Acta Scientiarum
http://periodicos.uem.br/ojs
ISSN on-line: 1807-8664
Doi: 10.4025/actascitechnol.v46i1.63058

CHEMISTRY

Discrimination of baru oil adulteration with soy oil using portable near infrared spectroscopy, Fourier-transform infrared spectroscopy and Partial Least-Squares Discriminant

Tamara Mendes Leite Silva Trindade[1], Laiz de Oliveira Magalhães[2], Stephane Bandeira[1], Carla Jovania Gomes Colares[3] and Nelson Luis Gonçalves Dias de Souza[1,3*]

[1]Programa de Pós-graduação em Química, Universidade Federal do Tocantins, Caixa Postal 66, 77402-970, Gurupi, Tocantins, Brazil. [2]Instituto de Química, Universidade de Brasilia, Brasília, Distrito Federal, Brazil. [3]Departamento de Ciências Exatas e Biotecnologia, Universidade Federal do Tocantins, Gurupi, Tocantins, Brazil. *Author for correspondence. E-mail: nelson.luis@uft.edu.br

ABSTRACT. Considered an important source of bioactive compounds and unsaturated fatty acids, Baru oil comes from the Brazilian Cerrado ecoregion and is sold with high value. Methods to detect tampering to certify the authenticity of this type of product are in great demand for quality control. Herein, this study used partial least squares for discriminant analysis (PLS-DA) combined with Fourier-transform infrared spectroscopy (FTIR) and portable near-infrared spectroscopy (NIR) for the detection of adulteration in Baru oil, using soybean oil as an adulterant. Adulteration concentrations were 10, 30, 50, 60, and 80% for analysis by NIR and 10, 30, 50, 60, 70, and 90% for FTIR. After the spectral range selection and pre-processing steps, discrimination models showed good discrimination, and the efficiency was confirmed from the figures of merit, which highlighted percentages above 99% with both techniques used. Anomalous samples were identified in discrimination models obtained with a portable NIR spectrometer in the outlier detection step. The methodology evaluated is direct, fast, and non-destructive, in which the PLS-DA models obtained from the sets of spectra by FTIR presented superior results than those evaluated with the portable NIR spectrometer; however, the latter was also efficient in discriminating as adulterated, reinforcing the feasibility of using spectroscopic techniques and chemometric tools.

Keywords: chemometrics, vibrational spectroscopy, cerrado, *Dipteryx alata vog.*

Received on March 28, 2022.
Accepted on March 30 2023.

Introduction

Cerrado is the second-largest Brazilian biome and provides a large diversity of fruits rich in nutritional and bioactive properties. Considered the richest savanna in the world, it contains about 12,000 cataloged species representing 5% of the planet's biodiversity (Sano et al., 2019). Thus, studies on the fruits of Cerrado and their uses have caused emerging interest.

Baru, a *Dipteryx alata vog* species, is a fruit of the family Fabaceae widely distributed in South and Central America. This fruit is brown ovoid with one almond and is generally used in food because of its nutritional quality (Oliveira-Alves et al., 2020; Pineli et al., 2015; Sousa, Fernandes, Alves, Freitas, & Naves, 2011). The oil is one of the Baru almond's co-products, known for its high content of phenolic compounds, antioxidant activity (Pineli et al., 2015) and its role in reducing cholesterol and triglycerides (dos Santos et al., 2016).

The production of oils, which play functional and sensory roles, is of interest in several industrial segments, such as the pharmaceutical and cosmetics industries (Fernandes, Freitas, Czeder, & Naves, 2010). However, its low extraction yield makes adulteration a common practice, in which oils with similar physical-chemical characteristics, such as soybean oil, are added (Al-Ahmed, Alsowaidi, & Vadas, 2008). This practice causes great concern due to its negative effects on the quality of the product, as its attributable and bioactive properties are compromised .

Instrumental analytical methodologies that do not require long execution times, that are easy to execute, and that minimize waste generation have stood out in the field of adulteration identification and quality control in several sectors. Among these, infrared spectroscopy stands out (Moreira, Machado, Almeida, & Braga, 2018).

However, these methodologies generate a significant amount of information, which hinders the processing and interpretation of the data obtained. Thus, Chemometrics, a science that employs a set of mathematical, statistical, and computational methods for analysis of data obtained in chemical experiments, solves these problems, making it a valuable tool to remove redundant information contained in the spectra, allowing the identification and the quantification of several parameters in different matrices (Frank & Friedman, 1993).

Among the most explored categories in the Chemometrics field, two of the best known for exploratory analysis are Principal Component Analysis (PCA) and Partial Least Squares for Discriminant analysis (PLS-DA). The first is usually aimed at visualizing the data structure, by transforming a large set of variables into a smaller one that still contains most of the information, and identifying anomalies in the dataset (Wold, Esbensen, & Geladi, 1987), while the latter seeks to discriminate samples based on their predicted values (0's or 1's) obtained by PLS algorithm (Ballabio & Consonni, 2013; Brereton & Lloyd, 2014).

PCA is a widely used technique to for reducing the dimensionality of datasets, and thus define those pertinent to the analysis without affecting relationships between the samples. In this way, this information is expanded to aid in visual inspection, enabling the visualization and interpretation of differences between samples and highlighting atypical behaviors, which become even more evident in such a technique. As an unsupervised methodology, PCA does not require initial information on the analyzed samples (Wold et al., 1987).

PLS-DA methodology is a commonly used method that allows the prediction of samples in a predefined target class while other samples (such as interferents or samples with different physical-chemical aspects) can be predicted in another class. This tool does not require exact knowledge of all the components present in the samples under study, as long as they are present in the construction of the model (Ballabio & Consonni, 2013).

There are several studies on the development of methodologies for the detection of adulteration of olive oil (Jiang & Chen, 2019; Wójcicki et al., 2015), biodiesel (Câmara et al., 2017) and copaiba oil (Moreira et al., 2018). There are also studies to evaluate drug authenticity (Magalhães, Arantes, & Braga, 2019), for discrimination of different wood-producing forest species (Pastore et al., 2011), the different stages of human milk lactation (Botelho, Reis, Oliveira, & Sena, 2015), non-destructive identification of different types and brands of inks (Silva et al., 2014), and the different adulterants in beef meat using infrared spectroscopy and chemometric tools (Nunes, Andrade, Almeida, & Sena, 2020). All these examples show the potential of these methodologies in the control of vegetable oils and other diverse applications.

To this end, this aimed to expand the knowledge in the detection of possible adulteration in Baru vegetable oil, combining chemometric models based on PLS-DA, with simplified sample preparation, and direct analysis using mid- and near-infrared spectroscopy portable equipment.

Material and methods

Raw material and processing

Baru seeds were obtained in the municipality of Gurupi, state of Tocantins, and initially washed, dried, and stored in a refrigerator until oil extraction. A Soxhlet extraction was conducted using a 365.1 g sample and ethanol (99.8%) for 6 hours. The solvent was evaporated, and the solids were removed by centrifugation. Other solvents could be used for extraction, such as hexane, however, ethanol was chosen because of its low toxicity (Lôbo, Ferreira, & Cruz, 2009).

Adulterated sample preparation

From pure Baru oil, adulterations carried out with different percentages (% v/v) of soybean oil (SO) were carried out to cover several levels of adulteration. Samples tested here are listed in Table 1. Adulterations of Baru oil with SO were carried out in quintuplicate, and samples were stored in a refrigerator until spectra acquisition. Samples with adulteration levels lower than 10% were produced (2.0, 4.0, 6.0, and 8.0%), however, the FTIR and NIR techniques could not discriminate them from the sample without adulteration, therefore, they were not used for the construction of the models.

Table 1. Percentages of adulteration of oils, spectral range, and number of variables.

Sample	Adulteration levels (%)	Region spectra	Number of variables
NIR	10, 30, 50, 60, 80	1,114.1-1,520.2 nm	116
FTIR	10, 30, 50, 60, 70, 90	1,250 - 850 and 2,451- 2,111 cm^{-1}	742

Infrared spectral analysis

Eight spectra were obtained for each sample (5 samples/adulteration) in each equipment, totaling 40 spectra measured in the portable NIR, and 40 spectra obtained in the FTIR equipment.

Analysis in the near-infrared region was performed using a portable NIR spectrometer from Texas Instruments - Dallas® (DLP NIRscan Model Nano Evaluation Module EVM, USA). The spectrometer has an Indium Gallium Arsenide (InGaAs) detector. Spectra were obtained in a spectral range from 900 to 1,700 nm (11,099 to 5,880 cm^{-1}), with an average resolution of 3.51 nm (17 cm^{-1}), 32 scans, absorbance mode. The software used was NIRscanNano GUI version 1.1.9, developed by Texas Instruments - Dallas®.

Analysis in the mid-infrared region was performed using a Perkin Elmer FTIR spectrometer, in absorption mode. Spectra were obtained in the range from 400 to 4,000 cm^{-1} with a resolution of 4 cm^{-1}, 32 scans, and using a CsI pallet. Blank measurement was achieved using a CsI pastille.

Spectra were obtained as follows: a drop of oil was placed on the pastille surface, and the excess oil was removed with rods of cotton. Afterward, the pastille was placed in the device to obtain the spectrum. At each new sample, the pastille was cleaned with acetone, and the blank was achieved again.

Construction of PLS-DA models

PLS-DA is a supervised classification method, whose objective is to build a classification model in which samples are divided into groups and have as a prerequisite the need for prior knowledge of the sample set to be analyzed. In short, for the construction of PLS-DA models, a calibration is carried out using all variables in relation to a binary vector, in which the property of interest to be determined consists of the samples belonging or not to a particular class. Thus, the class values are assigned to the binary vector, with value 1 referring to samples that will be discriminated, and value 0 assigned to other classes, with a model built for each class. Models with three or more classes are usually modeled as a group against all others in a PLS1-DA approach, i.e. model 1, which differentiates Class 1 from Classes 2 and 3, etc., model 2 which differentiates Class 2 from Classes 1 and 3 and finally, model 3 differentiates Class 3 from Classes 1 and 2 (Ballabio & Consonni, 2013; Brereton & Lloyd, 2014).

For PLS-DA, partial least squares decomposition is performed, simultaneously for matrix X and vector y. The PLS-DA model is built from the matrix X of type i x j (Xi,j), where i lines represent the samples and j columns represent the instrumental variables (absorbance, relative intensity, etc.). The vector y contains the reference values, that is, the binary numerical identification of each sample according to its class, as represented by Equations 1 and 2 (Ballabio & Consonni, 2013; Brereton & Lloyd, 2014).

$$X = \sum_{a-1}^{A} T_a P_a^T + E \qquad (1)$$

$$y = \sum_{a=1}^{A} U_a q_a^T + f \qquad (2)$$

where T_a represents the score vectors, P represents weights. The matrix E and vector f represent errors, that is, both contain the variance not explained by the PLS model for X and y, respectively. Finally, a identifies the number of latent variables, which are positioned to guarantee a smaller dimension to the original variable space and obtain a maximum possible covariance between matrix X and vector y.

For analysis of the obtained spectra and construction of the matrices for the PLS1-DA models, MATLAB software version 7.12 (R2011b) was used, and the PLS-Toolbox version 6.5 package was used for the construction of the models.

Initially, a selection of the spectral region was made, considering the most efficient bands. From the dataset obtained, approximately two-thirds of the spectra were randomly designed for the training set and one-third for the validation set. Thus, the training set consisted of 26 spectra, and the validation set, 14.

Subsequently, the reference class values to the training set were determined. In this step, value 1 was considered to the class to be discriminated against and value 0 to the others. To facilitate visualization, individual markers were determined for each class. The same was done for the validation set. The simultaneous decomposition of the X training set, which corresponds to the spectral data, and the y vector, containing values of each class, was performed. Thus, a model for each analyzed adulteration was obtained.

The preprocessing is a very important step for the construction of multivariate models to attenuate baseline deviations and spectral noise, and to highlight the variation of spectral bands. First-order Savitzky-Golay derivative (using a 15-point window and a second-order polynomial fit), Spectra Normalization, Autoscaling, Mean Centering, Standard Normal Variate (SNV), and Multiplicative Scatter Correction (MSC) were applied as preprocessing (Barnes, Dhanoa, & Lister, 1989; Geladi, MacDougall, & Martens, 1985;

Helland, Næs, & Isaksson, 1995; Jiao, Chen, & Fei, 2020; Rinnan, Berg, & Engelsen, 2009; Savitzky & Golay, 1964). The best preprocessing was selected considering the most efficient separation of samples from each model. The choice of the number of latent variables (LVs) was made based on the observation of the smallest value of the root mean square error of cross-validation (RMSECV).

In order to obtain the best model, both the number of latent variables (LVs) and the Root Mean Square Errors (RMSE) are required. PLS-DA uses the term latent variables to designate the main components that can describe the data without losing information after PLS. During the cross-validation process, the number of LVs that will be used is determined, based on lower RMSECV values (Botelho et al., 2015; Pastore et al., 2011).

Outlier-detection procedure

Outlier exclusion consists of detecting and excluding samples with strong influence and high values of non-modeled residuals in the data matrix. In the present study, a procedure for the detection and exclusion of outliers based on Borin and Poppi (Borin & Poppi, 2004) and addressed by Magalhães et al. (Magalhães et al., 2019) and Martins et al. (Martins, Talhavini, Vieira, Zacca, & Braga, 2017) was adopted. The Outlier exclusion was performed first by applying a method in the training set is based on evaluating Q (residues) and T^2 (Hotelling) statistics, so that, samples that simultaneously present Q residual values and Hotelling's T^2 above the critical values (Figure 1) are considered anomalous by the model and, consequently, excluded from the training and validation set.

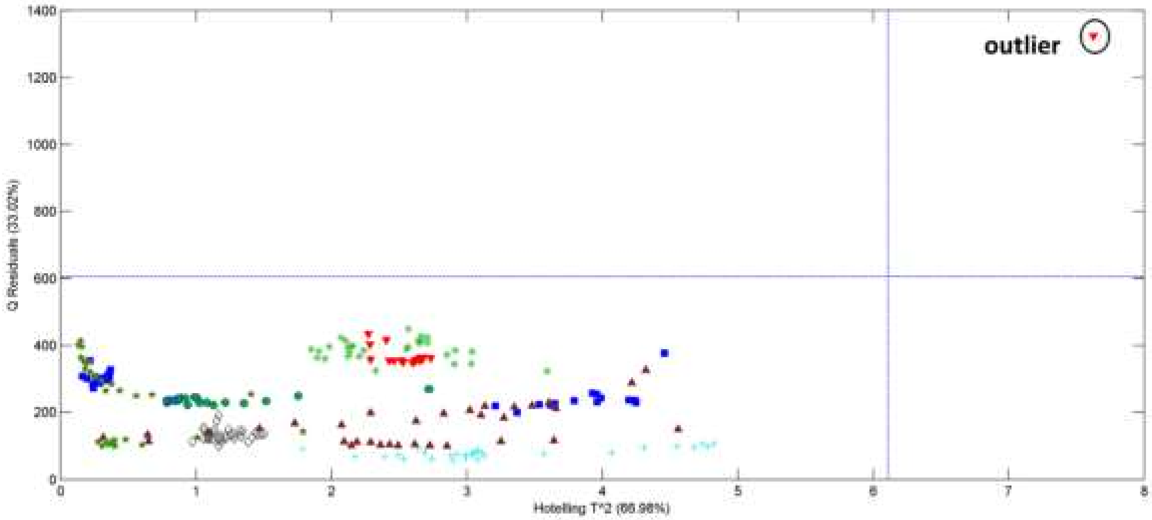

Figure 1. Decomposition residual values (Q) versus the influence (Hotelling's T^2).

Outlier exclusion was performed first by applying a method in the training set. If there are outliers, these samples are excluded and the model is reconstructed after outlier removal. Afterward, the model is submitted to a second exclusion step where the number of excluded outliers is observed and compared to the previous step. If the number is the same, then no exclusion is made and the method is finished. Otherwise, the second model is submitted to a third exclusion step, and this last model is considered the final model for the training set. For the validation set, the outliers are excluded at once, by the application of the method, and the prediction model is reconstructed (Borin & Poppi, 2004; Magalhães et al., 2019; Martins et al., 2017).

Figures of merit

Models were validated based on the calculations of figures of merit, which are statistical parameters used for validation of classification models, and evaluation of their performance. For this phase, it is necessary to know which samples are considered true positives, true negatives, false negatives, and false positives.

A sample is considered a true positive when belonging to its predefined class (target class) after discrimination, and in the final model, it is located above the threshold. The true negative is the opposite: samples belonging to the other class and located below the threshold. The sample described as false positive is the one that does not belong to the target class and even so, it is above the threshold. Furthermore, the false negative considers samples belonging to the target class but below the threshold, classified with different samples (Botelho et al., 2015).

The False Positive Rate (FPR) and the False Negative Rate (FNR) represent the percentage of samples considered as false positive and negative, respectively, in each class, and its calculation is represented by equations 3 and 4 (Botelho et al., 2015):

$$FPR = \frac{FP}{FP + TN} \times 100 \qquad (3)$$

$$FNR = \frac{FN}{FN + TP} \times 100 \qquad (4)$$

From the values found in the two previous equations, the Efficiency Rate (EFR) of the model was calculated by equation 5:

$$EFR = 100 - (FNR + FPR) \qquad (5)$$

Results and discussion

Oil extraction yield

A final volume of 72 mL Baru oil was obtained (20% yield), which is consistent with the literature on the yield of extracting this oil using ethanol at different temperatures, and obtained yield values between 17.43 and 28.95% (Souza, Miranda, Sousa, Vieira, & Coimbra, 2020). Although the Soxhlet method favors oil extraction, the refining steps to eliminate both residues of the solvent used and solids from the seeds, in addition to the toxicity of some solvents, are factors limiting this method (Herrero, Cifuentes, & Ibañez, 2006; Liu et al., 2009).

Infrared spectroscopy

As seen in Figures 2 and 3, Baru and soybean oil have a very similar spectral profile, both in the near- and mid-infrared, due to similarities in their chemical composition. It highlights the need for pre-processing not only to eliminate regions without relevant information but also to smooth out the noise in the raw spectra, which can affect the final efficiency of discriminant models.

Baru and soybean oil NIR spectral profiles (Figure 2) showed two main absorption bands, one from 1,100 to 1,300 nm with an increase in intensity, related to C-H bonds corresponding to the functional groups $-CH_2$, $-CH_3$ and -CH=CH, and another from 1,350 to 1,550 nm, representing a combination of C-H stretching vibrations (Basri et al., 2017; Galtier et al., 2007).

Baru and soybean oil FTIR spectral profiles (Figure 3) showed the main bands from 1,400 to 1,200 cm^{-1}; 1,750 cm^{-1}; between 2,922 and 2,800 cm^{-1}; and at 3,005 cm^{-1}. Angular deformations on CH_2 and CH_3 are responsible for bands in the region between 1,400 and 1,200 cm^{-1}, whereas a band observed at 1,750 cm^{-1} is due to carbonyl stretching. The band observed from 2,922 to 2,800 cm^{-1} is related to C-H stretching in the terminal methylene and methyl groups of triacylglycerol fatty acid chains (Lumakso, Rohman, Riyanto, & Yusof, 2015; Poiana et al., 2015).

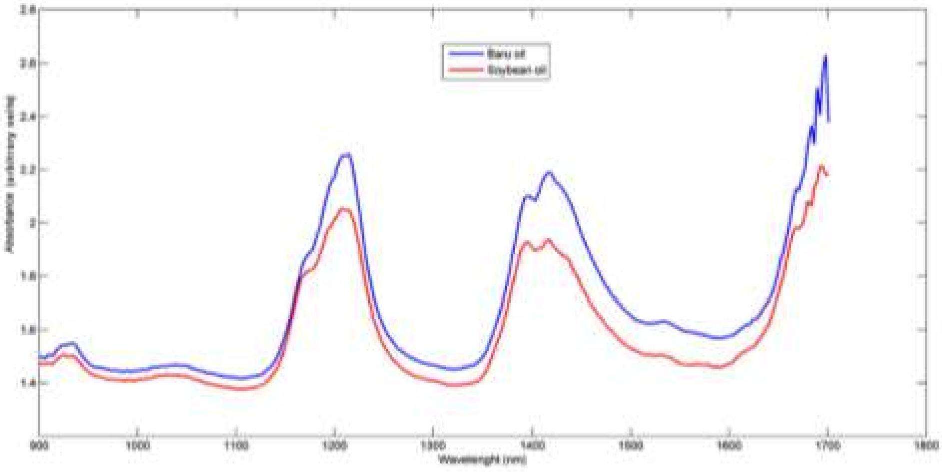

Figure 2. NIR spectra of Baru and soybean oils.

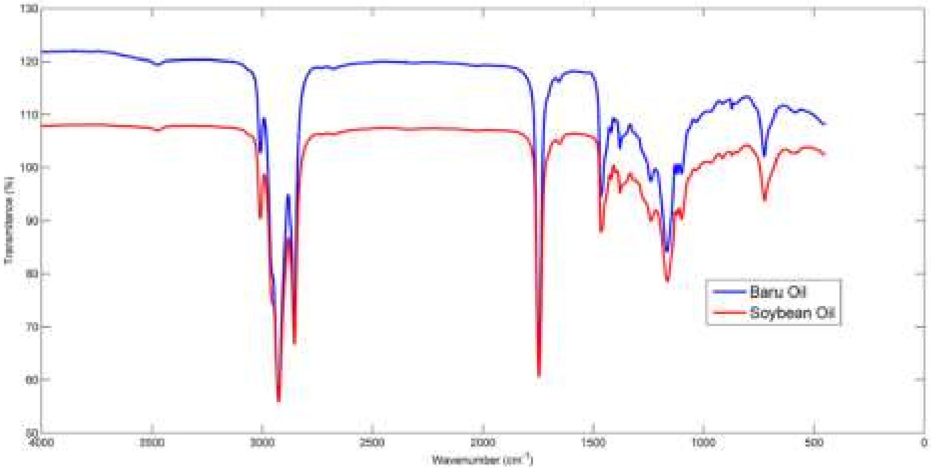

Figure 3. FTIR spectra of Baru and soybean oils.

Development of discrimination models

After spectra acquisition, 15 models were built, 7 models for the set of spectra obtained from the portable NIR, and 8 for the set of spectra obtained using the FTIR equipment.

The dispersion of samples in predicted values for each model is illustrated in Figures 4 and 5. The classes formed by adulteration with soybean oil (SO) were the most dispersed, in compared to the Baru oil (BO) class. This behavior can be associated with a little spectroscopic variability of Baru oil caused by adulteration that is enhanced in PLS-DA.

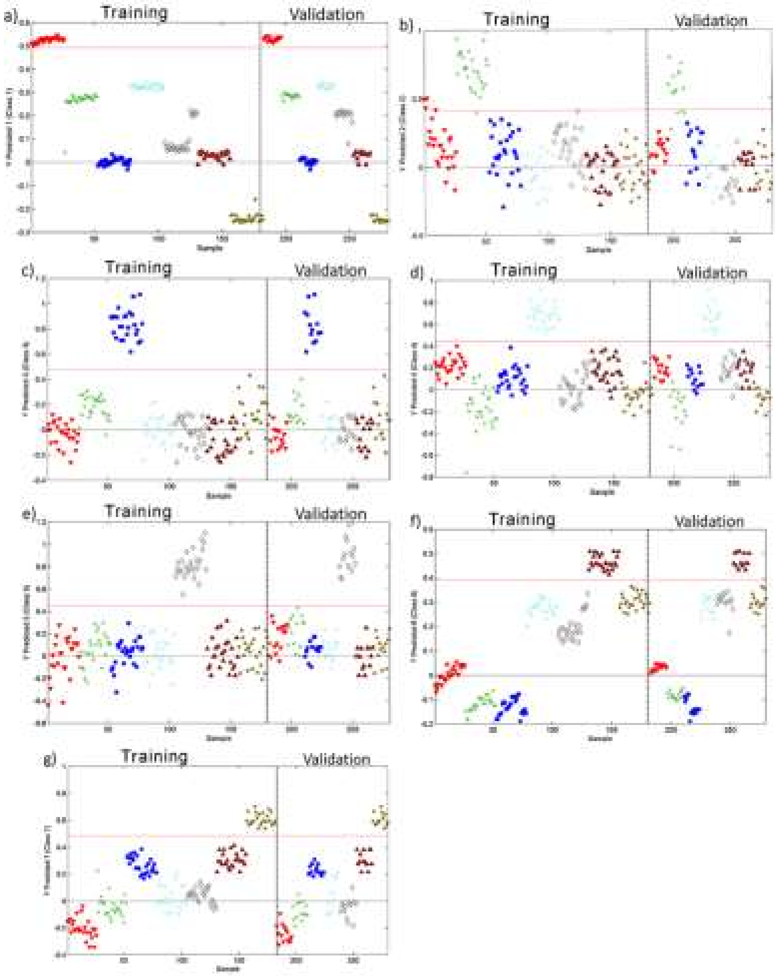

Figure 4. PLS1-DA NIR models for Baru oil adulterated with a) 0% (BO) (▼), b) 10% (*), c) 30% (■), d) 50% (+), e) 60% (◊), f) 80% (▲), g) 100% (SO) (★); threshold (– –).

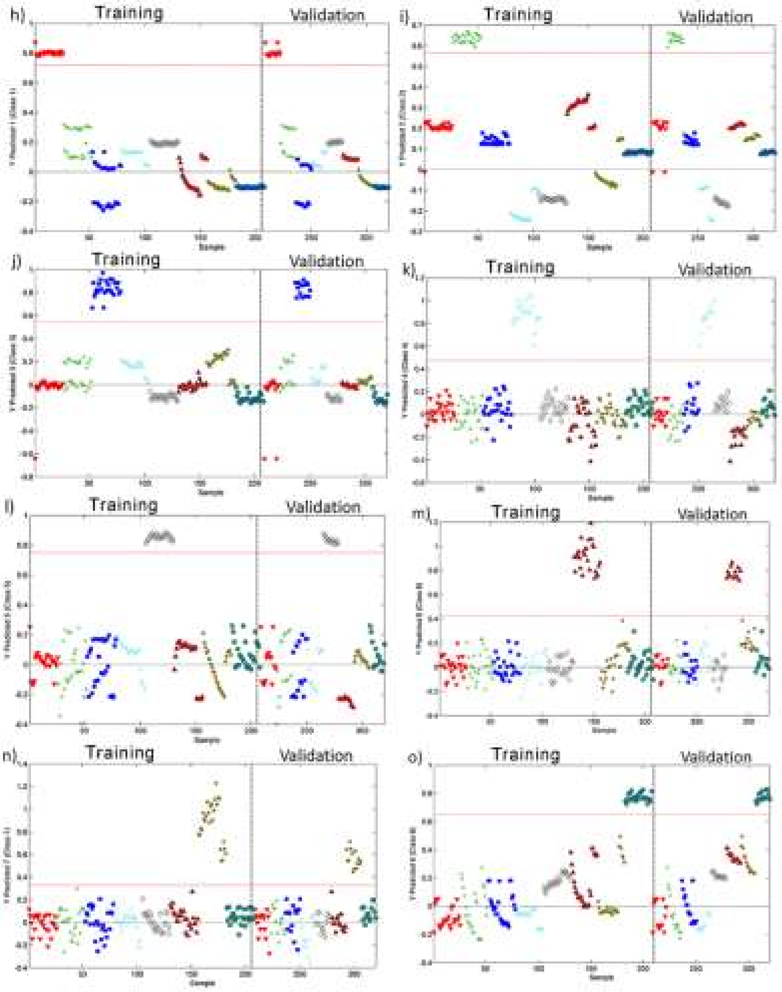

Figure 5. PLS1-DA FTIR models for Baru oil adulterated with h) 0% (▼), i) 10% (*), j) 30% (■), k) 50% (+), l) 60% (◊), m) 70% (▲), n) 90% (★), o) 100% (SO) (●); threshold (– –).

The predicted graph good resolution is explained by the separation of discriminated samples by the threshold, the dotted line determining a threshold value, calculated by the Bayes theorem. The calculation of the Bayesian threshold assumes that the predicted variance of the classes (y vector) follows a distribution similar to that observed in new samples. Thus, to build the PLS1-DA model, samples with predicted values above the threshold are considered to belong to the discriminated class, and samples below it, belong to other classes (Ballabio & Consonni, 2013; Brereton & Lloyd, 2014).

PLS1-DA models in Figures 4 and 5 showed good separation between the discriminated samples, with no classification errors in the models built with the spectra using FTIR. However, some samples from pure Baru oil in the training set were confused with the set discriminated in the 10% adulteration model (Figure 4b) using NIR, which also had misclassifications in the test set. However, other FTIR models did not have this problem.

Table 2 lists the obtained value for the root mean square error of calibration (RMSEC) and cross-validation (RMSECV). The similarity between values of RMSEC and RMSECV indicated that was no over-adjustment of the training and test models, which gives them efficiency, proving the results illustrated in the discrimination graphs. However, these errors represent random and systematic behaviors, thus requiring other indicators to validate the proposed model.

Outlier exclusion

In the second step of model development, the presence of outliers in the training and validation sets, and the calculation of figures of merit were done to improve the built models presented in Figures 6 and 7. Table 3 lists the number of outliers excluded from the sets and the value of the root mean square error of prediction (RMSEP).

For outlier exclusion, the 98% confidence interval was adopted. The best results were found in the validation sets, in which only one detected outliers (class with 50% adulteration in the NIR model). Among the training sets, the best results were observed in FTIR models, given the lower number of outliers.

The RMSEP is the error related to the predictive power of the models, low values demonstrate that the models were efficient in discriminating the samples. The low values obtained with the RMSEP referring to the built models indicate its good predictive power.

Table 2. Parameters of NIR and FTIR models of discrimination for Baru oil.

NIR				FTIR			
Adulteration (%)	LV*	RMSEC	RMSECV	Adulteration (%)	LV*	RMSEC	RMSECV
0 (BO)	2	0.26	0.26	O (BO)	3	0.15	0.16
10	12	0.20	0.23	10	4	0.21	0.22
30	7	0.15	0.17	30	5	0.14	0.14
50	6	0.21	0.24	50	7	0.09	0.10
60	12	0.16	0.22	60	9	0.11	0.11
80	3	0.28	0.28	70	12	0.10	0.12
OS	3	0.24	0.24	90	12	0.09	0.11
				OS	7	0.11	0.12

* Latent Variable

Table 3. Outlier detection and RMSEP values for NIR and FTIR models.

	NIR					FTIR					
Adulteration (%)	10	30	50	60	80	10	30	50	60	70	90
RMSEP	0.23	0.15	0.20	0.16	0.29	0.22	0.11	0.14	0.16	0.14	0.23
Number of outliers in the training set	4	2	1	6	2	4	1	1	0	1	6
Number of outliers in the validation set	0	0	1	0	0	0	0	0	0	0	0

RMSEP: Root mean square error of prediction.

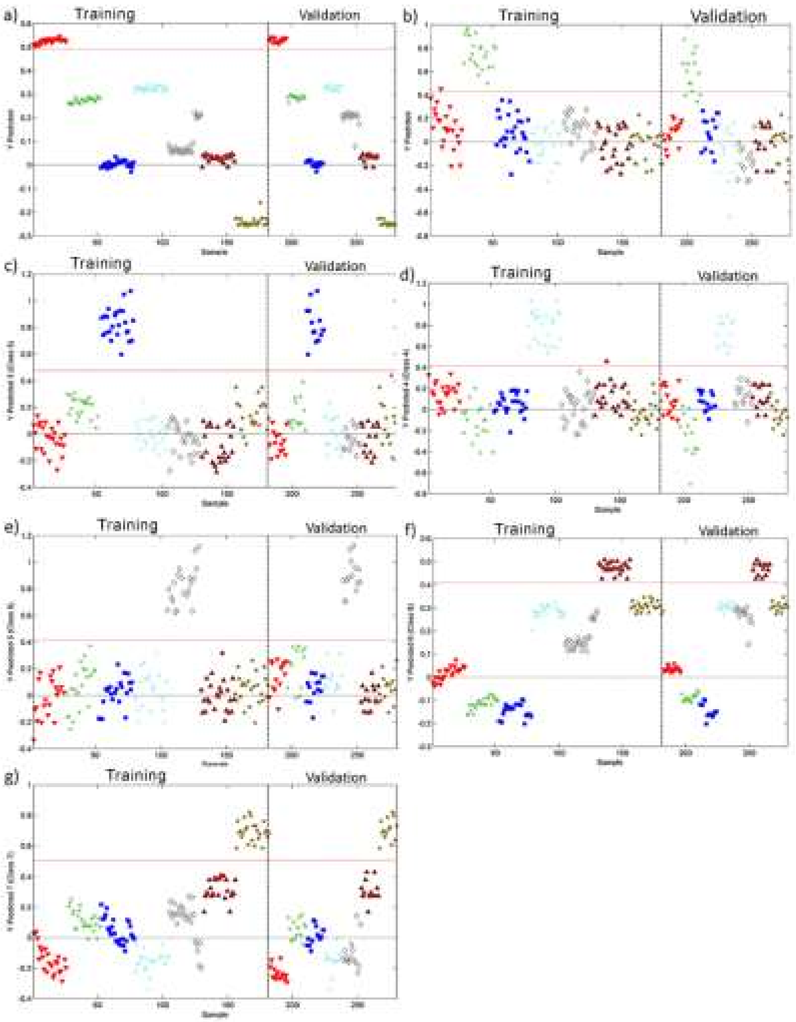

Figure 6. Outlier detection procedure performed on PLS1-DA NIR models for Baru oil adulterated with a) 0% (BO) (▼), b) 10% (*), c) 30% (■), d) 50% (+), e) 60% (◊), f) 80% (▲), g) 100% (SO) (★); threshold (– –).

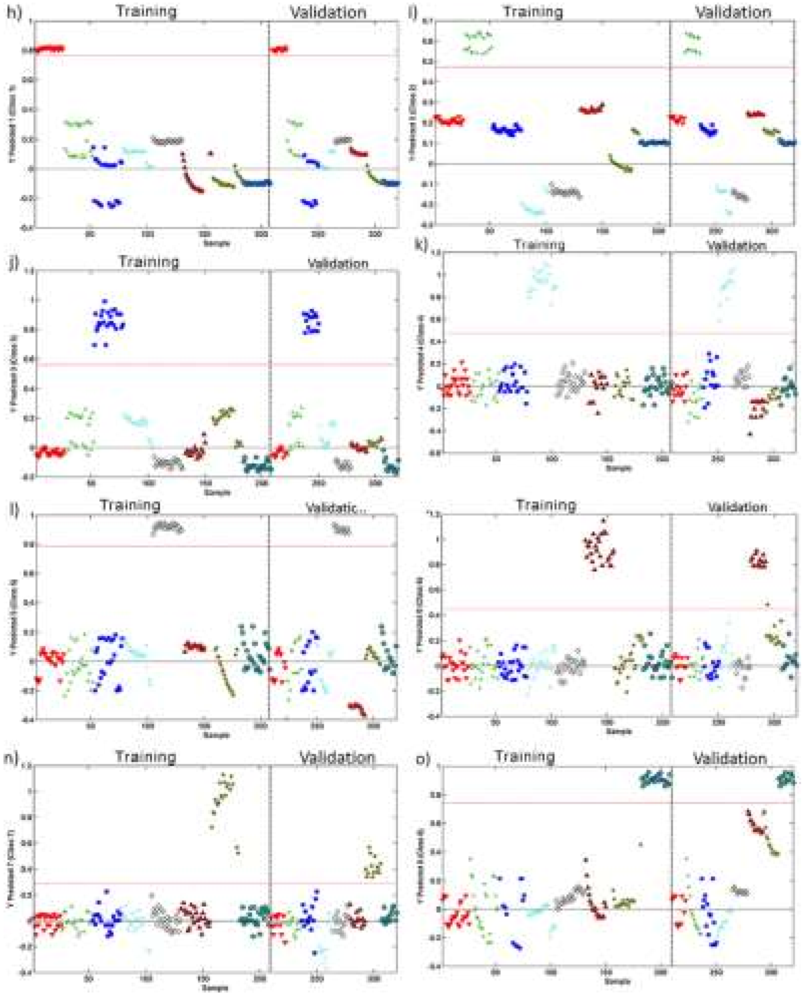

Figure 7. Outlier detection procedure performed on PLS1-DA FTIR models for Baru oil adulterated with h) 0% (▼), i) 10% (*), j) 30% (■), k) 50% (+), l) 60% (◊), m) 70% (▲), n) 90% (★), o) 100% (SO) (●); threshold (– –).

Validation

The values obtained are presented in Table 4. Most models presented a percentage of 0% in relation to FNR (False Negative Rate) and FPR (False Positive Rate), which resulted in efficiency rates with values above 98%. The NIR model with 10% adulteration obtained a 71.43% efficiency rate in the validation set, by presenting 4 samples as False Negative.

Table 4. Figures of merit for NIR and FTIR models with adulterated Baru oil.

	Adulteration (%)	FPR (%)		FNR (%)		EFR (%)	
		T	V	T	V	T	V
FTIR	0 (BO)	0.00	0.00	0.00	0.00	100.00	100.00
	10	0.00	0.00	0.00	0.00	100.00	100.00
	30	0.00	0.00	0.00	0.00	100.00	100.00
	50	0.00	0.00	0.00	0.00	100.00	100.00
	60	0.00	0.00	0.00	0.00	100.00	100.00
	70	0.00	1.05	0.00	0.00	100.00	98.95
	90	0.00	0.00	0.00	0.00	100.00	100.00
	100 (SO)	0.00	0.00	0.00	0.00	100.00	100.00
NIR	0 (BO)	0.00	0.00	0.00	0.00	100.00	100.00
	10	0.70	0.00	0.00	28.57	99,3	71.43
	30	0.00	0.00	0.00	0.00	100.00	100.00
	50	0.72	0.00	0.00	0.00	99.28	100.00
	60	0.00	0.00	0.00	0.00	100.00	100.00
	80	0.00	0.00	0.00	0.00	100.00	100.00
	100 (SO)	0.00	0.00	0.00	0.00	100.00	100.00

FPR: False Positive Rate, FNR: False Negative Rate, EFR: Efficiency Rate, T: Training phase, V: Validation phase.

The analysis of Figures 6 and 7 shows that outlier exclusion method, and calculations of the figures of merit indicated that the models built with the sets of spectra obtained by the FTIR equipment have better performance compared to the portable NIR. This is because the NIR region results in highly overlapped broad bands, which is a limiting factor in the identification and quantification of adulterants (Moreira et al., 2018). Nevertheless, the developed methodologies were able to discriminate the adulterations, obtaining efficient PLS-DA models capable of predicting the presence of this type of fraud in Baru oil.

Conclusion

PLS1-DA discrimination models built with Baru oil adulterated with soybean oil showed efficient responses when evaluated in the efficiency graphs combined with RMSEC, RMSECV, and RMSEP errors.

The validation of discrimination models based on the calculations of figures of merit provided evidence of high-efficiency rates while excluding a small number of samples using the outlier exclusion method. The application of spectroscopic techniques consists of a simple and fast methodology, increasingly widespread in academia and laboratory routines specialized at detecting frauds. The use of portable NIR allowed to explore the use of equipment still little applied for the detection of oil adulteration.

Acknowledgements

Brazilian National Council for Scientific and Technological Development – CNPq, Coordination for the Improvement of Higher Education Personnel – CAPES/ Brazil, Federal University of Tocantins – UFT/ Brazil and Research Productivity Scholarship by the Tocantins Research Support Foundation – FAPT.

References

Al-Ahmed, N., Alsowaidi, S., & Vadas, P. (2008). Peanut Allergy: An Overview. *The Journal of Allergy and Clinical Immunology*, *4*, 139-143. DOI: https://doi.org/10.1186/1710-1492-4-4-139

Ballabio, D., & Consonni, V. (2013). Classification tools in chemistry. Part 1: Linear models. PLS-DA. *Analytical Methods*, *5*, 3790-3798. DOI: https://doi.org/10.1039/c3ay40582f

Barnes, R. J., Dhanoa, M. S., & Lister, S. J. (1989). Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra. *Applied Spectroscopy*, *43*(5), 772-777. DOI: https://doi.org/10.1366/0003702894202201

Basri, K. N., Hussain, M. N., Bakar, J., Sharif, Z., Khir, M. F. A., & Zoolfakar, A. S. (2017). Classification and quantification of palm oil adulteration via portable NIR spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 173*, 335-342. DOI: https://doi.org/10.1016/j.saa.2016.09.028

Borin, A., & Poppi, R. (2004). Multivariate quality control of lubricating oils using Fourier transform infrared Spectroscopy. *Journal of the Brazilian Chemical Society*, *15*(4), 570-576. DOI: http://doi.org/10.1590/S0103-50532004000400020

Botelho, B. G., Reis, N., Oliveira, L. S., & Sena, M. M. (2015). Development and analytical validation of a screening method for simultaneous detection of five adulterants in raw milk using mid-infrared spectroscopy and PLS-DA. *Food Chemistry*, *181*, 31-37. DOI: https://doi.org/10.1016/j.foodchem.2015.02.077

Brereton, R., & Lloyd, G. (2014). Partial least squares discriminant analysis: Taking the magic away. *Journal of Chemometrics*, *28*(4), 213-225. DOI: http://doi.org/10.1002/cem.2609

Câmara, A. B. F., de Carvalho, L. S., de Morais, C. L. M., de Lima, L. A. S., de Araújo, H. O. M., de Oliveira, F. M., & de Lima, K. M. G. (2017). MCR-ALS and PLS coupled to NIR/MIR spectroscopies for quantification and identification of adulterant in biodiesel-diesel blends. *Fuel*, *210*, 497-506. DOI: https://doi.org/10.1016/j.fuel.2017.08.072

dos Santos, P., de Aguiar, A. C., Viganó, J., Boeing, J. S., Visentainer, J. V., & Martínez, J. (2016). Supercritical CO2 extraction of cumbaru oil (Dipteryx alata Vogel) assisted by ultrasound: Global yield, kinetics and fatty acid composition. *The Journal of Supercritical Fluids*, *107*, 75-83. DOI: https://doi.org/10.1016/j.supflu.2015.08.018

Fernandes, D., Freitas, J., Czeder, L., & Naves, M. (2010). Nutritional composition and protein value of the baru (Dipteryx alata Vog.) almond from the Brazilian Savanna. *Journal of the Science of Food and Agriculture*, *90*(10), 1650-1655. DOI: https://doi.org/10.1002/jsfa.3997

Frank, l. E., & Friedman, J. H. (1993). A Statistical View of Some Chemometrics Regression Tools. *Technometrics*, *35*(2), 109-135. DOI: https://doi.org/10.2307/1269658

Galtier, O., Dupuy, N., Le Dréau, Y., Ollivier, D., Pinatel, C., Kister, J., & Artaud, J. (2007). Geographic origins and compositions of virgin olive oils determinated by chemometric analysis of NIR spectra. *Anal. Chim. Acta*, *595*(1-2), 136-144. DOI: https://doi.org/10.1016/j.aca.2007.02.033

Geladi, P., MacDougall, D., & Martens, H. (1985). Linearization and Scatter-Correction for Near-Infrared Reflectance Spectra of Meat. *Applied Spectroscopy*, *39*(3), 491-500.

Helland, I. S., Næs, T., & Isaksson, T. (1995). Related versions of the multiplicative scatter correction method for preprocessing spectroscopic data. *Chemometrics and Intelligent Laboratory Systems*, *29*(2), 233-241. DOI: https://doi.org/10.1016/0169-7439(95)80098-T

Herrero, M., Cifuentes, A., & Ibañez, E. (2006). Sub- and supercritical fluid extraction of functional ingredients from different natural sources: Plants, food-by-products, algae and microalgae: A review. *Food Chemistry*, *98*, 136-148. DOI: https://doi.org/10.1016/j.foodchem.2005.05.058

Jiang, H., & Chen, Q. (2019). Determination of Adulteration Content in Extra Virgin Olive Oil Using FT-NIR Spectroscopy Combined with the BOSS–PLS Algorithm. *Molecules*, *24*(11), 2134-2144. DOI: https://doi.org/10.3390/molecules24112134

Jiao, Y., Li, Z., Chen, X., & Fei, S. (2020). Preprocessing methods for near-infrared spectrum calibration. *Journal of Chemometrics*, *34*(11), 1-19. DOI: http://doi.org/10.1002/cem.3306

Liu, S., Yang, F., Zhang, C., Ji, H., Hong, P., & Deng, C. (2009). Optimization of process parameters for supercritical carbon dioxide extraction of Passiflora seed oil by response surface methodology. *The Journal of Supercritical Fluids*, *48*, 9-14. DOI: https://doi.org/10.1016/j.supflu.2008.09.013

Lôbo, I. P., Ferreira, S. L. C., & Cruz, R. S. d. (2009). Biodiesel: parâmetros de qualidade e métodos analíticos. *Química Nova*, *32*(6), 1596-1608. DOI: https://doi.org/10.1590/S0100-40422009000600044

Lumakso, F., Rohman, A., M, H., Riyanto, S., & Yusof, F. (2015). Detection and quantification of soybean and corn oils as adulterants in avocado oil using fourier transform mid infrared (FT-MIR) spectroscopy aided with multivariate calibration. *Jurnal Teknologi*, *77*, 251-255. DOI: https://doi.org/10.11113/jt.v77.3640

Magalhães, L. d. O., Arantes, L. C., & Braga, J. W. B. (2019). Identification of NBOMe and NBOH in blotter papers using a handheld NIR spectrometer and chemometric methods. *Microchemical Journal*, *144*, 151-158. DOI: https://doi.org/10.1016/j.microc.2018.08.051

Martins, A. R., Talhavini, M., Vieira, M. L., Zacca, J. J., & Braga, J. W. B. (2017). Discrimination of whisky brands and counterfeit identification by UV–Vis spectroscopy and multivariate data analysis. *Food Chemistry*, *229*, 142-151. DOI: https://doi.org/10.1016/j.foodchem.2017.02.024

Moreira, A., Machado, A., Almeida, F., & Braga, J. (2018). Rapid Purity Determination of Copaiba Oils by a Portable NIR Spectrometer and PLSR. *Food Analytical Methods*, *11*, 1867-1877. DOI: https://doi.org/10.1007/s12161-017-1079-8

Nunes, K. M., Andrade, M. V. O., Almeida, M. R., & Sena, M. M. (2020). A soft discriminant model based on mid-infrared spectra of bovine meat purges to detect economic motivated adulteration by the addition of non-meat ingredients. *Food Analytical Methods*, *13*(9), 1699-1709. DOI: https://doi.org/10.1007/s12161-020-01795-3

Oliveira-Alves, S. C., Pereira, R. S., Pereira, A. B., Ferreira, A., Mecha, E., Silva, A. B., ... Bronze, M. R. (2020). Identification of functional compounds in baru (Dipteryx alata Vog.) nuts: Nutritional value, volatile and phenolic composition, antioxidant activity and antiproliferative effect. *Food Research International*, *131*, 109026. DOI: https://doi.org/10.1016/j.foodres.2020.109026

Pastore, T. C. M., Braga, J. W. B., Coradin, V. T. R., Magalhães, W. L. E., Okino, E. Y. A., Camargos, J. A. A., ... Davrieux, F. (2011). Near infrared spectroscopy (NIRS) as a potential tool for monitoring trade of similar woods: Discrimination of true mahogany, cedar, andiroba, and curupixá. *Holzforschung*, *65*, 73-80. DOI: https://doi.org/10.1515/hf.2011.010

Pineli, L. L. O., de Aguiar, L. A., de Oliveira, G. T., Botelho, R. B. A., Ibiapina, M. d. D. F. P., de Lima, H. C., & Costa, A. M. (2015). Use of Baru (Brazilian Almond) Waste from Physical Extraction of Oil to Produce Gluten Free Cakes. *Plant Foods for Human Nutrition*, *70*, 50-55. DOI: http://doi.org/10.1007/s11130-014-0460-7

Poiana, M.-A., Alexa, E., Munteanu, M.-F., Gligor, R., Moigradean, D., & Mateescu, C. (2015). Use of ATR-FTIR spectroscopy to detect the changes in extra virgin olive oil by adulteration with soybean oil and high temperature heat treatment. *Open Chemistry*, *13*, 689-698. DOI: https://doi.org/10.1515/chem-2015-0110

Rinnan, Å., Berg, F. v. d., & Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *Trends in Analytical Chemistry*, *28*(10), 1201-1222. DOI: https://doi.org/10.1016/j.trac.2009.07.007

Sano, E. E., Rodrigues, A. A., Martins, E. S., Bettiol, G. M., Bustamante, M. M. C., Bezerra, A. S., … Bolfe, E. L. (2019). Cerrado ecoregions: A spatial framework to assess and prioritize Brazilian savanna environmental diversity for conservation. *Journal of Environmental Management*, *232*, 818-828. DOI: https://doi.org/10.1016/j.jenvman.2018.11.108

Savitzky, A., & Golay, M. J. E. (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, *36*(8), 1627-1639. DOI: https://doi.org/10.1021/ac60214a047

Silva, V., Talhavini, M. r., Peixoto, I. C. F., Zacca, J. J., Maldaner, A. O., & Braga, J. W. B. (2014). Non-destructive identification of different types and brands of blue pen inks in cursive handwriting by visible spectroscopy and PLS-DA for forensic analysis. *Microchemical Journal*, *116*, 235-243. DOI: http://doi.org/10.1016/j.microc.2014.05.013

Sousa, A., Fernandes, D., Alves, A., Freitas, J., & Naves, M. (2011). Nutritional quality and protein value of exotic almonds and nut from the Brazilian Savanna compared to peanut. *International Food Research Journal*, *44*(7), 2319-2325. DOI: http://doi.org/10.1016/j.foodres.2011.02.013

Souza, A. L. S., Miranda, J. S., Sousa, R. C. S., Vieira, B. B., & Coimbra, J. S. R. (2020). Extraction of baru almond oil using alternative solvents to hexane: ethanol and isopropanol. *International Journal of Research - GRANTHAALAYAH*, *8*(8), 356-371. DOI: https://doi.org/10.29121/granthaalayah.v8.i8.2020.1197

Wójcicki, K., Khmelinskii, I., Sikorski, M., Caponio, F., Paradiso, V., Summo, C., … Sikorska, E. (2015). Spectroscopic Techniques and Chemometrics in Analysis of Blends of Extra Virgin With Refined and Mild Deodorized Olive Oils. *European Journal of Lipid Science and Technology*, *117*, 92-102. DOI: https://doi.org/10.1002/ejlt.201300402

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, *2*(1-3), 37-52. DOI: https://doi.org/10.1016/0169-7439(87)80084-9