

Predictive model of moisture content in dry parchment coffee beans using near-infrared spectroscopy (FT-NIR)

Keyv Sandrith Ordoñez Lozano¹ , Andrés Felipe Bahamón Monje^{1*} , Nelson Gutierrez Guzman¹ 

¹Universidad Surcolombiana, Departamento de Ingeniería Agrícola, Centro Surcolombiano de Investigación en Café (CESURCAFÉ), Neiva, Huila, Colombia

*Corresponding author: andres.bahamon.coffee@gmail.com

Editor: Samuel Pereira de Carvalho

Received in October 4, 2024 and approved in December 18, 2024

ABSTRACT

Moisture content is a key quality parameter in grain storage. Near infrared spectroscopy (NIRS) is a non-destructive technique, with fast and simple measurements, that can be a method to implement for monitoring this parameter. The objective of this research was to evaluate the NIRS technique for the prediction of the moisture content of dry samples of dry parchment coffee (DPC), for this it was necessary to carry out the study in different drying times in order to obtain several points of moisture of the grain that could exist, once the spectrum is taken in the FT-NIR, it is taken to the oven so that through the gravimetric method the real moisture of the grain can be known. The Spectrum Quant software was used to develop the predictive mathematical model by means of principal component regression (PCR) and partial least squares (PLS), using a set of randomly divided data for calibration and validation. The results showed that a better goodness of fit was found with a PLS model and the application of the baseline and second derivative correction, obtaining a coefficient of determination (R^2) of 0.99 and a predictive standard error (SEP) of 0.34. Finding a good correlation between the real data with those estimated by the NIR equipment, emerging a fast and practical way in full-scale monitoring in DPC grain moisture control.

Key words: Quality control; prediction; mathematical models; drying; arabica coffee.

1 INTRODUCTION

Coffee is a commodity of significant economic importance in developing countries. The majority of coffee production is concentrated in Brazil (32-34%), followed by Vietnam (12-13%) and Colombia (8-9%) (Barbosa et al., 2019). Coffee is one of the most exported products worldwide due to its high consumption, particularly in Europe as a hot beverage. There was an increase in coffee consumption in Europe between 2019 and 2020 (Tugnolo et al., 2021). The export of coffee represents a vital source of foreign exchange and substantial revenues, reaching up to 80% in less developed countries. Additionally, coffee contributes significantly to gross domestic product and income taxes in some countries (Figueroa-Hernández et al., 2019).

The drive to meet consumer demand and market requirements encourages producers to compete to produce high-quality products (Li et al., 2024). Therefore, post-harvest processing conditions must be optimized to ensure the desired quality (Munyendo et al., 2021; Tugnolo et al., 2021). Storage is a crucial aspect of maintaining cup quality, as it can influence the production and/or degradation of various chemical compounds, including acids, caffeine, lipids, and carbohydrates (Wójcicki, 2022). A critical parameter to consider during storage is the moisture content of the grain. This is one of the quality standards required for the export and import of the grain, which must be within a range of 10% to 12%. Exceeding this range can impair the quality and safety of the grain. At high moisture contents, discoloration, unpleasant

flavors, and thermostable mycotoxin-producing molds can occur. These molds can cause damage to consumer health, as demonstrated by Adnan et al. (2017). Consequently, the shelf life of the grain is diminished (Donovan et al., 2019). There are direct reference methods for determining moisture content, such as ISO 1446:2001, ISO 1447:1978, and ISO 6673:2003. The latter is equivalent to NTC 2325, which served as the basis for quantifying moisture content in this study. An alternative approach is the use of near-infrared spectroscopy (NIR) to model and predict grain moisture in a rapid, efficient, and straightforward manner (Patiño Velasco et al., 2016).

Near-infrared spectroscopy is a non-destructive, accurate, reliable, and reproducible test method for qualitative and quantitative analysis in a variety of sectors. It has been widely employed in agricultural products, food processing, pharmaceuticals, and other fields (Liu et al., 2022). In recent years, NIR spectroscopy has been of practical value and important in coffee bean analysis research (Li et al., 2024), this method allows for the analysis of a large number of samples without the need for prior processing. Furthermore, it is a relatively rapid approach compared to the reference method, facilitating the analysis of different parameters and enabling the classification of samples based on qualitative characteristics (Gómez; Ortiz; Osorio, 2022).

The technique is based on the Beer-Lambert law, which consists of irradiating a beam of infrared light (800 to 2500 nm) on the sample. The absorbed intensity of near-infrared light by the sample is related to the electrostatic bonds and charges present in its atoms and molecules. These values are

expressed in terms of reflectance (log standard reflectance/reflectance of the sample) and are transformed into a characteristic spectrum of the analyzed sample (Gómez; Ortiz; Osorio, 2022). Consequently, the light scattering is contingent upon the physical and chemical properties of the sample. The NIR technique is used to know and predict the quantitative or qualitative data of the analyzed sample, combining computer technology and chemometric methods, correlating spectral data with reference methods, from which predictive models are built, turning the NIR technique into a secondary method (Gómez *et al.*, 2023; Gómez *et al.*, 2021) offering a new opportunity to analyze sample structures and enhance development (Li *et al.*, 2024).

To date, research has been conducted utilizing the NIRS instrument to predict moisture content in green coffee beans (Adnan *et al.*, 2017; Escobar; Aleugene; Rodriguez, 2020; Kyaw *et al.*, 2020) and to monitor moisture in roasted coffee beans (Tugnolo *et al.*, 2021). Other applications of NIR in green coffee beans is with the detection of defects, geographical origin, processing method (Li *et al.*, 2024) variety, roasting time in roasted coffee (Guerrero-Peña *et al.*, 2023), as well as its use to distinguish coffee quality (Mutz *et al.*, 2023).

However, no studies have been conducted on dried parchment coffee, arising the need for the present study, as a tool to monitor the moisture content in coffee beans in real time during drying, this is a crucial quality parameter. The objective was to take spectra with the NIR technique at different drying times to monitor the moisture content and develop a mathematical model to predict the moisture content in dry parchment coffee.

2 MATERIAL AND METHODS

2.1 Raw material procurement

The cherry coffee was harvested in the year 2023 from August to December in different municipalities of the department of Huila: in the municipality of La Plata (Huila), the Castillo variety was harvested; in the municipality of Gigante (Huila), the Caturra variety was harvested; and in the municipality of Pitalito (Huila), the Colombia variety was harvested. The samples were processed in the facilities of the coffee research center - CESURCAFÉ (Neiva-Huila) by means of wet milling.

The dried parchment coffee beans (sample) were selected during the drying process at different times to provide a wide range of moisture content and feed the NIR library, for obtaining the best response variables, including three varieties from different regions. Drying was carried out in a mechanical dryer with a maximum capacity of 15 kg (CE-500 manufactured by Ingesec, Colombia) at a temperature of 40 °C, a flow rate of 36.6 Hz and agitation per hour (Figure 1).

Once the sample was removed from the mechanical dryer, a time of 10 minutes was allowed for the grains to stabilize before taking the spectra.

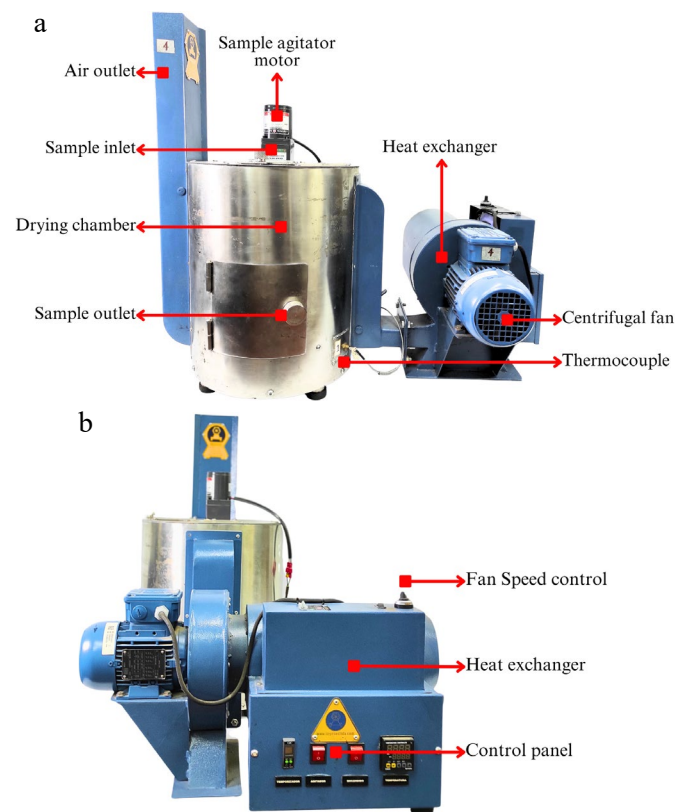


Figure 1: Mechanical dryer (CE-500 manufactured by Ingesec, Colombia).

2.2 Sample collection

During the drying process, 15 g samples were taken every 2 hours to assemble the library of spectra in the NIR, ensuring variable reference values that help develop viable and robust regression models (Zhu *et al.*, 2021). Once the spectrum was obtained, the sample was taken to an oven (UF160 manufactured by Mermmet, Germany) to determine the moisture by the gravimetric method according to NTC 2325:2005 (Café verde. Determinación de la pérdida de masa a 105 °C., 2005). The required number of samples were taken at different ranges of moisture content, starting at approximately 50%, which is the moisture content of the coffee after fermentation, until the coffee reaches the optimum range of moisture content (between 10 and 12%); each sample was taken in triplicate.

2.3 Near infrared spectroscopy collection

To build the library and develop the chemometric analysis, reference values were taken over a wide range of moisture content. A Fourier transform near-infrared (FT-NIR)

instrument (Spectrum Two manufactured by PerkinElmer, Inc., USA) with a rotating accessory and an interleaved scanner was used. Each spectrum was acquired in diffuse reflectance mode, using a high-resolution InGaAs detector over a wavelength range of 4000 to 12000 cm^{-1} (830-2500 nm) with an inter-data interval of 4 cm^{-1} , a resolution of 8 cm^{-1} , and 64 scans per spot. Each spectrum contains 2000 points, and three replicates were acquired for each sample (Zhu et al., 2021).

The spectra were collected using Spectrum IR software (Version 10, PerkinElmer, USA). Dried parchment coffee beans (15 g) were deposited in a petri dish and arranged on the rotating accessory for continuous rotation of the sample during spectral acquisition.

2.4 Moisture content measurement

The percentage of moisture content on a moisture basis was determined using the NTC 2325:2005 standard. It was performed by using an oven (UF160 manufactured by Mermmet, Germany) to dry approximately 5 g of DPC in open aluminum boxes at $105 \pm 1^\circ \text{C}$ for 16 h. The aluminum boxes were tempered in desiccators for 60 min to cool the samples to room temperature. Finally, samples were weighed to calculate % moisture content as a function of weight loss; data are given as the average of three replicates (Equation 1).

$$\%H = \frac{w_i - w_f}{w_i} \times 100 \quad (1)$$

Where %H is the percentage of moisture content, W_i is the initial weight and W_f is the final weight of the sample after it leaves the oven.

2.5 Data processing and analysis

Each of the spectral data obtained by diffuse reflectance was obtained by absorbance ($\log 1/R$), and then the data were transformed to the Kubelka-Munk (K-M) model using the advanced algorithms package of the Spectrum™10 Quant Advanced statistical software (<https://www.perkinelmer.com/category/ftir-software>). A combination of two regression models: principal component regression (PCR) and partial least squares (PLS), along with some mathematical pre-treatments, such as baseline correction, were used to develop the chemometric models for predicting the moisture content of parchment coffee samples. This process helped us to eliminate variations in the spectrum that were not related to the chemical characteristics of the samples, such as interferences due to instrumental noise or light scattering effects. In addition, it improved the error related to the shape of the coffee beans, since they are heterogeneous samples. First and second order derivatives, fundamental tools for highlighting specific spectral features and eliminating signal variations, were also applied as pretreatment. The first derivative was

used to correct for baseline shifts and highlight changes in the slope of the spectrum, while the second derivative was used to enhance small peaks and minimize signal overlap effects (Ozaki et al., 2021). This treatment was applied to the calibration and validation data; the data were randomly divided into 70% (60 spectra) for calibration used to develop the prediction model and the remaining 30% (24 spectra) to validate and test model performance (Munawar et al., 2024). The evaluation index of the predictive model accuracy in the NIR instrument, was established from the ratio of the resulting values of parameters such as the highest coefficient of determination (R^2), the estimated standard error (SEE), the predictive standard error (SEP) the standard error of cross-validation (SEPVC) and the lowest and homogeneous mean square error (RMSE) among themselves (Cortés et al., 2019; Molano et al., 2016), being the R^2 and RMSE the most used performance indicators (Chen et al., 2022), as a final evaluation of the model performance the RPD (Ratio between performance and deviation) is taken into account (Liu et al., 2022).

When the RPD is less than 1.4, the predictive effect of the model is very poor and cannot be used; when the RPD is between 1.4 and 1.8, the model can perform an approximate quantitative analysis of the sample. When the RPD is between 1.8 and 2, it indicates that the model has a better predictive effect; when the RPD is greater than 2, it indicates that the model has excellent predictive performance. The higher the RPD value, the better the robustness of the model (Liu et al., 2022; Wei et al., 2024).

3 RESULTS

3.1 Spectral data processing

Observation of the absorbance-transformed ($\log (1/R)$) (1a) and Kubelka-Munk (K-M) (1b) spectra (Figure 2) reveals that the most pronounced peaks are found around the 1450 nm and also at the 1950 nm bands, which are dominated by the water absorption bands due to the first O-H stretching tone. This is in accordance with the findings reported in the literature (Budiastra et al., 2018). This finding is consistent with the results of previous studies conducted Tugnolo et al. (2021) and Adnan et al. (2017), which indicate that the chemical information regarding the moisture content of the samples is located in this region.

3.2 Pre-treatment and analysis of moisture content prediction data from NIR spectra

In the analysis of each of the spectra, the full spectral range from 800 to 2500 nm was utilized (Munyendo et al., 2021). The statistical models of PCR and PLS were employed,

with the latter being a widely utilized statistical tool for chemometric models (Ribeiro *et al.*, 2021). Additionally, it is the most recommended modeling method (Cortés *et al.*, 2019). In each model, the parameters yielded by the Spectrum Quant software were compared. The raw data underwent a variety of pretreatments, including baseline correction, first and second derivative, and both derivatives, which collectively enhanced the spectral resolution and reduced the signal-to-noise ratio, thereby improving the accuracy of the predictive mathematical model (Pizarro *et al.*, 2004; Shan *et al.*, 2023). The resulting values of the statistical parameters are presented in Table 1, which depicts the calibration of all the chemometric models under investigation using the aforementioned spectra.

Predictive efficiency is demonstrated in several of the models, expressing low standard errors, as well as coefficients of determination $R^2 > 0.90$, expressing the good

fit of the models (Molano *et al.*, 2016). This demonstrates a high correlation between the prediction by NIRS and the data obtained by the gravimetric method (Cortés *et al.*, 2019).

As illustrated in Table 1, a variety of spectral pretreatment techniques were employed prior to the chemometric analysis. This was done to eliminate any additive or multiplicative light scattering, signal-to-noise reduction, and to diminish the impact of physical sample characteristics (Zhu *et al.*, 2021). The comparison between models and pretreatments was conducted by establishing PCR and PLS regression models with their pretreated and raw spectra across the full spectral range (800-2500 nm). The optimal model was selected by comparing R^2 , SEPVC, and RMSE. It can be observed that certain spectral pretreatments yielded positive outcomes, while others resulted in adverse effects.

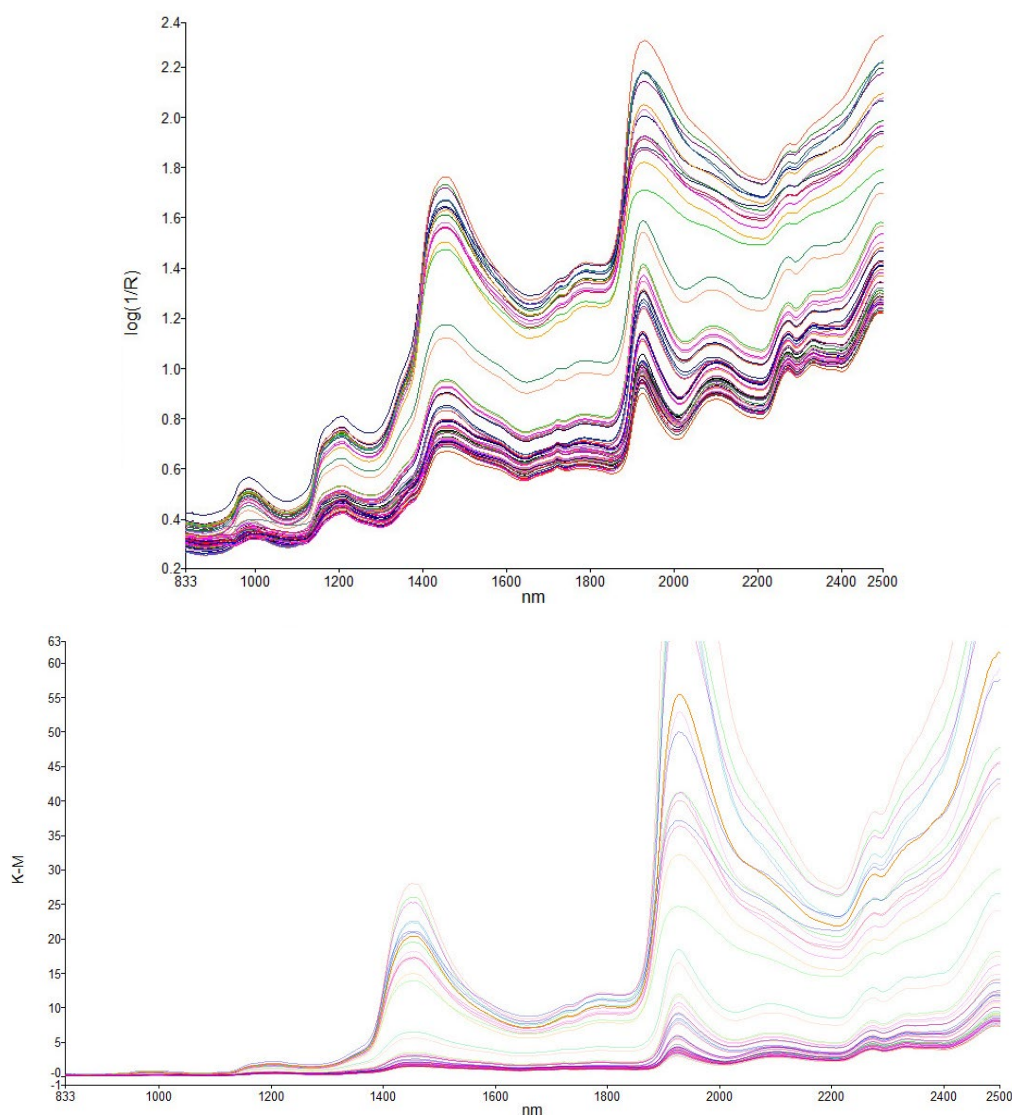


Figure 2: The graph is not numbered a and b. On the other hand, also set the same width in the two graphs.

Table 1: Predictive model performance metrics for moisture content in dried parchment coffee beans.

Models	Pretreatments	Calibration statistics					
		% R ²	SEE	SEP	SEPVC	RMSE	RPD
PCR ABS	Raw data	98.72	2.11	2.36	2.50	0.53	6.94
	LB Y D1	98.41	2.30	2.58	2.50	0.53	6.92
	LB Y D2	56.47	11.63	12.16	11.38	0.78	1.52
PCR Kubelka Munk	Raw data	89.65	6.00	8.65	7.89	0.26	2.20
	LB Y D1	33.73	14.23	14.49	14.33	0.39	1.21
	LB Y D2	26.21	15.14	16.89	14.62	0.39	1.19
PLS ABS	Raw data	98.20	2.43	2.61	2.98	0.14	5.81
	LB Y D1	99.97	0.31	0.34	1.86	0.08	9.32
	LB Y D2	99.98	0.24	0.27	5.12	0.10	3.39
PLS Kubelka Munk	ST	41.28	13.51	13.43	13.59	0.26	1.28
	LB Y D1	64.14	10.85	11.88	17.99	0.13	0.96
	LB Y D2	25.95	15.05	15.14	15.02	0.52	1.15

PCR: Principal component regression; ABS: Absorbance; PLS: Least squares regression; LB: Baseline correction; D1 and D2: Application of 1st and/or 2nd derivative; R²: Coefficient of determination; SEE: Estimated standard error; SEP: Predictive standard error; SEPVC: Cross-validation standard error; RMSE: Root mean square error; RPD: Residual predictive deviation.

The models with the optimal results were obtained from the spectra that underwent absorbance processing, resulting in a high R² value in the least squares regression (PLS) with LB preprocessing and application of D2. This model exhibited an R² value of 99.98 and a SEP value of 0.27, representing the lowest value among all models. Nevertheless, this model does not exhibit the lowest values of SEPVC and RMSE. Therefore, the model that most closely aligns with the desired fit and provides the greatest reliability, according to its statistical results, is the one obtained with the absorbance data in the least squares regression (PLS) and LB and D1 preprocessing. This model also demonstrates a high predictive performance. Figure 3 depicts the spectrum of the calibrated model, wherein all the utilized calibration samples are discernible, exhibiting clear delineation of the regions.

3.3 Effect of outliers on the prediction model.

Figure 4a) shows the good fit of the means of each of the data entered; however, dispersion is observed in Figure 4b), showing suspicious values considered as outliers for the predictive model. The literature mentions that this type of value can give erroneous conclusions, so it is suggested to eliminate these data so that the model has a more accurate prediction (Neira-Rodado et al., 2020). The inspection in Figure 4b) shows four outliers, of which two were eliminated, the decision was taken to eliminate these values as it improves the predictive capacity and stability of the model, these outliers may be deviations attributable to external factors such as errors during the measurement of the spectra, such as environmental conditions and probably some variations in the characteristics of the coffee such as the variety and its origin.

Excluding the data that were scattered (Figure 4. b) evidences a slight improvement in the R² of 99.98, SEE 0.24, and SEP 0.27. Adnan et al. (2017) in their research reveal that the suspected outliers were not real outliers as they did not represent significant variance, so they decided not to exclude these values.

4 DISCUSSION

The absorption peaks are found around 1450 and 1950 where the O-H stretching vibrations of the water molecules occur, this spectral fingerprint is used to predict the moisture content in dried parchment coffee beans through NIR. Studies such as Munawar et al. (2024), show the effectiveness of NIR in predicting moisture content in intact coffee beans, using models such as PCR and PLS. Both methods provided good prediction performance in determining moisture content with an R² correlation of 0.979 for PCR and 0.984 for PLSR, respectively. Within our results, it is found that better prediction was obtained with the PLS model; Munawar et al. (2024) mentions that PLS seem to be slightly more accurate in predicting water content with higher correlation and determination coefficients.

As stated by Adnan et al. (2017), the models developed with partial least squares (PLS) are characterized by robustness and accuracy. According to Tugnolo et al. (2021) have observed, PLS is designed to maximize the covariance between the NIR spectra and the results obtained through NTC 2325:2005. It is observed that the application of the Kubelka-Munk transform has a significant impact on

the resulting data, with a notable decrease in the goodness of fit. This is interpreted as a reduction in the predictive probability of the model if the transformed data are taken as a reference. Moreover, the PLS model developed from NIR data demonstrated a strong correlation between estimated values and reference concentrations in whole roasted and ground coffee beans, with prediction errors below 0.15% and coefficients of determination exceeding 0.95 after applying baseline and second derivative preprocessing. However, the

opposite is observed in the study by Adnan *et al.* (2017), which identifies the highest accuracy for predicting moisture content in whole green coffee beans with the PLS model without any preprocessing. Specifically, the raw data yielded the highest R^2 and the lowest mean square error compared to other preprocessing methods. In contrast, a study by Kyaw *et al.*, (2020) indicates that the Kubelka-Munk transformed data, pretreated with a second derivative, yield superior results in predicting the moisture content of Liberica coffee.

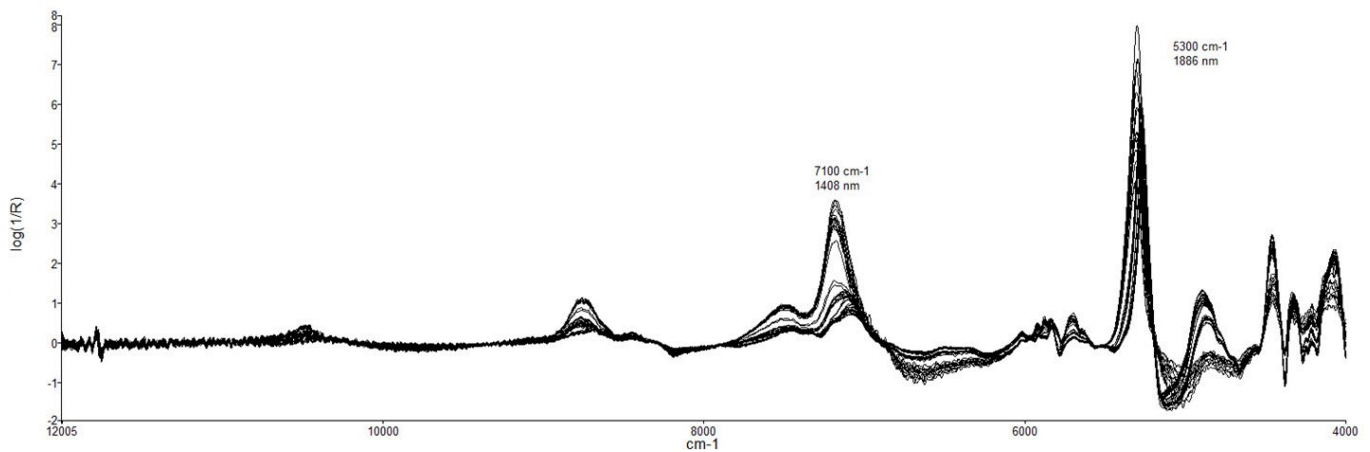


Figure 3: Spectrum with baseline and 1st derivative correction pretreatment.

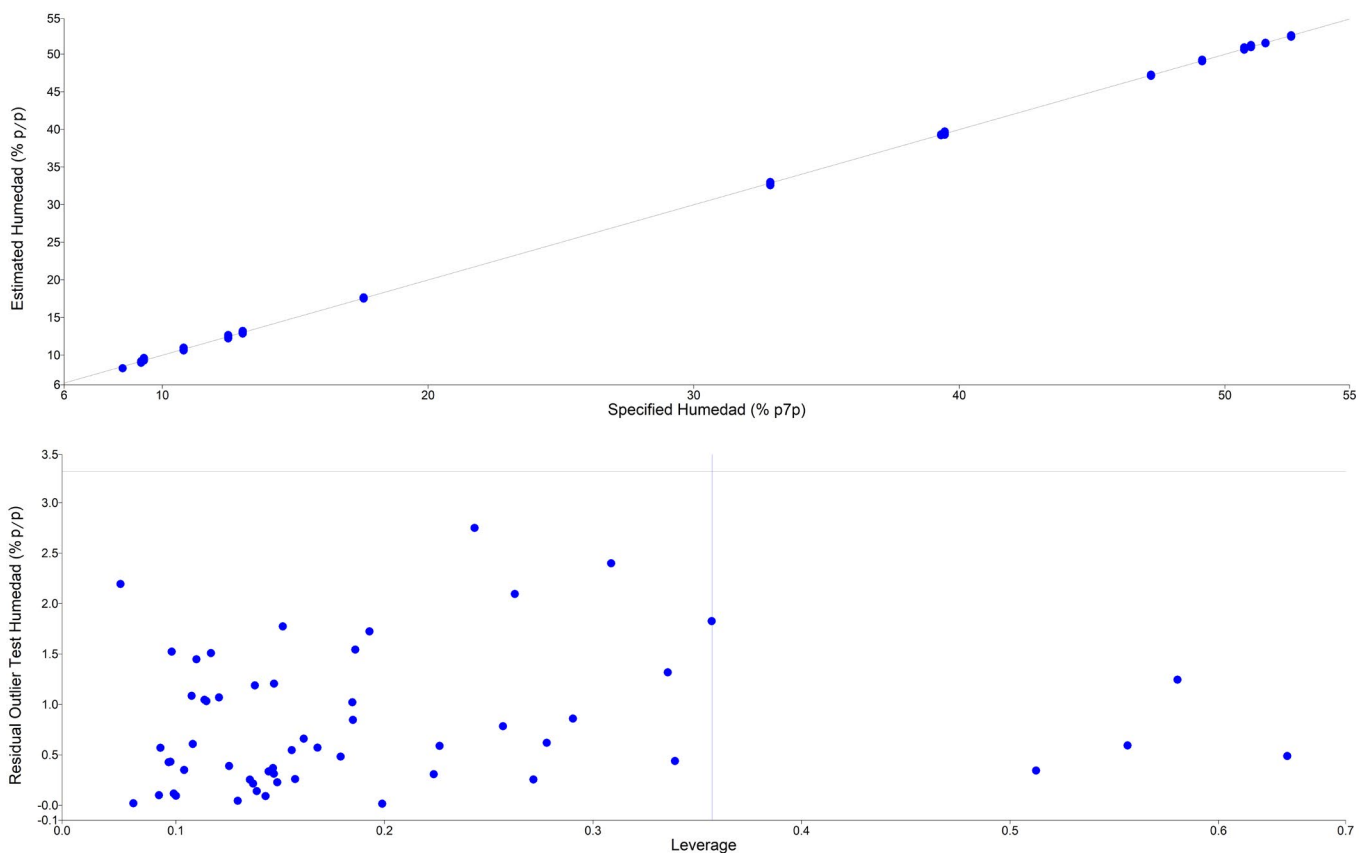


Figure 4: Goodness-of-fit with the elimination of two data points.

The results indicate that the developed chemometric models showed high predictive efficiency, with determination coefficients (R^2) greater than 0.90. This suggests a significant correlation between the predictions made by NIRS spectroscopy and the data obtained through gravimetric methods. The high accuracy of these models is crucial, as it allows coffee producers to monitor moisture content quickly and non-destructively, which is essential for grain quality and preventing deterioration during storage.

The results also highlight the importance of data pre-treatment in the quantitative determination of moisture content. Different pre-treatment methods, such as normalization and Kubelka-Munk transformation, showed significant variations in the model's accuracy. This underscores the need to carefully select the appropriate pre-treatment method to optimize calibration and improve the robustness of predictive models.

The ability of NIRS spectroscopy to provide real-time, non-destructive results represents a significant advantage over traditional moisture analysis methods, which are often slower and require sample destruction. This not only saves time but also allows for continuous monitoring of coffee quality during storage and transportation, which can positively impact the sector's economy.

The findings of this study have important implications for the coffee sector, especially regarding quality and sustainability. The ability to accurately predict moisture content can help producers optimize their drying and storage processes, thereby reducing losses due to deterioration and improving the quality of the final product. Moreover, the use of technologies like NIRS can contribute to more sustainable practices, minimizing resource waste and improving operational efficiency.

The results demonstrate that the PLS calibration model with a baseline and second derivative correction preprocessing is an effective method for predicting moisture content in dried parchment coffee beans. The real-time acquisition of data from the entire product represents a potential application of NIR in the coffee quality line, particularly as a follow-up in the drying process. As an optimal, non-invasive, and effective tool, its online application can be used to replace the slow and tedious conventional methods (Cortés et al., 2019).

The scalability of using near-infrared (NIR) technology for monitoring parchment coffee drying is faced with significant challenges. For instance, it is imperative to guarantee homogeneity or an understanding of the drying gradient zones in large batches within the drying equipment. To this end, a dynamic sampling system that achieves a representative sample or sensors distributed within the bean are required. This is necessary to take into account the effect of temperature as well as to achieve a chemometric model capable of handling the inherent variability of environmental conditions, sample, dryer design, technology, among others. Furthermore, the transfer of

calibration models between regions and cultivars can be unique for each zone due to differences in climatic conditions and grain characteristics. Initial investment and maintenance costs are also a significant barrier, particularly for small producers. Overcoming these challenges necessitates the development of adaptive solutions, the creation of specific chemometric models, and collaboration between producers, researchers, and industry stakeholders for the effective implementation of these solutions to control the moisture content of coffee during drying, optimize time, reduce loss of mass, maintain bean quality, and conserve economic resources.

5 CONCLUSIONS

This study demonstrates that the PLS model with baseline correction and first derivative preprocessing provides highly accurate predictions of moisture content in dry parchment coffee beans, with an R^2 of 0.99, with a predictive standard error (SEP) of 0.34, SEPVC of 1.86, and an RMSE of 0.08. This model offers a reliable, non-destructive alternative to traditional moisture measurement methods, enabling potential real-time monitoring in coffee processing. Further research is recommended to validate this approach across diverse coffee varieties and conditions to support broader industrial implementation.

6 ACKNOWLEDGEMENTS

The author would like to thank the Young Researchers Program of the Ministry of Science, the Universidad de la Amazonía, coordinator of the BPIN project: 2022000100076 and the laboratories of the Coffee Research Center - CESURCAFÉ of the Faculty of Engineering of the Universidad Surcolombiana-Neiva.

7 AUTHORS' CONTRIBUTION

Conceptual idea: Ordoñez, K.S.; Bahamón, A.F.; Methodology design: Ordoñez, K.S.; Bahamón, A.F.; Gutierrez N.; Data collection: Ordoñez, K.S.; Bahamón, A.F.; Data analysis and interpretation: Ordoñez, K.S.; Bahamón, A.F.; Gutierrez N.; Writing and editing: Ordoñez, K.S.; Bahamón, A.F.; Gutierrez N.].

8 REFERENCES

- ADNAN, A. et al. Rapid prediction of moisture content in intact green coffee beans using near infrared spectroscopy. **Foods**, 6:38, 2017.
- BARBOSA, M. de S.G. et al. Correlation between the composition of green Arabica coffee beans and the sensory quality of coffee brews. **Food Chemistry**, 292:275-280, 2019.

- Budiastira, I. W. et al. Prediction of Caffeine Content in Java Preanger Coffee Beans by NIR Spectroscopy Using PLS and MLR Method. **IOP Conference Series: Earth and Environmental Science**, 147, 012004, 2018.
- CHEN, S. et al. Digital mapping of *GlobaSoilMap* soil properties at a broad scale: A review. **Geoderma**, 409:115567, 2022.
- CORTÉS, V. et al. Monitoring strategies for quality control of agricultural products using visible and near-infrared spectroscopy: A review. **Trends in Food Science & Technology**, 85:138-148, 2019.
- DONOVAN, N. K. et al. Analysis of green coffee quality using hermetic bag storage. **Journal of Stored Products Research**, 80:1-9, 2019.
- ESCOBAR, M. M.; ALEUGENE, L. T.; RODRIGUEZ, N. J. M. Non-destructive prediction of moisture content of philippine coffea arabica and coffea liberica green beans using locally-developed NIR spectroscopy instrument. **Mindanao Journal of Science and Technology**, 18(2):208-223, 2020.
- FIGUEROA-HERNÁNDEZ, E. et al. Los precios de café en la producción y las exportaciones a nivel mundial. **Revista Mexicana de Economía y Finanzas**, 14:41-56, 2019.
- GÓMEZ, C. R. et al. 2023. Determinación de compuestos químicos del café tostado por Espectroscopia de Infrarrojo Cercano (NIRS). **Revista Cenicafé**, 74:e74104, 2023.
- GÓMEZ, C. R. et al. Validación de curvas de calibración por NIRS para la predicción de compuestos químicos de café almendra. **Revista Cenicafé**, 72:e72204, 2021.
- GÓMEZ, C. R.; ORTIZ, A.; OSORIO, V. Predicción del origen regional del café de Colombia a partir de la técnica de espectroscopia de infrarrojo cercano - NIRS. **Revista Cenicafé**, 73:e73205, 2002.
- GUERRERO-PEÑA, A. et al. Chemical analysis and NIR spectroscopy in the determination of the origin, variety and roast time of Mexican coffee. **Heliyon**, 9:e18675, 2023.
- KYAW, E. M. et al. Estimation of moisture content in Liberica coffee by using near infrared spectroscopy. **IOP Conference Series: Earth and Environmental Science**, 542:012013, 2020.
- LI, T. et al. Review: Application and research progress of near infrared spectroscopy for multidimensional assessment of coffee beans. **Journal of Near Infrared Spectroscopy**, 32:179-191, 2024.
- LIU, Z. et al. Research on moisture content detection method during green tea processing based on machine vision and near-infrared spectroscopy technology. **Spectrochim Acta A Mol Biomol Spectrosc**, 271:120921, 2022.
- MOLANO, M. L. et al. Near infrared spectroscopy (NIRS) calibration equations to predict nutritional quality parameters of tropical forages. **Tropical Grasslands-Forrajes Tropicales**, 4(3):139, 2016.
- MUNAWAR, A. A. et al. Near-infrared technology in agriculture: Rapid, simultaneous, and non-destructive determination of inner quality parameters on intact coffee beans. **Open Agriculture**, 9:20220290, 2024.
- MUNYENDO, L.; NJORGE, D.; HITZMANN, B. The potential of spectroscopic techniques in coffee analysis: A review. **Processes**, 10:71, 2021.
- MUTZ, Y. S. et al. 2023. Feasibility of NIR spectroscopy coupled with chemometrics for classification of Brazilian specialty coffee. **Food Control**, 149:109696.
- NEIRA-RODADO, D. et al. Evaluating the impact of a two-stage multivariate data cleansing approach to improve to the performance of machine learning classifiers: a case study in human activity recognition. **Sensors**, 20:1858, 2020.
- OZAKI, Y. et al. **Near-infrared spectroscopy**. Springer Singapore, Singapore. 2021.
- PATÍÑO VELASCO, M. M. et al. Determination of moisture content in dry coffee beans by means of dynamic speckle. **Biotechnología en el Sector Agropecuario y Agroindustrial**, 14(2):84-91, 2016.
- PIZARRO, C. et al. Influence of data pre-processing on the quantitative determination of the ash content and lipids in roasted coffee by near infrared spectroscopy. **Analytica Chimica Acta**, 509(2):217-227, 2004.
- RIBEIRO, J. S. et al. Prediction of a wide range of compounds concentration in raw coffee beans using NIRS, PLS and variable selection. **Food Control**, 125:107967, 2021.
- SHAN, P. et al. A novel infrared spectral preprocessing method based on self-deconvolution and differentiation in the frequency domain. **Vibrational Spectroscopy**, 127:103562, 2023.

- TUGNOLO, A. et al. A reliable tool based on near-infrared spectroscopy for the monitoring of moisture content in roasted and ground coffee: A comparative study with thermogravimetric analysis. **Food Control**, 130:108312, 2021.
- WÓJCICKI, K. Near-infrared spectroscopy as a green technology to monitor coffee roasting. **Foods and Raw Materials**, 10(2):295-303, 2022.
- WEI, Y. et al. Estimation for soluble solid content in hetian jujube using hyperspectral imaging with fused spectral and textural Features. **Journal of Food Composition and Analysis**, 128:106079, 2024.
- ZHU, M. et al. Fast determination of lipid and protein content in green coffee beans from different origins using NIR spectroscopy and chemometrics. **Journal of Food Composition and Analysis**, 102:104055, 2021.