remote sensing

MDPI

Article

Vis-NIR Spectroscopy and Machine Learning for Prediction of Soil Fertility Indicators and Fertilizer Recommendation in Andean Highland and Rainforest Agroecosystems

Samuel Pizarro [1,2], Dennis Ccopi [1], Kevin Ortega [3], Duglas Contreras [1], Javier Ñaupari [4], Deyvis Cano [5], Solanch Patricio [6], Hildo Loayza [7] and Orly Enrique Apolo-Apolo [8,*]

1 Estación Experimental Agraria Santa Ana, Dirección de Servicios Estratégico, Instituto Nacional de Innovción Agraria (INIA), Carretera Saños Grande—Hualahoyo Km 8 Santa Ana, Huancayo 12007, Peru; samuel.pizarro@untrm.edu.pe (S.P.); dennisccopit@gmail.com (D.C.); contreraspino17@gmail.com (D.C.)
2 Centro de Investigación en Geomática Ambiental (CIGA), Instituto de Investigación para el Desarrollo Sustentable de Ceja de Selva (INDES-CES), Universidad Nacional Toribio Rodríguez de Mendoza, Chachapoyas 01001, Peru
3 Estación Experimental Agraria Donoso, Dirección de Servicios Estratégicos Agrarios, Instituto Nacional de Innovación Agraria (INIA), Lima 15200, Peru; kevinorqu@gmail.com
4 Grupo en Ecología y Utilización de Pastizales, Universidad Nacional Agraria La Molina, Lima 15024, Peru; jnaupariv@lamolina.edu.pe
5 Postgraduate Program in Remote Sensing (PPGSR), Federal University of Rio Grande do Sul (UFRGS), Porto Alegre 91501-970, RS, Brazil; deyvis.cano@udh.edu.pe or deyvis.cano@ufrgs.br
6 Estación Experimental Agraria Santa Ana, Dirección de Recursos Genéticos y Biotecnología, Instituto Nacional de Innovación Agraria (INIA), Carretera Saños Grande-Hualahoyo Km 8 Santa Ana, Huancayo 12007, Peru; solanch.patricio.r@gmail.com
7 International Potato Center (CIP), Headquarters, Lima 15024, Peru; h.loayza@cgiar.org
8 IRTA, Postharvest, Fruitcentre, 25194 Lleida, Catalonia, Spain
* Correspondence: enrique.apolo@irta.cat

Highlights

What are the main findings?

- Vis-NIR spectroscopy combined with machine learning reliably predicts stable soil properties (pH, organic matter, texture) and nitrogen recommendations, but performs poorly for mobile nutrients such as phosphorus, potassium, and sodium.
- Differences between agroecosystems (Andean highlands vs. central rainforest) are larger than discrepancies between observed and predicted values, highlighting the dominant role of ecological context in soil management and the limitations of relying solely on statistical metrics for fertilizer recommendations.

What is the implication of the main finding?

- Vis-NIR spectroscopy can be reliably applied for rapid assessment of stable soil properties and for supporting nitrogen management, but caution is needed when estimating mobile nutrients or deriving fertilizer recommendations directly.
- Agronomic decision-making frameworks should incorporate ecosystem-specific context and validation steps to avoid misleading recommendations based solely on statistical prediction metrics.

Academic Editor: Mohammadmehdi Saberioon

Received: 5 March 2026
Revised: 22 April 2026
Accepted: 23 April 2026
Published: 26 April 2026

Abstract

This study evaluated the use of visible and near-infrared (Vis-NIR) spectroscopy combined with machine learning (ML) algorithms to predict soil fertility-related properties in two contrasting agroecological regions of Peru: the Highlands and the Rainforest. A total

of 297 soil samples were analyzed using portable spectroradiometers covering a spectral range of 350–2500 nm, applying transformations such as Savitzky–Golay smoothing, first derivative, and band depth. Predictive models were developed using PLSR, Random Forest, Support Vector Machines, and neural networks. Results show variable predictive performance across soil properties and ecosystems. Organic matter in Highland soils and calcium in Rainforest soils achieved the strongest test-set accuracy ($R^2 > 0.70$), while pH and texture fractions showed moderate performance ($R^2 = 0.42–0.67$), and mobile nutrients including phosphorus, potassium, and sodium showed limited predictive accuracy due to their weak spectral expression. Spectral predictions were further integrated into a structured nutrient balance framework to assess agronomic reliability. Nitrogen fertilizer recommendations showed the strongest agreement between observed and predicted values across both ecosystems, whereas K_2O and CaO recommendations in Highland soils were substantially underestimated, demonstrating that property-level statistical performance does not guarantee agronomic reliability. These findings confirm that Vis-NIR spectroscopy combined with ML represents a fast, cost-effective, and sustainable alternative to conventional soil analysis, especially in rural areas with limited laboratory infrastructure. Expanding regional calibration datasets and exploring mid-infrared FTIR spectroscopy as a complementary technology are identified as priority directions for improving predictions of agronomically critical nutrients.

Keywords: Vis-NIR spectroscopy; machine learning; Andean highlands; rainforest; soil fertility; prediction models; fertilizer recommendations; precision agriculture

1. Introduction

Soil is a fundamental resource for agricultural production and global food security, as its capacity to support crop growth is directly linked to its fertility [1]. Fertility plays a crucial role in creating optimal conditions for healthy plant development and is essential for ensuring efficient and sustainable agricultural production [2,3]. Without proper fertility management, soils can progressively degrade, losing their ability to sustain agricultural systems over the long term.

In the highlands and the Peruvian rainforest, soil management faces particular challenges. In the highlands, steep slopes, accelerated erosion, shallow soil profiles, and acidic or organic matter-poor soils pose serious constraints to agricultural productivity [4]. Meanwhile, rainforest soils, despite their rich biological diversity, typically exhibit low natural fertility, high acidity, and rapid nutrient mineralization, which complicates sustained agriculture without external inputs [5]. In both regions, local farmers need to frequently monitor key soil properties such as pH, texture, cation exchange capacity (CEC), and available nutrients, including phosphorus (P), potassium (K), calcium (Ca), magnesium (Mg), and sodium (Na), all of which are vital for crop productivity.

Traditional soil analysis methods, although accurate, present important limitations. Conventional laboratory methods often require chemical reagents, sophisticated equipment, and protocols that are costly and time-consuming [6]. Likewise, in areas with difficult access, long transport times from the field to the laboratory may become an additional constraint for the timely evaluation of samples [7]. Likewise, advanced techniques such as electron microscopy or gamma-ray spectrometry, while effective, demand expensive equipment and specialized personnel, limiting their field applicability.

In this context, visible and near-infrared (Vis-NIR) spectroscopy has emerged as an accessible and promising technological alternative for the rapid, non-destructive, and cost-effective estimation of certain soil properties [8]. This technique is based on the

interaction between radiation and the physical and chemical components of the soil, generating a spectral signature that can be correlated with conventional analytical data [9]. It has shown particularly strong performance in estimating soil solid-phase properties such as texture, organic carbon content, and moisture [10].

In Peru, visible and near-infrared spectroscopy (Vis-NIR) remains a little-known and rarely applied technology, particularly for extensive agricultural areas. Its implementation offers a strategic opportunity to improve agronomic decision-making, especially in regions with limited access to analytical infrastructure. According to Ahmadi et al. [8], Vis-NIR can accurately predict variables such as organic matter, clay content, and cation exchange capacity (CEC), which supports more efficient, real-time soil management. Furthermore, this technology can be applied directly in the field using portable sensors, simplifying data collection and enabling the generation of detailed soil variability maps for use in precision agriculture [11].

This simplifies data collection and supports the generation of high-resolution soil variability maps for use in precision agriculture. In vineyards of central and northwestern Spain, VIS–NIR–SWIR diffuse reflectance spectroscopy showed potential for soil characterization in precision viticulture [12]. At the farm scale in São Paulo, Brazil, Vis-NIR spectroscopy was used for robust soil property mapping, including particle-size fractions and exchangeable calcium, and made it possible to reduce the need for conventional chemical analyses while maintaining a given level of mapping accuracy [13].

Vis-NIR also presents a more environmentally sustainable alternative to conventional analysis methods, as it is non-destructive and chemical-free. It reduces the use of reagents and minimizes waste generation, which is particularly beneficial in degraded or resource-limited areas [12]. Moreover, incorporating broader spectral regions, such as shortwave infrared (SWIR, 1300–2500 nm), has been shown to improve predictions of clay minerals and carbonates, reducing RMSE by up to 25% compared with Vis-NIR alone [14]. This points to the potential of multispectral strategies for comprehensive soil characterization.

However, Vis-NIR spectroscopy is more limited for the direct determination of plant-available nutrients, as spectral soil sensors generally provide an indication rather than a direct measurement of nutrient availability, and only a few in-field methods currently allow direct determination of plant-available nutrients [15]. These variables tend to show inconsistent correlations with spectral data due to their high temporal and spatial variability, as well as the influence of moisture content or redox conditions [11].

Despite these constraints, recent advances in miniaturized and portable NIR spectroscopy have facilitated the transfer of this technology toward field applications. Recent studies indicate that spectrometers are becoming more portable and accessible, and that miniaturized, relatively inexpensive devices can support rapid and cost-effective soil assessment [16].

These tools have demonstrated high repeatability and reproducibility, even under variable conditions, making them suitable for rapid diagnostics on smallholder farms, cooperatives, or territories with limited access to laboratories. In the Peruvian context, this technology represents a strategic opportunity to bridge the gap in accessible and timely soil diagnostics, especially considering that the country has approximately 11.6 million hectares of agricultural land, much of it located in remote areas or regions with limited technical resources [17].

One of the main challenges for the practical application of Vis-NIR is the interpretation of spectral data, which typically requires advanced statistical processing. In this regard, machine learning (ML) techniques have proven highly useful for extracting complex patterns and improving the prediction of soil properties [18]. Machine learning algorithms have been widely applied to predict soil properties from infrared spectra, and recent deep-

learning approaches, such as convolutional neural networks, have further expanded the capacity to model soil spectral data [19]. The integration of Vis-NIR with ML has enabled significant advances not only in agriculture, but also in environmental monitoring and product quality assessment. For instance, it has been successfully applied to estimate heavy metals in soil [20] and to predict organic matter in ecologically complex environments [21]. Its potential has also been demonstrated in combination with other spectral or chemometric techniques, optimizing soil characterization in agropastoral landscapes and in horticultural products [22].

Nonetheless, the performance of Vis-NIR under field conditions still faces technical challenges. Factors such as soil moisture variability, mineralogical heterogeneity, and vegetation cover can affect spectral quality and model accuracy [23]. This is especially relevant in megadiverse countries like Peru, where locally calibrated spectral libraries tailored to the main soil types of each agroecological region are essential. Currently, the limited availability of representative national-level spectral databases is a major obstacle to the effective transfer and application of spectral models developed in other geographic and edaphoclimatic contexts [24]. This gap restricts the capacity to adequately calibrate models to local conditions, affecting their accuracy and usefulness for soil characterization and monitoring [25].

Beyond the accurate estimation of soil fertility attributes, the practical value of Vis-NIR spectroscopy lies in its ability to inform nutrient management decisions [26]. Translating predicted soil properties into fertilizer requirements enables the assessment of whether model-derived estimates preserve agronomic decision consistency under contrasting agroecological conditions. In this study, both laboratory-measured and spectrally predicted soil properties were integrated into a nutrient balance framework to estimate crop-specific fertilizer recommendations for Highland and Rainforest systems. By comparing fertilizer-equivalent nutrient supply derived from observed and predicted datasets, we evaluate whether prediction uncertainty meaningfully alters final fertilizer doses. This approach links digital soil characterization to site-specific nutrient management, providing a direct pathway from spectral modeling to agronomic application in smallholder production systems.

Given this context, the present study aims to develop robust predictive models for fast prediction of soil fertility-related properties using laboratory-based Vis-NIR spectroscopy on preprocessed soil samples, with emphasis on its applicability in highland and rainforest soils. Specifically, it seeks to establish optimal configurations for reliable field spectrum acquisition and to address the following key questions:

(i) Which spectral features correlate with soil fertility-related and texture properties that are relevant to local farmers?
(ii) What level of accuracy can different machine learning regression models achieve when estimating key soil fertility-related and texture properties from Vis-NIR spectra?
(iii) To what extent can spectrally predicted soil properties be reliably integrated into a nutrient balance framework to generate consistent fertilizer recommendations across contrasting agroecological regions?

2. Materials and Methods

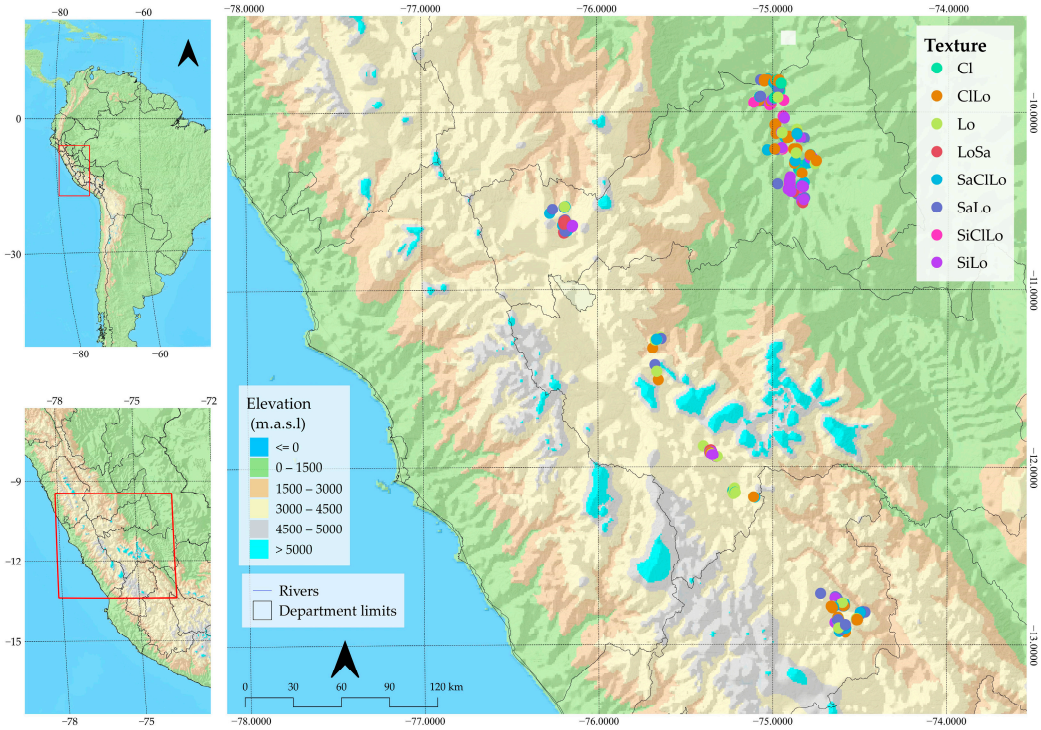
Soil analyses and spectral measurements were carried out at the Santa Ana Agricultural Experimental Station of the National Institute for Agrarian Innovation (INIA-Peru), located in the Mantaro Valley (75°13′17.60″W, 12°0′42.36″S), in the central highlands of Peru. This facility functioned as the central hub for the collection and analysis of 297 soil samples of topsoil (30 cm), obtained from three departments: Pasco, Junín, and Huancavelica. Of these, 168 samples were collected from the highlands (Junín, Pasco and

Huancavelica) and 129 from the central rainforest (Pasco). Figure 1 shows the spatial distribution of the samples by geographic origin.

In the humid montane forest of Pasco, the soil sampling sites are located between 250 and 1450 m a.s.l., within a tropical humid climate zone, with average annual temperatures around 23 °C and precipitation exceeding 2000 mm per year, exhibiting marked seasonality. The soils exhibit high textural variability, predominantly sandy clay loam (SaClLo), clay loam (ClLo), and sandy loam (SaLo), followed by loam-silt (LoSa) and loam (Lo) textures. These edaphoclimatic conditions, characteristic of the humid montane forest, favor the development of soils with good moisture retention capacity and moderate organic matter content [27,28].

In contrast, the highland sampling zones in Junín, Pasco and Huancavelica are situated between 3000 and 4500 m a.s.l., under a cold mountain climate, with average annual temperatures ranging from 4 to 12 °C, and annual rainfall between 700 and 1200 mm. In Junín, the samples span a gradient from the Mantaro Valley to highlands, with predominant silty loam (SiLo) and silty clay loam (SaClLo, SiClLo) textures, which are naturally fertile but show increasing acidity with elevation [29]. In Huancavelica, soils are predominantly clay loam (ClLo), with low organic matter content and high susceptibilityto erosion due to steep slopes and limited vegetation cover [30].

Soil taxonomic classification of the sampling zones was obtained from the Peruvian soils map (1:5,000,000) published by the National Institute of Natural Resources [31]. In the Highland zones, four soil type associations were identified across the 168 samples: Eutric Leptosol—Haplic Kastanozem ($n = 85$), Dystric Regosol—Dystric Cambisol ($n = 44$), Eutric Leptosol—Eutric Cambisol ($n = 30$), and Dystric Leptosol—Dystric Cambisol—Dystric Regosol ($n = 9$). In the Rainforest zone, two soil type associations were identified across the 129 samples: Dystric Cambisol—Haplic Alisol (n = 120) and Eutric Leptosol—Eutric Cambisol—Eutric Regosol ($n = 9$). This diversity in soil taxonomic units is consistent with the wide gradients in elevation, climate, slope, texture, and organic matter content observed across both agroecological regions.



https://doi.org/10.3390/rs18091331

Figure 1. Geographic distribution of the sampling sites in the central region of Peru. Shows the location of collected soil samples categorized by texture classes (e.g., Cl, ClLo, Lo), overlaid on an elevation gradient (in meters above sea level).

2.1. Methodological Workflow for Spectral Soil Analysis

Figure 2 shows the methodological workflow, encompassing soil sampling and laboratory analysis through the development and validation of predictive models (PLSR, RF, SVM, and NN) using 5-fold cross-validation, culminating in the generation of fertilizer recommendations.

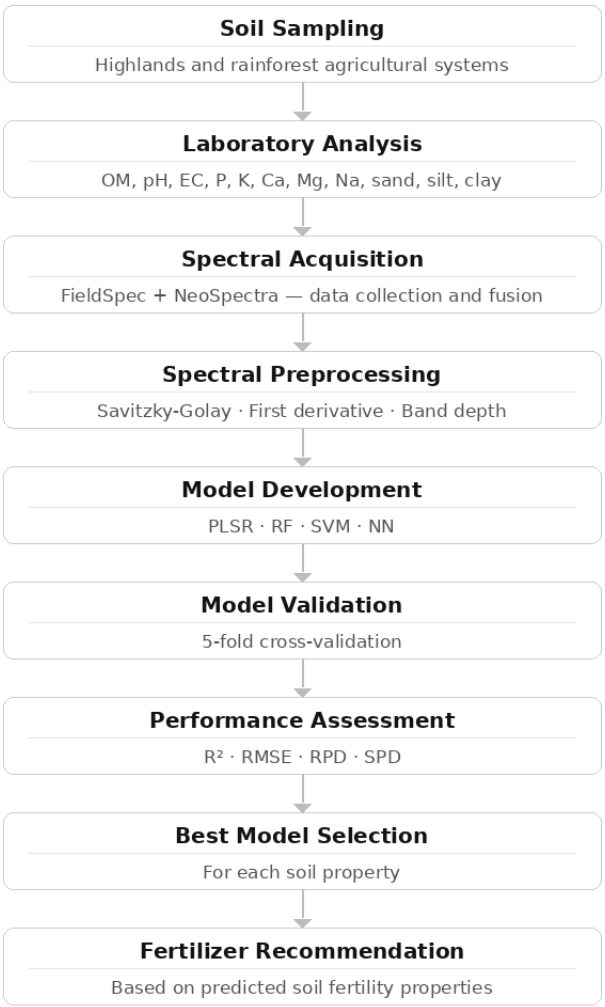


Figure 2. Methodological framework for soil property prediction using visible and near-infrared spectroscopy.

2.2. Sampling Sites and Soil Laboratory Analyses

Soil samples were collected using a targeted sampling strategy, focusing exclusively on active agricultural plots. In the Highland zones, sampling was conducted in potato plots and starch corn fields, the latter being commonly grown as a rotation crop with potato in smallholder farming systems of the central Highlands [29]. In the Rainforest zone, samples were collected from coffee and cacao plantations, as well as yellow corn fields, which are typically cultivated as a primary pioneer crop prior to the establishment of permanent coffee and cacao plantations, and which continue to be grown in association with these crops by smallholder farmers [32]. This purposive approach was adopted because not all land within the study area is under agricultural use, and the objective was to

characterize soil fertility-related properties specifically within productive agricultural lands relevant to local farming systems.

The chemical analysis of soil fertility was carried out at the Soil, Water and Foliar Analysis Laboratory (LABSAF) of the Santa Ana Experimental Station, INIA-Peru. Eleven parameters were analyzed: pH, electrical conductivity (EC), organic matter (OM), phosphorus (P) (Olsen method), potassium (K), calcium (Ca), magnesium (Mg), sodium (Na), and the percentages of sand, clay, and silt. The methods used included: the Walkley-Black method [33] for organic matter, the Olsen method for available phosphorus, ammonium acetate extraction for cation exchange capacity and exchangeable bases, and standard procedures for pH and EC.

2.3. Soil Spectral Collection

A total of 297 absolute soil reflectance measurements were conducted to characterize the optical properties of the samples. Prior to analysis, all samples were air-dried and sieved, then carefully placed into 15 cm diameter Petri dishes. Spectral measurements were performed using two spectroradiometers: the FieldSpec® HandHeld (Analytical Spectral Devices, Inc., Boulder, CO, USA) and the NeoSpectra Scanner (Si-Ware Systems, Cairo, Egypt). Technical specifications are showed in Table 1.

The FieldSpec® HandHeld is a portable and lightweight instrument equipped with a 512-channel photodiode array, covering the spectral range of 325 to 1075 nm. This model uses a fixed reflective holographic grating coupled to a 512-element silicon photodiode array detector.

In contrast, the NeoSpectra Scanner operates in the 1350–2500 nm spectral range and employs a miniaturized MEMS-based Fourier Transform Near-Infrared (FT-NIR) interferometer. This device enables rapid and non-destructive spectral acquisition, with a typical resolution of 16 nm (FWHM at 1550 nm) and guaranteed signal-to-noise ratio (SNR) exceeding 170:1 across the spectral range. In addition, it features a sample holder that defines a sampling area approximately 10 mm in diameter.

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

Specification	FieldSpec® HandHeld	NeoSpectra Scanner
Spectral Range (nm)	325–1075	1350–2500 (7400–4000 cm^{-1})
Sampling Interval (nm)	1.5	-
Spectral Resolution (FWHM)	3.5 @ 700 nm	16 nm @ 1550 nm
Signal-to-Noise Ratio (SNR)	-	>170:1 (Typical: 2000:1 at λ = 2350 nm, 2 s scan time)
Field of View/Sample Coverage	25°	~0.4 inch (~10 mm diameter)

Preprocessed soil samples were transported to the laboratory facilities of the Santa Ana Agricultural Experimental Station (INIA-Peru) prior to spectral acquisition. Spectral measurements using the FieldSpec spectroradiometer were conducted outdoors between 10:30 and 11:00 a.m. under direct sunlight. The sensor was positioned in the nadir direction relative to soil samples using a tripod, maintaining a fixed distance of 10 cm between the 1° FOV foreoptic and the soil sample surface. This configuration resulted in a measurement footprint of 1.7 mm diameter. Each measurement was performed at 80% of the instrument's radiometric resolution and consisted of at least five individual readings, which were averaged to obtain a single representative value. To calculate absolute reflectance, radiance measurements were normalized against a white reference panel (Spectralon, Labsphere Inc., North Sutton, NH, USA), known for its exceptional Lambertian reflectance properties. White reference readings were taken at 5-min intervals to minimize reflectance drift caused by variations in natural illumination. All radiance data were

processed using ViewSpec Pro v.6.2 software (Analytical Spectral Devices, Inc., Boulder, CO, USA) [34,35].

In the case of the NeoSpectra Scanner, measurements were conducted in parallel, by placing the device directly on the surface of the sample, without the need for a tripod or angular adjustments. The instrument operates in the 1350 to 2500 nm range using a miniaturized FT-NIR interferometer based on MEMS technology, and incorporates its own internal light source, making it independent of natural lighting conditions. Readings were processed through the manufacturer's mobile application, and a sampling footprint of approximately 10 mm in diameter was used.

In this study, we integrated spectral reflectance data from both instruments (FieldSpec and NeoSpectra) to identify optimal wavelength regions for predicting soil properties. The data fusion approach resulted in a spectral gap between 1000–1350 nm due to the non-overlapping measurement ranges of the two devices [17,32].

2.4. Spectral Preprocessing

Prior to model development, raw spectral reflectance data were subjected to three preprocessing transformations: Savitzky–Golay (SG) smoothing for noise reduction, first derivative (1D) transformation to enhance absorption features and remove baseline effects, and band depth (BD) transformation to normalize absorption intensities relative to a continuum baseline. Each transformation was applied independently across the full spectral range (350–2500 nm), and the best-performing input configuration for each soil property and ecosystem was selected based on cross-validation results.

Spectral data processing was performed using the hsdar package (v0.5.1) [36] within the R statistical environment (R Core Team, v4.3). A Savitzky–Golay filter with a window size of 15 was applied to preserve spectral signal shape while effectively reducing noise and baseline drift, making it particularly suitable for hyperspectral data [37].

Six distinct datasets were compiled for soil property prediction. The first comprised the full-spectrum SG-filtered data. The second incorporated the first-order derivative of reflectance filtered using the SG method (1D), while the third contained band depth (BD) spectrum measurements. BD represents wavelength-specific radiation absorption intensity relative to an estimated spectral continuum between adjacent wavelengths of the absorption feature [37]. The BD calculation involves comparative analysis between reflectance at the absorption feature center and an interpolated line between its shoulders, utilizing a continuum hull surrounding each SG spectrum, ensuring identification of discrete spectral features rather than consolidating them into broader features [38], as represented in Equation (1).

$$BD = \frac{Rc - Rb}{Rc} \quad (1)$$

where Rb represents the reflectance value of a specific spectral band, and Rc denotes the reflectance of the corresponding spectral continuum.

The remaining three datasets (4th to 6th) comprise normalized spectral indices derived from each aforementioned spectral form. The fourth dataset incorporates indices calculated from the SG filter (I-SG), the fifth from the first-order derivative values, and the sixth from the band depth spectrum. These normalized indices were computed according to Equation (2).

$$Index = \frac{(Ri - Rj)}{(Ri + Rj)} \quad (2)$$

where i and j ranged from 350 to 2500 nm, with $i \neq j$. Ri and Rj represent reflectance values at specific wavelengths obtained from either the SG smoothed spectrum depending on the dataset.

2.5. Variable Selection Method

For machine learning models, feature selection was carried out using the Successive Projections Algorithm (SPA), a forward variable selection method that identifies optimal, non-redundant reflectance values by projecting individual bands onto orthogonal spaces and selecting those with the highest vector magnitudes. To enhance computational efficiency while preserving spectral representativeness, the number of selected wavelength reflectance values was constrained to 1–30. The selected subsets were further evaluated through iterative modelling and 5-fold cross-validation to refine the variable set with minimal collinearity and maximal predictive relevance [39].

2.6. Prediction Models

Partial Least Squares Regression (PLSR) is a multivariate latent-variable method used to model the relationship between predictor variables and response variables, particularly when the data contain many and potentially collinear variables [40]. It is particularly useful when predictors are numerous and highly collinear. PLSR is commonly applied in spectroscopy, for example, to estimate soil properties from Vis-NIR spectral [41]. To balance model complexity and predictive accuracy, PLSR models were configured with 30 principal components and validated using 5-fold cross-validation, ensuring robust performance assessment and minimizing overfitting within the spectral analysis framework.

Additionally, three machine learning algorithms, Random Forest (RF), Support Vector Machines (SVM), and Neural Networks (NN) were evaluated for predicting soil properties from reflectance measurements.

RF constructs decision tree ensembles to identify patterns in soil parameters, excelling at capturing non-linear relationships even with heterogeneous data [42]. SVM establishes precise classification boundaries for soil categorization based on parameters like pH and cation exchange capacity, enabling detailed agricultural land segmentation [43]. Neural networks model complex non-linear relationships between spectral inputs and soil quality indicators by iteratively adjusting internal weights during training, enabling accurate prediction of fertility parameters across diverse soil conditions [44].

For Random Forest, the number of trees ranged from 50 to 1000 in increments of 50, with mtry fixed at 8 and bagging values between 0.1 and 1. For Support Vector Machines (SVM), the model was tuned with a cost parameter C = 2 and a radial kernel parameter gamma = 0. 0523. The Artificial Neural Network (ANN) was configured with a single hidden layer containing 3 neurons, and a weight decay parameter of 0.1, allowing for efficient learning while reducing overfitting.

Highlands and Rainforest datasets were analyzed independently, using 70% of the data for training and 30% for testing. A K-fold cross-validation (K = 5) was applied within the training set to optimize model hyperparameters, with each dataset divided into five subsets. In each iteration, one subset was used as the validation set, while the remaining K − 1 subsets were used for training [45,46]. This procedure provided a consistent basis for identifying the most suitable model for predicting soil quality and fertility in each agroecosystem.

2.7. Modelling and Performance Assessment

To evaluate the performance of the developed models, accuracy assessments related to error were conducted. Root Mean Squared Error (RMSE), and the Ratio of Performance to Deviation (RPD) were used to compare the predictive accuracy of the different models [41,47]. RMSE assessed the magnitude of error between the observed and predicted values; MAE evaluated the average magnitude of prediction errors, regardless of their direction. RPD, calculated as the ratio between the standard deviation of observed values and the RMSE, was included as an additional indicator of predictive capacity. RPD values

greater than 2 indicate excellent predictive performance, values between 1.4 and 2 are considered acceptable, and values below 1.4 suggest limited predictive ability [8,48]. RMSE and RPD express prediction error in the same units as the variable of interest.

Additionally, the Pearson coefficient (r) was used to measure the correlation degree between measured and predicted soil parameters.

The RMSE in Equation (3), and the RPD in Equation (4) [49] were calculated for both calibration and validation phases.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(yi - \bar{y}i)^2}{n}} \quad (3)$$

$$RPD = \frac{\sigma}{RMSE} \quad (4)$$

where n is the number of observations, $\hat{y}$ represents the predicted value, while $\bar{y}$ is the mean of the observed values, y corresponds to the observed (measured) value and σ the standard deviation [37].

R^2 (coefficient of determination) is a statistical measure that indicates how much of the variability in a dependent variable that is predictable from the independent variable in a model. It assesses how well the predictions fit the actual data. Its value ranges from 0 to 1, although it can be negative if the model performs worse than simply using the mean Equation (5) [50].

$$R^2 = 1 - \frac{RSS}{TSS} \quad (5)$$

where *TSS* measures the total variability in the data. *RSS* represents the model error. Their relationship is used to calculate R^2 and assess the model's fit.

To complement the error-based metrics, the Symmetric Percent Difference [51] in Equation (6) was calculated for each observation to assess the relative magnitude and direction of prediction errors across soil properties and ecosystems. *SPD* is defined as:

$$SPD(\%) = 200 \times \frac{\hat{y}_i - y_i}{\hat{y}_i + y_i} \quad (6)$$

where $\hat{y}_i$ is the predicted value and y_i is the observed value for the *i-th* sample. Unlike the standard percent error, SPD is symmetric with respect to over- and under-predictions, avoiding asymmetric penalization when predicted and observed values are exchanged. Values close to zero indicate high agreement between observed and predicted values, while large positive or negative values reflect systematic over- or under-prediction, respectively. The distribution of *SPD* values across all validation samples was visualized using violin plots, allowing a comprehensive assessment of prediction bias, spread, and the presence of outliers for each soil property.

2.8. Fertilizer Recommendation Framework

Fertilizer recommendations were estimated by integrating crop nutrient demand, soil nutrient supply, stoichiometric conversion factors, and fertilizer-use efficiency coefficients [52,53]. Parameter values are provided in Supplementary Tables S1–S5.

2.8.1. Crop Nutrient Demand

Maize was selected as the reference crop for the nutrient balance framework because it is cultivated in both agroecological zones studied, namely starch corn in the Highland zone and yellow hard corn in the Rainforest zone, enabling a consistent and comparable assessment of fertilizer recommendations across both ecosystems. Nutrient demand was calculated for a target grain yield of 2.5 t ha^{-1} under these two production systems [54,55],

using elemental nutrient extraction coefficients (Supplementary Table S1) multiplied by the target yield and adjusted by a biomass correction factor (1.55) to account for total aboveground nutrient requirements [56].

2.8.2. Soil Nutrient Supply

Soil nutrient supply was estimated for the 0–30 cm layer using either laboratory-measured (Observed) or spectrally predicted (Predicted) soil properties. Bulk density (BUD, t m^{-3}) was estimated from soil texture and organic matter using: [57,58].

$$BUD = 1.66 - 0.004 \cdot Clay - 0.002 \cdot Silt - 0.005 \cdot OM \quad (7)$$

where Clay, Silt, and OM are expressed in percentage (%).

Soil mass for the 0–30 cm layer was then calculated as the product of bulk density, sampling depth (0.30 m), and surface area (10,000 m^2 ha^{-1}), yielding soil mass in t ha^{-1} [59].

Nitrogen supply was estimated from organic matter mineralization, considering the proportion of N in organic matter (5%), an annual mineralization rate (1.3%), and seasonal adjustment according to cropping duration (4 months) [60].

For phosphorus and potassium, soil nutrient availability was adjusted using correction factors based on soil pH and cation exchange capacity (Supplementary Table S3). In Rainforest soils, additional adjustments were applied using a texture-pH dependent nutrient availability matrix (Supplementary Table S4), which accounts for differences in nutrient retention and loss processes under humid tropical conditions. These efficiency factors were used exclusively to estimate the plant-available soil nutrient fraction [61].

Exchangeable Ca and Mg were converted from cmolc kg^{-1} to kg ha^{-1} using atomic weight and valence corrections [53]. All nutrients were then expressed in fertilizer-equivalent forms using stoichiometric conversion factors (Supplementary Table S2).

2.8.3. Fertilizer Dose Calculation

Final fertilizer doses were determined as the positive difference between crop nutrient demand and estimated soil nutrient supply, divided by standardized fertilizer-use efficiency coefficients (Supplementary Table S4) [52,62]. These coefficients were applied uniformly across both Highlands and Rainforest systems during the final recommendation stage.

3. Results

3.1. Soil Physical and Chemical Characteristics

Table 2 shows the descriptive statistics of soil samples from the highlands and the rainforest, highlighting notable regional variability in fertility parameters. Soil pH values averaged 6.51 in the highlands and 5.66 in the rainforest, both with low coefficients of variation (CV ≈ 18%), indicating relatively stable acidity levels. Highland soils tended to be near neutral, while rainforest soils were more acidic.

EC showed a strong regional contrast. The highlands soils samples recorded a higher mean of 7.76 dS/m with an extremely high CV of 189.71%, compared with 3.51 dS/m (CV = 56.99%) in the rainforest, reflecting greater variation in soluble salt concentrations at higher elevations. OM content was higher in the highlands soil samples (4.14%) than in the rainforest (2.44%), with moderate to high variability (CV = 76.64% and 53.3%, respectively), which may be attributed to differences in climate, vegetation cover, and land use.

Regarding available nutrients, P and K levels were considerably higher in the highlands soil samples (P = 53.96 mg/kg, K = 234.57 cmol/kg) than in the rainforest (P = 22.96 mg/kg, K = 94.95 cmol/kg), with both nutrients showing high variability (CV > 100%), indicating heterogeneous fertility conditions. For exchangeable cations, Ca and Mg were also more abundant in the highlands soil samples (Ca = 19.82 cmol/kg, Mg = 2.62 cmol/kg)

compared with the rainforest (Ca = 9.42 cmol/kg, Mg = 2.07 cmol/kg). However, Na was higher and more variable in the rainforest soil samples (0.46 cmol/kg, CV = 372.32%) than in the highlands (0.22 cmol/kg, CV = 141.85%). Textural analysis revealed structural differences. Soils in the highlands contained more sand (51.82%) and less silt (30.15%) and clay (18.03%), whereas rainforest soils were richer in silt (39.72%) and clay (25.76%), and had a lower sand content (34.52%).

The high coefficient of variations, especially for EC, P, K, Mg, and Na, confirm the substantial heterogeneity of fertility conditions across and within both regions. These differences had direct implications for site-specific soil management and the tuning of predictive model hyperparameters to adequately capture the high variability in soil fertility-related properties across both agroecological zones.

Table 2. Physical and chemical variability of soil samples in highland and rainforest agroecosystems.

Variable	Highlands					Rainforest				
	Mean	Median	σ	Var	CV (%)	Mean	Median	σ	Var	CV (%)
pH	6.51	6.6	1.18	1.4	18.17	5.66	5.4	1	1	17.64
EC (dS/m)	7.76	5.22	14.72	216.78	189.71	3.51	3.06	2	4.01	56.99
OM (%)	4.14	3	3.17	10.07	76.64	2.44	2.3	1.3	1.69	53.3
P (mg/kg)	53.96	23.44	77.92	6071.67	144.41	22.96	17.22	23.35	545.04	101.69
K (mg/kg)	234.57	171.6	239.86	57,532.77	102.25	94.95	81.65	51.09	2610.54	53.81
Ca (cmol/kg)	19.82	13.63	17.15	294.04	86.5	9.42	7.51	8.58	73.6	91.09
Mg (cmol/kg)	2.62	2	2.28	5.18	86.82	2.07	1.17	3.13	9.82	151.16
Na (cmol/kg)	0.22	0.15	0.31	0.1	141.85	0.46	0.08	1.73	2.98	372.32
Sand (%)	51.82	51.52	15.2	231.04	29.33	34.52	31.8	15.52	240.77	44.95
Silt (%)	30.15	27.61	12.62	159.21	41.85	39.72	37.34	15	224.89	37.75
Clay (%)	18.03	18.16	9.77	95.54	54.2	25.76	25	9.77	95.46	37.93

Note: Descriptive statistics based on 297 soil samples where: EC = electrical conductivity; pH = hydrogen potential; OM = organic matter; P = phosphorus; K = potassium; Ca = exchangeable calcium content; Mg = exchangeable magnesium content; Na = exchangeable sodium content; var = variance; CV = coefficient of variation (%); σ = standard deviation.

Soil texture, a fundamental determinant of edaphic fertility and hydraulic properties, was quantitatively characterized through granulometric analysis of sand, silt, and clay fractions [63]. Figure 3a depicts the distribution of 297 soil samples within the USDA soil texture triangle classification framework [64], which delineates twelve distinct textural classes based on proportional particle size distribution. Analysis of the distribution revealed nine distinct textural classes present in the studied samples.

The analytical results reveal marked textural heterogeneity between the highlands soil samples (Figure 3a) and those from the rainforest region (Figure 3c). In the highlands, where agricultural plots cultivated with potatoes were evaluated, the predominant soil textures were sandy loam (SaLo), loam (Lo), and sandy clay loam (SaClLo), which accounted for the majority of the samples. These textures provide a favorable balance between hydraulic conductivity and moisture retention capacity, making them well suited for tuber cultivation under high-altitude conditions.

In the rainforest, where samples were collected from coffee and banana plots, finer textures such as clay loam (ClLo), silty clay loam (SiClLo), and clay (Cl) were more common. These soils typically exhibit high water and nutrient retention capacity, although with reduced permeability conditions that support the cultivation of perennial crops in humid tropical climates.

Extreme textural classes, such as sand (Sa) and silt (Si), were minimally represented in both regions, indicating a predominance of soils with intermediate physical properties.

Figure 3b displays representative photographs of each identified textural class, providing visual confirmation of the morphological characteristics associated with each soil type. This observed textural variability underscores the considerable edaphic complexity across the study area and highlights the importance of detailed pedological characterization to support regionally adapted and efficient agricultural management strategies.

Figure 3. Soil texture classification using the USDA soil texture triangle for two contrasting ecological regions in Peru: (**a**) Highlands (n = 168) and (**c**) Rainforest (n = 129). Each orange dot represents a soil sample positioned according to its percentage of sand, silt, and clay. Panel (**b**) displays photographic examples of representative soil samples corresponding to the most frequent textural classes identified in both regions, including their respective USDA texture classifications.

3.2. Normalized Weights of the Thematic Dimensions

Figure 4 shows the spectral signatures of agricultural soils from the rainforest and highland regions, processed using three spectral transformations: SG-smoothed reflectance (a and b), first derivative (c and d), and band depth (e and f). The SG spectral signatures exhibit an ascending pattern from the visible (VIS) region to the shortwave infrared (SWIR) region. In the highland soils (b), higher overall reflectance is observed, with more marked differences in the 600–1400 nm and 1900–2200 nm ranges. In contrast, the rainforest soils (a) show attenuated curves, particularly between 500 and 1000 nm, where reflectance is significantly lower. These differences highlight the spectral contrast between the two regions under different agricultural conditions.

Figure 4c,d, corresponding to the first derivative, reveal the rate of change along the spectral profile. In the highland soils (d), more pronounced inflection points are evident between 600–750 nm and 1900–2200 nm. In the rainforest soils (c), the most visible transitions occur between 550–700 nm and 1350–1900 nm, although with lower intensity, indicating a smoother spectral response. Figure 4e,f show the band depth (BD) transformation, which emphasizes specific absorption regions. In the highland soils (f), the deepest bands appear around 1400, 1900, and 2200 nm, with well-defined profiles. In the rainforest soils (e), bands are also observed in these regions, although they are broader and less defined, particularly between 1350–1550 nm and 1850–2100 nm. These absorption regions are relevant to the soil properties evaluated in this study. The features around 1400 and 1900 nm are commonly associated with water- and hydroxyl-related absorptions, which may influence the spectral response of properties such as OM, pH, Ca, Mg, and texture. Likewise, the spectral region near 2200 nm is usually linked to the fine mineral fraction of the soil, particularly clay-related responses, which helps explain its relevance for clay, silt, pH, and exchangeable **Ca**.

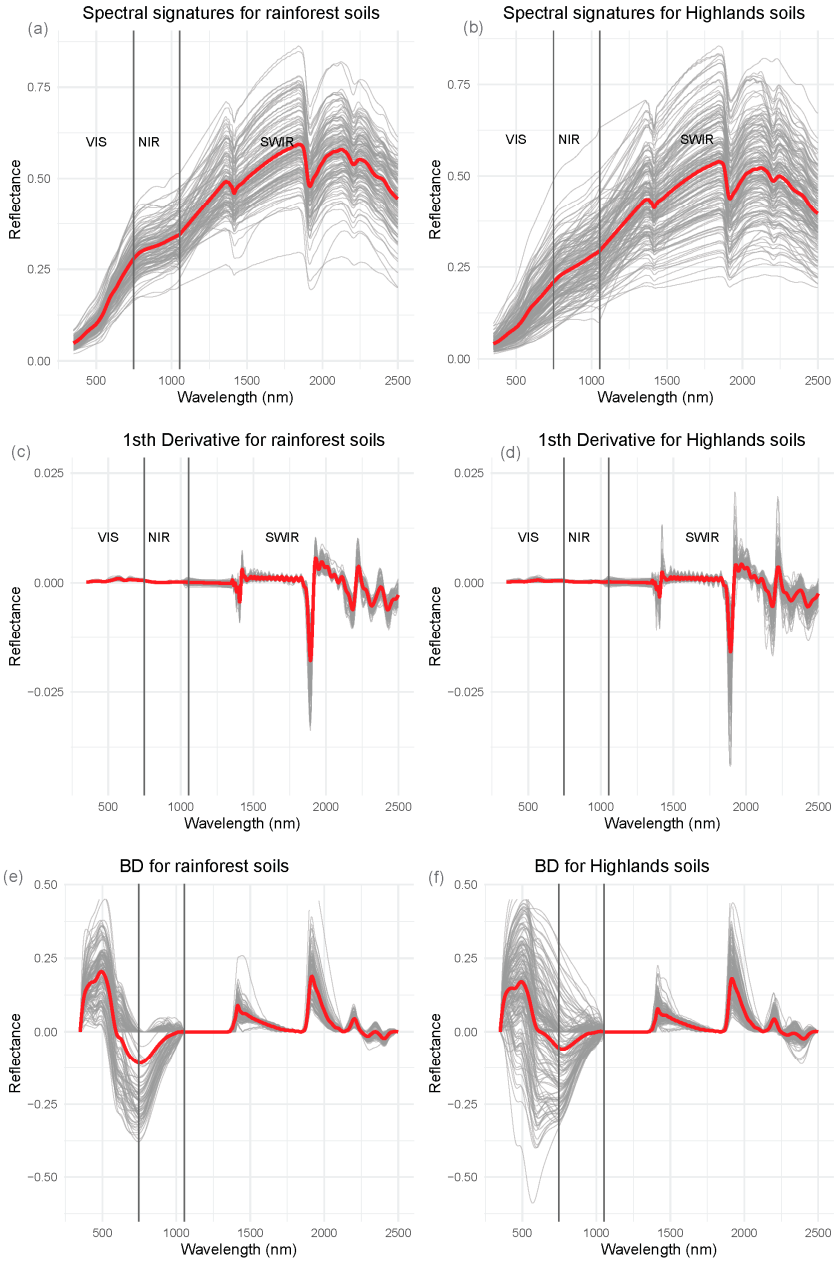


Figure 4. Spectral Data of soil samples; (**a**) Spectral signatures SG for rainforest soil; (**b**) Spectral signatures SG for Highlands soil; (**c**) 1st derived for rainforest soils; (**d**) 1st derived for Highlands soils (**e**) Band depth for rainforest soils; (**f**) Band depth for highlands soils. Red line: mean values.

Figure 5 presents the Pearson correlation coefficients between VNIR spectral bands and soil physicochemical properties, showing clear differences between soils from the highlands and the rainforest. In the case of the highlands (Figure 5a,c,e), a more consistent and defined spectral response is observed. In Figure 4a, which corresponds to the SG method, clay shows consistently positive correlations across the spectrum, exceeding $r > 0.25$ in the 750 to 1000 nm range.

Organic matter (OM) and silt also exhibit moderate correlations with well-defined spectral patterns, particularly between 620 and 820 nm. The application of the first derivative (D1), shown in Figure 4e, significantly enhances spectral sensitivity, with sharper correlation peaks. Clay and sand reach values of $r = 0.42$ and $r = 0.39$, respectively, between 680 and 720 nm, while OM reaches $r = 0.38$ within the 500 to 520 nm range.

Other properties such as electrical conductivity (EC) and pH also improve moderately, with r = 0.19 and r = 0.32. The band depth (BD) method, shown in Figure 4c, produces smoother patterns but still identifies key absorption zones. Clay remains the most responsive variable, with r > 0.12 between 450 and 700 nm. Overall, these results indicate greater spectral sensitivity in highland soils, especially for textural properties and organic matter.

In contrast, Figure 5b,d,f present the results for rainforest soils, where spectral correlations are generally weaker. In Figure 5b (SG), although clay retains a positive trend, correlations are less consistent and lower compared with the Highlands. Properties such as OM, K, and Na show diffuse spectral responses with no clear structure. The D1 transformation (Figure 5f) slightly improves correlations for some properties like clay, sand, and OM, but the peaks are weaker and more variable.

The BD method (Figure 5d) shows even smoother correlation patterns, with moderate responses for clay mainly between 1400 and 1900 nm, while other properties display weak or scattered signals. Overall, the spectral response in rainforest soils appears less pronounced. This may be due to higher moisture content, the presence of surface organic residues, or lower textural variability, which reduce the spectral discrimination capacity compared with highland soils.

Figure 6 shows Pearson correlation heatmaps between soil physicochemical properties and normalized spectral indices derived from all possible wavelength combinations within the VNIR-SWIR range (500–2500 nm), calculated from SG smoothed reflectance spectra (I-SG) [65]. In both regions, correlation coefficients range from −0.6 to 0.6, enabling the identification of spectral regions with greater sensitivity to specific soil variables. In the highlands (Figure 6a), stronger correlations were observed for OM and Sand, with prominent associations in the 1350–1850 nm range for OM and the 500–700 nm and 1800–2000 nm ranges for Sand. Clay exhibited consistently negative correlations between 500–700 nm, while Silt showed moderate positive correlations in the mid-spectral range. Among chemical properties, pH and Ca displayed notable correlations, especially around 1400 and 2200 nm. In contrast, P showed low to moderate correlations, mostly between 1400–1600 nm, without a clear spectral pattern, whereas K presented very weak, scattered, and inconsistent correlations.

In the rainforest soil samples (Figure 6b), distinct patterns were observed, with stronger correlations for Silt, pH, Ca, Mg, and P, primarily between 1400–1550 nm and 1900–2200 nm. Silt exhibited the highest coefficients (>0.6) in the 1400–1600 nm range, while pH showed strong positive associations between 1900–2200 nm. Ca and Mg showed mixed spectral responses around 1400 and 2200 nm. In contrast, variables such as OM, K, Na, and Sand recorded weak correlations with no defined patterns. Overall, the results demonstrate that SG-based normalized indices effectively identify spectral combinations relevant for predicting soil properties, although clear regional differences are evident: in the highland region, associations are stronger for textural variables (OM, Sand, Clay), whereas in the rainforest, spectral responses are more pronounced for chemical properties (pH, Ca, Mg, P) and Silt.

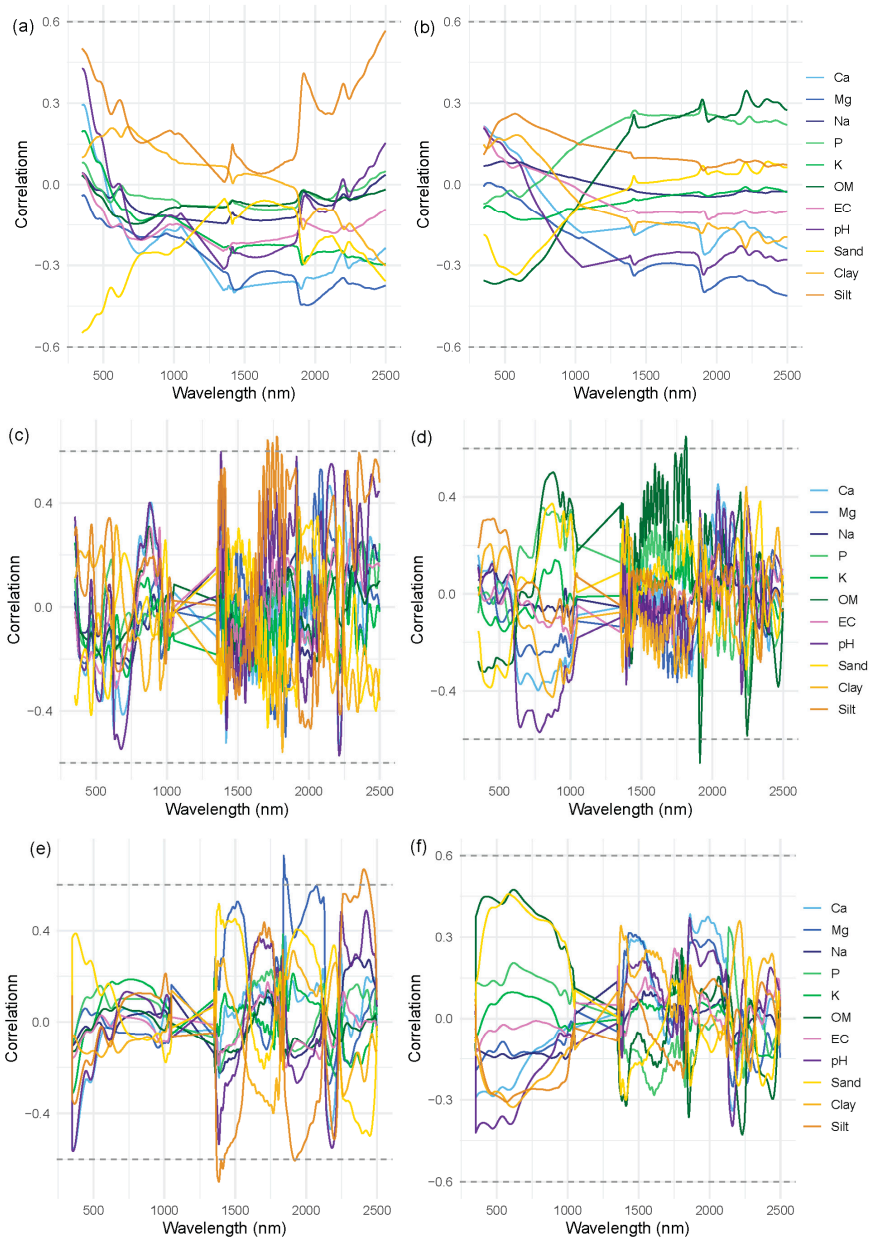


Figure 5. Pearson correlation coefficients between VNIR spectral wavelengths (350 to 2500 nm) and soil physicochemical properties in the Highlands (**a**,**c**,**e**) and rainforest (**b**,**d**,**f**) using three spectral preprocessing techniques: (**a**,**b**) Savitzky–Golay reflectance (SG), (**c**,**d**) Band Depth transformation (BD), and (**e**,**f**) First Derivative (D1).

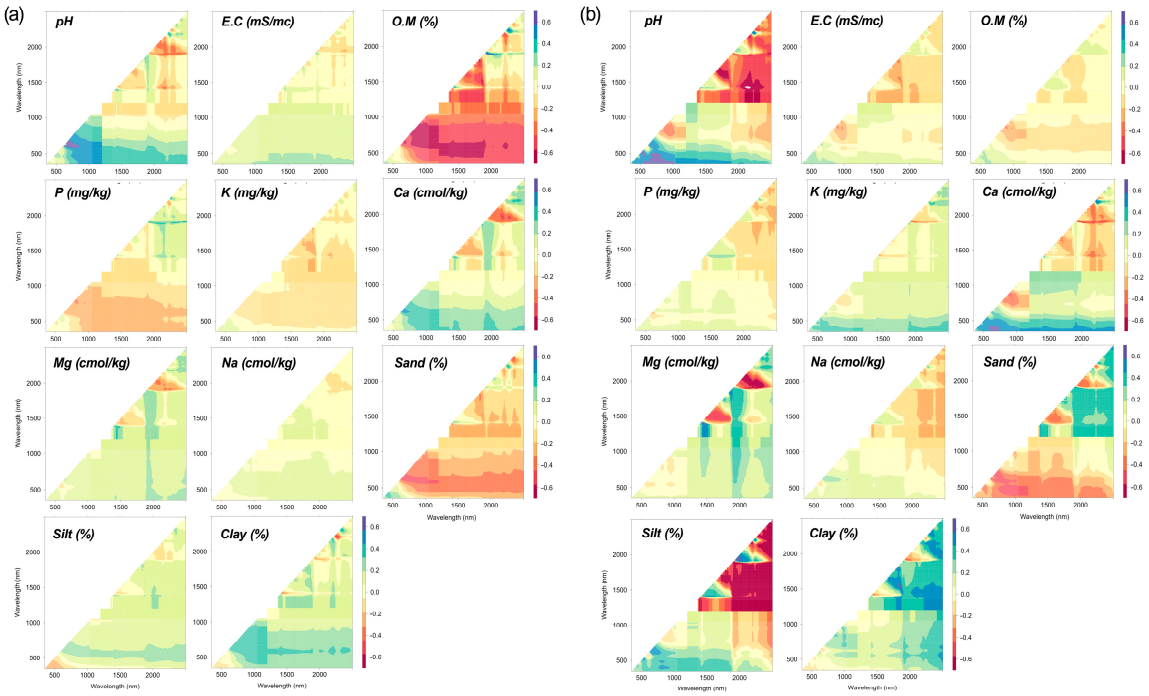


Figure 6. Pearson correlation heatmaps between Savitzky–Golay (SG) smoothed visible-near infrared (Vis-NIR) spectral reflectance and selected soil properties for 168 samples from the highlands (**a**) and 129 samples from the rainforest (**b**). Each panel shows the correlation (from −0.6 to 0.6) between normalized index based on spectral reflectance according to Equation (2), and individual soil attributes (e.g., pH, electrical conductivity, organic matter, macro- and micronutrients, and texture components), highlighting wavelength regions with higher predictive potential.

3.3. Spectral Modelling Results Using Machine Learning

Machine learning models applied to the highlands and rainforest dataset derived from I-SG, BD, and 1D showed variable performance depending on the soil property, with marked differences between training and validation phases. For this analysis, the best-performing results for each variable were selected based on the spectral input type, algorithm, and hyperparameters that showed the best fit according to the evaluated metrics, while datasets derived from indices calculated using 1D and BD spectra were excluded due to their poor performance. These results are summarized in Table 3.

Highlands. For pH, the NN model using I-SG showed good performance during training (R^2 = 0.80, RPD = 2.25), although its predictive ability decreased in validation (R^2 = 0.47, RPD = 0.89), indicating limited generalization. For OM, also modeled with NN and SG_IND, the model achieved robust performance in both phases (R^2 = 0.83 and 0.72; RPD = 2.41 and 1.74, respectively), reflecting greater predictive stability. Regarding textural components, the prediction of Clay using RF and I-SG achieved high accuracy during training (R^2 = 0.91, RPD = 2.52), but declined in testing (R^2 = 0.40, RPD = 1.28), suggesting overfitting or greater variability in unseen data. Sand, modeled with SVM and D1, showed moderate performance (R^2 = 0.72 in training; 0.55 in testing), while Silt yielded weaker results, especially in validation (R^2 = 0.17, RPD = 1.09).

For exchangeable cations, Mg modeled with SVM and BD showed excellent fit in training (R^2 = 0.93, RMSE = 3.09), but moderate validation performance (R^2 = 0.38, RPD = 1.26). Na, modeled with RF and I-SG, performed well in training (R^2 = 0.93, RPD = 1.99), but its predictive power dropped sharply in validation (R^2 = 0.017, RPD = 0.47). Ca, modeled with PLSR and SG, showed acceptable and consistent results (RPD = 1.69 in training and 1.43 in validation).

The prediction of available nutrients was particularly challenging. For P, the NN model with D1 showed low R^2 values in both phases (0.20 and 0.22), with RPD just above 1, confirming the difficulty of modeling this variable. For K, modeled with RF and BD, overfitting was evident: R^2 = 0.92 in training dropped to 0.13 in validation (RPD = 1.03), indicating low robustness.

The models applied to the rainforest soil samples exhibited highly variable performance depending on the soil property and model phase. As with the highlands, Table 3 presents the best combinations of spectral input, algorithm, and parameters for each variable, selected based on the best fit according to the evaluated metrics. For pH, the NN model using BD performed well during training (R^2 = 0.75, RPD = 2.03) and showed acceptable generalization in validation (R^2 = 0.62, RPD = 1.60), indicating stability across phases. In contrast, the EC model—PLSR with SG preprocessing—exhibited low predictive capacity (R^2 = 0.24, RPD = 1.11) in both training and validation phases, limiting its applicability for this variable under rainforest conditions.

Predictions for OM using NN and I-SG were poor, with R^2 values below 0.16 and RPD values barely above 1 in both phases, reflecting limited ability to capture this variable in rainforest soils. For available nutrients, predictions for P and K followed a common pattern: high R^2 in training (0.89 and 0.34, respectively), but poor generalization in validation (R^2 = 0.40 and 0.23), with RPD ≤ 1.17. This highlights the difficulty of modeling mobile nutrients, likely due to their high variability and weak spectral expression. Regarding exchangeable cations, Mg, modeled with RF and D1, showed strong overfitting (R^2 = 0.96 in training vs. 0.28 in validation), while Na, modeled with NN and I-SG, was unstable (R^2 = 0.67 in training and 0.18 in testing). In contrast, Ca showed solid results in both phases with SVM (R^2 = 0.79, RPD = 3.52 in validation), indicating strong predictive potential in rainforest soils.

For textural properties, sand was best modeled with RF and BD, achieving high accuracy in training (R^2 = 0.93), but a sharp decline in validation (R^2 = 0.38, RPD = 1.29). Silt, modeled with NN and BD, showed relatively consistent performance (R^2 = 0.71 in training and 0.67 in testing), with RPD values above 1.6. For Clay, the PLSR model with D1 also showed acceptable stability (R^2 = 0.56, RPD = 1.46 in validation).

Machine learning models showed better predictive performance for multiple soil properties, although this pattern was not consistent across all variables. In the Highlands, NN and SVM achieved better results for OM, pH, and sand, whereas PLSR showed more stable performance for Ca. For rainforest soils, machine learning models performed better for pH, P, Mg, Ca, and silt, while PLSR achieved the best results for EC, K, and clay. This behavior suggests that machine learning models were more effective in capturing complex and non-linear relationships for some soil properties, whereas PLSR retained predictive capacity for variables with more stable or linearly structured spectral responses.

Table 3. Predictive performance of ML models using spectral data for soil property estimation in the highlands and rainforest.

		Highlands		Train			Test		
Input	**Variable**	**Model**	**Hyperparameters**	**R^2**	**RMSE**	**RPD**	**R^2**	**RMSE**	**RPD**
I-SG	pH	NN	size = 1, decay = 0.1	**0.8004**	**0.5227**	**2.2472**	**0.4676**	**0.8912**	**1.3642**
1D	EC	SVM	C = 1, gamma = 0.0375	0.5998	14.8001	1.154	0.0476	5.4214	1.0003
I-SG	OM	NN	size = 1, decay = 0.1	**0.8275**	**1.3439**	**2.4117**	**0.723**	**1.7351**	**1.7486**
1D	P	NN	size = 1, decay = 0.0006	0.2017	70.4259	1.1238	0.2193	66.3275	1.139
BD	K	RF	mtry = 2	0.9227	136.5559	1.9449	0.1326	154.3905	1.0266
BD	Mg	SVM	C = 8, gamma = 0.0325	**0.9288**	**0.7768**	**3.0938**	**0.3762**	**1.5414**	**1.2596**
I-SG	Na	RF	mtry = 2	0.9265	0.1057	1.9905	0.0174	0.4693	1.0182
SG	Ca	PLSR	ncomp = 14	**0.6491**	**9.4937**	**1.6953**	**0.5248**	**13.7029**	**1.4386**

1D	Sand	SVM	C = 1, gamma = 0.0320	**0.7212**	**8.534**	**1.8021**	**0.5511**	**10.0094**	**1.4886**
I-SG	Silt	SVM	C = 1, gamma = 0.0362	0.5573	9.2287	1.4492	0.165	9.6598	1.0938
I-SG	Clay	RF	mtry = 2	**0.9101**	**3.721**	**2.5241**	**0.4034**	**8.4303**	**1.2779**
		Rainforest			**Train**			**Test**	
Input	**Variable**	**Model**	**Hyperparameters**	**R^2**	**RMSE**	**RPD**	**R^2**	**RMSE**	**RPD**
BD	pH	NN	size = 1, decay = 0.1	**0.7542**	**0.5127**	**2.0255**	**0.6206**	**0.568**	**1.5969**
SG	EC	PLSR	ncomp = 9	0.3643	1.6901	1.2611	0.2374	1.4789	1.113
I-SG	OM	NN	size = 1, decay = 0.0422	0.1555	1.148	1.0941	0.1203	1.3137	1.079
I-SG	P	RF	mtry = 2	0.8865	7.9101	2.1428	0.3987	29.8742	1.1673
I-SG	K	PLSR	ncomp = 6	0.3385	41.551	1.2362	0.231	45.9359	1.1124
1D	Mg	RF	mtry = 2	0.9596	1.3951	2.3876	0.2761	2.2272	1.1651
I-SG	Na	NN	size = 1, decay = 0.0422	0.6756	0.9826	1.7653	0.1801	2.1514	0.8044
1D	Ca	SVM	C = 2, gamma = 0.0297	**0.8234**	**3.9176**	**2.2886**	**0.7921**	**3.5178**	**2.1577**
BD	Sand	RF	mtry = 20	0.9279	4.9101	3.0053	0.381	13.6315	1.2864
BD	Silt	NN	mtry = 2	**0.7111**	**8.1583**	**1.8705**	**0.6666**	**8.8376**	**1.6409**
1D	Clay	PLSR	ncomp = 18	**0.788**	**4.5914**	**2.1834**	**0.5552**	**6.3181**	**1.4587**

Note: The table presents the models, input data types, hyperparameter adjustments, and performance metrics (R^2, RMSE, RPD) for both training and testing phases. Models include RF = Random Forest; SVM = Support Vector Machine; NN = Neural Network; BD = Band Depth feature; 1D = First Derivative; SG = Savitzky–Golay; I-SG = Indices calculated from Savitzky–Golay. Bold values indicate the best-performing configuration selected for acceptable predicted soil property based on cross-validation results.

To further evaluate prediction accuracy beyond R^2 and RPD, the distribution of the Symmetric Percent Difference (SPD) between observed and predicted values was examined for each soil property and ecosystem (Figure 7). Narrower violin shapes centered near zero indicate lower prediction error and greater model precision, whereas wide or asymmetric distributions reflect high variability and systematic bias. In the Highlands, Ca, Mg, OM, and pH exhibited relatively compact distributions centered around 0%, consistent with their moderate-to-good validation metrics. In contrast, Na, K, and P showed markedly wide and skewed distributions, confirming the overfitting and low generalization capacity observed in their R^2 and RPD values. EC and Silt also displayed broad spreads, corroborating their weak predictive performance. For the Rainforest, Ca stood out with a narrow, symmetric distribution, in agreement with its strong SVM validation results (R^2 = 0.79, RPD = 2.16). Silt and pH also showed relatively stable SPD distributions, while Na, Mg, and Sand exhibited pronounced spread and asymmetry, reflecting the sharp training-to-validation performance drops reported for those variables. Overall, the SPD distributions reinforce the quantitative metrics in Table 3 and provide a complementary visual assessment of prediction reliability across soil properties and ecosystems.

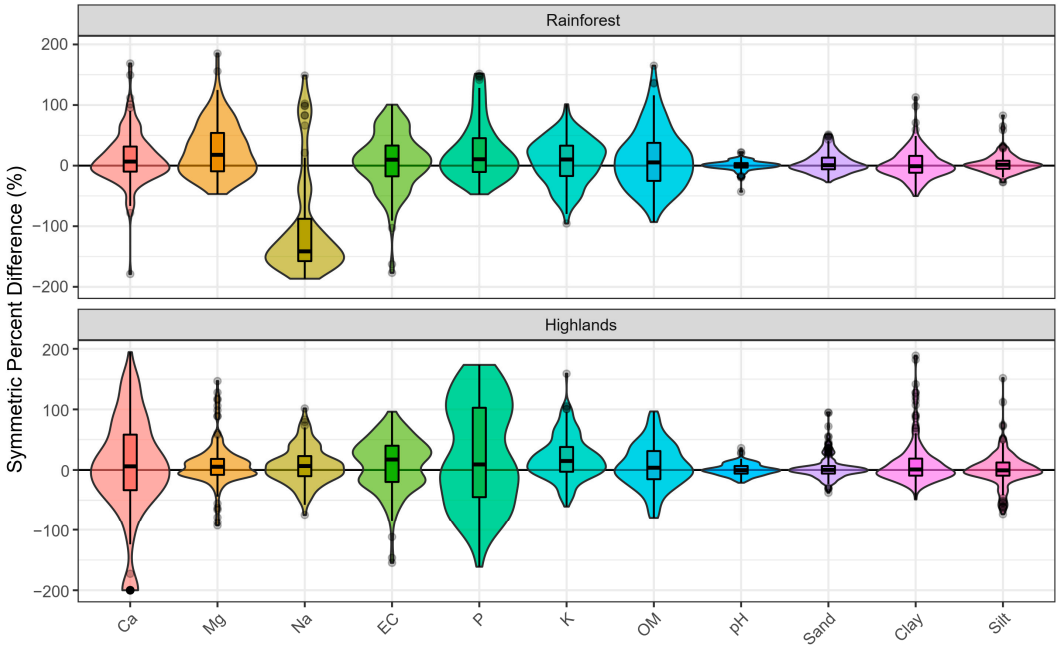


Figure 7. Violin plots of the Symmetric Percent Difference (SPD, %) between observed and predicted soil property values for the Rainforest (**top**) and Highlands (**bottom**). Boxes represent the interquartile range; horizontal lines indicate the median. Variables are ordered by exchangeable cations, available nutrients, organic matter, pH, and texture.

3.4. Fertilizer Recommendations

The agronomic analyses were performed using all samples from each agroecosystem, comparing observed laboratory measurements with soil property values predicted by the final fitted models. The comparison between laboratory-derived (Observed) and spectrally predicted (Predicted) fertilizer-equivalent nutrient availability revealed consistent regional patterns with nutrient-specific differences in model performance (Figure 8). In Rainforest soils, N (approximately 20 kg ha^{-1}), P_2O_5 (approximately 50 kg ha^{-1}), and K_2O (250–300 kg ha^{-1}) showed close alignment between Observed and Predicted medians and interquartile ranges. Minor attenuation of upper-tail values occurred for CaO (up to 20,000 kg ha^{-1}) and MgO (12,000 kg ha^{-1}), without altering the general nutrient ranking.

Highlands soils exhibited substantially greater dispersion, particularly for CaO (observed extremes reaching 1×10^5 kg ha^{-1}) and MgO (10,000 kg ha^{-1}). Predicted values reproduced central tendencies well but considerably compressed variability at higher concentrations. Agreement was strongest for N and P_2O_5, moderate for K_2O, and weakest for CaO and MgO in base-saturated Highland soils.

Notably, regional contrasts between Highlands and Rainforest were consistently larger than discrepancies between Observed and Predicted values within each zone, indicating that agroecological differentiation exerts greater influence on fertilizer-equivalent nutrient availability than prediction uncertainty.

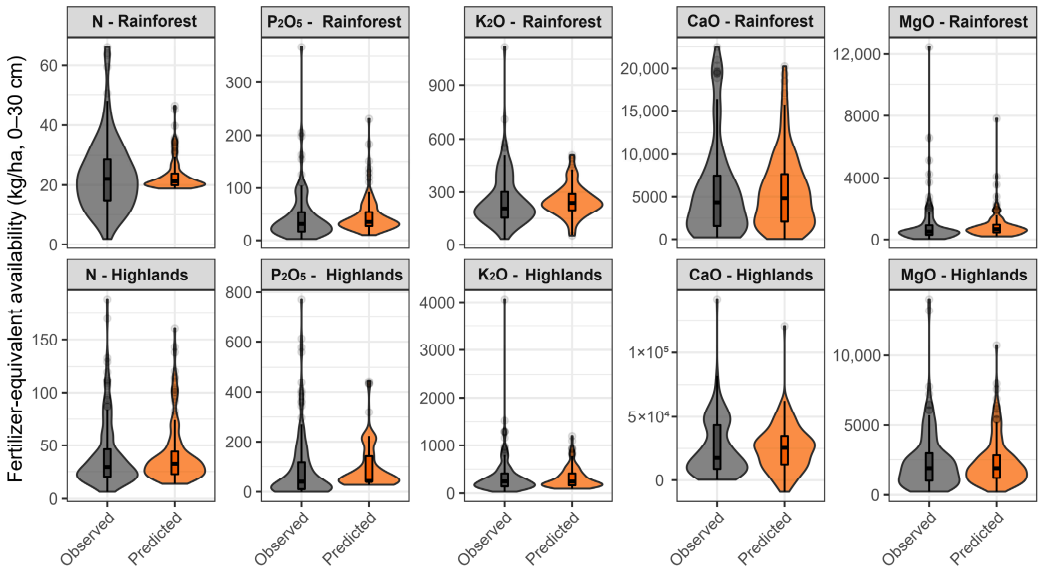


Figure 8. Distribution of fertilizer-equivalent nutrient availability (kg ha^{-1}, 0–30 cm) derived from observed laboratory measurements and soil property values predicted by the final fitted models for Highlands (n = 168) and Rainforest (n = 129) agroecosystems. Grey and orange violins represent observed and predicted values, respectively. Embedded boxplots indicate the median, interquartile range, and variability, while black points represent outliers.

The efficiency-adjusted fertilizer recommendations revealed clear regional contrasts and provided an agronomically interpretable comparison between Observed and Predicted soil information (Figure 9).

In Rainforest soils, N recommendations showed the closest alignment, with both distributions centered near 250 kg ha^{-1} and comparable dispersion. This consistency is supported by the acceptable OM prediction performance in Rainforest soils (R^2 = 0.72, RPD = 1.75 in Highlands), which underpins N mineralization estimates. For P_2O_5 and K_2O, Predicted doses exhibited markedly reduced variability—K_2O Predicted values collapsed near zero against Observed ranges reaching 100 kg ha^{-1}—consistent with the weak spectral models for these nutrients (K: R^2 = 0.23, RPD = 1.11; P: R^2 = 0.40, RPD = 1.17 in validation). CaO and MgO recommendations were negligible in both Observed and Predicted distributions, reflecting the lower base saturation typical of Rainforest soils and minimal Ca and Mg deficits relative to crop requirements.

In Highland soils, N again presented the strongest agreement between Observed and Predicted recommendations, with overlapping interquartile ranges and medians near 150–200 kg ha^{-1}, suggesting that OM prediction accuracy is sufficient for operational N fertilizer planning. P_2O_5 showed moderate alignment, though Predicted values exhibited a more constrained distribution. K_2O recommendations were consistently underpredicted, with Predicted values concentrated near zero (R^2 = 0.13, RPD = 1.03), limiting their practical utility. CaO exhibited the most critical divergence: Observed values reached extreme upper tails of approximately 20,000 kg ha^{-1}, while Predicted estimates remained near zero. Although Ca prediction showed acceptable validation metrics (R^2 = 0.52, RPD = 1.44), the model failed to reproduce the highest concentrations, directly compromising liming recommendations in base-saturated Highland soils.

Overall, regional contrasts between Highlands and Rainforest remained the dominant source of variation in recommended doses, exceeding Observed–Predicted discrepancies within each zone. These findings suggest that spectrally derived soil properties can support agronomically consistent recommendations for N across both environments, while caution is warranted for K_2O, P_2O_5, and particularly CaO in Highland systems.

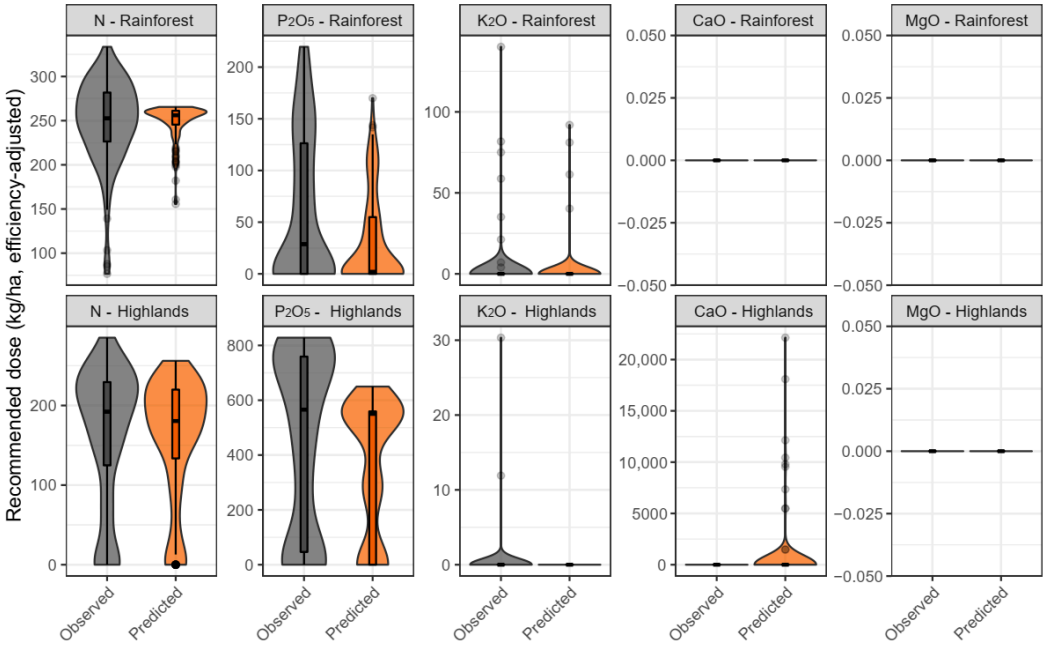


Figure 9. Recommended fertilizer doses (kg ha^{-1}, efficiency-adjusted) for Highlands (n = 168) and Rainforest (n = 129) soils based on observed laboratory data and soil property values predicted by the final fitted models. Grey and orange violins represent observed-based and predicted-based fertilizer doses, respectively. Embedded boxplots indicate the median, interquartile range, and variability, while black points represent outliers.

4. Discussion

Soil spectral libraries (SSLs) are a key component for building reproducible and scalable calibration models for rapid, cost-effective, and non-destructive estimation of soil properties [66]. Studies such as those by [67] and [68] highlight the potential of global libraries, such as the Open Soil Spectral Library (OSSL), and the use of mid-infrared (MIR) spectra to generate robust and reproducible models applicable across diverse agroecological settings. Moreover, hybrid approaches like those of [69], which combine VNIR, XRF, and LIBS, extend the capabilities of SSLs by adapting to tropical soils and variable instrumental conditions. However, despite these advances, the predictive capacity of SSLs depends critically on the representativeness and diversity of the reference datasets [70]. The need to harmonize reference methods and scanning protocols across laboratories may further restrict the interoperability of SSLs [66]. Nonetheless, the availability of open-access resources like the OSSL helps mitigate initial costs and enables countries with limited infrastructure to benefit from standardized, reusable calibration models [67].

The understanding of soil physicochemical properties related to fertility is crucial for effective crop management and yield optimization. While conventional soil analysis methods provide accurate results, they require specialized laboratory equipment that are both laborious and time-consuming, creating significant delays in implementing effective farming practices. Our findings align with growing evidence supporting Vis-NIR spectroscopy as a cost-effective alternative with reduced resource consumption. The integration of hyperspectral data with advances in artificial intelligence and machine learning (ML) frameworks has enabled the development of predictive models for soil fertility properties that can be effectively shared across diverse agricultural settings. These results complement the findings of studies that demonstrate data-driven approaches enhance the understanding of complex environmental systems, improve reproducibility, and facilitate model transferability across different spatial and temporal contexts.

Our study extends this framework specifically to soil fertility assessment, providing farmers with more accessible and timely information for small agricultural decision-making. A critical methodological consideration was the management of hyperspectral data, which typically comprise hundreds of spectral responses, not all correlating with soil properties of interest. Furthermore, scatter effects and baseline drift can introduce considerable noise into measurements. To address these challenges, preprocessing techniques—including SG filtering, which is widely adopted in spectroscopic studies for its effectiveness in smoothing and differentiating spectral data were employed [71,72]. Also, the implementation of first derivative transformations of the reflectance spectrum with respect to wavelength helped in removing slow variations due to background reflectance and highlighted subtle spectral features associated with key soil attributes such as texture and organic matter [37,73]. These preprocessing steps enhanced separation between similar spectral signatures and reduced multicollinearity, ultimately improving model performance. The transformation of band depth analysis to isolate key absorption regions within the Vis-NIR SWIR range facilitated the differentiation of soil types by amplifying spectral signals related to compositional properties and reducing interference from external factors [38,74]. The reflectance spectra showed systematic variation across soil types, with higher reflectance observed in sandy soils and lower values in samples with higher clay or organic matter contents—attributable to greater light absorption [75,76].

Soil sample analysis revealed a heterogeneous distribution of fertility properties across the central Peruvian highlands and rainforest basin. As shown in Table 2, key soil properties such as EC, P, K, and Mg exhibited coefficients of variation exceeding 100%. This marked variability is consistent with the contrasting agroecological settings and physiographic complexity of the study areas, in agreement with previous reports from topographically diverse regions [77]. The results show that various soil properties achieved high predictive performance, underscoring their relevance for the development of regional spectral libraries. In the highlands, key variables included organic matter (OM) (R^2 = 0.723, RPD = 1.73), pH (RPD = 0.89), calcium (RPD = 13.70), as well as sand (RPD = 10.01) and clay fractions (RPD = 8.43). In the rainforest region, strong results were also obtained for pH (R^2 = 0.6206), calcium (R^2 = 0.7921, RPD = 3.52), silt (RPD = 8.84), and clay (RPD = 6.32). These variables, due to their spectral stability and generalizability, are fundamental components for building functional libraries adapted to contrasting agroecological contexts. Likewise, certain properties such as electrical conductivity and mobile nutrients (e.g., P and Na) consistently showed low prediction accuracy. This limitation is attributed to the lack of distinctive spectral features for these nutrients within the Vis-NIR range, their occurrence in spectrally inactive forms, or their adsorption onto dominant soil components such as clay and organic matter [78]. These conditions hinder reliable signal detection and compromise calibration model accuracy.

Although this study focused on well-established models in soil spectroscopy (PLSR, RF, SVM, and NN), their selection responded to the objective of comparing linear and non-linear approaches within a robust and reproducible methodological framework suitable for a moderate-sized dataset. While recent reviews suggest that more advanced approaches, such as gradient boosting or one-dimensional convolutional neural networks, may offer advantages under specific conditions, these methods involve greater tuning complexity and, in the case of deep architectures, often require larger calibration datasets to ensure adequate generalization [79]. For this reason, the neural network was implemented with a single hidden layer as a parsimonious architecture capable of capturing relevant non-linear relationships without introducing unnecessary complexity. Gradient boosting represents a promising alternative that could be explored in future work [80].

In addition to model selection, the marked differences between training and validation performance for variables such as P, K, and Na indicate that overfitting remained an

important limitation for some models, particularly for mobile nutrients with weak or indirect spectral expression. Although validation was performed through independent 5-fold cross-validation within each agroecosystem-specific dataset, the observed drops in performance confirm that good calibration fit did not always translate into stable generalization. This pattern highlights the need to expand calibration datasets and to explore more robust strategies to improve model generalization in future studies, especially for nutrients with limited spectral sensitivity.

The integration of spectral data from two sensors with non-overlapping ranges resulted in a spectral gap between 1000 and 1350 nm. This discontinuity, together with differences in spectral resolution, measurement conditions, and acquisition geometry between instruments, may have limited the capture of intermediate spectral transitions and introduced additional inter-sensor variability in the fused dataset [81,82]. Nevertheless, the spectral regions preserved in the visible range and in the SWIR domain above 1350 nm showed relevant responses for OM, texture, pH, Ca, and Mg, indicating that the combined approach retained useful predictive information for several soil fertility attributes. These results suggest that sensor fusion is a viable strategy, although future studies should further investigate instrument harmonization and evaluate the specific effects of spectral discontinuities on model robustness and transferability [83].

Machine learning approaches such as RF, SVM, and NNs have shown strong potential in soil spectroscopy for modeling complex and non-linear relationships between spectral responses and soil attributes [84]. In the present study, these models performed well for several soil properties, although their superiority over PLSR was not consistent across all variables.

Beyond prediction accuracy, the agronomic relevance of Vis-NIR models depends on whether prediction uncertainty meaningfully alters fertilizer recommendations. By integrating observed and predicted soil properties into a structured nutrient balance framework, we assessed how model-derived differences propagate into fertilizer dose calculations across contrasting ecosystems, extending conventional evaluation beyond statistical metrics toward a more operational measure of predictive utility [26].

Fertilizer recommendations derived from predicted soil properties represent a practical application of Vis-NIR spectroscopy. Previous studies have shown that spectral soil analysis can be linked to agronomic recommendation frameworks [85], but also that uncertainty in predicted nutrient concentrations, particularly for P and K, can influence recommendation outcomes and their expected performance [86]. In the present study, fertilizer recommendations were estimated using predicted values as point inputs, so the extent to which uncertainty associated with those predictions may have influenced the final recommended doses was not explicitly evaluated. In this sense, future studies should examine more closely how prediction error may be transferred to the final agronomic recommendation by incorporating sensitivity analyses or formal uncertainty propagation approaches [87].

For stable and well-predicted variables, notably OM and texture, fertilizer-equivalent N supply remained largely consistent between datasets, with N recommendations showing strong agreement across both ecosystems. However, even variables with acceptable statistical performance produced agronomically consequential errors when cumulative effects of unit conversions, bulk density estimation, and soil-specific availability corrections amplified property-level uncertainties into recommendation errors. This was most critical for CaO in Highland soils, where predicted values collapsed near zero against observed extremes reaching 20,000 kg ha^{-1}. Vis-NIR spectroscopy has been shown to struggle with accurate prediction of total and exchangeable calcium, potassium, phosphorus, and total magnesium [88], consistent with our findings for these nutrients.

Improving performance for these challenging variables may require two complementary strategies. First, expanding the calibration dataset with a larger number of samples representing the full range of soil variability in these ecosystems, as larger calibration sets have been shown to substantially reduce prediction error for mobile nutrients. Second, exploring mid-infrared FTIR spectroscopy as a complementary or alternative technology: MIR generally produces better predictions than Vis-NIR for exchangeable cations and available phosphorus [89], and FTIR-PAS has shown good performance specifically for soil available phosphorus predictions [90]. These results suggest that Vis-NIR currently supports reliable N and OM-based fertility assessments, while liming and phosphorus recommendations in base-saturated or highly weathered soils require complementary analytical strategies.

5. Conclusions

This study showed that Vis-NIR spectroscopy captured relevant spectral responses associated with several soil fertility-related and textural properties across contrasting agroecosystems in the Andean Highlands and central Rainforest. These responses were mainly associated with organic matter, pH, exchangeable calcium, and textural fractions, and varied between agroecological regions, which could reflect the influence of ecological context on spectral behavior and model performance.

Regarding predictive accuracy, the results demonstrate that model performance depended on both the soil property and the agroecosystem. The most reliable predictions were obtained for OM in the Highlands, Ca in the Rainforest, and for selected texture fractions and pH in both regions, whereas mobile nutrients such as P, K, and Na showed limited predictive accuracy and poor generalization. No single model or preprocessing strategy was consistently superior across all variables; instead, performance varied according to the specific soil attribute being predicted.

When spectrally predicted soil properties were integrated into the nutrient balance framework, nitrogen recommendations showed the strongest agreement between observed and predicted values across both ecosystems, indicating that Vis-NIR can support operational N management under contrasting agroecological conditions. In contrast, predictions for K_2O and P_2O_5 were less robust, and CaO recommendations in Highland soils were severely underestimated, demonstrating that acceptable statistical performance at the property level does not necessarily guarantee agronomic reliability. Overall, these findings indicate that Vis-NIR spectroscopy is a promising and practical tool for rapid assessment of stable soil properties and N-related fertility management, but complementary analytical strategies are still required for agronomically critical mobile nutrients and liming-related recommendations.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/rs18091331/s1, Table S1: Crop nutrient extraction coefficients (kg element t^{-1} grain); Table S2: Stoichiometric conversion factors to fertilizer-equivalent forms; Table S3: Soil nutrient availability correction factors applied in Highland soils; Table S4: Texture-pH dependent nutrient availability matrix applied in Rainforest soils; Table S5: Fertilizer-use efficiency coefficients applied in final dose calculation; Figure S1: Selected spectral bands for each soil property across preprocessing transformations, input types, and agroecological zones; Figure S2: Scatter plots of observed versus predicted values for the best-performing model of each soil property in the Highlands and Rainforest zones.

Author Contributions: Conceptualization, S.P. (Samuel Pizarro), O.E.A.-A., and H.L.; methodology, S.P. (Solanch Patricio) and H.L.; software, D.C. (Duglas Contreras); validation; formal analysis, S.P. (Samuel Pizarro), D.C. (Dennis Ccopi), and K.O.; investigation, D.C. (Duglas Contreras), K.O., and

D.C. (Deyvis Cano); resources; data curation; writing—original draft preparation, S.P. (Samuel Pizarro) and D.C. (Dennis Ccopi); writing—review and editing, D.C. (Dennis Ccopi), K.O., J.Ñ., O.E.A.-A., S.P. (Solanch Patricio), and H.L.; visualization, S.P. (Samuel Pizarro), D.C. (Duglas Contreras), and D.C. (Deyvis Cano); supervision, H.L.; project administration and funding acquisition, O.E.A.-A., S.P. (Samuel Pizarro), and H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the INIA project "Mejoramiento de los servicios de investigación y transferencia tecnológica en el manejo y recuperación de suelos agrícolas degradados y aguas para riego en la pequeña y mediana agricultura en los departamentos de Lima, Áncash, San Martín, Cajamarca, Lambayeque, Junín, Ayacucho, Arequipa, Puno y Ucayali" with CUI N°2487112 of the Ministry of Agrarian Development and Irrigation (MIDAGRI) of the Peruvian Government.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable: This manuscript does not include human or animal research.

Data Availability Statement: The data that support the findings of this study are available on request from the corresponding author.

Conflicts of 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.

References

1. Cotler, H.; Corona, J.A.; Mauricio Galeana-Pizaña, J. Erosión de Suelos y Carencia Alimentaria en México: Una Primera Aproximación. *Investig. Geogr.* **2020**. https://doi.org/10.14350/RIG.59976.
2. Tomczyk, P.; Wdowczyk, A.; Wiatkowska, B.; Szymańska-Pulikowska, A.; Kuriqi, A. Fertility and Quality of Arable Soils in Poland: Spatial–Temporal Analysis of Long-Term Monitoring. *Ecol. Indic.* **2024**, *166*, 112375. https://doi.org/10.1016/J.ECOLIND.2024.112375.
3. Dupla, X.; Claustre, R.; Bonvin, E.; Graf, I.; Le Bayon, R.C.; Grand, S. Let the Dust Settle: Impact of Enhanced Rock Weathering on Soil Biological, Physical, and Geochemical Fertility. *Sci. Total Environ.* **2024**, *954*, 176297. https://doi.org/10.1016/J.SCITOTENV.2024.176297.
4. Barrera Mosquera, V.H.; Delgado, J.A.; Alwang, J.R.; Escudero López, L.O.; Cartagena Ayala, Y.E.; Domínguez Andrade, J.M.; D'adamo, R. Conservation Agriculture Increases Yields and Economic Returns of Potato, Forage, and Grain Systems of the Andes. *Agron. J.* **2019**, *111*, 2747–2753. https://doi.org/10.2134/AGRONJ2019.04.0280.
5. Glaser, B. Prehistorically Modified Soils of Central Amazonia: A Model for Sustainable Agriculture in the Twenty-First Century. *Philos. Trans. R. Soc. B Biol. Sci.* **2006**, *362*, 187–196. https://doi.org/10.1098/RSTB.2006.1978.
6. Barra, I.; Haefele, S.M.; Sakrabani, R.; Kebede, F. Soil Spectroscopy with the Use of Chemometrics, Machine Learning and Pre-Processing Techniques in Soil Diagnosis: Recent Advances–A Review. *TrAC Trends Anal. Chem.* **2021**, *135*, 116166. https://doi.org/10.1016/J.TRAC.2020.116166.
7. Lucas, Y.; Santin, R.C.; Silva, W.T.L. da; Merdy, P.; Melfi, A.J.; Pereira, O.J.R.; Montes, C.R. Soil Sample Conservation from Field to Lab for Heterotrophic Respiration Assessment. *Methods X* **2020**, *7*, 101039. https://doi.org/10.1016/J.MEX.2020.101039.
8. Ahmadi, A.; Emami, M.; Daccache, A.; He, L. Soil Properties Prediction for Precision Agriculture Using Visible and Near-Infrared Spectroscopy: A Systematic Review and Meta-Analysis. *Agronomy* **2021**, *11*, 433.
9. Demattê, J.A.M.; da Silveira Paiva, A.F.; Poppiel, R.R.; Rosin, N.A.; Chimelo Ruiz, L.F.; Alcantara de Oliveira Mello, F.; Minasny, B.; Grunwald, S.; Ge, Y.; Ben Dor, E.; et al. The Brazilian Soil Spectral Service (BraSpecS): A User-Friendly System for Global Soil Spectra Communication. *Remote Sens.* **2022**, *14*, 740. https://doi.org/10.3390/RS14030740/S1.
10. Viscarra Rossel, R.A.; Behrens, T.; Ben Dor, E.; Chabrillat, S.; Demattê, J.A.M.; Ge, Y.; Gomez, C.; Guerrero, C.; Peng, Y.; Ramirez-Lopez, L.; et al. Diffuse Reflectance Spectroscopy for Estimating Soil Properties: A Technology for the 21st Century. *Eur. J. Soil Sci.* **2022**, *73*, e13271. https://doi.org/10.1111/EJSS.13271.
11. Piccini, C.; Metzger, K.; Debaene, G.; Stenberg, B.; Götzinger, S.; Borůvka, L.; Sandén, T.; Bragazza, L.; Liebisch, F. In-Field Soil Spectroscopy in Vis–NIR Range for Fast and Reliable Soil Analysis: A Review. *Eur. J. Soil Sci.* **2024**, *75*, e13481. https://doi.org/10.1111/EJSS.13481.

12. Ramirez-Lopez, L.; Wadoux, A.M.J.C.; Franceschini, M.H.D.; Terra, F.S.; Marques, K.P.P.; Sayão, V.M.; Demattê, J.A.M. Robust Soil Mapping at the Farm Scale with Vis–NIR Spectroscopy. *Eur. J. Soil Sci.* **2019**, *70*, 378–393. https://doi.org/10.1111/ejss.12752.
13. Rodríguez-Pérez, J.R.; Marcelo, V.; Pereira-Obaya, D.; García-Fernández, M.; Sanz-Ablanedo, E. Estimating Soil Properties and Nutrients by Visible and Infrared Diffuse Reflectance Spectroscopy to Characterize Vineyards. *Agronomy* **2021**, *11*, 1895. https://doi.org/10.3390/agronomy11101895.
14. Wijewardane, N.K.; Hetrick, S.; Ackerson, J.; Morgan, C.L.S.; Ge, Y. VisNIR Integrated Multi-Sensing Penetrometer for in Situ High-Resolution Vertical Soil Sensing. *Soil Tillage Res.* **2020**, *199*, 104604. https://doi.org/10.1016/J.STILL.2020.104604.
15. Najdenko, E.; Lorenz, F.; Dittert, K.; Olfs, H.W. Rapid In-Field Soil Analysis of Plant-Available Nutrients and PH for Precision Agriculture—A Review. *Precis. Agric.* **2024**, *25*, 3189–3218. https://doi.org/10.1007/S11119-024-10181-6.
16. Salazar, O.; Benvenuto, A.; Fajardo, M.; Fuentes, J.P.; Nájera, F.; Celedón, A.; Pfeiffer, M.; Renwick, L.L.R.; Seguel, O.; Tapia, Y.; et al. Evaluation of a Miniaturized Portable NIR Spectrometer for the Prediction of Soil Properties in Mediterranean Central Chile. *Geoderma Reg.* **2023**, *34*, e00675. https://doi.org/10.1016/J.GEODRS.2023.E00675.
17. Ng, W.; Husnain; Anggria, L.; Siregar, A.F.; Hartatik, W.; Sulaeman, Y.; Jones, E.; Minasny, B. Developing a Soil Spectral Library Using a Low-Cost NIR Spectrometer for Precision Fertilization in Indonesia. *Geoderma Reg.* **2020**, *22*, e00319. https://doi.org/10.1016/J.GEODRS.2020.E00319.
18. de Souza, M.F.; Franco, H.C.J.; Do Amaral, L.R. Estimation of Soil Phosphorus Availability via Visible and Near-Infrared Spectroscopy. *Sci. Agric.* **2019**, *77*, e20180295. https://doi.org/10.1590/1678-992X-2018-0295.
19. Ma, Y.; Minasny, B.; Demattê, J.A.M.; McBratney, A.B. Incorporating Soil Knowledge into Machine-Learning Prediction of Soil Properties from Soil Spectra. *Eur. J. Soil Sci.* **2023**, *74*, e13438. https://doi.org/10.1111/EJSS.13438.
20. Tepanosyan, G.; Muradyan, V.; Tepanosyan, G.; Avetisyan, R.; Asmaryan, S.; Sahakyan, L.; Denk, M.; Gläßer, C. Exploring Relationship of Soil PTE Geochemical and "VIS-NIR Spectroscopy" Patterns near Cu–Mo Mine (Armenia). *Environ. Pollut.* **2023**, *323*, 121180. https://doi.org/10.1016/J.ENVPOL.2023.121180.
21. Wang, X.; Zhang, M.W.; Guo, Q.; Yang, H.L.; Wang, H.L.; Sun, X.L. Estimation of Soil Organic Matter by in Situ Vis-NIR Spectroscopy Using an Automatically Optimized Hybrid Model of Convolutional Neural Network and Long Short-Term Memory Network. *Comput. Electron. Agric.* **2023**, *214*, 108350. https://doi.org/10.1016/J.COMPAG.2023.108350.
22. Cevoli, C.; Iaccheri, E.; Fabbri, A.; Ragni, L. Data Fusion of FT-NIR Spectroscopy and Vis/NIR Hyperspectral Imaging to Predict Quality Parameters of Yellow Flesh "Jintao" Kiwifruit. *Biosyst. Eng.* **2024**, *237*, 157–169. https://doi.org/10.1016/J.BIOSYSTEMSENG.2023.12.011.
23. Munnaf, M.A.; Mouazen, A.M. Spectra Transfer Based Learning for Predicting and Classifying Soil Texture with Short-Ranged Vis-NIRS Sensor. *Soil Tillage Res.* **2023**, *225*, 105545. https://doi.org/10.1016/J.STILL.2022.105545.
24. Guerrero, C.; Wetterlind, J.; Stenberg, B.; Mouazen, A.M.; Gabarrón-Galeote, M.A.; Ruiz-Sinoga, J.D.; Zornoza, R.; Viscarra Rossel, R.A. Do We Really Need Large Spectral Libraries for Local Scale SOC Assessment with NIR Spectroscopy? *Soil Tillage Res.* **2016**, *155*, 501–509. https://doi.org/10.1016/J.STILL.2015.07.008.
25. Díaz-Guadarrama, S.; Varón-Ramírez, V.M.; Lizarazo, I.; Guevara, M.; Angelini, M.; Araujo-Carrillo, G.A.; Argeñal, J.; Armas, D.; Balta, R.A.; Bolivar, A.; et al. Improving the Latin America and Caribbean Soil Information System (SISLAC) Database Enhances Its Usability and Scalability. *Earth Syst. Sci. Data* **2024**, *16*, 1229–1246. https://doi.org/10.5194/essd-16-1229-2024.
26. Recena, R.; Fernández-Cabanás, V.M.; Delgado, A. *Soil Fertility Assessment by Vis-NIR Spectroscopy: Predicting Soil Functioning Rather Than Availability Indices*; FAO: Rome, Italy, 2018.
27. Espinoza, R.; Molina, J.; Horn, M.J.; Gómez, M. Conceptos Bioclimáticos y su Aplicabilidad a la Zona Rural Altoandina: Caso Comunidad San Francisco de Raymina (Sfr)-Ayacucho. *Technia* **2015**, *25*, 5.
28. INEI. Territorio y Suelos. 2015. Available online: https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1342/cap01.pdf (accessed on 22 April 2026).
29. Arce, A.; de Haan, S.; Juarez, H.; Burra, D.D.; Plasencia, F.; Ccanto, R.; Polreich, S.; Scurrah, M. The Spatial-Temporal Dynamics of Potato Agrobiodiversity in the Highlands of Central Peru: A Case Study of Smallholder Management across Farming Landscapes. *Land* **2019**, *8*, 169. https://doi.org/10.3390/land8110169.
30. Ccopi, T.D.; Barzola, R.B.; Ruiz, S.S.; Gabriel, C.E.; Ortega, Q.K.; Cordova, B.F. River Flood Risk Assessment in Communities of the Peruvian Andes: A Semiquantitative Application for Disaster Prevention. *Sustainability* **2023**, *15*, 13768. https://doi.org/10.3390/su151813768.
31. INRENA. *Mapa de Suelos del Peru (1:5,000,000)*; Ministry of Agriculture, General Directorate of Water and Soils: Lima, Peru, 1996. Available online: https://es.scribd.com/document/422549685/Mapa-de-suelos-del-Peru (accessed on).

32. MINAGRI. Análisis de La Cadena Productiva del Cacao: Con Enfoque en Los Pequeños Productores de Limitado Acceso al Mercado, Gobierno del Perú. 2018. Available online: http://repositorio.minagri.gob.pe:80/jspui/handle/MINAGRI/66 (accessed on).
33. Walkley, A.; Black, I.A. An Examination of the Degtjareff Method for Determining Soil Organic Matter and a Proposed Modification of the Chromic Acid Titration Method. *Soil Sci.* **1934**, *37*, 29–38. https://doi.org/10.1097/00010694-193401000-00003.
34. Barra, I.; Briak, H.; Kebede, F. The Application of Statistical Preprocessing on Spectral Data Does Not Always Guarantee the Improvement of the Predictive Quality of Multivariate Models: Case of Soil Spectroscopy Applied to Moroccan Soils. *Vib. Spectrosc.* **2022**, *121*, 103409. https://doi.org/10.1016/J.VIBSPEC.2022.103409.
35. Ben-Dor, E.; Chabrillat, S.; Demattê, J.A.M.; Taylor, G.R.; Hill, J.; Whiting, M.L.; Sommer, S. Using Imaging Spectroscopy to Study Soil Properties. *Remote Sens. Environ.* **2009**, *113*, S38–S55. https://doi.org/10.1016/J.RSE.2008.09.019.
36. Lukas, M.; Lehnert, W. *Package "hsdar" Type Package Title Manage, Analyse and Simulate Hyperspectral Data*; CRAN: Vienna, Austria, 2017.
37. Stenberg, B.; Viscarra Rossel, R.A.; Mouazen, A.M.; Wetterlind, J. Visible and Near Infrared Spectroscopy in Soil Science. *Adv. Agron.* **2010**, *107*, 163–215. https://doi.org/10.1016/S0065-2113(10)07005-7.
38. Tian, J.; Philpot, W.D. Relationship between Surface Soil Water Content, Evaporation Rate, and Water Absorption Band Depths in SWIR Reflectance Spectra. *Remote Sens. Environ.* **2015**, *169*, 280–289. https://doi.org/10.1016/J.RSE.2015.08.007.
39. Araújo, M.C.U.; Saldanha, T.C.B.; Galvão, R.K.H.; Yoneyama, T.; Chame, H.C.; Visani, V. The Successive Projections Algorithm for Variable Selection in Spectroscopic Multicomponent Analysis. *Chemom. Intell. Lab. Syst.* **2001**, *57*, 65–73. https://doi.org/10.1016/S0169-7439(01)00119-8.
40. Wold, S.; Sjöström, M.; Eriksson, L. PLS-Regression: A Basic Tool of Chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130. https://doi.org/10.1016/S0169-7439(01)00155-1.
41. Viscarra Rossel, R.A.; Walvoort, D.J.J.; McBratney, A.B.; Janik, L.J.; Skjemstad, J.O. Visible, near Infrared, Mid Infrared or Combined Diffuse Reflectance Spectroscopy for Simultaneous Assessment of Various Soil Properties. *Geoderma* **2006**, *131*, 59–75. https://doi.org/10.1016/J.GEODERMA.2005.03.007.
42. Zhao, G.; Liu, M.; Shi, P.; Zong, N.; Wang, J.; Wu, J.; Zhang, X. Spatial-Temporal Variation of ANPP and Rain-Use Efficiency along a Precipitation Gradient on Changtang Plateau, Tibet. *Remote Sens.* **2019**, *11*, 325. https://doi.org/10.3390/rs11030325.
43. Chang, R.L.; Nithiyanantham, S.; Huang, C.Y.; Pai, P.Y.; Chang, T.T.; Hu, L.C.; Chen, R.J.; VijayaPadma, V.; Kuo, W.W.; Huang, C.Y. Synergistic Cardiac Pathological Hypertrophy Induced by High-Salt Diet in IGF-IIRα Cardiac-Specific Transgenic Rats. *PLoS ONE* **2019**, *14*, e0216285. https://doi.org/10.1371/journal.pone.0216285.
44. Yang, X.M.; Feng, Q.; Zhu, M. Vegetation Characteristics and Soil Properties of Artificially Remediated Grasslands: The Case Study of the Shimenhe Mining Area in Qilian Mountains, Northwest China. *Res. Cold Arid. Reg.* **2024**, *16*, 9. https://doi.org/10.1016/J.RCAR.2024.09.001.
45. Lyu, X.; Li, M.; Li, X.; Li, S.; Yan, C.; Ma, C.; Gong, Z. Assessing the Systematic Effects of the Concentration of Nitrogen Supplied to Dual-Root Systems of Soybean Plants on Nodulation and Nitrogen Fixation. *Agronomy* **2020**, *10*, 763. https://doi.org/10.3390/agronomy10060763.
46. Kostrzewski, M.; Melnik, R. Condition Monitoring of Rail Transport Systems: A Bibliometric Performance Analysis and Systematic Literature Review. *Sensors* **2021**, *21*, 4710.
47. Chai, T.; Draxler, R.R. Root Mean Square Error (RMSE) or Mean Absolute Error (MAE)? Arguments against Avoiding RMSE in the Literature. *Geosci. Model Dev.* **2014**, *7*, 1247–1250. https://doi.org/10.5194/gmd-7-1247-2014.
48. Chang, C.-W.; Laird, D.A.; Mausbach, M.J.; Hurburgh, C.R. Near-Infrared Reflectance Spectroscopy–Principal Components Regression Analyses of Soil Properties. *Soil Sci. Soc. Am. J.* **2001**, *65*, 480–490. https://doi.org/10.2136/sssaj2001.652480x.
49. Canero, F.M.; Rodriguez-galiano, V.; Aragones, D. Heliyon Machine Learning and Feature Selection for Soil Spectroscopy. An Evaluation of Random Forest Wrappers to Predict Soil Organic Matter, Clay, and Carbonates. *Heliyon* **2024**, *10*, e30228. https://doi.org/10.1016/j.heliyon.2024.e30228.
50. Kvalseth, T.O. Cautionary Note about R 2. *Am. Stat.* **1985**, *39*, 279. https://doi.org/10.2307/2683704.
51. Nuzzo, R. Percent Differences: Another Look. *PM&R* **2018**, *10*, 661–664. https://doi.org/10.1016/j.pmrj.2018.05.003.
52. Capetillo-Burela, A.; López-Collado, C.J.; Zetina-Lezama, R.; Reynolds-Chávez, M.A.; Matilde-Hernández, C.; Cadena-Zapata, M.; López-López, J.A. Modelo Conceptual de Fertilización Nitrogenada Para Maiz (*Zea mays* L.) En Veracruz, México. *Rev. Iberoam. Bioecon. Cambio Clim.* **2021**, *7*, 1617–1631. https://doi.org/10.5377/ribcc.v7i14.12606.
53. Burt, R. Soil Survey Laboratory Methods Manual. 2004. Available online: https://www.govinfo.gov/content/pkg/GOVPUB-A57-PURL-gpo93947/pdf/GOVPUB-A57-PURL-gpo93947.pdf (accessed on 9 September 2025) .

54. INIA INDIA. Disponibilidad de Semilla Dirección General de Proyección y Servicios Agrarios Unidad de Medios y Comunicación Técnica, Huancayo. 2004; Available online: https://www.inia.gob.pe/disponibilidad-de-semillas/ (accessed on).
55. MIDAGRI (Ministerio de Desarrollo Agrario y Riego). *Manual de Producción de Maíz Amiláceo*; Platforma del Estado Peruano: Lima, Peru, 2020.
56. Maiti, D.; Das, D.K.; Pathak, H. Simulation of Fertilizer Requirement for Irrigated Wheat in Eastern India Using the QUEFTS Model. *Sci. World J.* **2006**, *6*, 231–245. https://doi.org/10.1100/tsw.2006.43.
57. Rawls, W.J.; Nemes, A.; Pachepsky, Y. Effect of Soil Organic Carbon on Soil Hydraulic Properties. In *Development of Pedotransfer Functions in Soil Hydrology*; Elsevier: Amsterdam, The Netherlands, 2004; pp. 95–114.
58. Manrique, L.A.; Jones, C.A. Bulk Density of Soils in Relation to Soil Physical and Chemical Properties. *Soil Sci. Soc. Am. J.* **1991**, *55*, 476. https://doi.org/10.2136/sssaj1991.03615995005500020030x.
59. Bernoux, M.; Cerri, C.; Arrouays, D.; Jolivet, C.; Volkoff, B. Bulk Densities of Brazilian Amazon Soils Related to Other Soil Properties. *Soil Sci. Soc. Am. J.* **1998**, *62*, 743–749. https://doi.org/10.2136/sssaj1998.03615995006200030029x.
60. Raun, W.R.; Johnson, G.V. Improving Nitrogen Use Efficiency for Cereal Production. *Agron. J.* **1999**, *91*, 357–363. https://doi.org/10.2134/agronj1999.00021962009100030001x.
61. Smil, V. Nitrogen in Crop Production: An Account of Global Flows. *Glob. Biogeochem. Cycles* **1999**, *13*, 647–662. https://doi.org/10.1029/1999GB900015.
62. Rangaiah, K.M.; Nagaraju, B.; Shankaraiah, S.K.; Kasturappa, G.; Kadappa, B.P.; Sugatur Narayanaswamy, U.K.; Saqeebulla, M.; Dey, P. Enhancing Yield, Uptake and Nutrient Use Efficiency of Brinjal Through Soil Test Crop Response Approach. *Commun. Soil Sci. Plant Anal.* **2024**, *55*, 998–1014. https://doi.org/10.1080/00103624.2023.2289988.
63. FAO. *Una Introducción al Análisis de Suelo Utilizando Espectroscopía Visible e Infrarrojo Cercano (Vis-NIR) y Espectroscopía de Infrarrojo Medio (MIR)*; FAO: Rome, Italy, 2024.
64. USDA. *Soil Survey Manual Soil Science Division Staff Agriculture*; USDA: Washington, DC, USA, 2017; 18p.
65. Gallagher, N.B. *Savitzky-Golay Smoothing and Differentiation Filter*; Eigen Vector Inc.: Manson, WA, USA, 2020. https://doi.org/10.13140/RG.2.2.20339.50725.
66. Shepherd, K.D.; Walsh, M.G. Development of Reflectance Spectral Libraries for Characterization of Soil Properties. *Soil Sci. Soc. Am. J.* **2002**, *66*, 988–998. https://doi.org/10.2136/sssaj2002.9880.
67. Safanelli, J.L.; Hengl, T.; Parente, L.L.; Minarik, R.; Bloom, D.E.; Todd-Brown, K.; Gholizadeh, A.; de Sousa Mendes, W.; Sanderman, J. Open Soil Spectral Library (OSSL): Building Reproducible Soil Calibration Models through Open Development and Community Engagement. *PLoS ONE* **2025**, *20*, e0296545. https://doi.org/10.1371/journal.pone.0296545.
68. Terhoeven-Urselmans, T.; Vagen, T.-G.; Spaargaren, O.; Shepherd, K.D. Prediction of Soil Fertility Properties from a Globally Distributed Soil Mid-Infrared Spectral Library. *Soil Sci. Soc. Am. J.* **2010**, *74*, 1792–1799. https://doi.org/10.2136/sssaj2009.0218.
69. Tavares, T.R.; Molin, J.P.; Nunes, L.C.; Alves, E.E.N.; Krug, F.J.; de Carvalho, H.W.P. Spectral Data of Tropical Soils Using Dry-Chemistry Techniques (VNIR, XRF, and LIBS): A Dataset for Soil Fertility Prediction. *Data Brief* **2022**, *41*, 108004. https://doi.org/10.1016/j.dib.2022.108004.
70. Du, C.; Zhou, J. Evaluation of Soil Fertility Using Infrared Spectroscopy: A Review. *Environ. Chem. Lett.* **2009**, *7*, 97–113.
71. Clingensmith, C.M.; Grunwald, S. Predicting Soil Properties and Interpreting Vis-NIR Models from across Continental United States. **2022**, *22*, 3187.
72. Vestergaard, R.-J.; Vasava, H.B.; Aspinall, D.; Chen, S.; Adamchuk, V.; Biswas, A.; Evaluation of Optimized Preprocessing and Modeling Algorithms for Prediction of Soil Properties Using VIS-NIR Spectroscopy. *Sensors* **2021**, *21*, 20**.** https://doi.org/10.3390/s21206745.
73. Neale, P.A.; Escher, B.I. In Vitro Bioassays to Assess Drinking Water Quality. *Curr. Opin. Environ. Sci. Health* **2019**, *7*, 1–7. https://doi.org/10.1016/J.COESH.2018.06.006.
74. Mutanga, O.; Skidmore, A.K. Hyperspectral Band Depth Analysis for a Better Estimation of Grass Biomass (*Cenchrus ciliaris*) Measured under Controlled Laboratory Conditions. *Int. J. Appl. Earth Obs. Geoinf.* **2004**, *5*, 87–96. https://doi.org/10.1016/J.JAG.2004.01.001.
75. Murad, M.O.F.; Jones, E.J.; Minasny, B.; McBratney, A.B.; Wijewardane, N.; Ge, Y. Assessing a VisNIR Penetrometer System for In-Situ Estimation of Soil Organic Carbon under Variable Soil Moisture Conditions. *Biosyst. Eng.* **2022**, *224*, 197–212. https://doi.org/10.1016/J.BIOSYSTEMSENG.2022.10.011.
76. Liu, Y.; Shen, L.; Zhu, X.; Xie, Y.; He, S. Spectral Data-Driven Prediction of Soil Properties Using LSTM-CNN-Attention Model. *Appl. Sci.* **2024**, *14*, 11687. https://doi.org/10.3390/APP142411687.

77. Zhou, W.; Li, C.; Zhao, W.; Stringer, L.C.; Fu, B. Spatial Distributions of Soil Nutrients Affected by Land Use, Topography and Their Interactions, in the Loess Plateau of China. *Int. Soil Water Conserv. Res.* **2024**, *12*, 227–239. https://doi.org/10.1016/j.iswcr.2023.02.005.
78. Kawamura, K.; Tsujimoto, Y.; Nishigaki, T.; Andriamananjara, A.; Rabenarivo, M.; Asai, H.; Rakotoson, T.; Razafimbelo, T. Laboratory Visible and Near-Infrared Spectroscopy with Genetic Algorithm-Based Partial Least Squares Regression for Assessing the Soil Phosphorus Content of Upland and Lowland Rice Fields in Madagascar. *Remote Sens.* **2019**, *11*, 506. https://doi.org/10.3390/RS11050506.
79. Shin, S.K.; Lee, S.J.; Park, J.H. Prediction of Soil Properties Using Vis-NIR Spectroscopy Combined with Machine Learning: A Review. *Sensors* **2025**, *25*, 5045. https://doi.org/10.3390/s25165045.
80. Nawar, S.; Mouazen, A. Comparison between Random Forests, Artificial Neural Networks and Gradient Boosted Machines Methods of On-Line Vis-NIR Spectroscopy Measurements of Soil Total Nitrogen and Total Carbon. *Sensors* **2017**, *17*, 2428. https://doi.org/10.3390/s17102428.
81. Ge, Y.; Morgan, C.L.S.; Grunwald, S.; Brown, D.J.; Sarkhot, D.V. Comparison of Soil Reflectance Spectra and Calibration Models Obtained Using Multiple Spectrometers. *Geoderma* **2011**, *161*, 202–211. https://doi.org/10.1016/j.geoderma.2010.12.020.
82. Lei, T.; Sun, D.-W. Achieving Joint Calibration of Soil Vis-NIR Spectra across Instruments, Soil Types and Properties by an Attention-Based Spectra Encoding-Spectra/Property Decoding Architecture. *Geoderma* **2022**, *405*, 115449. https://doi.org/10.1016/j.geoderma.2021.115449.
83. Almeida Silva, F.H.C.; Wijewardane, N.K.; Cox, M.S.; Zhang, X. Assessment of Different VisNIR and MIR Spectroscopic Techniques and the Potential of Calibration Transfer between MIR Laboratory and Portable Instruments to Estimate Soil Properties. *Soil Tillage Res.* **2025**, *251*, 106555. https://doi.org/10.1016/j.still.2025.106555.
84. de Santana, F.B.; Otani, S.K.; de Souza, A.M.; Poppi, R.J. Comparison of PLS and SVM Models for Soil Organic Matter and Particle Size Using Vis-NIR Spectral Libraries. *Geoderma Reg.* **2021**, *27*, e00436. https://doi.org/10.1016/j.geodrs.2021.e00436.
85. Breure, T.S.; Webster, R.; Haefele, S.M.; Hannam, J.A.; Corstanje, R.; Milne, A.E. The Effect of Uncertainty in Predictions of Nutrient Concentrations from Soil Spectra on Variable-Rate Fertilizer Applications. *Geoderma* **2025**, *462*, 117504. https://doi.org/10.1016/j.geoderma.2025.117504.
86. Breure, T.S.; Haefele, S.M.; Hannam, J.A.; Corstanje, R.; Webster, R.; Moreno-Rojas, S.; Milne, A.E. A Loss Function to Evaluate Agricultural Decision-Making under Uncertainty: A Case Study of Soil Spectroscopy. *Precis. Agric.* **2022**, *23*, 1333–1353. https://doi.org/10.1007/s11119-022-09887-2.
87. Suárez-Rey, E.M.; Gallardo, M.; Romero-Gámez, M.; Giménez, C.; Rueda, F.J. Sensitivity and Uncertainty Analysis in Agro-Hydrological Modelling of Drip Fertigated Lettuce Crops under Mediterranean Conditions. *Comput. Electron. Agric.* **2019**, *162*, 630–650. https://doi.org/10.1016/j.compag.2019.05.011.
88. Metzger, K.; Liebisch, F.; Herrera, J.M.; Guillaume, T.; Walder, F.; Bragazza, L. The Use of Visible and Near-infrared Spectroscopy for In-situ Characterization of Agricultural Soil Fertility: A Proposition of Best Practice by Comparing Scanning Positions and Spectrometers. *Soil Use Manag.* **2024**, *40*, 12952. https://doi.org/10.1111/sum.12952.
89. Soriano-Disla, J.M.; Janik, L.J.; Viscarra Rossel, R.A.; Macdonald, L.M.; McLaughlin, M.J. The Performance of Visible, Near-, and Mid-Infrared Reflectance Spectroscopy for Prediction of Soil Physical, Chemical, and Biological Properties. *Appl. Spectrosc. Rev.* **2014**, *49*, 139–186. https://doi.org/10.1080/05704928.2013.811081.
90. Mokere, R.; Ghassan, M.; Barra, I. Soil Spectroscopy Improves Mid Infrared Soil Property Prediction through Optimized Preprocessing and Variable Selection. *Front. Soil Sci.* **2026**, *6*, 1760011. https://doi.org/10.3389/fsoil.2026.1760011.