

SW-NIR spectroscopy and PLSR for predicting cocoa nibs quality across processing stages

RIANA LISTANTI^{1,2} , RUDIATI EVI MASITHOH^{1*}, ARIFIN DWI SAPUTRO¹,
HANIM ZUHROTUL AMANAH¹, SONNY ILHAM WICAKSONO³

¹Department of Agricultural and Biosystems Engineering, Faculty of Agricultural Technology,
Universitas Gadjah Mada, Yogyakarta, Indonesia

²Department of Agricultural Engineering, Faculty of Agriculture, Universitas Jenderal, Purwokerto,
Indonesia

³Department of Physics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada,
Yogyakarta, Indonesia

*Corresponding author: evi@ugm.ac.id

Citation: Listanti R., Masithoh R.E., Saputro A.D., Amanah H.Z., Wicaksono S.I. (2026): SW-NIR spectroscopy and PLSR for predicting cocoa nibs quality across processing stages. *Czech J. Food Sci.*, 44: 424–440.

Abstract: Cocoa beans quality parameters are among other things biochemical properties such as moisture content, reducing sugars, fat, protein and total phenolic content (TPC) changed during postharvest processes. Current analytical approaches to these parameters are usually not only destructive and time-consuming but also laboratory-based analyses. This study aimed at assessing the potential of short-wave near infrared (SW-NIR) spectroscopy (954–1 700 nm) combined with partial least squares regression (PLSR) for the prediction of these quality parameters in both fresh, fermented, dried and roasted Forastero cocoa nibs. In total, a portable Flame-NIR+ spectrometer analysed 400 samples. Before developing the model, we have compared several spectral preprocessing methods like Savitzky–Golay derivatives, multiple scatter correction (MSC), standard normal variate (SNV), and normalisation. The spectra exhibited various absorption features linked to O-H, C-H, and N-H links reflective of evolving compositions across proximate processing stages. The prediction coefficients (R^2P) obtained for moisture content (0.984), reducing sugars (0.958), protein (0.992), and total phenolic content (0.954) of the cocoa nibs was good, while fat prediction showed lower predictive performance than the other evaluated parameters. Savitzky-Golay derivative preprocessing consistently improved the performance of models across all evaluated parameters. These results show that SW-NIR spectroscopy with PLSR can be applied for fast, non-destructive and simultaneous estimations of multiple quality attributes in cocoa nibs during different stages of postharvest processing.

Keywords: biochemical composition; chemometrics; model interpretation; non-destructive analysis; spectral preprocessing

Cocoa (*Theobroma cacao* L.) (Figure 1) is one of the most important commodities around the world and its main end-use is as a raw material for chocolate

production (Perez et al. 2022). Its quality depends on the post-harvest processes (fermentation, drying, and roasting) that define the physicochemical and

Supported by the the Centre for Higher Education Funding (BPPT) and the Indonesian Endowment Fund for Education (LPDP) through a doctoral scholarship program.

© The authors. This work is licensed under a Creative Commons Attribution 4.0 International Licence (CC BY 4.0).

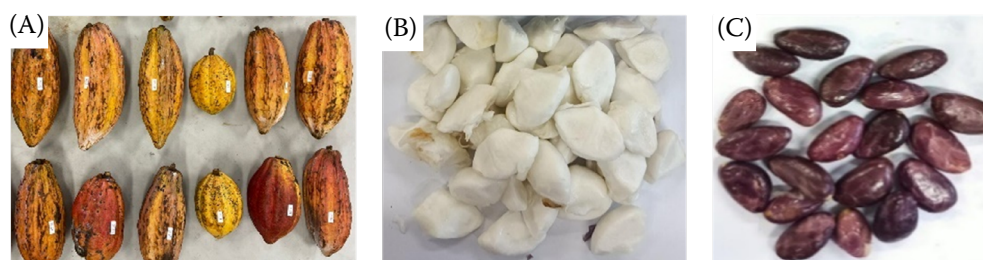


Figure 1. Cocoa samples used in this study

(A) Forastero cocoa pods at different ripening stages; (B) fresh unpeeled cocoa beans with pulp; (C) peeled cocoa beans (cocoa nibs)

sensory characteristics of cocoa products (Calvo et al. 2021). Different quality parameters such as moisture content (Jiménez-Rodríguez et al. 2024), reducing sugars and protein (Peña-Correa et al. 2024), fat content (Afoakwa et al. 2007), and total phenolic content (das Virgens et al. 2021) are important in storage stability, flavour development, and nutritional value (Ramón et al. 2023).

Typical analytical approaches to measure these parameters are accurate but also time-consuming, destructive, and not appropriate for real-time monitoring (Devi et al. 2025). The limitations of these methods

have been an impetus for the development of non-destructive analysis technologies that can assess the quality of a material without destroying the sample. Rapid surface analysis is achievable with computer vision system, but they cannot bring additional information about internal chemical composition (Brosnan and Sun 2004). Hyperspectral imaging provides more detailed spectral information, but with high investment costs (Manley 2014).

Near-infrared (NIR) spectroscopy is an efficient non-destructive approach to evaluating chemical composition by measuring the O-H, C-H and N-H bonds

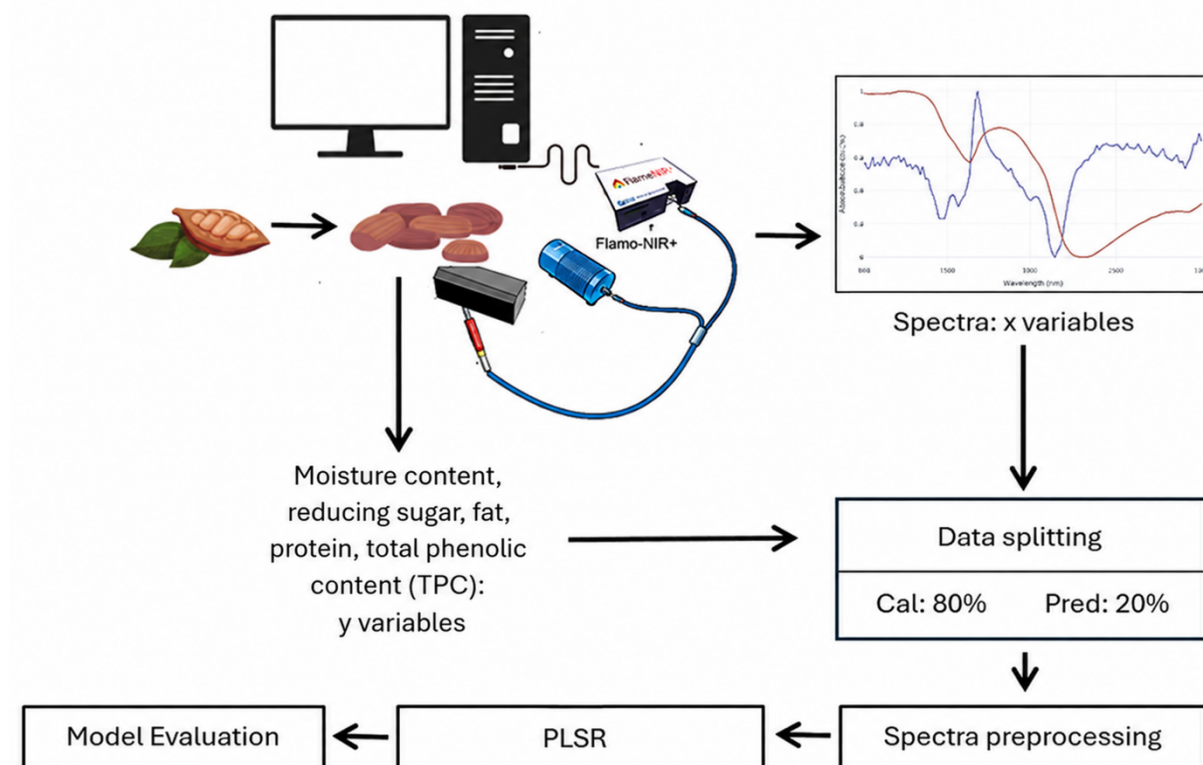


Figure 2. Flow of determination of cocoa nibs quality parameters based on short-wave near infrared (SW-NIR) spectra and chemometrics

PLSR – partial least squares regression

interactions (Nadimi and Paliwal 2024). Still, traditional NIR systems are quite expensive and tend to be limited to laboratory environment. Consequently, portable short-wave near infrared (SW-NIR) offers a viable alternative for the rapid and economical evaluation of quality (Pahlawan et al. 2022).

As examples of cocoa applications, infrared spectroscopy coupled to chemometric approaches has been extensively used for predicting chemical constituents and quality attributes (moisture content, fat, proteins, sugars, acidity, polyphenols and other bioactive compounds) (Krähmer et al. 2015; Batista et al. 2021; Hernández-Hernández et al. 2022). These studies successfully show the consistency of infrared spectroscopy as an efficient analytical method for rapid and non-destructive analysis of cocoa monitoring, quality classification and compositional analysis under multiple specific postharvest processing conditions.

While infrared-based prediction of chemical constituents using partial least squares regression (PLSR) has been reported for cocoa bean separation, it has predominantly been limited to quantification of singular quality parameters and spectroscopic systems employed at various stages of cocoa processing (most studies used FTIR and FT-NIR based laboratory-constructed spectroscopies). Previously, only pH, moisture content, or bioactive compounds were evaluated as cocoa bean quality under specific processing conditions (Batista et al. 2021; Hernández-Hernández et al. 2022). Yet, little is known regarding the feasibility of implementing SW-NIR fibre-optic spectroscopy to predict multiple biochemical quality parameters simultaneously along the whole cocoa postharvest processing chain, including fresh and fermented, dried, and roasted samples of cocoa nibs. Because fermentation, drying, and roasting include extensive biotransformations (Calvo et al. 2021; Jiménez-Rodríguez et al. 2024), the spectral responses as well as model performance may vary significantly over processing steps (Forte et al. 2023). Thus, this work fills an existing knowledge gap using a systematic evaluation of SW-NIR (950–1 700 nm) combined with PLSR for predicting moisture content, reducing sugars, fat, protein and total phenolic content across four cocoa processing stages in a unified experimental and modelling framework. This study investigates the potential for using a single SW-NIR fibre-optic system to predict multiple postharvest quality attributes at different processing stages in comparison with previous studies that derived specific estimates of one or just a few individual quality non-confor-

mities and/or from laboratory-based spectroscopic systems.

Hence, the present study aimed to build and validate PLSR models using SW-NIR spectra data (950–1 700 nm) for predicting moisture content, reducing sugars, fat, protein, and total phenolic content in fresh, fermented, dried and roasted cocoa nibs. Moreover, the work also studied which spectral regions were responsible for model development and how the models performed between cocoa processing stages. This multi-parameter approach therefore offers a practical framework for rapid, non-destructive evaluation of key cocoa quality parameters over the postharvest processing value chain.

MATERIAL AND METHODS

Figure 2 shows the flow of determination of cocoa nibs quality parameters based on near-infrared spectra and chemometric model.

Raw materials and sample preparation. Forastero cocoa pods were obtained from a plantation in Nglangeran Village, Special Region of Yogyakarta, Indonesia. Twenty cocoa pods with different ripening levels were collected to represent natural biological variability among samples. Approximately 20 cocoa beans were randomly taken from each pod, resulting in a total of 400 individual cocoa bean samples for spectral acquisition and laboratory analysis

The cocoa beans were manually peeled to obtain cocoa nibs (dehulled cotyledons), which were subsequently classified into four postharvest processing stages: fresh (RCB), fermented (FCB), dried (DCB), and roasted (RSCB). Fermentation was carried out in tiered wooden fermenters covered with banana leaves for seven days (168 h), during which the temperature naturally increased to approximately 45 °C due to microbial activity. Drying was conducted at 50 °C using a controlled hot-air dryer until the samples reached a final moisture content of approximately 7% (wet basis). Roasting was performed at 130 °C for 25–30 min.

The cocoa nib samples were grouped according to processing stage and randomly selected within each group to ensure representative variability for chemometric modelling. All spectral measurements were conducted on intact cocoa nibs (dehulled cotyledons) at each processing stage. No grinding or powder preparation was applied before SW-NIR spectral acquisition. Each sample was scanned twice using the SW-NIR spectrometer, and the averaged spectrum was used for further analysis.

Chemical analysis. Chemical parameters, including moisture content, reducing sugars, fat, protein, and total phenolic compounds, were measured using AOAC-recommended reference methods. All chemical analyses were performed in triplicate. Moisture content was determined by the oven-drying method (AOAC 934.01) at 105 °C until constant weight (Forte et al. 2023). Fat content was analysed using the Soxhlet extraction method (AOAC 963.15) with petroleum ether solvent (Batista et al. 2021). Protein content was determined using the Kjeldahl method (AOAC 979.09) with a nitrogen conversion factor of 6.25. Reducing sugars were analysed using the DNS method with absorbance measurements at a wavelength of 543 nm (Batista et al. 2021). Total phenolic content (TPC) was analysed using the Folin–Ciocalteu method and expressed as gallic acid equivalents (mg GAE·g⁻¹ sample) (Hernández-Hernández et al. 2022).

SW-NIR spectral acquisition. Spectral measurements were performed using a Flame-NIR+Spectrometer (Ocean Optics, USA) in the wavelength range of 954–1 700 nm in reflectance mode. Measurements were conducted under laboratory conditions at 25 ± 2 °C. The probe was positioned at a fixed distance (~10 mm) from the sample surface. The integration time was set to 400 ms, and 35 scans were averaged for each spectrum. Each sample was measured at two different positions, and the spectra were averaged to reduce local heterogeneity.

Chemometric and statistical analysis. Spectral data were pre-processed using four different mathematical pretreatment methods: Savitzky–Golay derivatives, multiple scatter correction (MSC), standard normal variate (SNV), and normalisation. The dataset was divided into calibration (80%) and validation (20%) sets using stratified sampling to maintain proportional representation across processing stages. The number of calibration and validation samples varied among quality parameters because several cocoa nib samples did not have complete laboratory reference measurements for all chemical analyses. Consequently, each PLSR model was developed using the maximum number of valid observations available for the corresponding parameter, ensuring that only complete paired spectral and reference data were included in model calibration and validation. Outlier detection was performed using principal component analysis (PCA) based on Hotelling's T^2 and Mahalanobis distance. Partial least squares regression (PLSR) models were developed separately for each quality parameter. The optimal number of latent variables was select-

ed based on the lowest cross-validation error. Model performance was evaluated using the coefficient of determination (R^2), root mean square error (RMSE), and residual prediction deviation (RPD) (Masithoh et al. 2024). The goal of PLSR is to find the linear relationship between the X (independent variable) and Y variables (dependent variables) using an $m \times q$ matrix B of regression coefficients and an error matrix E (Kandpal et al. 2016).

Regression coefficients were used to evaluate the relative contribution of individual wavelengths to each PLSR model. Within a given model, wavelengths with larger absolute β -coefficients were interpreted as having a stronger relative contribution to prediction, whereas coefficients close to zero indicated a smaller contribution. The sign of the coefficient indicates the direction of the relationship between spectral response and the predicted reference value. Because adjacent SW-NIR wavelengths are highly collinear, β -coefficients were interpreted primarily within individual models and in relation to broader spectral regions rather than as independent effects of individual wavelengths.

Pearson correlation analysis was performed using laboratory reference measurements of moisture content, fat content, protein content, reducing sugar, and total phenolic content (TPC). Correlation coefficients were calculated separately for each processing stage (fresh, fermented, dried, and roasted cocoa nibs) to evaluate pairwise relationships among the measured quality parameters. Statistical significance was assessed at $P < 0.05$.

All spectral preprocessing, principal component analysis (PCA), partial least squares regression (PLSR), and statistical analyses were performed using Python 3.11. Data processing and model development were conducted using NumPy, Pandas, SciPy, Scikit-learn, and Matplotlib libraries.

RESULTS AND DISCUSSION

SW-NIR spectral characteristics. Similar overall trends of the SW-NIR spectra (954–1 700 nm) were observed across all processing stages, with differences in absorption intensities reflecting compositional changes. O-H, C-H, and N-H functional groups were associated with major absorption bands.

Figure 3B shows the spectra pre-processed as the Savitzky–Golay second derivatives for four different post-harvest cocoa processing stages: fresh beans, fermentation, drying and roasting in their own interval

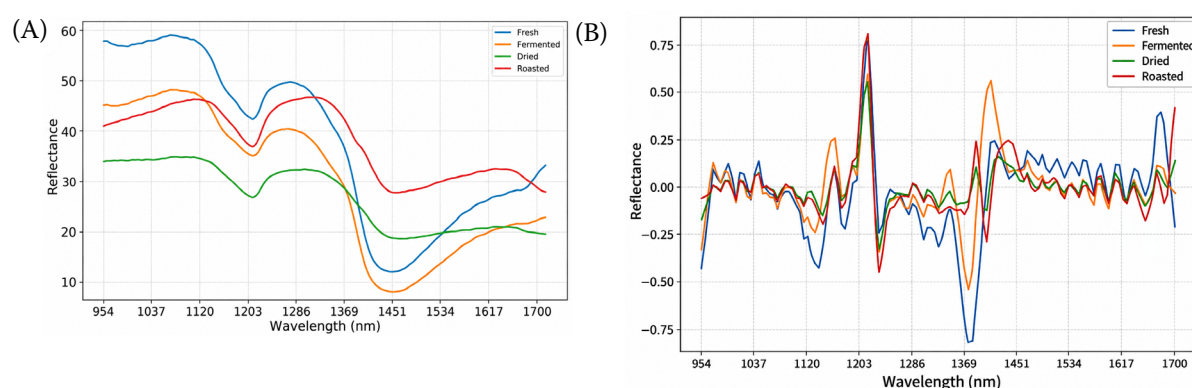


Figure 3. The averaged SW-NIR reflectance spectra of four stages of cocoa nibs using (A) original spectra and (B) SGD2 spectra

SW-NIR – short-wave near infrared; SGD2 – Savitzky-Golay second derivative

(900–1 700 nm). Savitzky–Golay second derivatives enhance spectral resolution and remove baseline drift as well as scattering effects, separating overlapping absorption bands. The indicators of moisture appeared as peaks that occurred at approximately 970–1 000 nm and 1 440–1 470 nm, which indicate decreasing trend in intensity from fresh to roasted samples (Anyidoho et al. 2021). The bands around 1 200–1 250 nm associated

with reducing sugars were related to the carbohydrate transformations during fermentation and roasting (Agussabti et al. 2020). Absorption due to fats was observed at 1 200–1 300 nm and 1 700 nm (overtone of C-H) (Castillejos-Mijangos et al. 2023). Protein signals around 1 510–1 535 nm could be identified (Battista et al. 2021), and TPC related features were obtained in the areas of 1 100–1 150 nm and 1 400–1 450 nm (Xu

Table 1. Statistical characteristics of calibration and validation datasets for cocoa nib quality parameters at different processing stages

Component	Sample	Calibration set				Validation set			
		<i>n</i>	Min–Max	Mean	SD	<i>n</i>	Min–Max	Mean	SD
Water content	RCB	140	28.4–51.0	33.69	4.71	38	28.5–51.0	37.4	7.34
	FCB	133	39.2–47.7	43.33	2.69	36	39.4–47.4	43.95	2.62
	DCB	144	4.0–5.6	4.83	0.63	40	4.0–5.6	4.71	0.65
	RSCB	142	2.5–5.6	3.93	1.31	39	2.5–5.6	4.36	1.33
Reducing sugar	RCB	129	9.3–12.8	10.49	0.81	36	9.4–12.7	10.39	0.73
	FCB	144	7.2–12.9	10.87	1.3	40	7.1–12.9	10.88	1.52
	DCB	78	2.7–13.3	8.27	3.15	21	3.7–13.5	7.27	3.16
	RSCB	141	3.0–7.3	4.81	1.28	39	3.9–7.3	5.26	1.34
Fat	RCB	102	57.2–74.1	68.33	4.24	28	57.4–74.2	68.38	4.89
	FCB	100	70.2–75.9	73.22	1.93	27	70.2–75.8	72.88	2.01
	DCB	144	72.7–75.7	74.04	0.49	40	71.8–75.8	73.94	0.75
	RSCB	131	70.5–75.9	72.77	0.81	36	70.5–75.9	72.74	0.91
Protein	RCB	136	1.8–2.8	2.13	0.25	37	1.8–2.7	2.14	0.26
	FCB	144	1.2–2.1	1.51	0.33	40	1.2–2.0	1.61	0.35
	DCB	126	1.1–1.3	1.24	0.04	34	1.1–1.3	1.23	0.04
	RSCB	136	1.1–1.2	1.2	0.03	38	1.1–1.2	1.19	0.04
Total phenolic content (TPC)	RCB	144	1.3–14.4	6.04	3.7	40	1.3–14.6	5.4	3.56
	FCB	111	6.0–8.8	7.37	0.79	31	6.0–8.9	7.38	0.84
	DCB	144	16.6–36.3	28.45	5.76	40	16.7–36.2	29.2	5.59
	RSCB	142	2.5–5.6	3.93	1.31	39	2.5–5.6	4.36	1.33

RCB – fresh cocoa nibs; FCB – fermented cocoa nibs; DCB – dried cocoa nibs; RSCB – roasted cocoa nibs

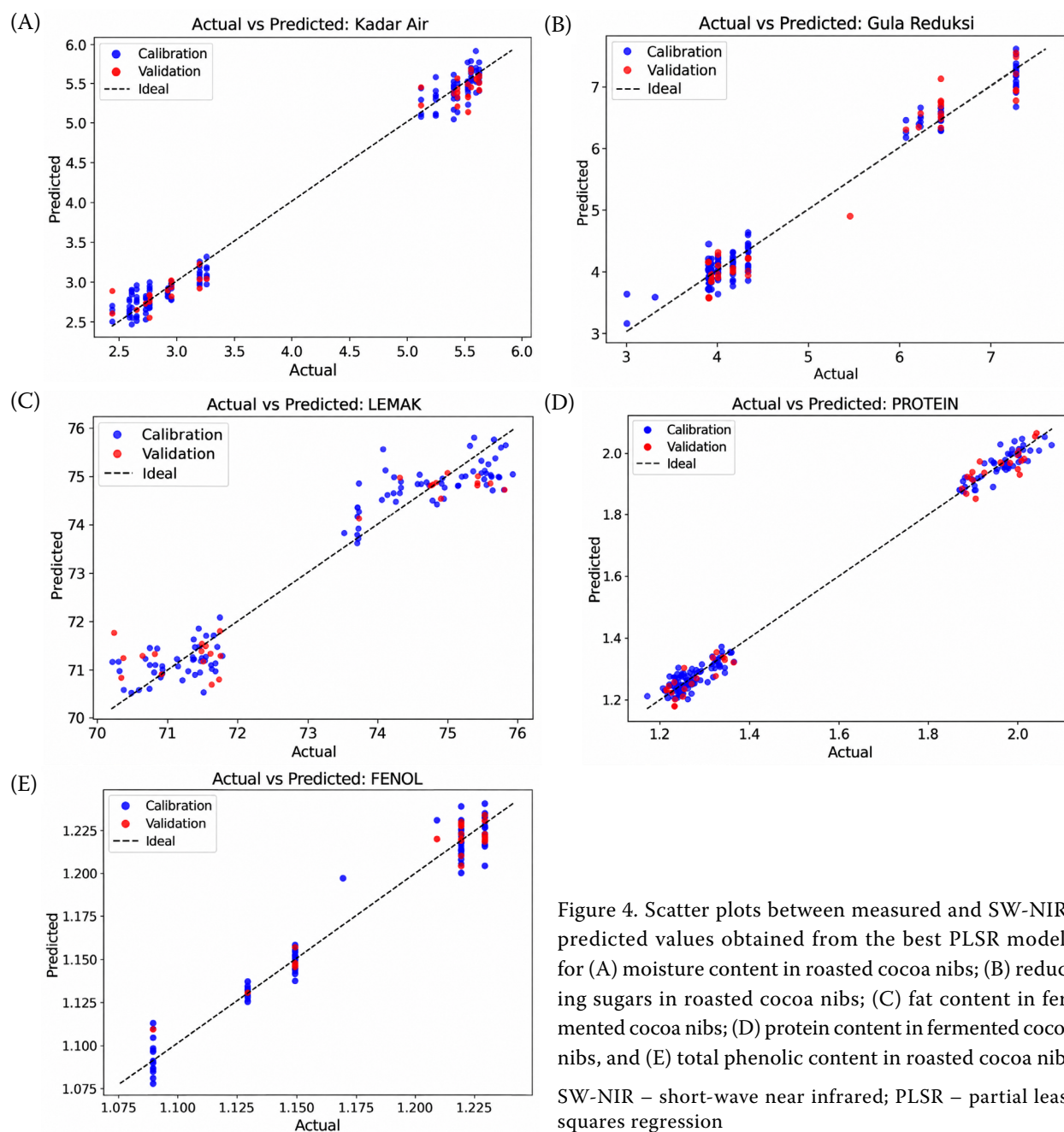


Figure 4. Scatter plots between measured and SW-NIR-predicted values obtained from the best PLSR models for (A) moisture content in roasted cocoa nibs; (B) reducing sugars in roasted cocoa nibs; (C) fat content in fermented cocoa nibs; (D) protein content in fermented cocoa nibs, and (E) total phenolic content in roasted cocoa nibs. SW-NIR – short-wave near infrared; PLSR – partial least squares regression

et al. 2025). These spectral characteristics show that the SW-NIR is adequate for the identification of biochemical changes that occur during cocoa processing.

Moisture content, reducing sugars, fat, protein and total phenolic content of cocoa nib samples taken from the processing stages are expressed as means $\pm$ SD (Table 1). These values are cubature measurements taken from laboratory sample of cocoa nib at each stage. Since the cocoa pods were harvested at different stages of ripening then subjected to diverse postharvest processing, such as fermentation, drying, and roasting, some variability

between samples was anticipated. Thus, in order to summarise the overall compositional profile and the variation of cocoa nibs through each processing stage, the average values presented in Table 1 were considered.

The spectral data were combined with their reference values and divided into calibration set and validation set in which specific information is displayed in Table 1 before PLS regression. Different statistical properties of the calibration and validation sets (including moisture, reducing sugar, fat, protein and total phenolic content) for each processing stage are shown in Table 1.

The ranges and standard deviations of fat and protein across stages were relatively small, indicating a more constant distributions compared to the other nutrients. In contrast, TPC showed a larger variation after drying. While this numerical variability is useful for model building, if the data are not well modelled they may be subjected to smaller but higher prediction errors (Thukral et al. 2019). Conversely, if the variability is too little, it may be hard for the model to identify any worthwhile trends (Li 2016). In general, a balanced dataset is important to have reliable model calibration.

PLSR model performance. A Pearson correlation was performed to determine the relationships among moisture content, fat, protein, reducing sugars and total phenolic content for each processing step (Table 2). In summary, the associations tended to be very weak. Very few relationships were statistically significant ($P < 0.05$), such as moisture content vs. TPC in fresh beans ($r = 0.187$) and reducing sugars vs. TPC in dried beans ($r = 0.182$).

One expects that the diagonal of a correlation matrix would contain scores of 1.000, because it indicates

each variable correlated with itself (and also always exactly correlates perfectly). However, the low coefficients of correlation generally show that most quality parameters changed rather independently during cocoa processing. The only meaningful correlations we found indicate weak associations between some quality attributes, but the low correlation values suggest that such associations were weak.

PLSR models performed well overall, although performance was parameter-dependent and particularly improved after preprocessing of the spectra. Moisture content presented the maximum variability, it was strongly calibrated and generally improved prediction. Fat and protein were moderate predictors as their ranges of variation offered less (in comparison) variability for prediction. The decrease of reducing sugars seen during stages of processing reflects fermentation and Maillard reactions, which these trends suggest have been well captured through the spectrum. Of note, the predictions of proteins were very accurate, substantiating a constant correlation with respect to the N-H absorption bands.

Table 2. Pearson correlation coefficients calculated from laboratory reference measurements of cocoa nib quality parameters at different postharvest processing stages

Sample	Component	Moisture	Fat	Protein	Reducing sugar	TPC
RCB	Moisture	1.000	−0.098	0.043	−0.048	0.187
	Fat	−0.098	1.000	0.063	−0.038	−0.106
	Protein	0.043	0.063	1.000	0.037	0.095
	Reducing sugar	−0.048	−0.038	0.037	1.000	0.055
	TPC	0.187	−0.106	0.095	0.055	1.000
FCB	Moisture	1.000	−0.052	−0.051	0.110	0.014
	Fat	−0.052	1.000	−0.083	0.003	0.150
	Protein	−0.051	−0.083	1.000	−0.039	0.054
	Reducing sugar	0.110	0.003	−0.039	1.000	0.058
	TPC	0.014	0.150	0.054	0.058	1.000
DCB	Moisture	1.000	−0.143	0.173	0.138	−0.043
	Fat	−0.143	1.000	0.087	−0.073	0.062
	Protein	0.173	0.087	1.000	0.010	−0.060
	Reducing sugar	0.138	−0.073	0.010	1.000	0.182
	TPC	−0.043	0.062	−0.060	0.182	1.000
RSCB	Moisture	1.000	0.061	−0.045	−0.001	−0.012
	Fat	0.061	1.000	−0.109	0.090	0.033
	Protein	−0.045	−0.109	1.000	−0.054	−0.002
	Reducing sugar	−0.001	0.090	−0.054	1.000	0.104
	TPC	−0.012	0.033	−0.002	0.104	1.000

TPC – total phenolic content; RCB – fresh cocoa nibs; FCB – fermented cocoa nibs; DCB – dried cocoa nibs; RSCB – roasted cocoa nibs

Table 3. Best performance of PLSR (partial least squares regression) models for all components of cocoa nibs at each process stage

Sample	Component	Preprocess- ing	R^2C	RMSEC	R^2P	RMSEP	RPD
RCB	1. Water content	SG1	0.947	1.081	0.919	2.058	3.522
	2. Reducing sugar	SG2	0.949	0.183	0.867	0.262	2.741
	3. Fat	SG2	0.835	1.713	0.741	2.442	1.967
	4. Protein	SG1	0.920	0.071	0.877	0.090	2.847
	5. Total phenolic content	SG1	0.976	0.567	0.945	0.825	4.262
FCB	1. Water content	SG1	0.944	0.668	0.863	1.234	2.700
	2. Reducing sugar	SG1	0.951	0.287	0.944	0.357	4.209
	3. Fat	SG1	0.937	0.481	0.910	0.593	3.329
	4. Protein	SG1	0.994	0.025	0.992	0.031	10.992
	5. Total phenolic content	SG1	0.922	0.221	0.874	0.294	2.817
DCB	1. Water content	SG1	0.976	0.097	0.957	0.133	4.816
	2. Reducing sugar	SG1	0.965	0.585	0.891	1.021	3.023
	3. Fat	SG1	0.914	0.143	0.902	0.233	3.198
	4. Protein	SG1	0.944	0.009	0.870	0.015	2.774
	5. Total phenolic content	SG1	0.950	1.280	0.824	2.318	2.383
RSCB	1. Water content	SG1	0.989	0.140	0.984	0.164	7.998
	2. Reducing sugar	SG1	0.975	0.201	0.958	0.272	4.860
	3. Fat	SG1	0.925	0.221	0.863	0.332	2.698
	4. Protein	SG1	0.979	0.005	0.966	0.007	5.413
	5. Total phenolic content	SG1	0.973	0.008	0.954	0.008	4.646

SG1 – Savitzky–Golay first derivative; SG2 – Savitzky–Golay second derivative; R^2C and R^2P – coefficient of determination in calibration and prediction of the regression line, respectively; $RMSEC$ and $RMSEP$ – root mean square error in calibration and prediction, respectively; RPD – residual prediction deviation; RCB – fresh cocoa nibs; FCB – fermented cocoa nibs; DCB – dried cocoa nibs; RSCB – roasted cocoa nibs

These findings are consistent with previous studies reporting the potential of NIR spectroscopy for predicting multiple cocoa quality attributes, including moisture and protein (Hashimoto et al. 2018; Hernández-Hernández et al. 2022). The comparatively lower performance of fat prediction in the present study may be related to overlapping C-H absorptions, matrix effects in intact cocoa nibs, and the restricted spectral range, which may limit the availability of more specific lipid-related spectral information. Total phenolic content varied markedly across processing stages, with the lowest mean value observed after roasting. This stage-dependent variation is consistent with the complex transformations of cocoa polyphenols during fermentation, drying, and roasting, including oxidation, polymerisation, and thermal degradation (Calvo et al. 2021; Alean et al. 2022; La Mantia et al. 2023). In summary, spectral preprocessing reduces noise and the spectra tend to highlight information that is

chemically meaningful which leads to better model performances. These results indicate that SW-NIR spectroscopy and PLSR can be employed as non-destructive methods to predict several quality of cocoa nibs in postharvest processes.

The best PLSR models for the quality parameter of cocoa across all processing stages can be found in Table 3. The most robust predictions derived from moisture more generally, particularly for roasted cocoa nibs ($R^2P = 0.984$; $RPD = 7.998$). This may be associated with the strong O-H absorption bands found in SW-NIR range. Additionally, in fermented and roasted samples, protein and reducing sugars showed R^2P values greater than 0.94 and RPD values over 4.0, indicating excellent calibration accuracy.

In contrast, fat and total phenolic content had lower predictive performances across several stages. For example, fat in fresh cocoa nibs ($R^2P = 0.741$; $RPD = 1.967$); total phenolic content in dried cocoa nibs ($R^2P = 0.824$; $RPD = 2.383$). This might be attributable to overlap-

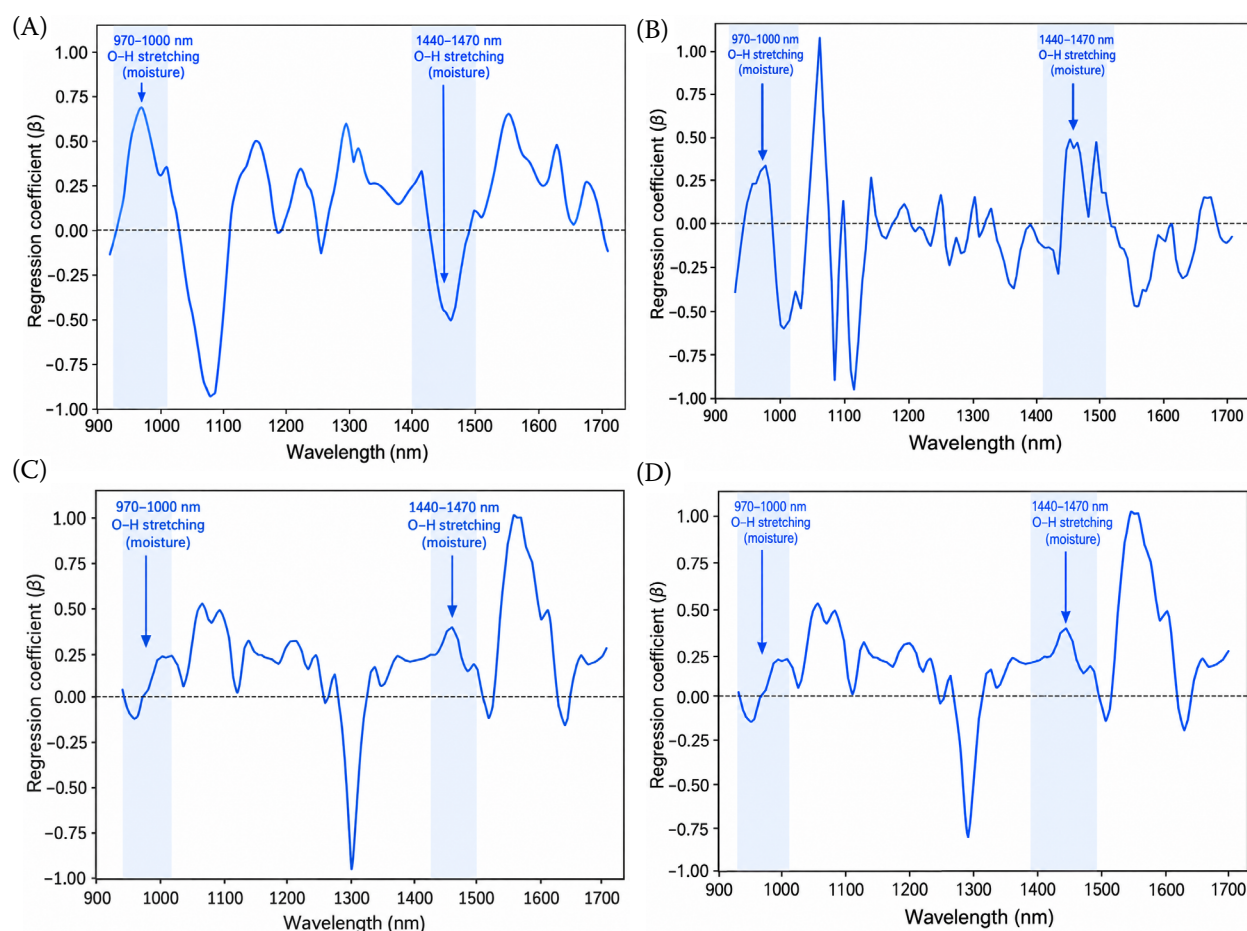


Figure 5. β -coefficients from PLSR models based on SW-NIR spectra for moisture content of (A) fresh cocoa nibs; (B) fermented cocoa nibs; (C) dried cocoa nibs; (D) roasted cocoa nibs

SW-NIR – short-wave near infrared; PLSR – partial least squares regression

ping absorption bands together with lipid and phenolic compounds – whose spectra are less attractive in the SW-NIR region. Such conclusions have been reported in previous cocoa spectroscopy studies (Hashimoto et al. 2018; Batista et al. 2021).

Savitzky–Golay derivative preprocessing also consistently ranked as the highest for most quality parameters across the board. This indicates that the derivatives with reduced baseline shifts and scattering effects also enhanced chemical information extraction from spectra.

Figure 4 presents the measured versus predicted plots obtained from the best PLSR models for each evaluated quality parameter. The selected models were chosen based on the highest R^2 values and lowest prediction errors among the evaluated preprocessing techniques and cocoa processing stages.

Lastly, the predicted values were well within the regression line, demonstrating that our results are in good agreement with those that we have predicted. While moisture, protein and reducing sugars tended

to be very close clustered/packed around the regression line; fat and total phenolic content on the other hand seemed to deviate more from that regression line in some of the processing stages. These trends corroborate the patterns observed in the statistical results presented in Table 3 and lend support to our hypothesis that representational meaning is being captured in the developed PLSR models.

Interpretation of regression coefficients. Regression coefficient (β -coefficient) analysis was performed to determine the significant wavelengths associated with each PLSR prediction model (Mahesh et al. 2015). The β -coefficient profiles derived from the optimal preprocessing models on cocoa nibs at different processing stages are depicted in Figures 5–9. In this context, β -coefficients emphasise the wavelengths that contribute most to predicting chemical parameters and illustrate how sample composition relates to specific absorption band regions (Jiang et al. 2023).

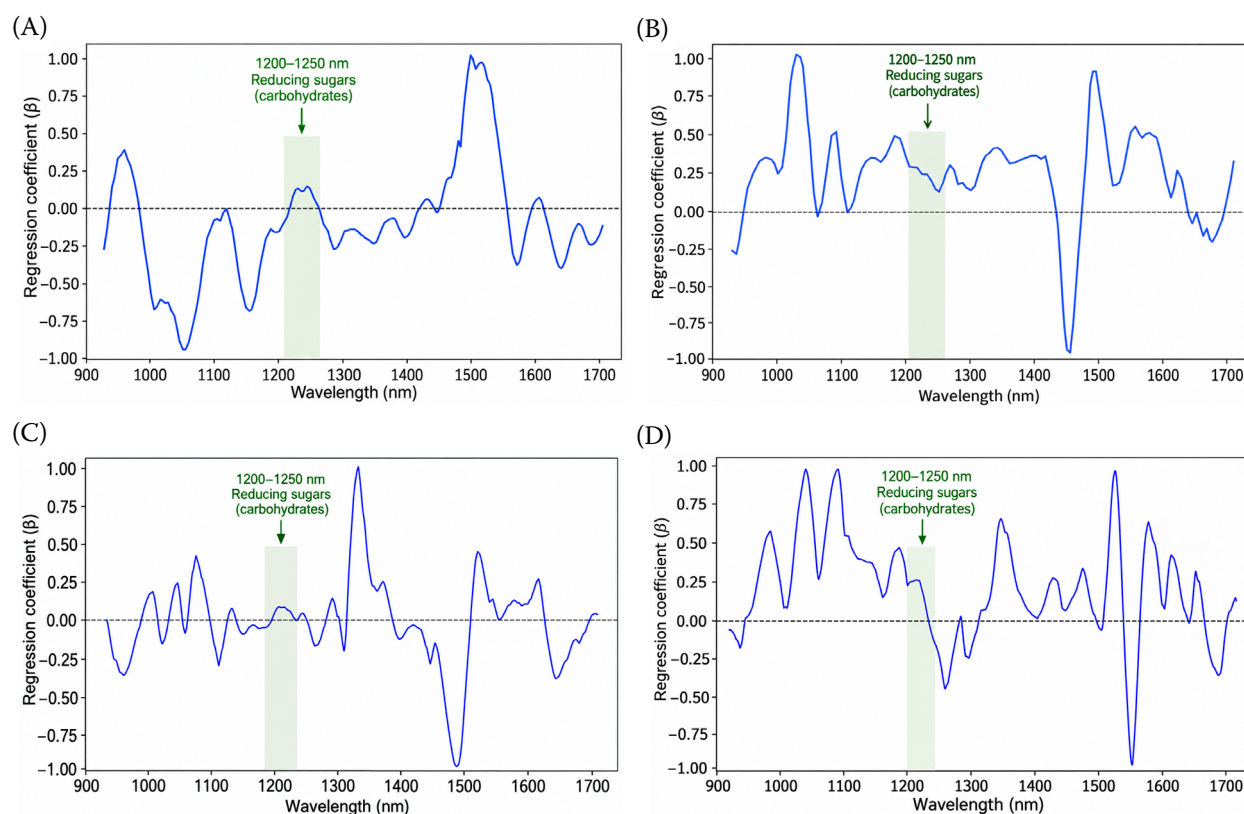


Figure 6. β -coefficients from PLSR models based on SW-NIR spectra for reducing sugar of (A) fresh cocoa nibs; (B) fermented cocoa nibs; (C) dried cocoa nibs; (D) roasted cocoa nibs

SW-NIR – short-wave near infrared; PLSR – partial least squares regression

These profiles also suggest how the spectral (β -coefficient) plots were inspected to interpret how the PLSR models distributed predictive information across the SW-NIR spectral range, rather than merely assigning signals to absorption bands previously identified in earlier work. In general, regression coefficients in PLSR can be interpreted to estimate how influential each individual wavelengths is: larger absolute coefficient values indicate greater relative influence of the associated spectral variables on the predicted response (Wu et al. 2023). Positive and negative coefficients indicate an opposite direction of association between the spectral variables and the response. However, these signals and magnitudes should be cautiously interpreted due to influences by spectral scaling, preprocessing PSR latent-variable structures, and interference among correlated wavelengths (Reggente et al. 2019). Typically, neighbouring wavelengths are strongly correlated in spectroscopic datasets. Thus, pronounced oscillations in the regression-coefficient profiles may represent no true chemical and wavelength-specific information but multicollinearity or spectral interferences (Zheng et al. 2024). Consequently, localised co-

efficient maxima ought to be considered as belonging to larger, chemically relevant spectral domains and interpreted in the context of established O-H, C-H, and N-H absorption bands. Rather, β -coefficients were considered as complementary evidence for an understanding of the model behaviour, while independent validation statistics (i.e. R^2P , $RMSEP$, and RPD) were primarily used to assess reliability and predictive performance (Fodor et al. 2024).

Regression coefficient analysis indicated that the most important wavelengths for predictions related to O-H, C-H, and N-H absorption bands. As for moisture, various O-H regions associated with water were detected as similar values of β -coefficients in the vicinity of 970–1 000 nm and 1 440–1 470 nm among all four processing types (Figure 5). We find that spectral regions correspond to known water-sensitive NIR bands: near 970 nm due to O-H absorption from water and the range of 1 400–1 490 nm driven by O-H-related absorptions and moisture (Teye 2022; Forte et al. 2023). The importance of these regions in the current models indicates that there were spectral changes associated with moisture present, and these contribute positively

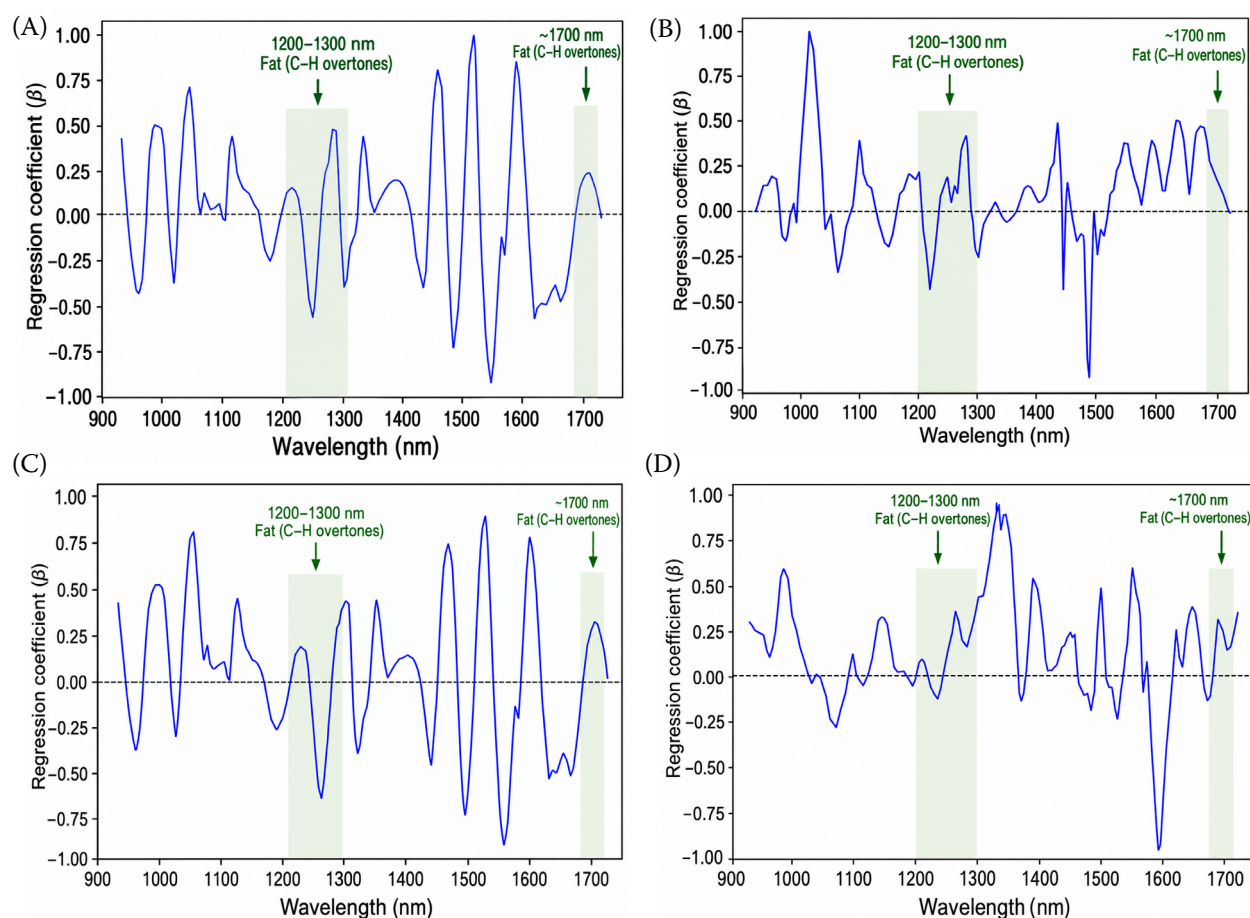


Figure 7. β -coefficients from PLSR models based on SW-NIR spectra for fat of (A) fresh cocoa nibs; (B) fermented cocoa nibs; (C) dried cocoa nibs; (D) roasted cocoa nibs

SW-NIR – short-wave near infrared; PLSR – partial least squares regression

to PLSR predictions; it is not just that the model captured features corresponding with absorption bands already reported. Such an interpretation also agrees with the large differences between reference moisture values at different stages, i.e. higher in fresh and fermented nibs than in dried and roasted nibs (Table 1). Moisture profiles of cocoa beans are altered over the course of fermentation as structural changes to the cellular matrix take place triggering mass transfer between cells. Subsequently, a clear decrease in moisture content is even obtained through drying (Ackah and Dompey 2021; Calvo et al. 2021). Roasting favours water removal from cocoa nibs, as recently demonstrated by de Souza et al. (2024): subjected to under roasting at 135 °C for 35 min resulted in moisture levels of almost equal to 3%. Despite the large drop in moisture after drying and roasting, the O-H-related regions still contributed significance to each PLSR model. The same effect was reflected in the predictive performance: R^2P and RPD for moisture models changed

from 0.919/3.522 (fresh nibs), to 0.957/4.816 (dried nibs), to 0.984/7.998 for roasted nibs (Table 3). In general, these results suggest that the moisture models adequately explained systematic, water-related spectral variability over the full spectrum of moisture conditions. Nevertheless, those patterns of β -coefficient should be interpreted jointly with the independent prediction statistics.

In the case of reducing sugars, the most important β -coefficient features were also detected in ranges including ~1 200–1 250 nm (Figure 6). The particular of the spectrum corresponds to NIR signals that react with carbohydrates. Recent works have associated features from the ~1 176–1 212 nm range with the second overtone of C-H stretching in carbohydrates and reported additional carbohydrate-related features around 1 200–1 240 nm (López-Calabozo et al. 2024). As NIR absorption bands are broad and frequently overlap, however, these coefficients cannot be regarded as unique signatures for reducing sug-

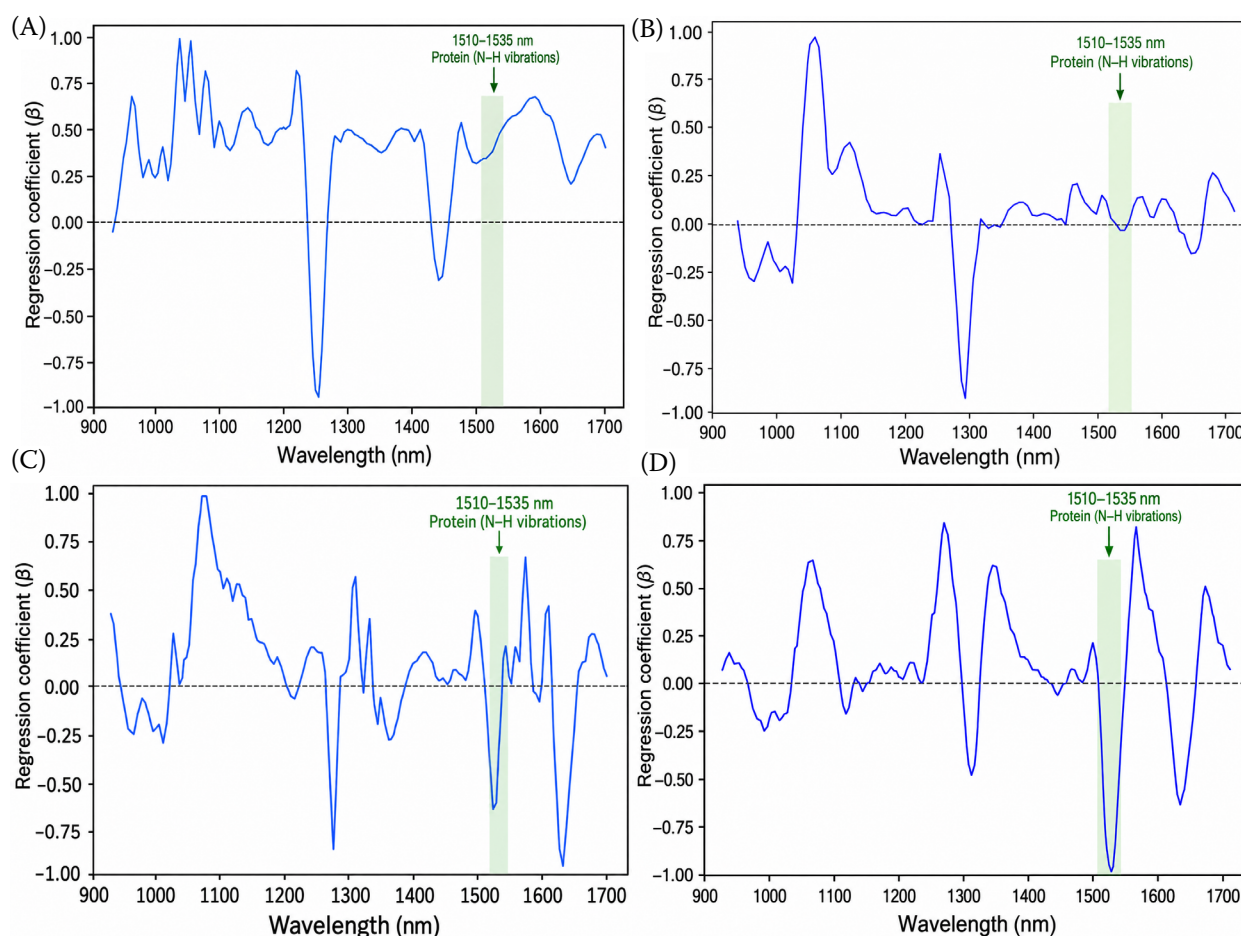


Figure 8. β -coefficients from PLSR models based on SW-NIR spectra for protein of (A) fresh cocoa nibs; (B) fermented cocoa nibs; (C) dried cocoa nibs; (D) roasted cocoa nibs

SW-NIR – short-wave near infrared; PLSR – partial least squares regression

ars. Rather, the pattern is indicative of PLSR models drawing on carbohydrate-related spectral information in conjunction with co-varying changes in the cocoa matrix. The differences in β -coefficient patterns between the fresh, fermented-dried and roasted nibs also support that reducing sugars are liable during cocoa processing. At fermentation, although reducing sugars in the bean may be decreased because they are used up in biochemical reactions, it can also increase due to carbohydrate hydrolysis and diffusion of reducing sugars from surrounding pulp. Thus, the net change depends on which processes take over (Calvo et al. 2021). That balance produced reference mean values for fresh and fermented nibs that were similar (10.49 and 10.87, respectively), followed by lower values after drying (8.27) and roasting processes (4.81) in the study (Table 1). Reducing sugars are major Maillard reaction precursors and can decrease in roasting as aroma compounds form. Reducing sugars have also been found to decrease during roast-

ing in experimental cocoa studies (Andi Pallawa et al. 2023; Moreno-Trujillo et al. 2024). Despite these substantial changes in composition and matrix types, prediction ability of the fermented and roasted reducing-sugar models remained relatively good, with $R^2P/$ RPD values of 0.944/4.209 and 0.958/4.860 (Table 3). In general, these results indicate that the β -coefficient spatial distributions contain relevant predictive information associated with reducing-sugar variations at loaf and bran processing transitions. However, they do not provide support for a single sugar-specific absorption band.

The contributions to β -coefficient profiles from fat occurred mainly in the C-H-related region at approximately 1 200–1 300 nm, with strength, distribution varying by processing stages (Figure 7). This is chemically reasonable for predicting fat as a lower fraction of lipid in cocoa beans and NIR absorption ~1 200 nm has been assigned to C-H vibrations related to fat in cocoa (Forte et al. 2023). However, the absorp-

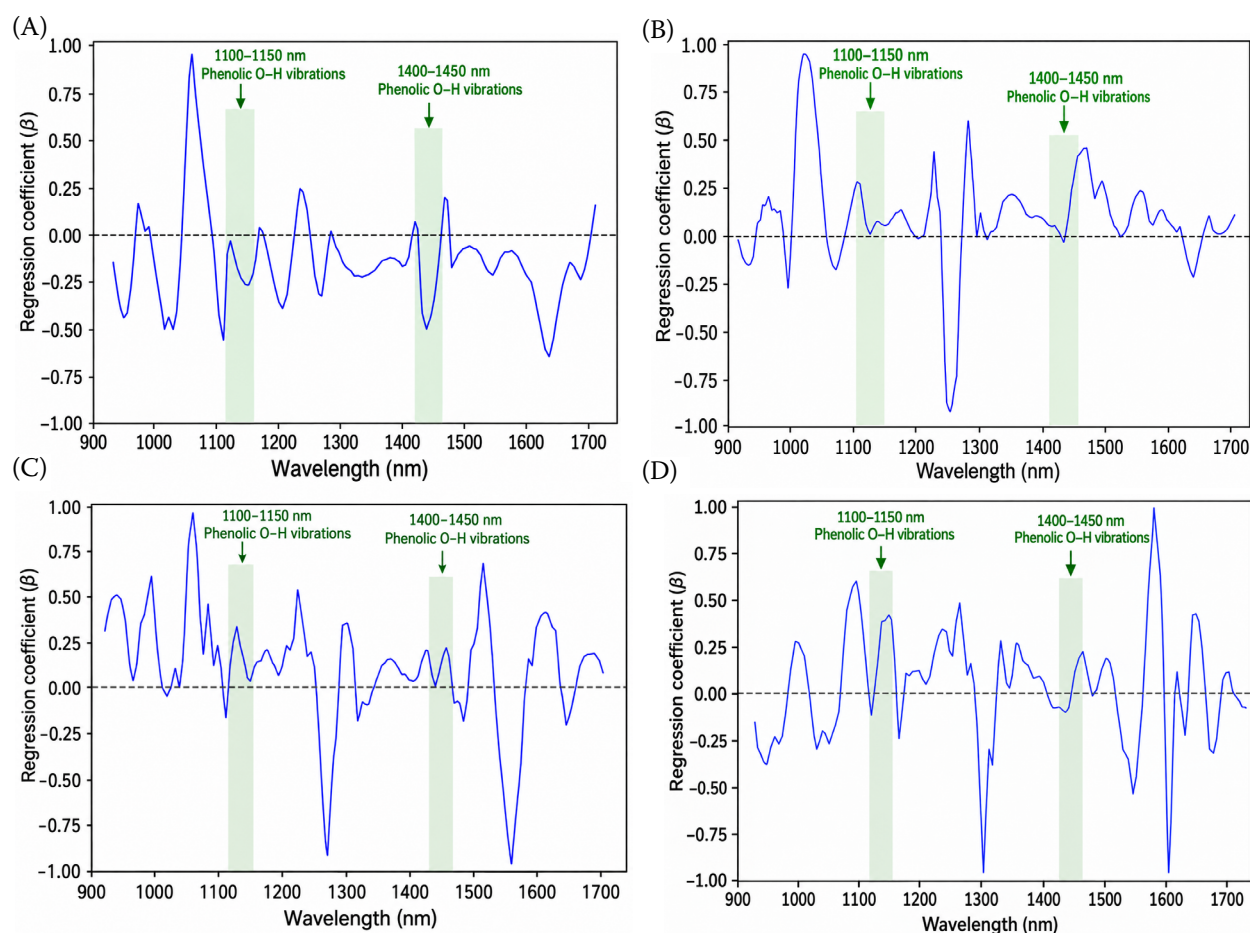


Figure 9. β -coefficients from PLSR models based on SW-NIR spectra for total phenolic content of (A) fresh cocoa nibs; (B) fermented cocoa nibs; (C) dried cocoa nibs; (D) roasted cocoa nibs

SW-NIR – short-wave near infrared; PLSR – partial least squares regression

tions related to the lipids do not take place in only one area. We reported that first-overtone C-H bands from CH_2 groups are generally stronger at 1 725–1 744 nm (Forte et al. 2023). The SW-NIR range used in the current work extended only to 1 700 nm, thus those relatively stronger lipid bands fall outside of the measured spectral window. Consequently, the fat models relied primarily on short-wavelength C-H information and multivariate covariance with extraneous spectral variation in the cocoa matrix alone. This restricted spectral coverage (combined with C-H signal overlapping those from lipids and other organic constituents) may partly account for the lower fat prediction accuracy in comparison to moisture, reducing sugar, protein and total phenolic content.

The variation in stage-specific β -coefficients also must be considered contextually where shifts in cocoa matrix do not directly serve as evidence towards denouncing large lipid synthesis or degradation via processing. Fermentation of cocoa has been shown

to induce small changes in total fat and fatty-acid composition (Calvo et al. 2021; Herrera-Rocha et al. 2022), while roasting too has only been reported to induce limited relative changes in total lipid content on a dry-matter basis (Moussy Moundanga et al. 2024). However, in this study, mean fat values increased from 68.33 in the fresh nibs to 73.22 and 74.04 in the fermented and dried nibs, respectively, and finally decreasing slightly to 72.77 in roasted nibs (Table 1). Thus, these differences need to be interpreted with caution as relative composition and spectral responses can change due to variations in moisture content and/or matrix structure during fermentation, drying, and roasting that may or may not correspond directly to changes in the absolute lipid fraction. Indices obtained with PLSR exhibited this variability: $R^2P/$ $RPD = 0.741/1.967$ (fresh), $0.910/3.329$ (fermented), $0.902/3.198$ (dried) and $0.863/2.698$ (roasted) respectively (Table 3). The worse performance for fresh nibs – on the contrary even when chemically plausible

C-H-related coefficients are present – demonstrates that assigning coefficients to the 'correct' region is not sufficient. Instead, the prediction of lipid content was dependent on the accessibility of spectral information related to lipids alone (not overlapping with other constituents) and how the physical and chemical matrix of the intact cocoa nibs changed upon processing.

In the protein, the significant β -coefficient contributions occurred in 1 510–1 535 nm that is associated with N-H region (Figure 8). It is a well-known protein sensitive region in SW-NIR associated with N-H absorption (Forte et al. 2023; van Rooyen et al. 2023). The shifts in coefficients through processing stages reflect what happens to proteins within cocoa while it is processed. Peptides and free amino acids are generated due to these storage proteins underwent proteolysis through fermentation. Then, these compounds enter Maillard-related reactions during roasting (Calvo et al. 2021; Peña-Correa et al. 2023; Balcázar-Zumaeta et al. 2024). Protein predictions were strongest for fermented and roasted nibs, R^2P ; RPD : 0.992;10.992 and 0.966;5.413, indicating protein spectral properties were captured at those stages. However, the fact that dried nibs ($R^2P = 0.870$; $RPD = 2.774$) were accompanied with less good performance shows that simply finding β -coefficient for chemically plausible N-H wavelengths does not automatically mean excellent predictive accuracy. The β -coefficient profiles should therefore be treated as component signatures of the overall multivariate response of the cocoa matrix, and not as isolated protein absorption markers.

The β -coefficient profiles for TPC were distributed over several spectral regions, including approximately 1 100–1 150 and 1 400–1 450 nm (Figure 9). The region around 1 400–1 450 nm is a chemically logical basis for phenolic-difference variation, as NIR responses in the 1 400 and 1 440 nm have been associated with aliphatic alcohols and phenolic O-H groups existing in cocoa. Nonetheless, signals close to 1 450 nm may be affected by contributions overlapping with water and other O-H-containing compounds (Forte et al. 2023). Consequently, the coefficients observed in those regions should be interpreted as multivariate descriptors of phenolic-like compositional changes and not for specific phenolic compounds. This explanation is consistent with the significant changes observed in cocoa polyphenols during processing. Polyphenols will diffuse during fermentation and be oxidised by enzymes converting flavanols and procyanidins into oxidised,

polymerised products (Calvo et al. 2021). Oxidative changes can continue during drying, and catechins and procyanidins retention or degradation is affected by the drying temperature (Alean et al. 2022; Maldonado and Figueroa 2023). The phenolic profile can be modified also further during roasting, with reductions of the total phenolic content and also several flavanols having been observed in response to thermal processing (La Mantia et al. 2023). TPC models performed better for both fresh and roasted nibs in this study (Table 3). These differences illustrate the same point – these differences demonstrate that the presence of chemically meaningful β -coefficient regions alone may not be sufficient to achieve reliable prediction. The quality of prediction, therefore, is only dependent on the transformation of phenolics between the raw and processed cocoa beans, the extent with which spectral signals overlap (if present), and upon how processing affects the remaining cocoa matrix.

The β -coefficient profiles inferred from the PLSR models demonstrated that the spectral information was utilised differently in relation to quality parameters and processing stages examined. Overall, the areas with chemically plausible coefficients matched more fully with higher performing models. However, even having coefficients in the 'right-looking' wavelength ranges were often not sufficient for accurate predictions – for instance, the fat model using fresh nibs and TPC with dried nibs performed worse than expected. Regression-coefficient patterns can also be influenced by spectral collinearity, sample-matrix features, preprocessing options and biologic variability (Hayati et al. 2021; Hernández-Hernández et al. 2022; Zheng et al. 2024). In that regard, β -coefficients were primarily used to assist in the interpretation of model behaviour whereas R^2P , $RMSEP$ and RPD remained our main metrics for predictive performance.

CONCLUSION

In this study, we evaluated the ability of spectroscopy (954–1 700 nm) and partial least squares regression (PLSR) to predict moisture content, reducing sugars, fat, protein and total phenolic content in cocoa nibs from all postharvest processing stages. The resulting models has high accuracy for moisture, reducing sugars, protein, and total phenolic content prediction while fat prediction was the least accurate. In this analysis, the contributions of the different wavelengths were correlated to O-H, C-H, and N-H absorption bands are

evident as they are associated with key biochemical constituents in cocoa nibs. Savitzky–Golay derivatives were consistently the best preprocessing method out of those tested. In general, the results demonstrate that SW-NIR spectroscopy could evaluate varying biochemical quality traits in cocoa nibs, providing a practical non-destructive tool to assist with monitoring of quality during processing. More work using larger datasets, additional cocoa varieties and processing conditions is needed to test the level of transferability and generality of the models.

Acknowledgement. The authors would like to acknowledge Universitas Jenderal Soedirman and Universitas Gadjah Mada for providing research facilities and academic support throughout this study.

REFERENCES

- Ackah E., Dompey E. (2021): Effects of fermentation and drying durations on the quality of cocoa (*Theobroma cacao* L.) beans during the rainy season in the Juaboso District of the Western-North Region, Ghana. *Bulletin of the National Research Centre*, 45: 175.
- Afoakwa E.O., Paterson A., Fowler M. (2007): Factors influencing rheological and textural qualities in chocolate – A review. *Trends in Food Science & Technology*, 18: 290–298.
- Agussabti, Rahmaddiansyah, Satriyo P., Munawar A.A. (2020): Data analysis on near infrared spectroscopy as a part of technology adoption for cocoa farmer in Aceh Province, Indonesia. *Data in Brief*, 29: 105251.
- Alean J., Chejne F., Ramírez S., Rincón E., Alzate-Arbelaez A.F., Rojano B. (2022): Proposal of a method to evaluate the *in-situ* oxidation of polyphenolic during the cocoa drying. *Drying Technology*, 40: 559–570.
- Alonso M.J., Wolf S., Jørgensen R.B., Madsen H., Mathisen H.M. (2022): A methodology for the selection of pollutants for ensuring good indoor air quality using the de-trended cross-correlation function. *Building and Environment*, 209: 108668.
- Andi Pallawa N.B., Salengke, Pirman (2023): Effects of roasting process on the precursor components for Maillard reaction in fermented cocoa beans. In: Mubarak H., Ramli A.R., Muhpidah, Arifin A.S., Hardinasinta G., Rizal M., Rais A., Hatmus A.S.R., Rahmatika N.D., Anshari M.A., Hidayat S.H., Yolanda D.S., Ainani A.F., Kamaruddin I., Kasmira (eds): *Proceedings of the 1st Unhas International Conference on Agricultural Technology (UICAT 2021)*, Makassar, Indonesia, Oct 27–28, 2021: 020002.
- Anyidoho E.K., Teye E., Agbemaflle R. (2021): Differentiation of organic cocoa beans and conventional ones by using handheld NIR spectroscopy and multivariate classification techniques. *International Journal of Food Science*, 2021: 1844675.
- Balcázar-Zumaeta C.R., Fernández-Romero E., Santos Lopes A., Ferreira N.R., Chagas-Júnior G.C.A., Yoplac I., López-Trigoso H.A., Tuesta-Occ M.L., Maldonado-Ramírez I., Maicelo-Quintana J.L., Cayo-Colca I.S., Castro-Alayo E.M. (2024): Amino acid profile behavior during the fermentation of Criollo cocoa beans. *Food Chemistry*, X, 22: 101486.
- Batista A.S., de Freitas Oliveira T., de Oliveira Pereira I., Santos L.S. (2021): Identification of cocoa bean quality by near infrared spectroscopy and multivariate modeling. *Research, Society and Development*, 10: e64101522732.
- Brosnan T., Sun D.W. (2004): Improving quality inspection of food products by computer vision – A review. *Journal of Food Engineering*, 61: 3–16.
- Calvo A.M., Botina B.L., García M.C., Cardona W.A., Montenegro A.C., Criollo J. (2021): Dynamics of cocoa fermentation and its effect on quality. *Scientific Reports*, 11: 16746.
- Castillejos-Mijangos L.A., Meza-Márquez O.G., Osorio-Revilla G., Jiménez-Martínez C., Gallardo-Velázquez T. (2023): Identification of variety and prediction of chemical composition in cocoa beans (*Theobroma cacao* L.) by FT-MIR spectroscopy and chemometrics. *Foods*, 12: 4144.
- das Virgens I.A., Pires T.C., de Santana L.R.R., Soares S.E., Maciel L.F., Ferreira A.C.R., Biasoto A.C.T., Bispo E.S. (2021): Relationship between bioactive compounds and sensory properties of dark chocolate produced from Brazilian hybrid cocoa. *International Journal of Food Science & Technology*, 56: 1905–1917.
- Devi N., Cruz-Martins N., Malik T., Kumar V., Islam M.T., Verma R., Singh G.P., Kumar D. (2025): Phytochemical profiling and anti-bacterial activity of Red Delicious apple pomace with integrated hydrolysis to obtain fermentable sugars: A biorefinery approach. *Food Science & Nutrition*, 13: e70543.
- Fodor M., Matkovits A., Benes E.L., Jókai Z. (2024): The role of near-infrared spectroscopy in food quality assurance: A review of the past two decades. *Foods*, 13: 3501.
- Forte M., Currò S., Van de Walle D., Dewettinck K., Mirisola M., Fasolato L., Carletti P. (2023): Quality evaluation of fair-trade cocoa beans from different origins using portable near-infrared spectroscopy (NIRS). *Foods*, 12: 4.
- Hashimoto J.C., Lima J.C., Celeghini R.M.S., Nogueira A.B., Efraim P., Poppi R.J., Pallone J.A.L. (2018): Quality control of commercial cocoa beans (*Theobroma cacao* L.) by near-infrared spectroscopy. *Food Analytical Methods*, 11: 1510–1517.
- Hayati R., Zulfahrizal, Munawar A.A. (2021): Robust prediction performance of inner quality attributes in intact cocoa

- beans using near infrared spectroscopy and multivariate analysis. *Heliyon*, 7: e06286.
- Hernández-Hernández C., Fernández-Cabanás V.M., Rodríguez-Gutiérrez G., Fernández-Prior Á., Morales-Sillero A. (2022): Rapid screening of unground cocoa beans based on their content of bioactive compounds by NIR spectroscopy. *Food Control*, 131: 108347.
- Herrera-Rocha F., Cala M.P., León-Inga A.M., Aguirre Mejía J.L., Rodríguez-López C.M., Florez S.L., Chica M.J., Olarte H.H., Duitama J., González Barrios A.F., Fernández-Niño M. (2022): Lipidomic profiling of bioactive lipids during spontaneous fermentations of fine-flavor cocoa. *Food Chemistry*, 397: 133845.
- Jiang H., Zhou Y., Zhang C., Yuan W., Zhou H. (2023): Evaluation of dual-band near-infrared spectroscopy and chemometric analysis for rapid quantification of multi-quality parameters of soy sauce stewed meat. *Foods*, 12: 2882.
- Jiménez-Rodríguez D.J., García-Alamilla P., Márquez-Rocha F.J., Vázquez-Medina R., Carrera-Lanestosa A., González-Alejo F.A., Sánchez-Ramos C.A., Ruiz-Santiago F.L. (2024): Temperature effect of cocoa (*Theobroma cacao* L.) drying on energy consumption, bioactive composition and vibrational changes. *Processes*, 12: 2523.
- Kandpal L.M., Lohumi S., Kim M.S., Kang J.S., Cho B.K. (2016): Near-infrared hyperspectral imaging system coupled with multivariate methods to predict viability and vigor in muskmelon seeds. *Sensors and Actuators B: Chemical*, 229: 534–544.
- 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.
- La Mantia A., Ianni F., Schoubben A., Cespi M., Lisjak K., Guarnaccia D., Sardella R., Blasi P. (2023): Effect of cocoa roasting on chocolate polyphenols evolution. *Antioxidants*, 12: 469.
- Li J. (2016): Assessing spatial predictive models in the environmental sciences: Accuracy measures, data variation and variance explained. *Environmental Modelling & Software*, 80: 1–8.
- López-Calabozo R., Liberal Â., Fernandes Â., Revilla I., Ferreira I.C.F.R., Barros L., Vivar-Quintana A.M. (2024): Determination of carbohydrate composition in lentils using near-infrared spectroscopy. *Sensors*, 24: 4232.
- Mahesh S., Jayas D.S., Paliwal J., White N.D.G. (2015): Comparison of partial least squares regression (PLSR) and principal components regression (PCR) methods for protein and hardness predictions using near-infrared (NIR) hyperspectral images of bulk samples of Canadian wheat. *Food and Bioprocess Technology*, 8: 31–40.
- Maldonado Y.E., Figueroa J.G. (2023): Microwave-assisted extraction optimization and effect of drying temperature on catechins, procyanidins and theobromine in cocoa beans. *Molecules*, 28: 3755.
- Manley M. (2014): Near-infrared spectroscopy and hyperspectral imaging: Non-destructive analysis of biological materials. *Chemical Society Reviews*, 43: 8200–8214.
- Masithoh R.E., Pahlawan M.F.R., Arifani E.N., Amanah H.Z., Cho B.K. (2024): Determination of the betacyanin and betaxanthin contents of red beet (*Beta vulgaris*) powder using partial least square regression based on visible-near infrared spectra. *Trends in Sciences*, 21: 7639.
- Moreno-Trujillo T.R., Perez E., Verardo V., García-Villanova B., Guerra-Hernández E.J. (2024): Impact of technological processes on the formation of furosine, acrylamide and furan in traditional Venezuelan cocoa. *Foods*, 13: 829.
- Moussyoi Moundanga S., Petit J., Ndangui C.B., Scher J., Nzikou J.-M. (2024): Impact of cocoa variety on merchant quality and physicochemical characteristics of raw cocoa beans and roasted cocoa mass. *Discover Food*, 4: 115.
- Nadimi M., Paliwal J. (2024): Recent applications of near-infrared spectroscopy in food quality analysis. *Foods*, 13: 2633.
- Pahlawan M.F.R., Murti B.M.A., Masithoh R.E. (2022): The potency of Vis/NIR spectroscopy for classification of soybean based of colour. *IOP Conference Series: Earth and Environmental Science*, 1018: 012015.
- Peña-Correa R.E., Ataç Mogol B., Frygas C., Fogliano V. (2023): Fluidized-bed-roasted cocoa has different chemical characteristics than conventionally roasted cocoa. *Journal of Agricultural and Food Chemistry*, 71: 10203–10211.
- Peña-Correa R.E., Ataç Mogol B., Fogliano V. (2024): The impact of roasting on cocoa quality parameters. *Critical Reviews in Food Science and Nutrition*, 64: 4348–4361.
- Perez M., Lopez-Yerena A., Vallverdú-Queralt A. (2022): Traceability, authenticity and sustainability of cocoa and chocolate products: A challenge for the chocolate industry. *Critical Reviews in Food Science and Nutrition*, 62: 475–489.
- Ramón V., Hernández H.E., Polania P., Suárez J.C. (2023): Spatial distribution of cocoa quality: Relationship between physicochemical, functional and sensory attributes of clones from southern Colombia. *Agronomy*, 13: 15.
- Reggente M., Dillner A.M., Takahama S. (2019): Analysis of functional groups in atmospheric aerosols by infrared spectroscopy: Systematic intercomparison of calibration methods for US measurement network samples. *Atmospheric Measurement Techniques*, 12: 2287–2312.
- de Souza M.S.d.S., Borges L.L.R., de Barros Kobi H., Freitas V.V., Rigolon T.C.B., Olegário L.S., Gutiérrez E.A., Felix P.H.C., Vidigal M.C.T.R., Stringheta P.C. (2024): Roasted and unroasted cocoa nibs: Bioactive compounds analysis and application in cereal bars. *Foods*, 13: 3510.

<https://doi.org/10.17221/58/2026-CJFS>

- Teye E. (2022): Mini shortwave spectroscopic techniques and multivariate statistical analysis as a tool for testing intact cocoa beans at farmgate for quality control in Ghana. *Infrared Physics & Technology*, 122: 104092.
- Thukral A.K., Bhardwaj R., Kumar V., Sharma A. (2019): New indices regarding the dominance and diversity of communities, derived from sample variance and standard deviation. *Heliyon*, 5: e02606.
- van Rooyen J., Marini F., Orth S.H., Oyeyinka S.A., Simsek S., Manley M. (2023): Effect of wheat roasting conditions and wheat type on short-wave infrared (SWIR) spectral data of whole and milled wheat by ANOVA-simultaneous component analysis. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 303: 123160.
- Wu Q., Mousa M.A.A., Al-Qurashi A.D., Ibrahim O.H.M., Abo-Elyousr K.A.M., Rausch K., Abdel Aal A.M.K., Kamruzzaman M. (2023): Global calibration for non-targeted fraud detection in quinoa flour using portable hyperspectral imaging and chemometrics. *Current Research in Food Science*, 6: 100483.
- Xu Y., Kong T., Ma Y., Zhao Y., Chu L., Zheng M. (2025): Near-infrared spectroscopy: Application in ensuring food quality and safety. *Analytical Methods*, 17: 3381–3406.
- Zheng R., Jia Y., Ullagaddi C., Allen C., Rausch K., Singh V., Schnable J.C., Kamruzzaman M. (2024): Optimizing feature selection with gradient boosting machines in PLS regression for predicting moisture and protein in multi-country corn kernels via NIR spectroscopy. *Food Chemistry*, 456: 140062.

Received: April 6, 2026

Accepted: August 25, 2026

Published online: October 23, 2026