

Portable NIR Spectroscopy for soil property prediction and fertility assessment in coffee and cocoa systems

Marlen A. Grandez-Alberca^{a,*}, Julio Puscan-Rojas^a, Teodoro B. Silva-Melendez^a, Lily del Pilar Juarez-Contreras^a, Angel J. Medina-Medina^a, Jhon A. Zabaleta-Santisteban^a, Abner S. Rivera-Fernandez^a, Katerin M. Tuesta-Trauco^a, Mirella N. Coronel-Alcalde^a, Alcides Roman^b, Samuel Pizarro^a, Elgar Barboza^{a,c,*}

^a Instituto de Investigación para el Desarrollo Sustentable de Ceja de Selva (INDES-CES), Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas, Chachapoyas, 01001, Peru

^b Estación Experimental Agraria Amazonas, Dirección de Investigación y Desarrollo Tecnológico, Instituto Nacional de Innovación Agraria (INIA), Chachapoyas, 01001, Peru

^c Facultad de Ingeniería Ambiental, Biosistemas y de la Energía (FIABE), Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas, Chachapoyas, 01001, Peru

ARTICLE INFO

Keywords:

Precision agriculture
Machine learning
Near-infrared (NIR) spectroscopy
Proximal soil sensing
Agroforestry systems

ABSTRACT

The study evaluated the potential of portable Near-Infrared (NIR) spectroscopy to predict soil physicochemical properties in coffee and cocoa production systems in Amazonas, Peru. A total of 126 topsoil samples were initially collected using a stratified sampling design. A total of 125 surface soil samples were analyzed using a stratified sampling design. Soil properties were determined using standard laboratory reference methods, and spectral data were acquired with a NeoSpectra portable spectrometer operating in the 1350–2550 nm (FT-NIR) range. Predictive models were developed using Random Forest (RF), Support Vector Machine (SVM), Neural Networks (NN), and Partial Least Squares Regression (PLSR), and their performance was evaluated using nested five-fold cross-validation (CV).

The predictive performance of portable NIR spectroscopy varied according to the soil property, production system, preprocessing technique, and machine learning algorithm. Coffee soils showed higher predictive performance than cocoa soils. The best results were obtained for clay ($R^2_{CV} = 0.775$; $RPD_{CV} = 2.276$) using RF with first-derivative preprocessing, followed by phosphorus ($R^2_{CV} = 0.772$; $RPD_{CV} = 2.037$) using PLSR with band-depth preprocessing. Soil pH and silt also showed satisfactory predictive performance in coffee soils. In cocoa soils, potassium exhibited the highest predictive capability ($R^2_{CV} = 0.454$; $RPD_{CV} = 1.378$) using SVM with second-derivative preprocessing, although the overall model performance was lower than that observed for coffee soils. Portable NIR spectroscopy provided potentially useful estimates for selected soil properties, particularly clay, phosphorus, pH, and silt in coffee soils. However, its predictive performance was strongly dependent on the soil property, preprocessing strategy, machine learning algorithm, and production system.

1. Introduction

Soil is an essential component of agricultural production, as its physical, chemical, and biological properties directly influence crop growth, productivity, and the sustainability of agricultural systems (García et al., 2012). In economically important perennial crops such as

coffee (*Coffea arabica* L.) and cocoa (*Theobroma cacao* L.), soil characterization is essential for guiding agronomic management and improving crop productivity (Pérez-Burillo et al., 2022). Variables such as soil texture, Soil Organic Matter (SOM), pH, and nutrient availability allow the identification of soil constraints that may limit plant development and physiological performance (Vogel et al., 2019).

* Corresponding authors.

E-mail addresses: marlen.grandez.epg@untrm.edu.pe (M.A. Grandez-Alberca), 7278447092@untrm.edu.pe (J. Puscan-Rojas), 7680096361@untrm.edu.pe (T.B. Silva-Melendez), lily.juarez@untrm.edu.pe (L.P. Juarez-Contreras), angel.medina@untrm.edu.pe (A.J. Medina-Medina), jhon.zabaleta.epg@untrm.edu.pe (J.A. Zabaleta-Santisteban), abner.rivera.epg@untrm.edu.pe (A.S. Rivera-Fernandez), katerin.tuesta.epg@untrm.edu.pe (K.M. Tuesta-Trauco), 7246139592@untrm.edu.pe (M.N. Coronel-Alcalde), aroman@inia.gob.pe (A. Roman), samuel.pizarro@untrm.edu.pe (S. Pizarro), elgar.barboza@untrm.edu.pe (E. Barboza).

<https://doi.org/10.1016/j.soisec.2026.100250>

Received 18 June 2026; Received in revised form 11 August 2026; Accepted 1 September 2026

Available online 1 September 2026

2667-0062/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

However, conventional soil analysis methods rely on specialized laboratory procedures, which involve higher costs, long processing times, and limited capacity for frequent monitoring. These limitations have driven the search for faster, more efficient, and reliable methodologies for the assessment of agricultural soils (Partida et al., 2025). In this context, Near-Infrared (NIR) spectroscopy emerges as a promising alternative for soil characterization due to its ability to rapidly and non-destructively generate information on various soil properties.

NIR spectroscopy is an analytical technique based on the interaction of electromagnetic radiation with molecular bonds present in materials. Through this interaction, it is possible to obtain spectral signatures associated with the composition and structure of samples, enabling the estimation of various physicochemical properties with minimal sample preparation and without altering the sample (Park et al., 2024). In complex matrices such as soil, the spectral response is mainly influenced by SOM, clay minerals, and water molecules that largely control absorption and scattering in the NIR range.

This information can be used to estimate soil properties through calibration models; therefore, NIR spectroscopy has become an important tool in soil science (Thingujam et al., 2025). In addition, it allows the analysis of a large number of samples in short periods of time. This capability represents a significant advantage over conventional laboratory methods, particularly in studies that require high spatial coverage and frequent soil monitoring (Wang et al., 2022). Several studies have demonstrated that NIR spectroscopy can predict soil properties with varying levels of accuracy. Among the most frequently reported properties are SOM, Total Carbon (TC), soil texture fractions, and certain nutrients, due to their direct or indirect relationship with spectrally active compounds. Consequently, its use has increased in the assessment of agricultural soils as a complementary alternative to traditional analytical methods (Prananto et al., 2020).

Recent advances in portable and miniaturized NIR spectrometers have facilitated the widespread application of proximal soil sensing for precision agriculture, enabling rapid, non-destructive, and cost-effective soil characterization. Recent advances in miniaturized spectrometers have facilitated the use of proximal soil sensing for precision agriculture, supporting real-time decision-making and reducing the dependence on conventional laboratory analyses. Combined with chemometric and machine learning approaches, portable NIR spectroscopy has demonstrated promising performance for predicting soil texture, pH, soil organic carbon, total nitrogen, cation exchange capacity, and exchangeable nutrients across different agricultural environments (Mitu et al., 2024; Partida et al., 2025).

Among portable NIR devices, the NeoSpectra spectrometer has gained increasing attention because of its compact design, low cost, and ability to acquire spectra within the 1350–2550 nm FT-NIR range. Recent studies have shown that NeoSpectra-based systems can accurately estimate several soil physicochemical properties, including soil organic carbon, total nitrogen, pH, clay, silt, sand, cation exchange capacity, and exchangeable nutrients using machine learning algorithms such as Partial Least Squares Regression (PLSR), Cubist, Random Forest (RF), and Memory-Based Learning (MBL) (Mitu et al., 2024; Partida et al., 2025). Furthermore, the development of open spectral libraries based on NeoSpectra has expanded the applicability of portable spectroscopy for large-scale soil monitoring and digital soil assessment (Sanderman et al., 2024).

Most studies have been conducted using temperate soils or large regional soil spectral libraries, whereas the application of portable NeoSpectra spectroscopy in tropical agroecosystems remains limited. In particular, little information is available regarding its performance in coffee and cocoa production systems, where marked variations in soil texture, mineral composition, organic matter, and management practices can substantially influence spectral responses and prediction accuracy. Although portable NIR spectroscopy has demonstrated considerable potential for soil characterization, its application using the NeoSpectra spectrometer in coffee and cocoa production systems of the

Peruvian Amazon remains largely unexplored. Therefore, this study contributes to expanding the application of portable NIR spectroscopy combined with multiple machine learning algorithms for rapid soil assessment under highly heterogeneous tropical conditions, thereby supporting precision soil management strategies.

The integration of NIR spectroscopy with reference laboratory analyses is particularly relevant for coffee and cocoa production systems, where soil properties play a decisive role in crop development, productivity, and bean quality. These perennial crops are of high economic and social importance in tropical regions, particularly in Latin America, where they represent a major source of income for small- and medium-scale producers (International Trade Centre, 2021). Factors such as soil fertility, nutrient availability, texture, and soil structure directly influence plant growth, water and nutrient uptake, and crop responses to environmental and management conditions (Haro-Altamirano et al., 2025). In tropical regions, soils often exhibit high spatial variability associated with geological, topographic, climatic, and agronomic management factors (Márquez et al., 2020), which can affect plantation productivity and the accuracy of decisions related to fertilization, soil management, and soil fertility conservation. Consequently, the use of spectroscopic technologies combined with precision agriculture tools provides an effective alternative for improving soil monitoring and supporting more efficient management of coffee and cocoa production systems (Recena et al., 2019; Javaid et al., 2022). However, one of the main challenges in applying NIR spectroscopy is the development of reliable and robust calibration models. Their accuracy depends on representative datasets, reference laboratory analyses, and an adequate representation of soil physicochemical variability (Schmidinger et al., 2024). In addition, the quality of the spectral signal and the pre-processing techniques applied strongly influence model performance; therefore, rigorous calibration and validation are essential (Nocita et al., 2015). By integrating laboratory reference data with NIR spectroscopy, it is possible to develop robust predictive models that enable the rapid estimation of multiple soil properties while maintaining analytical reliability (FAO, 2024; Minasny et al., 2024). This capability can strengthen soil diagnosis and support agronomic decision-making in agricultural systems characterized by high spatial variability (Reza et al., 2025).

Despite this potential, the effective application of NIR spectroscopy requires rigorous validation against reference methods to ensure accurate and consistent estimates under different soil conditions (FAO, 2024). In this process, indicators such as the coefficient of determination (R^2), Root Mean Square Error (RMSE), and the ratio of Performance to Standard Deviation (RPD) are used to assess the accuracy, robustness, and reliability of predictive models (Bellon-Maurel et al., 2010). Building upon these knowledge gaps, this study investigated the performance of a portable NeoSpectra NIR spectrometer combined with machine learning algorithms for characterizing and predicting soil physicochemical properties in coffee (*Coffea arabica* L.) and cocoa (*Theobroma cacao* L.) production systems in the Amazonas region of Peru. Laboratory reference analyses were used to calibrate and validate the predictive models. Specifically, this study aimed to: (i) characterize the physicochemical properties of soils from both production systems; (ii) examine the relationships between NIR spectral signatures and laboratory-determined soil properties; and (iii) evaluate and compare the predictive performance of different machine learning models for estimating key soil variables.

2. Materials and methods

2.1. Study area

The Amazonas region is located in the northeastern part of Peru, between the Andean mountain range and the Amazon lowlands (Fig. 1). It covers an area of 39,249.13 km², equivalent to 3.5% of the Peruvian territory, and has a total population of 379,384 inhabitants, with an

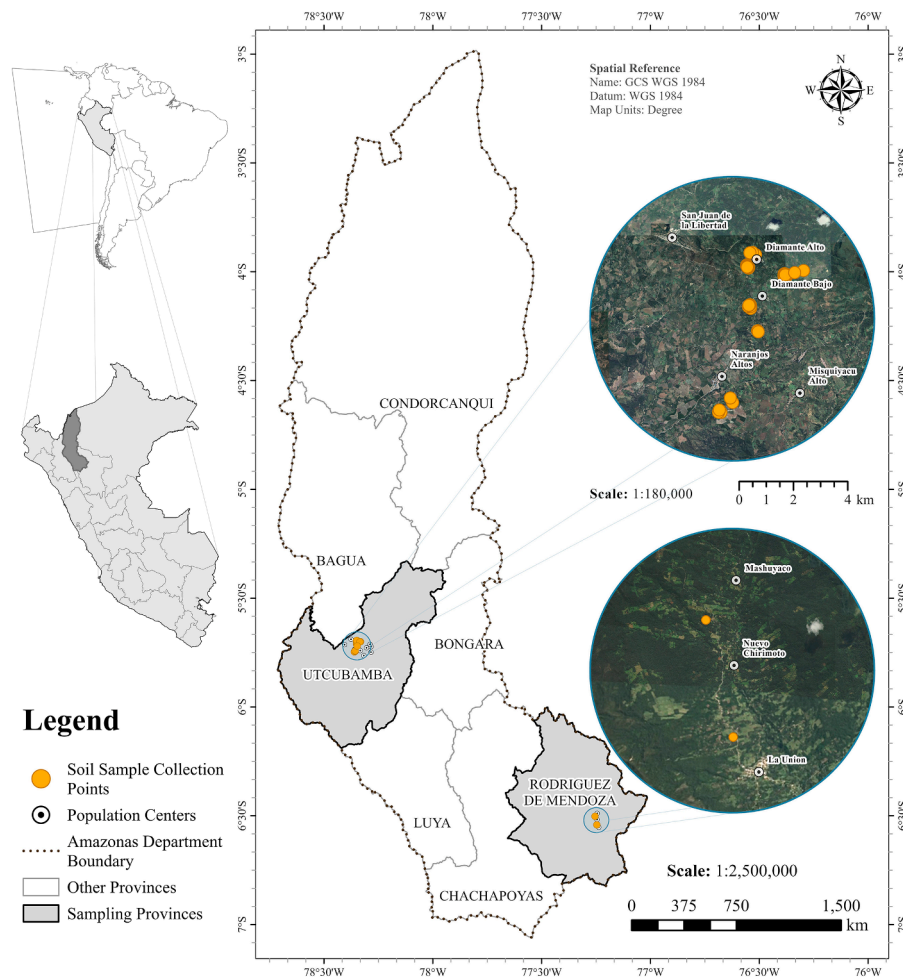


Fig. 1. Study area and soil sampling points. (a) Sampling points in Amazonas; (b) boundary of the Amazonas region; (c) location of Amazonas in northeastern Peru.

approximate density of 9.6 inhabitants km^{-2} (MIDAGRI, 2021; INEI, 2018). Due to its altitudinal and ecological transition, it presents high biophysical and sociocultural diversity, as well as significant agricultural activity. Its main crops include rice, coffee, cocoa, potato, cassava, and banana, along with extensive livestock production. It also records an illiteracy rate of 8.3% among agricultural producers (IIAP, 2010).

2.2. Methodological process

The study methodology is shown in Fig. 2. It was developed sequentially, beginning with the delimitation of the study area, corresponding to coffee (*Coffea arabica* L.) and cocoa (*Theobroma cacao* L.) production systems located within the study region. Subsequently, soil sampling was carried out using a stratified approach, considering altitudinal gradients and agronomic management conditions to represent the soil variability of both crops.

Once collected, the samples were subjected to a preparation process that included oven-drying, homogenization, and sieving, in order to ensure uniformity prior to analysis. From this stage, the study was divided into two complementary processes: laboratory analysis, aimed at determining soil physicochemical properties, and spectral measurement, performed using NIR spectroscopy with a portable spectrometer.

The resulting data were integrated into a spectral preprocessing phase, applying techniques designed to reduce noise and improve spectral signal quality. Subsequently, predictive models based on machine learning algorithms were developed to estimate soil properties from spectral information.

Finally, the models were subjected to a validation and evaluation

stage using statistical performance indicators, followed by a comparison of results between coffee and cocoa systems, identifying variables with the highest predictive potential and the applicability of NIR spectroscopy for rapid soil characterization in agricultural systems.

2.3. Sampling

For spectral analysis and predictive modeling, a total of 125 soil samples (69 from coffee and 56 from cacao) were collected, following standardized protocols for agricultural soil sampling (MIDAGRI, 2024a). To capture the spatial variability of the study area, a stratified sampling design was employed, considering altitudinal gradients and different agronomic management conditions. Sampling points were georeferenced using a South Galaxy G2 GNSS device to record the spatial location of each sample and ensure the reproducibility of the study. At each sampling point, samples were collected using a composite sampling scheme, integrating several subsamples collected within a previously defined area. This procedure helped reduce local variability and improve sample representativeness, enabling a more consistent characterization of the evaluated soil properties (FAO, 2024). The samples were collected from the surface soil layer (0–30 cm), considered the most relevant depth for assessing fertility and nutrient availability in perennial crops such as coffee and cocoa (Weil and Brady, 2017).

2.4. Laboratory analysis

After collection, the soil samples were oven-dried and homogenized. Conventional laboratory analyses were performed on subsamples

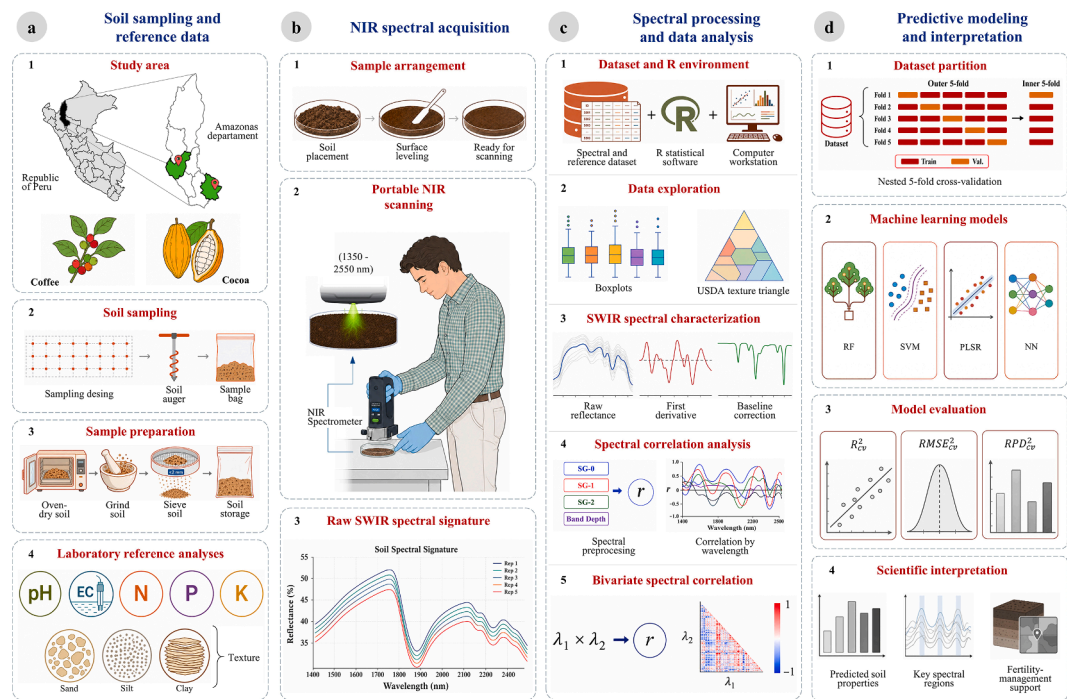


Fig. 2. Methodological workflow of the study, including soil sampling, laboratory analysis, NIR spectral acquisition, spectral preprocessing, nested five-fold cross-validation, and predictive modeling in coffee and cocoa systems.

following established reference procedures to determine the main physicochemical properties evaluated in this study, including soil pH, electrical conductivity (EC), total nitrogen (N), available phosphorus (P), potassium (K), and particle-size distribution expressed as sand, silt, and clay fraction (Nocita et al., 2015). Separately, an aliquot intended for NIR spectral acquisition was sieved through a 106- μ m mesh (0.106 mm) to obtain a fine and homogeneous sample matrix (MIDAGRI, 2024b). During this preparation, coarse materials and root residues that could interfere with spectral measurements were removed. This conditioning improved sample homogeneity and reduced particle-size-related variability in the spectral signal (Stenberg et al., 2010). The prepared NIR aliquots were subsequently stored in properly labeled airtight containers until spectral acquisition.

2.5. Soil spectral data acquisition

NIR spectral measurements were available for 125 soil samples. These samples constituted the final dataset used for spectral analysis and predictive modeling. For spectral acquisition, the previously prepared samples were evenly placed in 10 cm Petri dishes to maintain consistent conditions during spectral measurement (Soriano-Disla et al., 2014).

Measurements were carried out using a portable ProxiScout by NeoSpectra® spectrometer based on NIR technology, operating in a spectral range of 1350–2550 nm. The device has a spectral resolution of 16 nm (FWHM at 1500 nm) and an effective measurement area of

approximately 10 mm in diameter (Table 1). These characteristics allowed the spectral response of the samples to be recorded under controlled conditions (Malley et al., 1999).

For each sample, five consecutive scans were performed at two-second intervals. The resulting spectra were then averaged to reduce instrumental noise and improve signal stability. This procedure enabled the generation of a consistent spectral database, which was used for correlation analyses and predictive modeling of soil physicochemical properties (Mouazen et al., 2010).

2.6. NIR analysis

The obtained spectra were subjected to preprocessing techniques aimed at improving signal quality and reducing variations not associated with soil composition. Smoothing using the Savitzky–Golay (SG) filter, first- and second-order derivatives, and Band Depth transformation were applied. These procedures helped reduce instrumental noise, correct baseline variations, and enhance spectral regions potentially related to soil physicochemical properties (Rinnan, 2014; Savitzky and Golay, 1964). Once preprocessed, the spectra were organized into spectral data matrices for integration with soil properties determined using reference laboratory methods. This stage enabled the structuring of inputs used in subsequent spectral correlation analyses, bispectral correlation, and predictive modeling.

Prior to model calibration, spectral preprocessing was performed to reduce noise and enhance absorption features. Savitzky–Golay (SG) smoothing was applied using a window size of 11 spectral bands and a third-order polynomial. Based on the same configuration, the first (1D) and second (2D) derivatives were calculated. In addition, the Band Depth (BD) transformation was obtained by continuum removal using the continuumRemoval() function from the prosprctr package (type = "R"), and band depth was subsequently calculated as $BD = 1 - CR$, where CR represents the continuum-removed spectrum. All preprocessing procedures were performed in R prior to model calibration.

Table 1
Technical specifications of the NIR spectroradiometer used for soil spectral measurements.

Technical characteristic	Specification
Equipment	ProxiScout by NeoSpectra®
Instrument type	Portable NIR spectrometer
Spectral range	1350-2550 nm
Technology	Portable NIR system for soil sample spectral measurement
Spectral resolution	16 nm (FWHM a 1500 nm)

2.7. Statistical analysis

Statistical processing was carried out to integrate the physicochemical properties determined in the laboratory with the spectral information obtained via NIR. Analyses were performed in R v.4.6.0 using RStudio, applying specific routines for descriptive analysis, predictive modeling, and model performance evaluation (R Core Team, 2026).

For predictive model development and evaluation, nested five-fold cross-validation was applied independently to the coffee and cocoa datasets. In the outer loop, the samples were divided into five folds, with four folds used for model training and the remaining fold used for validation. This procedure was repeated until each fold had been used once for validation. Within each outer training subset, an additional five-fold cross-validation was used for hyperparameter optimization. The 40 most informative wavelengths were selected using only the samples included in each outer training subset. Model performance was summarized as the mean $\pm$ standard deviation across the five outer validation folds.

The predictive models were developed using RF, SVM, NN, and PLSR. Before model calibration, the 40 most informative spectral wavelengths were selected according to the absolute Pearson correlation with each response variable. Hyperparameter optimization was performed within the five-fold cross-validation framework using predefined tuning grids. For RF, the mtry parameter was evaluated using values of 2, 5, 10, and 20 (limited by the number of selected predictors). For SVM, combinations of the cost parameter ($C = 1, 2, 4$, and 8) and the radial basis function kernel parameter ($\sigma = 0.001, 0.005, 0.01, 0.03$, and 0.05) were evaluated. For NN, hidden-layer sizes of 1, 3, and 5 neurons and weight decay values of 0.0001, 0.001, 0.01, and 0.1 were tested. For PLSR, the number of latent variables (ncomp) was optimized by evaluating values from 1 to the maximum number permitted by the training dataset. Predictor variables were centered and scaled before fitting the SVM, NN, and PLSR models (Belgiu and Drăgu, 2016; Breiman, 2001).

Model performance was evaluated using the R^2 , RMSE, and RPD. For each soil property, the best-performing model was selected primarily according to the highest mean cross-validation R^2 , followed by the lowest mean cross-validation RMSE and the highest mean cross-validation RPD. Results are presented as mean $\pm$ standard deviation across the five cross-validation folds, providing a robust assessment of predictive accuracy and model stability across different validation subsets (Stenberg et al., 2010).

2.8. Fertility management interpretation framework for coffee and cocoa systems

The fertility-management interpretation framework integrated laboratory-determined soil properties with the predictive reliability of crop-specific NIR models to identify indicators suitable for rapid soil-fertility assessment and management-zone delineation within a soil-security perspective (McBratney et al., 2014).

The evaluated variables were interpreted according to four agro-nomic domains: soil reaction, represented by pH because acidity and alkalinity influence nutrient availability, root activity, aluminum toxicity, and the need for corrective practices; soluble-salt conditions, represented by EC; macronutrient status, represented by N, P, and K; and the physical soil matrix, represented by sand, silt, and clay because particle-size distribution influences water retention, nutrient retention, cation-exchange processes, and spectral scattering (Corwin and Lesch, 2005; Neina, 2019; Saxton and Rawls, 2006; Stenberg et al., 2010).

For coffee, the interpretation prioritized soil-acidity diagnosis, available-P management, and texture-based zoning because laboratory analyses indicated strongly acidic conditions, substantial variability in available P, and marked differences in particle-size fractions. For cocoa, the interpretation prioritized soil-reaction assessment, potassium-balance evaluation, and clay-associated nutrient retention because these soils showed a moderately alkaline reaction, very low EC, higher K

concentrations, and greater clay content. EC was retained as a complementary indicator to confirm the absence of soluble-salt constraints. The agronomic usefulness of each indicator was determined by combining the observed soil condition with the mean cross-validated performance of the corresponding NIR model. Use classes were assigned according to mean RPD_{CV}: indicators with RPD_{CV} ≥ 2.0 were classified as high-use indicators, those with $1.4 \leq \text{RPD}_{\text{CV}} < 2.0$ were classified as moderate-use indicators, and those with RPD_{CV} < 1.4 were classified as low-use indicators requiring laboratory or foliar confirmation before fertility-management decisions. The variability among folds and the corresponding R^2_{CV} were additionally considered when interpreting model stability (Bellon-Maurel et al., 2010; Stenberg et al., 2010). Accordingly, the framework was used to derive soil-analysis and fertility-management implications rather than direct fertilizer-dose prescriptions, because operational recommendations additionally require information on crop age, planting density, target yield, nutrient extraction, soil organic matter, bulk density, exchangeable bases, cation-exchange capacity, foliar status, fertilizer-use efficiency, and field-response validation.

3. Results

3.1. Physical and chemical characteristics of the soil

The soil physicochemical properties revealed marked edaphic contrasts between the coffee and cocoa systems (Fig. 3). Coffee soils showed a mean pH of 4.80 ± 0.81 , indicating predominantly acidic conditions, whereas cocoa soils showed a mean pH of 8.11 ± 0.47 , corresponding to a moderately alkaline reaction.

Electrical conductivity showed a marked contrast between production systems. Coffee soils presented a mean EC of $1.59 \pm 0.25 \text{ dS m}^{-1}$, ranging from 0.82 to 2.13 dS m^{-1} . In contrast, cocoa soils presented a mean EC of $0.144 \pm 0.030 \text{ dS m}^{-1}$, with values ranging from 0.056 to 0.217 dS m^{-1} . These values indicate low soluble-salt concentrations and no apparent salinity constraints under the evaluated conditions.

Regarding macronutrients, coffee and cocoa soils showed mean N concentrations of $2.00 \pm 0.59\%$ and $1.73 \pm 0.47\%$, respectively. Available P was higher in coffee soils ($22.28 \pm 12.46 \text{ ppm}$) than in cocoa soils ($14.39 \pm 6.41 \text{ ppm}$), whereas K was higher in cocoa soils ($294.80 \pm 69.18 \text{ ppm}$) than in coffee soils ($189.15 \pm 84.34 \text{ ppm}$). These differences may be associated with soil reaction, fertilization history, mineral composition, and nutrient-retention processes.

Soil texture also differed between production systems. Coffee soils contained, on average, 39.93% sand, 26.89% silt, and 33.18% clay. Cocoa soils showed lower sand content (30.48%) and a greater proportion of clay (45.82%), accompanied by 23.70% silt. These differences indicate a finer soil matrix in cocoa systems, with implications for water retention, nutrient retention, cation-exchange processes, and spectral response.

3.2. Soil texture class distribution

The USDA soil texture triangles show the distribution of soil samples in the coffee and cocoa systems (Fig. 4). In coffee, samples were mainly concentrated in the loam (Lo), clay loam (CLLo), and sandy clay loam (SaCLLo) classes, with a secondary presence of sandy loam (SaLo). This distribution indicates a predominance of intermediate textures, associated with a relative balance between water retention and drainage.

In cocoa, the samples showed a greater concentration in fine-textured classes, mainly clay (Cl), clay loam (CLLo), and silty clay loam (SiCLLo). This pattern indicates a higher clay content compared to coffee soils, suggesting potentially greater water and nutrient retention and lower permeability.

In both systems, the sand (Sa) and silt (Si) textural classes were poorly represented, confirming the predominance of intermediate to fine-textured soils. Overall, these results reflect a clear textural

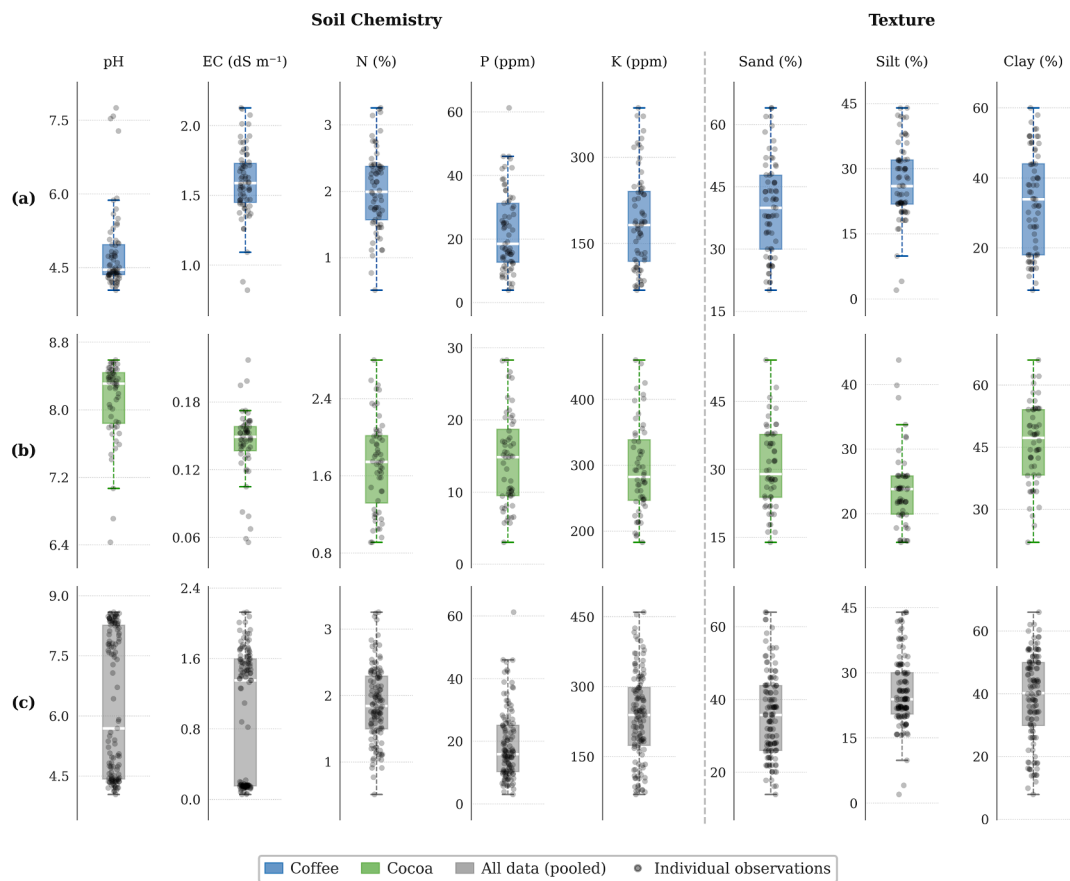


Fig. 3. Boxplot comparing the physicochemical properties of soil in coffee and cocoa systems.

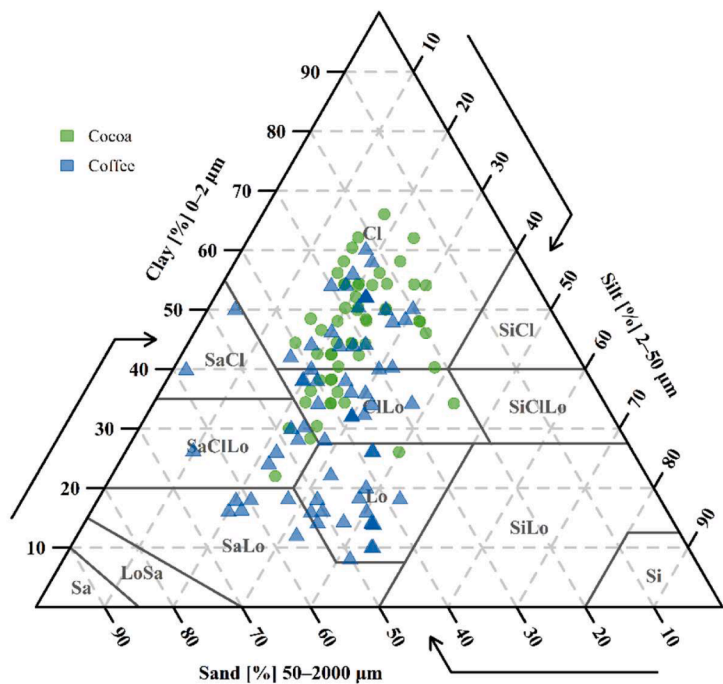


Fig. 4. Distribution of soil samples in the USDA textural triangle for coffee and cocoa agroforestry systems. Points represent individual samples classified according to their percentage of sand, silt, and clay content.

differentiation between cropping systems. Coffee soils showed a higher proportion of loamy classes, whereas cocoa soils tended toward more clay-rich textures.

3.3. Spectral correlation in the NIR range

The spectral signatures in the NIR region showed differences between soils associated with cocoa and coffee, depending on the type of transformation applied (Fig. 5).

The spectral response was evaluated using raw spectral reflectance, first-derivative spectra, and baseline-corrected spectra within the NIR range.

In cocoa soils (Fig. 5a, 5b, and 5c), the spectral curves exhibited lower dispersion among samples. The spectral reflectance showed moderate values, with absorption features observed around 1400, 1900, and 2200 nm. These regions are mainly associated with water, OH groups, and clay minerals. The smaller amplitude between curves suggests a relatively homogeneous spectral response.

In coffee soils (Fig. 5d, 5e, and 5f), the spectral curves showed greater dispersion among samples, particularly in spectral reflectance and derivative transformations. This variability was more evident between 1800 and 2300 nm, where changes associated with absorption features related to water and clay minerals were observed. The greater amplitude of the curve set suggests higher heterogeneity in the physicochemical properties of coffee soils.

The first-derivative transformation helped to highlight spectral differences that were not fully visible in the raw reflectance spectra. In coffee soils, this transformation revealed more pronounced changes in spectral slope, particularly around 1900 and 2200 nm. In cocoa soils, the spectral responses were more compact, indicating lower variability among samples.

3.4. Spectral correlation between soil properties and wavelengths in coffee and cocoa systems

The correlations between soil physicochemical properties and

wavelengths showed variable responses depending on the crop, the evaluated property, and the applied spectral preprocessing (Fig. 6). Overall, the curves revealed differences between coffee and cocoa soils, as well as among SG-0, SG-1, SG-2, and Band Depth transformations. This indicates that the spectral sensitivity of soil properties was not homogeneous across the NIR range.

In the SG-0 preprocessing (Fig. 6a), the curves showed smoother and more stable patterns in both crops. In cocoa, correlations remained mostly close to zero, although some textural variables showed gradual changes across the spectrum. In coffee, a greater separation among variables was observed, particularly for sand, silt, and clay. This suggests that soil texture fractions exhibited a more differentiated spectral response under this preprocessing approach.

The SG-1 (Fig. 6b) and SG-2 (Fig. 6c) derivative transformations increased the variability of the curves. This effect was more evident in coffee soils, where correlations showed pronounced oscillations across the spectrum. Although these transformations helped highlight local changes associated with specific absorption regions, they also increased signal irregularity, particularly in SG-2. In cocoa soils, the curves were less extreme than in coffee, but still exhibited variability among soil properties.

The Band Depth transformation (Fig. 6d) allowed the identification of more structured spectral patterns in some variables. In cocoa, well-defined changes were observed in properties such as pH, P, K, and soil texture fractions. In coffee, correlations showed greater amplitude, particularly for textural variables and nutrients. This suggests that Band Depth can highlight sensitive spectral regions, although its response depends on both the crop type and the analyzed property.

The results reveal spectral differentiation between both systems, associated with variations in pH, texture, and nutrients. The identified NIR regions, related to OH groups and clay minerals, support the prediction of phosphorus, silt, clay, and pH. These differences may be

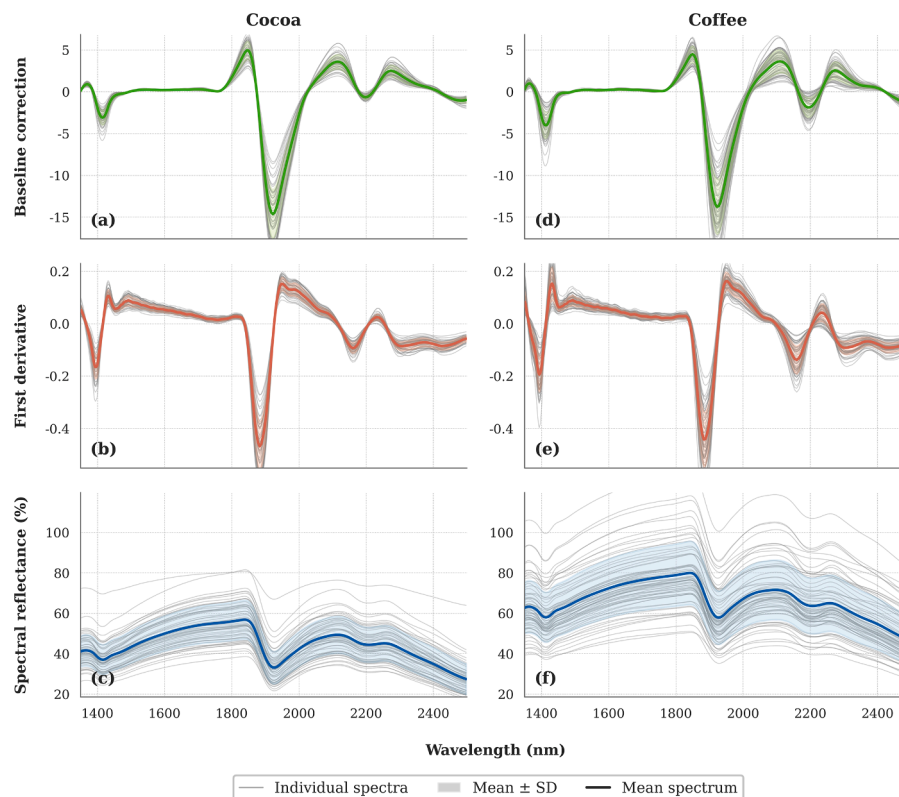


Fig. 5. NIR spectral signatures of agricultural soils associated with cocoa and coffee production systems. Panels (a–c) show baseline-corrected spectra, first-derivative spectra, and raw spectral reflectance for cocoa soils, respectively, while panels (d–f) show the corresponding transformations for coffee soils. Colored lines represent the mean spectra, while gray lines correspond to individual samples.

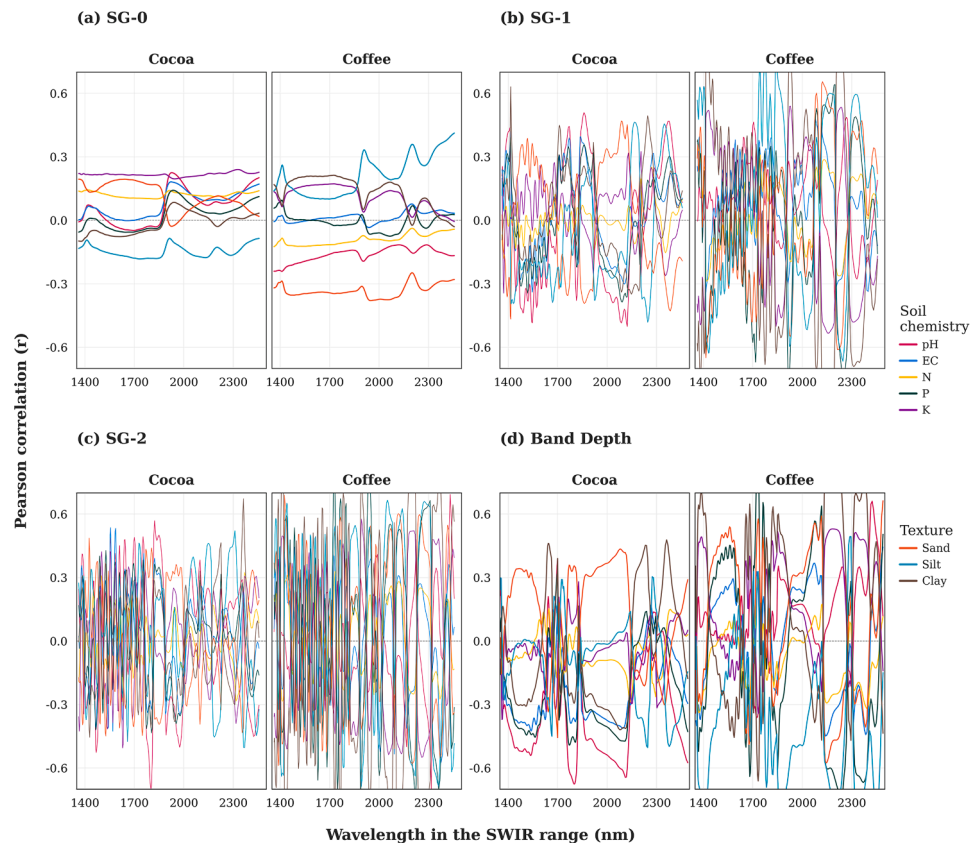


Fig. 6. Spectral correlation between soil physicochemical properties and wavelengths in cocoa and coffee samples. The panels show results obtained using SG-0 (a), SG-1 (b), SG-2 (c), and Band Depth (d).

linked to soil physicochemical variability, mineral composition, moisture, and organic matter content. Therefore, the selection of spectral preprocessing is crucial to improve the interpretation of correlations and to guide the development of predictive models.

3.5. Bispectral correlation between NIR reflectance and soil properties

Pearson correlation heatmaps revealed associations between combinations of NIR wavelengths and soil physicochemical properties in coffee and cocoa systems (Fig. 7). Red tones represented positive correlations, while blue tones indicated negative correlations. This representation allowed the identification of spectral regions with greater sensitivity to specific soil properties.

In both crops, the regions near 1400 and 1900 nm showed relevant correlation patterns, mainly associated with water absorption bands. Likewise, the intervals between 2100 and 2300 nm exhibited responses linked to organic matter functional groups and clay minerals. These spectral regions concentrated a significant portion of the observed variability, suggesting their usefulness for constructing bispectral indices aimed at soil characterization.

In coffee soils, P showed broad positive correlations within the bispectral space, indicating favorable potential for indirect estimation. N also exhibited well-defined patterns, mainly in regions near 2100 and 2300 nm, associated with nitrogen-containing organic compounds. In contrast, K and clay showed predominantly negative correlations in some spectral regions, possibly related to differences in mineral composition and cation exchange processes. Silt displayed relatively homogeneous positive correlations, while sand showed a weaker spectral response.

In cocoa soils, the patterns were partially similar, although with variations in correlation intensity. P exhibited broad positive correlation patterns across some regions of the bispectral space; however, these

associations did not translate into robust cross-validated predictive performance ($R^2_{CV} = 0.248 \pm 0.192$; $RPD_{CV} = 1.181 \pm 0.163$). N showed more localized spectral responses in regions associated with organic matter, but its predictive performance was also low ($R^2_{CV} = 0.117 \pm 0.196$; $RPD_{CV} = 1.014 \pm 0.015$). K, in contrast, exhibited weaker correlations than in coffee soils, suggesting lower spectral sensitivity under this production system. Regarding soil texture, sand showed predominantly negative responses, while clay displayed patterns related to spectral regions sensitive to silicates and hydroxyl groups.

Overall, the results indicate that wavelength combinations derived from bispectral analysis allow the identification of sensitive regions for both chemical and textural soil properties. This approach complements conventional spectral analysis and provides useful information for guiding the selection of spectral bands in the predictive modeling of soil properties in coffee and cocoa systems.

3.6. Predictive performance of spectral models in coffee and cocoa systems

The predictive performance of the spectral models varied according to soil property and production system (Table 2). Overall, coffee soils showed better predictive performance than cocoa soils, particularly for texture-related properties and phosphorus (P).

In coffee soils, the strongest results were obtained for clay ($R^2 = 0.775 \pm 0.096$; $RPD = 2.276 \pm 0.498$), P ($R^2 = 0.772 \pm 0.108$; $RPD = 2.037 \pm 0.417$), pH ($R^2 = 0.665 \pm 0.221$; $RPD = 2.085 \pm 0.921$), and silt ($R^2 = 0.679 \pm 0.154$; $RPD = 1.822 \pm 0.450$). Clay was predicted using RF with first-derivative preprocessing, whereas P and silt were modeled using PLSR with band-depth preprocessing, and pH using SVM with second-derivative preprocessing. Sand showed moderate performance ($R^2 = 0.571 \pm 0.224$; $RPD = 1.630 \pm 0.596$). In contrast, EC, N,

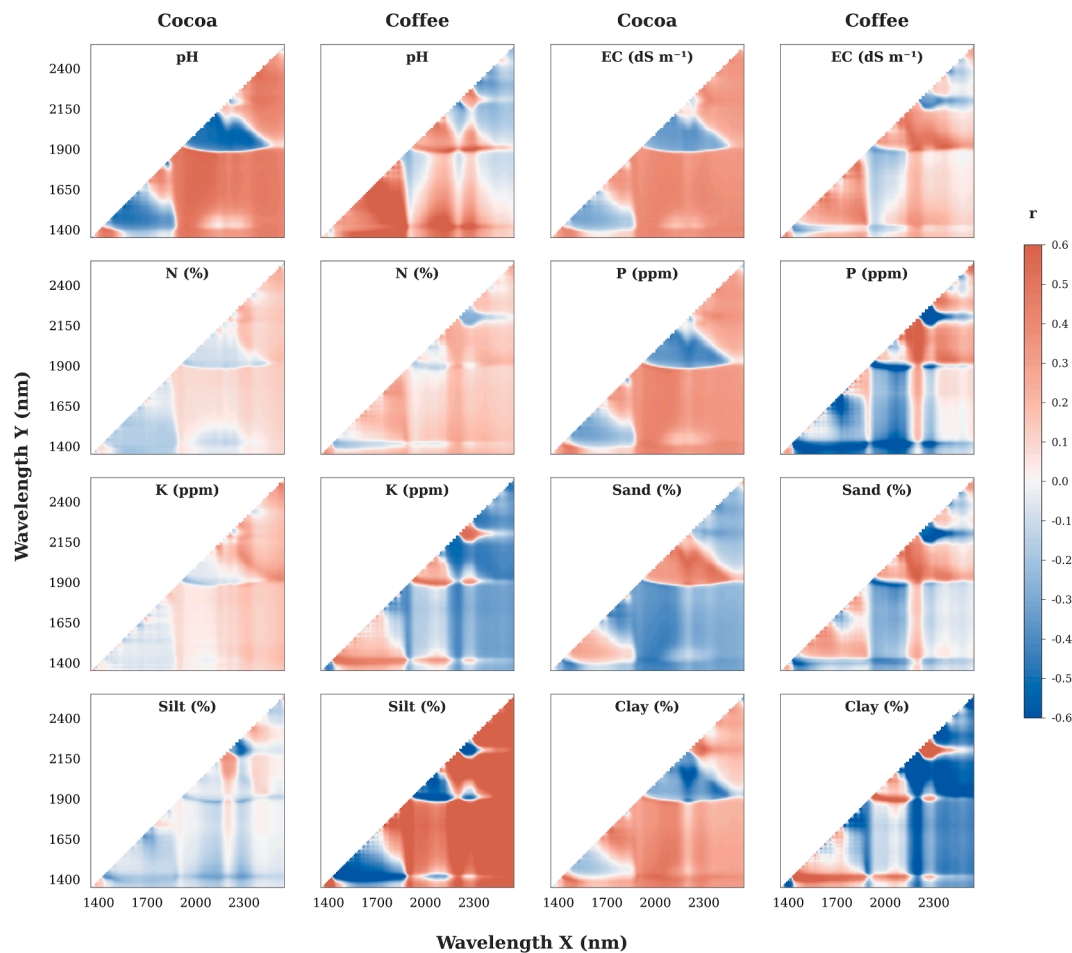


Fig. 7. Bispectral correlation heatmap between NIR wavelengths and soil physicochemical properties in coffee and cocoa systems.

Table 2

Nested five-fold cross-validated performance of the selected spectral models for predicting soil physicochemical properties in coffee and cocoa production systems.

Crop	Soil property	Preprocessing	Model	R^2_{train}	RMSE _{train}	RPD _{train}	R^2_{cv}	RMSE _{cv}	RPD _{cv}
Coffee	pH	2D	SVM	0.953 ± 0.022	0.183 ± 0.057	4.759 ± 1.367	0.665 ± 0.221	0.404 ± 0.122	2.085 ± 0.921
Coffee	EC (dS m ⁻¹)	SG	NN	0.416 ± 0.095	0.190 ± 0.018	1.325 ± 0.128	0.256 ± 0.211	0.218 ± 0.070	1.195 ± 0.221
Coffee	N (%)	1D	PLSR	0.119 ± 0.050	0.552 ± 0.030	1.076 ± 0.031	0.132 ± 0.159	0.567 ± 0.139	1.044 ± 0.125
Coffee	P (ppm)	BD	PLSR	0.794 ± 0.034	5.572 ± 0.273	2.241 ± 0.180	0.772 ± 0.108	6.341 ± 1.516	2.037 ± 0.417
Coffee	K (ppm)	BD	PLSR	0.298 ± 0.042	70.130 ± 2.712	1.205 ± 0.035	0.292 ± 0.179	71.650 ± 10.420	1.220 ± 0.146
Coffee	Sand (%)	BD	SVM	0.726 ± 0.065	5.948 ± 0.609	1.943 ± 0.210	0.571 ± 0.224	7.745 ± 2.124	1.630 ± 0.596
Coffee	Silt (%)	BD	PLSR	0.690 ± 0.043	4.865 ± 0.406	1.824 ± 0.124	0.679 ± 0.154	5.156 ± 1.748	1.822 ± 0.450
Coffee	Clay (%)	1D	RF	0.954 ± 0.005	3.143 ± 0.175	4.590 ± 0.274	0.775 ± 0.096	6.715 ± 1.307	2.276 ± 0.498
Cocoa	pH	2D	SVM	0.632 ± 0.094	0.294 ± 0.054	1.617 ± 0.247	0.379 ± 0.195	0.357 ± 0.096	1.328 ± 0.234
Cocoa	EC (dS m ⁻¹)	2D	RF	0.929 ± 0.005	0.0105 ± 0.0007	2.888 ± 0.208	0.332 ± 0.263	0.0264 ± 0.0065	1.120 ± 0.163
Cocoa	N (%)	SG	PLSR	0.028 ± 0.017	0.464 ± 0.025	1.026 ± 0.009	0.117 ± 0.196	0.475 ± 0.087	1.014 ± 0.015
Cocoa	P (ppm)	BD	SVM	0.276 ± 0.065	5.443 ± 0.323	1.182 ± 0.054	0.248 ± 0.192	5.724 ± 1.109	1.181 ± 0.163
Cocoa	K (ppm)	2D	SVM	0.739 ± 0.070	35.570 ± 4.431	1.968 ± 0.212	0.454 ± 0.148	52.276 ± 9.226	1.378 ± 0.197
Cocoa	Sand (%)	2D	RF	0.921 ± 0.005	3.207 ± 0.195	2.787 ± 0.173	0.260 ± 0.265	7.687 ± 1.707	1.213 ± 0.252
Cocoa	Silt (%)	BD	RF	0.905 ± 0.006	2.273 ± 0.233	2.565 ± 0.166	0.402 ± 0.177	4.614 ± 1.279	1.319 ± 0.163
Cocoa	Clay (%)	1D	PLSR	0.514 ± 0.070	6.761 ± 0.751	1.461 ± 0.107	0.387 ± 0.202	7.814 ± 2.025	1.304 ± 0.316

Note. Values are presented as mean $\pm$ standard deviation across the five outer cross-validation folds for the coffee and cocoa datasets evaluated independently. For each soil property, the best-performing model was selected primarily based on the highest mean cross-validation R^2 , followed by the lowest mean cross-validation RMSE and the highest mean cross-validation RPD. SG = Savitzky–Golay smoothing; 1D = first derivative; 2D = second derivative; BD = band depth; RF = Random Forest; SVM = Support Vector Machine; NN = Neural Network; PLSR = Partial Least Squares Regression.

and K showed limited predictive capability, with RPD values of 1.195 ± 0.221 , 1.044 ± 0.125 , and 1.220 ± 0.146 , respectively.

In cocoa soils, predictive performance was generally lower. K showed the highest R^2 (0.454 ± 0.148), followed by silt (0.402 ± 0.177), clay (0.387 ± 0.202), and pH (0.379 ± 0.195). However, none of the evaluated properties reached an RPD above 1.4. The lowest

predictive performance was observed for N ($R^2 = 0.117 \pm 0.196$; RPD = 1.014 ± 0.015), followed by P and sand. These results indicate greater limitations in estimating soil properties from NIR spectra in the cocoa dataset.

The contrasting results between production systems suggest that the predictive capability of portable NIR spectroscopy depends on the soil

property and the physicochemical variability of the evaluated soils, as well as on the preprocessing and modeling strategy. The relatively stronger performance for texture-related properties in coffee soils may reflect their closer relationships with spectrally active soil constituents, whereas the weaker performance for N and several nutrient variables suggests limited spectral sensitivity under the evaluated conditions.

The prediction of N, P, and K should be interpreted cautiously because their estimation from NIR spectra does not necessarily represent direct detection of these elements. Their predictive relationships may instead arise indirectly through associations with organic matter, clay minerals, and other spectrally active soil properties. Therefore, the observed predictive performance should be understood as statistical estimation under the evaluated conditions rather than evidence of direct NIR absorption features for these nutrients.

Overall, the nested five-fold cross-validation results demonstrate that portable NIR spectroscopy has potential for the rapid estimation of selected soil properties, particularly clay, P, pH, and silt in coffee soils. However, its performance was strongly property- and production-system dependent, highlighting the need for appropriate preprocessing, model selection, and rigorous cross-validation when applying portable NIR spectroscopy to heterogeneous tropical soils.

3.7. Fertility management guidance based on NIR-informed soil indicators

Following the framework described in Section 2.8, selected laboratory indicators and the best crop-specific NIR models were integrated to identify their potential applications in soil fertility management (Table 3). Each implication was derived from the observed magnitude of the indicator, its mean five-fold cross-validation performance, and its agronomic relevance to coffee or cocoa production. This synthesis evaluates spectral prediction not only according to statistical performance but also according to its capacity to support agronomically consistent decision-making.

4. Discussion

The results obtained show that the performance of spectral models varied substantially depending on the soil property, the applied preprocessing, and the algorithm used, which is consistent with previous studies in NIR spectroscopy (Park et al., 2024). Overall, predictive performance was higher in coffee soils than in cocoa soils, particularly for texture-related properties and phosphorus, suggesting that differences in soil composition influenced the spectral response. Similar findings have been reported by Shin et al. (2025), who indicated that soil properties associated with mineralogy and organic matter generally

Table 3

Potential agronomic applications of NIR-informed soil indicators for fertility management in coffee and cocoa systems.

Crop	Management target	Numerical evidence from this study	Best cross-validated NIR evidence	Use class	Fertility-management implication
Coffee	Acidity/pH	pH = 4.80 ± 0.81 (4.04–7.75)	SVM-2D: $R^2_{CV} = 0.665 \pm 0.221$; $RPD_{CV} = 2.085 \pm 0.921$	High (variable)	Supports the preliminary identification of areas requiring acidity diagnosis and liming assessment. The high variability among folds indicates that laboratory confirmation remains necessary before corrective applications.
Coffee	Available P	P = 22.28 ± 12.46 ppm (3.87–61.25)	PLSR-BD: $R^2_{CV} = 0.772 \pm 0.108$; $RPD_{CV} = 2.037 \pm 0.417$	High	Supports preliminary P-management zoning and targeted laboratory confirmation before establishing P ₂ O ₅ application rates through a complete fertility assessment.
Coffee	Texture-based zoning	Silt = $26.89 \pm 8.83\%$ (2.00–44.00); clay = $33.18 \pm 14.36\%$ (7.93–60.00)	Silt, PLSR-BD: $R^2_{CV} = 0.679 \pm 0.154$; $RPD_{CV} = 1.822 \pm 0.450$. Clay, RF-1D: $R^2_{CV} = 0.775 \pm 0.096$; $RPD_{CV} = 2.276 \pm 0.498$	Moderate for silt; high for clay	Supports sampling stratification, interpretation of water and nutrient retention, and preliminary delineation of differentiated management zones.
Coffee	N and K planning	N = $2.00 \pm 0.59\%$ (0.51–3.26); K = 189.15 ± 84.34 ppm (68.54–385.68)	N, PLSR-1D: $R^2_{CV} = 0.132 \pm 0.159$; $RPD_{CV} = 1.044 \pm 0.125$. K, PLSR-BD: $R^2_{CV} = 0.292 \pm 0.179$; $RPD_{CV} = 1.220 \pm 0.146$	Low	N and K fertilizer planning should not rely on NIR predictions alone. Laboratory analysis, crop demand, yield targets, foliar diagnosis, and management history are required.
Cocoa	Soil reaction/pH	pH = 8.11 ± 0.47 (6.43–8.59)	SVM-2D: $R^2_{CV} = 0.379 \pm 0.195$; $RPD_{CV} = 1.328 \pm 0.234$	Low	NIR may support exploratory identification of soil variation, but conventional pH measurement is required before managing alkalinity or nutrient-availability constraints.
Cocoa	Electrical conductivity	EC = 0.144 ± 0.030 dS m ⁻¹ (0.056–0.217)	RF-2D: $R^2_{CV} = 0.332 \pm 0.263$; $RPD_{CV} = 1.120 \pm 0.163$	Low	Laboratory EC values indicated no apparent soluble-salt constraints under the evaluated conditions. However, because the NIR model showed low predictive performance, conventional EC measurements remain necessary when salinity assessment or monitoring is required.
Cocoa	K and cation balance	K = 294.80 ± 69.18 ppm (182.91–459.85)	SVM-2D: $R^2_{CV} = 0.454 \pm 0.148$; $RPD_{CV} = 1.378 \pm 0.197$	Low	K management should therefore be supported by laboratory K measurements, assessment of the Ca–Mg–K balance, CEC or ECEC, and foliar diagnosis.
Cocoa	Texture and nutrient retention	Silt = $23.70 \pm 5.79\%$ (15.57–43.78); clay = $45.82 \pm 9.81\%$ (22.00–66.07)	Silt, RF-BD: $R^2_{CV} = 0.402 \pm 0.177$; $RPD_{CV} = 1.319 \pm 0.163$. Clay, PLSR-1D: $R^2_{CV} = 0.387 \pm 0.202$; $RPD_{CV} = 1.304 \pm 0.316$	Low	Texture remains agronomically relevant for interpreting water and nutrient retention, but NIR estimates require confirmation through conventional particle-size analysis.
Cocoa	N and P planning	N = $1.73 \pm 0.47\%$ (0.91–2.80); P = 14.39 ± 6.41 ppm (3.04–28.31)	N, PLSR-SG: $R^2_{CV} = 0.117 \pm 0.196$; $RPD_{CV} = 1.014 \pm 0.015$. P, SVM-BD: $R^2_{CV} = 0.248 \pm 0.192$; $RPD_{CV} = 1.181 \pm 0.163$	Low	N and P fertilization planning should be based on conventional soil analysis, foliar diagnosis, crop demand, and production context rather than NIR predictions alone.

Note. Laboratory values are presented as mean $\pm$ standard deviation, followed by the observed range. R^2_{CV} and RPD_{CV} are presented as mean $\pm$ standard deviation across the five outer cross-validation folds. Use classes were assigned according to mean RPD_{CV} : high when $RPD_{CV} \geq 2.0$, moderate when $1.4 \leq RPD_{CV} < 2.0$, and low when $RPD_{CV} < 1.4$. Variability among folds was additionally considered when interpreting model stability. SG = Savitzky–Golay smoothing; 1D = first derivative; 2D = second derivative; BD = band depth; RF = random forest; SVM = support vector machine; PLSR = partial least squares regression; P = phosphorus; N = nitrogen; K = potassium; EC = electrical conductivity; CEC = cation exchange capacity; ECEC = effective cation exchange capacity; NIR = near-infrared spectroscopy. These implications represent potential applications for soil screening and fertility-management support and do not constitute direct fertilizer-dose prescriptions.

exhibit stronger spectral responses in the NIR region. Likewise, Vyavahare et al. (2026) suggested that soil pH can be estimated indirectly through its relationships with spectrally active constituents such as clay minerals and organic matter.

The observed differences in predictive performance can be explained by the physicochemical mechanisms governing the interaction between NIR radiation and soil constituents. Soil texture strongly influences radiation scattering because particle-size distribution modifies the optical path length and the intensity of reflected radiation, making texture-related properties, particularly clay and silt, more spectrally responsive (Stenberg et al., 2010; Viscarra Rossel et al., 2006). In addition, clay minerals and soil organic matter exhibit characteristic spectral features that generate absorption bands in the NIR-SWIR region, particularly around 1400, 1900, and 2200 nm, providing the basis for the indirect estimation of several soil properties (Park et al., 2024; Stenberg et al., 2010; Viscarra Rossel et al., 2006). In contrast, nutrients such as N, P, and K do not exhibit direct absorption features in the NIR region. Consequently, their prediction is primarily based on indirect relationships with clay minerals, soil organic matter, moisture, and other spectrally active soil constituents rather than with the nutrients themselves (Nocita et al., 2015; Park et al., 2024; Stenberg et al., 2010). Therefore, differences in mineralogical composition, organic matter content, and soil heterogeneity between the coffee and cocoa production systems likely influenced the spectral responses and, ultimately, model performance. These mechanisms provide a plausible explanation for the differences in predictive performance observed among soil properties and between production systems.

The lower predictive performance observed in cocoa soils may be related to their narrower response ranges, reduced spectral contrast for several properties, differences in mineralogical composition, and the limited size of the crop-specific calibration dataset. These factors may have weakened the relationships between spectral information and laboratory reference values. Representative calibration datasets remain essential for improving the robustness and transferability of soil spectral models (Nocita et al., 2015).

No single modeling algorithm consistently outperformed the others across all soil properties. Instead, model performance depended on the target variable, indicating that different algorithms capture different spectral soil relationships. Likewise, first- and second-derivative preprocessing and band depth transformation frequently improved predictive performance, supporting previous findings that these techniques enhance relevant absorption features while reducing baseline shifts and scattering effects (Savitzky and Golay, 1964).

Overall, these findings demonstrate that the predictive capability of portable NIR spectroscopy depends on the interaction between soil physicochemical characteristics, spectral preprocessing, and calibration strategy. Improving calibration datasets and expanding spectral libraries will be essential to increase the robustness and transferability of predictive models across heterogeneous tropical soils.

This study has several limitations. The models were developed using relatively small crop-specific datasets and were evaluated through nested cross-validation but not through an independent external dataset. In addition, spectral measurements were obtained from oven-dried and finely sieved samples under controlled conditions; therefore, model performance may differ under field-moist or less standardized sample conditions. The absence of detailed mineralogical, soil organic matter, cation-exchange, and foliar information also limited the interpretation of indirect spectral relationships. Consequently, the models should be considered site- and dataset-specific until their transferability is assessed across additional locations, seasons, soil conditions, and independent samples.

5. Conclusions

Portable NIR spectroscopy combined with machine-learning algorithms showed potential as a rapid and complementary tool for

predicting selected soil physicochemical properties and supporting preliminary fertility-management decisions in coffee and cocoa production systems in the Peruvian Amazon. Its performance was strongly dependent on the evaluated soil property, production system, spectral preprocessing method, and predictive algorithm. The most reliable results were obtained in coffee soils, particularly for clay, available phosphorus, pH, and silt, indicating potential applications for acidity assessment, phosphorus-management zoning, sampling stratification, and the preliminary delineation of texture-based management zones. In cocoa soils, predictive performance was lower, and although potassium achieved the best result, none of the evaluated properties reached sufficient accuracy to support stand-alone management decisions.

The contrasting acidic conditions of coffee soils and the moderately alkaline, clay-rich, and low-electrical-conductivity conditions observed in cocoa soils demonstrate the importance of developing crop- and soil-specific calibration models. In coffee systems, NIR predictions may help prioritize areas requiring liming assessment, phosphorus management, and differentiated soil management. In cocoa systems, conventional analyses remain essential for verifying soil reaction, potassium balance, nutrient availability, and clay-associated retention processes before selecting fertilizer sources or management practices.

Overall, portable NIR spectroscopy should be used to complement rather than replace conventional laboratory and foliar analyses. The fertility implications derived from this study support soil screening, management-zone identification, and targeted laboratory confirmation, but they do not constitute direct fertilizer-dose prescriptions. The study did not experimentally validate fertilizer rates, crop response, yield response, or sustainability outcomes. Therefore, future research should incorporate larger and more representative calibration datasets, independent external validation, cation-exchange and foliar information, and field fertilization experiments to determine the agronomic effectiveness of NIR-supported fertility-management strategies under heterogeneous tropical conditions.

Funding

This work was subsidised by CONCYTEC through the PROCENCIA programme under the Technology Development Through Consortia for Value Chains (E041-2024-02-BM) competition, under contract N° PE501093902-2024-PROCENCIA-BM, and from the GEOMATICA public investment project (CUI N° 2255626). The APC was funded by the Vice-Rectorate for Research of the Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas.

CRedit authorship contribution statement

Marlen A. Grandez-Alberca: Writing – review & editing, Visualization, Validation, Supervision, Software, Methodology, Investigation, Data curation. **Julio Puscan-Rojas:** Writing – review & editing, Methodology, Investigation. **Teodoro B. Silva-Melendez:** Visualization, Validation, Methodology, Investigation. **Lily del Pilar Juarez-Contreras:** Visualization, Methodology, Investigation, Data curation. **Angel J. Medina-Medina:** Writing – review & editing, Visualization, Validation, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Jhon A. Zabaleta-Santisteban:** Writing – review & editing, Visualization, Validation, Methodology, Investigation, Data curation. **Abner S. Rivera-Fernandez:** Writing – review & editing, Visualization, Validation, Methodology, Investigation, Data curation. **Katerin M. Tuesta-Trauco:** Writing – review & editing, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Mirella N. Coronel-Alcalde:** Writing – review & editing, Visualization, Methodology, Investigation. **Alcides Roman:** Methodology, Investigation. **Samuel Pizarro:** Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Elgar Barboza:** Writing – review & editing, Supervision, Methodology,

Investigation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors acknowledge and express their gratitude to the farmers who facilitated access to their coffee and cocoa plots for evaluation, to Valeria Montano López for her assistance in providing soil sample results, and to Cecibel Portocarrero Díaz for her logistical and administrative support during the project implementation. We also thank the Soil and Water Research Laboratory (LABISAG) for its support in the analysis of soil samples from the evaluated plots. We confirm that the manuscript has been read and approved by all listed authors and that there are no other individuals who meet the authorship criteria but are not included in the author list. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

Data availability

Data will be made available on request.

References

- Belgiu, M., Drăgu, L., 2016. Random forest in remote sensing: a review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* 114, 24–31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>.
- Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, B., Roger, J.M., McBratney, A., 2010. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. *TrAC Trends Anal. Chem.* 29 (9), 1073–1081. <https://doi.org/10.1016/j.trac.2010.05.006>.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45 (1), 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Corwin, D.L., Lesch, S.M., 2005. Apparent soil electrical conductivity measurements in agriculture. *Comput. Electron. Agric.* 46 (1–3), 11–43. <https://doi.org/10.1016/j.compag.2004.10.005>.
- Food and Agriculture Organization of the United Nations, 2024. Una Introducción al Análisis de Suelo Utilizando Espectroscopía Visible e Infrarrojo Cercano (vis-NIR) y Espectroscopía de Infrarrojo Medio (MIR). <https://doi.org/10.4060/cb9005es>.
- García, Y., Ramírez, W., Sánchez, S., 2012. Indicadores de la calidad de los suelos: una nueva manera de evaluar este recurso. *Pastos Forrajes* 35 (2), 125–138. http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-03942012000200001&lng=es&sknm=iso&tng=es.
- Haro-Altamirano, J.P., Carrillo-Barahona, W.E., Fernández-Quinde, L.M., Toalombo-Vargas, V.M., Jara-Ruiz, C.D., 2025. Determination of the sustainability of soil quality and crop health of two varieties of coffee, Rio Blanco, Ecuador. *Trop. Subtrop. Agroecosystems* 28 (1), 55207. <https://doi.org/10.56369/tsaes.5520>.
- Instituto de Investigaciones de la Amazonía Peruana, 2010. Estudios Temáticos para la Zonificación Ecológica y Económica del departamento de Amazonas. https://terra.iiaip.gob.pe/assets/files/macro/zee-amazonas/02_Geologia_2010.pdf.
- Instituto Nacional de Estadística e Informática, 2018. Amazonas: Resultados definitivos. https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1567/01TOMO_01.pdf.
- International Trade Centre, 2021. The Coffee Guide, 4th ed. International Trade Centre.
- Javaid, M., Haleem, A., Singh, R.P., Suman, R., 2022. Enhancing smart farming through the applications of Agriculture 4.0 technologies. *Int. J. Intell. Netw.* 3 (2), 150–164. <https://doi.org/10.1016/j.ijin.2022.09.004>.
- Malley, D.F., Yesmin, L., Wray, D., Edwards, S., 1999. Application of near-infrared spectroscopy in analysis of soil mineral nutrients. *Commun. Soil Sci. Plant Anal.* 30 (7–8), 999–1012. <https://doi.org/10.1080/00103629909370263>.
- Márquez, F.R., Quispe, P., Molleapaza, N., Cabrera, S., Peña, J., 2020. Relación entre las características del suelo y altitud con la calidad sensorial de café cultivado bajo sistemas agroforestales en Cusco, Perú. *Sci. Agropecu.* 11 (4), 529–536. <https://doi.org/10.17268/sci.agropecu.2020.04.08>.
- McBratney, A., Field, D.J., Koch, A., 2014. The dimensions of soil security. *Geoderma* 213, 203–213. <https://doi.org/10.1016/j.geoderma.2013.08.013>.
- Minasny, B., Bandai, T., Ghezzehei, T.A., Huang, Y.C., Ma, Y., McBratney, A.B., Ng, W., Norouzi, S., Padarian, J., Rudiyanto, Sharififar, A., Styc, Q., Widyastuti, M., 2024. Soil science-informed machine learning. *Geoderma* 452, 117094. <https://doi.org/10.1016/j.geoderma.2024.117094>.
- Ministerio de Desarrollo Agrario y Riego, 2021. Amazonas: Información Territorial. https://www.midagri.gob.pe/portal/download/pdf/herramientas/organizaciones/dgpa/documentos/estudio_cacao/4_3_1amazonas_informefinal.pdf.
- Ministerio de Desarrollo Agrario y Riego, 2024a. Manual para el Análisis de Suelos Agrícolas y Agua Para Riego. <https://repositorio.midagri.gob.pe/handle/20.500.13036/1745>.
- Ministerio de Desarrollo Agrario y Riego, 2024b. Muestreo de Suelos y Servicio de Análisis. <https://repositorio.midagri.gob.pe/handle/20.500.13036/1716>.
- Mitu, S.M., Smith, C., Sanderman, J., Ferguson, R.R., Shepherd, K., Ge, Y., 2024. Evaluating consistency across multiple NeoSpectra (compact Fourier transform near-infrared) spectrometers for estimating common soil properties. *Soil Sci. Soc. Am. J.* 88 (4), 1324–1339. <https://doi.org/10.1002/saj2.20678>.
- Mouazen, A.M., Kuang, B., De Baerdemaeker, J., Ramon, H., 2010. Comparison among principal component, partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy. *Geoderma* 158 (1–2), 23–31. <https://doi.org/10.1016/j.geoderma.2010.03.001>.
- Neina, D., 2019. The role of soil pH in plant nutrition and soil remediation. *Appl. Environ. Soil Sci.* 2019, 1–9. <https://doi.org/10.1155/2019/5794869>.
- Nocita, M., Stevens, A., van Wesemael, B., Aitkenhead, M., Bachmann, M., Barthès, B., Ben-Dor, E., Brown, D.J., Clairotte, M., Csorba, A., Dardenne, P., Demattè, J.A.M., Genot, V., Guerrero, C., Knadel, M., Montanarella, L., Noon, C., Ramirez-Lopez, L., Robertson, J., Wetterlind, J., 2015. Soil spectroscopy: an alternative to wet chemistry for soil monitoring. *Adv. Agron.* 132, 139–159. <https://doi.org/10.1016/bs.agron.2015.02.002>.
- Park, S., Jeon, S., Kwon, N.H., Kwon, M., Shin, J.H., Kim, W.C., Lee, J.G., 2024. Application of near-infrared spectroscopy to predict chemical properties in clay rich soil: a review. *Eur. J. Agron.* 159, 127228. <https://doi.org/10.1016/j.eja.2024.127228>.
- Partida, C., Safanelli, J.L., Mitu, S.M., Murad, M.O.F., Ge, Y., Ferguson, R., Shepherd, K., Sanderman, J., 2025. Building a near-infrared (NIR) soil spectral dataset and predictive machine learning models using a handheld NIR spectrophotometer. *Data Br.* 58, 111229. <https://doi.org/10.1016/j.dib.2024.111229>.
- Pérez-Burillo, S., Cervera-Mata, A., Fernández-Arteaga, A., Pastoriza, S., Rufián-Henares, J.A., Delgado, G., 2022. Why should we be concerned with the use of spent coffee grounds as an organic amendment of soils? A narrative review. *Agronomy* 12 (11), 2771. <https://doi.org/10.3390/agronomy12112771>.
- Prananto, J.A., Minasny, B., Weaver, T., 2020. Near infrared (NIR) spectroscopy as a rapid and cost-effective method for nutrient analysis of plant leaf tissues. *Adv. Agron.* 164, 1–49. <https://doi.org/10.1016/bs.agron.2020.06.001>.
- R Core Team, 2026. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.
- Recena, R., Fernández-Cabanás, V.M., Delgado, A., 2019. Soil fertility assessment by Vis-NIR spectroscopy: Predicting soil functioning rather than availability indices. *Geoderma* 337, 368–374. <https://doi.org/10.1016/j.geoderma.2018.09.049>.
- Reza, M.N., Lee, K.H., Karim, M.R., Haque, M.A., Bicomakubwa, E., Dey, P.K., Jang, Y. Y., Chung, S.O., 2025. Trends of soil and solution nutrient sensing for open field and hydroponic cultivation in facilitated smart agriculture. *Sensors* 25 (2), 453. <https://doi.org/10.3390/s25020453>.
- Rinnan, Å., 2014. Pre-processing in vibrational spectroscopy—when, why and how. *Anal. Methods* 6 (18), 7124–7129. <https://doi.org/10.1039/C3AY42270D>.
- Sanderman, J., Partida, C., Safanelli, J.L., Mitu, S.M., Ge, Y., Murad, O., Shepherd, K., 2024. Near-infrared (NIR) soil spectral library using the NeoSpectra Handheld NIR Analyzer by Si-Ware [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.13122321>.
- Savitzky, A., Golay, M.J.E., 1964. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* 36 (8), 1627–1639. <https://doi.org/10.1021/ac60214a047>.
- Saxton, K.E., Rawls, W.J., 2006. Soil water characteristic estimates by texture and organic matter for hydrologic solutions. *Soil Sci. Soc. Am. J.* 70 (5), 1569–1578. <https://doi.org/10.2136/sssaj2005.0117>.
- Schmidinger, J., Barkov, V., Tavakoli, H., Correa, J., Ostermann, M., Atzmüller, M., Gebbers, R., Vogel, S., 2024. Which and how many soil sensors are ideal to predict key soil properties: a case study with seven sensors. *Geoderma* 450, 117017. <https://doi.org/10.1016/j.geoderma.2024.117017>.
- Shin, S.K., Lee, S.J., Park, J.H., 2025. Prediction of soil properties using vis-NIR spectroscopy combined with machine learning: a review. *Sensors* 25 (16), 5045. <https://doi.org/10.3390/s25165045>.
- Soriano-Disla, J.M., Janik, L.J., Viscarra Rossel, R.A., MacDonald, L.M., McLaughlin, M. J., 2014. The performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties. *Appl. Spectrosc. Rev.* 49 (2), 139–186. <https://doi.org/10.1080/05704928.2013.811081>.
- Stenberg, B., Viscarra Rossel, R.A., Mouazen, A.M., Wetterlind, J., 2010. Visible and near infrared spectroscopy in soil science. *Adv. Agron.* 107, 163–215. [https://doi.org/10.1016/S0065-2113\(10\)07005-7](https://doi.org/10.1016/S0065-2113(10)07005-7).
- Thingujam, U., Prabha, D., Ghosh Bag, A., Thingujam, V., Darshan, N.P., Dutta, S., Gorain, S., 2025. From point sensing to intelligent systems: a comprehensive review on advanced sensor technologies for soil health monitoring. *Discov. Sens.* 1, 27. <https://doi.org/10.1007/s44397-025-00028-8>.
- Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J., Skjemstad, J.O., 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* 131 (1–2), 59–75. <https://doi.org/10.1016/j.geoderma.2005.03.007>.
- Vogel, H.J., Eberhardt, E., Franko, U., Lang, B., Ließ, M., Weller, U., Wiesmeier, M., Wollschläger, U., 2019. Quantitative evaluation of soil functions: potential and state. *Front. Environ. Sci.* 7, 164. <https://doi.org/10.3389/fenvs.2019.00164>.
- Vyavahare, G.D., Yun, J.J., Park, J.H., Shim, J.H., Kim, S.H., Kim, K., Roh, A., Kim, S.H., Jang, H.J., Ng, W., Jeon, S., 2026. Applications and challenges of visible-near-infrared and mid-infrared spectroscopy in soil analysis: Chemometric approaches

- and data fusion. *Agriculture* 16 (1), 135. <https://doi.org/10.3390/agriculture16010135>.
- Wang, Z., Wu, Q., Kamruzzaman, M., 2022. Portable NIR spectroscopy and PLS based variable selection for adulteration detection in quinoa flour. *Food Control* 138, 108970. <https://doi.org/10.1016/j.foodcont.2022.108970>.
- Weil, R.R., Brady, N.C., 2017. *The Nature and Properties of Soils*, 15th ed. Pearson.