

Identification of cocoa bean quality by near infrared spectroscopy and multivariate modeling

Identificação da qualidade de amêndoas de cacau por espectroscopia de infravermelho próximo e modelagem multivariada

Identificación de la calidad del grano de cacao mediante espectroscopia de infrarrojo cercano y modelización multivariante

Received: 11/03/2021 | Reviewed: 11/10/2021 | Accept: 11/11/2021 | Published: 11/20/2021

Acsa Santos Batista

ORCID: <https://orcid.org/0000-0003-0071-302X>
University of Southwestern Bahia, Brazil
E-mail: acsang.alimentos@gmail.com

Thinara de Freitas Oliveira

ORCID: <https://orcid.org/0000-0001-9950-9322>
University of Southwestern Bahia, Brazil
E-mail: thinarafreitas@live.com

Ivan de Oliveira Pereira

ORCID: <https://orcid.org/0000-0003-1344-2994>
Federal Institute Baiano, Brazil
E-mail: pereiraio@hotmial.com

Leandro Soares Santos

ORCID: <https://orcid.org/0000-0001-7431-318X>
University of Southwestern Bahia, Brazil
E-mail: leosoassantos@yahoo.com.br

Abstract

Cocoa is a commodity responsible for the income of millions of people and the manufacture of several important products for the food, pharmaceutical, and cosmetic industries. Its quality is associated with several factors involved in the processing steps, mainly in fermentation and drying. The objective of this study was to evaluate the application of near-infrared spectroscopic data associated with multivariate analysis to classify cocoa beans according to their quality and predict attributes such as pH and total acidity by PLS-DA and PLS, respectively. The pH values (4.4-6.7) and total acidity (6.12-29.9) were determined by conventional methods. The PLS-DA proved to be effective in differentiating the classes of cocoa samples with superior and inferior quality, presenting in the validation 100% and 71.43% correct cocoa bean classification with inferior Quality and Higher Quality, respectively. The models obtained by PLS presented satisfactory parameters, being classified as having moderate practical utility and excellent predictive capacity for pH and moderate practical utility and reasonable predictive capacity for total acidity. Thus, the potential of the NIRS technology associated with chemometrics was found and showed efficiency in the classification and prediction of attributes in cocoa beans.

Keywords: Quality control; NIR spectroscopy; PLS-DA; PLS; Chemometry.

Resumo

O cacau é uma commodity responsável pela renda de milhões de pessoas e pela fabricação de diversos produtos importantes para a indústria alimentícia, farmacêutica, cosmética. A sua qualidade está associada a vários fatores envolvidos nas etapas do beneficiamento, principalmente na fermentação e secagem. A espectroscopia no infravermelho próximo (NIR) associada à análise estatística multivariada mostrou eficiência na classificação e predição de atributos em grãos de cacau. O objetivo deste estudo foi avaliar a aplicação de dados espectroscópicos no infravermelho próximo associados à análise multivariada para classificar os grãos de cacau de acordo com sua qualidade e prever atributos como pH e acidez total por PLS-DA e PLS, respectivamente. Os valores de pH (4,4-6,7) e acidez total (6,12-29,9) foram determinados pelos métodos convencionais. O PLS-DA mostrou-se eficaz em diferenciar as classes de amostras de cacau com qualidade superior e inferior, apresentando na validação 100% e 71,43% de classificação correta das amêndoas de cacau com qualidade inferior e qualidade superior, respectivamente. Os modelos obtidos por PLS apresentaram parâmetros satisfatórios, sendo classificados como de utilidade prática moderada e excelente capacidade preditiva para pH e utilidade prática moderada e capacidade preditiva razoável para acidez total. Desta forma, o potencial da tecnologia NIRS associada com a quimiometria foi constatado e pode ser utilizada para classificação das amostras de acordo com a sua qualidade e para predição dos atributos de qualidade das

amêndoas de cacau.

Palavras-chave: Controle de qualidade; Espectroscopia NIR; PLS-DA; PLS; Quimiometria.

Resumen

El cacao es un producto básico responsable de los ingresos de millones de personas y de la fabricación de varios productos importantes para las industrias alimentaria, farmacéutica y cosmética. Su calidad está asociada a varios factores que intervienen en las fases de elaboración, principalmente en la fermentación y el secado. La espectroscopia de infrarrojo cercano (NIR) asociada al análisis estadístico multivariante mostró su eficacia en la clasificación y predicción de atributos en los granos de cacao. El objetivo de este estudio fue evaluar la aplicación de los datos espectroscópicos del infrarrojo cercano asociados al análisis multivariante para clasificar los granos de cacao según su calidad y predecir atributos como el pH y la acidez total mediante PLS-DA y PLS, respectivamente. Los valores de pH (4,4-6,7) y de acidez total (6,12-29,9) se determinaron por métodos convencionales. El PLS-DA demostró ser eficaz en la diferenciación de las clases de muestras de cacao con calidad superior e inferior, presentando en la validación un 100% y un 71,43% de clasificación correcta de los granos de cacao con calidad inferior y superior, respectivamente. Los modelos obtenidos por PLS presentaron parámetros satisfactorios, siendo clasificados como de moderada utilidad práctica y excelente capacidad predictiva para el pH y de moderada utilidad práctica y razonable capacidad predictiva para la acidez total. Así, se verificó el potencial de la tecnología NIRS asociada a la quimiometría, que puede utilizarse para clasificar las muestras según su calidad y para predecir los atributos de calidad de los granos de cacao.

Palabras clave: Control de calidad; Espectroscopia NIR; PLS-DA; PLS; Quimiometria.

1. Introduction

The cacao tree is a plant of the family Sterculiaceae, genus *Theobroma*, of the species *Theobroma cacao* L. that grows in tropical regions around the world. Although this species is native to central and northern South America, almost 70% of the world's crop is currently produced in West Africa (Diomande et al., 2015). The fruit of the cacao tree is cocoa and it is divided into a shell, pulp, and seeds. The seed is the part of greatest interest to the chocolate manufacturing industries (Sunoj et al., 2016). Cocoa is a commodity responsible for the income of millions of people and the manufacture of several important products for the food, pharmaceutical, and cosmetics industries, highlighting the need to promote discussions about the quality of cocoa beans as a way to foster alternatives that increase the profitability of cocoa farms and valorize the raw material (Beg et al., 2017).

The process from cocoa beans to the production of their derivatives is complex. Before the cocoa beans can be marketed and processed into their final industrial products, they undergo post-harvest processing that comprises the steps of pod opening and berry removal, fermentation for 6 to 7 days, and drying to a moisture content of 7% (Kadow et al., 2013). The quality of cocoa is associated with several factors involved in the processing stages. When there is control, especially in fermentation and drying, it is possible to obtain a quality product. This quality is due to the occurrence of some changes, such as acidification and temperature increase, which are necessary for the reactions to process satisfactorily and promote the development of characteristics pertinent to cocoa and its derivatives (Krähmer et al., 2015; Oliveira et al., 2016).

During the early fermentation stages, the cocoa beans are purple due to the presence of anthocyanin pigments. During fermentation, these pigments are broken down and form other condensation products, which result in a brownish-brown colour (Afoakwa et al., 2013). The Fermentation Index (FI) is associated with kernel quality and is obtained by cutting the kernels lengthwise and counting the proportion of purple and brown coloured kernels, among other external and internal characteristics of a representative dry sample of 300 kernels. This process is also known as the cut test (CoEx, 2017). The FI comprises the percentage of brown and partly brown cocoa beans and is an indication of the quality of the beans in relation to the conduct of fermentation and consequently the content of bioactive compounds in cocoa (Crafack et al., 2014). The higher the FI, the higher the technological quality of the cocoa beans.

There are a number of conventional methods to determine the biomarkers and centesimal composition of cocoa beans and indicate the quality of this product, resulting in a large number of analyses to be performed and statistically evaluated

(Lohumi et al., 2015). These methods are time-consuming, expensive, and destructive to the sample, highlighting the need for a fast, inexpensive, and non-destructive tool to evaluate these quality parameters. Fourier transform near-infrared spectroscopy (FT-NIR), associated with chemometrics, may be a suitable strategy as it is considered to be fast, sensitive, non-invasive, non-destructive, relatively low cost, providing a large amount of information with only one test, as it requires little or no sample pre-treatment. It is clean, since it eliminates the use of chemical reagents and avoids the generation of waste that is harmful to the environment, and the equipment is relatively easy to handle (Lima et al., 2020). Infrared spectroscopy and chemometrics are techniques that are always allied, because the spectra have wide and overlapping bands and multivariate analysis becomes a valuable tool to remove the information contained therein, allowing the identification and quantification of various parameters in different matrices (Souza, 2013).

NIR combined with chemometrics have been commonly used to determine the composition of cocoa products. (Hernández-Hernández et al., 2021) highlighted NIR as a fast and easy method to identify cocoa genotypes according to the content of bioactive compounds. Veselá et al. (2007) used NIR to determine the fat, nitrogen, and moisture content of the cocoa powder. Whitacre et al. (2003) detected protein, fat, starch, and proanthocyanidins in cocoa. Barbin et al. (2018) highlighted the potential of NIR for the classification and differentiation of cocoa varieties and for predicting the chemical composition of the spectra obtained for intact and ground cocoa beans. Álvarez et al. (2012) detected theobromine and epicatechin in cocoa using NIR. However, the feasibility of employing NIR to classify cocoa beans according to their quality has not yet been fully explored.

The objective of this study was to evaluate the application of near-infrared spectroscopic data associated with multivariate analysis to classify cocoa beans according to their quality and predict attributes such as pH and titratable acidity by PLS-DA and PLS, respectively.

2. Methodology

2.1 Cocoa bean samples

A total of 78 samples of bulk cocoa beans (approximately 5 kg), collected from April to August in the cocoa producing regions of Bahia, were used in conducting this research.

2.2 Chemical and physical-chemical characterization of cocoa beans

A sample (300 cocoa beans) was collected for evaluation by the longitudinal cut for commercial classification as recommended by CoEx (2017) and the Technical Regulation of Cocoa Almond. Then, the samples were classified into two groups according to their quality: Higher Quality (HQ) and Inferior Quality (IQ). For a sample to be considered HQ it should have a fermentation index (percentage of brown and partially brown almonds) equal to or greater than 60% and defect limits of 4% and 5% for mold and slate, respectively, in addition to the absence of smoke. If they did not meet these criteria, they were classified as IQ.

For the physical-chemical and near infrared spectroscopic analyses, samples were dried and ground in a portable grain grinder (model DCG 20). Physicochemical analyses were performed according to the methods described by AOAC (2010), being pH by method 981.12 and total acidity by 942.15.

2.3 Near-infrared spectroscopy (NIRS)

Spectra were collected in an NIR spectrometer (SpectraStar 2500XL, Unity Scientific, Brookfield, CT, USA) equipped with a tungsten halogen lamp as light source, an indium-gallium-arsenide (InGaAs) detector. The signals were generated in reflectance (% R) mode and transformed into absorbance by using $\log 1/R$. Samples were poured into the sample

cup and scanned over the range of 1100–2500 nm at 1 nm intervals. During collecting, the temperature was kept at about 25°C in the laboratory. The Unity InfoStar V3.11.3 software was used for spectrometer configuration, control, and data acquisition (Lima et al., 2020; Santos et al., 2021). The spectral regions and the main functional groups of the samples analyzed in the NIRS were identified in the literature (Sunoj et al., 2016).

2.4 Chemometric analysis

The data obtained were analyzed using Chemoface software version 1.63 (Nunes et al., 2012) for Partial Least Squares-Discriminant Analysis (PLS-DA) and Partial Least Squares (PLS). For statistical analysis, spectral data obtained in the NIR were used with and without mathematical methods of data pre-processing (None, Standard Normal Variate - SNV and Multiplicative Scatter Correction - MSC). The data were pre-processed using the Chemoface program.

2.4.1 Partial Least Squares-Discriminant Analysis (PLS-DA)

The PLS-DA chemometric analysis is based on the conventional PLS regression method, in which two matrices are used, a prediction matrix ($I \times J$) X and another response matrix ($I \times K$) Y that presents the categorical variables and that describe class participation. K is the number of classes and if only two classes are considered, the matrix Y is reduced to a vector y ($I \times 2$). When developing the PLS regression, the response value Y_{pred} is predicted for a new sample. The decision is made by comparing Y_{pred} with certain categorical variables in Y , and the sample is ranked based on a minimum distance value between Y and Y_{pred} (Cruz-Tirado et al., 2020).

PLS-DA was performed using Chemoface software version 1.63 (Nunes et al., 2012). The data were divided into two groups using the Kennard-Stone (KS) algorithm (Lopes et al., 2022). A set consisting of 70% of the samples from each class was used to optimize the classification model by a single cross-validation. The other 30% of the data was used as an external validation set.

2.4.2 Partial Least Squares (PLS)

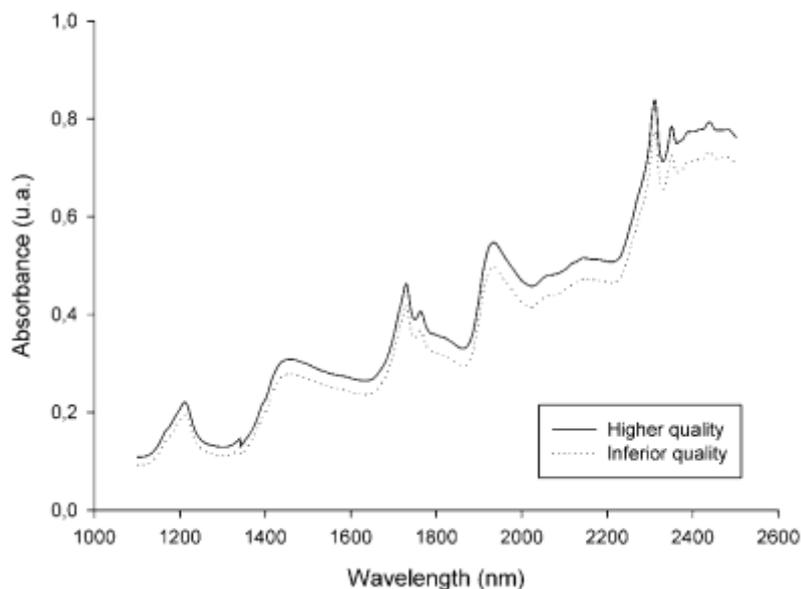
Partial least squares (PLS) is the multivariate tool commonly used to develop a calibration model to relate the information of interest of a sample to its spectrum. The data obtained with NIR were used to build the regression model, which used the information from the data matrix X (full spectrum) and the response matrix Y (pH and total acidity values) to obtain new variables, which were called latent variables, components or factors. The absorbance values from the full spectrum were used as variables (Anyidoho et al., 2021; Hayati et al., 2021; Santos et al., 2021).

The number of latent variables and the performance of the obtained PLS models were evaluated considering the correlation coefficient (R), the Root-Mean-Square Error (RMSE), the ratio of performance to deviation (RPD) and the range error ratio (RER) (Lima et al., 2020). The analysis was performed using Chemoface software version 1.63.

3. Results and Discussion

The spectra presented in Figure 1 were obtained from the mean absorbance of the NIR of cocoa beans of the higher quality and inferior quality. It can be observed that the spectral behavior of the samples is similar. What differentiates them in each peak formed is the absorbance that indicates a greater or lesser amount of the compound responsible for the vibration generated.

Figure 1. Mean absorbance spectrum of cocoa beans of higher quality and inferior quality.



Source: Authors.

The PLS-DA technique is capable of generating functions or models that classify the samples into their relevant classes. The performance of the developed models was evaluated by their ability to separate the samples into their classes (higher or inferior). Table 1 presents the percentage of the correctness of the cocoa sample classes by the developed model with eight latent variables.

It can be seen that there was an excellent classification rate of the samples in their respective classes in the training stage. In the validation stage, the generalization capacity of the model to correctly classify new samples can be verified. The technique can be used to discriminate new cocoa samples and to define the best destination for their application. Cocoa beans in the high-quality group can be used to make chocolate, the most demanded cocoa product, which requires properly processed cocoa beans, with low acidity, bitterness, and astringency. Inferior quality cocoa beans, on the other hand, will present higher intensity in these aspects and thus provide increased nutritional value due to the amount of bioactive antioxidant compounds in their composition.

Table 1. Classification of higher and inferior quality cocoa samples by PLS-DA from NIR data.

Stage	Factors	Model	Higher Quality	Inferior Quality
Training	Samples	53	30	23
	Samples correctly classified	43	26	17
	% of success	88.68	90	86.96
Validation	Samples	25	14	11
	Samples correctly classified	21	10	11
	% of success	84	71,43	100

Source: Authors.

Table 2 shows the average pH and total acidity values of the cocoa beans. The fermentation stage promotes an increase in pH and consequently a reduction in acidity due to biochemical reactions which occur in this process. Thus, pH values can be used as an indication of the quality of cocoa processing. Sunoj et al. (2016) found values between 4.26 and 6.13

which were similar to those obtained in the present study. Both results were close to pH 5.0, which indicates that the samples were subjected to the proper fermentation process (Barrientos et al., 2019).

Variability was observed in the total acidity of the cocoa samples, thus showing the representativeness of these values for multivariate modeling. During fermentation, microorganisms produce acids that raise acidity and lower the pH values. The values obtained in the present study are similar to those found by Peláez et al. (2016) and Sunoj et al. (2016).

Table 2. Values ph and total acidity obtained by the reference methods for cocoa bean samples.

Quality parameters	Max	Min	Mean	SD
pH	6.777	4.447	5.366	0.514
TA (mEqNaOH/ 100g)	29.990	6.128	14.959	5.440

N=78 samples. Max= maximum value; Min= minimum value; SD= standard deviation; TA= Total Acidity.
Source: Authors.

The Partial Least Squares (PLS) technique was used in the development of mathematical modeling able to predict the total acidity and pH variables of cocoa samples as a function of near-infrared spectroscopy. These models were adjusted with and without mathematical data preprocessing methods (None, Standard Normal Variate - SNV, Multiplicative Scatter Correction - MSC). The fit parameters of the optimized models for predicting pH and total acidity variables are presented in Table 3.

Table 3. PLS calibration parameters of the NIR spectra to predict pH and total acidity values.

Spectra correction	Calibration parameters					Validation parameters			
	LV	R	RMSE	RER	RPD	R	RMSE	RER	RPD
<i>pH</i>									
None	8	0.859	0.273	8.532	1.883	0.687	0.384	6.063	1.338
	12	0.923	0.205	11.360	2.508	0.790	0.313	7.451	1.645
	20	0.996	0.050	46.881	10.349	0.794	0.275	8.479	1.872
MSC	6	0.839	0.291	8.018	1.770	0.662	0.410	5.682	1.254
	11	0.933	0.192	12.129	2.678	0.778	0.330	7.071	1.561
	20	0.998	0.035	66.571	14.696	0.793	0.281	8.283	1.829
SNV	7	0.855	0.277	8.402	1.855	0.724	0.392	5.947	1.313
	11	0.935	0.189	12.335	2.723	0.790	0.297	7.845	1.732
	20	0.996	0.048	48.441	10.694	0.824	0.251	9.298	2.053
<i>Total acidity</i>									
None	9	0.842	2.863	8.335	1.900	0.859	3.457	6.903	1.574
	12	0.920	2.080	11.473	2.616	0.803	3.358	7.107	1.620
	17	0.973	1.217	19.613	4.471	0.776	3.532	6.757	1.540
MSC	7	0.835	2.919	8.176	1.864	0.862	3.465	6.886	1.570
	11	0.926	2.003	11.916	2.717	0.802	3.486	6.845	1.561
	18	0.992	0.673	35.457	8.084	0.861	2.813	8.482	1.934
SNV	7	0.837	2.904	8.218	1.874	0.861	3.642	6.553	1.494
	11	0.927	1.990	11.989	2.733	0.795	3.548	6.726	1.533
	18	0.992	0.659	36.189	8.251	0.830	3.340	7.145	1.629

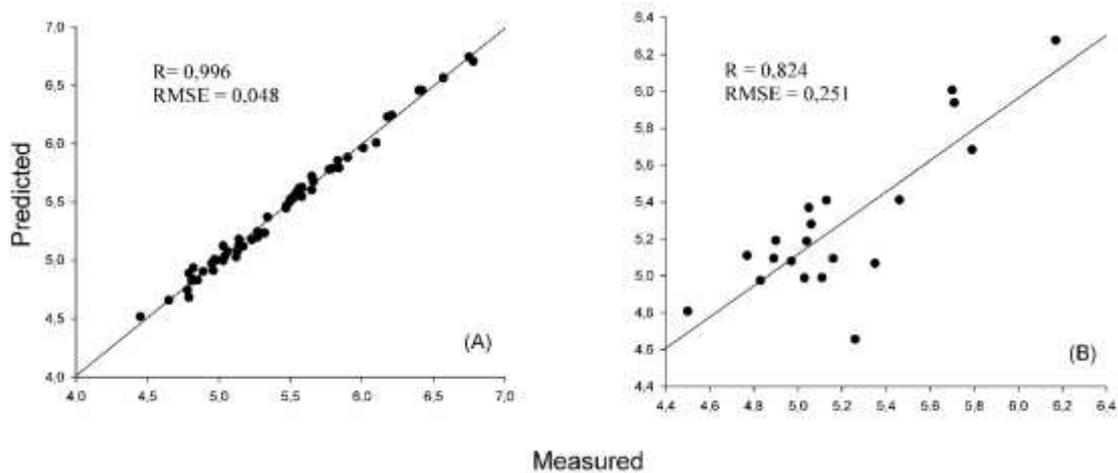
LV: latent variables, R: coefficient of correlation, RMSE: root mean square error, RER: range to error ratio, RPD: residual predictive deviation. Source: Authors.

Different numbers of latent variables were tested and the parameters (Table 3) were evaluated to choose the best model. The choice was based on the external validation data. R values close to 1 and low RMSE values indicate the good performance of the model to predict the parameters pH and titratable acidity. Another very relevant fit parameter is the ratio of performance to deviation (RPD), where models with $RPD > 2$ can be considered excellent; reasonable models when $1.4 < RPD < 2$; and unreliable models when $RPD < 1.4$ (Lima et al., 2020). The ratio of an analyte's concentration range to the root mean square error (RER) is also a relevant parameter, where a model with RER values < 3 shows low predictive ability, models with RER between 3 and 10 have moderate practical utility, and RER values > 10 indicate good model utility (Lima et al., 2020).

According to this classification, the models generated by the PLS technique can be considered reasonable for predicting the quality attributes. However, the model generated by SNV-treated data with 20 VL has moderate practical utility and excellent predictive ability for pH. For total acidity, the model generated by MSC data with 18LV has moderate utility and reasonable predictive ability. These were considered the best models to predict these quality parameters.

The scatter plots show the correlation between the experimental reference values and the values predicted by the optimized models for pH (Figure 2) and total acidity (Figure 3). It is possible to observe a high correlation between the actual and predicted data for all models, which may indicate that the models generated from data with higher sample dispersion over the model range have an easier fit. Therefore, it can be seen that the evaluation parameters of the models cited above indicate the possibility of satisfactorily determining the parameters acidity and pH of the cocoa bean.

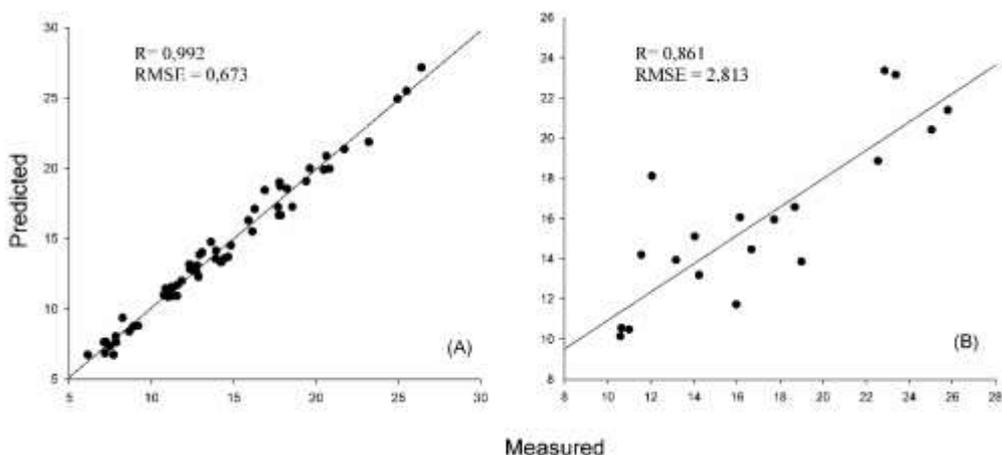
Figure 2. Correlation graph for the real and predicted pH values by the PLS model in the calibration (A) and validation (B), using data obtained by NIR for the samples cocoa beans.



Source: Authors.

The evaluation parameters of the models indicate the possibility to simultaneously determine two quality attributes of cocoa beans by using the NIR infrared spectroscopy technique associated with multivariate calibration models. Thus, chemometric models can be successfully used as a routine analysis to evaluate the quality of cocoa bean samples.

Figure 3. Correlation graph for the real and predicted total acidity values by the PLS model in the calibration (A) and validation (B), using data obtained by NIR for the samples cocoa beans.



Source: Authors.

4. Conclusion

This work provides a strategic alternative for the quick and simple analysis and classification of cocoa beans. PLS-DA was effective in differentiating the classes of higher and lower quality samples, determining to which class an unknown sample belonged based on the information provided to the system. PLS-DA obtained over 86% accuracy in classifying the samples into their respective classes. The application of PLS to the NIR dataset provided models with moderate practical utility and excellent predictive ability for pH and moderate utility and modest predictive for total acidity.

The authors indicate the use of the calibrated model by PLS-DA and PLS with NIR data as a form of quality control for the classification of higher and inferior quality samples. In addition to providing values of qualitative and essential attributes for the proper destination of the industrial application of cocoa beans.

Acknowledgments

This study was partially funded by the Foundation for Research Support of Bahia (FAPESB). The authors are grateful to CEPLAC (Uruçuca and Itabuna, BA, Brazil) and to University of Southwestern Bahia, Brazil.

References

- Afoakwa, E. O. (2011). Cocoa cultivation, bean composition and chocolate flavor precursor formation and character. *Chocolate Science and Technology*, 12-34.
- Alvarez, C., Perez, E., Cros, E., Lares, M., Assemat, S., Boulanger, R. & Davrieux, F. (2012). The use of near infrared spectroscopy to determine the fat, caffeine, theobromine and epicatechin contents in unfermented and sun-dried beans of criollo cocoa. *J Near Infrared Spectrosc*, 20, 307–315.
- Anyidoho, E. K., Teye, E., Agbemaflé, R., Amuah, C. L. Y. & Boadu, V. G. (2021). Application of portable near infrared spectroscopy for classifying and quantifying cocoa bean quality parameters. *Journal of Food Processing and Preservation*, 45(5), 1–13.
- AOAC. (2010). Official Methods of Analysis of Association of Official Analytical Chemists (18th ed.).
- Barbin, D. F., Maciel, L. F., Bazoni, C. H. M., Ribeiro, M. D. S., Carvalho, E. D. S., Bispo, M. D. P. S. & Miranda, E.Y. (2018). Classification and compositional characterization of different varieties of cocoa beans by near infrared spectroscopy and multivariate statistical analyses. *Journal of Food Science & Technology*, 55, 2457-2466.
- Barrientos, L. D. P., Oquendo, J. D. T., Garzón, M. A. G. & Álvarez, O. L. M. (2019). Effect of the solar drying process on the sensory and chemical quality of cocoa (*Theobroma cacao* L.) cultivated in Antioquia, Colombia. *Food Research International*, 115(52), 259–267.
- Beg, M. S., Ahmad, S., Jan, K. & Bashir, K. (2017). Status, supply chain and processing of cocoa - A review. *Trends in Food Science and Technology*, 66, 108–116.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento-Instrução Normativa nº 38, de 23 de jun. de 2008. Regulamento Técnico da Amêndoa de Cacau. Diário Oficial da União, Brasília, 30 jun. 2008. Seção I, pg.2.

CoEx. (2017). Elements of harmonized international standards for cocoa quality and flavour assessment.

Crafack, M., Keul, H., Eskildsen, C. E., Petersen, M. A., Saerens, S., Blennow, A., Skovmand-Larsen, M., Swiegers, J. H., Petersen, G. B., Heimdal, H. & Nielsen, D. S. (2014). Impact of starter cultures and fermentation techniques on the volatile aroma and sensory profile of chocolate. *Food Research International*, 63, 306–316.

Cruz-Tirado, J. P., Fernández Pierna, J. A., Rogez, H., Barbin, D. F. & Baeten, V. (2020). Authentication of cocoa (*Theobroma cacao*) bean hybrids by NIR-hyperspectral imaging and chemometrics. *Food Control*, 118, 107445.

Diomande, D., Antheaume, I., Leroux, M., Lalande, J., Balayssac, S., Remaud, G. S. & Tea, I. (2015). Multi-element, multi-compound isotope profiling as a means to distinguish the geographical and varietal origin of fermented cocoa (*Theobroma cacao* L.) beans. *Food Chemistry*, 188, 576–582.

Hayati, R., Zulfahrizal, Z. & Munawar, A. A. (2021). Robust prediction performance of inner quality attributes in intact cocoa beans using near infrared spectroscopy and multivariate analysis. *Heliyon*, 7(2), e06286.

Hernandez-Hernandez, C., Fernandez-Cabanas, V. M., Rodríguez-Gutierrez, G., Fernandez-Prior, A. & Morales-Sillero, A. (2022). Rapid screening of unground cocoa beans based on their content of bioactive compounds by NIR spectroscopy. *Food Control*, 131,1-9.

Kadow, D., Bohlmann, J., Phillips, W. & Lieberei, R. (2013). Identification of main fine or flavour components in two genotypes of the cocoa tree (*Theobroma cacao* L.). *Journal of Applied Botany and Food Quality*, 86, 90–98.

Krähmer, A., Engel, A., Kadow, D., Ali, N., Umaharan, P., Kroh, L. W. & Schulz, H. (2015). Fast and neat-determination of biochemical quality parameters in cocoa using near infrared spectroscopy. *Food Chemistry*, 181,152-159.

Lima, A. B. S. de, Batista, A. S., Jesus, J. C. de, Silva, J. de J., Araújo, A. C. M. & Santos, L. S. (2020). Fast quantitative detection of black pepper and cumin adulterations by near-infrared spectroscopy and multivariate modeling. *Food Control*, 107.

Lohumi, S. H., Lee, S., Lee, H. & Cho, B-K. (2015). A review of vibrational spectroscopic techniques for the detection of food authenticity and adulteration. *Trends in Food Science e Technology*, 46 (1), 85-98.

Lopes, J., Lima, A., Ribeiro, R., Santos, L. S., Viana, M., Passini, S. & Ferr, B. (2022). Application of spectroscopic techniques and chemometric methods to differentiate between true cinnamon and false cinnamon. *Food Chemistry*, 368.

Nogueira, B.L. (2015). Processamento do cacau: avaliação do teor nutricional do chocolate e dos outros derivados do cacau. (TCC) em Engenharia Bioquímica, Universidade de São Paulo.

Nunes, C. A., Freitas, M. P., Pinheiro, A. C. M. & Bastos, S. C. (2012). Chemoface: A novel free user-friendly interface for chemometrics. *Journal of the Brazilian Chemical Society*, 23(11), 2003–2010.

Oliveira, L. F., Braga, S. C. G. N., Augusto, F., Hashimoto, J. C., Efraim, P. & Poppi, R. J. (2016). Differentiation of cocoa nibs from distinct origins using comprehensive twodimensional gas chromatography and multivariate analysis. *Food Research International*, 90, 133-138.

Peláez, P. P., Guerra, S. & Contreras, D. (2016). Changes in physical and chemical characteristics of fermented cocoa (*Theobroma cacao*) beans with manual and semi-mechanized transfer, between fermentation boxes. *Scientia Agropecuaria*, 07(02), 111–119.

Santos, I. A., Conceição, D. G., Viana, M. B., Silva, G. de J., Santos, L. S. & Ferrão, S. P. B. (2021). NIR and MIR spectroscopy for quick detection of the adulteration of cocoa content in chocolates. *Food Chemistry*, 349.

Souza, A. M. D., Breikreitz, M. C., Filgueiras, P. R., Rohwedder, J. J. R. & Poppi, R. J. (2013). Experimento didático de quimiometria para calibração multivariada na determinação de paracetamol em comprimidos comerciais utilizando espectroscopia no infravermelho próximo: um tutorial, parte II. *Química Nova*, 36 (7), 1057-1065.

Sunoj, S., Igathinathane, C. & Visvanathan, R. (2016). Nondestructive determination of cocoa bean quality using FT-NIR spectroscopy. *Computers and Electronics in Agriculture*, 124, 234–242.

Veselá, A., Barros, A. S., Synytsya, A., Delgadillo, I., Čopíková, J. & Coimbra, M. A. (2007). Infrared spectroscopy and outer product analysis for quantification of fat, nitrogen, and moisture of cocoa powder. *Anal Chim Acta*, 601(1), 77–86.

Whitacre, E., Oliver, J., Van, D. B. R., Van E. P., Kremers, B., Van, D. H. B., Stewart, M. & Jansen-Beuvinck, A. (2003) Predictive analysis of cocoa procyanidins using near-infrared spectroscopy techniques. *J Food Sci*, 68, 2618–2622.