IOSR Journal of Applied Chemistry (IOSR-JAC)
e-ISSN: 2278-5736.Volume 9, Issue 7 Ver. I (July 2016), PP 97-103
www.iosrjournals.org

Quantifying roasted date seed coffee in a binary mixture with Arabica coffee by Mid Infrared Spectroscopy and Chemometrics

W.Terouzi, 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

Abstract: *This paper proposes a new method for the quantitative analysis of date seed coffee in a binary mixture with Arabica coffee by applying Attenuated Total Reflectance-Fourier Transform Mid Infrared Spectroscopy (ATR-FTMIR) associated with chemometric tools. Blends of Arabica coffee with different percentages of Date seed coffee 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 < 2.31, Limit of Detection 6.912%, and Relative Prediction Errors as low as 0.21. This result demonstrated the feasibility of ATR-FTIR spectroscopy combined with chemometrics to quantify successfully binary mixtures of Arabica coffee in the 7–50 % weight ratio range of Date seed coffee with a reliable, rapid and inexpensive tool without the need for sample preparation.*

Keywords: *Arabica coffee, chemometrics, date sees coffee, infrared spectroscopie, quantification.*

I. Introduction

Coffee is one of the most preferred and consumed beverages around the world due to its pleasant taste and flavor [1]. Also, The specialty coffee market is constantly growing,which is driven by changing consumers preferences [2]. An increasing demand for such coffees on the world market resulted in a segmentation of the coffee market and higher prices are paid for specialty coffees [3]. In worldwide,two coffee varieties are of the highest economic importance: Arabic and Robusta. Arabica beans are more expensive and valuable because of their preferable flavor in comparison to Robusta beans [1]. Additionally, The date seed have been roasted and grinded into smaller size to turn it into caffeine-free coffee substitute, which have been commercialized by the Arabs in two types, whether plain or mixed with coffee [4], [5].

On the other hand, the authenticity of raw materials and food products presents a huge importance for regulatory agencies, consumers, food processors, and industries, in order to satisfy food quality and safety requirements [6], [7]. In particular, assurance of the quality of roasted coffees has attracted widespread attention as a means for controlling and preventing adulterated or mixed coffee, and also given the great difference in the final sale price depending on a wide range of factors, including coffee varietal and geographic origin.

According to littérature, different analytical techniques are often employed for coffee analysis, including chromatographic analysis [8], UV–Vis spectroscopy [9], nuclear magnetic resonance [10]. 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 [11] or blends of arabica and robusta grains [12], [13], [14] and to detect adulteration or defects in coffees [15], [16]. 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.

In this context, the objective of this study was to examine the ability and feasibility of ATR-FTMIR spectroscopy coupled with chemometric methods to predict the actual content of Date seed coffee in the binary mixtures with Arabica coffee. In fact, 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 Date seed coffee in the binary blend samples.

II. Materials And Methods

In the following subsections, the different coffees, materials, and methods employed are detailed.

2.1. Samples preparation

In this study, to prepare the coffee binary mixtures 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.
- Pure Date seed coffee: Two kilograms of Moroccan date fruits from the "Bouslikhan" variety picked up in Rachidia. The seeds were manually separated from date fruits, thoroughly washed with normal tap water and then with distilled water to remove the adhering dirt, and finally dried in an oven set at 80°C for 8 hours. Washed and dried date seeds were roasted at 220°C for 6 hours. The roasted date seeds were grinded with an electric coffee grinder and preserved at 17°C until preparation of blends.

Samples were prepared by mixing Arabica coffee (A) with Date seed coffee (D). Samples with a final mass of 10 g were prepared in different percentages in the 7–50 % weight ratio range of Date seed coffee. All the samples were stored in a dry and dark location at ambient temperature (25°C) until analysis.
The final data base consists of 30 samples, containing spectroscopic and compositional information of the analyzed mixtures. Among which 20 samples (calibration set) were randomly selected for establishing principal component analysis and partial least square regression (PLSR) models. Other 10 samples were used to test the applicability of the regression model (prediction set).

2.2. ATR-FTIR analysis

ATR-FTIR spectra were obtained using a PerkinElmer spectrum, Version 10.5.1 equipped with an attenuated total reflectance accessory with DTGS detector, Globar (MIR) Source and KBr Germanium separator, with a resolution of 4 cm^{-1} at 80 scans. Spectra were scanned in the absorbance mode from 4000 to 450 cm^{-1} and the data were handled with PerkinElmer logiciel. About 1g of each binary blend powder samples of coffees from date seed and Arabica variety 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.

2.3. 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 [17] and Norris gap [18] algorithms were tested for data derivatisation. Standard normal variate (SNV) and multiple scatter correction (MSC) [19] were also tested. For data pre-treatment giving best result is the derivative function. In all PCA and PLSR, second derivative through the Gap algorithm has been applied as preprocessing technique with centered data, in order to correct the spectrum by separating overlapping peaks and to enhance spectral differences.

2.4. Multivariate analysis

2.4.1. Principal Component Analysis (PCA)

Principal component analysis (PCA) is an unsupervised technique commonly used for quantifiaction, characterization and classification of data. It is based on variance, transforms the original measurement variables into new uncorrelated variables called principal components [20], [21]. 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 [22].

2.4.2. Partial least squares regression (PLSR)

Partial least squares regression (PLSR) [23] is popular and the most commonly used multivariate calibration chemometrics methods. It is able to resolve overlapping spectral responses [24]. It assumes a linear relationship between the measured sample parameters (for example, concentration or contenent) 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 [25].
In this study, the collected MIR spectra will be used as the X matrix, and the Date seed coffee compositions of the different samples will be used as the Y vector.

2.5. 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

3.1. Data acquisition

ATR-FTIR spectra of 30 samples of the studied binary mixtures, were recorded and divided in two sets: a calibration set of 20 samples and an external validation set of 10 samples. One spectrum is the average of 80 scans of the same sample of coffee blend. The average spectra of all considered samples in calibration set are presented in **Fig.1**.

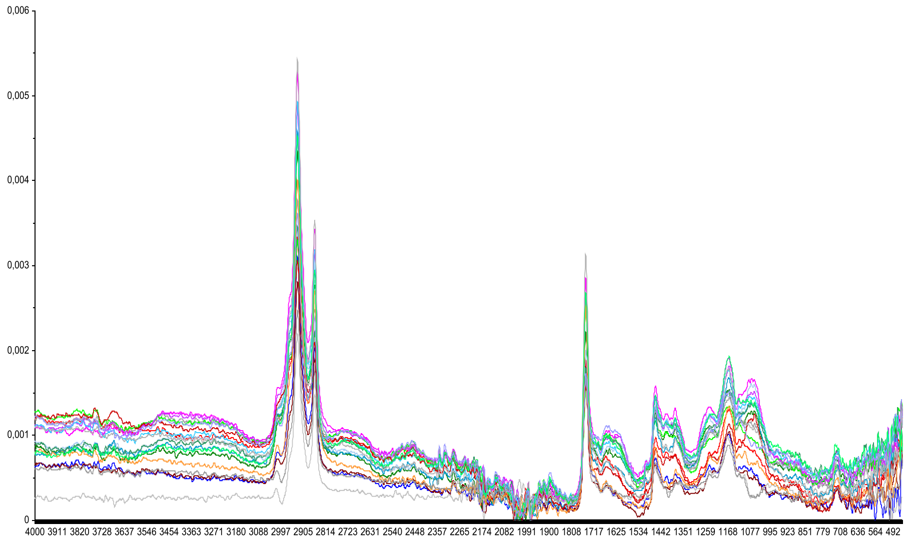

Fig.1. ATR-FTIR spectra of the binary mixture (Arabica coffee - Date seed coffee : A-D) samples of calibration set in the 7–50 % weight ratio range

Fig.1 shows the ATR-FTIR spectra of the studied coffee blends at frequency region of 4000–450 cm^{-1}. The obtained spectra are dominated by typical bands of holocellulosic materials in the 1200 - 900 cm^{-1} region [19]. In the 4000-1200 cm^{-1} region, All the spectra were dominated by two peaks at 2860 and 2850cm^{-1}, due to bands arising from asymmetrical and symmetrical stretching vibrations of methylene (–CH2) groups. The peak at 3008 cm^{-1} could be assigned to the functional group (trans =C–H stretch), and strong bands between 1750 and 1650cm^{-1} 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^{-1}. The bands from 1480 to 1450cm^{-1} could be assigned to the bending vibrations of the –CH2 and –CH3 aliphatic groups [26], [27].

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 centered data.

3.2. Statistical analysis

3.2.1. 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 [28].

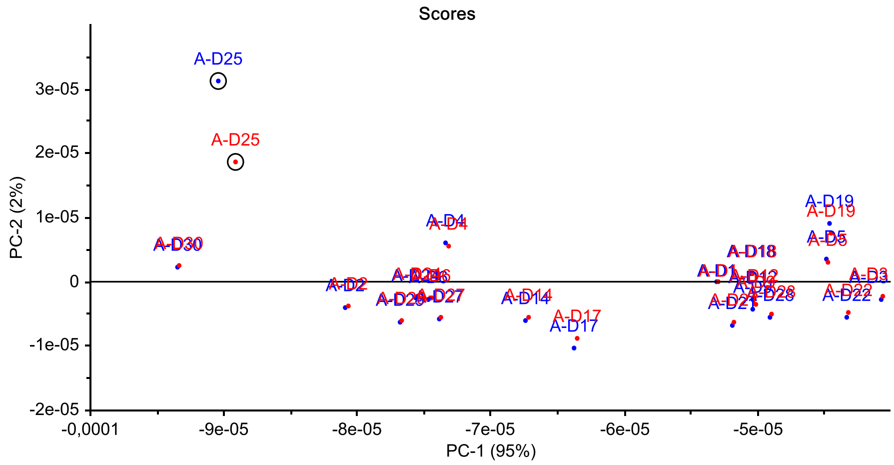

Fig. 2. PC1 / PC2 Score plot by PCA analysis on the calibration set of binary mixtures (Arabica coffee- Date seed coffee) samples

According to **Fig.2** of PCA score plot, the data set contained one spectral « outlier » (A-D25). However, at first, the prediction model (PLSR) was building with all samples including this sample to insure his nature (outlier or extreme sample).

3.2.2. 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 Date seed coffee in coffee blends was carried out using PLS algorithm. The PLS model is built by considering the all spectra range 4000–450 cm^{-1} with X as variable and the Y variables is associated to the different percentages of the Date seed coffee. The data set contained 20 coffee samples including the « outlier » spectral identified by PCA (**Fig.2**) because it is considered extreme by PLS. The PLSR model was evaluated using coefficient of determination (R^2) 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^2, RMSEP and the residual prediction deviation (RPD). Here, the criteria of classifying RPD values [29] 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^2 and RPD, and low values of RMSEC, RMSECV and RMSEP.

Fig.3 shows the PLSR model which correlates the « actual » and « predicted » values of Date seed coffee percentages obtained from FT-MIR spectra. The term « actual » refers to the known percentage of Date seed coffee. 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^2) values 0.99 with calibration set (**Fig.3.(a)**) and 0.985 with internal validation (**Fig.3.(b)**). The low value RMSEC (< 1.13) indicates the good performance of PLS model [30].

Additionally, validity of the model was checked by running several diagnostics including R^2, 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^2) were used as parameters to determine appropriate number of latent variables (LV) [31], [32].

The determination of latent variables number was based on the statistical parameters that they offer the highest values of R^2 and the lowest values of error, either in calibration or in prediction models. The statistical parameters RMSEC, RMSECV and R^2 are summarized in **Fig.3**. The coefficient of determination (R^2) of 0.99, RMSEC lower than 1.13 and RMSECV lower than 2.61, could be considered satisfactory. The two latent variables (factors) were sufficient for describing PLS model, with an explained variances above 99% (**Table 1**).

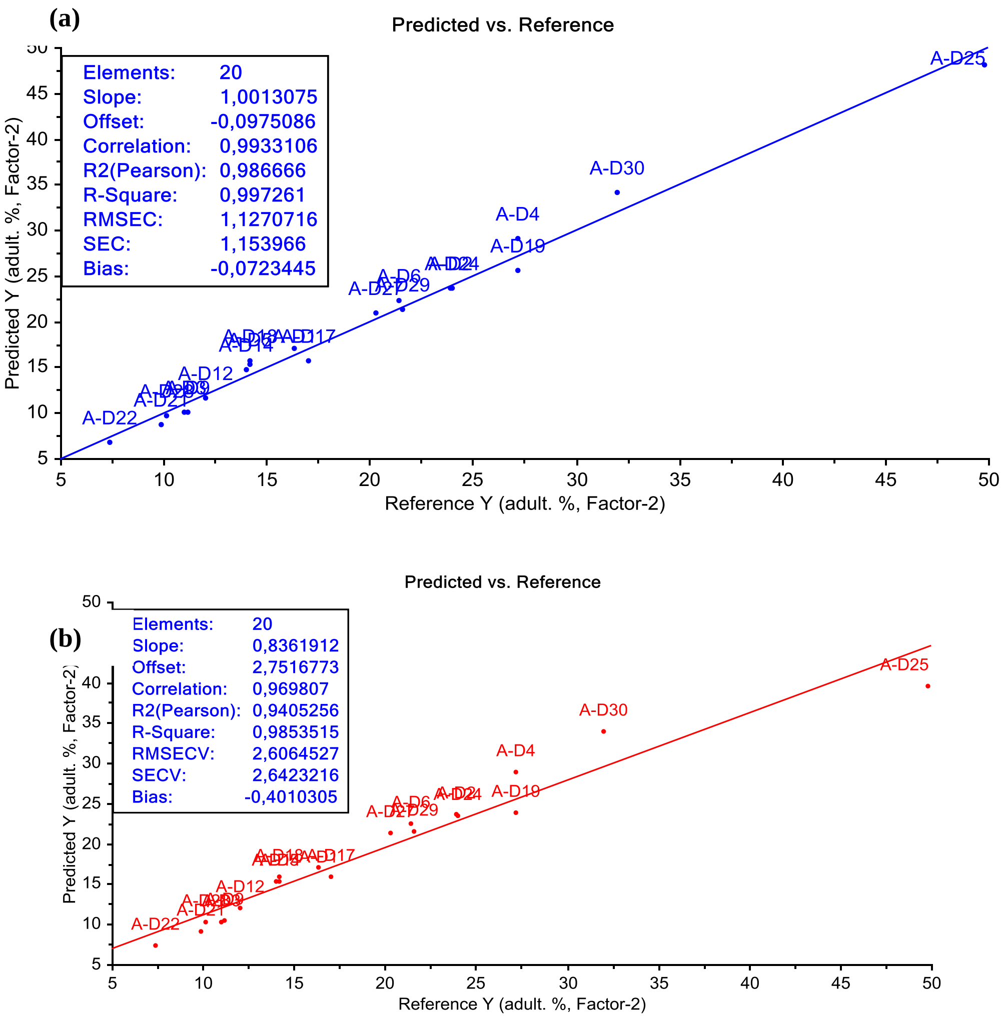

Fig. 3. Measured vs. Predicted values for date seed coffee in the studied binary mixtures, obtained from the final PLS model developed from the FT-MIR spectra: (a) calibration set; (b) Internal validation set (Cross validation).

Table 1. Explained variances (%) of PCs used in the PLSR model.

	Explained	Factor 1	Factor 2
PLSR	Calibration	89.8321	99.7261
	Validation	87.69585	98.53515

3.2.3. Prediction of Date seed coffee 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 Date seed coffee in binary mixtures with Arabica coffee, the external validation process was carried out.

PLS models are used to predict percentage of Date seed coffee 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 were applied to a group of external samples, the results are listed in **Table 2**. The deviations of prediction of Date seed coffee composition in the blend samples by FT-MIR spectroscopy were between 2.05 and 2.79, which were very satisfied (**Table 2**).

Table 2. Prediction results of date seed coffee content (%) in the binary mixtures by FT-MIR spectroscopy coupled with PLSR.

Samples of Validation	Pred.	Dev.	Ref.
VS1	11.5098	2.4123	12.0414
VS2	14.6731	2.2237	14.0275
VS3	15.6796	2.7912	17.0197
VS4	25.5403	2.3128	24.0156

VS5	8.7636	2.0521	9.8632
VS6	23.6430	2.3864	24.0129
VS7	48.0663	2.6583	49.0459
VS8	20.9651	2.4152	21.9985
VS9	22.2541	2.2307	23.0895
VS10	34.0898	2.2507	32.0342

Pred. : Predicted ; Dev. : deviation ; Ref. : Reference

According to **Table 2**, there is no significant difference between the reference methods and the proposed one. The PLS model for the ATR-FTIR data treatment appears to be appropriate therefore.
In addition, **Fig.4** shows the PLSR model reconstructed by external validation samples, following the same previous pre-treatments. Which correlates the « actual » and « predicted » values of Date seed coffee 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 **Table 3**. As can be seen, PLSR model offered good values for the different multivariate parameters.

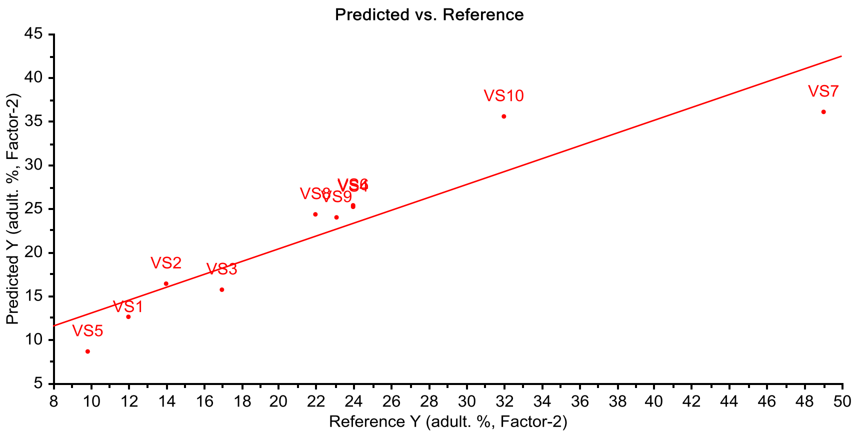

Fig. 4. Measured vs. Predicted values for date seed coffee in binary mixtures Arabica coffee-Date seed coffee of external validation set.

Table 3. Statistical parameters carried out by external validation on PLSR

External validation	LVs	Rp^2	RMSEP	Bias	SEP	REP %	RPD	LD%
	2	0.99	2.3040	- 0.00944	2.4286	0.2072	2.6933	6.912

IV. Conclusions

This paper proposes a new method for the quantitative analysis of Date seed coffee in binary blend with Arabica coffee, by applying Attenuated Total Reflectance-Fourier Transform Mid Infrared Spectroscopy (ATR-FTMIR) 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 date seed coffee in the studied binary blend 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 2.3040.

Finally, we arrived to develop a new application of the ATR-FTIR associated with PLSR technique as a rapid, inexpensive and non destructive authenticity measuring tool, useful to determine the percentage of Date seed coffee 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. Thus ensuring authenticity, quality, safety and efficacy of final products to be commercialised.

References

[1]. Belitz, H. D., Grosch, W., & Schieberle, P. Food chemistry. Berlin-Heidelberg: Springer-Verlag, 2009.
[2]. Oberthür, T., Läderach,P., Jurgen, P.H.A., Cock, J.H. Specialty coffee- managing quality, Secondary, International Plant Nutrition Institute, Southeast Asia Program, 2012.
[3]. Labouisse, J.P., Bellachew, B., Kotecha, S., Bertrand, B. Springer Science and Business Media, B.V 200, 2008.
[4]. Rahman, M. S., Kasapis, S., Al-Kharusi, N. S. Z., Al-Marhubi, I. M. and Khan, A. J. Composition characterization and thermal transition of date pits powders. Journal of Food Engineering 80, 2007, 1-10.
[5]. Al-Farsi, M. A. and Lee, C. Y. Usage of date (Phoenix Dactylifera L.) seeds in human health and animal feed. In Preedy, V.R., Watson, R.R.; Patel, V.B. (eds). Nuts and Seeds in Health and Disease Prevention. USA: Elsevier, 2011, 447-452.

[6]. Singhal, R.S., Kulkarni, P.R., Rege, D.V. Handbook of Indices of Food Quality and Authenticity, Woodhead Publishing Ltd., Cambridge, UK, 1997.
[7]. Lees, M. (Ed.), Food Authenticity and Traceability, Woodhead Publishing Ltd., Cambridge, UK, 2003.
[8]. 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.
[9]. 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 UVeVis spectroscopy and SPA-LDA. LWT – Food Sci. Technol. 63, 2015, 1037-1041.
[10]. 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.
[11]. Marquetti, I.,Varaschim Link,J.,Guimarães Lemes,A.L., Dos Santos Scholz, M.B., Patrícia 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
[12]. Esteban-Díez, I., Gonzáliz-Sáiz, J.M., Sáenz-González, C., Pizarro, C. Coffee varietal differentiation based on near infrared spectroscopy. Talanta 71, 2007, 221-229.
[13]. 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.
[14]. Bertone, E., Venturello, A., Giraudo, 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.
[15]. Pizarro, C., Esteban-D´ıez, 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
[16]. 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.
[17]. Savitzky, A. & Golay, M.J.E. Smoothing and differentiation of data by simplified least-squares procedures. Analytical Chemistry, 36, 1964, 1627–1639.
[18]. 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.
[19]. 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.
[20]. 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.
[21]. Wold, S., Esbensen, K. & Geladi, P. Principal component analysis. Chemom. Intell. Lab. Syst., 2, 1987, 37-52.
[22]. Brown, S.D., Tauler, R., Walczak, B. Comprehensive Chemometrics: Chemical and Biochemical Data Analysis, vol. 2. Elsevier, Amsterdam, 2009.
[23]. Martens, H. & Naes, T. Multivariate calibration. Chichester: Wiley. 1989.
[24]. 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.
[25]. Geladi, P. & Kowalski, B. (1986) Partial least-squares regression: a tutorial. Anal Chim Acta, 185, 1986, 1–17.
[26]. 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
[27]. 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.
[28]. 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.
[29]. 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.
[30]. 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
[31]. 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.
[32]. 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.