>2</sup>) were used as parameters to determine appropriate number of latent variables (LV) [29], [30].

The determination of latent variables number was based on the statistical parameters that they offer the highest values of R<sup>2</sup> and the lowest values of error, either in calibration or in prediction models. The statistical parameters RMSEC,

## Mid-infrared spectroscopy and chemometric tools applied to quantification of adulteration in Arabica coffee by addition of roasted barley

RMSECV and  $R^2$  are summarized in Fig.3. The coefficient of determination ( $R^2$ ) of 0.99, RMSEC lower than 0.624 and RMSECV lower than 2.295, could be considered satisfactory. The six latent variables (factors) were sufficient for describing PLS model, with explained variances above 99% (Table1).



**Fig.3.** Measured vs. Predicted values for roasted Barley in the studied binary mixtures obtained from the final PLS model developed from the FT-MIR spectra: (a) calibration set; (b) Internal validation set (Cross validation).

**Table1.** Explained variances (%) of LVs used in the PLSR model.

Expl.	LV 1	LV2	LV3	LV4	LV5	LV6
Cal.	59.090	72.767	82.80	98.07	99.43	99.863
Val.	51.792	61.515	69.535	90.141	95.47	98.149

Expl.: Explained; Cal.; Calibration, Val.: Validation

•Prediction of roasted Barley content in the new binary blend samples (External validation)

In order to verify the applicability, performance and how reliable this model in estimating the percentage of roasted

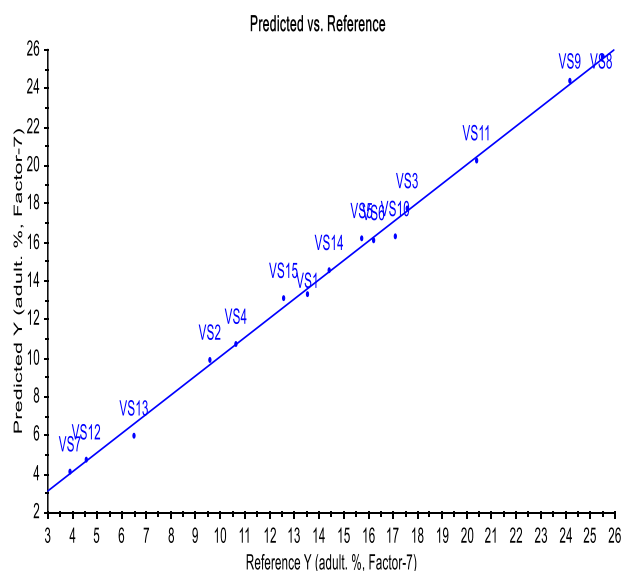
Barley in binary mixtures with Arabica coffee, the external validation process was carried out.

PLSR model was used to predict percentage of roasted Barley in new blend samples. The new samples were prepared within the range considered by the original database. These samples have the same matrix effects as samples of calibration set. In this step, the models were subdued to validation procedure by quantifying the new objects.

The PLSR model was applied to a group of external samples (15 samples), the results are shown in Fig.4.

Fig.4 shows the PLSR model reconstructed by external validation samples, following the same previous pre-treatments. This PLSR model correlates the « actual » and « predicted » values of roasted Barley percentages obtained from ATR-FTMIR spectra. The difference between the actual and the predicted percentage is relatively small.

Figures of merit of the calibration graphs are summarized in Table2. As can be seen, PLSR model offered good values for the different multivariate parameters.



**Fig.4.** Measured vs. Predicted values for roasted Barley in binary mixtures Arabica coffee-roasted Barley of external validation set.

**Table2.** Statistical parameters carried out by external validation on PLSR

LVs	Rp <sup>2</sup>	RMSEP	Bias	SEP	REP %	RPD	LD%
7	0.99	0.337	0.0017	0.35	2.053	17.9	1.012

## IV. CONCLUSION

Quantitative analysis of food adulterants is an important for health, wealth and economic issue that needs to be fast and simple. This work proposes a new method to quantify the adulteration of Arabica coffee with roasted Barley; based on FT-MIR spectroscopy using ATR accessory coupled with multivariate analysis tool (PLSR).

In the light of the statistical results, it has been proved that the proposed method allow the correct quantification of roasted Barley in the studied coffees. The PLSR model obtained from transformed infrared spectra gave correlation coefficients of 0.99 and root mean square errors of prediction (RMSEP) value of 0.337.

Finally, we arrived to develop a new application of the ATR-FTMIR associated with PLSR technique as a rapid, inexpensive and non destructive adulteration measuring tool, useful to determine the percentage of roasted Barley in the binary mixture with Arabica coffee.

In fact, this approach can be used in food industry for the reliable, cheap and fast quality control of raw material. Also, able to ensure authenticity, quality, safety and efficacy of final products to be commercialized.

## REFERENCES

- [1] Singhal, R.S., Kulkarni, P.R., Rege, D.V. Handbook of Indices of Food Quality and Authenticity, Woodhead Publishing Ltd., Cambridge, UK, 1997.
- [2] Lees, M. (Ed.), Food Authenticity and Traceability, Woodhead Publishing Ltd., Cambridge, UK, 2003.
- [3] Labouisse, J.P., Bellachew, B., Kotecha, S., Bertrand, B. Springer Science and Business Media, B.V 200, 2008.
- [4] Jham, G. N., Winkler, J. K., Berhow, M. A., & Vaughn, S. F. Gamma-tocopherol as a marker of Brazilian coffee (*Coffea arabica* L.) adulteration by corn. *Journal of Agricultural and Food Chemistry*, 55(15), 2007, 5995-5999.
- [5] Toci, A. T., Farah, A., Pezza, H. R., & Pezza, L. Coffee Adulteration: More than two decades of research. *Critical Reviews in Analytical Chemistry*, 46(2), 2016, 83-92.
- [6] Novaes, F.J.M., Oigman, S.S., Souza, R.O.M.A., Rezende, C.M., Neto, F.R.A. New approaches on the analyses of thermolabile coffee diterpenes by gas chromatography and its relationship with cup quality. *Talanta* 139, 2015, 159–166.
- [7] Souto, U.T.C.P., Barbosa, M.F., Dantas, H.V., Pontes, A.S., Lyra, W.S., Diniz, P.H.G.D., Araujo, M.C.U., Silva, E.C. Identification of adulteration in ground roasted coffees using UV-Vis spectroscopy and SPA-LDA. *LWT – Food Sci. Technol.* 63, 2015, 1037-1041.
- [8] Arana, V.A., Medina, J., Alarcon, R., Moreno, E., Heintz, L., Schafer, H., Wist, J. Coffee's country of origin determined by NMR: the Colombian case. *Food Chem.* 175, 2015, 500-506.
- [9] Marquetti, I., Varaschim Link, J., Guimaraes Lemes, A.L., Dos Santos Scholz, M.B., Patricia Valderrama, P., Bona, E. Partial least square with discriminant analysis and near infrared spectroscopy for evaluation of geographic and genotypic origin of arabica coffee. *Computers and Electronics in Agriculture* 121, 2016, 313–319
- [10] Esteban-Díez, I., González-Sáiz, J.M., Sáenz-González, C., Pizarro, C. Coffee varietal differentiation based on near infrared spectroscopy. *Talanta* 71, 2007, 221-229.
- [11] Santos, J.R., Sarraguça, M.C., Rangel, A.O.S.S., Lopes, J.A. Evaluation of green coffee beans quality using near infrared spectroscopy: a quantitative approach. *Food Chem.* 135, 2012, 1828-1835.
- [12] Bertone, E., Venturello, A., Giraud, A., Pellegrino, G., Geobaldo, F. Simultaneous determination by NIR spectroscopy of the roasting degree and Arabica/Robusta ratio in roasted and ground coffee. *Food Control* 56, 2016, 683-689.
- [13] Pizarro, C., Esteban-Díez, I., González-Sáenz, J.M. 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, 2007, 266–276
- [14] Craig, A.P., Franca, A.S., Oliveira, L.S., Irudayaraj, J., Ileleji, K. Fourier transform infrared spectroscopy and near infrared spectroscopy for the quantification of defects in roasted coffees. *Talanta* 134, 2015, 379–386.
- [15] Savitzky, A. & Golay, M.J.E. Smoothing and differentiation of data by simplified least-squares procedures. *Analytical Chemistry*, 36, 1964, 1627–1639.
- [16] Norris, K.H. & Williams, P.C. Optimization of mathematical treatments of raw near infrared signal in the measurement of protein in hard Red Spring wheat, I: influence of particle size. *Cereal Chemistry*, 62, 1984, 158–165.
- [17] Iñón, F. A., Garrigues, J. M., Garrigues, S., Molina, A., and de la Guardia, M. Selection of calibration set samples in determination of olive oil acidity by partial least squares-attenuated total reflectance-Fourier transform infrared spectroscopy. *Analytica Chimica Acta*, 489, 2003, 59–75.
- [18] Vandeginste, B.G.M., Massart, D.L., Buydens, L.M.C., Jong, S.D.E., Lewi, P.J., Smeyers Verbeke, J. Handbook of Chemometrics and Qualimetrics: Part B, Elsevier, Amsterdam, The Netherlands, 1998, p. 88.
- [19] Wold, S., Esbensen, K. & Geladi, P. Principal component analysis. *Chemom. Intell. Lab. Syst.*, 2, 1987, 37-52.
- [20] Brown, S.D., Tauler, R., Walczak, B. Comprehensive Chemometrics: Chemical and Biochemical Data Analysis, vol. 2. Elsevier, Amsterdam, 2009.
- [21] Martens, H. & Naes, T. Multivariate calibration. Chichester: Wiley, 1989.
- [22] Khanmohammadi, M., Bagheri Garmarudi, A., Ghasemi, K., Garrigues, S., De la Guardia, M. Artificial neural network for quantitative determination of total protein in yogurt by infrared spectrometry. *Microchemical Journal* 91, 2009, 47–52.
- [23] Geladi, P. & Kowalski, B. (1986) Partial least-squares regression: a tutorial. *Anal Chim Acta*, 185, 1986, 1–17.
- [24] Rohman, A., & Che Man, Y.B. Simultaneous quantitative analysis of two functional food oils, extra virgin olive oil and virgin coconut oil using FTIR spectroscopy and multivariate calibration. *International Food Research Journal* 18(4), 2011, 1231-1235
- [25] López-Sánchez M., Ayora-Cañada M.J., Molina-Díaz A. Olive fruit growth and ripening as seen by vibrational spectroscopy. *J Agric Food Chem.*, 58, 2010, 82–87.
- [26] Pirie, A.; Singh, B. & Islam, K. Ultra-violet, visible, near-infrared, and mid-infrared diffuse reflectance spectroscopic techniques to predict several soil properties. *Aust J Soil Res*, 43, 2005, 713-721.
- [27] Mouazen, A.M., Baerdemaeker, J.D. and Ramon, H. Effect of wavelength range on the measurement accuracy of some selected soil constituents using visual-near infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, 14, (3), 2006, 189-99.
- [28] Rohman, A., & Che Man, Y.B. The use of Fourier transform mid infrared (FT-MIR) spectroscopy for detection and quantification of adulteration in virgin coconut oil. *Food Chemistry* 129, 2011, 583–588
- [29] Durante, C., Cocchi, M., Grandi, M., Marchetti, A., & Bro, R. Application of N-PLS to gas chromatographic and sensory data of traditional balsamic vinegars of Modena. *Chemometrics and Intelligent Laboratory Systems*, 83(1), 2006, 54–65.
- [30] Rezaei, B., Khayamian, T., & Mokhtari, A. (2009). Simultaneous determination of codeine and noscapine by flow-injection chemiluminescence method using N-PLS regression. *Journal of pharmaceutical and biomedical analysis*, 49(2), 2009, 234–239.